PNAS nexus

PNAS Nexus, 2023, **2**, 1–16
https://doi.org/10.1093/pnasnexus/pgad242
Advance access publication 22 August 2023
Research Report

Long-term causal effects of far-right terrorism in New Zealand

Joseph A. Bulbulia[a,b,c,*], M. Usman Afzali[c], Kumar Yogeeswaran[c] and Chris G. Sibley[d]

[a]School of Psychology, Victoria University of Wellington, Wellington, New Zealand
[b]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[c]School of Psychology, Speech, Hearing, University of Canterbury, Christchurch, New Zealand
[d]School of Psychology, University of Auckland, Auckland, New Zealand
*To whom correspondence should be addressed: Email: joseph.bulbulia@vuw.ac.nz
Edited By: M. Gelfand

Abstract

The Christchurch mosque attacks in 2019, committed by a radical right-wing extremist, resulted in the tragic loss of 51 lives. Following these events, there was a noticable rise in societal acceptance of Muslim minorities. Comparable transient reactions have been observed elsewhere. However, the critical questions remain: can these effects endure? Are enduring effects evident across the political spectrum? It is challenging to answer such questions because identifying long-term causal effects requires estimating unobserved attitudinal trajectories without the attacks. Here, we use six preattack waves of Muslim acceptance responses from the New Zealand Attitudes and Values Study (NZAVS) to infer missing counterfactual trajectories (NZAVS cohort 2012, $N = 4{,}865$; replicated in 2013 cohort, $N = 7{,}894$). We find (1) the attacks initially boosted Muslim acceptance; (2) the magnitude of the initial Muslim acceptance boost was similar across the political spectrum; (3) no changes were observed in negative control groups; and (4) two- and three-year effects varied by baseline political orientation: liberal acceptance was stable, conservative acceptance grew relative to the counterfactual trend. Overall, the attacks added five years of growth in Muslim acceptance, with no regression to preattack levels over time. Continued growth among conservatives highlights the attack's failure to divide society. These results demonstrate the utility of combining methods for causal inference with national-scale panel data to answer psychological questions of basic human concern.

Keywords: causal inference, counterfactual, far-right extremism, longitudinal, prejudice

Significance Statement

The enduring impact of terrorist attacks on the perceptions of targeted minorities remains an unexplored area of study, despite its profound theoretical and practical implications. In this study, we leverage a causal identification strategy derived from quantitative epidemiology and extensive longitudinal data from the New Zealand Attitudes and Value Study to infer the multiyear influence of the 2019 Christchurch Mosque attacks on the acceptance of Muslims in New Zealand. Our findings reveal that acceptance of Muslims persisted for at least three years following the Christchurch mosque attacks. Contrary to the divisive intentions of the terrorist, our study discovers a long-term increase in acceptance.

On 2019 March 15, a far-right extremist attacked two mosques in Christchurch, New Zealand, killing 51 Muslims. Another 49 Muslims were injured (1). Over 250 survivors witnessed the murders first-hand. The attacks were broadcast globally through a Facebook live feed (2, 3).

Apart from their brutality, the attacks put a spotlight on New Zealand's Muslim minority, which totals 1.3% of the general population (4), and has a long history of peaceful settlement. Despite their small population and peaceful history, anti-Muslim prejudice before the attacks exceeded that of any other measured group (4, 5) (online Supplementary material S1). Indeed, at the time of the attacks, Muslims in New Zealand were perceived to be more threatening than any other religious or ethnic group (6) (online Supplementary material S1).

Terrorists attempt to harm targeted communities in two ways: (1) directly by injury and (2) indirectly by shifting public opinion against the targeted communities (7, 8). However, psychological research suggests that terrorists do not generally succeed in their second objective: initially, attacked minorities experience greater acceptance (8, 9). Terrorism diminishes minority acceptance only when the minority group is associated with the violence (10–15). A fundamental question is whether the boost in acceptance of targeted minorities that follows extremist attacks is transitory.

Meaning of the term "attack"

Here, we use the term "attack" to denote both (1) an act or acts of violence against members of a targeted group and (2) the public,

Competing Interest: The authors declare no competing interest.
Received: July 25, 2022. **Revised:** June 19, 2023. **Accepted:** July 19, 2023

government, and media response to these acts. That is, we use "attack" to denote a complex attack–response–sequence (online Supplementary material S1). Disentangling the direct and indirect causal effects of the components of this complex attack–response–sequence is not possible (16).[a] Nevertheless, it may be possible to estimate the total effects of this attack–response–sequence. The present study attempts to estimate these total effects on Muslim acceptance at +1, +2, and +3 years following the attacks.

Theoretical interest of the long-term marginal effect of the attacks on Muslim acceptance

Although previous work has considered the short-term effects of terrorism on attitudes toward the targeted group, it is unclear how sustained such effects may be. For example, Shanaah et al. investigated the short-term effects of the 2019 Christchurch mosque attacks on prejudice by applying regression discontinuity analysis to a rolling sample in the New Zealand Attitudes and Values Study (NZAVS) ($n_{\text{postattack}} = 8{,}180$) (20). In the 90 days following the attacks, there was an 8% increase in warmth ratings of Muslims (20). Notably, the authors observed that the initial boost to Muslim acceptance declined among political conservatives.

In support of a transitory acceptance-effect hypothesis, research suggests that attitudes to the government following terrorist attacks are transitory (21). For example, the September 2001 attacks in the United States initially boosted President George W. Bush's approval ratings to 90% favorable (22); however, Bush's approval steadily declined thereafter (23). Moreover, following the 2019 Christchurch mosque attacks in New Zealand, a substantial increase in government satisfaction lasted for only three months (24). Thus, the durability of any effects following a terrorist attack remains unclear.

