Causal effects of forgiveness on multi-dimensional well-being

An outcome-wide study

Richard Cowden

Don E Davis

Chris G. Sibley

AUTHORS (TYLER?, EV?)

Joseph A. Bulbulia

We model a causal effect of forgiveness on outcome-wide well-being. Our estimand: the effect in the target population of moving from -1 SD on foregiveness to average forgiveness. We use doubly-robust estimation to recover this effect. Our sample is a large, national probability panel of New Zelanders (N= 34,748), measured over three waves (years 2018-2021). We find that forgiveness affects most-dimensions of well-being.

## Introduction

Here, we use three waves of panel data from a national probability study in New Zealand to investigate the causal effects of forgiveness on multi-dimensional well-being.

We assessed participants’ forgiveness using three items, respectively adapted from Caprara (1986) and Berry et al. (2005), and developed for NZAVS: (1) Sometimes I can’t sleep because of thinking about past wrongs I have suffered; (2) I can usually forgive and forget when someone does me wrong; (3) I find myself regularly thinking about past times that I have been wronged. Participants indicated their agreement with these items (1 = Strongly Disagree to 7 = Strongly Agree). The values for the first and the third items were reversely coded.

We generally follow Vanderweele’s “outcomewide approach” (Tyler J. VanderWeele, Mathur, and Chen 2020) (deviations from this approach, and rational noted in the text.)

## Method

### Sample

Data were collected as part of The New Zealand Attitudes and Values Study (NZAVS) is an annual longitudinal national probability panel study of social attitudes, personality, ideology and health outcomes. The NZAVS began in 2009. It includes questionnaire responses from more than 70,000 New Zealand residents. The study includes researchers from many New Zealand universities, including the University of Auckland, Victoria University of Wellington, the University of Canterbury, the University of Otago, and Waikato University. Because the survey asks the same people to respond each year, it can track subtle change in attitudes and values over time, and is an important resource for researchers both in New Zealand and around the world. The NZAVS is university-based, not- for-profit and independent of political or corporate funding.https://doi.org/10.17605/OSF.IO/75SNB

### Eligibility criteria

The sample consisted of respondents to NZAVS waves 10, 11, and 12 (years 2018-2021) (See Appendix A.)

1. Included: those participants who at the baseline wave (NZAVS wave 10, years 2018-2019).
2. Included: missing data for all variables except the exposure. This includes loss-to-follow up in wave 12 (the outcome year). Missing values were multiply imputed conditional on all covariates (see explanation for details).
3. Excluded: those with missing data on the exposure in wave 10 (baseline) and wave 11 (exposure)  
   There were **34,748** NZAVS participants who met these criteria.

### The exposure: foregiveness

[Table 1](#tbl-transition) is a transition matrix describes the shifts from one state of forgiveness to another state between the baseline wave and the following wave. The numbers in the cells represent the number of individuals who transitioned from one state (rows) to another (columns). For example, the cell in the first row and second column shows the number of individuals who transitioned from the first state (indicated by the left-most cell in the row) to the second state. The top left cell shows the number of individuals who remained in the first state. For the continuous exposure analysis, we fit regressions with splines to avoid linearity assumptions. We find evidence for both change in the exposure between the baseline wave and the exposure wave (wave 11) among the N= 34,748 participants, thus satisfying the positivity assumption (see below). In our continuous models we ask: what would happen if we were to move everyone in the population from an average scores on forgiveness to +1 standard deviation higher score on forgiveness.

Table 1: Transition matrix for continuous stability and change in forgiveness (rounded to nearest whole number)

| From | State 1 | State 2 | State 3 | State 4 | State 5 | State 6 | State 7 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| State 1 | 80 | 72 | 49 | 30 | 10 | 10 | 3 |
| State 2 | 94 | 289 | 317 | 198 | 107 | 27 | 3 |
| State 3 | 59 | 358 | 911 | 961 | 542 | 196 | 26 |
| State 4 | 31 | 222 | 920 | 1935 | 1820 | 803 | 114 |
| State 5 | 9 | 120 | 467 | 1743 | 3722 | 3013 | 485 |
| State 6 | 6 | 27 | 163 | 735 | 2637 | 5582 | 1849 |
| State 7 | 3 | 7 | 34 | 111 | 440 | 1449 | 1959 |

### Indicators of well-being.

We assessed well-being following Tyler J. VanderWeele, Mathur, and Chen (2020) outcome-wide template. Outcomewide studies argue that, rather than cherry-picking one or several domains of well-being, science may advance more rapidly, and with greater hope for replication, by assess well-being across as many range of indicators the data may afford. To assist with interpretation, Tyler J. VanderWeele, Mathur, and Chen (2020) groups well-being into larger dimensions of interest. Here, we identify five-domains: health, embodied well-being, practical well-being, reflective well-being, and social well-being (see Appendix A.)

### Assumptions for causal inference

In this study, we aim to find out if changing one thing (an intervention) would cause a different outcome. However, in reality, we can only observe what happens when we intervene, not what would have happened if we did not, or vice versa. So we face a fundamental problem in showing cause and effect directly. To get around this, we estimate what we call the “average treatment effect” (ATE), which compares the average outcomes of those who received the intervention and those who did not.

For us to say something is a cause, we rely on three important assumptions:

1. **Causal consistency**: We satisfy the causal consistency assumption when the potential or counterfactual outcome under exposure corresponds to the observed outcome .
2. **Exchangability**:When we assume exchangability, we assume that the treatment assignment is independent of the potential outcomes, given a set of observed covariates – that is, *no unmeasured confounding.*
3. **Positivity**: The positivity assumption is satisfied if there is a positive probability of receiving the exposure or non-receiving the exposure within every level of the the covariates.
   * **Random non-positivity**: the casual effect of ageing with observations missing within our data, but may be assumed to exist.
   * **Deterministic non-positivity**:the causal effect is inconceivable. For example, the causal effect of hysterectomy in biological males violates deterministic non-positivity.

### Causal identification strategy

Effects must follow causes. To avoid the problems of reverse causation, we measured outcomes during the year following the exposure (NZAVS wave 2020). The causal graph presented in [Figure 1](#fig-outcomewide-dag) describes our method for confouding control. We follow Tyler J. VanderWeele, Mathur, and Chen (2020) in adopting a modified disjunctive cause criterion which states:

1. **Identify all relevant factors**: first, find every covariates that can influence either the exposure or the outcomes (across the five domains), or both. These factors are any variable that can have an impact on the exposure or outcome, are that might be the effect of such a factor.
2. **Remove instrumental variables**: next, take out any factors that are known to be instrumental variables. These are factors that cause the exposure but do not affect the outcome. Including instrumental variables reduces efficiency.
3. **Include proxy variables for unmeasured common causes**: if there are any unmeasured factors that influence both the exposure and outcome, but we don not have direct measurements for them, we should try to include a proxy for these. A proxy is an effect of the variable.
4. **Control for prior exposure**: Controlling for prior exposure assesses the effects of “incident exposure” rather than “prevalent exposure” - and is a critical step in causal inference(Danaei, Tavakkoli, and Hernán 2012; Hernan and Robins 2023). By including prior exposure in the analysis, we can more effectively emulate a controlled trial. This approach not only helps interpret the effect of exposure changes but also strengthens confounding control. It aids in avoiding reverse causation and managing other forms of unmeasured confounding. This setup ensures that any unmeasured confounder would have to influence both the outcome and initial exposure, irrespective of previous exposure levels, to explain an observed exposure-outcome association.
5. **Control for prior outcome**: It is also vital to control for the outcome measured at baseline – the ‘baseline outcome’. This tactic aims to rule out reverse causation by ensuring that the cause-effect relationship follows the right temporal order. Even though it does not eliminate the possibility of reverse causation, controlling for the baseline outcome helps mitigate its effects. Hence, along with a rich set of covariates, the baseline outcome should be included in the covariate set to make the confounding control assumption as plausible as possible. It is often the strongest confounder affecting both the exposure and subsequent outcome. (For a detailed account of confounding control in three-wave panel designs see Tyler J. VanderWeele, Mathur, and Chen (2020))

To avoid bias, we must also handle missing data arising from non-response or panel attrition (loss-to-follow up). Selection bias occurs when… [jb to say more] To address selection bias we perform multiple imputation *separately* by each exposure condition, and combine the imputations after modelling missingness conditional on the observed covariates (see: (Zhang et al. 2023; Westreich et al. 2015)) . We implement multiple imputation using the mice package in R (Van Buuren 2018). We imputed 10 x missing data sets which were passed separately to the the MatchThem and WeightIt packages for propensity score matching. There are several ways in which selection bias can occur from missing responses and loss to follow up (see (Hernán et al. 2016; Hernan and Robins 2023), [Figure 1](#fig-outcomewide-dag) presents a scenario in which the exposure affects the process of selection. For example … (say more).

|  |
| --- |
| Figure 1: Causal graph: three-wave panel design with selection bias |

### Causal estimation

We use Doubly Robust Estimation to estimate the impact of increased forgiveness in people. This method combines two different statistical techniques and is resilient to errors: if one of them goes wrong, the other can still provide accurate results.

