Your Title

YOUR NAME

2025-05-20

Abstract

Background: (Brief few sentences) **Objectives**: 1. Estimate the causal effect of YOUR EXPOSURE on YOUR OUTCOMES measured one year later. 2. Evaluate whether these effects vary across the population. 3. Provide policy guidance on which individuals might benefit most. **Method**: We conducted a three-wave retrospective cohort study (waves XX-XXX, October XXXX–October XXXX) using data from the New Zealand Attitudes and Values Study, a nationally representative panel. Participants were eligible if they participated in the NZAVS in the baseline wave (XXXX, were under the age of 62, and were employed > 20 hours per week. We defined the exposure as (XXXX > NUMBER on a 1-7 Likert Scale (1 = yes, 0 = no)). To address attrition, we applied inverse probability of censoring weights; to improve external validity, we applied weighted to the population distribution of Age, Ethnicity, and Gender. We computed expected mean outcomes for the population in each exposure condition (high XXXX/low XXXXX). Under standard causal assumptions of unconfoundedness, the contrast provides an unbiased average treatment effect. We then used causal forests to detect heterogeneity in these effects and employed policy tree algorithms to identify individuals ("strong responders") likely to experience the greatest benefits. **Results**: Increasing XXXXX leads to XXXXX. Heterogeneous responses to (e.g. *Forgiveness*, *Personal Well-Being*, and *Life-Satisfaction*…) reveal structural variability in subpopulations… **Implications**: (Brief few sentences) **Keywords**: *Causal Inference*; *Cross-validation*; *Distress*; *Employment*; *Longitudinal*; *Machine sLearning*; *Religion*; *Semi-parametric*; *Targeted Learning*.

Introduction

The fundamental question in social psychology is whether political identification leads to conspiracies about political out-groups (Allen et al., 2000; Editorial, 2025).

Previous research falls into two categories, experimental and observational.

Experimetaea

Three challenges remains: First, to identify causality Second, to characterise responses in a national population, not a student sample. Third, to evaluate a "phenotype" of response across multiple measuers of conspiracy....

To address these challenges we apply modern causal inference methods to a (simulated) national dataset that repeatedly measures the same individuals over time. ...

Method

Sample

Data were collected as part of the New Zealand Attitudes and Values Study (NZAVS), an annual longitudinal national probability panel assessing New Zealand residents' social attitudes, personality, ideology, and health outcomes. The panel began in 2009 and has since expanded to include over fifty researchers, with responses from 40,000 participants to date. The study operates independently of political or corporate funding and is based at a university. It employs prize draws to incentivise participation. The NZAVS tends to slightly under-sample males and individuals of Asian descent and to over-sample females and Māori (the Indigenous people of New Zealand). To enhance the representativeness of our sample population estimates for the target population of New Zealand, we apply census-based survey weights that adjust for age, gender, and ethnicity (New Zealand European, Asian, Māori, Pacific) (Sibley, 2021). For more information about the NZAVS, visit: OSF.IO/75SNB. Note, the data used in this study were simulated from the NZAVS for the purposes of a student assessment and therefore are not valid for the general population of New Zealand. They are for illustrative purposes only.

Target Population

The target population for this study comprises New Zealand residents as represented in the 2018 of the New Zealand Attitudes and Values Study (NZAVS) during the years 2018 weighted by New Zealand Census weights for age, gender, and ethnicity (refer to Sibley (2021)). The NZAVS is a national probability study designed to reflect the broader New Zealand population accurately. Despite its comprehensive scope, the NZAVS has some limitations in its demographic representation. Notably, it tends to under-sample males and individuals of Asian descent while over-sampling females and Māori (the indigenous peoples of New Zealand). To address these disparities and enhance the accuracy of our findings, we apply New Zealand Census survey weights to the sample data (again, caveates about simulation noted.)

Eligibility Criteria

To be included in the analysis of this study, participants needed to participate in the 2018 of the study and respond to the baseline measure of Extraversion.

Participants may have been lost to follow-up at the end of the study if they met eligibility criteria at 2018. We adjusted for attrition and non-response using censoring weights, described below.

A total of 39,635 individuals met these criteria and were included in the study.

Average Treatment Effect

Researchers often want to know what might happen if we could change (or 'intervene on') a particular variable for everyone in a study—much like testing a new treatment in a randomised trial. Because we cannot always run an actual trial, we imagine a **target trial** (Hernán et al., 2016), a hypothetical experiment that clarifies exactly which cause-and-effect question we are trying to answer.

Here, we ask:

> 'How would the outcomes of interest change if, for everyone in the population, we set the exposure to **>4, scale range 1-7**, compared with setting it to **<=4, scale range 1-7**, given each individual's characteristics?'

Thus we compare two scenarios:

1. **1**: Everyone receives exposure level >4, scale range 1-7.
2. **0**: Everyone receives exposure level <=4, scale range 1-7.

The difference between these two population means is the **Average Treatment Effect (ATE)**. Because we evaluate several outcomes, ATE confidence intervals were corrected for multiplicity with bonferroni at $\alpha = 0.05$.

By combining time-series data with a rich baseline covariate set, we may, under the identifications assumptions described below, separate the causal effects of the exposure from spurious associations. Measuring demographics, personality traits, and other background factors at baseline helps ensure that, conditional on those covariates, assignment to the two exposure levels is 'as good as random.' (See Appendix D for a full statement of the required assumptions.)

Heterogeneous Treatment Effects and Treatment Policies

After estimating the overall average treatment effect (ATE) for the population, we turn to the question of whether different people respond differently. We investigate effect modifiers (or moderators)—factors that make the intervention more or less effective for certain subgroups—by estimating the Conditional Average Treatment Effect (CATE) using a causal forest approach. While the ATE reflects the overall impact, the CATE reveals how that impact can vary across individuals with different baseline characteristics. We denote the individual-level estimated treatment effect as $\hat{\tau}(x)$, which represents the predicted benefit for an individual with covariates x. A notable advantage of causal forests is that we do not have to specify potential moderators in advance; the algorithm uncovers them automatically. We can also apply search algorithms to derive priority treatment rules that target the intervention to those most likely to benefit.

First, we standardised effect directions by inverting outcomes where lower scores were preferable so that positive values always indicated improvement. Specifically, we inverted Anxiety, Depression, Rumination.

Next, to reduce overfitting and separate real heterogeneity from noise, we split the data into training and validation folds (70/30). A causal forest trained on the first fold produced out-of-sample CATE predictions on the second. We then computed Rank-Weighted Average Treatment Effect (RATE) metrics—AUTOC and Qini—which quantify the gain from targeting the highest-ranked individuals (Tibshirani et al., 2024; Wager & Athey, 2018). Their p-values were corrected with Benjamini–Hochberg false-discovery-rate adjustment at q = 0.1 to control the exploratory false-discovery rate (Benjamini & Hochberg, 1995). Where heterogeneity remained reliable, we fitted depth-2 policy trees (Athey & Wager, 2021a, 2021b; Sverdrup et al., 2024) to distil transparent 'treat-if' rules (e.g., *treat if baseline score* > X).

All heterogeneity steps—calibration tests, RATE, Qini curves, and policy-tree learning—were implemented in R with **grf** (Tibshirani et al., 2024), **policytree** (Sverdrup et al., 2024), using graphical and summary functions from **margot** (Bulbulia, 2024a). This workflow identifies individualised effects, quantifies the policy value of targeting, and delivers practical decision rules. See Appendix D for full methodological details.

Exposure Indicator

The New Zealand Attitudes and Values Study assesses Extraversion using the following question:

Mini-IPIP6 Extraversion dimension: (i) I am the life of the party. (ii) I don't talk a lot. (r) (iii) I keep in the background. (r) (iv) I talk to a lot of different people at parties.(Refer to Appendix A).

Causal Identification Assumptions

This study relies on the following identification assumptions for estimating the causal effect of Extraversion:

1. **Consistency**: the observed outcome under the observed Extraversion is equal to the potential outcome under that exposure level. As part of consistency, we assume no interference: the potential outcomes for one individual are not affected by the Extraversion status of other individuals.

2. **No unmeasured confounding**: all variables that affect both Extraversion and the outcome have been measured and accounted for in the analysis.

3. **Positivity**: there is a non-zero probability of receiving each level of Extraversion for every combination of values of Extraversion and confounders in the population. Positivity is the only fundamental casual assumption that can be evaluated with data (refer to Appendix E).

Confounding Control

To manage confounding in our analysis, we implement VanderWeele (2019)'s *modified disjunctive cause criterion* by following these steps:

1. **Identified all common causes** of both the treatment and outcomes.
2. **Excluded instrumental variables** that affect the exposure but not the outcome. Instrumental variables do not contribute to controlling confounding and can reduce the efficiency of the estimates.
3. **Included proxies for unmeasured confounders** affecting both exposure and outcome. According to the principles of d-separation Pearl (2009), using proxies allows us to control for their associated unmeasured confounders indirectly.
4. **Controlled for baseline exposure** and **baseline outcome**. Both are used as proxies for unmeasured common causes, enhancing the robustness of our causal estimates, refer to VanderWeele et al. (2020).

