Causal Effects of Religious Service Attendance on Charity and Volunteering: Evidence Using Novel Measures From A National Longitudinal Panel

Joseph A. Bulbulia

Don E Davis

Kenneth G. Rice

Chris G. Sibley

Geoffrey Troughton

2024-04-21

Abstract

Causal investigations for the effects of religion on prosociality must be precise. One should articulate a specific causal contrast for a feature of religion, select appropriate prosociality measures, define the target population, gather time-series data, and, only after identification assumptions are met, conduct statistical and sensitivity analyses. Here, we examine three distinct interventions on religious service attendance (increase, decrease, maintain) in a longitudinal sample of 33,198 New Zealanders (years 2018 to 2021). Study 1 investigates effects of religious service on charitable contributions and volunteerism. Studies 2 and 3 investigate effects of religious service on the relative risks of *receiving* aid or financial support from others during the past week – measures designed to minimise self-reporting bias. Across all studies, inferred causal effects are substantially less pronounced than observed cross-sectional associations. Nonetheless, regular attendance across the population would enhance charitable donations by 4% of the New Zealand Government's annual spending. This research underscores the essential role of formulating precise causal questions and recommends a workflow for answering them in the scientific study of cultural practices.

Introduction

A central question in the scientific study of religion is whether religion fosters prosociality (D. D. Johnson 2005; Norenzayan et al. 2016; J. Watts et al. 2016; Joseph Watts et al. 2015; Sosis and Bressler 2003; Whitehouse et al. 2023; Schloss and Murray 2011; Swanson 1967; De Coulanges 1903; Wheatley 1971). However, quantifying causal effects for religion, and many other social behaviours presents significant challenges. Investigators have limited scope to randomise supernatural beliefs, community worship, and personal prayer. On the other hand, valid causal inferences from non-experimental or observational data must combine high-resolution repeated-measures time-series data with robust methods for causal inference. Few studies meet this standard. Indeed, a recent survey of the religion and prosociality literature reveals that nearly all observational studies assessing links between prosociality and religion are associational Kelly, Kramer, and Shariff (2024). To our knowledge, studies that draw on longitudinal (panel) data have yet to appropriately leverage their repeated measures data to obtain reliable causal inferences.

An encouraging recent attempt to obtain valid causal inference is Major-Smith's thoughtful investigation of the relationships between religious attendance, beliefs, and affiliation on blood donations among pregnant women and

their partners who were residents of Bristol, United Kingdom, in the early 1990s and participated in the Avon Longitudinal Study of Parents and Children, $N = 13,477$ mothers and $N = 13,424$ partners (Major-Smith 2023). Major-Smith's study begins with a careful overview of the threats to causal inference from confounding and selection bias. Next, a series of cross-sectional regression analyses observe associations between the stated features of religion and self-reported blood donations in this Bristol cohort. Unfortunately, because the outcome is a retrospective question asked during pregnancy about past blood donations, the analysis cannot capitalise on the time-series data of the Avon panel. There is neither control for baseline measures of religion variables (the treatments) nor for baseline blood donations (the outcome), and the study cannot evaluate outcomes after initiation into religious service or its cessation. Although we may sometimes use cross-sectional associations to obtain credible suggestions about causality, we cannot typically attach causal interpretations, at least not without strong assumptions about the relative order and timing of events (Tyler J. VanderWeele 2021a). Indeed, below we report an analysis restricted to baseline New Zealand Attitudes and Values Study data that observes a 2.89 times overstatement of monthly vs zero charitable donations causal contrast and a 2.58 times overstatement of the monthly vs zero volunteering causal contrast. A virtue of Major-Smith (2023), however, is that the study makes its causal assumptions explicit, and provides several sensitivity analyses to frame its results.

Here, to obtain causal inferences from time-series data, we leverage comprehensive panel data from 33,198 participants in the New Zealand Attitudes and Values Study from 2018-2021 to quantify the effects of clearly defined interventions in religious attendance across the population of New Zealanders on two features of prosociality: charitable financial donations and volunteering, measured by stated charitable donations and volunteering as well as by revealed measures of help received from the community. We define a causal effect as a quantitative contrast between mean outcomes for a population from pre-specified interventions. These outcomes are called "counterfactual" or "potential" outcomes, terms we use interchangeably (Rubin 2005; Splawa-Neyman, Dabrowska, and Speed 1990; J. Robins 1986; Pearl 2009; Van Der Laan and Rose 2018).[1] A fundamental challenge in observational studies is to ensure *balance* between the variables in interventions or "treatments" to be compared that might affect both treatment and the potential outcomes under treatment (Shiba and Kawahara 2021). We call the state of imbalance *confounding*, and the strategy for ensuring balance, *confounding control.* In this study, we express the interventions on religious service as "modified treatment policies" (Haneuse and Rotnitzky 2013; Díaz et al. 2023, 2021; Hoffman et al. 2023). We obtain causal inferences by contrasting inferred population averages under different modified treatment policies.

Our initial causal contrast investigates: "What would be the average difference across the New Zealand population if everyone attended religious services regularly (at least four times per month) versus if no one attended?" This theoretical question simulates a hypothetical experiment with random assignment to regular or non-attendance. This contrast addresses the scientifically interesting all/none contrast that is common in experimental designs (Hernán et al. 2016).

A second causal contrast investigates: "What would be the average difference across the New Zealand population if everyone attended religious services regularly compared with maintaining the status quo?" Here, we contrast regular religious service and New Zealand society as it was at the end of the study (2021) without modification. This causal contrast may inform practical policies relevant to non-regular attendees who might start.

Our third causal contrast examines: "What would be the average difference across the New Zealand population if no one attended religious services compared with the status quo?" Here, we contrast the loss of any religious service with New Zealand society as it was at the end of the study (2021), again without modification. This causal contrast may inform practical policies relevant to regular attendees who stop attending.

Although the set of causal contrasts analysts might consider is unbounded, those we have selected address these specific scientific and policy interests.

Note that our approach does not focus on testing specific hypotheses; instead, we aim to compute our pre-specified causal contrasts with high accuracy by combining appropriate time-series data and robust methods for causal in-

---

[1]Philosophical disagreements about the meanings assigned to "potential" and "counterfactual" outcomes do not affect our use.

ference (Hernán and Greenland 2024).

Method

Sample

Data were collected by the New Zealand Attitudes and Values Study (NZAVS), an annual longitudinal national probability panel study of social attitudes, personality, ideology, and health outcomes in New Zealand. Chris G. Sibley started the New Zealand Attitudes and Values Study in 2009, which has grown to include a community of over fifty researchers. Since its inception, The New Zealand Attitudes and Values Study has accumulated questionnaire responses from 72,910 New Zealand residents. The study operates independently of political or corporate funding and is based in a university setting. Data summaries for our study sample on all measures used in this study are found in **Appendices B-D**. For more details about the New Zealand Attitudes and Values Study see: OSF.IO/75SNB.

Treatment Indicator

Religious service attendance is assessed in the New Zealand Attitudes and Values Study with the following questions:

- *Do you identify with a religion and/or spiritual group? If yes...How many times did you attend a church or place of worship during the last month?*

We rounded responses to the nearest whole number. Because there were few responses greater than eight, we coded all above eight as eight (see *Appendix B*: we label this variable religion_church_round). Note that contrasts with responses greater than four were not intervened upon in the regular church service condition; we decided to ease the computational burden during estimation.

Measures of Prosociality

Study 1: Self-reported charity The New Zealand Attitudes and Values Study includes two self-reported measures of pro-sociality:

- Volunteering: *"Please estimate how many hours you spent doing each of the following things last week...Volunteer/charitable work"*
- Annual charitable financial donations: *"How much money have you donated to charity in the last year?"*

Study 2: Help received from others in the last week: *time*

Participants were asked:

"Please estimate how much help you have received from the following sources in the last week."

- *Family...TIME (hours)*
- *Friends...TIME (hours)*
- *Community...TIME (hours)*

Owing to the high variability of responses, we transformed responses into binary indicators: *0 = none/ 1 = any.*

Study 3: Help received from others in the last week: *money*

Similarly, participants were asked:

"Please estimate how much help you have received from the following sources in the last week."

- *Family...MONEY (dollars)*
- *Friends...MONEY (dollars)*
- *Community...MONEY (dollars)*

These measures were also highly variable. Hence, we converted responses to binary indicators: *0 = none/ 1 = any*.

Studies 2 and 3 aim to minimise self-presentation bias by using revealed measures of prosocial exposure. Our approach assumes that if religious institutions foster prosociality, the initiation of regular attendance –controlling for past religious service, past measures of the prosocial outcomes, and a rich array of demographic, personality, and health measures recorded at baseline – will increase exposure to prosocial behaviours. Notably, our revealed measures of prosociality rely on stated help received are robust to self-presentation biases that might associate religious service attendance with indicators of prosociality in the absence of causation. Such an association could occur if initiation into treatment affected the subsequent error term of one or more of the outcomes we contrast. Note that we include baseline outcomes and treatments in all studies, which mitigates the threat of undirected correlated errors.

We provide comprehensive details of all measures in **Appendix A**.

Causal Interventions

We define three targeted causal contrasts (*causal estimands*) as interventions on prespecified modified treatment policies (refer to Haneuse and Rotnitzky (2013); Díaz et al. (2021); Díaz et al. (2023)). Let A_t denote the treatment – monthly frequency of religious service. There are three time points: $t \in 0, 1, 2$, where $t = 0$ denotes the baseline wave, $t = 1$, the treatment wave, and $t = 2$ at the end of the study. $\mathbf{d}(\cdot)$ denotes a modified treatment policy $f_\mathbf{d}$. When a treatment is fixed to a level defined by the modified treatment policy, perhaps contrary to a participant's observed level of treatment, we use the lowercase symbol a_1. Here, the functions defined by modified treatment policies $f_\mathbf{d}$ are interventions that fix A_1 to a_1.

1. **Regular Religious Service Treatment**: Administer regular religious service attendance to everyone in the adult population. If an individual's religious service attendance is below four times per month, shift to four; otherwise, maintain their current attendance:

$$\mathbf{d}^\lambda(a_1) = \begin{cases} 4 & \text{if } a_1 < 4 \\ a_1 & \text{otherwise} \end{cases}$$

2. **Zero Religious Service Treatment**: Ensure no religious service attendance for everyone in the adult population of New Zealand. If an individual's religious service attendance is greater than zero, shift to zero; otherwise, make no change:

$$\mathbf{d}^\phi(a_1) = \begin{cases} 0 & \text{if } a_1 > 0 \\ a_1 & \text{otherwise} \end{cases}$$

3. **Status Quo – No Treatment**: Apply no treatment. Each expected mean outcome is calculated using each individual's natural (observed) value of religious service attendance.

$$\mathbf{d}(a_1) = a_1$$

Causal Contrasts

From these policies, we compute the following causal contrasts.

Target Contrast A: 'Regular vs. Zero': How do the prosocial effects of a society with regular religious service attendance differ from those of a society with zero religious service attendance?

$$\text{Regular Religious Service vs. Zero Religious Service} = E[Y(\mathbf{d}^\lambda) - Y(\mathbf{d}^\phi)]$$

This contrast simulates a scientifically interesting hypothetical experiment where we could randomise individuals to either regular religious service or none, assessing the differences in prosociality outcomes measured one year after the intervention.

Target Contrast B: 'Regular vs. Status Quo': How does a society with regular religious service attendance compare to its status quo?

$$\text{Regular Religious Service vs. No Treatment} = E[Y(\mathbf{d}^\lambda) - Y(\mathbf{d})]$$

This contrast reflects a policy-relevant hypothetical experiment examining the effect of transitioning to regular religious service if one does not already regularly attend, allowing us to quantitatively assess how much a society in which everyone attends would differ from a society in its current state.

Target Contrast C: 'Zero vs. Status Quo': What are the social consequences for society of zero religious service attendance compared to its status quo?

$$\text{Zero Religious Service vs. No Treatment} = E[Y(\mathbf{d}^\phi) - Y(\mathbf{d})]$$

This contrast investigates the policy implications of eliminating religious services entirely, questioning whether such a shift would meaningfully affect average levels of charitable donations and volunteering.

Identification Assumptions

To consistently estimate a causal effect, investigators must satisfy three assumptions:

1. **Causal consistency:** potential outcomes must correspond with observed outcomes under the treatments in the data. Essentially, we assume potential outcomes do not depend on how the treatment was administered, conditional on measured covariates (Tyler J. VanderWeele 2009; Tyler J. VanderWeele and Hernan 2013).

2. **Exchangeability**: given observed covariates, we assume treatment assignment is independent of the potential outcomes to be contrasted. In other words, there is "no unmeasured confounding" (Hernan and Robins 2024; Chatton et al. 2020).

3. **Positivity:** every individual must have a non-zero chance of receiving the treatment, regardless of their covariate values Westreich and Cole (2010). We evaluate this assumption in each study by examining changes in religious service attendance from baseline (NZAVS time 10) to the treatment wave (NZAVS time 11). For further discussion of these assumptions in the context of NZAVS studies, see Joseph A. Bulbulia et al. (2023).

Target Population

The target population for this study comprises New Zealand residents as represented in the baseline wave of the New Zealand Attitudes and Values Study (NZAVS) during the years 2018-2019, weighted by New Zealand Census weights for age, gender, and ethnicity (refer to Chris G. Sibley (2021)). The NZAVS is a national probability study designed to reflect the broader New Zealand population accurately. Despite its comprehensive scope, the NZAVS does have 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. These weights adjust for variations in age, gender, and ethnicity to better approximate the national demographic composition (Chris G. Sibley 2021). Survey weights were integrated into statistical models using the `weights` option in `lmtp` (Williams and Díaz 2021), following protocols stated in J. Bulbulia (2024).