The first part of our process is calculating a propensity score for each individual. This is probability of someone experiencing a change in their “forgiveness” based on certain characteristics they have. To get this score, we use a statistical model that estimates the contributions of these characteristics.

Then, we assign a weight to each person in our study based on their propensity score. Those who have higher probability of experiencing a change get a higher weight and vice versa. This helps balance out our sample in respect of confounders that might induce a non-causal association between the exposure (forgiveness) and the aspects of multi-dimensional well-being that we here investigate.

Next, we use these weights to create another statistical model, but this time focusing on the outcome – about 35 measures of well-being in five domains. This second model considers the exposure to the change, as well as the person’s characteristics, and it takes into account the weight we assigned to each person earlier.

Then, we use this outcome model to predict what would happen if everyone in our sample had the same level of change in their forgiveness, no matter what their actual level of change was. Predicting and contrasting counterfactual outcomes is at the heart of causal estimation.

Finally, we estimate the overall effect of changing the “forgiveness” level by calculating the difference in these predicted outcomes between two specific levels of change. For our study, we looked at two scenarios: one where the forgiveness increases by one standard deviation from the average and another where it moves from the lower quartile to the third quartile of response on seven-point scale.

We used simulation methods to calculate our standard errors and confidence intervals. This allowed us to obtain a margin of error in which we can confidently say our actual result lies (Greifer et al. 2023).

### Baseline confounders, exposure, and outcome measures

( *Say more.* *Tables here* )

((*Insert section on why the assumptions of factor models are much stronger than people are aware, and that – despite a loss of efficiency – it is often best to use single item measures except when there are clear* conceptual\* reasons to do otherwise.\* ))

**SEE**: (Tyler J. VanderWeele 2022)

## Results: the continuous model

This set of results reports the hypothetical effect of intervening on forgiveness by setting the population to -1 SD forgiveness and increasing forgiveness to the population average level of forgiveness. Outcomes are measured one-year after this hypothetical intervention.

## Results for the contiuous model (intervention from -1 SD from average forgiveness)

### Effects on health

As indicate in [Figure 2](#fig-results-health_con), the expected + one-year effect of the contrast from average to +1 SD forgiveness is as follows:

|  |
| --- |
| Figure 2: Causal effects of religious loss on reported physical health (continuous exposure) |

Table 2: Table of results for the health domain (continuous exposure)

|  | E[Y(1)]-E[Y(0)] | 2.5 % | 97.5 % | E\_Value | E\_Val\_bound |
| --- | --- | --- | --- | --- | --- |
| BMI (sd) | 0.0027 | -0.0062 | 0.0112 | 1.052 | 1.000 |
| Alcohol frequency (sd) | -0.0130 | -0.0299 | 0.0069 | 1.122 | 1.000 |
| Alcohol intensity (sd) | -0.0244 | -0.0455 | -0.0040 | 1.174 | 1.061 |
| Hours exercise (log sd) | 0.0149 | -0.0112 | 0.0407 | 1.131 | 1.000 |
| Your health (sd) | 0.0473 | 0.0199 | 0.0789 | 1.258 | 1.145 |
| Get sick easier (reversed sd) | 0.0512 | 0.0233 | 0.0772 | 1.271 | 1.174 |
| Expect worse health (sd) | 0.0540 | 0.0311 | 0.0780 | 1.280 | 1.199 |
| Sleep hours (sd) | 0.0161 | -0.0089 | 0.0422 | 1.137 | 1.000 |

[Table 2](#tbl-results-health_con) presents the Population Average Treatment Effect (PATE) which is the expected difference in outcomes between treatment and control groups for the New Zealand population. We observed the following:

For the outcome ‘Expect worse health (sd)’, the PATE causal contrast is 0.054. The confidence interval ranges from 0.031 to 0.078. The E-value for this outcome is 1.28, indicating reliable evidence for causality.

For the outcome ‘Get sick easier (reversed sd)’, the PATE causal contrast is 0.051. The confidence interval ranges from 0.023 to 0.077. The E-value for this outcome is 1.271, indicating reliable evidence for causality.

For the outcome ‘Your health (sd)’, the PATE causal contrast is 0.047. The confidence interval ranges from 0.02 to 0.079. The E-value for this outcome is 1.258, indicating reliable evidence for causality.

For the outcome ‘Sleep hours (sd)’, the PATE causal contrast is 0.016. The confidence interval ranges from -0.009 to 0.042. The E-value for this outcome is 1.137, indicating evidence for causality is weak.

For the outcome ‘Hours exercise (log sd)’, the PATE causal contrast is 0.015. The confidence interval ranges from -0.011 to 0.041. The E-value for this outcome is 1.131, indicating evidence for causality is weak.

For the outcome ‘BMI (sd)’, the PATE causal contrast is 0.003. The confidence interval ranges from -0.006 to 0.011. The E-value for this outcome is 1.052, indicating no reliable evidence for causality.

For the outcome ‘Alcohol frequency (sd)’, the PATE causal contrast is -0.013. The confidence interval ranges from -0.03 to 0.007. The E-value for this outcome is 1.122, indicating evidence for causality is weak.

For the outcome ‘Alcohol intensity (sd)’, the PATE causal contrast is -0.024. The confidence interval ranges from -0.046 to -0.004. The E-value for this outcome is 1.174, indicating evidence for causality is weak.

### Effects on embodied well-being continuous model (intervention from -1 SD from average forgiveness)

|  |
| --- |
| Figure 3: Causal effects of religious loss on embodied well-being (continuous exposure) |

Table 3: Table of results for the embodied well-being domain (continuous exposure)

|  | E[Y(1)]-E[Y(0)] | 2.5 % | 97.5 % | E\_Value | E\_Val\_bound |
| --- | --- | --- | --- | --- | --- |
| Fatigue (sd) | -0.0562 | -0.0799 | -0.0312 | 1.287 | 1.204 |
| Rumination (sd) | -0.1435 | -0.1714 | -0.1175 | 1.538 | 1.465 |
| Kessler depressed (sd) | -0.0881 | -0.1208 | -0.0627 | 1.384 | 1.297 |
| Kessler effort (sd) | -0.0752 | -0.1005 | -0.0482 | 1.346 | 1.264 |
| Kessler hopeless (sd) | -0.1132 | -0.1381 | -0.0882 | 1.455 | 1.385 |
| Kessler nervous (sd) | -0.1001 | -0.1313 | -0.0711 | 1.419 | 1.331 |
| Kessler restless (sd) | -0.0647 | -0.0927 | -0.0309 | 1.314 | 1.211 |
| Kessler worthless (sd) | -0.0937 | -0.1282 | -0.0643 | 1.400 | 1.305 |

[Table 3](#tbl-results-embodied_con) presents the Population Average Treatment Effect (PATE) for the embodied domain.

For the outcome ‘Fatigue (sd)’, the PATE causal contrast is -0.056. The confidence interval ranges from -0.08 to -0.031. The E-value for this outcome is 1.287, indicating reliable evidence for causality.

For the outcome ‘Kessler restless (sd)’, the PATE causal contrast is -0.065. The confidence interval ranges from -0.093 to -0.031. The E-value for this outcome is 1.314, indicating reliable evidence for causality.

For the outcome ‘Kessler effort (sd)’, the PATE causal contrast is -0.075. The confidence interval ranges from -0.1 to -0.048. The E-value for this outcome is 1.346, indicating reliable evidence for causality.

For the outcome ‘Kessler depressed (sd)’, the PATE causal contrast is -0.088. The confidence interval ranges from -0.121 to -0.063. The E-value for this outcome is 1.384, indicating reliable evidence for causality.

For the outcome ‘Kessler worthless (sd)’, the PATE causal contrast is -0.094. The confidence interval ranges from -0.128 to -0.064. The E-value for this outcome is 1.4, indicating reliable evidence for causality.

For the outcome ‘Kessler nervous (sd)’, the PATE causal contrast is -0.1. The confidence interval ranges from -0.131 to -0.071. The E-value for this outcome is 1.419, indicating reliable evidence for causality.

For the outcome ‘Kessler hopeless (sd)’, the PATE causal contrast is -0.113. The confidence interval ranges from -0.138 to -0.088. The E-value for this outcome is 1.455, indicating reliable evidence for causality.