Theoretical interest of the long-term conditional effect of the attacks on Muslim acceptance within strata of political orientation

As the 2019 Christchurch mosque attacks were carried out by a far-right extremist to rally political conservatives against Muslims, it is important to examine if the attacks had varied implications across the political spectrum. Did the terrorism evoke this intended response, and if so, for how long? Evidence from the European Social Survey (2014–2019) and the Anti-Refugee Violence dataset in Germany (2014–2017) suggests that exposure to collective violence can foster closer alignment with radical right parties among those holding prejudicial views towards immigrants (25). The generality of such responses, however, remains unclear.

Correlational evidence based on data from Germany reveals an association between higher rates of antirefugee violence and increased negative attitudes towards refugees among German natives (26). These findings suggest that for people on the right, seeing antirefugee violence increased their antirefugee sentiments. However, both causality and generalization remain unclear.

It is theoretically possible that conservatives expressed greater acceptance of Muslims following the attacks, perhaps as a way of distancing themselves from violent far-right extremism. As mentioned, a recent study found that there was, contrary to the terrorist's intention, a boost to Muslim acceptance among conservatives, but that this boost declined during the 90 days following the attacks (20). Unfortunately, the sample was relatively small. Moreover, the reliability of conditional comparisons in a rolling sample may be questioned: if the attacks were to affect political orientation, conditioning on postexposure political orientation would bias effect estimates (16, 27). Thus, assessing longer term attitudes, free from the immediate effects of the attack, is necessary to provide a clearer picture of the magnitude and duration of effect modification among populations who differ in political orientation.

In short, previous studies find that the impact of far-right terrorist attacks might vary among populations with differing political orientations; however, the direction, magnitude, and duration of such effects remain unclear. Some research suggests that conservatives will show greater acceptance of Muslims to distance themselves from the attack's extreme violence. Other research suggests that exposure to such violence might push those with prejudicial views closer to far-right ideologies. To examine these possibilities, we use a combination of national-scale panel data collected over 10 years and robust causal estimation techniques.

Causal identification strategy

Measuring the enduring psychological effects of terrorism is a complex task, primarily because it necessitates answering two questions: first, "What occurred in the population exposed to the terror?" and second, "What would have occurred if the population had not been exposed to terror?" The basis of all causal estimation methodologies, including randomized experiments, is the exchangeability assumption between the exposed and unexposed groups. In the context of the 2019 Christchurch mosque attacks, we assume that the entire New Zealand population was exposed to the events (online Supplementary material S1 for more details). It is plausible to consider that there is exchangeability between the population just before and just after the attacks (28, 29). However, we must also take into account the fact that Muslim acceptance had been on the rise in New Zealand even before the attacks (5). As such, estimating the long-term attack effects without considering preattack acceptance trajectories would be unwise. This frames our challenge: to accurately quantify the lasting psychological effects of the 2019 Christchurch mosque attacks on New Zealanders, we need to derive counterfactual comparisons following a population-wide exposure event.

Here, we utilize 10 years (2012–2022) of longitudinal data from the NZAVS with a cohort size of $N = 4{,}865$. Our analysis builds on a clock-like linear pattern of increasing acceptance of Muslims before the attacks. It aims to discern the effects of a population-wide event—in this case, a terrorist attack—on societal attitudes.

Our analysis is grounded in the principles of causal inference. Even though causal inference comes with certain assumptions, our robust, long-term panel data lends credibility to its use. In essence, causal inference pertains to the unobservable—we compare potential outcomes for the entire target population (or for a full stratum of the population) under two different conditions at most only one of which may be experienced by individuals in the target population. In our study, we must credibly predict how acceptance of Muslims might have evolved if the attack had not taken place.

To accomplish this, we assume that the observed linear preattack trend for increasing acceptance of Muslims would have persisted unchanged. This assumption may appear strong but is backed by observations from previous years (30). Moreover, we test this assumption in preattack responses by deleting and then recovering missing responses using our imputation strategy (online Supplementary material S8). In the Discussion section, we consider how our study's conclusions might be affected if the counterfactual acceptance trend were not linear. We hope that our research provides meaningful insights not just into the specific question of Muslim acceptance in New Zealand, but also into how causal inference can be applied more extensively in the behavioral sciences to move beyond simple reporting of longitudinal correlations. Again, we extend our focus beyond the marginal effects of the attacks on acceptance across the population and develop a conditional analysis that enables us to examine changes among people who occupy different positions on the political spectrum. We examine "effect modification" rather than "interaction" or "moderation" because the analysis of effect modification requires fewer assumptions than does the analysis of causal interaction. To better explain, consider the idea of causal interaction. If we were to investigate the causal interaction between political orientation and the attacks, we would have to assume the possibility of independently manipulating both variables. Moreover, because the attacks might influence political orientation, we would require an analysis known as "causal mediation with interaction." However, causal mediation carries strong assumptions. It presupposes the inclusion of common causes of the mediator (in our case, political orientation) and the outcome (in our case, warmth towards Muslims) in the analysis. Furthermore, causal mediation analysis assumes that the primary exposure (the attacks) does not affect these common causes (16). Given the complexity of real-world social dynamics, we consider these assumptions too strong to warrant credible inference. Therefore, we limit our analysis to investigating effect modification within preattack levels of political orientation. We aim to uncover how exposure to the attack influenced people across the political spectrum using data about political orientation measured in the NZAVS wave the year before the attacks.

Method

Sample

The NZAVS is a national-scale probability panel study in New Zealand (online Supplementary material S2–S4) that started in 2009. In Time 4 (2012–2013), a Muslim acceptance measure was introduced. Here, we use repeated measures data from the Time 4 cohort to obtain conditionally unbiased counterfactual contrasts for postattack acceptance. Fig. 1A presents a histogram of responses for this cohort from 2013 to 2022. Fig. 1B presents a regression discontinuity over the raw response data, implying a strong boost in acceptance following the attacks. There is no evidence in the raw data for regression to a preattack mean for Muslim acceptance. However, statistical associations in the observed data may lead to biased causal effect estimates. We next consider the general assumptions required for the causal identification of long-term causal effects and explain how this study addresses these assumptions.

