# Pre-analysis plan: Group B Energy billing guideline research

## Policy problem

Across the energy market consumers experience confusion reading and understanding their energy bills. The current rules for what is required in energy bills are almost a decade old, and fail to reflect the changes to the market that have occurred during this period. This means that consumers often receive bills that are unnecessarily complex and struggle to comprehend the information. This creates problems for consumers trying to pay their bills and understand their energy use, as well as for retailers who have increasing costs from customer complaints.

## Trial aim

We are aiming to inform the development of new rules for energy bills, designed to improve customer understanding.

This research will involve three randomised trials run sequentially on the same sample via an online survey. Each trial will vary particular elements of bills and test these variations for comprehension. Specifically, we will test:

* the effect of including definitions on bills
* the impact of cheapest offer messaging and plan information boxes
* whether people understand historical usage, benchmark, and solar usage charts.

## Interventions

Below, we provide a summary of the interventions in each treatment arm of the three trials. See Interim report for images of each treatment arm.

### Trial 1 (Plan summaries, market engagement and definitions)

This will be a five arm trial with the following groups:

* Control (C) = Detailed charges table only
* Treatment 1 (T1) = Detailed charges table + Plan summary
* Treatment 2 (T2) = Detailed charges table + Could you save money?
* Treatment 3 (T3) = Detailed charges table + Plan summary + Could you save money?
* Treatment 4 (T4) = Detailed charges table + Plan summary + Could you save money? + Definitions

### Trial 2 (Benchmarks)

This will be a five arm trial with the following groups:

* Control (C) = usage chart only
* Treatment 1 (T1) = usage chart + benchmark table
* Treatment 2 (T2) = usage chart + benchmark vertical bar graph
* Treatment 3 (T3) = usage chart + benchmark infographic
* Treatment 4 (T4) = usage chart + benchmark simple infographic

### Trial 3 (Energy consumption and generation charts and definitions)

This will be a 5x2 factorial design. Our first independent variable (A) has five levels and will vary energy consumption and generation charts. Our second independent variable has two levels and will test the impact of providing additional definitions for technical terms. The table below summarises the intervention associated with each factor and defines the individual cells formed by each independent variable.

**Table 1** Trial 3 factorial design

| Chart types | B0 = Without definitions | B1 =With definitions |
| --- | --- | --- |
| A0 = Complex consumption chart, solar generation table | A0B0 | A0B1 |
| A1 = Simple consumption column chart, solar generation table | A1B0 | A1B1 |
| A2 = Two column charts | A2B0 | A2B1 |
| A3 = Combined bar chart | A3B0 | A3B1 |
| A4 = Combined line chart | A4B0 | A4B1 |

## Outcome measures

For all questions in the survey we will allow people to skip answering. A skipped response will be coded as 0 (see ‘Trial Threats’ for more discussion). See Appendix A for actual questions and response options.

Box 1: Outcome measures for Trial 1 (Plan summaries, market engagement and definitions)

Primary outcomes

* Comprehension - Able to understand your contract. Number of correct answers (0-4).
* Intention - Advises to switch plans or compare plans. Free text coded as binary.
* Intention - Advises using Energy made easy. Binary.
* Intention - Advises contacting own retailer. Binary.
* Comprehension - Able to identify cheapest plan. Score (0-3).

Secondary outcomes

* Time taken - Able to understand your contract.
* Confidence to choose a better plan. Binary (Very confident or confident = 1, all other responses = 0).
* Bill is easy to understand. Binary (very or fairly easy = 1, all other responses = 0).
* Would value having plan summary on their bill. Binary (any agree = 1, all other responses = 0).
* Would value having information about other plans on their bill. Binary (any agree = 1, all other responses = 0).

Box 2: Outcome measures for Trial 2 (Benchmarks)

Primary outcomes

* Comprehension - Able to understand how they compare to benchmark. Binary.
* Intention - Advises to save energy. Free text coded as binary.

Secondary outcomes

* Time taken - Able to respond to comprehension questions.
* Comprehension - Able to understand that benchmark measures usage, not price. Binary.
* Confidence to find a cost-saving strategy. Binary (Very confident or confident = 1, all other responses = 0).
* Bill is easy to understand. Binary (very or fairly easy = 1, all other responses = 0).
* Agrees that benchmarks help their household choose how much energy to use. Binary (any agree = 1, all other responses = 0).
* Would value having benchmark on their bill. Binary (any agree = 1, all other responses = 0).