For the outcome ‘Rumination (sd)’, the PATE causal contrast is -0.144. The confidence interval ranges from -0.171 to -0.117. The E-value for this outcome is 1.538, indicating reliable evidence for causality.

### Effects on practical well-being contiuous model (intervention from -1 SD from average forgiveness)

|  |
| --- |
| Figure 4: Causal effects of religious loss on practical well-being (continuous exposure) |

Table 4: Table of results for the practical well-being domain (continuous exposure)

|  | E[Y(1)]-E[Y(0)] | 2.5 % | 97.5 % | E\_Value | E\_Val\_bound |
| --- | --- | --- | --- | --- | --- |
| Sexual satisfaction (sd) | 0.0430 | 0.0169 | 0.0703 | 1.244 | 1.138 |
| Perfectionism (sd) | -0.1017 | -0.1228 | -0.0821 | 1.423 | 1.365 |
| Body satisfaction (sd) | 0.0671 | 0.0405 | 0.0947 | 1.322 | 1.233 |
| Power self nocontrol (sd) | -0.1015 | -0.1333 | -0.0696 | 1.423 | 1.330 |
| Power others control (sd) | -0.1039 | -0.1313 | -0.0767 | 1.429 | 1.351 |
| Selfesteem satself (sd) | 0.0789 | 0.0530 | 0.1036 | 1.357 | 1.279 |
| Selfesteem postiveself (sd) | 0.1156 | 0.0895 | 0.1397 | 1.462 | 1.391 |
| Selfesteem failure (reversed, sd) | 0.1104 | 0.0839 | 0.1373 | 1.448 | 1.372 |
| Self control have lots (sd) | 0.0789 | 0.0556 | 0.1039 | 1.357 | 1.283 |
| Self control wish more (reversed, sd) | 0.0463 | 0.0220 | 0.0674 | 1.255 | 1.171 |
| Emotion reg out control (sd) | -0.1322 | -0.1576 | -0.1082 | 1.508 | 1.440 |
| Emotion reg hide neg emotions (sd) | -0.0406 | -0.0682 | -0.0131 | 1.235 | 1.122 |
| Emotion reg change thinking to calm (sd) | 0.0331 | 0.0030 | 0.0621 | 1.208 | 1.061 |

[Table 4](#tbl-results-practical_con) presents the PATE for the practical domain.

For the outcome ‘Selfesteem postiveself (sd)’, the PATE causal contrast is 0.116. The confidence interval ranges from 0.09 to 0.14. The E-value for this outcome is 1.462, indicating reliable evidence for causality.

For the outcome ‘Selfesteem failure (reversed, sd)’, the PATE causal contrast is 0.11. The confidence interval ranges from 0.084 to 0.137. The E-value for this outcome is 1.448, indicating reliable evidence for causality.

For the outcome ‘Selfesteem satself (sd)’, the PATE causal contrast is 0.079. The confidence interval ranges from 0.053 to 0.104. The E-value for this outcome is 1.357, indicating reliable evidence for causality.

For the outcome ‘Self control have lots (sd)’, the PATE causal contrast is 0.079. The confidence interval ranges from 0.056 to 0.104. The E-value for this outcome is 1.357, indicating reliable evidence for causality.

For the outcome ‘Body satisfaction (sd)’, the PATE causal contrast is 0.067. The confidence interval ranges from 0.04 to 0.095. The E-value for this outcome is 1.322, indicating reliable evidence for causality.

For the outcome ‘Self control wish more (reversed, sd)’, the PATE causal contrast is 0.046. The confidence interval ranges from 0.022 to 0.067. The E-value for this outcome is 1.255, indicating reliable evidence for causality.

For the outcome ‘Sexual satisfaction (sd)’, the PATE causal contrast is 0.043. The confidence interval ranges from 0.017 to 0.07. The E-value for this outcome is 1.244, indicating evidence for causality is weak.

For the outcome ‘Emotion reg change thinking to calm (sd)’, the PATE causal contrast is 0.033. The confidence interval ranges from 0.003 to 0.062. The E-value for this outcome is 1.208, indicating evidence for causality is weak.

For the outcome ‘Emotion reg hide neg emotions (sd)’, the PATE causal contrast is -0.041. The confidence interval ranges from -0.068 to -0.013. The E-value for this outcome is 1.235, indicating evidence for causality is weak.

For the outcome ‘Power self nocontrol (sd)’, the PATE causal contrast is -0.102. The confidence interval ranges from -0.133 to -0.07. The E-value for this outcome is 1.423, indicating reliable evidence for causality.

For the outcome ‘Perfectionism (sd)’, the PATE causal contrast is -0.102. The confidence interval ranges from -0.123 to -0.082. The E-value for this outcome is 1.423, indicating reliable evidence for causality.

For the outcome ‘Power others control (sd)’, the PATE causal contrast is -0.104. The confidence interval ranges from -0.131 to -0.077. The E-value for this outcome is 1.429, indicating reliable evidence for causality.

For the outcome ‘Emotion reg out control (sd)’, the PATE causal contrast is -0.132. The confidence interval ranges from -0.158 to -0.108. The E-value for this outcome is 1.508, indicating reliable evidence for causality.

### Effects on reflective well-being continuous model (intervention from -1 SD from average forgiveness)

|  |
| --- |
| Figure 5: Causal effects of religious loss on reflective well-being (continuous exposure) |

Table 5: Table of results for the reflective well-being domain (continuous exposure)

|  | E[Y(1)]-E[Y(0)] | 2.5 % | 97.5 % | E\_Value | E\_Val\_bound |
| --- | --- | --- | --- | --- | --- |
| Gratitude (sd) | 0.1285 | 0.1002 | 0.1591 | 1.497 | 1.416 |
| Pwi health (sd) | 0.0560 | 0.0335 | 0.0776 | 1.287 | 1.211 |
| Pwi relationships (sd) | 0.0731 | 0.0468 | 0.0990 | 1.340 | 1.257 |
| Pwi security (sd) | 0.0643 | 0.0381 | 0.0873 | 1.313 | 1.232 |
| Pwi standardliving (sd) | 0.0487 | 0.0228 | 0.0716 | 1.263 | 1.174 |
| Lifesat satlife (sd) | 0.0910 | 0.0643 | 0.1173 | 1.393 | 1.314 |
| Lifesat ideal (sd) | 0.0783 | 0.0544 | 0.1018 | 1.355 | 1.282 |

[Table 5](#tbl-results-reflective_con) presents the results for the reflective domain.

For the outcome ‘Gratitude (sd)’, the PATE causal contrast is 0.128. The confidence interval ranges from 0.1 to 0.159. The E-value for this outcome is 1.497, indicating reliable evidence for causality.

For the outcome ‘Lifesat satlife (sd)’, the PATE causal contrast is 0.091. The confidence interval ranges from 0.064 to 0.117. The E-value for this outcome is 1.393, indicating reliable evidence for causality.

For the outcome ‘Lifesat ideal (sd)’, the PATE causal contrast is 0.078. The confidence interval ranges from 0.054 to 0.102. The E-value for this outcome is 1.355, indicating reliable evidence for causality.

For the outcome ‘Pwi relationships (sd)’, the PATE causal contrast is 0.073. The confidence interval ranges from 0.047 to 0.099. The E-value for this outcome is 1.34, indicating reliable evidence for causality.

For the outcome ‘Pwi security (sd)’, the PATE causal contrast is 0.064. The confidence interval ranges from 0.038 to 0.087. The E-value for this outcome is 1.313, indicating reliable evidence for causality.

For the outcome ‘Pwi health (sd)’, the PATE causal contrast is 0.056. The confidence interval ranges from 0.034 to 0.078. The E-value for this outcome is 1.287, indicating reliable evidence for causality.

For the outcome ‘Pwi standardliving (sd)’, the PATE causal contrast is 0.049. The confidence interval ranges from 0.023 to 0.072. The E-value for this outcome is 1.263, indicating reliable evidence for causality.