Statistical Estimation

We estimate heterogeneous treatment effects with Generalized Random Forests (GRF) (Tibshirani et al., 2024). GRF extends random forests for causal inference by focusing on conditional average treatment effects (CATE). It handles complex interactions and non-linearities without explicit model specification, and it provides 'honest' estimates by splitting data between model-fitting and inference. GRF is doubly robust because it remains consistent if either the outcome model or the propensity model is correct. We evaluate policies with the `policytree` package (Athey & Wager, 2021b; Sverdrup et al., 2024) and visualise results with `margot` (Bulbulia, 2024a). (Refer to Appendix D for a detailed explanation of our approach.)

Missing Data

The GRF package accepts missing values at baseline. To obtain valid inference for missing responses we computed inverse probability of censoring weights for censoring of the exposure, given that systematic censoring following the baseline wave may lead to selection bias that limit generalistion to the baseline target population (Bulbulia, 2024b). See Appendix D.

Sensitivity Analysis

We perform sensitivity analyses using the E-value metric (Linden et al., 2020; VanderWeele & Ding, 2017). The E-value represents the minimum association strength (on the risk-ratio scale) that an unmeasured confounder would need with both exposure and outcome—after adjusting for measured covariates—to explain away the observed association (Linden et al., 2020; VanderWeele et al., 2020). Confidence intervals for each E-value were derived from the multiplicity-adjusted confidence intervals of the corresponding coefficient estimates (bonferroni, $\alpha = 0.05$), so the sensitivity analysis obeys the same error-control framework as the main results.

Results

Average Treatement Effects

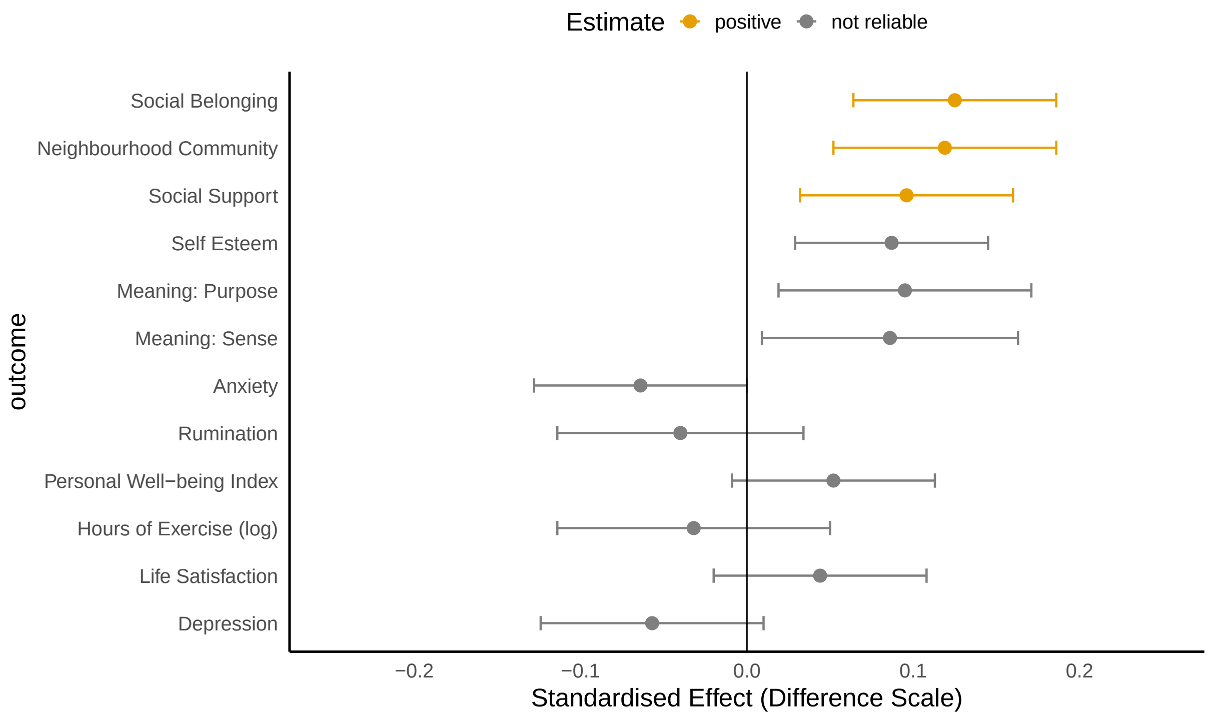

Figure 1: Average Treatment Effects on Multi-dimensional Wellbeing

Table 1: Average Treatment Effects on Multi-dimensional Wellbeing

Outcome	ATE	2.5 %	97.5 %	E-Value	E-Value bound
Social Belonging	**0.125**	**0.083**	**0.167**	**1.488**	**1.372**
Neighbourhood Community	**0.119**	**0.073**	**0.165**	**1.471**	**1.343**
Social Support	**0.096**	**0.052**	**0.14**	**1.407**	**1.277**
Self Esteem	**0.087**	**0.047**	**0.127**	**1.381**	**1.26**
Meaning: Purpose	**0.095**	**0.043**	**0.147**	**1.404**	**1.241**
Meaning: Sense	**0.086**	**0.033**	**0.138**	**1.378**	**1.208**
Anxiety	**-0.064**	**-0.107**	**-0.02**	**1.312**	**1.159**
Depression	**-0.057**	**-0.103**	**-0.011**	**1.29**	**1.116**
Personal Well-being Index	**0.052**	**0.011**	**0.094**	**1.274**	**1.11**
Rumination	-0.04	-0.091	0.01	1.233	1
Hours of Exercise (log)	-0.032	-0.088	0.024	1.204	1
Life Satisfaction	0.044	0	0.088	1.247	1

Confidence intervals were adjusted for multiple comparisons using bonferroni correction ($\alpha = 0.05$). E-values were also adjusted using bonferroni correction ($\alpha = 0.05$).

The following outcomes showed reliable causal evidence (E-value lower bound > 1.2): - Social Belonging: 0.125(0.064,0.186); on the original scale, 0.136 (0.07,0.203). E-value bound = 1.311 - Neighbourhood Community: 0.119(0.052,0.186); on the original scale, 0.187 (0.082,0.292). E-value bound = 1.273 - Social Support: 0.096(0.032,0.16); on the original scale, 0.107 (0.036,0.179). E-value bound = 1.203

Heterogeneous Treatment Effects

RATE AUTOC and RATE Qini

Rate Test

The RATE metric shows how much extra gain (or avoided loss) we achieve by **targeting** instead of treating everyone identically.

Technical note: In code we always set `policy = "treat_best"`; for harmful exposures this is interpreted as *'treat-those-most-sensitive'* (i.e., prioritise protection or withholding).

- **Beneficial exposure:** we rank by positive CATEs and deliver the exposure to those predicted to **benefit most**.
- **Detrimental exposure:** we rank by increasingly **positive** CATEs (more predicted harm) and identify those who should be protected or withheld from the exposure.

Either way, a larger **absolute** RATE shows that a CATE-based targeting rule 'outperforms' a one-size-fits-all policy—by boosting outcomes for beneficial exposures or – in the case where we are explore sensitivity to harm – evaluating increasing harms for detrimental ones.

Recall we flipped Anxiety, Depression, Rumination so **'higher' always tracks the analysis goal: higher = more benefit for beneficial exposures, higher = more harm for detrimental exposures.**

Because we test several outcomes, RATE p-values are adjusted with Benjamini–Hochberg false-discovery-rate adjustment (q = 0.1) before we decide whether heterogeneity is actionable.

Comparison of targeting operating characteristic (TOC) by rank average treatment effect (RATE): AUTOC vs QINI

We applied two TOC by RATE methods to the same causal-forest $\hat{\tau}$(x) estimates:

- **AUTOC** intensifies focus on top responders via logarithmic weighting.
- **QINI** balances effect size and prevalence via linear weighting.

We treated the RATE analysis as exploratory. All 12 outcomes were prespecified. To flag promising signals we controlled the false discovery rate at q=0.20.

When QINI and AUTOC disagree on positive RATE (only AUTOC yields a positive RATE for **Hours of Exercise (log)**; only QINI yields a positive RATE for Meaning: Sense), choose **QINI** to maximise overall benefit or **AUTOC** to focus on top responders.

Refer to Appendix E for details.

RATE AUTOC Results

Evidence for heterogeneous treatment effects (policy = treat best responders) using AUTOC

AUTOC uses logarithmic weighting to focus treatment on top responders. We treated the RATE analysis as exploratory. All 12 outcomes were prespecified. To flag promising signals we controlled the false discovery rate at q=0.20.

Positive RATE estimates for: **Hours of Exercise (log)**.