Measures

The NZAVS panel survey measures attitudes toward Muslims using a "feeling thermometer" scale. Participants were asked "Please rate your feelings of WARMTH towards the following groups using the 'feelings thermometer scale' for each group." Responses ranged on a 7-point scale from "Feel LEAST WARM Toward This Group = 1 to "Neutral" = 4 to Feel MOST WARM Toward this group = 7. In addition to ratings of Muslims, participants were asked to rate warmth to 11 other groups. Here, we use the terms "warmth" and "acceptance" interchangeably. Below, we also investigate postattack responses to the three stigmatized nonethnic/cultural groups for which the NZAVS had information in the year of the attacks: the Elderly, People with Mental Illness, and the Overweight. Unlike ethnic and cultural groups that may be psychologically represented as overlapping with Muslims (e.g. Arabs and Muslims), people with mental illness, elderly, and overweight serve as negative controls for the causal effects of the attacks on Muslim prejudice. The NZAVS panel survey measures political orientation using a single item: "Please rate how politically liberal versus conservative you see yourself as being. Responses ranged from on a 7-point scale with 1 = *extremely liberal* to 7 = *extremely conservative*" (online Supplementary material S3).

Causal assumptions

Here, our goal is to infer a causal effect of the Christchurch mosque attacks on attitudes, or "warmth," towards Muslims at +1, +2, and +3 years after the attacks. In this case, we can denote warmth ratings as Y, and define two possible conditions of exposure: $A = 1$ if an individual experienced the attack–response–sequence, and $A = 0$ if they did not. We use $Y_i^{a=1}$ and $Y_i^{a=0}$ to represent the potential outcomes for each individual's exposure to the attack and nonattack conditions, respectively.

Researchers can only observe one of the two necessary outcomes to compute individual-level causal effects at any given time. This limitation is common to both observational studies and randomized experiments, often referred to as "The fundamental problem of causal inference," and likened to a missing data problem (31–34). We present the fundamental problem of causal inference as a missing data problem in Table 1 (see online Supplementary material S13).

Although we cannot generally access individual-level causal effects, we can compute average or marginal causal effects for the target population as well as conditional causal effects within certain strata of a population (assuming the fundamental identification assumptions described below hold). These are the contrasts between the condition where everyone in a population (or stratum) is exposed and the condition where everyone in a population (or stratum) is unexposed (36). We express the marginal contrast as: $E(Y^{a=1}) - E(Y^{a=0})$. We express the conditional contrast within population strata V as: $E(Y^{a=1} \mid V = v) - E(Y^{a=0} \mid V = v)$ (online Supplementary material S14). Note that estimating causal contrasts requires something more than data science; it requires what might be called *counterfactual data science*. This is because, unlike traditional data science which relies primarily on observed data, counterfactual data science involves the estimation of potential outcomes under different exposure scenarios, which may not be directly observed in the data.

We emphasize the connection between causal estimation and missing data for two reasons. First, although there are traditions for causal estimation in some social sciences (35, 37) and epidemiology (38), many areas of social sciences—including psychological science—do not regularly use these methods. Thus, although in observational settings the counterfactual contrast required for causal inference, $E[Y^{a=1}] - E[Y^{a=0}]$, is not generally equivalent to the observed contrast $E[Y \mid A = 1] - E[Y \mid A = 0]$, many psychological scientists operate under the assumption of an equivalence (16, 39–41). Yet to identify causal effects one must appreciate the need for a systematic counterfactual data science. Any assumed