### Effects social well-being continous model (intervention from -1 SD from average forgiveness)

|  |
| --- |
| Figure 6: Causal effects of religious loss on social well-being (continuous exposure) |

Table 6: Table of results for the reflective well-being domain (continuous exposure)

|  | E[Y(1)]-E[Y(0)] | 2.5 % | 97.5 % | E\_Value | E\_Val\_bound |
| --- | --- | --- | --- | --- | --- |
| Gratitude (sd) | 0.1285 | 0.1002 | 0.1591 | 1.497 | 1.416 |
| Pwi health (sd) | 0.0560 | 0.0335 | 0.0776 | 1.287 | 1.211 |
| Pwi relationships (sd) | 0.0731 | 0.0468 | 0.0990 | 1.340 | 1.257 |
| Pwi security (sd) | 0.0643 | 0.0381 | 0.0873 | 1.313 | 1.232 |
| Pwi standardliving (sd) | 0.0487 | 0.0228 | 0.0716 | 1.263 | 1.174 |
| Lifesat satlife (sd) | 0.0910 | 0.0643 | 0.1173 | 1.393 | 1.314 |
| Lifesat ideal (sd) | 0.0783 | 0.0544 | 0.1018 | 1.355 | 1.282 |

@ tbl-results-social-con for the social domain are as follows. For the outcome ‘Support noguidance (reversed, sd)’, the PATE causal contrast is 0.107. The confidence interval ranges from 0.074 to 0.139. The E-value for this outcome is 1.437, indicating reliable evidence for causality.

For the outcome ‘Belong accept (sd)’, the PATE causal contrast is 0.103. The confidence interval ranges from 0.071 to 0.135. The E-value for this outcome is 1.426, indicating reliable evidence for causality.

For the outcome ‘Permeability individual (sd)’, the PATE causal contrast is 0.087. The confidence interval ranges from 0.065 to 0.112. The E-value for this outcome is 1.381, indicating reliable evidence for causality.

For the outcome ‘Belong outsider (reversed, sd)’, the PATE causal contrast is 0.081. The confidence interval ranges from 0.056 to 0.104. The E-value for this outcome is 1.363, indicating reliable evidence for causality.

For the outcome ‘Support turnto (sd)’, the PATE causal contrast is 0.08. The confidence interval ranges from 0.05 to 0.11. The E-value for this outcome is 1.362, indicating reliable evidence for causality.

For the outcome ‘Support help (sd)’, the PATE causal contrast is 0.072. The confidence interval ranges from 0.046 to 0.099. The E-value for this outcome is 1.338, indicating reliable evidence for causality.

For the outcome ‘Belong beliefs (sd)’, the PATE causal contrast is 0.059. The confidence interval ranges from 0.027 to 0.09. The E-value for this outcome is 1.295, indicating reliable evidence for causality.

For the outcome ‘Neighbourhood community (sd)’, the PATE causal contrast is 0.043. The confidence interval ranges from 0.02 to 0.069. The E-value for this outcome is 1.245, indicating evidence for causality is weak.

For the outcome ‘Impermeability group (sd)’, the PATE causal contrast is -0.032. The confidence interval ranges from -0.063 to -0.004. The E-value for this outcome is 1.202, indicating evidence for causality is weak.

Here, we combined rigorous methods from causal epidemiology with national scale time-series data to estimate the causal effects of forgiveness on multidimensional well-being. We used doubly robust methods that combine propensity score weights with regression stratification. By controlling for measures of all outcomes as well as the exposure at baseline, our estimation approach reduces the probability of unmeasured confounding. Because we cannot know whether unmeasured confounding has been controlled, we report E-values, a sensitivity analysis that clarifies the “worst case” scenario for an unmeasured confounder to explain away the results.

### Generalisability and Transportability

These results generalise to the New Zealand population.

### Assumptions and Limitations

### Theoretical Relevance

This study is important both for its methods and findings.

Methods: - The bar for causality in this study very high…

Findings: - Forgiveness has total well-being effects.

### Future Research

### Author contributions

### Ethics Approval Details

The NZAVS is reviewed every three years by the University of Auckland Human Participants Ethics Committee. 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.

### Acknowledgements

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

## Appendix A. Measures

### Baseline confounding control

#### Age (waves: 1-15)

We asked participants’ age in an open-ended question (“What is your age?” or “What is your date of birth”).

#### Disability (waves: 5-15)

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

#### Education Attainment (waves: 1, 4-15)

Participants were asked “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 quals coded as Level 3, and all other ancillary categories coded as missing) See:https://www.nzqa.govt.nz/assets/Studying-in-NZ/New-Zealand-Qualification-Framework/requirements-nzqf.pdf

#### Employment (waves: 1-3, 4-11)

We asked participants “Are you currently employed? (This includes self-employed or casual work)”. \* note: This question disappeared in the updated NZAVS Technical documents (Data Dictionary).

#### European (waves: 1-15)

Participants were asked “Which ethnic group do you belong to (NZ census question)?” or “Which ethnic group(s) do you belong to? (Open-ended)” (wave: 3). Europeans were coded as 1, whereas other ethnicities were coded as 0.

#### Ethnicity (waves: 3)

Based on the New Zealand Cencus, we asked participants “Which ethnic group(s) do you belong to?”. The responses were: (1) New Zealand European; (2) Māori; (3) Samoan; (4) Cook Island Māori; (5) Tongan; (6) Niuean; (7) Chinese; (8) Indian; (9) Other such as DUTCH, JAPANESE, TOKELAUAN. Please state:. We coded their answers into four groups: Maori, Pacific, Asian, and Euro (except for Time 3, which used an open-ended measure).

#### Gender (waves: 1-15)

We asked participants’ gender in an open-ended question: “what is your gender?” or “Are you male or female?” (waves: 1-5). Female was coded as 0, Male was coded as 1, and gender diverse coded as 3 (Fraser et al. 2020). (or 0.5 = neither female nor male)

#### Income (waves: 1-3, 4-15)

Participants were asked “Please estimate your total household income (before tax) for the year XXXX”. To stablise this indicator, we first took the natural log of the response + 1, and then centred and standardised the log-transformed indicator.

#### Number of Children (waves: 1-3, 4-15)

We measured number of children using one item from Bulbulia (2015). We asked participants “How many children have you given birth to, fathered, or adopted. How many children have you given birth to, fathered, or adopted?” or ““How many children have you given birth to, fathered, or adopted. How many children have you given birth to, fathered, and/or parented?” (waves: 12-15).

#### Political Orientation

We measured participants’ political orientation using a single item adapted from Jost (2006).

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

(1 = Extremely Liberal to 7 = Extremely Conservative)

#### NZSEI-13 (waves: 8-15)

We assessed occupational prestige and status using the New Zealand Socio-economic Index 13 (NZSEI-13) (Fahy, Lee, and Milne 2017). This index uses the income, age, and education of a reference group, in this case the 2013 New Zealand census, to calculate an 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.

Participants were asked “If you are a parent, what is the birth date of your eldest child?”.

#### Living with Partner

Participants were asekd “Do you live with your partner?” (1 = Yes, 0 = No).

#### Living in an Urban Area (waves: 1-15)

We coded whether they are living in an urban or rural area (1 = Urban, 0 = Rural) based on the addresses provided.

We coded whether they are living in an urban or rural area (1 = Urban, 0 = Rural) based on the addresses provided.

#### NZ Deprivation Index (waves: 1-15)

We used the NZ Deprivation Index to assign each participant a score based on where they live (Atkinson, Salmond, and Crampton 2019). This score combines data such as income, home ownership, employment, qualifications, family structure, housing, and access to transport and communication for an area into one deprivation score.

#### NZ-Born (waves: 1-2,4-15)

We asked participants “Which country were you born in?” or “Where were you born? (please be specific, e.g., which town/city?)” (waves: 6-15).

#### Mini-IPIP 6 (waves: 1-3,4-15)

We measured participants personality with the Mini International Personality Item Pool 6 (Mini-IPIP6) (Chris G. Sibley et al. 2011) which consists of six dimensions and each dimensions is measured with four items:

1. agreeableness,
   1. I sympathize with others’ feelings.
   2. I am not interested in other people’s problems. (r)
   3. I feel others’ emotions.
   4. I am not really interested in others. (r)
2. conscientiousness,
   1. I get chores done right away.
   2. I like order.
   3. I make a mess of things. (r)
   4. I ften forget to put things back in their proper place. (r)
3. extraversion,
   1. I am the life of the party.
   2. I don’t talk a lot. (r)
   3. I keep in the background. (r)
   4. I talk to a lot of different people at parties.
4. honesty-humility,
   1. I feel entitled to more of everything. (r)
   2. I deserve more things in life. (r)
   3. I would like to be seen driving around in a very expensive car. (r)
   4. I would get a lot of pleasure from owning expensive luxury goods. (r)
5. neuroticism, and
   1. I have frequent mood swings.
   2. I am relaxed most of the time. (r)
   3. I get upset easily.
   4. I seldom feel blue. (r)
6. openness to experience
   1. I have a vivid imagination.
   2. I have difficulty understanding abstract ideas. (r)
   3. I do not have a good imagination. (r)
   4. I am not interested in abstract ideas. (r)

Each dimension was assessed with four items and participants rated the accuracy of each item as it applies to them from 1 (Very Inaccurate) to 7 (Very Accurate). Items marked with (r) are reverse coded.