Box 3: Outcome measures for Trial 3 (Energy consumption and generation charts, and definitions)

Primary outcomes

* Comprehension - Able to understand usage chart. Number of correct answers (0-4).
* Comprehension - Able to understand solar generation chart. Number of correct answers (0-4).
* Intention - Advises to use solar more efficiently. Free text coded as binary.

Secondary outcomes

* Time taken - Able to respond to understand usage chart.
* Time taken - Able to understand solar generation chart.
* Confidence to find a cost-saving strategy. Binary (Very confident or confident = 1, all other responses = 0).
* Bill is easy to understand. Binary (very or fairly easy = 1, all other responses = 0).
* Would value having solar information on their bill. Binary (any agree = 1, all other responses = 0).

For the ‘free text coded as binary’ outcomes, we will manually classify free text responses to a binary variable where 1 equals “advises switching or comparing plans / advises saving energy/advises using solar more efficiently/” for trials 1/2/3 respectively. To reduce the chances of any bias in the classification process a staff member not involved in analysis will conduct the classification using a dataset that contains only the text field and a de-identified ID code, this staff member will be blind to treatment allocation.

## Population and sample selection

Our population of interest is adults (18 and over) residing in areas covered by the National Electricity Market who are also covered by the National Energy Consumer Framework. This excludes people residing in WA, NT, and VIC. We will use an online survey panel provider to recruit participants from this target population.

While the survey panel provider will be able to restrict invites to members who meet our requirements, due to lags in updating profiles there is still a chance that we could recruit participants who live in WA, NT, or VIC. To avoid this we will also screen out any responses to our demographics questions who indicate they currently live in any of these states.

The overall sample will be 7,500 individuals who meet the above requirements. We will also use quotas for gender to ensure the sample is broadly representative of the Australian population on this dimension.

## Hypotheses

### Trial 1 (Plan summary, encouragement to switch & definitions)

#### Plan summaries

H1a: **Plan summaries will improve plan comprehension**

 Bills with a plan summary (T1 and T3 pooled) will result in higher plan comprehension than bills without a plan summary (C and T2 pooled)

T1 and T3 pooled > C and T2 pooled

H1b: **Plan summaries will result in choosing a cheaper plan**

 Bills with a plan summary (T1 and T3 pooled) will result in choosing a cheaper plan than bills without a plan summary (C and T2 pooled)

T1 and T3 pooled > C and T2 pooled

Both of these hypotheses will be assessed with a one-tailed hypothesis test. We will correct for the two multiple comparisons that comprise this family of tests.

#### Plain language definitions

H2a: **Plain language definitions will improve plan comprehension**

 Bill with definitions (T4) will result in higher plan comprehension than equivalent bill without definitions (T3).

T4>T3

H2b: **Plain language definitions will result in choosing a cheaper plan**

Bill with definitions (T4) will result in choosing a cheaper plan than the equivalent bill without definitions (T3)

T4>T3

Both of these hypotheses will be assessed with a one-tailed hypothesis test. We will correct for the two multiple comparisons that comprise this family of tests.

#### Encouragement to switch

H3a: **Encouragement to choose cheaper plan will increase switching intention**

 Bill with encouragement (T2 and T3 and T4 pooled) will result in a higher switching intention than those without encouragement (C and T1 pooled)

T2 and T3 and T4 pooled > C and T1 pooled

H3b: **Encouragement to choose cheaper plan will increase recommendation to use a government comparison website to find a better deal**

Bill with encouragement (T2 and T3 and T4 pooled) will result in a higher proportion recommending a government comparison website than those without encouragement (C and T1 pooled)

 T2 and T3 and T4 pooled > C and T1 pooled

H3c: **Encouragement to choose cheaper plan will increase recommendation to contact own retailer for a better deal.**

Bill with encouragement (T2 and T3 and T4 pooled) will result in a higher proportion recommending contacting your own retailer than those without encouragement (C and T1 pooled)

 T2 and T3 and T4 pooled > C and T1 pooled

We will assess these hypotheses with a one-tailed hypothesis test, using a bonferroni correction for the three comparisons that comprise this family of tests.