Estimates (**Hours of Exercise (log)**: 0.065 (95% CI 0.018, 0.112)) show robust heterogeneity.

For outcomes with adjusted p-values not meeting the FDR threshold of q = 0.20 (Meaning: Sense, Rumination, Anxiety, Personal Well-being Index, Self Esteem, Social Belonging, Social Support, Life Satisfaction, Meaning: Purpose, Depression, Neighbourhood Community), evidence is inconclusive.

Figure 2 presents the RATE AUTOC curve for Hours of Exercise (log)

RATE Qini Results

Evidence for heterogeneous treatment effects (policy = treat best responders) using QINI

QINI uses linear weighting to balance effect size and prevalence. We treated the RATE analysis as exploratory. All 12 outcomes were prespecified. To flag promising signals we controlled the false discovery rate at q=0.20.

Positive RATE estimates for: **Meaning: Sense**.

Estimates (**Meaning: Sense**: 0.020 (95% CI 0.002, 0.038)) show robust heterogeneity.

Negative RATE estimates for: Neighbourhood Community.

Estimates (Neighbourhood Community: -0.026 (95% CI -0.042, -0.010)) caution against CATE prioritisation.

For outcomes with adjusted p-values not meeting the FDR threshold of q = 0.20 (Hours of Exercise (log), Personal Well-being Index, Anxiety, Self Esteem, Social Support, Life Satisfaction, Social Belonging, Rumination, Meaning: Purpose, Depression), evidence is inconclusive.

QINI Curve Analysis

Qini Curves

The Qini curve shows the cumulative **gain** as we expand a targeting rule down the CATE ranking.

- **Beneficial exposure:** we add individuals from the top positive CATEs downward; the baseline is 'expose everyone.'
- **Detrimental exposure:** we first flip outcome direction (so higher values represent **more harm**; see Anxiety, Depression, Rumination), then *add* the exposure starting with individuals whose CATEs show the **greated harm**, gradually including those predicted to be more resistant to harm; the baseline is 'expose everyone.' The curve therefore quantifies the harm by when those most suceptible to harm are exposed.

If the Qini curve stays above its baseline, a targeted policy increases the outcome more than a one-size-fits-all alternative. (Outcome directions were flipped where needed—Anxiety, Depression, Rumination—so the positively valenced exposures always have positively valanced outcomes and negative exposures always have negatively valenced outcomes.)

We computed the cumulative benefits as we increase the treated fraction by prioritising conditional average treatment effects (CATE) at two different spend levels: 20% of a total budget and 50% of a total budget, where the contrast is no priority assignment. **Meaning Sense** At 20% spend: CATE prioritisation is beneficial (diff: 0.04 [95% CI: 0.01, 0.07]). At 50 % spend: No reliable benefits from CATE prioritisation.

Table 2 presents results for our Qini curve analysis at different spend rates.

Table 2: Qini Curve Results

Model	Spend 20%	Spend 50%
Meaning Sense	**0.04 [0.01, 0.07]**	-0.00 [-0.06, 0.05]