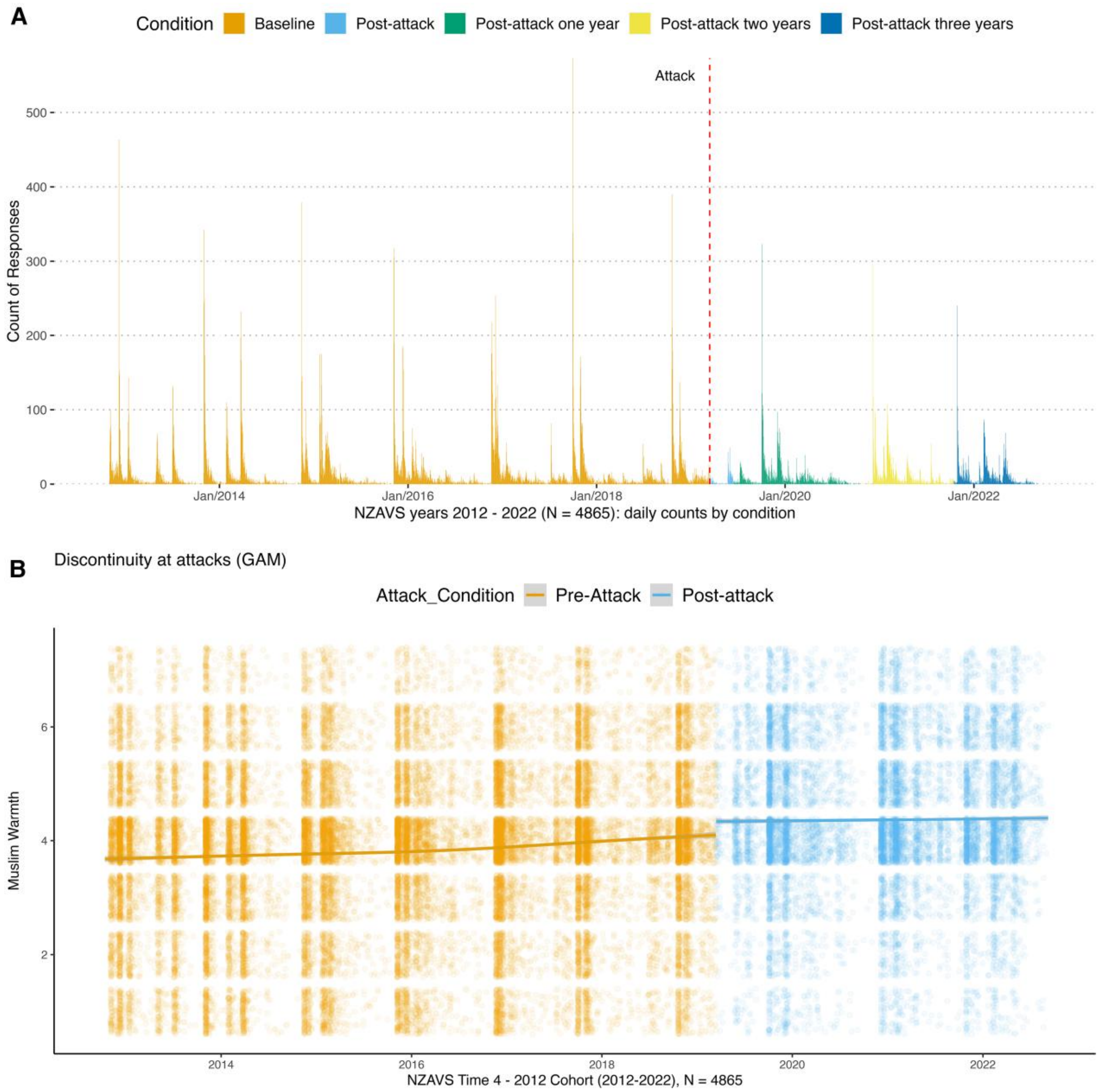

Fig. 1. A) Histogram of rolling responses from the NZAVS. Responses in sequential order by periods of interest: (i) the preattack baseline (before 15 March 2019); (ii) the postattack interval in NZAVS Time 10 (2019 March 16 and the following three months); (iii) the + one-year post-attack period; (iv) the + two-years post-attack period; and (iii) the + three-years-post-attack period. B) Regression discontinuity analysis reveals a sharp increase in average acceptance of Muslims immediately after the attack, with no evidence for regression in the observed sample to the preattack acceptance average.

equivalence between observed and counterfactual contrasts can lead to misleading conclusions about causal relationships. Second, understanding that the fundamental problem of causal inference is a missing data problem clarifies the motivation for predicting missing responses for everyone in a target population under each level of a causal contrast, in this case, the +1, +2, +3 year acceptance of Muslims in the attack and nonattack exposures (36, 42, 43).

Table 1. Observed outcomes and inherently missing counterfactual outcomes presented by condition.

Condition	$Y_i^{a=1}$	$Y_i^{a=0}$
Exposure ($A_i = 1$)	Observed Y_i	Counterfactual Y_i
No exposure ($A_i = 0$)	Counterfactual Y_i	Observed Y_i

At least half the data needed for inferring causal effects are missing. It is only with assumptions that researchers may identify average causal effects from individual-level observations (see online Supplementary material S13). This table is adapted from Morgan and Winship (35).

We next explain our method for identifying the counterfactual contrasts of interest to this study. Because psychological scientists might not be familiar with the identification assumptions required for causal inference, we clarify our specific approach by referring to how our study addresses these identification assumptions.

The causal consistency assumption

The values of exposure under comparisons correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data. (43, 44)

It is the satisfaction of the causal consistency condition that allows researchers to recover counterfactual outcomes from observed data. When all conditions for causal estimation are satisfied, we may say that, by the causal consistency condition, an individual's observed outcome under the exposure that they received is equal to their counterfactual outcome under that exposure.

This causal consistency assumption can be understood as follows (35):

$$Y^{obs} = AY^{a=1} + (1 - A)Y^{a=0} \tag{1}$$

For individuals with exposure level $A = 1$, if the causal consistency condition is satisfied, we may infer that:

$$\begin{aligned}(Y^{obs} \mid A = 1) &= 1 \times A \times Y^{a=1} + (1 - 1) \times Y^{a=0} \\ &= 1 \times Y^{a=1} + 0 \times Y^{a=0} \\ &= Y^{a=1}\end{aligned} \tag{2}$$

Similarly, for individuals with exposure level $A = 0$, if the causal consistency condition is satisfied, we may infer that:

$$\begin{aligned}(Y^{obs} \mid A = 0) &= 0 \times A \times Y^{a=1} + (1 - 0) \times Y^{a=0} \\ &= 0 \times Y^{a=1} + 1 \times Y^{a=0} \\ &= Y^{a=0}\end{aligned} \tag{3}$$

Which implies:

$$\begin{aligned} Y_i &= Y_i^{a=1} \quad \text{if } A_i = 1 \\ Y_i &= Y_i^{a=0} \quad \text{if } A_i = 0 \end{aligned} \tag{4}$$

The causal consistency assumption can be difficult to understand. Data scientists estimate parameters for observed data. In causal inference, we must go one step further, and estimate contrasts involving counterfactual parameters. First, we must compute the average response when the entire target population is exposed. Then we must compute the average response when the entire target population is unexposed. Next, we must contrast these predicted averages on some scale, here the difference scale. Then, we must assess uncertainty in these contrasts by applying the delta method, simulation, working with Bayesian posterior distributions, or bu apply another method. Notice that the steps for causal inference are not merely more involved than for correlational analysis. Causal inference requires linking counterfactual outcomes to data. It is the causal consistency assumption that allows us to do that.