### Trial 2 (Benchmark charts)

H1: **Viewing benchmarks will result in greater understanding of how individual electricity usage compares to average usage.**

 Any bill showing benchmark data (T1, T2, T3, T4, not pooled) will result in greater understanding than the control condition (C)

T1>C

T2>C

T3>C

T4>C

H2: **Graphical presentation of a benchmark (rather than a table) will result in greater understanding of how individual electricity usage compares to average usage.**

 Any bill showing benchmark data as a chart or infographic (T2-T3-T4, pooled) will result in greater understanding than the bill which shows benchmark data presented as a table (T1)

 T2 and T3 and T4 pooled >T1

H3: **Viewing benchmarks will result in higher energy-saving intention**

 Any bill showing benchmark data (T1, T2, T3, T4, not pooled) will result in than the control condition (C)

T1>C

T2>C

T3>C

T4>C

The three hypotheses in this trial will be assessed using one-tailed hypothesis tests. We will use a bonferroni adjustment to correct for the three main comparisons that comprise this family of tests. We do not correct for the comparison of multiple arms against a shared control group due to correlation between comparisons.

If we reject the null for hypothesis 1 for multiple treatment arms we will assess the overall best performing group based on effect size without a corresponding statistical test. We expect such differences to be small and a formal test will lack power. There is little risk in a false positive when selecting the best performing group given that superiority over control has been demonstrated.

### Trial 3 (Energy consumption and generation charts, and definitions)

#### Energy usage patterns

H1: **Simple energy consumption charts will improve understanding of energy usage patterns over time compared to a more complex chart**

 Any simple energy usage chart (A1, A2, A3, A4, not pooled) will result in higher comprehension than a complex chart (A0).

A1>A0

A2>A0

A3>A0

A4>A0

This hypotheses will be assessed with a series of one sided tests. We will not correct for multiple comparisons due to the shared control group.

#### Solar generation

H2a: **Solar generation charts will improve understanding of solar energy generation patterns over time compared to a table**

 Any bill with a solar chart (A2, A3, A4, not pooled) will result in higher comprehension than a table (A0 and A1 pooled).

 A2> A0 and A1 pooled

 A3> A0 and A1 pooled

 A4> A0 and A1 pooled

H2b: **Solar generation charts will result in higher intention to use solar more efficiently than a table**

 Any bill with a solar chart (T2, T3, T4, not pooled) will result in a higher proportion that advises using solar more efficiently than those seeing a solar table (T0 and T1 pooled).

 A2> A0 and A1 pooled

 A3> A0 and A1 pooled

 A4> A0 and A1 pooled

Both of these hypotheses will be assessed with a series one-tailed hypothesis test. We will correct for two multiple comparisons for this family of tests but will not correct for the multiple comparisons against a shared control group.

#### Plain language definitions

H3a: **Plain language definitions will improve understanding of energy usage patterns over time**

Any bill with definitions (B1) will result in higher comprehension than a bill without definitions (B0)

 B1>B0

H3b: **Plain language definitions will improve understanding of solar energy generation patterns over time.**

Any bill with definitions (B1) will result in higher comprehension than a bill without definitions (B0)

 B1>B0

H3c: **Plain language definitions will result in higher intention to use solar more efficiently**

 Any bill with definitions (B) will result in a higher proportion that advises using solar more efficiently than a bill without definitions (A).

 B1>B0

We will assess these hypotheses with a one-tailed hypothesis test, using a bonferroni correction for the three comparisons that comprise this family of tests.

## Randomisation

Randomisation will occur within the Qualtrics survey platform. After demographics are collected, all respondents are individually assigned a random number from 0 to 5 (with a 0.2 probability of assignment) for each trial indicating treatment arm, and an additional number from 0 to 1 (with a 0.5 probability of assignment) indicating assignment to arms A or B in trial 4 (presence of definitions). Sample size may not be perfectly balanced between groups.

The order in which participants undertake the trials will also be randomised to allow averaging over any order effects. There are 6 possible orders that participants can experience, as below. This will be implemented using Qualtrics “Randomly present elements” (in which each element is a block containing one trial), specifying that they must evenly present three of three elements.