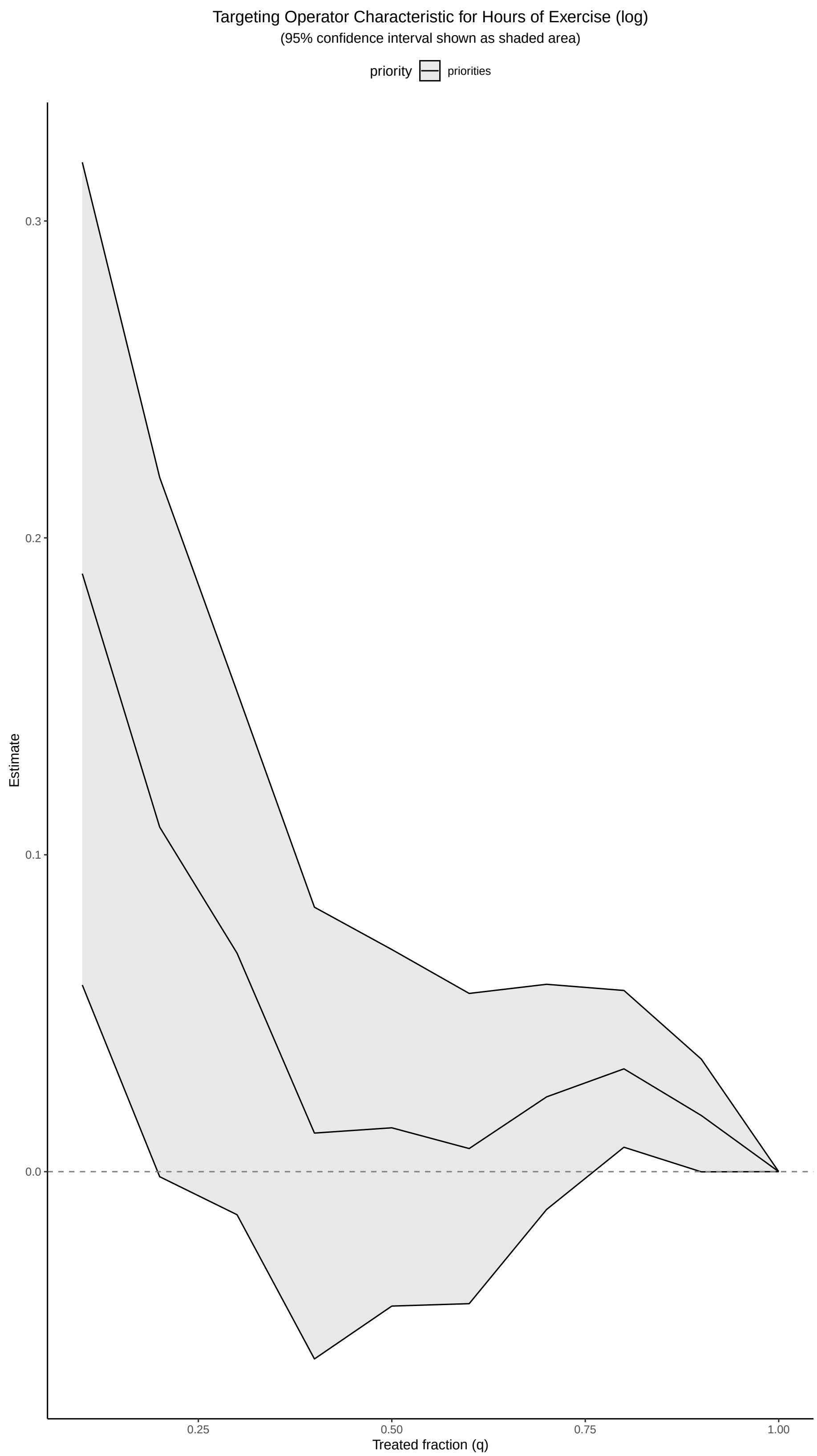

Figure 2: RATE AUTOC Graphs

Figure 3 presents the Qini curve for Meaning: Sense

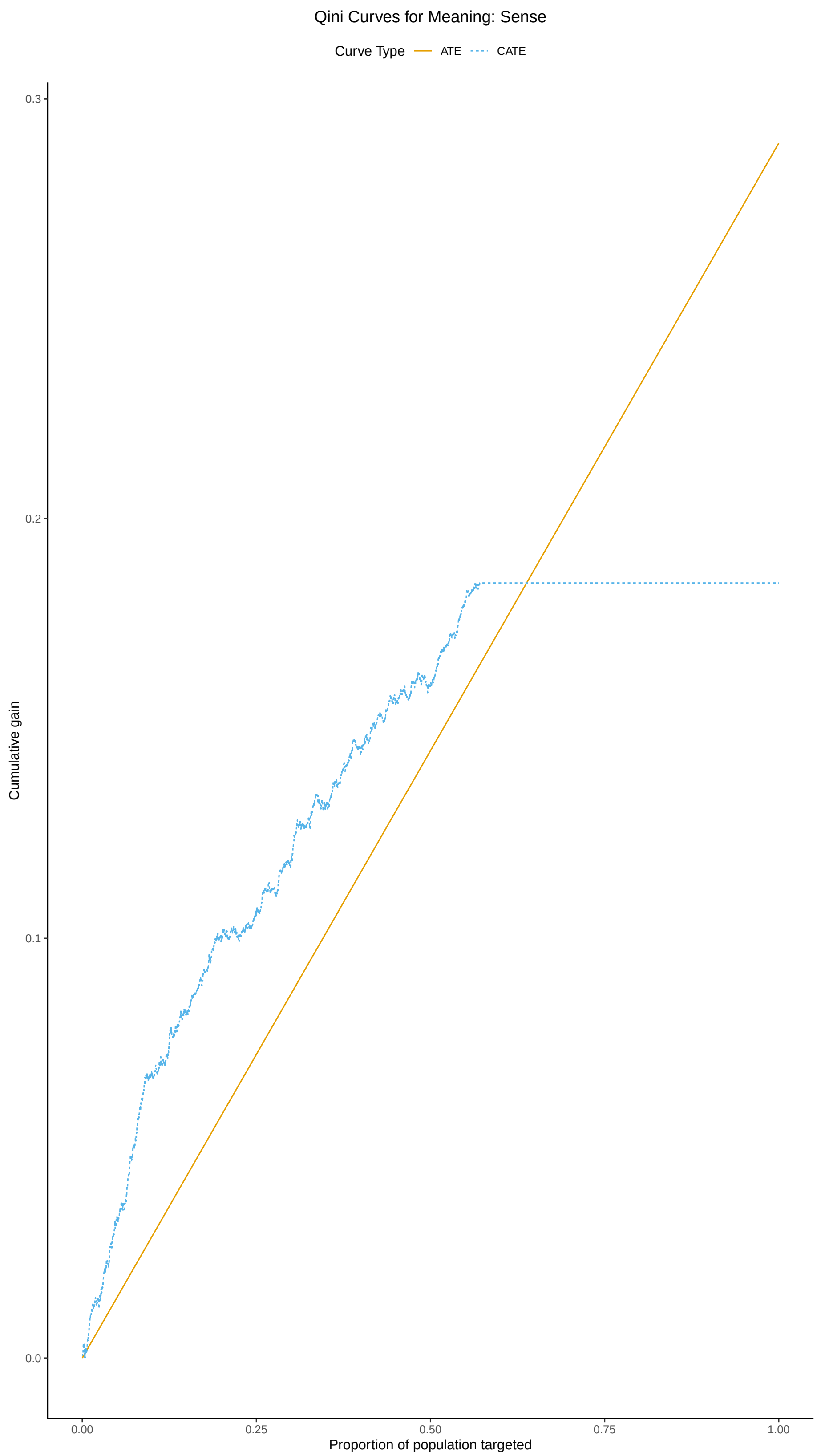

Figure 3: Qini Graph

Decision Rules (Who is Most Sensitive to Treatment?)

Policy Trees

We used policy trees (Athey & Wager, 2021a, 2021b; Sverdrup et al., 2024) to find straightforward 'if-then' rules for who benefits most from treatment, based on participant characteristics. Because we flipped some measures, a higher predicted effect always means greater improvement. Policy trees can uncover small but important subgroups whose treatment responses stand out, even when the overall differences might be modest.

Policy Tree Interpretations (depth 2)

A shallow policy tree recommends actions based on two splits for depth=2, or one split for depth=1. We trained on 50% of the data and evaluated on the rest.

Findings for log Hours Exercise:

Split 1: Short Form Health ≤ -0.333. Within that subgroup, split 2a: Belong ≤ -0.441, ⟶ **Control**; Belong > -0.441 ⟶ **Treated**.

Split 2: Short Form Health > -0.333. Within that subgroup, split 2b: Lifesat ≤ -0.268, ⟶ **Control**; Lifesat > -0.268 ⟶ **Treated**.

Findings for Meaning Sense:

Split 1: Alcohol Intensity ≤ -0.313. Within that subgroup, split 2a: log Hours Commute ≤ -0.496, ⟶ **Control**; log Hours Commute > -0.496 ⟶ **Treated**.

Split 2: Alcohol Intensity > -0.313. Within that subgroup, split 2b: Age ≤ 0.612, ⟶ **Treated**; Age > 0.612 ⟶ **Control**.

Discussion

Ethics

The University of Auckland Human Participants Ethics Committee reviews the NZAVS every three years. Our most recent ethics approval statement is as follows: The New Zealand Attitudes and Values Study was approved by the University of Auckland Human Participants Ethics Committee on 26/05/2021 for six years until 26/05/2027, Reference Number UAHPEC22576.

Author Statement

Acknowledgements

The New Zealand Attitudes and Values Study is supported by a grant from the Templeton Religious Trust (TRT0196; TRT0418). JB received support from the Max Plank Institute for the Science of Human History. The funders had no role in preparing the manuscript or deciding to publish it.

Data Availability

The data described in the paper are part of the New Zealand Attitudes and Values Study. Members of the NZAVS management team and research group hold full copies of the NZAVS data. A de-identified dataset containing only the variables analysed in this manuscript is available upon request from the corresponding author or any member of the NZAVS advisory board for replication or checking of any published study using NZAVS data. The code for the analysis can be found at OSF link.