How our study satisfies the causal consistency assumption

Our study follows the theory of causal inference with multiple treatment variations, which requires "treatment-variation-irrelevance" (18, 45). Here, the term "attack" refers to the sequence of responses to the Christchurch mosque attacks. We contemplate several versions of this attack sequence, each associated with a different outcome. We assume no confounding for the relationship between these versions and outcomes given the measured variables in our model.

We assume all individuals experienced a version of the postattack exposure from the first to third postattack waves. Because everyone in the sample from which the target population is drawn experienced this exposure, the causal consistency assumption is weaker for the exposure condition. However, the causal consistency assumption for the no-attack sequence depends on the stronger assumption that we can extrapolate the contrast condition from each individual's preattack responses.

We test this stronger assumption in three ways. First, we simulate responses in the pre-exposure timeline and compare them with observed responses. Results provide a close match between the observed and simulated data (online Supplementary material S8).

Second, we investigate whether the attacks influenced attitudes towards unrelated stigmatized groups. If attitudes towards these groups also changed, other societal changes might explain observed acceptance among Muslims. However, our results do not suggest this (see Results section).

Third, we examine the potential impact of the global pandemic on Muslim acceptance. We compare responses pre and post-COVID-19 in the postattack condition and find no significant change. Although we cannot examine whether the pandemic affected attitudes toward Muslims among those who did not experience the attacks, we can evaluate differences in the postattack condition by comparing pre- and post-COVID-19 responses. We find no evidence for differences in the exposed sample (see Results section).

Of course, we cannot know whether people would have responded differently to the COVID-19 pandemic had they not also experienced the Christchurch mosque attacks. In the Discussion section, we consider how our inferences might change if our method for recovering the postattack counterfactual contrast trajectory were wrong.

The exchangeability assumption

The conditional probability of receiving every value of an exposure level, though not decided by the investigators, depends only on the measured covariates. (36, 43)

When the conditional exchangeability condition holds, counterfactual outcomes are independent of actual exposures received:

$$Y^{a=1}, Y^{a=0} \perp\!\!\!\perp A \mid L \tag{5}$$

or equivalently

$$A \perp\!\!\!\perp Y^{a=1}, Y^{a=0} \mid L \tag{6}$$

$Y^a \perp\!\!\!\perp A \mid L$ (or equivalently, $A \perp\!\!\!\perp Y^a \mid L$), where L is the set of covariates sufficient to ensure the independence of the counterfactual outcomes and the exposure.

How our study satisfies the exchangeability assumption

The attacks were random concerning NZAVS data collection. However, because participants might selectively not respond to Muslim warmth questions, there is potential for selection bias at every wave, including the exposure wave (NZAVS Time 10). Moreover, predicting any missing responses requires satisfaction of the missing at random (MAR) assumptions (online Supplementary material S13). If those who did not respond were to differ from those who responded, causal effect estimation would be biased. Thus, in addition to missing individual-level potential outcomes, we encounter potential selection bias from nonresponse and panel attrition (46). Fig. 2A and B and Table 2 clarify the structure of our missing data challenges by exposure condition. $\boxed{S}$ denotes a selection on observed responses. We assume the exposure condition may affect the observed responses (i.e. $A \to \boxed{S}$). We omit the arrows between the exposure and the outcomes to clarify the threat of bias. Conditioning on $\boxed{S}$ may induce a collider bias by producing a statistical association between A and Y without any causal association. As mentioned, the one-, two-, and three-year postattack no-exposure responses are entirely missing. This is indicated in Fig. 2B. Identifying the causal effects of the attacks on Muslim acceptance during the three

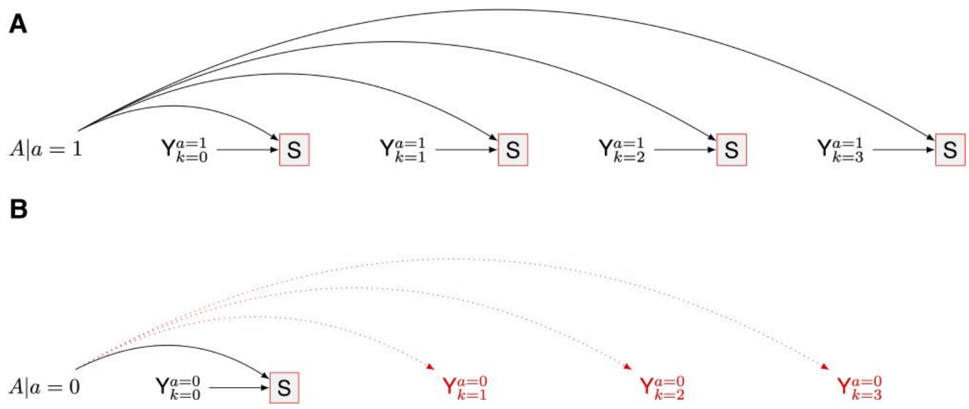

Fig. 2. A) and B) Two causal diagrams, one for each exposure. These graphs clarify the different missing data challenges for each exposure condition. If those prone to panel attrition differ in their expected warmth ratings of Muslims, then there is a differential loss to follow-up or selection bias. By conditioning on $\boxed{S}$—the observed responses—there is a risk of opening a path between the exposure condition A and the outcome Y (Muslim prejudice). Any common cause of nonresponse and Muslim warmth that is differential concerning the timing of exposure may induce such a bias (e.g. the panel study retains more accepting people). However, to simplify the graph, we denote the threat to inference from selection bias by drawing an arrow from Y to $\boxed{S}$ and omitting arrows from A to Y. A) This is a causal diagram for the attack condition ($A = 1$). It focuses on the problem of possible selection bias in the attack condition (NZAVS Time 10) and at the following three waves. B) This is a causal diagram for the no-attack condition ($A = 0$). It shows that selection bias is possible in the no-attack condition (NZAVS Time 10). However, responses in the no-attack condition for the following three years (Time 11, Time 12, Time 13) are fully missing. These fully missing responses are indicated by $Y^{a=0}_{k>0}$. (Note: these causal diagrams are not strictly causal DAGS because they do not present a Markov Factorization of exposure and outcome. Instead, as each graph presents a separate counterfactual exposure and counterfactual response: they are Single World Intervention Graphs (SWIGs). For more information about SWIGs, see Richardson and Robins (47).

years following the attacks is only possible if we may credibly impute the counterfactual arc of Muslim acceptance had the attacks never occurred. Again, we recover the missing acceptance trajectories by imputing the counterfactual outcomes using seven waves of panel data from the NZAVS 2012 cohort (33, 34).