Eligibility Criteria

To be included in the analysis of this study, participants needed to meet the following eligibility criteria:

Inclusion Criteria

- Enrolled in the 2018 wave of the New Zealand Attitudes and Values Study (NZAVS time 10).
- Missing covariate data at baseline was permitted, and the data was subjected to imputation methods to reduce bias. Only information obtained at baseline was used for such imputation (refer to Zhang et al. (2023)). Participants may have been lost to follow-up the end of study NZAVS time 12 if they met eligibility criteria at NZAVS time 11 (the treatment wave). We adjusted for attrition and non-response using censoring weights, described below.

Exclusion Criteria

- Did not answer the religious service attendance question at New Zealand Attitudes and Values Study at time 10 (the baseline wave) and NZAVS time 11 (the treatment wave).

A total of 33,198 individuals met these criteria and were included in the study.

Causal Identification

Table 1 presents three Single World Intervention Graphs (SWIGs) that describe our confounding control (identification strategy) (J. M. Robins and Richardson 2010; Richardson and Robins 2013, 2023; Shpitser, Richardson, and Robins 2022; Richardson et al. 2023; Shpitser and Tchetgen 2016). Our approach consistently applies the same identification strategy across all functions estimated in this study. Unlike standard causal diagrams, SWIGs allow us to *separately* read the factorisation of the conditional dependencies for the distribution of each set of counterfactual outcomes under each modified treatment policy (Richardson and Robins 2013). Note, that the natural value of the treatment A is obtained both from its observed instances and from baseline historical data, including the baseline treatment. This method ensures that our analysis accurately captures the causal effects of flexible treatment regimes that rely on levels of religious service attendance at the treatment wave, while ensuring balance for each treatment function that we compare (Muñoz and Van Der Laan 2012; Young, Hernán, and Robins 2014; Díaz et al. 2021).

Confounding Control

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

1. **Identified all common causes** of both the treatment and outcomes to ensure a comprehensive approach to confounding control.
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, 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.

Appendix B details the covariates we included for confounding control. These methods adhere to the guidelines provided in (J. Bulbulia 2024) and were pre-specified in our study protocol https://osf.io/ce4t9/.

Missing Data

To mitigate bias from missing data, we implement the following strategies:

Baseline missingness: we employed the ppm algorithm from the mice package in R (Van Buuren 2018) to impute missing baseline data. This method allowed us to reconstruct incomplete datasets by estimating a plausible value for missing observation. Because we could only pass one data set to the lmtp, we employed single imputation.

Table 1: This table presents three Single World Intervention Graphs (SWIGs), one for each treatment condition we compare. Note that we obtain robust confounding control by including baseline measures for both the treatments and outcomes (refer to Tyler J. VanderWeele, Mathur, and Chen (2020), protocols described in J. Bulbulia (2024)).We recommend using SWIGs because they are more precise and general than standard causal diagrams (refer to Richardson and Robins (2013)).

	Identification Task	Single World Intervention Graphs (SWIGs)
1	Control for common causes of A_1 and Y_2, compute outcome under $\mathbf{d}^\lambda$.	$U \longrightarrow \boxed{\begin{array}{c}L_0\\A_0\\Y_0\end{array}} \longrightarrow A_1 \mid \mathbf{d}^\lambda(a_1) \longrightarrow Y_2(\mathbf{d}^\lambda)$
2	Control for common causes of A_1 and Y_2, compute outcome under $\mathbf{d}^\phi$.	$U \longrightarrow \boxed{\begin{array}{c}L_0\\A_0\\Y_0\end{array}} \longrightarrow A_1 \mid \mathbf{d}^\phi(a_1) \longrightarrow Y_2(\mathbf{d}^\phi)$
3	Control for common causes of A_1 and Y_2, compute outcome under $\mathbf{d}$.	$U \longrightarrow \boxed{\begin{array}{c}L_0\\A_0\\Y_0\end{array}} \longrightarrow A_1 \mid \mathbf{d}(a_1) \longrightarrow Y_2(\mathbf{d})$

Target Contrast A	$\mathbf{E}[Y(\mathbf{d}^\lambda) - Y(\mathbf{d}^\phi)]$
Target Contrast B	$\mathbf{E}[Y(\mathbf{d}^\lambda) - Y(\mathbf{d})]$
Target Contrast C	$\mathbf{E}[Y(\mathbf{d}^\phi) - Y(\mathbf{d})]$

Key:
$\mathbf{d}(\cdot)$ denotes a modified treatment policy $f_{\mathbf{d}}$
U denotes unmeasured confounders, which we assume are captured in measured variables.
L_0 denotes baseline confounders (may be time-varying).
L_1 denotes confounders of A_1.
A_0 denotes the value of the treatment at the initial wave.
A_1 denotes "the natural value treatment" evaluated at treatment wave $t_1 > t_0$ (sometimes written A_t^*)
$Y_2(f_{\mathbf{d}})$ denotes the potential outcome under a pre-specified treatment policy $f_{\mathbf{d}}$, where the outcome wave: $t_2 > t_1$.
$\boxed{X}$ indicates conditioning on confounding control vector X.
$\longrightarrow$ asserts causality.
Note: Single World Intervention Graphs (SWIGs) clarify demands for confounding control under each modified treatment policy. We apply census weights at baseline to obtain valid estimates for the target population. Because missing responses can bias this estimate, we impute missing values at baseline ($t = 0$) and employ censoring weights to address attrition/non-response.

About 2% of covariate values were missing at baseline. Eligibility for the study required fully observed baseline treatment measures as well as treatment wave treatment measures. Again, we only used baseline data to impute baseline missingness (refer to Zhang et al. (2023)).

Outcome missingness: to address confounding and selection bias arising from missing responses and panel attrition, we applied censoring weights obtained using nonparametric machine learning ensembles afforded by the lmtp package (and its dependencies) in R (Williams and Díaz 2021).

Statistical Estimator

We perform statistical estimation using semi-parametric Targeted Learning, specifically a Targeted Minimum Loss-based Estimation (TMLE) estimator. TMLE is a robust method that combines machine learning techniques with traditional statistical models to estimate causal effects while providing valid statistical uncertainty measures for these estimates (Van der Laan 2014; Laan and Gruber 2012).

TMLE operates through a two-step process that involves modelling both the outcome and treatment (exposure). Initially, TMLE employs machine learning algorithms to flexibly model the relationship between treatments, covariates, and outcomes. This flexibility allows TMLE to account for complex, high-dimensional covariate spaces *efficiently* without imposing restrictive model assumptions (Laan, Luedtke, and Díaz 2014; Van Der Laan and Rose 2011, 2018). The outcome of this step is a set of initial estimates for these relationships.

The second step of TMLE involves "targeting" these initial estimates by incorporating information about the observed data distribution to improve the accuracy of the causal effect estimate. TMLE achieves this precision through an iterative updating process, which adjusts the initial estimates towards the true causal effect. This updating process is guided by the efficient influence function, ensuring that the final TMLE estimate is as close as possible, given the measures and data, to the targeted causal effect while still being robust to model-misspecification in either the outcome or the treatment model (Laan, Luedtke, and Díaz 2014).

Again, a central feature of TMLE is its double-robustness property. If either the treatment model or the outcome model is correctly specified, the TMLE estimator will consistently estimate the causal effect. Additionally, we used cross-validation to avoid over-fitting, following the pre-stated protocols in J. Bulbulia (2024). The integration of TMLE and machine learning technologies reduces the dependence on restrictive modelling assumptions and introduces an additional layer of robustness. For further details of the specific targeted learning strategy we favour, see (Hoffman et al. 2022, 2023; Díaz et al. 2021). We perform estimation using the lmtp package (Williams and Díaz 2021). We used the superlearner library for semi-parametric estimation with the predefined libraries SL.ranger, SL.glmnet, and SL.xgboost (Polley et al. 2023; T. Chen et al. 2023; Wright and Ziegler 2017). We created graphs, tables and output reports using the margot package (Joseph A. Bulbulia 2024).

Sensitivity Analysis Using the E-value

To assess the sensitivity of results to unmeasured confounding, we report VanderWeele and Ding's "E-value" in all analyses (Tyler J. VanderWeele and Ding 2017). The E-value quantifies the minimum strength of association (on the risk ratio scale) that an unmeasured confounder would need to have with both the exposure and the outcome (after considering the measured covariates) to explain away the observed exposure-outcome association (Tyler J. VanderWeele, Mathur, and Chen 2020; Linden, Mathur, and VanderWeele 2020). To evaluate the strength of evidence, we use the bound of the E-value 95% confidence interval closest to 1.

Scope of Interventions

To illustrate the magnitude of the shift interventions we contrast, we provide histograms in Figure 1, that display the distribution of treatments during the treatment wave. Figure 1 *A*: The intervention for regular religious service, represented in these histograms, affects a larger portion of the sample than the zero religious service intervention. Figure 1 *B*: presents the intervention for zero religious service. It involves shifting a smaller portion of the sample

than the regular religious service intervention. Again, the comparative analysis of the 'Regular' versus 'Zero' interventions addresses the scientifically intriguing question: what is the effect difference in a scenario where religious service is universal versus completely absent? The intervention that increases attendance to regular service levels from the status quo allows us to consider the potential costs and benefits of widespread religious practice across society. The intervention that eliminates religious service allows us to consider the potential costs and benefits of the widespread loss of religious practice across society.

A

Histogram of Monthly_Religious_Service Shift Intervention
Gold region denotes population shifted up to grey
Red dashed line shows the average value

B

Histogram of Monthly_Religious_Service Shift Intervention
Blue region denotes population shifted down to grey
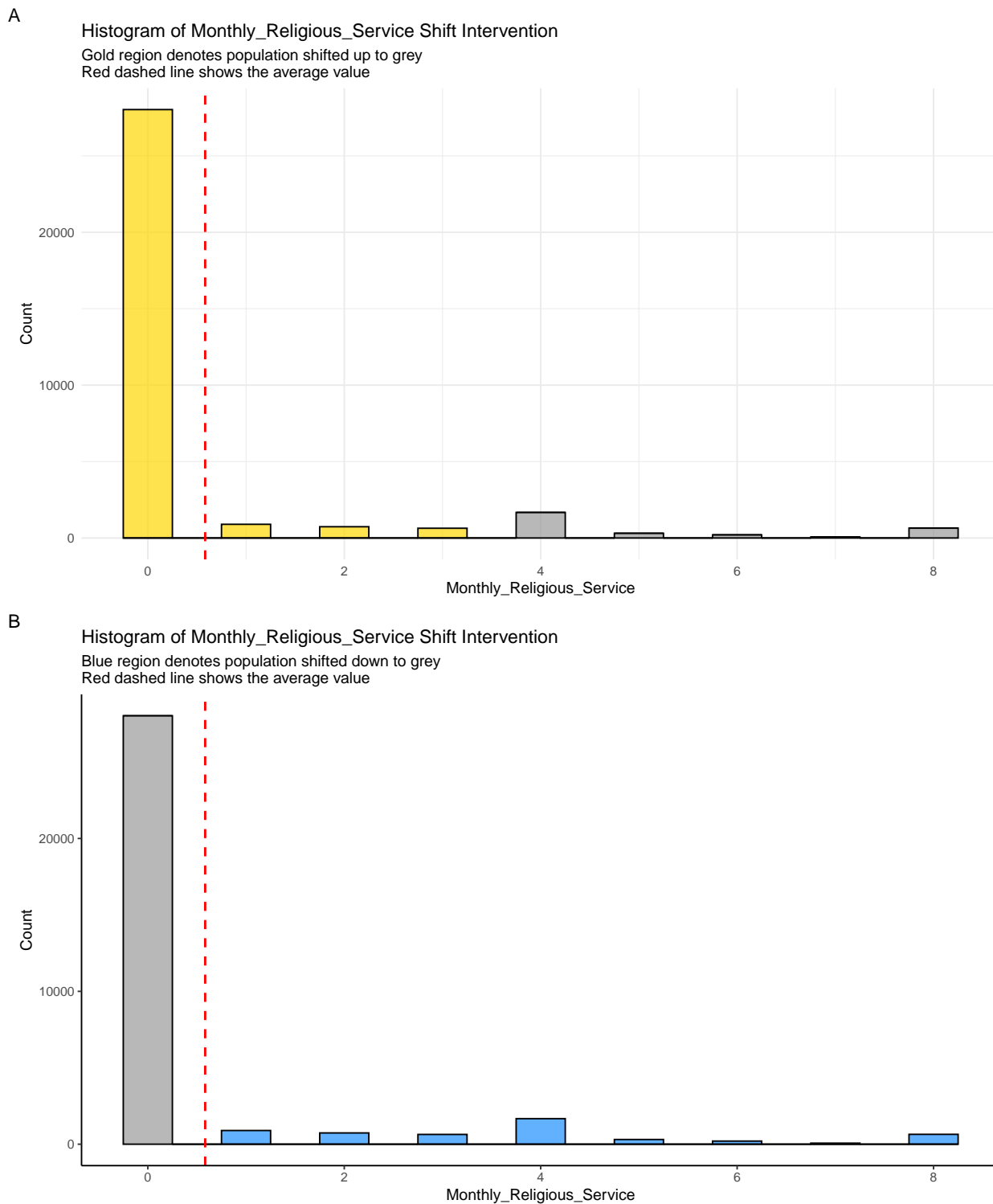Red dashed line shows the average value

Figure 1: This figure shows a histogram of responses to religious service frequency in the baseline + 1 wave. Responses above eight were assigned to eight, and values were rounded to the nearest whole number. The red dashed line shows the population average. (A) Responses in the gold bars are shifted to four on the Regular Religious Service intervention. All those responses in grey (four and above) remain unchanged. (B) On the zero-intervention, responses in the blue bars denote those shifted under the zero-intervention treatment.

Evidence for Change in the Treatment Variable

Table 2 clarifies the change in the treatment variable from the baseline wave to the baseline + 1 wave across the sample. Assessing change in a variable is essential for evaluating the positivity assumption and recovering evidence for the incident exposure effect of the treatment variable (Tyler J. VanderWeele, Mathur, and Chen 2020; Danaei, Tavakkoli, and Hernán 2012; Hernan and Robins 2024). We find that state 4 (weekly attendance) and state 0 present the highest overall. However, movement between these states reveals they are not deterministic. States 1, 2, 3, and 5 exhibit more frequent jumps in and out of these states, suggesting lower stability and/or measurement error.