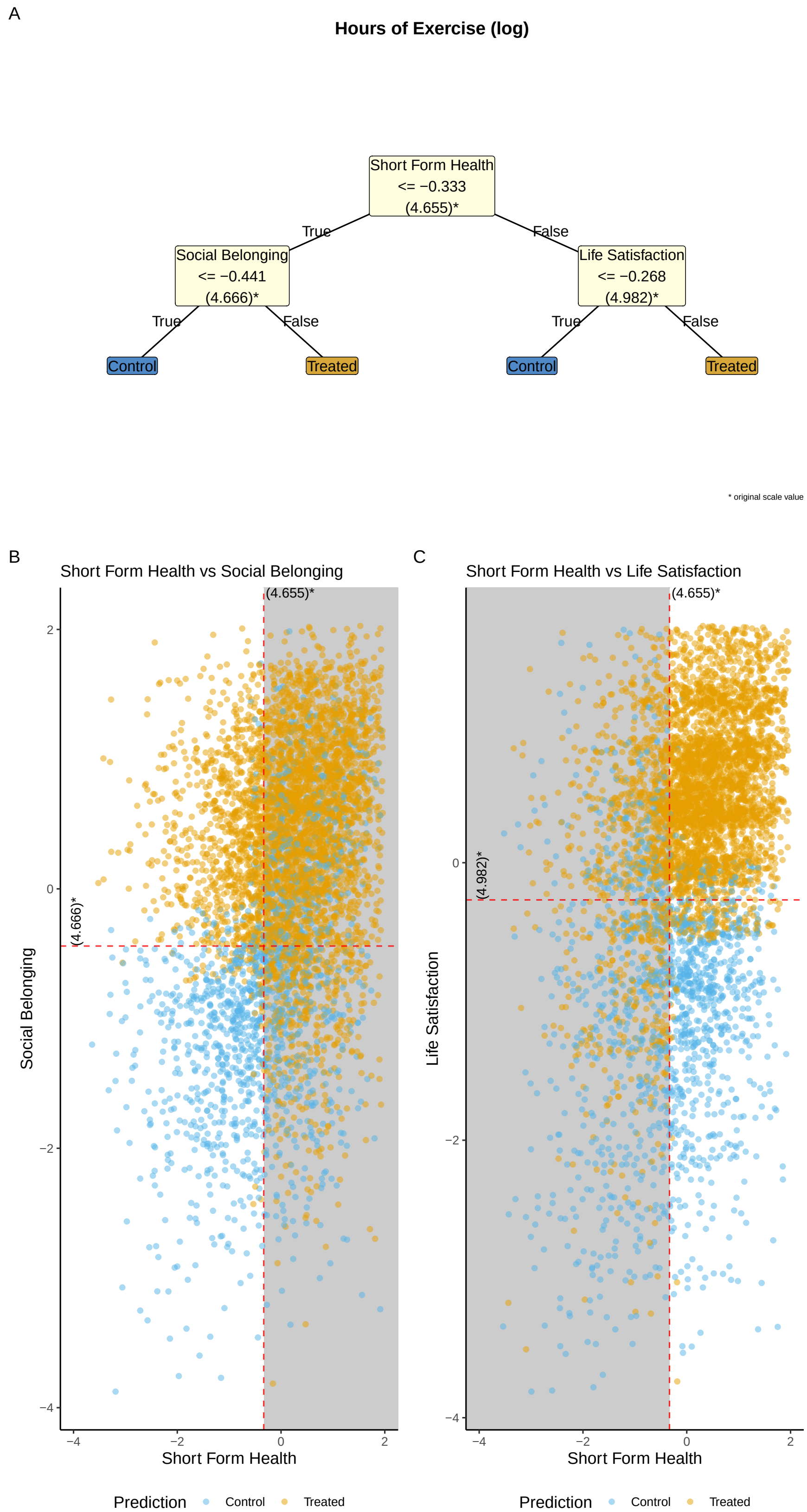

Figure 4: Decision Tree: {glued_name_policy_1}

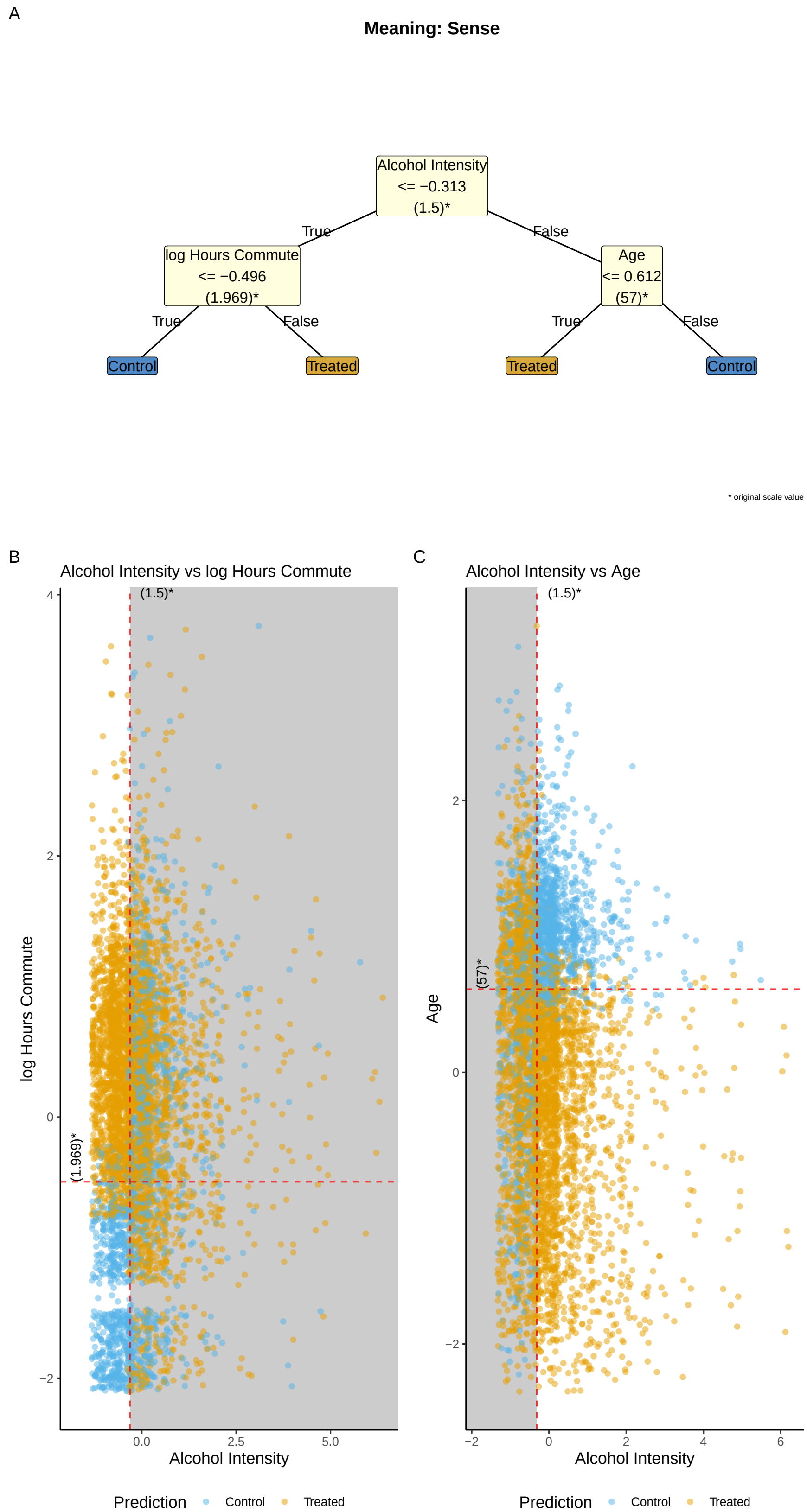

Figure 5: Decision Tree: {glued_name_policy_2}

Appendix A: Measures

Measures

Baseline Covariate Measures

Baseline Covariates

Age

What is your date of birth?

We asked participants' ages in an open-ended question ("What is your age?" or "What is your date of birth"). (**?** Developed for the NZAVS.)

Agreeableness

I sympathize with others' feelings. I am not interested in other people's problems. I feel others' emotions. I am not really interested in others (reversed).

Mini-IPIP6 Agreeableness dimension: (i) I sympathize with others' feelings. (ii) I am not interested in other people's problems. (r) (iii) I feel others' emotions. (iv) I am not really interested in others. (r) (Sibley et al., 2011)

Alcohol Frequency

"How often do you have a drink containing alcohol?"

Participants could chose between the following responses: '(1 = Never - I don't drink, 2 = Monthly or less, 3 = Up to 4 times a month, 4 = Up to 3 times a week, 5 = 4 or more times a week, 6 = Don't know)' (Health, 2013)

Alcohol Intensity

"How many drinks containing alcohol do you have on a typical day when drinking alcohol? (number of drinks on a typical day when drinking)"

Participants responded using an open-ended box. (Health, 2013)

Social Belonging

Know that people in my life accept and value me. Feel like an outsider (reversed). Know that people around me share my attitudes and beliefs.

We assessed felt belongingness with three items adapted from the Sense of Belonging Instrument (Hagerty & Patusky, 1995): (1) "Know that people in my life accept and value me"; (2) "Feel like an outsider"; (3) "Know that people around me share my attitudes and beliefs". Participants responded on a scale from 1 (Very Inaccurate) to 7 (Very Accurate). The second item was reversely coded. (Hagerty & Patusky, 1995)

Born in Nz

Where were you born? (please be specific, e.g., which town/city?)

Coded binary (1 = New Zealand; 0 = elsewhere.) (**?** Developed for the NZAVS.)

Conscientiousness

I get chores done right away. I like order. I make a mess of things. I often forget to put things back in their proper place.

Mini-IPIP6 Conscientiousness dimension: (i) I get chores done right away. (ii) I like order. (iii) I make a mess of things. (r) (iv) I often forget to put things back in their proper place. (r) (Sibley et al., 2011)

Education Level

What is your highest level of qualification?

We asked participants, "What is your highest level of qualification?". We coded participans highest finished degree according to the New Zealand Qualifications Authority. Ordinal-Rank 0-10 NZREG codes (with overseas school qualifications coded as Level 3, and all other ancillary categories coded as missing) (**?** Developed for the NZAVS.)

Employed

Are you currently employed (This includes self-employed of casual work)?

Binary response: (0 = No, 1 = Yes) (**?** Stats NZ Census Question)

Ethnicity

Which ethnic group(s) do you belong to?

Coded string: (1 = New Zealand European; 2 = Māori; 3 = Pacific; 4 = Asian) (**?** NZ Census coding.)

Disability Status

Do you have a health condition or disability that limits you and that has lasted for 6+ months?

We assessed disability with a one-item indicator adapted from Verbrugge (1997). It asks, "Do you have a health condition or disability that limits you and that has lasted for 6+ months?" (1 = Yes, 0 = No). (Verbrugge, 1997)

Log Hours with Children

Hours spent…looking after children.

We took the natural log of the response + 1. (Sibley et al., 2011)

Log Hours Commuting

Hours spent…travelling/commuting.

We took the natural log of the response + 1. (**?** Developed for the NZAVS.)

Log Hours of Exercise

Hours spent…exercising/physical activity.

We took the natural log of the response + 1. (Sibley et al., 2011)

Log Hours on Housework

Hours spent…housework/cooking.

We took the natural log of the response + 1. (Sibley et al., 2011)

Log Household Income

Please estimate your total household income (before tax) for the year XXXX.

We took the natural log of the response + 1. (**?** Developed for the NZAVS.)

Male

We asked participants' gender in an open-ended question: "what is your gender?"

Here, we coded all those who responded as Male as 1, and those who did not as 0. (Fraser et al., 2020)

Neuroticism

I have frequent mood swings. I am relaxed most of the time (reversed). I get upset easily. I seldom feel blue (reversed).

Mini-IPIP6 Neuroticism dimension: (i) I have frequent mood swings. (ii) I am relaxed most of the time. (r) (iii) I get upset easily. (iv) I seldom feel blue. (r) (Sibley et al., 2011)

Non Heterosexual

How would you describe your sexual orientation? (e.g., heterosexual, homosexual, straight, gay, lesbian, bisexual, etc.)

Open-ended question, coded as binary (not heterosexual = 1). (Greaves et al., 2017)

Nz Deprivation Index

New Zealand Deprivation - Decile Index - Using 2018 Census Data

Numerical: (1-10) (Atkinson et al., 2019)

Occupational Prestige Index

We assessed occupational prestige and status using the New Zealand Socio-economic Index 13 (NZSEI-13).

This index uses the income, age, and education of a reference group, in this case, the 2013 New Zealand census, to calculate a score for each occupational group. Scores range from 10 (Lowest) to 90 (Highest). This list of index scores for occupational groups was used to assign each participant a NZSEI-13 score based on their occupation. (Fahy et al., 2017)

Openness

I have a vivid imagination. I have difficulty understanding abstract ideas (reversed). I do not have a good imagination (reversed). I am not interested in abstract ideas (reversed).

Mini-IPIP6 Openness to Experience dimension: (i) I have a vivid imagination. (ii) I have difficulty understanding abstract ideas. (r) (iii) I do not have a good imagination. (r) (iv) I am not interested in abstract ideas. (r) (Sibley et al., 2011)

Parent

If you are a parent, in which year was your eldest child born?

Parents were coded as 1, while the others were coded as 0. (**?** for the NZAVS.)