#### Honesty-Humility-Modesty Facet (waves: 10-14)

Participants indicated the extent to which they agree with the following four statements from Campbell et al. (2004) , and Chris G. Sibley et al. (2011) (1 = Strongly Disagree to 7 = Strongly Agree)

i. I want people to know that I am an important person of high status, (Waves: 1, 10-14)  
ii. I am an ordinary person who is no better than others.  
iii. I wouldn't want people to treat me as though I were superior to them.  
iv. I think that I am entitled to more respect than the average person is.

### Exposure variable

#### Meaning of Life (waves: 10-15)

We assessed participants’ levels of life meaning using two items from Steger et al. (2006):

1. My life has a clear sense of purpose;  
2. I have a good sense of what makes my life meaningful.

Participants indicated their agreement with these items (1 = Strongly Disagree to 7 = Strongly Agree).

### Health well-being outcomes

#### Alcohol Frequency (waves: 6-15)

We measured participants’ frequency of drinking alcohol using one item adapted from Health (2013) . Participants were asked “How often do you have a drink containing alcohol?” (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).

#### Alcohol Intensity (waves: 6-15)

We measured participants’ intensity of drinking alcohol using one item adapted from (Health 2013). Participants were asked “How many drinks containing alcohol do you have on a typical day when drinking alcohol? (number of drinks on a typical day when drinking)”

#### Body Mass Index (waves: 2-3, 4-15)

Participants were asked “What is your height? (metres)” and “What is your weight? (kg)”. Based on participants indication of their height and weight we calculated the BMI by dividing the weight in kilograms by the square of the height in meters.

#### Short-Form Subjective Health (waves: 5-15)

Participants’ subjective health was assessed by three items selected from the MOS 36-item short-form health survey (Ware Jr and Sherbourne 1992). The items were

1. "In general, would you say your health is...";  
2. "I seem to get sick a little easier than most people.";  
3. "I expect my health to get worse." Participants responded to those items on a scale (1 = Poor to 7 = Excellent).

The second and third items were negatively-worded, so we reversed the responses.

#### Hours of Exercise (waves: 1, 4-15)

We measured hours of exercising using one item from Chris G. Sibley et al. (2011). We asked participants to estimate and report how many hours they spend in exercise/physical activity last week. To stablise this indicator, we first took the natural log of the response + 1, and then centred and standardised the log-transformed indicator.

#### Hours of Sleep (waves: 5-15)

Participants were asked “During the past month, on average, how many hours of *actual sleep* did you get per night”.

#### Smoker (waves: 4-15)

We asked participants whether they are currently smoking or not (1 = Yes or 0 = No), using a single item: “Do you currently smoke?” or “Do you currently smoke tobacco cigarettes?” (waves: 10-15) from Muriwai, Houkamau, and Sibley (2018).

### Embodied well-being outcomes

#### Kessler-6 (waves: 2-3,4-15)

We measured psychological distress using the Kessler-6 scale (R.  C. Kessler et al. 2002), which exhibits strong diagnostic concordance for moderate and severe psychological distress in large, cross-cultural samples (R. C. Kessler et al. 2010; Prochaska et al. 2012). Participants rated during the past 30 days, how often did… (

1. "... you feel hopeless";  
2. "... you feel so depressed that nothing could cheer you up";  
3. "... you feel restless or fidgety";  
4. "... you feel that everything was an effort";  
5. "... you feel worthless";  
6. " you feel nervous?"

Ordinal response options for the Kessler-6 are: “None of the time”; “A little of the time”; “Some of the time”; “Most of the time”; “All of the time.”

#### Fatigue (waves: 5-15)

We assessed subjective fatigue by asking participants, “During the last 30 days, how often did … you feel exhausted?” Responses were collected on an ordinal scale (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).

#### Rumination

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

Ordinal response options for the Kessler-6 are: “None of the time”; “A little of the time”; “Some of the time”; “Most of the time”; “All of the time.”

### Practical well-being outcomes

#### Body Satisfaction (waves: 2-3, 4-15)

We measured body satisfaction with one item from Stronge et al. (2015): “I am satisfied with the appearance, size and shape of my body”, which participants rated from 1 (very inaccurate) to 7 (very accurate).

#### Emotional Regulation (waves: 10-13)

We measured participants’ levels of emotional regulation using three items adpated from Gratz and Roemer (2004) and Gross and John (2003):

1. "When I feel negative emotions, my emotions feel out of control.";  
2. "When I feel negative emotions, I suppress or hide my emotions.";  
3. "When I feel negative emotions, I change the way I think to help me stay calm."

Participants were asked to indicate the extent to which they agree with these items (1 = Strongly Disagree to 7 = Strongly Agree).

#### Perfectionism (waves: 10-15)

We assessed participants’ perfectionism using three items from Rice, Richardson, and Tueller (2014): (1) Doing my best never seems to be enough; (2) My performance rarely measures up to my standards; (3) I am hardly ever satisfied with my performance. Participants indicated the extent to which they agree with these items (1 = Strongly Disagree to 7 = Strongly Agree).

#### Power Dependence

Participants’ Power dependence was measured using two items:

1." I do not have enough power or control over important parts of my life."  
2". Other people have too much power or control over important parts of my life.

Participants indicated their agreement with these items” (1 = Strongly Disagree to 7 = Strongly Agree).

#### Self-Control (waves: 5-15)

Participants were asked to indicate the extent to which they endorse the two items

1. "In general, I have a lot of self-control"  
2. "I wish I had more self-discipline"

The scale is from Tangney, Baumeister, and Boone (2004). The responses to the items ranged from 1 (Strongly Disagree) to 7 (Strongly Agree).

#### Self-Esteem (waves: 1-3, 4-15)

We measured participants’ self-esteem using three items adapted from Rosenberg (1965). Participants were instructed to circle the number that best represents how accurately each statement describes them. Participants responded to the items

1. "On the whole am satisfied with myself"  
2. "Take a positive attitude toward myself"  
3. "Am inclined to feel that I am a failure") on a likert-type scale (1 = Very inaccurate to 7 = Very accurate).

#### Sexual Satisfaction (waves: 10-15)

Participants were asked “How satisfied are you with your sex life?” (1 = Not satisfied to 7 = Very satisfied).

#### Vengeful Rumination (waves: 10-15)

We assessed participants’ vengeful rumination using three items, respectively adapted from Caprara (1986) and Berry et al. (2005), and developed for NZAVS: (1) Sometimes I can’t sleep because of thinking about past wrongs I have suffered; (2) I can usually forgive and forget when someone does me wrong; (3) I find myself regularly thinking about past times that I have been wronged. Participants indicated their agreement with these items (1 = Strongly Disagree to 7 = Strongly Agree). The values for the second item were reversely coded.

### Reflective well-being

#### Satisfaction with Life (waves: 1-3,4-15)

We measured life satisfaction with two items adapted from the Satisfaction with Life Scale (Diener et al. 1985):

1. "I am satisfied with my life" and  
2. "In most ways my life is close to ideal".

Participants responded on a scale from 1 (Strongly Disagree) to 7 (Strongly Agree).

#### Personal Wellbeing (waves: 1-3, 4-15)

We measured participants’ subjective wellbeing using three items from the Australian Unity Wellbeing Index (Cummins et al. 2003):

1. your health;  
2. Your standard of living;  
3. Your future security; 4 Your personal relationships.

Participants read an instruction (“The following items assess your current satisfaction with different aspects of your life and aspects of New Zealand more generally”) and indicated their satisfaction with those items (0 = Completely Dissatisfied to 10 = Completely Satisfied).

#### Standard Living

We measured participants’ satisfaction with their standard of living using an item from the Australian Unity Wellbeing Index (Cummins et al. 2003). Participants read an instruction (“Please rate your level of satisfaction with the following aspects of your life and New Zealand.”) and responded to an item

- "Your standard of living"

on a 10-point scale (0 = completely dissatisfied to 10 = completely satisfied).

### Social well-being outcomes

#### Charity Donation (waves: 1-3, 4-15)

We asked participants “How much money have you donated to charity in the last year?”. To stablise this indicator, we first took the natural log of the response + 1, and then centred and standardised the log-transformed indicator.

#### Felt Belongingness (waves: 1-3, 4-15)

We assessed felt belongingness with three items adapted from the Sense of Belonging Instrument (Hagerty and Patusky 1995):

1. "Know that people in my life accept and value me";  
2. "Feel like an outsider";

1. “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.

#### Ethnic group impermeability (waves: 9-13)

The current income gap between New Zealand Europeans and other ethnic groups would be very hard to change.

#### Individual Permeability (waves: 9-13)

Participants indicated the extent to which they agree with the statement, “I believe I am capable, as an individual of improving my status in society.”, from Tausch, Saguy, and Bryson (2015) (1 = Strongly Disagree to 7 = Strongly Agree).