Table 2: This transition matrix captures stability and change in religious service between the baseline and treatment wave. Each cell in the matrix represents the count of individuals transitioning from one state to another. The rows correspond to the state at baseline (From), and the columns correspond to the state at the treatment wave (To). **Diagonal entries** (in **bold**) signify the number of individuals who remained in their initial state across both waves. **Off-diagonal entries** signify the transitions of individuals from their baseline state to a different state in the treatment wave. A higher number on the diagonal relative to the off-diagonal entries in the same row indicates greater stability in a state. Conversely, higher off-diagonal numbers suggest more frequent shifts in the sample from the baseline state to other states.

From	State 0	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8
State 0	**26762**	405	174	71	126	26	13	8	68
State 1	647	**235**	85	44	46	5	2	3	10
State 2	236	105	**188**	104	96	12	13	2	21
State 3	112	54	110	**164**	173	18	8	4	15
State 4	150	71	127	205	**881**	124	64	16	91
State 5	24	7	17	17	145	**61**	25	7	33
State 6	14	5	13	17	84	22	**29**	5	37
State 7	9	0	6	3	16	6	9	**6**	19
State 8	74	14	17	14	105	34	42	17	**351**

Results

Study 1: Causal Effects of Regular Church Attendance on Self-Reported Volunteering and Self-Reported Volunteering and Donations

Regular Religious Service vs. Zero Treatment Contrast for Donations and Volunteering

Results for the treatment contrasts between Regular Religious Service and Zero Religious Service, focusing on self-reported volunteering and charitable donations, are displayed in Figure 2 *A* and Table 3. These results are measured on the difference scale.

Table 3: This table reports the results of model estimates for the causal effects of a universal gain of weekly religious service vs. a universal loss of weekly religious service on reported charitable behaviours at the end of the study. Contrasts are expressed in standard deviation units.

	E[Y(1)]-E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
donations	0.132	0.102	0.161	1.507	1.426
hours volunteer	0.123	0.090	0.156	1.482	1.389

For 'donations', the effect estimate is 0.132 [0.102, 0.161]. The E-value for this estimate is 1.507, with a lower bound of 1.426. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.426 to negate the observed effect. Weaker associations would not overturn it. We infer **evidence for causality**. On the data scale, this intervention represents a difference of **NZD 656.58 per adult per year** in charitable giving compared with the zero attendance intervention.

The effect estimate for 'hours volunteer' is 0.123 [0.09, 0.156]. The E-value for this estimate is 1.482, with a lower bound of 1.389. We infer **evidence for causality**. On the data scale, this intervention represents a difference of **NZD 30.21 minutes** per adult per week in volunteering compared with the zero attendance intervention.

Regular Religious Service vs. Status Quo Treatment Contrast for Donations and Volunteering

Figure 2 *B* and Table 4 present results for the treatment contrasts between Regular Religious Service and Status Quo, focusing on self-reported volunteering and charitable donations. These results are measured on the difference scale.

Table 4: This table reports results of model estimates for the causal effects of a universal gain of weekly religious service vs. the status quo on reported charitable behaviours at the end of the study. Contrasts are expressed in standard deviation units.

	E[Y(1)]-E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
donations	0.121	0.102	0.140	1.477	1.422
hours volunteer	0.095	0.066	0.123	1.404	1.317

For 'donations', the effect estimate is 0.121 [0.102, 0.14]. The E-value for this estimate is 1.477, with a lower bound of 1.422. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.422 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**. On the data scale, this intervention represents an increase of **NZD 601.87 per adult per year** in expected charitable giving over the status quo.

For 'hours volunteer', the effect estimate is 0.095 [0.066, 0.123]. The E-value for this estimate is 1.404, with a lower bound of 1.317. At this lower bound, unmeasured confounders would need a minimum association strength with

both the intervention sequence and outcome of 1.317 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**. On the data scale, this intervention represents an increase of **23.33 minutes** per adult per week in hours volunteering over the status quo.

Zero Religious Service vs. The Status Quo Treatment Contrast for Donations and Volunteering

Figure 2 *C* and Table 5 present results for the treatment contrasts between Zero Religious Service and Status Quo, focusing on self-reported volunteering and charitable donations. These results are measured on the difference scale.

Table 5: This table reports the results of model estimates for the causal effects of a universal loss of weekly religious service vs. the status quo on reported charitable behaviours at the end of the study. Contrasts are expressed in standard deviation units.
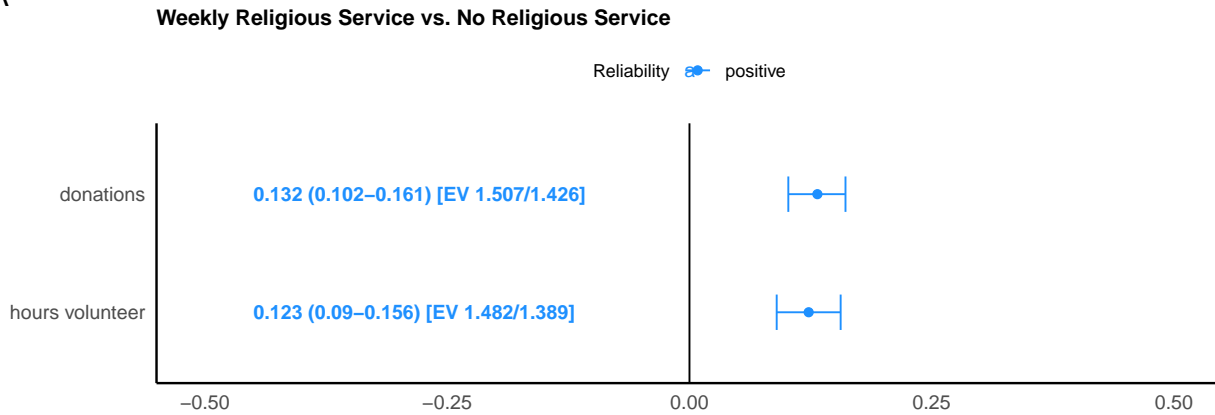
	E[Y(1)]-E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
donations	-0.011	-0.029	0.008	1.111	1.000
hours volunteer	-0.028	-0.042	-0.014	1.189	1.128

For 'donations', the effect estimate is -0.011 [-0.029, 0.008]. The E-value for this estimate is 1.111, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer that **the evidence for causality is not reliable**. On the data scale, this intervention represents a difference of NZD -54.72 per adult per year in charitable giving compared to the status quo. Still, again, this effect is not reliable.
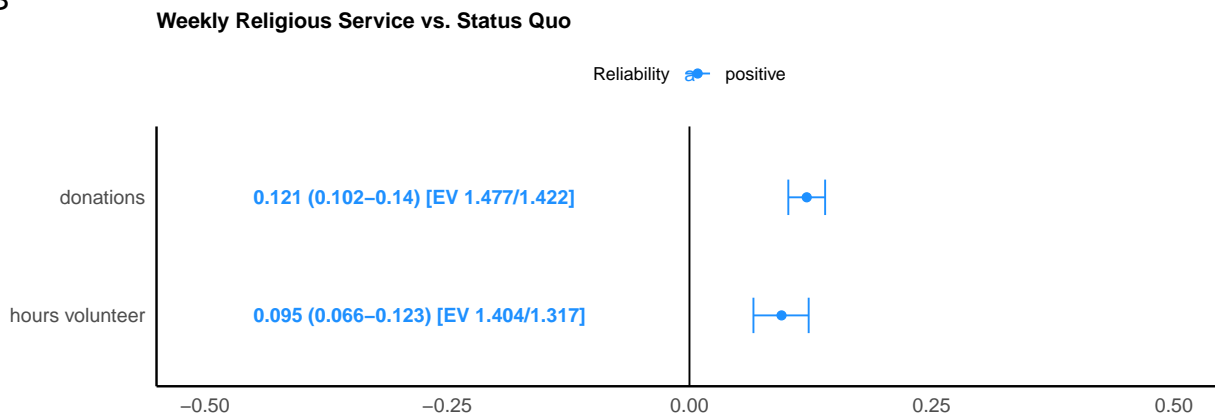
The effect estimate for 'hours volunteer' is -0.028 [-0.042, -0.014]. The E-value for this estimate is 1.189, with a lower bound of 1.128. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.128 to negate the observed effect. Again, weaker confounding would not overturn it. We infer **evidence for causality**. **On the data scale, this intervention represents a difference of -6.88 in volunteering minutes compared with the status quo.**

Study 1: Self–Reported Prosociality

A

Weekly Religious Service vs. No Religious Service

Reliability ● positive

donations	0.132 (0.102–0.161) [EV 1.507/1.426]
hours volunteer	0.123 (0.09–0.156) [EV 1.482/1.389]

−0.50   −0.25   0.00   0.25   0.50

B

Weekly Religious Service vs. Status Quo

Reliability ● positive

donations	0.121 (0.102–0.14) [EV 1.477/1.422]
hours volunteer	0.095 (0.066–0.123) [EV 1.404/1.317]

−0.50   −0.25   0.00   0.25   0.50

C

No Religious Service vs. Status Quo

Reliability ● negative ● zero_crossing

donations	−0.011 (−0.029−0.008) [EV 1.111/1]
hours volunteer	−0.028 (−0.042−−0.014) [EV 1.189/1.128]

−0.50   −0.25   0.00   0.25   0.50
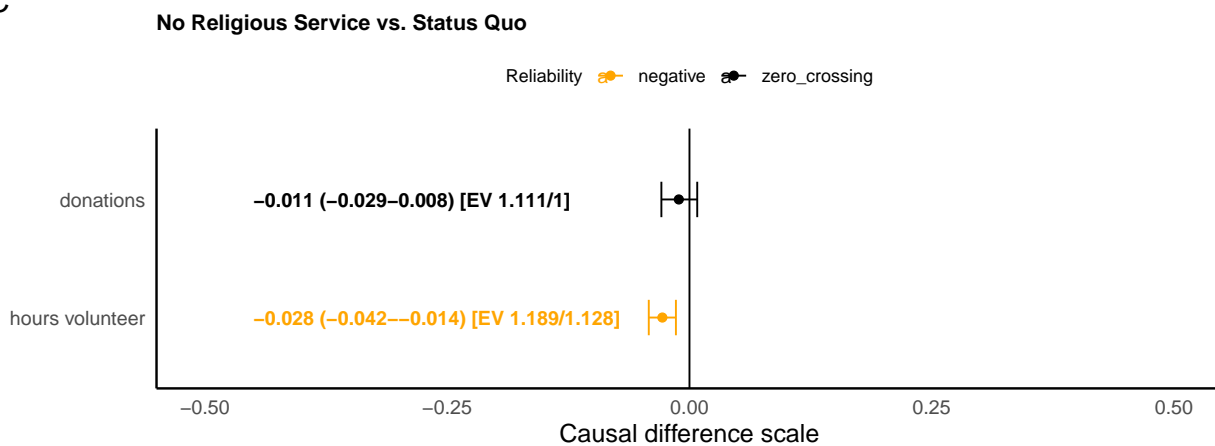Causal difference scale

Figure 2: This figure graphs the results of model estimates for the causal effects of the three causal contrasts of interest on reported charitable behaviours at the study's end. The causal contrasts are: (A) Regular vs. Zero Religious Service, (B) Regular Religious Service vs. Status Quo, and (C) Zero Religious Service vs. Status Quo. Contrasts are expressed in standard deviation units.

Study 2: Causal Effects of Regular Church Attendance on Support Received From Others – Time

Regular vs. Zero Causal Treatment Contrast for Time Received From Others

Figure 3 *A* and Table 6 present results for the treatment contrasts between Regular Religious Service and Zero, focusing on voluntary help received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 6: This table reports the results of model estimates for the causal effects of a universal gain of weekly religious service vs. a universal loss of weekly religious service on voluntary help received from others during the past week (yes/no) at the end of the study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives time	0.950	0.901	1.003	1.288	1.000
friends give time	1.187	1.108	1.271	1.658	1.454
community gives time	1.378	1.231	1.541	2.100	1.764

For 'community gives time', the effect estimate is 1.378 [1.231, 1.541]. The E-value for this estimate is 2.1, with a lower bound of 1.764. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.764 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'friends give time', the effect estimate is 1.187 [1.108, 1.271]. The E-value for this estimate is 1.658, with a lower bound of 1.454. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.454 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'family gives time', the effect estimate is 0.95 [0.901, 1.003]. The E-value for this estimate is 1.288, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

Regular Religious Service vs. Status Quo Treatment Contrast for Time Received From Others

Figure 3*B* and Table 7 present results for the treatment contrasts between Regular Religious Service and Status Quo, focusing on voluntary help received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 7: This table reports the results of model estimates for the causal effects of a universal gain of weekly religious service vs. the status quo on voluntary help received from others during the past week (yes/no) at the end of the study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives time	0.958	0.913	1.006	1.258	1.000
friends give time	1.128	1.061	1.199	1.508	1.315
community gives time	1.289	1.174	1.415	1.899	1.626

For 'community gives time', the effect estimate is 1.289 [1.174, 1.415]. The E-value for this estimate is 1.899, with a lower bound of 1.626. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.626 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'friends give time', the effect estimate is 1.128 [1.061, 1.199]. The E-value for this estimate is 1.508, with a lower bound of 1.315. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.315 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'family gives time', the effect estimate is 0.958 [0.913, 1.006]. The E-value for this estimate is 1.258, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

Zero Religious Service vs. Status Quo Treatment Contrast for Time Received From Others

Figure 3 C and Table 8 present results for the treatment contrasts between Zero Religious Service and Status Quo, focusing on voluntary help received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 8: This table reports results of model estimates for the causal effects of a universal loss of weekly religious service vs. the status quo on voluntary help received from others during the past week (yes/no) at the end of the study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives time	1.008	0.991	1.026	1.098	1.000
friends give time	0.950	0.928	0.973	1.288	1.197
community gives time	0.936	0.889	0.985	1.339	1.140