The positivity assumption

The probability of receiving every value of the exposure within all strata of covariates is greater than zero. (44, 48)

$$0 < \Pr(A = a \mid L) < 1, \quad \forall a \in A, \ \forall a \in L \tag{7}$$

There are two types of violations to the positivity assumption. The first type of violation is called random nonpositivity. It occurs when the sample data do not contain all levels of exposure within the strata for whom exposures are defined. For example, if participants aged 22–24 years old were missing from the data, we would need to use information from respondents at ages other than 22–24 years to model continuous age effects. In our study, there are no observations for the no-exposure condition in postattack waves $k > 0$. Thus, our study is subject to random nonpositivity:

$$\Pr(A = 0 \mid K = k > 0) = 0 \tag{8}$$

The second type of violation of the positivity condition is called deterministic nonpositivity. This occurs when it is scientifically impossible for certain strata to receive specific levels of exposure. For example, biological males cannot receive hysterectomy. Violations of deterministic nonpositivity imply restricting the analysis to scientifically plausible cases.

In cases where there is random nonpositivity, but not deterministic nonpositivity, researchers typically rely on modeling assumptions to estimate causal effects (36, 43). For this reason, satisfying random nonpositivity using models carries additional modeling assumptions.

How our study satisfies the positivity assumption

Here, we assume what nearly everyone wishes, that the terror attacks might not have happened. That is, we assume the occurrence of the attacks was not deterministically fated. This assumption is plausible. However, as mentioned above, we must use parametric models to recover the +1, +2, and +3 counterfactual contrasts for the no-attack conditions, and the imputation models we use to simulate these counterfactual outcomes carry extra assumptions. Although we test these extra assumptions, like many assumptions in causal inference, they cannot be verified. For this reason, our Discussion section considers how the inference from this study might change if the assumption or methods we use for counterfactual recovery were wrong.

Table 2. Missing potential or counterfactual outcomes.

	Time 10		Time 11		Time 12		Time 13	
	0	1	0	1	0	1	0	1
	(N = 4,865)	(N = 4,865)	(N = 4,836)	(N = 4,836)	(N = 4,795)	(N = 4,795)	(N = 4,777)	(N = 4,777)
Y_Warm.Muslims								
Mean (SD)	4.06 (1.45)	4.38 (1.36)	NA (NA)	4.33 (1.42)	NA (NA)	4.38 (1.38)	NA (NA)	4.39 (1.36)
Median [Min, Max]	4.00 [1.00, 7.00]	4.00 [1.00, 7.00]	NA [NA, NA]	4.00 [1.00, 7.00]	NA [NA, NA]	4.00 [1.00, 7.00]	NA [NA, NA]	4.00 [1.00, 7.00]
Missing	1555 (32.0%)	3840 (78.9%)	4836 (100%)	902 (18.7%)	4795 (100%)	957 (20.0%)	4777 (100%)	1210 (25.3%)

Causal estimation requires a method for assigning values to all missing observations. Many social and behavioral scientists are trained to apply these strategies without explicitly understanding missing observations' role in causal inference. Note that declining wave totals over time arise from known mortality.

Table 3. Key for NZAVS panel.

NZAVS	Years	Study label	Observed exposure	Notes
Time 4	2012–2013	Preattack wave −6	No-attack only	Discarded after imputation
Time 5	2013–2014	Preattack wave −5	No-attack only	Discarded after imputation
Time 6	2014–2015	P-attack wave −4	No-attack only	Discarded after imputation
Time 7	2015–2016	Preattack wave −3	No-attack only	Discarded after imputation
Time 8	2016–2017	P-attack wave −2	No-attack only	Discarded after imputation
Time 9	2017–2018	P-attack wave −1	No-attack only	Baseline: indicators for imputation incl. baseline political orientation
Time 10	2018–2019	Attack wave 0	Both no-attack & postattack	Imputed missingness for both conditions
Time 11	2019–2020	P-attack wave +1	Both	Imputed missingness for both conditions, fully for $A = 0$
Time 12	2020–2021	Postattack wave +2	Postattack only	Imputed missingness for both conditions, fully for $A = 0$
Time 13	2021–2022	Postattack wave +3	Postattack only	Imputed missingness for both conditions, fully for $A = 0$

Need for clearly defined causal estimands

In addition to satisfying the three fundamental conditions for causal inference, causal inference requires clearly defined causal contrasts (estimands).

A causal estimand names the causal effect whose magnitude, δ, we seek to compute, here, on the difference scale:

Where $E[\delta_{k,v}] \neq 0$ indicates a causal effect of $A = a$ at time $K = k$ within stratum of political orientation $V = v$, we assess:

$$E[\delta_{kv}] = E[Y^{a=1} \mid K = k, V = v] - E[Y^{a=0} \mid K = k, V = v] \tag{9}$$

Where $E[\delta_{k,v}] \neq 0$ indicates a conditional causal effect for the exposure at time $K = k$ within political stratum $V = v$.