#### Sense of Community (waves: 6-15)

We measured sense of community with a single item from Sengupta et al. (2013): “I feel a sense of community with others in my local neighbourhood.” Participants answered on a scale of 1 (strongly disagree) to 7 (strongly agree).

#### Support (waves: 1-3, 4-15)

Participants’ perceived social support was measured using three items from Cutrona and Russell (1987) and Williams, Cheung, and Choi (2000):

1. "There are people I can depend on to help me if I really need it";  
2. "There is no one I can turn to for guidance in times of stress";  
3. "I know there are people I can turn to when I need help."

Participants indicated the extent to which they agree with those items (1 = Strongly Disagree to 7 = Strongly Agree).

The second item was negatively-worded, so we reversely recorded the responses to this item.

APPENDIX B. Sample {.appendix}

| |baseline |  
|:-----------------------------|:-------------------|  
| |(N=34748) |  
|AGE | |  
|Mean (SD) |50.5 (13.5) |  
|Median [Min, Max] |53.0 [18.1, 95.5] |  
|AGREEABLENESS | |  
|Mean (SD) |5.37 (0.982) |  
|Median [Min, Max] |5.50 [1.00, 7.00] |  
|Missing |313 (0.9%) |  
|BORN\_NZ | |  
|0 |7251 (20.9%) |  
|1 |27411 (78.9%) |  
|Missing |86 (0.2%) |  
|CHILDREN\_NUM | |  
|Mean (SD) |1.76 (1.44) |  
|Median [Min, Max] |2.00 [0, 14.0] |  
|Missing |13 (0.0%) |  
|CONSCIENTIOUSNESS | |  
|Mean (SD) |5.14 (1.04) |  
|Median [Min, Max] |5.25 [1.00, 7.00] |  
|Missing |306 (0.9%) |  
|EDUCATION\_LEVEL\_COARSEN | |  
|Mean (SD) |3.48 (1.47) |  
|Median [Min, Max] |4.00 [1.00, 7.00] |  
|Missing |259 (0.7%) |  
|ETH\_CAT | |  
|Mean (SD) |1.28 (0.710) |  
|Median [Min, Max] |1.00 [1.00, 4.00] |  
|Missing |260 (0.7%) |  
|EXTRAVERSION | |  
|Mean (SD) |3.88 (1.19) |  
|Median [Min, Max] |3.75 [1.00, 7.00] |  
|Missing |306 (0.9%) |  
|FORGIVENESS | |  
|Mean (SD) |5.08 (1.26) |  
|Median [Min, Max] |5.33 [1.00, 7.00] |  
|FORGIVENESS\_COARSEN | |  
|q1 |10144 (29.2%) |  
|q2 |9602 (27.6%) |  
|q3 |7830 (22.5%) |  
|q4 |7172 (20.6%) |  
|GRATITUDE | |  
|Mean (SD) |5.91 (0.875) |  
|Median [Min, Max] |6.00 [1.00, 7.00] |  
|Missing |5 (0.0%) |  
|HLTH\_DISABILITY | |  
|Mean (SD) |0.231 (0.422) |  
|Median [Min, Max] |0 [0, 1.00] |  
|Missing |600 (1.7%) |  
|HONESTY\_HUMILITY | |  
|Mean (SD) |5.48 (1.15) |  
|Median [Min, Max] |5.75 [1.00, 7.00] |  
|Missing |310 (0.9%) |  
|HOURS\_CHILDREN\_LOG | |  
|Mean (SD) |1.10 (1.58) |  
|Median [Min, Max] |0 [0, 5.13] |  
|Missing |960 (2.8%) |  
|HOURS\_HOUSEWORK\_LOG | |  
|Mean (SD) |2.15 (0.769) |  
|Median [Min, Max] |2.20 [0, 5.13] |  
|Missing |960 (2.8%) |  
|HOURS\_WORK\_LOG | |  
|Mean (SD) |2.64 (1.59) |  
|Median [Min, Max] |3.56 [0, 4.62] |  
|Missing |960 (2.8%) |  
|HOUSEHOLD\_INC\_LOG | |  
|Mean (SD) |11.4 (0.758) |  
|Median [Min, Max] |11.5 [0.693, 14.9] |  
|Missing |1604 (4.6%) |  
|LIFESAT\_IDEAL | |  
|Mean (SD) |4.93 (1.43) |  
|Median [Min, Max] |5.00 [1.00, 7.00] |  
|Missing |1886 (5.4%) |  
|LIFESAT\_SATLIFE | |  
|Mean (SD) |5.71 (1.19) |  
|Median [Min, Max] |6.00 [1.00, 7.00] |  
|Missing |410 (1.2%) |  
|MALE | |  
|0 |12592 (36.2%) |  
|1 |22156 (63.8%) |  
|MEANING\_PURPOSE | |  
|Mean (SD) |5.22 (1.41) |  
|Median [Min, Max] |5.00 [1.00, 7.00] |  
|Missing |814 (2.3%) |  
|MEANING\_SENSE | |  
|Mean (SD) |5.74 (1.20) |  
|Median [Min, Max] |6.00 [1.00, 7.00] |  
|Missing |114 (0.3%) |  
|MODESTY | |  
|Mean (SD) |6.02 (0.919) |  
|Median [Min, Max] |6.25 [1.00, 7.00] |  
|Missing |17 (0.0%) |  
|NEUROTICISM | |  
|Mean (SD) |3.46 (1.15) |  
|Median [Min, Max] |3.50 [1.00, 7.00] |  
|Missing |315 (0.9%) |  
|NZ\_DEP2018 | |  
|Mean (SD) |4.70 (2.70) |  
|Median [Min, Max] |4.00 [1.00, 10.0] |  
|Missing |239 (0.7%) |  
|NZSEI13 | |  
|Mean (SD) |54.8 (16.4) |  
|Median [Min, Max] |56.0 [10.0, 90.0] |  
|Missing |317 (0.9%) |  
|OPENNESS | |  
|Mean (SD) |4.98 (1.12) |  
|Median [Min, Max] |5.00 [1.00, 7.00] |  
|Missing |308 (0.9%) |  
|PARTNER | |  
|0 |8158 (23.5%) |  
|1 |25163 (72.4%) |  
|Missing |1427 (4.1%) |  
|POL\_ORIENT | |  
|Mean (SD) |3.57 (1.39) |  
|Median [Min, Max] |4.00 [1.00, 7.00] |  
|Missing |2086 (6.0%) |  
|POL\_WING | |  
|Mean (SD) |3.70 (1.31) |  
|Median [Min, Max] |4.00 [1.00, 7.00] |  
|Missing |2396 (6.9%) |  
|PWI\_HEALTH | |  
|Mean (SD) |6.79 (2.30) |  
|Median [Min, Max] |7.00 [0, 10.0] |  
|Missing |128 (0.4%) |  
|PWI\_RELATIONSHIPS | |  
|Mean (SD) |7.78 (2.21) |  
|Median [Min, Max] |8.00 [0, 10.0] |  
|Missing |127 (0.4%) |  
|PWI\_SECURITY | |  
|Mean (SD) |6.35 (2.33) |  
|Median [Min, Max] |7.00 [0, 10.0] |  
|Missing |108 (0.3%) |  
|PWI\_STANDARDLIVING | |  
|Mean (SD) |7.67 (2.00) |  
|Median [Min, Max] |8.00 [0, 10.0] |  
|Missing |125 (0.4%) |  
|RELIGION\_CHURCH\_BINARY2 | |  
|Mean (SD) |0.159 (0.366) |  
|Median [Min, Max] |0 [0, 1.00] |  
|Missing |300 (0.9%) |  
|RELIGION\_IDENTIFICATION\_LEVEL | |  
|Mean (SD) |2.35 (2.17) |  
|Median [Min, Max] |1.00 [1.00, 7.00] |  
|Missing |458 (1.3%) |  
|SAMPLE\_ORIGIN\_YEAR | |  
|1-2 |2251 (6.5%) |  
|3-3.5 |1693 (4.9%) |  
|4 |2023 (5.8%) |  
|5-6-7 |3246 (9.3%) |  
|8-9 |4372 (12.6%) |  
|10 |21163 (60.9%) |  
|SAMPLE\_WEIGHTS | |  
|Mean (SD) |0.987 (0.527) |  
|Median [Min, Max] |0.892 [0.617, 2.59] |  
|TOTAL\_SIBLINGS | |  
|Mean (SD) |2.53 (1.81) |  
|Median [Min, Max] |2.00 [0, 23.0] |  
|Missing |747 (2.1%) |  
|URBAN | |  
|Mean (SD) |1.81 (0.393) |  
|Median [Min, Max] |2.00 [1.00, 2.00] |  
|Missing |235 (0.7%) |

## Appendix C Propensity score analysis

### Propensity score analysis for health

Following (Thoemmes and Kim 2011) we describe our method for obtaining and verifying our propensity score analysis using the WeightIt and Cobalt packages in R. z

**1. Information about data collection**

Information about data collection in the New Zealand Attitudes and Values Study can be obtained from (Chris G. Sibley 2021).