For 'family gives time', the effect estimate is 1.008 [0.991, 1.026]. The E-value for this estimate is 1.098, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

For 'friends give time', the effect estimate is 0.95 [0.928, 0.973]. The E-value for this estimate is 1.288, with a lower bound of 1.197. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.197 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'community gives time', the effect estimate is 0.936 [0.889, 0.985]. The E-value for this estimate is 1.339, with a lower bound of 1.14. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.14 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

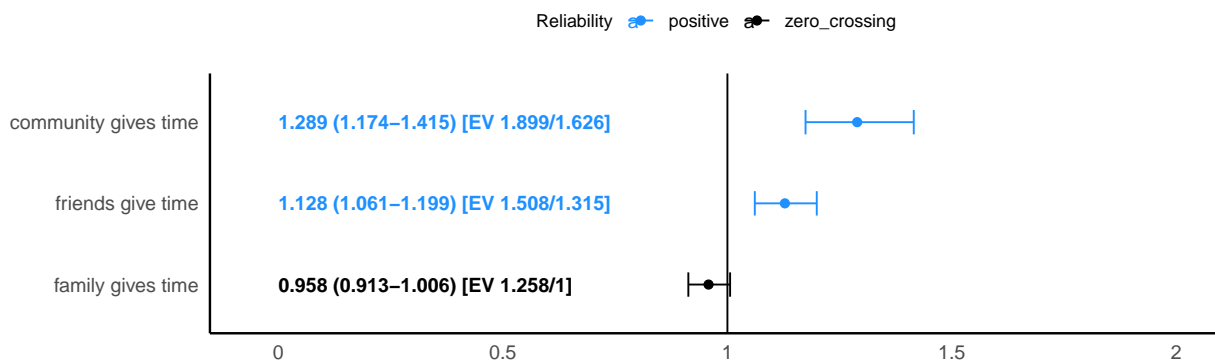## Study 2: Relative Risk of Help Received Last Week: Time
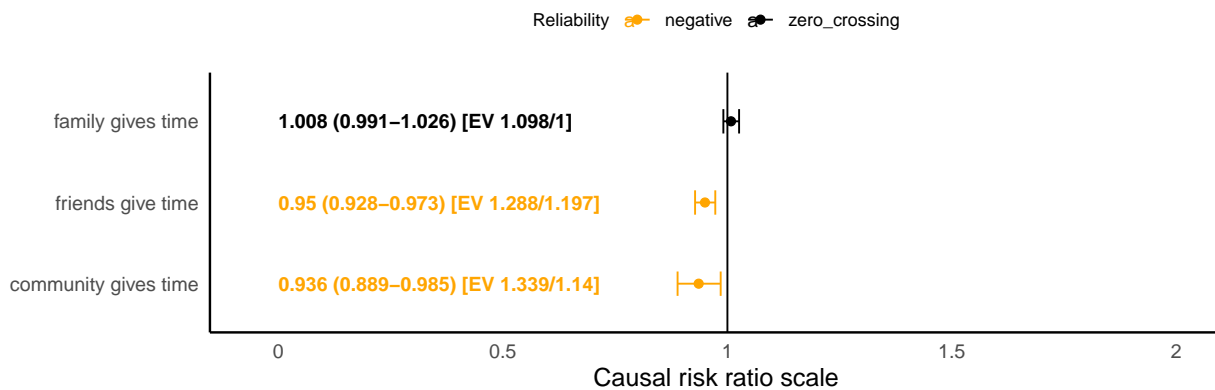
A
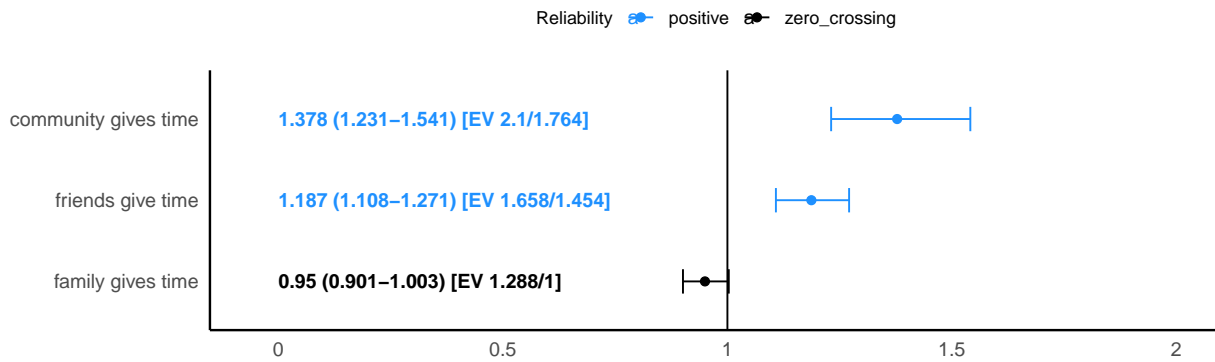
Figure 3: This figure reports the results of model estimates for the three causal contrasts of interest on help received from others during the past week (yes/no). The causal contrasts are (A) Regular vs. Zero Religious Service, (B) Regular Religious Service vs. Status Quo, and (C) Zero Religious Service vs. Status Quo. Contrasts are expressed on the risk ratio scale.

Study 3: Causal Effects of Regular Church Attendance on Support Received From Others – Money

Regular vs. Zero Causal Contrast on Money Received From Others

Figure 4 A and Table 9 present results for the treatment contrasts between Regular Religious Service and Zero, focusing on money received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 9: This table reports the results of model estimates for the causal effects of a universal gain of weekly religious service vs. a universal loss of weekly religious service on financial help received from others during the past week (yes/no) at the end of the study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives money	1.137	1.028	1.258	1.532	1.198
friends give money	1.137	0.964	1.342	1.532	1.000
community gives money	1.376	1.112	1.703	2.095	1.465

For 'community gives money', the effect estimate is 1.376 [1.112, 1.703]. The E-value for this estimate is 2.095, with a lower bound of 1.465. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.465 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'family gives money', the effect estimate is 1.137 [1.028, 1.258]. The E-value for this estimate is 1.532, with a lower bound of 1.198. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.198 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'friends give money', the effect estimate is 1.137 [0.964, 1.342]. The E-value for this estimate is 1.532, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

Regular vs. Status Quo Causal Contrast on Money Received From Others

Figure 4 B and Table 10 present results for the treatment contrasts between Regular Religious Service and Status Quo, focusing on money received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 10: This table reports the results of model estimates for the causal effects of a universal gain of weekly religious service vs. the status quo on financial help received from others during the past week (yes/no) at the end of the study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives money	1.130	1.037	1.232	1.513	1.233
friends give money	1.041	0.951	1.139	1.248	1.000
community gives money	1.254	1.098	1.432	1.818	1.426

For 'community gives money', the effect estimate is 1.254 [1.098, 1.432]. The E-value for this estimate is 1.818, with a lower bound of 1.426. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.426 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'family gives money', the effect estimate is 1.13 [1.037, 1.232]. The E-value for this estimate is 1.513, with a lower bound of 1.233. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1.233 to negate the observed effect. Weaker confounding would not overturn it. We infer **evidence for causality**.

For 'friends give money', the effect estimate is 1.041 [0.951, 1.139]. The E-value for this estimate is 1.248, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

Zero vs. Status Quo Causal Contrast on Money Received From Others

Figure 4 C and Table 11 present results for the treatment contrasts between Zero Religious Service and Status Quo, focusing on money received from others during the past week (yes/no). These results are measured on the risk ratio scale.

Table 11: Table reports results of model estimates for the causal effects of a universal loss of weekly religious service vs. the status quo on financial help received from others during the past week (yes/no) at the end of study. Contrasts are expressed on the risk ratio scale.

	E[Y(1)]/E[Y(0)]	2.5 %	97.5 %	E_Value	E_Val_bound
family gives money	0.993	0.953	1.035	1.091	1
friends gives money	0.915	0.809	1.036	1.412	1
community gives money	0.911	0.796	1.042	1.425	1

For 'family gives money', the effect estimate on the risk ratio scale is 0.993 [0.953, 1.035]. The E-value for this estimate is 1.091, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.
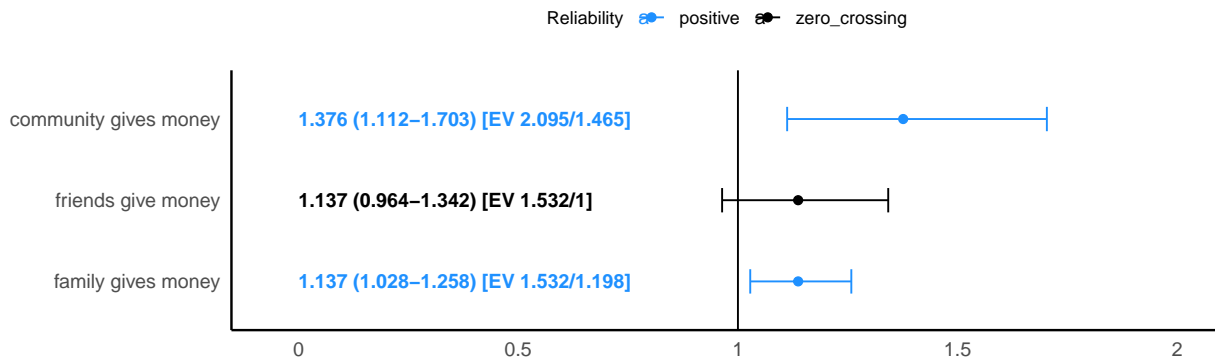
For 'friends give money', the effect estimate on the risk ratio scale is 0.915 [0.809, 1.036]. The E-value for this estimate is 1.412, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.

For 'community gives money', the effect estimate on the risk ratio scale is 0.911 [0.796, 1.042]. The E-value for this estimate is 1.425, with a lower bound of 1. At this lower bound, unmeasured confounders would need a minimum association strength with both the intervention sequence and outcome of 1 to negate the observed effect. Weaker confounding would not overturn it. We infer **that evidence for causality is not reliable**.
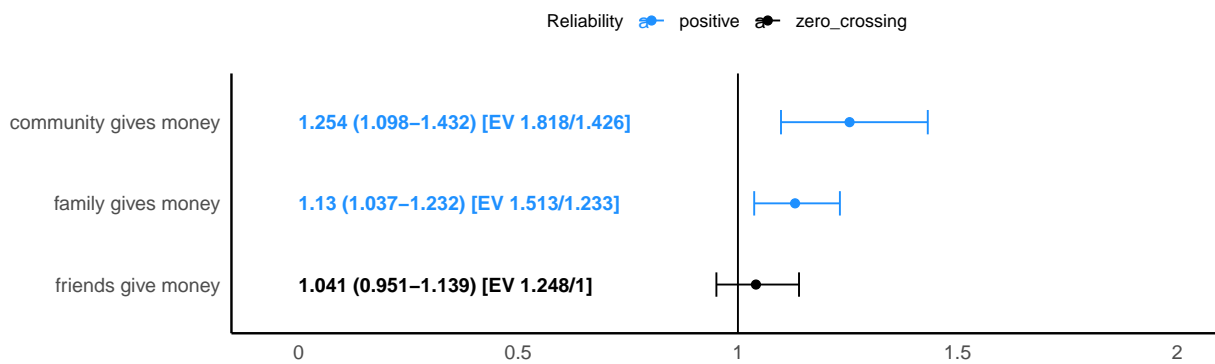
Study 3: Relative Risk of Help Received Last Week: Money

A

Weekly Religious Service vs. No Religious Service

Reliability ⊶ positive ⊶ zero_crossing

community gives money	1.376 (1.112–1.703) [EV 2.095/1.465]
friends give money	1.137 (0.964–1.342) [EV 1.532/1]
family gives money	1.137 (1.028–1.258) [EV 1.532/1.198]

0   0.5   1   1.5   2

B

Weekly Religious Service vs. Status Quo

Reliability ⊶ positive ⊶ zero_crossing

community gives money	1.254 (1.098–1.432) [EV 1.818/1.426]
family gives money	1.13 (1.037–1.232) [EV 1.513/1.233]
friends give money	1.041 (0.951–1.139) [EV 1.248/1]

0   0.5   1   1.5   2

C

No Religious Service vs. Status Quo

Reliability ⊶ zero_crossing

family gives money	0.993 (0.953–1.035) [EV 1.091/1]
friends gives money	0.915 (0.809–1.036) [EV 1.412/1]
community gives money	0.911 (0.796–1.042) [EV 1.425/1]

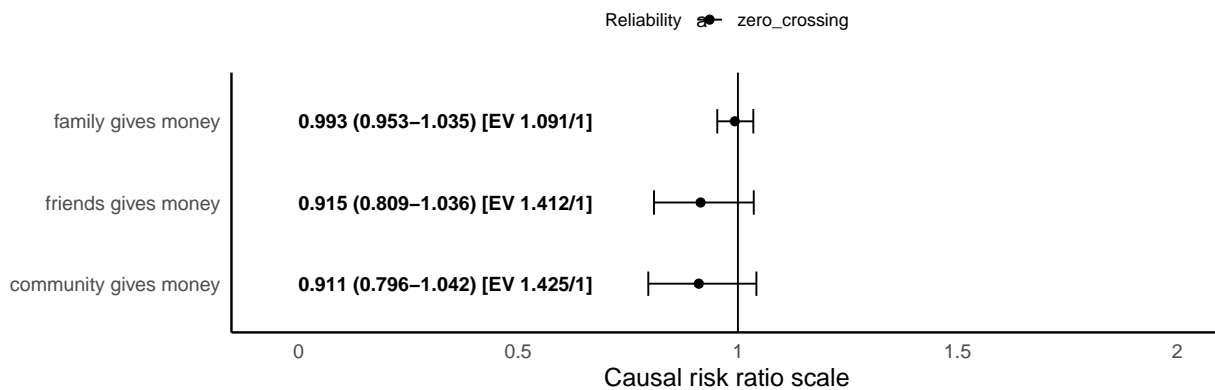0   0.5   1   1.5   2

Causal risk ratio scale

Figure 4: This figure reports the results of model estimates for the three causal contrasts of interest on help received from others during the past week (yes/no). The causal contrasts are: (A) Regular vs. Zero Religious Service (B) Regular Religious Service vs. Status Quo; (C) Zero Religious Service vs. Status Quo. Contrasts are expressed on the risk ratio scale.