**Table 2** Trial order

| A | B | C | D | E | F |
| --- | --- | --- | --- | --- | --- |
| Trial 1 | Trial 1 | Trial 2 | Trial 2 | Trial 3 | Trial 3 |
| Trial 2 | Trial 3 | Trial 1 | Trial 3 | Trial 1 | Trial 2 |
| Trial 3 | Trial 2 | Trial 3 | Trial 1 | Trial 2 | Trial 1 |

## Sample size and power calculations

We performed power calculations using a standard alpha of 5%, and a standard power of 80% for a one tail test.

Approximate sample size will be 1500 participants per arm. We estimate that for continuous outcomes we will be able to detect a standardised effect of approximately 0.1 SD unit.

For our hypotheses that compare proportions we present a minimum detectable effect based on a conservative assumption of a 50% baseline. With a sample size of 1500 we will be powered to detect 4.55pp increase over the baseline.

## Method of analysis

The principal analysis of the effect of the interventions will consist of a covariate-adjusted comparison of our primary outcomes. This estimate, confidence intervals and p-values will be derived from an ordinary least squares regression model using robust (HC2) standard errors and with the following specification (for trial 1 & 2):

$$Y\_{i}=β\_{0}+ β\_{1}Z\_{i}+ β\_{2}X\_{i}+ β\_{3}Z\_{i}X\_{i}+ \in \_{i}$$

Where $i$ is an index for each individual in the trial, $Y$ is the primary outcome in question, $β\_{0}$ is the intercept, $Z$ is a vector of four treatment assignment indicators , $β\_{1}$ is a vector of coefficients representing the average treatment effect (relative to the control group), $X$ is a vector of mean‑centred trial order indicators to account for the randomised trial order, and $ZX$ is the interaction of the treatment indicator vector with the mean centred trial order indicator vector[[1]](#footnote-1) and $\in $ is the error term.

For trial 3 (a factorial design) the specification is as follows:

$$Y\_{i}=β\_{0}+ β\_{1}A\_{i}+β\_{2}B\_{i}+β\_{3}X\_{i}+β\_{4}A\_{i}X\_{i } + β\_{5}B\_{i}X\_{i }+ \in \_{i}$$

Where $i$ is an index for each individual enrolled in the trial, $Y$ is the outcome, $X$ is a vector of mean‑centred trial order indicators to account for the randomised trial order, and $β\_{0}$ is the intercept. The coefficient on $A$ is a vector of 4 main effects of factor A (varying the presentation of bills) and the coefficient on $B$ is the main effect of factor B (including definitions), and $\in $ is the error term. Both $A$ and $B$ will be interacted with $X$ the treatment order indicator.

For the factorial design, we do not expect interactions between our independent variables and our design is not powered to detect them. However, we will estimate and report interactions as an exploratory analysis.

We will conduct separate analysis on the sub-groups who did each trial as their first trial. This subgroup will not be impacted by order effects for that trial. We will compare the results for these ‘first trial’ subgroups to the results for ‘second trial’ and ‘third trial’ subgroups to assess whether the order of trials taken by participants impacts their responses.

## Trial threats

Attrition related to treatment status is plausible in this trial. Some interventions presented will be harder to comprehend than others. If difficulty understanding a given intervention results in attrition (i.e., if people leave the survey because it is too difficult) then this could lead to bias in our estimates.

We will include a ‘don’t know’ option for participants to use when they are not confident in the answer. We will include anybody who was randomised into a trial in the analysis and record any unanswered questions as zero.

We will assess attrition, questions skips and ‘don’t know’ responses to see if there is suggestive evidence that these are related to assignment. We will take the results of this robustness check into account when interpreting and reporting our findings.

## Interpretation of results

Although we will use p-values 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. Where primary outcomes are inconclusive, we will look at effect sizes, secondary outcomes and subgroup analyses to determine whether there are grounds for recommending any particular treatments, either because they can be comprehended more quickly, perform better for vulnerable groups (older Australians, lower education levels, experiencing financial hardship) or are more likely to be valued, preferred or rated as ‘easy to understand’ in comparison to alternatives.