**2. List of all covariates used to estimate the propensity score**

The baseline covariates used in this study are detailed in Appendix A. These include:

- male  
- age  
- education\_level\_coarsen  
- eth\_cat  
- employed  
- nz\_dep2018  
- nzsei13  
- total\_siblings  
- born\_nz  
- hlth\_disability  
- partner  
- parent  
- religion\_religious  
- religion\_identification\_level  
- religion\_church\_binary  
- pol\_orient  
- pol\_wing  
- sample\_origin  
- urban  
- children\_num  
- household\_inc\_log  
- hours\_children\_log  
- hours\_work\_log  
- hours\_housework\_log  
- agreeableness  
- conscientiousness  
- extraversion  
- honesty\_humility  
- openness  
- neuroticism  
- modesty

1. **Method for determining the set of covariates**

Covariates were selected based on their likelihood of association with the exposure and the outcome, or with an unmeasured confounder.

1. **Inclusion of polynomial or interaction terms**

Following the guidance of “Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique” (n.d.), we included an interaction term for the exposure and baseline covariates.

1. **Estimation method for propensity scores**

Standard inverse probability weighting and the ebalance method from the WeightIt package were used for estimation. The ebalance method consistently performed better and its performance is reported here.

1. **Conditioning strategy**

A combination of weighting and stratification was used to obtain doubly robust estimation.

1. **Region of common support**

We did not use histograms to assess regions of common support as we did not apply matching. However, both the propensity score analysis and descriptive results in Appendix A show very good overlap.

1. **Details on weighting**

We included post-stratification census weights after the WeightIt method to obtain a population estimate for New Zealand. This method multiplies the propensity scores by the census weights to obtain a single vector of weights for all participants. We used the age x gender x nzeuropean census weights (Sibley, 2021).

1. **Sample size before and after conditioning**

The effective sample sizes before and after weighting are reported below.

1. **Standardized difference before and after matching**

Below, we report standardised differences before and after matching on the propensity score.

1. **Point estimate of treatment effect and associated standard error**

These are reported in the main results.

1. **Inclusion of covariates in outcome model**

All those used in the exposure (propensity score model), interacted with the exposure. Additionally, we weighted the regression using the final output of the WeightIt and MatchThem`package (Greifer 2023b, 2023a; Pishgar et al. 2021), which multiplies propensity scores x census weights.

|  |
| --- |
| Figure 7: Love plot for propensity score analysis: health outcomes. |

|  |
| --- |
| Figure 8: Distribution of propensity scores by condition: health domain |

**Summary of propensity score weights: health well-being domain**

### Propensity score analysis for embodied well-being

|  |
| --- |
| Figure 9: Love plot for propensity score analysis: embodied domain |

|  |
| --- |
| Figure 10: Distribution of propensity scores by condition: embodied domain |

**Summary of propensity score weights: embodied well-being domain**

### Propensity score analysis for practical well-being

|  |
| --- |
| Figure 11: Love plot for propensity score analysis: practical domain |

[Figure 12](#fig-propensity-dis-practical) shows propensity score weights in the practical domain study

|  |
| --- |
| Figure 12: Distribution of propensity scores by condition: practical domain |

**Summary of propensity score weights: Practical well-being domain**

|  |
| --- |
| Figure 13: Love plot for propensity score analysis: reflective domain |

### **Summary of propensity score weights: reflective well-being domain**

|  |
| --- |
| Figure 14: Love plot for propensity score analysis: social domain |

**Summary of propensity score weights: social well-being domain**

## Appendix D. Multiple comparisons in outcomewide studies

The concern for multiple comparisons is legitimate in many research settings. However, there are compelling reasons not to adjust for it in the case of outcome-wide science, as proposed by Tyler VanderWeele (Tyler J. VanderWeele, Mathur, and Chen 2020).

1. **Nature of the analysis:** Outcome-wide studies are inherently exploratory. They aim to generate hypotheses rather than testing pre-specified ones. In such a scenario, adjusting for multiple comparisons is out of place. Such testing might limit our ability to discover.
2. **False negatives vs. false positives:** Adjusting for multiple comparisons often results in an increased risk of Type II errors (false negatives). In the context of public health, false negatives could be more problematic than false positives. We might overlook potentially significant associations that could lead to beneficial interventions.
3. **Independence of outcomes:** The standard corrections for multiple comparisons, such as the Bonferroni or the Holm method, assume that tests are independent. In an outcome-wide study, outcomes are likely to be correlated, so these corrections could be overly conservative.
4. **Magnitude of effects:** Outcome-wide studies do not only focus on p-values, but also the magnitude of effects, confidence intervals, and their scientific or clinical significance. We advocate assessing E-values, or unmeasured confounding, in place of assessing p-values. Adjusting for multiple comparisons focuses primarily on p-values, potentially undermining the importance of effect sizes. P-values are often a measure of sample size.
5. **Replication and robustness:** Findings from outcome-wide studies are not intended to be conclusive, but rather to guide further research. Consequently, potential false positives should be addressed in future replication studies and through robustness checks.

For for outcome-wide studies, then, p-value corrections may limit the capacity to generate new hypotheses, increase the risk of missing potential public health interventions, and over-emphasize p-values at the expense of sensitivity analyses and E-values. In causal inference, the main worry is assessing the robustness of results to unmeasured confounding.

## Appendix E. Population Average Treatment Effect

As indicated in the main manuscript, the Average Treatment Effects is obtained by contrasting the expected outcome when a population sampled is exposed to an exposure level, , with the expected outcome under a different exposure level, .

For a binary treatment with levels and , the Average Treatment Effect (ATE), on the difference scale, is expressed:

On the risk ratio scale, the ATE is expressed:

Other effect scales, such as the incidence rate ratio, incidence rate difference, or hazard ratio, might also be of interest.

Here we estimate the Population Average Treatment Effect (PATE), which denotes the effect the treatment would have on the New Population if applied universally. This quantity can be expressed:

where is a function that incorporates post-stratification weights into the estimation of the expected outcomes from which we obtain causal contrasts. Because the NZAVS is national probability sample, i.e. inverse probability of being sampled 1. However, to incorporate gender, age, and ethnic differences we include post-stratification weight into our outcome wide models.

## References

“Agnostic Notes on Regression Adjustments to Experimental Data: Reexamining Freedman’s Critique.” n.d. <https://projecteuclid.org/journals/annals-of-applied-statistics/volume-7/issue-1/Agnostic-notes-on-regression-adjustments-to-experimental-data--Reexamining/10.1214/12-AOAS583.full>.

Atkinson, J, C Salmond, and P Crampton. 2019. “NZDep2018 Index of Deprivation, User’s Manual.” Wellington.

Berry, Jack W., Everett L. Worthington Jr., Lynn E. O’Connor, Les Parrott III, and Nathaniel G. Wade. 2005. “Forgivingness, Vengeful Rumination, and Affective Traits.” *Journal of Personality* 73 (1): 183–226. <https://doi.org/10.1111/j.1467-6494.2004.00308.x>.

Bulbulia, Shaver, J. A. 2015. “Religion and Parental Cooperation: An Empirical Test of Slone’s Sexual Signaling Model.” In *The Attraction of Religion: A Sexual Selectionist Account*, edited by & Van Slyke J. Slone D., 29–62. Bloomsbury Press.

Campbell, W. Keith, Angelica M. Bonacci, Jeremy Shelton, Julie J. Exline, and Brad J. Bushman. 2004. “Psychological entitlement: interpersonal consequences and validation of a self-report measure.” *Journal of Personality Assessment* 83 (1): 29–45. <https://doi.org/10.1207/s15327752jpa8301_04>.

Caprara, Gian Vittorio. 1986. “Indicators of Aggression: The Dissipation-Rumination Scale.” *Personality and Individual Differences* 7 (6): 763–69. <https://doi.org/10.1016/0191-8869(86)90074-7>.

Cummins, Robert A., Richard Eckersley, Julie Pallant, Jackie van Vugt, and RoseAnne Misajon. 2003. “Developing a National Index of Subjective Wellbeing: The Australian Unity Wellbeing Index.” *Social Indicators Research* 64 (2): 159–90. <https://doi.org/10.1023/A:1024704320683>.