Additional Study: Comparison of Causal Inference Results with Cross-Sectional Regressions

To better evaluate the contributions of our methodology to current practice, we conducted a series of cross-sectional analyses using the baseline wave data. We quantified the statistical associations between religious service attendance and our focal prosocial outcomes. We included all regression covariates from the causal models (including sample weights) for each analysis, obviously omitting the outcome measured at baseline, i.e. the response variable.

Cross-sectional volunteering result: the change in expected hours of volunteer work for a one-unit increase in religious service attendance is b = 0.31; (95% CI 0.28, 0.34). Multiplying this by 4.2 gives a monthly estimate of 77.95 minutes. This result is 2.58 per cent greater than the effect estimated from the 'regular vs. zero' causal contrast, indicating an **overstatement in the cross-sectional regression model.**

Cross-sectional charitable donations result: The coefficient for religious service on annual charitable donations suggests a change in expected donation amount per unit increase in attendance is: b = 451; (95% CI 408, 494). When adjusted to a monthly rate by multiplying by 4.2, this value equals NZ Dollars 1894.45. It is 2.89 per cent greater than our causal contrast estimate, again indicating **overstatement in the cross-sectional regression model.**

For Studies 2 and 3, which focus on community help received, we adjusted our analysis for the non-collapsibility of odds ratios by assuming a Poisson distribution for the outcome variables, obtaining a rate ratio that approximates a risk ratio (Huitfeldt, Stensrud, and Suzuki 2019; Tyler J. VanderWeele, Mathur, and Chen 2020):

Cross-sectional community assistance received result: Time: the exponentiated change in expectation for a one-unit change in religious service attendance is b = 1.17; (95% CI 1.14, 1.19) approximate risk rate ratio. The monthly rate ratio derived by multiplying this coefficient by 4.2 is 1.921. This estimate is 1.39 per cent greater than the 'regular vs. zero' causal estimate, pointing to **overstatement in the cross-sectional regression model.**

Cross-sectional community assistance received result: Money: similarly, the exponentiated change for money received yields an approximate risk ratio of b = 1.18; (95% CI 1.08, 1.27). The monthly risk ratio, after adjustment, is 1.996. This rate ratio is 1.45 per cent greater than the causal estimate, again revealing an **overstatement in the cross-sectional regression model.**

These findings underscore that the results of cross-sectional regressions, although suggestive, can considerably diverge from those obtained from the causal analysis of panel data.

What is the "Cash Value" of Religious Service Attendance for Charitable Donations in New Zealand?

We leveraged results to estimate the approximate economic value of religious service attendance under different scenarios— 'Regular Religious Service', 'Zero Religious Service', and the 'Status Quo', focusing on charitable donations.

- **Regular Religious Service**: an increase in religious service attendance yields an individual average donation sum of **NZD 1638.98.**
- **Zero Religious Service**: Reducing religious service attendance to zero yields an average donation sum of **NZD 984.59.**
- **Status Quo**: the expected individual average donation sum is currently **NZD 1037.14.**

With 3,989,000 adult residents in New Zealand in 2021:[2]

- Multiplying the adult population by the average donation sum gives a status quo national estimate for charitable giving of **NZD 4,137,151,460.**
- The net gain to charity from country-wide regular attendance at religious services, compared to the status quo, is **NZD 2,400,739,760.**
- Conversely, although the net cost to charity from a complete cessation of regular religious service attendance is NZD -209,621,950, recall the confidence interval crosses zero, and this effect is not reliable.

To provide context, consider these economic consequences against the New Zealand government's annual budget in year outcomes were measured (2021-2022) is NZD 57,976,000,000

- **The expected gain from a nationwide adoption of regular religious service represents 4.1 percent of New Zealand's annual government budget 2021.**
- We do not obtain a reliable effect from the loss intervention.

Thus, focussing on the individual-level and aggregating effects across the adult population, the counterfactual scenario in which all New Zealand adults regularly attend religious services projects a substantial increase in society-wide charitable support compared to the status quo, one year after the intervention.

However, suppose New Zealand were to experience a complete cessation of religious service attendance. We do not find evidence that the landscape of charitable giving would change from its current state one year later.[3]

Discussion

Considerations

First, as stated above, our causal inferences turn on three assumptions, which are worth revisiting:

(i). **Unmeasured confounding**: although we employ robust methods for causal inference, our results depend on the effectiveness of our strategy to control for confounding. Our sensitivity analyses address the potential impacts of unmeasured confounders. Nevertheless, the presence and influence of such confounders are uncertain and unverifiable from our data.

(ii). **Causal consistency**: the observed outcomes must correspond to the counterfactual treatments contrasted. Although "religious service attendance" may appear conditionally independent of the outcomes given baseline covariates, interpreting these interventions remains challenging. "Religious service attendance" varies widely, encompassing everything from informal gatherings at homes to formal services in cathedrals across New Zealand's religious diversity. This type of "treatment" does not mirror a straightforward medical intervention like a vaccine.

---

[2]National Population Estimates at 30 June 2021

[3]We emphasise that failure to rule out an effect is not the same as ruling it out. The relatively low levels of religious service attendance in New Zealand mutes differences in expected giving. In our view, some amount of charity is always better than none. We do not devalue any form of charitable giving. Furthermore, our population-wide estimates for the aggregated effects of *religious behaviour* do not clarify the effects of gaining or losing *religious institutions*; our analysis reflects the contributions made by those participating in religious institutions.

The heterogeneity of "religious service" limits the clarity of our results, an issue no amount of data or analysis can resolve because "religious service" reflects a broad spectrum of community activities.

(iii). **Positivity**: we have confirmed that religious service attendance varies within our sample and have used semi-parametric models with ensemble learning and cross-validation to prevent data over-extrapolation and model over-fitting. Conceptually, it is crucial for valid causal inference that every potential level of "treatment" to religious services is realistically possible (refer to discussion in Tyler J. VanderWeele (2017)). Although there are instances realised in our data of secular individuals initiating religious service and frequent attendees stopping (refer to Table 2), it might stretch credulity too far to imagine that such changes are possible for everyone.

Second, our study confronts the spectre of **measurement error**: both direct and correlated measurement errors can introduce biases, either by implying effects where none exist or by attenuating true effects (Tyler J. VanderWeele and Hernán 2012). Importantly, evaluating prosociality using multiple measures while also controlling for these measures and the treatment at baseline helps to mitigate measurement error concerns. Nevertheless, unknown combinations of measurement error might nevertheless bias our results. The outcomes and estimates we report here are best-considered approximations.

Third, we do not examine **treatment effect heterogeneity**: identifying which subgroups experience the strongest responses remains a task for future research. Such investigations are crucial for making informed policy decisions and tailoring advice relevant to those subgroups of the population who might benefit most. Perhaps the most obvious stratum is religious affiliates (Tyler J. VanderWeele 2017).

Fourth, the **transportability of our findings remains unclear**: New Zealand is our target population. Our findings generalise to this population. However, the transportability of our findings to other settings—whether our results generalise beyond our targeted New Zealand population—remains an open question, a matter for future investigations.

Observations and Recommendations

First, it is essential to notice that "Does religion cause prosociality?" lacks specificity. To refine this question, we must articulate clear causal contrasts and their scale, select specific measures of "prosociality," define our target population, gather appropriate time-series data, and, only after causal assumptions and identification criteria are satisfied and threats made explicit, calculate statistical estimates. Having obtained these estimates, which rely on assumptions, we must evaluate robustness using sensitivity analyses (Hernan and Robins 2024; Ogburn and Shpitser 2021; J. A. Bulbulia 2022; Linden, Mathur, and VanderWeele 2020; Tyler J. VanderWeele, Mathur, and Chen 2020).

Here, we obtain statistical estimates for the causal effects of religious service on charity and volunteering by (1) stating contrasts for specific interventions that increase and decrease religious service attendance across the population; (2) balancing the interventions to be compared on baseline confounders measured one year before treatment (including baseline measures of the treatment and outcomes) using flexible doubly robust machine learning ensembles with cross-validation; and (3) evaluating contrasts in expected average outcomes under different treatments one year after treatment.

In the scientifically interesting contrast condition, 'Regular vs Zero', we find considerable social benefits of regular religious attendance for charitable donations and volunteering. Moreover, we also find that regular religious service enacted in contemporary New Zealand would strongly enhance charity and volunteering above the status quo, a point that may inform discussions among critics of religion. Conversely, the overall social effects of altogether ceasing religious services are minimal compared with the status quo. Were New Zealand society to entirely forgo regular religious attendance, differences in charitable donations and volunteering one year later are unreliable. Although we did not investigate the long-term effects of the interventions considered here, this finding should alleviate some concerns among those wary of New Zealand's longstanding secular trend. Looking ahead, **we recommend that future research distinguish between models of treatment-gain and models of treatment-loss** (as exemplified by Van Tongeren et al. (2020)).

Second, we caution that **classical measures of effect, such as Cohen's *D* or R^2, may present a misleading indication of practical significance.** Our study computed standardised differences in the continuous charity and volunteering outcomes. The contrast between regular religious service and no service yields and effects size of 0.132 [0.102, 0.161]. This effect would be categorised as "small" effect by contemporary experimental conventions. Nonetheless, we project that the difference in annual charitable donations between the regular service condition and the status quo amounts to NZD 1638.98 versus NZD 1037.14. Not only is this effect practically significant at the individual level, summed over the population the difference amounts to 4% of New Zealand's Annual Government Budget (2021). **We recommend using causal inference to prioritise, evaluate and communicate practical effect sizes. Conventional statistical measures should not be mistaken for metrics of practical significance.**

Third, our novel measures of prosociality—**assistance and financial support received from one's community**– align with the expectations of self-reported data on charitable donations and volunteering. We detect marginal causal effects of religious service attendance on **measures of prosocial benefit and dependence**. Notice that just as our measures of prosocial actions (charity/volunteering) do not capture the targets of prosocial help, likewise our measures of community help received (time/money) do not capture the specific sources of help beyond the categories of "community," "friends," and "family". Nevertheless, the evolutionary theories of religious prosociality that motivate this study are grounded on within-group cooperative benefits (Sosis and Bressler 2003; D. D. Johnson 2015; J. Bulbulia 2012; Whitehouse et al. 2023; Schloss and Murray 2011). This consistency between self-reported giving and help received aligns with public data in New Zealand;[4] where religious institutional charity accounts for 40% of the charitable sector (refer to McLeod (2020), p.17, and also Brooks (2004); Woodyard and Grable (2014); Monsma (2007)). Additionally, religious institutions appear particularly efficient owing to low administrative costs and high volunteer engagement (Khanna, Posnett, and Sandler 1995; Bekkers and Wiepking 2011; McLeod 2020, 26). We note that the community signals of religious giving are consistent with the theory that religion promotes altruism outside families, friendships, and even among strangers (McCullough 2020). **In contexts where investigators are concerned about self-presentation biases when surveying charitable donations and volunteering, we recommend piloting and deploying measures that capture community-received assistance.**

Fourth, our secondary analysis reveals that standard cross-sectional regressions overstate the causal effects estimates obtained from a conscientious application of causal methods to time-series data. Importantly, associational models might also *understate* true causal relationships, especially if adjustments involve mediators (Westreich and Greenland 2013; McElreath 2020). Despite their effectiveness and relevance, causal inference methods remain uncommon in many human sciences, including the psychological sciences, where traditional associational methods still dominate. Transitioning to causal data science might seem daunting. Yet, causation inherently unfolds in time, with causes preceding effects. Estimating average causal effects involves multi-stepped workflows and robust time-series data (Neal 2020; J. A. Bulbulia 2023). Although the need for researchers to develop new skills imposes considerable demands, it is encouraging that methods for causal inference have become standard in epidemiology (Lash et al. 2020) and are rapidly gaining traction in economics (Angrist and Pischke 2009; Athey, Tibshirani, and Wager 2019; Athey and Wager 2021) and political science (Montgomery, Nyhan, and Torres 2018).

Moreover, a global panel studies suited for researching religion B. R. Johnson and VanderWeele (2022) offers exciting opportunities for psychological scientists to apply causal methods to those psychological questions at the heart of our discipline.

Also encouraging, well-designed panel studies investigating the causal effects of religious service attendance on health that capitalise on the affordances of time-series data provide models for studies that may investigate the social consequences of religion (refer to Y. Chen, Kim, and VanderWeele (2020), Pawlikowski et al. (2019),Li et al. (2016),Tyler J. VanderWeele (2021b), Kim and VanderWeele (2019)). At present, then, the intellectual benefits of retooling would seem considerable (refer to Major-Smith (2023)).

Setting intellectual benefits aside, it is vital to remember that investigators should be able to evaluate the practical interest of their findings coherently. Such evaluation relies on quantifying the effects of interventions. However,

---

[4]https://www.charities.govt.nz/view-data/

associational methods cannot deliver this understanding where causal inferential workflows are absent because association is not guaranteed to be causation (Westreich and Greenland 2013). What might be called a *practical significance gap* arises from associational models because we cannot interpret associations as causation unless we have obtained balance on the treatments to be compared, know that causes in our data precede effects in time, and ensure that identification conditions have been met. Notably, the *practical significance gap* remains for associational methods even when they are applied to time series data, and even when they are sophisticated—as in the statistical structural equation and multilevel modelling traditions (refer to Tyler J. VanderWeele (2021a); Hazlett and Wainstein (2022)). Yet, because investigators typically seek both scientific knowledge and practical understanding, **we strongly recommend the broader adoption of causal inferential methods in the psychological sciences.**

Fifth, and finally, we frame this study's contribution in light of future horizons. This study combined robust causal methods with national panel data to quantitatively investigate the social consequences of religious behaviour. Using both stated and revealed indicators, we provide evidence that religious service attendance causally affects charitable donations and volunteering. These findings offer insights into *how much* Religious service affects charity and volunteering in New Zealand and possibly in similar cultures. We hope this research will serve as a guide for studies in other settings.