## 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.’

## Appendix A- Actual outcome measures

### Trial 1 (Plan summaries, market engagement and definitions)

Primary outcomes

**Comprehension - Able to understand your contract**

Number of correct answers (0-4)

1. "How much money did my solar panels save me this bill?"
	1. Less than $75
	2. **More than $75**
	3. It doesn’t say
	4. I’m not sure
2. "I get a 15% discount, so it costs less than $1 a day just to stay connected to the grid. Is that right?"
	1. Yes
	2. **No**
	3. It doesn’t say
	4. I’m not sure
3. "I'm thinking about switching to a plan that charges a flat rate of 18 cents per kilowatt-hour. All the other costs and discounts are the same. Do you think that would that save me money?"
	1. **Yes**
	2. No
	3. It doesn’t say
	4. I’m not sure
4. "Could I save money by running my dishwasher at midnight instead of at 8pm?"
	1. **Yes**
	2. No
	3. It doesn’t say
	4. I’m not sure

**Intention - Advises to switch plans or compare plans**

Free text coded as 1 if it mentions switching or comparing plans, other answer or no answer coded as 0.

"What would you do to save some money on electricity, if you were in my position?"

* 1. I suggest… <Free text entry
	2. I wouldn’t know what to do

**Intention - Advises using Energy made easy**

Binary

"I'd like to try to find a cheaper plan. What should I try first?"

* 1. Contact my electricity company and request a cheaper plan
	2. **Visit a government comparison website**
	3. Call a few different electricity companies or check out their websites
	4. Visit a commercial comparison website
	5. I don’t know

**Intention - Advises contacting own retailer**

Binary

"I'd like to try to find a cheaper plan. What should I try first?"

* 1. **Contact my electricity company and request a cheaper plan**
	2. Visit a government comparison website
	3. Call a few different electricity companies or check out their websites
	4. Visit a commercial comparison website
	5. I don’t know

**Comprehension - Able to identify cheapest plan**

Score (0-3)

Respondent is shown the Trial 2 intervention plus two screenshots (below).





"Which of these three plans do you think will work out cheapest for me?"

* 1. My current plan (Simple Saver at EnergyCo) (Score=1)
	2. Dynamic Energy flat rate plan (Score=2)
	3. Verve Energy Ultra low rate plan (Score=3)
	4. I don’t feel confident to say (Score=1)

**Secondary outcomes**

**Time taken - Able to understand your contract**

Continuous (time stamps)

**Confidence to choose a better plan**

Binary

"How confident do you feel about this advice?"

1. Very confident =1
2. Confident =1
3. Not very confident =0
4. Not at all confident =0

**Bill is easy to understand**

Binary

To understand William's bill was...

1. Very easy =1
2. Fairly easy =1
3. Okay =0
4. A bit difficult =0
5. Very difficult =0

**Would value having plan summary on their bill**

Binary

I would value having this plan information on my bill.

1. Strongly agree =1
2. Moderately agree =1
3. Slightly agree =1
4. Neutral =0
5. Slightly disagree =0
6. Moderately disagree =0
7. Strongly disagree =0

**Would value having information about other plans on their bill**

Binary

I would value having this information about other plans on the market on my bill.

1. Strongly agree =1
2. Moderately agree =1
3. Slightly agree =1
4. Neutral =0
5. Slightly disagree =0
6. Moderately disagree =0

### Trial 2 (Benchmarks)

**Primary outcomes**

**Comprehension - Able to understand how they compare to benchmark**

Binary

"For the month of May, was my electricity usage about average?"

1. **Higher than other people**
2. Yes, an average amount
3. More efficient than other people
4. I couldn’t say

**Intention - Advises to save energy**

Free text coded as 1 if it mentions saving energy or ways of saving energy, other answer or no answer coded as 0.

"What would you do to save some money on electricity, if you were in my position?"

1. I suggest… <Free text entry
2. I wouldn’t know what to do

**Secondary outcomes**

**Time taken – Able to respond to comprehension questions**

Continuous (time stamp)

**Comprehension - Able to understand that benchmark measures usage, not price**

Binary

"I pay quite a bit more than my neighbour in the apartment next to mine. Why do you think this is?"