Cutrona, Carolyn E, and Daniel Wayne Russell. 1987. “The Provisions of Social Relationships and Adaptation to Stress.” *Advances in Personal Relationships* 1: 37–67.

Danaei, Goodarz, Mohammad Tavakkoli, and Miguel A. Hernán. 2012. “Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins.” *American Journal of Epidemiology* 175 (4): 250–62. <https://doi.org/10.1093/aje/kwr301>.

Diener, Ed, Robert A Emmons, Randy J Larsen, and Sharon Griffin. 1985. “The Satisfaction With Life Scale.” *Journal of Personality Assessment* 49 (1): 71–75.

Fahy, Katie M., Alan Lee, and Barry J. Milne. 2017. *New Zealand Socio-Economic Index 2013*. Wellington, New Zealand: Statistics New Zealand-Tatauranga Aotearoa.

Fraser, Gloria, Joseph Bulbulia, Lara M. Greaves, Marc S. Wilson, and Chris G. Sibley. 2020. “Coding Responses to an Open-Ended Gender Measure in a New Zealand National Sample.” *The Journal of Sex Research* 57 (8): 979–86. <https://doi.org/10.1080/00224499.2019.1687640>.

Gratz, Kim L., and Lizabeth Roemer. 2004. “Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale.” *Journal of Psychopathology and Behavioral Assessment* 26 (1): 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>.

Greifer, Noah. 2023a. *Cobalt: Covariate Balance Tables and Plots*.

———. 2023b. *WeightIt: Weighting for Covariate Balance in Observational Studies*.

Greifer, Noah, Steven Worthington, Stefano Iacus, and Gary King. 2023. *Clarify: Simulation-Based Inference for Regression Models*. <https://iqss.github.io/clarify/>.

Gross, James J., and Oliver P. John. 2003. “Individual Differences in Two Emotion Regulation Processes: Implications for Affect, Relationships, and Well-Being.” *Journal of Personality and Social Psychology* 85 (2): 348–62. <https://doi.org/10.1037/0022-3514.85.2.348>.

Hagerty, Bonnie M. K., and Kathleen Patusky. 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.

Hernan, M. A., and J. M. Robins. 2023. *Causal Inference*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. Taylor & Francis. <https://books.google.co.nz/books?id=\_KnHIAAACAAJ>.

Hernán, Miguel A, Brian C Sauer, Sonia Hernández-Díaz, Robert Platt, and Ian Shrier. 2016. “Specifying a Target Trial Prevents Immortal Time Bias and Other Self-Inflicted Injuries in Observational Analyses.” *Journal of Clinical Epidemiology* 79: 7075.

Jost, John T. 2006. “The End of the End of Ideology.” *American Psychologist* 61 (7): 651–70. <https://doi.org/10.1037/0003-066X.61.7.651>.

Kessler, R. C., G. Andrews, L. J. Colpe, E. Hiripi, D. K. Mroczek, S.-L. T. Normand, E. E. Walters, and A. M. Zaslavsky. 2002. “Short Screening Scales to Monitor Population Prevalences and Trends in Non-Specific Psychological Distress.” *Psychological Medicine* 32 (6): 959–76. <https://doi.org/10.1017/S0033291702006074>.

Kessler, Ronald C., Jennifer Greif Green, Michael J. Gruber, Nancy A. Sampson, Evelyn Bromet, Marius Cuitan, Toshi A. Furukawa, et al. 2010. “Screening for Serious Mental Illness in the General Population with the K6 Screening Scale: Results from the WHO World Mental Health (WMH) Survey Initiative.” *International Journal of Methods in Psychiatric Research* 19 (S1): 4–22. <https://doi.org/10.1002/mpr.310>.

Muriwai, E., C. A. Houkamau, and C. G. Sibley. 2018. “Looking Like a Smoker, a Smokescreen to Racism? Māori Perceived Appearance Linked to Smoking Status.” *Ethnicity & Health* 23 (4): 353–66. <https://doi.org/10.1080/13557858.2016.1263288>.

Pishgar, Farhad, Noah Greifer, Clémence Leyrat, and Elizabeth Stuart. 2021. “MatchThem:: Matching and Weighting After Multiple Imputation.” *The R Journal* 13 (2): 292305. <https://doi.org/10.32614/RJ-2021-073>.

Prochaska, Judith J., Hai-Yen Sung, Wendy Max, Yanling Shi, and Michael Ong. 2012. “Validity Study of the K6 Scale as a Measure of Moderate Mental Distress Based on Mental Health Treatment Need and Utilization: The K6 as a Measure of Moderate Mental Distress.” *International Journal of Methods in Psychiatric Research* 21 (2): 88–97. <https://doi.org/10.1002/mpr.1349>.

Rice, Kenneth G., Clarissa M. E. Richardson, and Stephen Tueller. 2014. “The Short Form of the Revised Almost Perfect Scale.” *Journal of Personality Assessment* 96 (3): 368–79. <https://doi.org/10.1080/00223891.2013.838172>.

Rosenberg, M. 1965. *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Sengupta, Nikhil K, Nils Luyten, Lara M Greaves, Danny Osborne, Andrew Robertson, Colmar Brunton, Gavin Armstrong, and Chris G Sibley. 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, Chris G. 2021. “Sampling Procedure and Sample Details for the New Zealand Attitudes and Values Study.” <https://doi.org/10.31234/osf.io/wgqvy>.

Sibley, Chris G, Nils Luyten, Missy Purnomo, Annelise Mobberley, Liz W Wootton, Matthew D Hammond, Nikhil Sengupta, et al. 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): 142–59.

Steger, Michael F., Patricia Frazier, Shigehiro Oishi, and Matthew Kaler. 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>.

Stronge, Samantha, Lara M. Greaves, Petar Milojev, Tim West-Newman, Fiona Kate Barlow, and Chris G. Sibley. 2015. “Facebook Is Linked to Body Dissatisfaction: Comparing Users and Non-Users.” *Sex Roles: A Journal of Research* 73 (5): 200–213. <https://doi.org/10.1007/s11199-015-0517-6>.

Tangney, June P., Roy F. Baumeister, and Angie Luzio Boone. 2004. “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success.” *Journal of Personality* 72 (2): 271–324. <https://doi.org/10.1111/j.0022-3506.2004.00263.x>.

Tausch, Nicole, Tamar Saguy, and Jeff Bryson. 2015. “How Does Intergroup Contact Affect Social Change? Its Impact on Collective Action and Individual Mobility Intentions Among Members of a Disadvantaged Group.” *Journal of Social Issues* 71 (3): 536–53. <https://doi.org/10.1111/josi.12127>.

Thoemmes, Felix J., and Eun Sook Kim. 2011. “A Systematic Review of Propensity Score Methods in the Social Sciences.” *Multivariate Behavioral Research* 46 (1): 90–118. <https://doi.org/10.1080/00273171.2011.540475>.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. CRC press.

VanderWeele, Tyler J. 2022. “Constructed Measures and Causal Inference: Towards a New Model of Measurement for Psychosocial Constructs.” *Epidemiology* 33 (1): 141. <https://doi.org/10.1097/EDE.0000000000001434>.

VanderWeele, Tyler J, Maya B Mathur, and Ying Chen. 2020. “Outcome-Wide Longitudinal Designs for Causal Inference: A New Template for Empirical Studies.” *Statistical Science* 35 (3): 437466.

Verbrugge, Lois M. 1997. “A Global Disability Indicator.” *Journal of Aging Studies* 11 (4): 337–62. <https://doi.org/10.1016/S0890-4065(97)90026-8>.

Ware Jr, John E, and Cathy Donald Sherbourne. 1992. “The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual Framework and Item Selection.” *Medical Care*, 473483.

Westreich, Daniel, Jessie K Edwards, Stephen R Cole, Robert W Platt, Sunni L Mumford, and Enrique F Schisterman. 2015. “Imputation Approaches for Potential Outcomes in Causal Inference.” *International Journal of Epidemiology* 44 (5): 17311737.

Williams, Kipling D., Christopher K. T. Cheung, and Wilma Choi. 2000. “Cyberostracism: Effects of Being Ignored over the Internet.” *Journal of Personality and Social Psychology* 79 (5): 748–62. <https://doi.org/10.1037/0022-3514.79.5.748>.

Zhang, Jiaxin, S. Ghazaleh Dashti, John B. Carlin, Katherine J. Lee, and Margarita Moreno-Betancur. 2023. “Should Multiple Imputation Be Stratified by Exposure Group When Estimating Causal Effects via Outcome Regression in Observational Studies?” *BMC Medical Research Methodology* 23 (1): 42. <https://doi.org/10.1186/s12874-023-01843-6>.