However, even within New Zealand, much remains to be discovered. We have not attempted to elucidate heterogeneity in the treatment effects we observe. Nor have we investigated the pathways by which religious service attendance affects social behaviours. A substantial body of research has examined the psychological and cultural features through which religious behaviours affect individuals (Norenzayan et al. 2016; Shaver et al. 2020; J. Watts et al. 2016; McCullough et al. 2016; Konvalinka et al. 2011; Lang et al. 2015; Cristofori et al. 2016; Shaver 2015; Sosis 2005). Yet obtaining a causal understanding presents conceptual, modelling, and data challenges (J. M. Robins and Greenland 1992; Tyler J. VanderWeele 2015; Díaz, Williams, and Rudolph 2023; Díaz et al. 2021). In the years ahead, **we recommend the wider development and application of causal inferential methods employing repeated measurements over more than three years to investigate variation, interaction, and causal mediation.**

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.

Data Availability

The data described in the paper are part of the New Zealand Attitudes and Values Study (NZAVMembers 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: https://github.com/go-bayes/models/blob/main/scripts/24-bulbulia-church-prosocial.R.

Appendix A: Measures

Age (waves: 1-15)

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

Born in New Zealand

Charitable Donations (Study 1 outcome)

Using one item from Hoverd and Sibley (2010), we asked participants, "How much money have you donated to charity in the last year?".

Charitable Volunteering (Study 1 outcome)

We measured hours of volunteering using one item from Chris G. Sibley et al. (2011): "Hours spent … voluntary/charitable work."

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

We measured the number of children using one item from J. A. Bulbulia et al. (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).

Disability

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).

Education Attainment (waves: 1, 4-15)

We asked participants, "What is your highest level of qualification?". We coded participants' 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)".

Ethnicity

Based on the New Zealand Census, 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).

Fatigue

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).