* 1. Plan is expensive
	2. **Electricity usage is high**
	3. May have been overcharged
	4. I couldn’t say

**Confidence to find an energy saving strategy**

Binary

"How confident do you feel about this advice?"

1. Very confident =1
2. Confident =1
3. Not very confident =0
4. Not at all confident =0

**Bill is easy to understand**

Binary

To understand Ana's bill was...

1. Very easy =1
2. Fairly easy =1
3. Okay =0
4. A bit difficult =0
5. Very difficult =0

**Agrees that benchmarks help their household choose how much energy to use**

Binary

This comparison with other households helps my household make a choice about how much electricity to use.

1. Strongly agree =1
2. Moderately agree =1
3. Slightly agree =1
4. Neutral =0
5. Slightly disagree =0
6. Moderately disagree =0
7. Strongly disagree =0

**Would value having benchmark on their bill**

Binary

I would value having this comparison on my bill.

1. Strongly agree =1
2. Moderately agree =1
3. Slightly agree =1
4. Neutral =0
5. Slightly disagree =0
6. Moderately disagree =0
7. Strongly disagree =0

### Trial 3 (Energy consumption and generation charts, and definitions)

**Primary outcomes**

**Comprehension - Able to understand usage chart**

Number of correct answers (0-4)

1. "According to this chart, what happened from February to March?"
	1. Electricity prices went up
	2. **Electricity usage went up**
	3. Electricity prices went down
	4. Electricity usage went down
	5. It doesn’t say
	6. I’m not sure
2. "Is my usage highest in January when the kids turn on the air-conditioner?"
	1. Yes
	2. **No**
	3. It doesn’t say
	4. I’m not sure
3. "I was working from home in May this year. How much electricity from the grid did I use?"
	1. **Around 23 kWh a day**
	2. Around 8 kWh a day
	3. Around 28 kWh a day
	4. It doesn’t say
	5. I’m not sure
4. "Did my electricity usage go up compared to the same time last year?"
	1. Yes, it went up
	2. **No, it came down**
	3. It’s virtually the same
	4. It doesn’t say
	5. I’m not sure

**Comprehension - Able to understand solar generation chart**

Number of correct answers (0-4)

1. "Why did I earn more money from my solar in January?"
	1. The price was higher
	2. **I sold more solar electricity to the grid**
	3. I used less electricity
	4. It doesn’t say
	5. I’m not sure
2. "How much electricity do you think my solar panels generated in May?"
	1. Probably less than 10 kWh a day
	2. **Probably more than 10 kWh a day**
	3. It doesn’t say
	4. I’m not sure
3. "Should I expect my solar exports to be lower this June than they were in May?"
	1. **Yes this is likely**
	2. No, this is unlikely
	3. It doesn’t say
	4. I’m not sure
4. "At any point in the year, did I sell more electricity than I bought?"
	1. Yes, for about half the year
	2. **Yes, from November to February**
	3. No
	4. It doesn’t say
	5. I’m not sure

**Intention - Advises to use solar more efficiently**

Free text coded as 1 if it mentions using solar more efficiently, broadly defined, such as by shifting more energy usage to daytime or storing solar in a battery, other answer or no answer coded as 0.

"What would you do to save some money on electricity, if you were in my position?"

1. I suggest… <Free text entry>
2. I wouldn’t know what to do

**Secondary outcomes**

**Time taken – Able to respond to understand usage chart**

Continuous (time stamp)

**Time taken – Able to respond to understand solar generation**

Continuous (time stamp)

**Confidence to find a cost saving strategy**

Binary

"How confident do you feel about this advice?"

1. Very confident =1
2. Confident =1
3. Not very confident =0
4. Not at all confident =0

**Bill is easy to understand**

Binary

To understand Isaac's bill was...

1. Very easy =1
2. Fairly easy =1
3. Okay =0
4. A bit difficult =0
5. Very difficult =0

**Would value having solar information on their bill**

Binary

I would value having this comparison on my bill.

1. Strongly agree =1
2. Moderately agree =1
3. Slightly agree =1
4. Neutral =0
5. Slightly disagree =0
6. Moderately disagree =0
7. Strongly disagree =0
1. Lin, Winston. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman’s critique." Annals of Applied Statistics 7, no. 1 (2013): 295-318. [↑](#footnote-ref-1)