Here, we compute causal contrasts for the attack vs. no-attack exposure at four-time points: the attack wave 0 through postattack wave +3 (NZAVS Time 10–Time 13, years 2018–2022). We compute these causal contrasts within five levels of preattack political orientation:

- extreme liberal = − 1.92 SD political orientation (note that −1.92 SD is at the farthest point on the liberal end of the data range)
- liberal = − 1 SD political orientation
- moderate = 0 SD political orientation
- conservative = + 2 SD political orientation
- extreme conservatives = + 2 SD political orientation.

We use graphical as well as tabular methods to present these causal contrasts.

Need for clearly defined target population for inference

To which population do our results generalize? Our study generalizes to the population from which the sample was drawn. The inclusion criteria for this study are:

1. participated in the NZAVS Times 4–9 (Time 9 is the year before exposure: 2017–2018);
2. may have been lost to follow-up after Time 9 (i.e. allow that the attack exposure may have caused loss-to-follow-up);
3. full information at baseline for the following indicators: age, gender (male, another gender), born in NZ (coded yes/no), ethnicity, employed (yes/no), occupation, personality (agreeableness, conscientiousness, openness, honesty-humility, extraversion, neuroticism), parent (yes/no), residential address (for regional, regional deprivation, and rural status), racial rejection anxiety, religious identification, and political orientation (see online Supplementary material S3).

Table 3 describes the structure of our data. We investigate dynamics in the oldest possible cohort for Muslim-acceptance data (Wave 4) because it has the longest history of within-person preattack responses from which to estimate postattack counterfactually missing responses. Additionally, we present an analysis of the Time 5 NZAVS (closed) cohort because this cohort is larger. The Time 5 cohort results replicate the Time 4 cohort results (online Supplementary material S9).

It is important to keep in mind that our pre-exposure exclusion criteria limit generalizations. Our results generalize to a probability sample of the New Zealand population that was sampled in 2012 and remained in the NZAVS through its 10th wave (years 2012–2019). Although our methods may consistently estimate causal effects without bias for this population, this population does not accurately reflect the full diversity of the New Zealand population.

Modeling approach

This study applies a variant of G-computation as described by Westreich et al. (34) for imputing and simulating potential responses for all study participants across various exposure conditions.

Recall that conditional exchangeability necessitates the independence of potential results from the exposure condition given a set of measured confounders, expressed as $Y^{a=1}, Y^{a=0} \perp\!\!\!\perp A \mid L$. If the counterfactual consistency assumption is also satisfied, then we may say: $E[Y^a \mid L] = E[Y \mid A, L]$, and

$$E[Y^{a=1} - Y^{a=0}] = \sum_l E[Y \mid A = 1, L = l]\Pr[L = l] - \sum_l E[Y \mid A = 0, L = l]\Pr[L = l] \tag{10}$$

By calculating the expectation of the difference between predicted responses for the population under the two exposure conditions (weighted by the distributions of baseline covariates), we can derive a weighted mean for the confounder distribution (online Supplementary material S14). Assuming the set of confounders L adequately controls for confounding, we achieve an unbiased causal estimate by computing:

$$E[Y^{a=1} - Y^{a=0}] = E[E[Y \mid A = 1, \mathbf{L}] - E[Y \mid A = 0, \mathbf{L}]] \tag{11}$$

Typically, G-computation realizes causal estimation through four steps:

1. Develop a regression model for the outcome: $E[Y \mid a, l] = \beta_0 + \beta_1 A + \beta_2 L'$.[b]
2. Replicate the dataset, removing everyone's observed outcome, and predict everyone's response when A is set to 1. Keep the predicted outcomes: $\hat{E}[Y^{a=1}] = \hat{E}[Y \mid A = 1, L = l]$.

3. Replicate the dataset, removing everyone's observed outcome, and predict everyone's response when A is set to 0. Keep the predicted outcomes: $\hat{E}[Y^{a=0}] = \hat{E}[Y \mid A = 0, L = l]$.
4. Compute the difference in standardized means between the two sets of imputed outcomes: $\hat{E}[Y^{a=1} - Y^{a=0}] = \hat{E}[\hat{E}[Y \mid A = 1, L = l] - \hat{E}[Y \mid A = 0, L = l]]$.[c]

Westreich et al. (34) outline an equivalent mathematical approach where outcomes are predicted through multiple imputations instead of regression. The authors also explore potential combinations of G-computation and imputation. In this study, we adopt a hybrid approach as follows:

1. For the condition $A = 0$, impute missing observations by fitting a Bayesian multilevel regression model for $E[Y^{\text{obs,missing}} \mid a = 0] = \gamma_0 + \boldsymbol{\gamma_0} Z'$ (49–51). This model incorporates an interaction term for time and all indicators used to predict missing responses. We presume all missing responses to be MAR, conditional on the observed outcomes and a covariate set Z. This set is deemed sufficient for imputing missing response values: $Y^{\text{missing}} \perp\!\!\!\perp R \mid Y^{\text{obs}}, A, Z$. This covariate set includes indicators for baseline political orientation V: $V \in Z$.[d]
2. Utilize this model to impute outcomes for everyone under the condition $A = 0$ for NZAVS Times 10–13 (attack wave 0 to attack wave +3). All Times 4–9 observations were then disregarded except for the baseline indicators of political orientation. Given $Y \mid A = 0 \perp\!\!\!\perp Z, A$ and following counterfactual consistency, $\hat{E}[Y \mid (A = 0, Z = z)] = \hat{E}[Y^{a=0}]$.
3. Duplicate the original dataset (clone the cohort) and repeat steps 1–2 for the condition $A = 1$, utilizing observations for condition $A = 1$ NZAVS Time 10–Time 13 (attack wave = 0—postattack wave +3). Employ the same baseline predictors used in the imputation model for the $A = 0$ condition. Again, the imputed outcomes are estimated for everyone in this condition. Given $Y \mid A = 1 \perp\!\!\!\perp Z, A$ and by counterfactual consistency, $\hat{E}[Y \mid (A = 1, Z = z)] = \hat{E}[Y^{a=1}]$.
4. Remember, our goal is to identify the conditional causal effects within the levels of K and V. According to the law of total probability and the partition theorem, $\hat{E}[\hat{E}[Y \mid (A, V, K) \mid A = a] = \hat{E}[Y \mid A = a]$ (online Supplementary material S14). To extract conditional causal contrasts, merge the imputed results by each condition and apply a marginal structural model to the unified data. This model computes the desired conditional causal effects $\hat{E}[Y \mid (A, K, V)] = \beta_0 + \beta_1 A + \beta_2 K + \beta_3 V + \boldsymbol{\beta}_4 S'$. In this instance, $\boldsymbol{\beta}_4'$ signifies the matrix of coefficients for the interplays of attack circumstance × wave × political inclination. Standard errors for the causal contrasts are obtained using a sandwich estimator from the `marginaleffects` package in R (55), or by operating directly with predictive posterior distributions (online Supplementary material S7).