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.
```

Hours of Childcare

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): 'Hours spent ... looking after children."

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Housework

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): "Hours spent ... housework/cooking"

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Exercise

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): "Hours spent ... exercising/physical activity"

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Childcare

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): 'Hours spent ... looking after children."

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Exercise

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): "Hours spent ... exercising/physical activity"

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Housework

We measured hours of exercising using one item from Chris G. Sibley et al. (2011): "Hours spent ... housework/cooking"

To stabilise this indicator, we took the natural log of the response + 1.

Hours of Sleep

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

Hours of Work

We measured work hours using one item from Chris G. Sibley et al. (2011): "Hours spent ... working in paid employment."

To stabilise this indicator, we took the natural log of the response + 1.

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

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

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

We measured psychological distress using the Kessler-6 scale (kessler2002?), which exhibits strong diagnostic concordance for moderate and severe psychological distress in large, crosscultural samples (kessler2010?; prochaska2012?). 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 alternatives 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."

Male 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 as 1, and gender diverse coded as 3 (Fraser et al. 2020). (or 0.5 = neither female nor male)

Here, we coded all those who responded as Male as 1, and those who did not as 0.

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

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

1. agreeableness,

    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)

2. conscientiousness,

    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)

3. extraversion,

    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.

4. honesty-humility,

    i. I feel entitled to more of everything. (r)
    ii. I deserve more things in life. (r)
    iii. I would like to be seen driving around in a very expensive car. (r)
    iv. I would get a lot of pleasure from owning expensive luxury goods. (r)

5. neuroticism, and

    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)

6. openness to experience

    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)

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.

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).

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.

NZSEI Occupational Prestige and Status (waves: 8-15)

We assessed occupational prestige and status using the New Zealand Socio-economic Index 13 (NZSEI-13) (Fahy, Lee, and Milne 2017a). 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 an NZSEI-13 score based on their occupation.

We assessed occupational prestige and status using the New Zealand Socio-economic Index 13 (NZSEI-13) (Fahy, Lee, and Milne 2017b). 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 an NZSEI-13 score based on their occupation.

Opt-in

The New Zealand Attitudes and Values Study allows opt-ins to the study. Because the opt-in population may differ from those sampled randomly from the New Zealand electoral roll; although the opt-in rate is low, we include an indicator (yes/no) for this variable.

Partner (No/Yes)

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

Politically Conservative

We measured participants' political conservative 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)

Religious Service Attendance

If participants answered *yes* to "Do you identify with a religion and/or spiritual group?" we measured their frequency of church attendence using one item from Sibley C. G. and Bulbulia (2012): "how many times did you attend a church or place of worship in the last month?". Those participants who were not religious were imputed a score of "0".

Rural/Urban Codes

Participants residence locations were coded according to a five-level ordinal categorisation ranging from "Urban" to Rural, see Chris G. Sibley (2021).

Short-Form Health

Participants' subjective health was measured using one item ("Do you have a health condition or disability that limits you, and that has lasted for 6+ months?"; 1 = Yes, 0 = No) adapted from Verbrugge (1997).

Sample Origin

Wave enrolled in NZAVS, see Chris G. Sibley (2021).

Support received: money (waves 10-12) (Study 4 outcomes)

The NZAVS has a 'revealed' measure of received help and support measured in hours of support in the previous week. The items are:

Please estimate how much help you have received from the following sources in the last week?

- *family...MONEY (hours)*
- *friends...MONEY (hours)*
- *members of my community...MONEY (hours)*

Because this measure is highly variable, we convert responses to binary indicators: *0 = none/1 any*

Support received: time (waves 10-13) (Study 3 outcomes)

Please estimate how much help you have received from the following sources in the last week.

- *family...TIME (hours)*
- *friends...TIME (hours)*
- *members of my community...TIME (hours)*

Because this measure is highly variable, we convert responses to binary indicators: *0 = none/1 any*

Total Siblings

Participants were asked the following questions related to sibling counts:

- Were you the 1st born, 2nd born, or 3rd born, etc, child of your mother?
- Do you have siblings?
- How many older sisters do you have?
- How many younger sisters do you have?
- How many older brothers do you have?
- How many younger brothers do you have?

A single score was obtained from sibling counts by summing responses to the "How many…" items. From these scores, an ordered factor was created ranging from 0 to 7, where participants with more than 7 siblings were grouped into the highest category.

Appendix B. Baseline Demographic Statistics

Table 12: Baseline demographic statistics

Exposure + Demographic Variables	N = 33,198
Age	NA
Mean (SD)	51 (14)
Range	18, 96
IQR	41, 61
Agreeableness	NA
Mean (SD)	5.37 (0.98)
Range	1.00, 7.00
IQR	4.75, 6.00
Unknown	272
Born Nz	26,197 (79%)
Unknown	34
Children Num	NA
Mean (SD)	1.76 (1.44)
Range	0.00, 14.00
IQR	0.00, 3.00
Conscientiousness	NA
Mean (SD)	5.14 (1.04)
Range	1.00, 7.00
IQR	4.50, 6.00
Unknown	266
Education Level Coarsen	NA
no_qualification	769 (2.3%)
cert_1_to_4	11,278 (34%)
cert_5_to_6	4,281 (13%)
university	8,947 (27%)
post_grad	3,892 (12%)
masters	2,956 (9.0%)
doctorate	891 (2.7%)
Unknown	184
Employed	26,379 (80%)
Unknown	26
Eth Cat	NA
euro	27,404 (83%)
maori	3,424 (10%)
pacific	707 (2.1%)
asian	1,438 (4.4%)
Unknown	225
Extraversion	NA
Mean (SD)	3.88 (1.20)
Range	1.00, 7.00
IQR	3.00, 4.75
Unknown	266
Hlth Disability	7,558 (23%)
Unknown	561
Hlth Fatigue	NA

Table 12: Baseline demographic statistics

Exposure + Demographic Variables	N = 33,198
0	5,289 (16%)
1	10,940 (33%)
2	10,196 (31%)
3	4,862 (15%)
4	1,577 (4.8%)
Unknown	334
Hlth Sleep Hours	NA
Mean (SD)	6.95 (1.11)
Range	2.50, 16.00
IQR	6.00, 8.00
Unknown	1,528
Honesty Humility	NA
Mean (SD)	5.49 (1.15)
Range	1.00, 7.00
IQR	4.75, 6.50
Unknown	269
Hours Children log	NA
Mean (SD)	1.10 (1.58)
Range	0.00, 5.13
IQR	0.00, 2.20
Unknown	875
Hours Exercise log	NA
Mean (SD)	1.57 (0.83)
Range	0.00, 4.39
IQR	1.10, 2.08
Unknown	875
Hours Housework log	NA
Mean (SD)	2.15 (0.77)
Range	0.00, 5.13
IQR	1.79, 2.71
Unknown	875
Hours Work log	NA
Mean (SD)	2.64 (1.59)
Range	0.00, 4.62
IQR	1.10, 3.71
Unknown	875
Household Inc log	NA
Mean (SD)	11.41 (0.76)
Range	0.69, 14.92
IQR	11.00, 11.92
Unknown	1,352
Kessler6 Sum	NA
Mean (SD)	5 (4)
Range	0, 24
IQR	2, 7
Unknown	297
Male	11,975 (36%)

Table 12: Baseline demographic statistics

Exposure + Demographic Variables	N = 33,198
Modesty	NA
Mean (SD)	6.03 (0.90)
Range	1.00, 7.00
IQR	5.50, 6.75
Unknown	11
Neuroticism	NA
Mean (SD)	3.45 (1.15)
Range	1.00, 7.00
IQR	2.50, 4.25
Unknown	274
Nz Dep2018	NA
Mean (SD)	4.69 (2.70)
Range	1.00, 10.00
IQR	2.00, 7.00
Unknown	233
Nzsei 13 l	NA
Mean (SD)	55 (16)
Range	10, 90
IQR	42, 69
Unknown	172
Openness	NA
Mean (SD)	4.99 (1.12)
Range	1.00, 7.00
IQR	4.25, 5.75
Unknown	267
Partner	24,869 (76%)
Unknown	422
Political Conservative	NA
1	1,777 (5.6%)
2	6,563 (21%)
3	6,505 (20%)
4	9,373 (29%)
5	4,813 (15%)
6	2,378 (7.5%)
7	483 (1.5%)
Unknown	1,306
Religion Church Round	NA
0	27,653 (83%)
1	1,077 (3.2%)
2	777 (2.3%)
3	658 (2.0%)
4	1,729 (5.2%)
5	336 (1.0%)
6	226 (0.7%)
7	74 (0.2%)
8	668 (2.0%)
Rural Gch 2018 l	NA

Table 12: Baseline demographic statistics

Exposure + Demographic Variables	N = 33,198
1	20,361 (62%)
2	6,390 (19%)
3	4,020 (12%)
4	1,816 (5.5%)
5	380 (1.2%)
Unknown	231
Sample Frame Opt in	1,107 (3.3%)
Sample Origin	NA
1-2	2,191 (6.6%)
3-3.5	1,664 (5.0%)
4	1,987 (6.0%)
5-6-7	3,203 (9.6%)
8-9	4,264 (13%)
10	19,889 (60%)
Short Form Health	NA
Mean (SD)	5.06 (1.16)
Range	1.00, 7.00
IQR	4.33, 6.00
Unknown	5
Total Siblings	NA
Mean (SD)	2.52 (1.80)
Range	0.00, 23.00
IQR	1.00, 3.00
Unknown	689

Table 12 baseline demographic statistics for couples who met inclusion criteria.

Appendix C: Treatment Statistics

Table 13: Exposures at baseline and baseline + 1 (treatment) wave

Exposure Variables by Wave	2018, N = 33,198	2019, N = 33,198
Religion Church Round	NA	NA
0	27,653 (83%)	28,028 (84%)
1	1,077 (3.2%)	896 (2.7%)
2	777 (2.3%)	737 (2.2%)
3	658 (2.0%)	639 (1.9%)
4	1,729 (5.2%)	1,672 (5.0%)
5	336 (1.0%)	308 (0.9%)
6	226 (0.7%)	205 (0.6%)
7	74 (0.2%)	68 (0.2%)
8	668 (2.0%)	645 (1.9%)
Unknown	0	0
Alert Level Combined	NA	NA
no_alert	33,198 (100%)	23,751 (72%)
early_covid	0 (0%)	3,643 (11%)
alert_level_1	0 (0%)	2,821 (8.5%)
alert_level_2	0 (0%)	836 (2.5%)
alert_level_2_5_3	0 (0%)	552 (1.7%)
alert_level_4	0 (0%)	1,595 (4.8%)
Unknown	0	0

tbl-table-exposures-code presents baseline (NZAVS time 10) and exposure wave (NZAVS time 11) statistics for the exposure variable: religious service attendance (range 0-8). Responses coded as eight or above were coded as "8". This decision to avoid spare treatments was based on theoretical grounds, namely, that daily exposure would be similar in its effects to more than daily exposure. We note that causal contrasts were obtained for projects with either no attendance or four or more visits per month. Hence this simplification of the measure is unlikely to affect theoretical and practical inferences. All models adjusted for the pandemic alert level because the treatment wave (NZAVS time 11) occurred during New Zealand's COVID-19 pandemic. The pandemic is not a "confounder" because a confounder must be related to the treatment and the outcome. At the end of the study, all participants had been exposed to the pandemic. However, to satisfy the causal consistency assumption, all treatments must be conditionally equivalent within levels of all covariates (Tyler J. VanderWeele and Hernan 2013). Because COVID affected the ability or willingness of individuals to attend religious service, we included the lockdown condition as a covariate (Chris G. Sibley 2021). To better enable conditional independence within levels of the treatment variable, we conditioned on the lead value of COVID-alert level at baseline. To mitigate systematic biases arising from attrition and missingness, the lmtp package uses inverse probability of censoring weights, which were used when estimating the causal effects of the exposure on the outcome.

Binary Transition Table for The Treatment

Table 14: Transition table for stability and change in regular religious service (4x per month) between baseline and treatment wave.

From	>=4	< 4
>=4	**29496**	669

Table 14: Transition table for stability and change in regular religious service (4x per month) between baseline and treatment wave.

From	>=4	< 4
< 4	804	**2229**

Table 14 presents a transition matrix to evaluate treatment shifts between baseline and treatment wave. Here, we focus on the shift from/to monthly attendance at four or more visits per month. Entries along the diagonal (in bold) indicate the number of individuals who **stayed** in their initial state. By contrast, the off-diagonal shows the transitions from the initial state (bold) to another state in the following wave (off diagonal). Thus the cell located at the intersection of row i and column j, where $i \neq j$, gives us the counts of individuals moving from state i to state j.

Table 15: Transition table for stability and change in zero religious service (0 x per month) between baseline and treatment wave.

From	0	> 0
0	**26762**	891
> 0	1266	**4279**

Table 15 presents a transition matrix to evaluate treatment shifts between baseline and treatment wave. Here, we focus on the shift from/to zero religious service attendance. Again, entries along the diagonal (in bold) indicate the number of individuals who **stayed** in their initial state. By contrast, the off-diagonal shows the transitions from the initial state (bold) to another state in the following wave (off diagonal). Thus the cell located at the intersection of row i and column j, where $i \neq j$, gives us the counts of individuals moving from state i to state j.

Imbalance of Confounding Covariates Treatments

Figure 5 shows imbalance of covariates on the treatment at the treatment wave. The variable on which there is strongest imbalance is the baseline measure of religious service attendance. It is important to adjust for this measure both for confounding control and to better estimate an incident exposure effect for the religious service at the treatment wave (in contrast to merely estimating a prevalence effect). See Tyler J. VanderWeele, Mathur, and Chen (2020).

Appendix D: Baseline and End of Study Outcome Statistics

Table 16: Outcomes at baseline and end-of-study

Outcome Variables by Wave	**2018**, N = 33,198	**2020**, N = 33,198
Annual Charity	150 (40, 500)	200 (20, 600)
Unknown	1,076	6,730
Community Gives Money Binary	135 (0.4%)	118 (0.4%)
Unknown	669	6,959
Community Gives Time Binary	1,702 (5.2%)	1,669 (6.4%)
Unknown	669	6,959
Family Gives Money Binary	1,782 (5.5%)	1,236 (4.7%)
Unknown	669	6,959
Family Gives Time Binary	9,539 (29%)	7,600 (29%)

Table 16: Outcomes at baseline and end-of-study

Outcome Variables by Wave	**2018**, N = 33,198	**2020**, N = 33,198
Unknown	669	6,959
Friends Give Money Binary	372 (1.1%)	270 (1.0%)
Unknown	669	6,959
Friends Give Time	5,765 (18%)	4,855 (19%)
Unknown	669	6,959
Sense Neighbourhood Community	NA	NA
1	1,976 (6.0%)	1,124 (4.2%)
2	4,037 (12%)	2,703 (10%)
3	4,796 (15%)	3,561 (13%)
4	6,840 (21%)	5,809 (22%)
5	7,088 (21%)	6,477 (24%)
6	5,753 (17%)	5,060 (19%)
7	2,550 (7.7%)	2,104 (7.8%)
Unknown	158	6,360
Social Belonging	5.33 (4.33, 6.00)	5.33 (4.33, 6.00)
Unknown	268	6,418
Social Support	6.33 (5.33, 7.00)	6.33 (5.33, 7.00)
Unknown	19	6,286
Volunteering Hours	0.00 (0.00, 1.00)	0.00 (0.00, 1.00)
Unknown	875	6,863
Volunteers Binary	9,443 (29%)	6,881 (26%)
Unknown	875	6,863

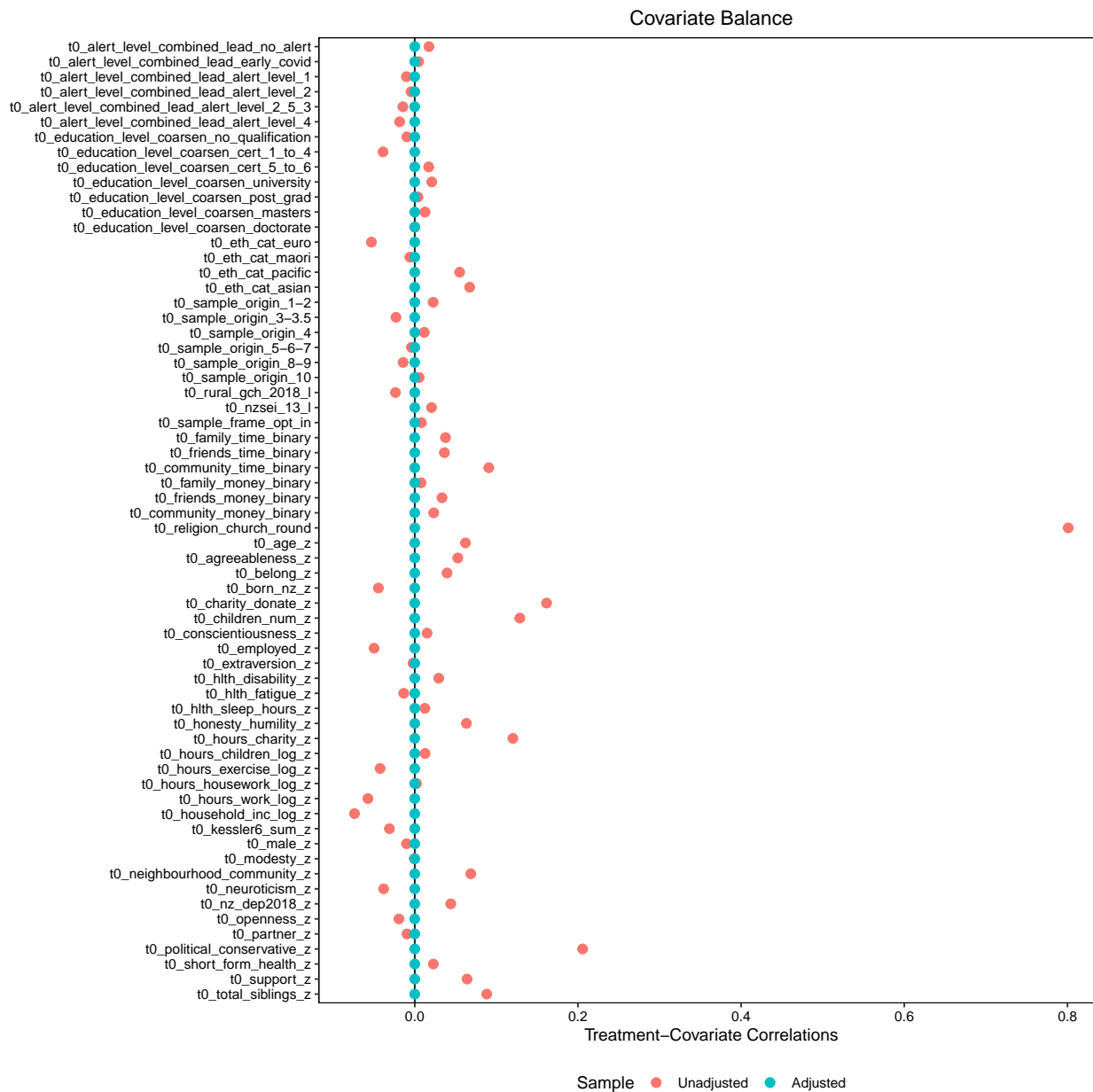Table 16 presents baseline and end-of-study descriptive statistics for the outcome variables.

Figure 5: This figure shows the imbalance in covariates on the treatment

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton university press.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *The Annals of Statistics* 47 (2): 1148–78. https://doi.org/10.1214/18-AOS1709.

Athey, Susan, and Stefan Wager. 2021. "Policy Learning With Observational Data." *Econometrica* 89 (1): 133–61. https://doi.org/10.3982/ECTA15732.

Atkinson, J, C Salmond, and P Crampton. 2019. "NZDep2018 Index of Deprivation, User's Manual." Wellington.

Bekkers, René, and Pamala Wiepking. 2011. "A Literature Review of Empirical Studies of Philanthropy: Eight Mechanisms That Drive Charitable Giving." *Nonprofit and Voluntary Sector Quarterly* 40 (5): 924–73.

Brooks, Arthur C. 2004. "Faith, Secularism, and Charity." *Faith & Economics* 43 (Spring): 1–8.

Bulbulia, J. A. 2022. "A Workflow for Causal Inference in Cross-Cultural Psychology." *Religion, Brain & Behavior* 0 (0): 1–16. https://doi.org/10.1080/2153599X.2022.2070245.

———. 2023. "Causal Diagrams (Directed Acyclic Graphs): A Practical Guide."

Bulbulia, J. A., J. H. Shaver, L. Greaves, R. Sosis, and C. G. Sibley. 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 Slone J. D. amd Van Slyke J., 29–62. Bloomsbury Press.

Bulbulia, Joseph. 2012. "Spreading Order: Religion, Cooperative Niche Construction, and Risky Coordination Problems." *Biology & Philosophy* 27: 1–27.

———. 2024. "A Practical Guide to Causal Inference in Three-Wave Panel Studies." *PsyArXiv Preprints*, February. https://doi.org/10.31234/osf.io/uyg3d.

Bulbulia, Joseph A. 2024. *Margot: MARGinal Observational Treatment-Effects.* https://doi.org/10.5281/zenodo.10907724.