Has Partner

What is your relationship status? (e.g., single, married, de-facto, civil union, widowed, living together, etc.)

Coded as binary (has partner = 1). (**?** Developed for the NZAVS.)

Political Conservatism

Please rate how politically liberal versus conservative you see yourself as being.

Ordinal response: (1 = Extremely Liberal, 7 = Extremely Conservative) (Jost, 2006)

Religious Identification

How important is your religion to how you see yourself?

Ordinal response: (1 = Not Important, 7 = Very Important) (**?** Developed for the NZAVS.)

Rural Classification

High Urban Accessibility = 1, Medium Urban Accessibility = 2, Low Urban Accessibility = 3, Remote = 4, Very Remote = 5.

"Participants residence locations were coded according to a five-level ordinal categorisation ranging from Urban to Rural." (Whitehead et al., 2023)

Sample Frame Opt in

Participant was not randomly sampled from the New Zealand Electoral Roll.

Code string (Binary): (0 = No, 1 = Yes) (**?** Developed for the NZAVS.)

Short Form Health

In general, would you say your health is...

Ordinal response: (1 = Poor, 7 = Excellent) (Instrument Ware Jr & Sherbourne, 1992)

Smoker

Do you currently smoke tobacco cigarettes?

Binary smoking indicator (0 = No, 1 = Yes). (**?** Developed for NZAVS.)

Exposure Measures

Exposure Variable

Extraversion

I am the life of the party. I don't talk a lot (reversed). I keep in the background (reversed). I talk to a lot of different people at parties.

Mini-IPIP6 Extraversion dimension: (i) I am the life of the party. (ii) I don't talk a lot. (r) (iii) I keep in the background. (r) (iv) I talk to a lot of different people at parties. (Sibley et al., 2011)

Outcome Measures

Outcome Variables

Social Belonging

Know that people in my life accept and value me. Feel like an outsider (reversed). Know that people around me share my attitudes and beliefs.

We assessed felt belongingness with three items adapted from the Sense of Belonging Instrument (Hagerty & Patusky, 1995): (1) "Know that people in my life accept and value me"; (2) "Feel like an outsider"; (3) "Know that people around me share my attitudes and beliefs". Participants responded on a scale from 1 (Very Inaccurate) to 7 (Very Accurate). The second item was reversely coded. (Hagerty & Patusky, 1995)

Anxiety

During the past 30 days, how often did...you feel restless or fidgety? During the past 30 days, how often did...you feel that everything was an effort? During the past 30 days, how often did...you feel nervous?

Ordinal response: (0 = None Of The Time; 1 = A Little Of The Time; 2= Some Of The Time; 3 = Most Of The Time; 4 = All Of The Time) (Kessler et al., 2002)

Depression

During the past 30 days, how often did...you feel hopeless? During the past 30 days, how often did...you feel so depressed that nothing could cheer you up? During the past 30 days, how often did...you feel you feel restless or fidgety?

Ordinal response: (0 = None Of The Time; 1 = A Little Of The Time; 2= Some Of The Time; 3 = Most Of The Time; 4 = All Of The Time) (Kessler et al., 2002)

Life Satisfaction

I am satisfied with my life. In most ways my life is close to ideal.

Ordinal response (1 = Strongly Disagree to 7 = Strongly Agree). (Diener et al., 1985)

Log Hours of Exercise

Hours spent...exercising/physical activity.

We took the natural log of the response + 1. (Sibley et al., 2011)

Meaning Purpose

My life has a clear sense of purpose

Ordinal response (1 = Strongly Disagree to 7 = Strongly Agree). (Steger et al., 2006)

Meaning Sense

I have a good sense of what makes my life meaningful.

Ordinal response (1 = Strongly Disagree to 7 = Strongly Agree). (Steger et al., 2006)

Neighbourhood Community

I feel a sense of community with others in my local neighbourhood.

Ordinal response (1 = Strongly Disagree to 7 = Strongly Agree). (Sengupta et al., 2013)

Personal Well Being Index

'Your standard of living.' 'Your health.' 'Your future security.' 'Your personal relationships.'

The Personal Well-Being Index consists of three items, asking 'How satisfied are you with…' (Cummins et al., 2003)

Rumination

During the last 30 days, how often did…you have negative thoughts that repeated over and over?

Ordinal responses: 0 = None of The Time, 1 = A little of The Time, 2 = Some of The Time, 3 = Most of The Time, 4 = All of The Time. (Nolen-hoeksema & Morrow, 1993)

Self Esteem

On the whole am satisfied with myself. Take a positive attitude toward myself. Am inclined to feel that I am a failure (reversed).

Ordinal response (1 = Very inaccurate to 7 = Very accurate). (Rosenberg, 1965)

Social Support

There are people I can depend on to help me if I really need it. There is no one I can turn to for guidance in times of stress (reversed). I know there are people I can turn to when I need help.

Ordinal response: (1 = Strongly Disagree, 7 = Strongly Agree) (Cutrona & Russell, 1987)

Appendix B: Sample Characteristics

Sample Statistics: Baseline Covariates

Table 3 presents sample demographic statistics.

Table 3: Demographic statistics for New Zealand Attitudes and Values Cohort: {baseline_wave_glued}.

	2018
	(N=39635)
Age	
Mean (SD)	48.5 (13.9)
Median [Min, Max]	51.0 [18.0, 99.0]
Agreeableness	
Mean (SD)	5.35 (0.988)
Median [Min, Max]	5.47 [1.00, 7.00]
Missing	9 (0.0%)
Alcohol Frequency	
Mean (SD)	2.16 (1.34)
Median [Min, Max]	2.00 [0, 5.00]
Missing	1342 (3.4%)
Alcohol Intensity	
Mean (SD)	2.15 (2.09)
Median [Min, Max]	2.00 [0, 15.0]
Missing	2348 (5.9%)
Belong	
Mean (SD)	5.14 (1.07)
Median [Min, Max]	5.31 [1.00, 7.00]
Missing	7 (0.0%)
Born in NZ	
0	8510 (21.5%)
1	30670 (77.4%)
Missing	455 (1.1%)
Conscientiousness	
Mean (SD)	5.10 (1.06)
Median [Min, Max]	5.23 [1.00, 7.00]
Education Level	
no_qualification	1003 (2.5%)
cert_1_to_4	13801 (34.8%)
cert_5_to_6	4953 (12.5%)
university	10400 (26.2%)
post_grad	4220 (10.6%)
masters	3297 (8.3%)
doctorate	930 (2.3%)
Missing	1031 (2.6%)
Employed	
0	8111 (20.5%)
1	31475 (79.4%)
Missing	49 (0.1%)
Ethnicity	
euro	31454 (79.4%)

	2018
maori	4561 (11.5%)
pacific	971 (2.4%)
asian	2124 (5.4%)
Missing	525 (1.3%)
Disability Status	
Mean (SD)	0.223 (0.416)
Median [Min, Max]	0 [0, 1.00]
Missing	745 (1.9%)
Log Hours with Children	
Mean (SD)	1.18 (1.61)
Median [Min, Max]	0.0341 [0, 5.13]
Missing	1242 (3.1%)
Log Hours Commuting	
Mean (SD)	1.50 (0.832)
Median [Min, Max]	1.61 [0, 4.40]
Missing	1242 (3.1%)
Log Hours Exercising	
Mean (SD)	1.55 (0.846)
Median [Min, Max]	1.61 [0, 4.40]
Missing	1242 (3.1%)
Log Hours on Housework	
Mean (SD)	2.14 (0.782)
Median [Min, Max]	2.20 [0, 5.13]
Missing	1242 (3.1%)
Log Household Income	
Mean (SD)	11.4 (0.765)
Median [Min, Max]	11.5 [0.685, 14.9]
Missing	3067 (7.7%)
Male	
0	24766 (62.5%)
1	14767 (37.3%)
Missing	102 (0.3%)
Neuroticism	
Mean (SD)	3.49 (1.15)
Median [Min, Max]	3.48 [1.00, 7.00]
Missing	10 (0.0%)
Non-heterosexual	
0	35100 (88.6%)
1	2562 (6.5%)
Missing	1973 (5.0%)
NZ Deprivation Index	
Mean (SD)	4.77 (2.73)
Median [Min, Max]	4.05 [1.00, 10.0]
Missing	255 (0.6%)
Occupational Prestige Index	
Mean (SD)	54.1 (16.5)
Median [Min, Max]	54.0 [10.0, 90.0]
Missing	536 (1.4%)
Openness	

	2018
Mean (SD)	4.96 (1.12)
Median [Min, Max]	5.00 [1.00, 7.00]
Missing	3 (0.0%)
Parent	
0	11539 (29.1%)
1	27776 (70.1%)
Missing	320 (0.8%)
Has Partner	
Mean (SD)	0.752 (0.432)
Median [Min, Max]	1.00 [0, 1.00]
Missing	1244 (3.1%)
Political Conservatism	
Mean (SD)	3.59 (1.38)
Median [Min, Max]	3.97 [1.00, 7.00]
Missing	2682 (6.8%)
Religious Identification	
Mean (SD)	2.36 (2.18)
Median [Min, Max]	1.00 [1.00, 7.00]
Missing	1050 (2.6%)
Rural Classification	
High Urban Accessibility	24406 (61.6%)
Medium Urban Accessibility	7431 (18.7%)
Low Urban Accessibility	4818 (12.2%)
Remote	2241 (5.7%)
Very Remote	486 (1.2%)
Missing	253 (0.6%)
Sample Frame Opt-In	
0	38485 (97.1%)
1	1150 (2.9%)
Short Form Health	
Mean (SD)	5.05 (1.17)
Median [Min, Max]	5.04 [1.00, 7.00]
Missing	6 (0.0%)
Smoker	
0	35771 (90.3%)
1	2880 (7.3%)
Missing	984 (2.5%)

Sample Statistics: Exposure Variable

Table 4: Demographic statistics for New Zealand Attitudes and Values Cohort waves 2018.

	2018	2019
	(N=39635)	(N=39635)
Extraversion		
Mean (SD)	3.91 (1.20)	3.86 (1.19)
Median [Min, Max]	3.96 [1.00, 7.00]	3.79 [1.00, 7.00]

	2018	2019
Missing	0 (0%)	11117 (28.0%)
Extraversion (binary)		
[1.0,4.0]	21138 (53.3%)	15637 (39.5%)
(4.0,7.0]	18497 (46.7%)	12881 (32.5%)
Missing	0 (0%)	11117 (28.0%)

Sample Statistics: Outcome Variables

Table 5: Outcome variables measured at

	2018	2020	Overall
	(N=39635)	(N=39635)	(N=79270)
Social Belonging			
Mean (SD)	5.14 (1.07)	5.06 (1.09)	5.11 (1.08)
Median [Min, Max]	5.31 [1.00, 7.00]	5.05 [1.00, 7.00]	5.30 [1.00, 7.00]
Missing	7 (0.0%)	13278 (33.5%)	13285 (16.8%)
Anxiety			
Mean (SD)	1.21 (0.774)	1.17 (0.756)	1.19 (0.767)
Median [Min, Max]	1.00 [0, 4.00]	1.00 [0, 4.00]	1.00 [0, 4.00]
Missing	51 (0.1%)	13275 (33.5%)	13326 (16.8%)
Depression			
Mean (SD)	0.584 (0.751)	0.550 (0.723)	0.571 (0.740)
Median [Min, Max]	0.333 [0, 4.00]	0.333 [0, 4.00]	0.333 [0, 4.00]
Missing	54 (0.1%)	13273 (33.5%)	13327 (16.8%)
Life Satisfaction			
Mean (SD)	5.30 (1.20)	5.25 (1.23)	5.28 (1.21)
Median [Min, Max]	5.50 [1.00, 7.00]	5.50 [1.00, 7.00]	5.50 [1.00, 7.00]
Missing	260 (0.7%)	13560 (34.2%)	13820 (17.4%)
Hours of Exercise (log)			
Mean (SD)	1.55 (0.846)	1.63 (0.839)	1.58 (0.844)
Median [Min, Max]	1.61 [0, 4.40]	1.78 [0, 4.40]	1.61 [0, 4.40]
Missing	1242 (3.1%)	13770 (34.7%)	15012 (18.9%)
Meaning: Purpose			
Mean (SD)	5.20 (1.41)	5.15 (1.44)	5.18 (1.42)
Median [Min, Max]	5.05 [1.00, 7.00]	5.04 [1.00, 7.00]	5.04 [1.00, 7.00]
Missing	1010 (2.5%)	13650 (34.4%)	14660 (18.5%)
Meaning: Sense			
Mean (SD)	5.71 (1.22)	5.71 (1.19)	5.71 (1.20)
Median [Min, Max]	5.99 [1.00, 7.00]	5.99 [1.00, 7.00]	5.99 [1.00, 7.00]
Missing	128 (0.3%)	13162 (33.2%)	13290 (16.8%)
Neighbourhood Community			
Mean (SD)	4.19 (1.66)	4.38 (1.57)	4.27 (1.63)
Median [Min, Max]	4.03 [1.00, 7.00]	4.95 [1.00, 7.00]	4.04 [1.00, 7.00]
Missing	212 (0.5%)	13202 (33.3%)	13414 (16.9%)
Pwi			
Mean (SD)	7.09 (1.66)	7.18 (1.63)	7.12 (1.65)
Median [Min, Max]	7.29 [0, 10.0]	7.47 [0, 10.0]	7.46 [0, 10.0]
Missing	41 (0.1%)	13120 (33.1%)	13161 (16.6%)
Rumination			
Mean (SD)	0.853 (1.00)	0.797 (0.959)	0.831 (0.987)
Median [Min, Max]	0.955 [0, 4.00]	0.0495 [0, 4.00]	0.953 [0, 4.00]
Missing	135 (0.3%)	13335 (33.6%)	13470 (17.0%)
Self Esteem			
Mean (SD)	5.14 (1.28)	5.13 (1.27)	5.14 (1.28)
Median [Min, Max]	5.34 [1.00, 7.00]	5.34 [1.00, 7.00]	5.34 [1.00, 7.00]
Missing	11 (0.0%)	13280 (33.5%)	13291 (16.8%)

	2018	2020	Overall
Social Support			
Mean (SD)	5.95 (1.12)	5.94 (1.12)	5.95 (1.12)
Median [Min, Max]	6.30 [1.00, 7.00]	6.29 [1.00, 7.00]	6.30 [1.00, 7.00]
Missing	30 (0.1%)	13112 (33.1%)	13142 (16.6%)

Appendix C: Transition Matrix to Check The Positivity Assumption

Table 6: Transition Matrix Showing Change

From / To	State 0	State 1	Total
State 0	17572	2271	19843
State 1	2400	6275	8675

These transition matrices capture shifts in states between consecutive waves. Each cell shows the count of individuals transitioning from one state to another. Rows are the initial state (From), columns the subsequent state (To). **Diagonal entries** (in **bold**) mark those who stayed in the same state.

Appendix D: Approach to Heterogeneity

Appendix X. Estimating and Interpreting Heterogeneous Treatment Effects with grf

Here we explain a heterogeneous-treatment-effect (HTE) analysis using causal forests (Tibshirani et al., 2024). In our workflow, we move from the average treatment effect (ATE) to individualised effects, quantify the practical value of targeting, and finish with interpretable decision rules.

1 Average Treatment Effect (ATE)

The ATE answers: *'What would happen, on average, if everyone received treatment versus no one?'*

$$\text{ATE} = E[Y(1) - Y(0)].$$

Using the `grf` package, we estimate the ATE doubly-robustly. Because we analyse several outcomes, we adjust ATE p-values with bonferroni ($\alpha = 0.05$) to control the family-wise error rate.

2 Do Effects Vary? Formal Test of Heterogeneity

Define the conditional average treatment effect (CATE)

$$\tau(x) = E[Y(1) - Y(0) \mid X = x].$$

If $\tau(x)$ is constant, effects are homogeneous; otherwise they vary. Classical interaction models impose strong forms; **grf** uses *causal forests* to discover complex, nonlinear heterogeneity (Wager & Athey, 2018). We assess heterogeneity with RATE p-values corrected via Benjamini–Hochberg false-discovery-rate adjustment (q = 0.1), controlling the false-discovery rate (Benjamini & Hochberg, 1995).

3 Causal Forests for Individualised Estimates

A causal forest is an ensemble of 'honest' causal trees that split on covariates to maximise treated–control contrasts. For each unit i we obtain

$$\hat{\tau}(x_i)$$

Strengths are flexibility, orthogonalisation, and per-person estimates.

4 Built-in Protection Against Over-fitting

Honesty (split half/estimate half) plus out-of-bag (OOB) predictions yield unbiased $\hat{\tau}(x)$ and standard errors without manual hyper-tuning.

5 Missing Data Handling

grf deploys 'Missing Incorporated in Attributes' (MIA): missingness is a valid split, so cases stay in the analysis – no ad-hoc imputation required.

6 Testing for Actionable Heterogeneity: the TOC & RATE Metrics

Ranking units by $\hat{\tau}$ defines a **Targeting Operator Characteristic** (TOC) curve: the cumulative gain from treating the top fraction q of predicted responders. Two scalar summaries:

- **RATE AUTOC** – area under the entire TOC; emphasises the very highest responders.
- **RATE Qini** – weighted area with weight q; rewards sustained gains across larger coverage (Yadlowsky et al., 2021).

Under $H_0 : \tau(x)$ constant, both equal 0. `grf::rank_average_treatment_effect()` supplies point estimates, standard errors, and t-tests.

Multiplicity control: We adjust AUTOC and Qini p-values with Benjamini–Hochberg false-discovery-rate adjustment (q = 0.1) before declaring actionable heterogeneity.