Bulbulia, Joseph A, M Usman Afzali, Kumar Yogeeswaran, and Chris G Sibley. 2023. "Long-Term Causal Effects of Far-Right Terrorism in New Zealand." *PNAS Nexus* 2 (8): pgad242.

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.

Chatton, Arthur, Florent Le Borgne, Clémence Leyrat, Florence Gillaizeau, Chloé Rousseau, Laetitia Barbin, David Laplaud, Maxime Léger, Bruno Giraudeau, and Yohann Foucher. 2020. "G-Computation, Propensity Score-Based Methods, and Targeted Maximum Likelihood Estimator for Causal Inference with Different Covariates Sets: A Comparative Simulation Study." *Scientific Reports* 10 (1): 9219. https://doi.org/10.1038/s41598-020-65917-x.

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, et al. 2023. *Xgboost: Extreme Gradient Boosting.* https://CRAN.R-project.org/package=xgboost.

Chen, Ying, Eric S Kim, and Tyler J VanderWeele. 2020. "Religious-Service Attendance and Subsequent Health and Well-Being Throughout Adulthood: Evidence from Three Prospective Cohorts." *International Journal of Epidemiology* 49 (6): 2030–40.

Cristofori, Irene, Joseph Bulbulia, John H Shaver, Marc Wilson, Frank Krueger, and Jordan Grafman. 2016. "Neural Correlates of Mystical Experience." *Neuropsychologia* 80: 212–20.

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.

De Coulanges, Fustel. 1903. *La Cité Antique: Étude Sur Le Culte, Le Droit, Les Institutions de La Grèce Et de Rome.* Hachette.

Díaz, Iván, Nicholas Williams, Katherine L. Hoffman, and Edward J. Schenck. 2021. "Non-Parametric Causal Effects Based on Longitudinal Modified Treatment Policies." *Journal of the American Statistical Association.* https://doi.org/10.1080/01621459.2021.1955691.

———. 2023. "Nonparametric Causal Effects Based on Longitudinal Modified Treatment Policies." *Journal of the American Statistical Association* 118 (542): 846–57. https://doi.org/10.1080/01621459.2021.1955691.

Díaz, Iván, Nima S Hejazi, Kara E Rudolph, and Mark J van Der Laan. 2021. "Nonparametric Efficient Causal

Mediation with Intermediate Confounders." *Biometrika* 108 (3): 627–41.

Díaz, Iván, Nicholas Williams, and Kara E. Rudolph. 2023. *Journal of Causal Inference* 11 (1): 20220077. https://doi.org/doi:10.1515/jci-2022-0077.

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

———. 2017a. *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.

Haneuse, Sebastian, and Andrea Rotnitzky. 2013. "Estimation of the Effect of Interventions That Modify the Received Treatment." *Statistics in Medicine* 32 (30): 5260–77.

Hazlett, Chad, and Leonard Wainstein. 2022. "Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches." *Political Analysis* 30 (1): 46–65.

Hernan, M. A., and J. M. Robins. 2024. *Causal Inference: What If?* Chapman & Hall/CRC Monographs on Statistics & Applied Probab. Taylor & Francis. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

Hernán, Miguel A, and Sander Greenland. 2024. "Why Stating Hypotheses in Grant Applications Is Unnecessary." *JAMA* 331 (4): 285–86.

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: 70–75.

Hoffman, Katherine L., Diego Salazar-Barreto, Kara E. Rudolph, and Iván Díaz. 2023. "Introducing Longitudinal Modified Treatment Policies: A Unified Framework for Studying Complex Exposures," April. https://doi.org/10.48550/arXiv.2304.09460.

Hoffman, Katherine L., Edward J. Schenck, Michael J. Satlin, William Whalen, Di Pan, Nicholas Williams, and Iván Díaz. 2022. "Comparison of a Target Trial Emulation Framework Vs Cox Regression to Estimate the Association of Corticosteroids with COVID-19 Mortality." *JAMA Network Open* 5 (10): e2234425. https://doi.org/10.1001/jamanetworkopen.2022.34425.

Hoverd, William James, and Chris G Sibley. 2010. "Religious and Denominational Diversity in New Zealand 2009." *New Zealand Sociology* 25 (2): 59–87.

Huitfeldt, Anders, Mats J Stensrud, and Etsuji Suzuki. 2019. "On the Collapsibility of Measures of Effect in the Counterfactual Causal Framework." *Emerging Themes in Epidemiology* 16: 1–5.

Johnson, Byron R, and Tyler J VanderWeele. 2022. "The Global Flourishing Study: A New Era for the Study of Well-Being." *International Bulletin of Mission Research* 46 (2): 272–75.

Johnson, Dominic DP. 2005. "God's Punishment and Public Goods: A Test of the Supernatural Punishment Hypothesis in 186 World Cultures." *Human Nature* 16: 410–46.

———. 2015. "Big Gods, Small Wonder: Supernatural Punishment Strikes Back." *Religion, Brain & Behavior* 5 (4): 290–98.

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.

Kelly, John Michael, Stephanie R Kramer, and Azim F Shariff. 2024. "Religiosity Predicts Prosociality, Especially When Measured by Self-Report: A Meta-Analysis of Almost 60 Years of Research." *Psychological Bulletin* 150 (3): 284–318.

Khanna, Jyoti, John Posnett, and Todd Sandler. 1995. "Charity Donations in the UK: New Evidence Based on Panel Data." *Journal of Public Economics* 56 (2): 257–72.

Kim, Eric S, and Tyler J VanderWeele. 2019. "Mediators of the Association Between Religious Service Attendance and Mortality." *American Journal of Epidemiology* 188 (1): 96–101.

Konvalinka, Ivana, Dimitris Xygalatas, Joseph Bulbulia, Uffe Schjoedt, Else-Marie Jegindo, Sebastian Wallot, Guy Van Orden, and Andreas Roepstorff. 2011. "Synchronized Arousal Between Performers and Related Spectators in a Fire-Walking Ritual." *Proceedings of the National Academy of Sciences* 108 (20): 8514–19.

Laan, Mark J van der, and Susan Gruber. 2012. "Targeted Minimum Loss Based Estimation of Causal Effects of

Multiple Time Point Interventions." *The International Journal of Biostatistics* 8 (1).

Laan, Mark J van der, Alexander R Luedtke, and Iván Díaz. 2014. "Discussion of Identification, Estimation and Approximation of Risk Under Interventions That Depend on the Natural Value of Treatment Using Observational Data, by Jessica Young, Miguel Hernán, and James Robins." *Epidemiologic Methods* 3 (1): 21–31.

Lang, Martin, Jan Krátkỳ, John H Shaver, Danijela Jerotijević, and Dimitris Xygalatas. 2015. "Effects of Anxiety on Spontaneous Ritualized Behavior." *Current Biology* 25 (14): 1892–97.

Lash, T. L., K. J. Rothman, T. J. VanderWeele, and S. Haneuse. 2020. *Modern Epidemiology*. Wolters Kluwer. https://books.google.co.nz/books?id=SiTSnQEACAAJ.

Li, Shanshan, Meir J. Stampfer, David R. Williams, and Tyler J. VanderWeele. 2016. "Association of Religious Service Attendance With Mortality Among Women." *JAMA Internal Medicine* 176 (6): 777–85. https://doi.org/10.1001/jamainternmed.2016.1615.

Linden, Ariel, Maya B Mathur, and Tyler J VanderWeele. 2020. "Conducting Sensitivity Analysis for Unmeasured Confounding in Observational Studies Using e-Values: The Evalue Package." *The Stata Journal* 20 (1): 162–75.

Major-Smith, Daniel. 2023. "Exploring Causality from Observational Data: An Example Assessing Whether Religiosity Promotes Cooperation." *Evolutionary Human Sciences* 5: e22.

McCullough, Michael E. 2020. *The Kindness of Strangers: How a Selfish Ape Invented a New Moral Code*. Simon; Schuster.

McCullough, Michael E, Paul Swartwout, John H Shaver, Evan C Carter, and Richard Sosis. 2016. "Christian Religious Badges Instill Trust in Christian and Non-Christian Perceivers." *Psychology of Religion and Spirituality* 8 (2): 149.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan*. CRC press.

McLeod, John. 2020. "The New Zealand Support Report: The Current State and Significance of Giving in New Zealand and the Outlook for Recipients." JBWere. https://www.jbwere.co.nz/media/1qudxw3q/jbwere-nz-support-report-digital.pdf.

Monsma, Stephen V. 2007. "Religion and Philanthropic Giving and Volunteering: Building Blocks for Civic Responsibility." *Interdisciplinary Journal of Research on Religion* 3.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. 2018. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do about It." *American Journal of Political Science* 62 (3): 760–75. https://doi.org/10.1111/ajps.12357.

Muñoz, Iván Díaz, and Mark Van Der Laan. 2012. "Population Intervention Causal Effects Based on Stochastic Interventions." *Biometrics* 68 (2): 541–49.

Neal, Brady. 2020. "Introduction to Causal Inference from a Machine Learning Perspective." *Course Lecture Notes (Draft)*. https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf.

Norenzayan, Ara, Azim F. Shariff, Will M. Gervais, Aiyana K. Willard, Rita A. McNamara, Edward Slingerland, and Joseph Henrich. 2016. "The Cultural Evolution of Prosocial Religions." *Behavioral and Brain Sciences* 39 (January): e1. https://doi.org/10.1017/S0140525X14001356.

Ogburn, Elizabeth L., and Ilya Shpitser. 2021. "Causal Modelling: The Two Cultures." *Observational Studies* 7 (1): 179–83. https://doi.org/10.1353/obs.2021.0006.

Pawlikowski, Jakub, Piotr Białowolski, Dorota Węziak-Białowolska, and Tyler J VanderWeele. 2019. "Religious Service Attendance, Health Behaviors and Well-Being—an Outcome-Wide Longitudinal Analysis." *European Journal of Public Health* 29 (6): 1177–83.

Pearl, Judea. 2009. "Causal Inference in Statistics: An Overview." https://doi.org/10.1214/09-SS057.

Polley, Eric, Erin LeDell, Chris Kennedy, and Mark van der Laan. 2023. *SuperLearner: Super Learner Prediction*. https://CRAN.R-project.org/package=SuperLearner.

Richardson, Thomas S, Robin J Evans, James M Robins, and Ilya Shpitser. 2023. "Nested Markov Properties for Acyclic Directed Mixed Graphs." *The Annals of Statistics* 51 (1): 334–61.

Richardson, Thomas S, and James M Robins. 2013. "Single World Intervention Graphs: A Primer." In *Second UAI Workshop on Causal Structure Learning, Bellevue, Washington*. Citeseer. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=07bbcb458109d2663acc0d098e8913892389a2a7.

———. 2023. "Potential Outcome and Decision Theoretic Foundations for Statistical Causality." *Journal of Causal Inference* 11 (1): 20220012.

Robins, James. 1986. "A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect." *Mathematical Modelling* 7 (9-12): 1393–1512.

Robins, James M, and Sander Greenland. 1992. "Identifiability and Exchangeability for Direct and Indirect Effects." *Epidemiology* 3 (2): 143–55.

Robins, James M, and Thomas S Richardson. 2010. "Alternative Graphical Causal Models and the Identification of Direct Effects." *Causality and Psychopathology: Finding the Determinants of Disorders and Their Cures* 84: 103–58.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions." *Journal of the American Statistical Association* 100 (469): 322–31. https://www.jstor.org/stable/27590541.

Schloss, Jeffrey P, and Michael J Murray. 2011. "Evolutionary Accounts of Belief in Supernatural Punishment: A Critical Review." *Religion, Brain & Behavior* 1 (1): 46–99.

Shaver, John H. 2015. "The Evolution of Stratification in Fijian Ritual Participation." *Religion, Brain & Behavior* 5 (2): 101–17.

Shaver, John H, Eleanor A Power, Benjamin G Purzycki, Joseph Watts, Rebecca Sear, Mary K Shenk, Richard Sosis, and Joseph A Bulbulia. 2020. "Church Attendance and Alloparenting: An Analysis of Fertility, Social Support and Child Development Among English Mothers." *Philosophical Transactions of the Royal Society B* 375 (1805): 20190428.

Shiba, Koichiro, and Takuya Kawahara. 2021. "Using Propensity Scores for Causal Inference: Pitfalls and Tips." *Journal of Epidemiology* 31 (8): 457–63.

Shpitser, Ilya, Thomas S Richardson, and James M Robins. 2022. "Multivariate Counterfactual Systems and Causal Graphical Models." In *Probabilistic and Causal Inference: The Works of Judea Pearl*, 813–52.

Shpitser, Ilya, and Eric Tchetgen Tchetgen. 2016. "Causal Inference with a Graphical Hierarchy of Interventions." *Annals of Statistics* 44 (6): 2433.

Sibley, C. G., and J. A. Bulbulia. 2012. "Healing Those Who Need Healing: How Religious Practice Affects Social Belonging." *Journal for the Cognitive Science of Religion* 1: 29–45.

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.

Sosis, Richard. 2005. "Does Religion Promote Trust?: The Role of Signaling, Reputation, and Punishment." *Interdisciplinary Journal of Research on Religion* 1.

Sosis, Richard, and Eric R Bressler. 2003. "Cooperation and Commune Longevity: A Test of the Costly Signaling Theory of Religion." *Cross-Cultural Research* 37 (2): 211–39.

Splawa-Neyman, Jerzy, Dorota M Dabrowska, and Terrence P Speed. 1990. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science*, 465–72.

Swanson, Guy E. 1967. "Religion and Regime: A Sociological Account of the Reformation."

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

Van der Laan, Mark J. 2014. "Targeted Estimation of Nuisance Parameters to Obtain Valid Statistical Inference." *The International Journal of Biostatistics* 10 (1): 29–57.

Van Der Laan, Mark J., and Sherri Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer Series in Statistics. New York, NY: Springer. https://link.springer.com/10.1007/978-1-4419-9782-1.

———. 2018. *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies.* Springer Series in Statistics. Cham: Springer International Publishing. http://link.springer.com/10.1007/978-3-319-65304-4.

Van Tongeren, Daryl R, C Nathan DeWall, Zhansheng Chen, Chris G Sibley, and Joseph Bulbulia. 2020. "Religious Residue: Cross-Cultural Evidence That Religious Psychology and Behavior Persist Following Deidentification." *Journal of Personality and Social Psychology*.

VanderWeele, Tyler J. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction.* Oxford University Press.

———. 2017. "Causal Effects of Religious Service Attendance?" *Social Psychiatry and Psychiatric Epidemiology* 52: 1331–36.

———. 2019. "Principles of Confounder Selection." *European Journal of Epidemiology* 34 (3): 211–19.

———. 2021a. "Can Sophisticated Study Designs with Regression Analyses of Observational Data Provide Causal Inferences?" *JAMA Psychiatry* 78 (3): 244–46.

———. 2021b. "Effects of Religious Service Attendance and Religious Importance on Depression: Examining the Meta-Analytic Evidence." *The International Journal for the Psychology of Religion* 31 (1): 21–26.

VanderWeele, Tyler J. 2009. "Concerning the Consistency Assumption in Causal Inference." *Epidemiology* 20 (6): 880. https://doi.org/10.1097/EDE.0b013e3181bd5638.

VanderWeele, Tyler J., and Peng Ding. 2017. "Sensitivity Analysis in Observational Research: Introducing the e-Value." *Annals of Internal Medicine* 167 (4): 268–74. https://doi.org/10.7326/M16-2607.

VanderWeele, Tyler J, and Miguel A Hernan. 2013. "Causal Inference Under Multiple Versions of Treatment." *Journal of Causal Inference* 1 (1): 1–20.

VanderWeele, Tyler J., and Miguel A. Hernán. 2012. "Results on Differential and Dependent Measurement Error of the Exposure and the Outcome Using Signed Directed Acyclic Graphs." *American Journal of Epidemiology* 175 (12): 1303–10. https://doi.org/10.1093/aje/kwr458.

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): 437–66.

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.

Watts, J., Bulbulia J. A., R. D. Gray, and Q. D. Atkinson. 2016. "Clarity and Causality Needed in Claims about Big Gods" 39: 41–42. https://doi.org/DOI:10.1017/S0140525X15000576.

Watts, Joseph, Simon J Greenhill, Quentin D Atkinson, Thomas E Currie, Joseph Bulbulia, and Russell D Gray. 2015. *Broad Supernatural Punishment but Not Moralizing High Gods Precede the Evolution of Political Complexity in Austronesia. Proceedings of the Royal Society B: Biological Sciences.* Vol. 282. 1804. The Royal Society.

Westreich, Daniel, and Stephen R. Cole. 2010. "Invited commentary: positivity in practice." *American Journal of Epidemiology* 171 (6). https://doi.org/10.1093/aje/kwp436.

Westreich, Daniel, and Sander Greenland. 2013. "The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients." *American Journal of Epidemiology* 177 (4): 292–98.

Wheatley, Paul. 1971. *The Pivot of the Four Quarters : A Preliminary Enquiry into the Origins and Character of the Ancient Chinese City.* Edinburgh University Press. https://cir.nii.ac.jp/crid/1130000795717727104.

Whitehouse, Harvey, Pieter Francois, Patrick E. Savage, Daniel Hoyer, Kevin C. Feeney, Enrico Cioni, Rosalind Purcell, et al. 2023. "Testing the Big Gods Hypothesis with Global Historical Data: A Review and Retake." *Religion, Brain & Behavior* 13 (2): 124–66.

Williams, Nicholas T., and Iván Díaz. 2021. *lmtp: Non-Parametric Causal Effects of Feasible Interventions Based on Modified Treatment Policies.* https://doi.org/10.5281/zenodo.3874931.

Woodyard, Ann, and John Grable. 2014. "Doing Good and Feeling Well: Exploring the Relationship Between Charitable Activity and Perceived Personal Wellness." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 25: 905–28.

Wright, Marvin N., and Andreas Ziegler. 2017. "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77 (1): 1–17. https://doi.org/10.18637/jss.v077.i01.

Young, Jessica G, Miguel A Hernán, and James M Robins. 2014. "Identification, Estimation and Approximation of Risk Under Interventions That Depend on the Natural Value of Treatment Using Observational Data." *Epidemiologic Methods* 3 (1): 1–19.

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.