Here is an **interpretation tip**:

- AUTOC answers *'How sharply can we prioritise?'*
- Qini answers *'How valuable is targeting when budgets are modest but not tiny?'*

7 Visualising Policy Value: the Qini Curve

Plotting the Qini curve (cumulative gain vs q) reveals where returns plateau. Investigators (and policy audiences) can see at a glance whether benefits concentrate in, say, the top 20 % or persist up to 50 %.

8 Valid Inference for RATE / Qini

Although OOB predictions are out-of-sample per tree, they inherit forest-level dependence. We use an explicit **sample split**:

1. **Train set**: fit the causal forest and compute $\hat{\tau}(x)$.
2. **Test set**: compute RATE AUTOC/Qini and run H_0 tests.

This second split yields honest policy evaluation and guards against optimistic bias (Tibshirani et al., 2024).

9 From Black Box to Simple Rules: Policy Trees

Stakeholders value transparent criteria. The **policytree** algorithm takes $\hat{\tau}(x)$ or doubly-robust scores and learns a shallow decision tree that maximises expected welfare (Sverdrup et al., 2024).

Advantages: interpretability, the possibility of fairness constraints, and easy communication (e.g., *'treat if age < 25 and baseline severity high'*).

Training mirrors the split above: learn the tree on one fold, evaluate welfare on another.

> **Caveat** Splits identify predictors of *effect variation*, not causal levers. Changing a covariate in the tree does **not** guarantee an effect on $\tau(x)$.

10 Ethical and Practical Considerations

Statistical optimisation rarely aligns perfectly with equity or political feasibility. Decisions about who *should* receive treatment belong to democratic processes that weigh fairness, cost, and broader social values.

Summary Checklist

Table 7: This workflow delivers both rigorous inference and clear guidance for applied researchers: *How large is heterogeneity? When does targeting pay off? And can we express a good policy in a few, defensible rules?*

Stage	Tool	Key Output	Guard-rail
1 ATE	`average_treatment_effect`	$\widehat{\text{ATE}}$	Doubly-robust
2 – 3 CATE	`causal_forest`	$\hat{\tau}(x)$	Honest trees
6 Heterogeneity test	`rank_average_treatment_effect`	AUTOC, Qini + p	Sample split
7 Visualise	Qini curve	Gain vs q	Same test fold
9 Policy tree	`policy_tree`	Decision rule	Cross-validation + test fold

References

Allen, M. W., Wilson, M., Ng, S. H., & Dunne, M. (2000). Values and beliefs of vegetarians and omnivores. *The Journal of Social Psychology*, *140*(4), 405–422.

Athey, S., & Wager, S. (2021a). Policy Learning With Observational Data. *Econometrica*, *89*(1), 133–161. https://doi.org/10.3982/ECTA15732

Athey, S., & Wager, S. (2021b). Policy learning with observational data. *Econometrica*, *89*(1), 133–161. https://doi.org/https://doi.org/10.3982/ECTA15732

Atkinson, J., Salmond, C., & Crampton, P. (2019). *NZDep2018 index of deprivation, user's manual.*

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bulbulia, J. A. (2024a). *Margot: MARGinal observational treatment-effects*. https://doi.org/10.5281/zenodo.10907724

Bulbulia, J. A. (2024b). Methods in causal inference part 3: Measurement error and external validity threats. *Evolutionary Human Sciences*, *6*, e42. https://doi.org/10.1017/ehs.2024.33

Cummins, R. A., Eckersley, R., Pallant, J., Vugt, J. van, & Misajon, R. (2003). Development of a national index of subjective wellbeing: The Australian Unity Wellbeing Index. *Social Indicators Research*, *64*, 159–190.

Cutrona, C. E., & Russell, D. W. (1987). The provisions of social relationships and adaptation to stress. *Advances in Personal Relationships*, *1*, 37–67.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75.

Editorial, N. (2025). Supreme court protected status venezuelans. *The New York Times*. https://www.nytimes.com/2025/05/19/us/politics/supreme-court-protected-status-venezuelans.html

Fahy, K. M., Lee, A., & Milne, B. J. (2017). *New Zealand socio-economic index 2013*. Statistics New Zealand-Tatauranga Aotearoa.

Fraser, G., Bulbulia, J., Greaves, L. M., Wilson, M. S., & Sibley, C. G. (2020). Coding responses to an open-ended gender measure in a New Zealand national sample. *The Journal of Sex Research*, *57*(8), 979–986. https://doi.org/10.1080/00224499.2019.1687640

Greaves, L. M., Barlow, F. K., Lee, C. H., Matika, C. M., Wang, W., Lindsay, C.-J., Case, C. J., Sengupta, N. K., Huang, Y., Cowie, L. J., et al. (2017). The diversity and prevalence of sexual orientation self-labels in a New Zealand national sample. *Archives of Sexual Behavior*, *46*, 1325–1336.

Hagerty, B. M. K., & Patusky, K. (1995). Developing a Measure Of Sense of Belonging: *Nursing Research*, *44*(1), 9–13. https://doi.org/10.1097/00006199-199501000-00003

Health, Ministry of. (2013). *The New Zealand Health Survey: Content guide 2012-2013*. Princeton University Press.

Hernán, M. A., Sauer, B. C., Hernández-Díaz, S., Platt, R., & Shrier, I. (2016). Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *Journal of Clinical Epidemiology*, *79*, 70–75.

Instrument Ware Jr, J., & Sherbourne, C. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473–483.

Jost, J. T. (2006). The end of the end of ideology. *American Psychologist*, *61*(7), 651–670. https://doi.org/10.1037/0003-066X.61.7.651

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., Walters, E. E., & Zaslavsky, A. M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*(6), 959–976. https://doi.org/10.1017/S0033291702006074

Linden, A., Mathur, M. B., & VanderWeele, T. J. (2020). Conducting sensitivity analysis for unmeasured confounding in observational studies using e-values: The evalue package. *The Stata Journal*, *20*(1), 162–175.

Nolen-hoeksema, S., & Morrow, J. (1993). Effects of rumination and distraction on naturally occurring depressed mood. *Cognition and Emotion*, *7*(6), 561–570. https://doi.org/10.1080/02699939308409206

Pearl, J. (2009). *Causality*. Cambridge University Press.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.

Sengupta, N. K., Luyten, N., Greaves, L. M., Osborne, D., Robertson, A., Brunton, C., Armstrong, G., & Sibley, C. G. (2013). Sense of Community in New Zealand Neighbourhoods: A Multi-Level Model Predicting Social

Capital. *New Zealand Journal of Psychology*, *42*(1), 36–45.

Sibley, C. G. (2021). *Sampling procedure and sample details for the New Zealand Attitudes and Values Study*. https://doi.org/10.31234/osf.io/wgqvy

Sibley, C. G., Luyten, N., Purnomo, M., Mobberley, A., Wootton, L. W., Hammond, M. D., Sengupta, N., Perry, R., West-Newman, T., Wilson, M. S., McLellan, L., Hoverd, W. J., & Robertson, A. (2011). The Mini-IPIP6: Validation and extension of a short measure of the Big-Six factors of personality in New Zealand. *New Zealand Journal of Psychology*, *40*(3), 142–159.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology*, *53*(1), 80–93. https://doi.org/10.1037/0022-0167.53.1.80

Sverdrup, E., Kanodia, A., Zhou, Z., Athey, S., & Wager, S. (2024). *Policytree: Policy learning via doubly robust empirical welfare maximization over trees*. https://CRAN.R-project.org/package=policytree

Tibshirani, J., Athey, S., Sverdrup, E., & Wager, S. (2024). *Grf: Generalized random forests*. https://github.com/grf-labs/grf

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, *34*(3), 211–219.

VanderWeele, T. J., & Ding, P. (2017). Sensitivity analysis in observational research: Introducing the E-value. *Annals of Internal Medicine*, *167*(4), 268–274. https://doi.org/10.7326/M16-2607

VanderWeele, T. J., Mathur, M. B., & Chen, Y. (2020). Outcome-wide longitudinal designs for causal inference: A new template for empirical studies. *Statistical Science*, *35*(3), 437–466.

Verbrugge, L. M. (1997). A global disability indicator. *Journal of Aging Studies*, *11*(4), 337–362. https://doi.org/10.1016/S0890-4065(97)90026-8

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Whitehead, J., Davie, G., Graaf, B. de, Crengle, S., Lawrenson, R., Miller, R., & Nixon, G. (2023). Unmasking hidden disparities: A comparative observational study examining the impact of different rurality classifications for health research in aotearoa new zealand. *BMJ Open*, *13*(4), e067927.

Yadlowsky, S., Fleming, S., Shah, N., Brunskill, E., & Wager, S. (2021). Evaluating treatment prioritization rules via rank-weighted average treatment effects. *arXiv Preprint arXiv:2111.07966*. https://doi.org/10.48550/arXiv.2111.07966