We obtained covariates for imputation for all participants from responses at baseline (NZAVS Time 9, preattack wave −1). The covariates set **Z** consists of age, education, ethnicity, has-partner, identity-male, is-employed, is-parent, urban residence, region of habitation, neighborhood deprivation, occupational prestige, racial rejection anxiety, political orientation, religious identification, and the big six personality traits (agreeableness, conscientiousness, extraversion, honesty-humility, openness, and neuroticism).

We included these demographic variables in the imputation model because previous research finds they predict loss-to-follow-up (56) and might be related to prejudice. We include personality measures because personality traits have been shown to correlate with various psychological outcomes (41). Our models incorporated indicators for objective job status (NZSEI) and geographical region (30). In our model, we assume that the indicators, as modeled, are jointly sufficient to impute missing responses without bias (MAR). Although this assumption is unverifiable, it is arguably plausible given previous studies on the causes of Muslim prejudice in New Zealand (4, 5, 57) and research on NZAVS panel attrition (56).

To handle missing values, we extracted posterior distributions from the model for each individual and averaged them using the BRMS's "predict" function. This approach follows the third strategy that the package's author suggests for imputing missing responses (58). We then merged the two separately imputed datasets into a single dataset (online Supplementary material S12). The same analysis was performed on the larger NZAVS Time 5 Cohort, as reported in online Supplementary material S7.

We emphasize that missing values must be imputed separately for each exposure condition because the counterfactuals outcomes are not jointly observed (34). Note, this is true for any causal inferential study. The demands on multiple imputations in counterfactual data science differ from the demands in other settings. Methods for appropriately imputing missing responses when determining causal effect estimates are the focus of active inquiry (59). For example, Zhang et al. have demonstrated that when both the imputation model and the analysis model are incorrectly specified, inferences will be biased. Their simulations also suggest that imputing missing responses separately within each exposure group provides the most accurate results across all scenarios (59). One may argue that within the separate exposure groups, restricting imputations to pretreatment variables, as we have done here, could limit the precision of the imputations, leading to increased uncertainty. However, there are instances where outcomes can influence the missingness of specific covariates, which, in turn, affect the missingness in outcomes, thus giving rise to potential bias. This can occur when post-treatment variables influenced by the outcome are included in the imputation model, leading to "collider-stratification bias." In situations where an outcome affects a post-treatment covariate that, in turn, affects the availability of outcome data, conditioning on a post-treatment covariate during imputation, stratification, or regression adjustment creates a backdoor path that can bias causal estimates. Hence, although including more information in the imputation model can sometimes increase precision (i.e. reduce variance), it also has the potential to introduce bias when the mechanisms of missingness are complex and interrelated with the outcomes of interest.[e] Given that approximately 80% of the NZAVS sample is retained each year (20% is lost to follow-up), we opted for a conservative approach, restricting our imputation model to condition solely on pre-exposure indicators.[f]

Statistical model

We estimated the main effects and interactions of Attack × Wave × Political Orientation on Muslim Acceptance. The statistical model is described in (online Supplementary material S5). Multilevel models may be biased where group-level indicators are correlated with exposure-level errors (62). Because the attacks were random with respect to NZAVS data collection, such bias is unlikely for multilevel models. The main article reports the results for generalized estimating equations with an autoregressive error term. The online Supplementary material also reports the Bayesian multilevel model with a random intercept term (online Supplementary material S6). We find results do not depend on this statistical modeling choice. We also replicate findings in a larger, more recent NZAVS cohort (NZAVS time 5, N = 7,824), again with no difference to inference (online Supplementary material S7). We obtain contrasts, contrasts plots, and prediction plots using the marginaleffects (55) and ggeffects (63) packages in R (49). We are grateful to all software developers whose packages we used (online Supplementary material S15).

Results

Model results

We first investigated the marginal effect of the attack on warmth to Muslims in the years following the Christchurch mosque attacks. Table 4 and Fig. 3 present the statistical model results. As indicated in Fig. 4A, the attacks strongly boosted warmth to Muslims. Moreover, in the years following the attacks, there is no indication of regression to preattack means levels of warmth to Muslims. Indeed, the lower bound of the confidence interval of the expected average warmth rating three years after the attacks is higher than the upper bound of the expected average warmth rating immediately following the attacks. However, as indicated in Fig. 3B, the relative magnitude of the increase in the average attack acceptance trended downward during the following three years.

Causal effect estimates

We next compute the conditional causal effects of the attack over time as modified by baseline political orientation. We obtain contrasts at five levels of political orientation: extreme liberal = − 1.92 SD political orientation (the end of the data range); liberal = − 1 SD political orientation; moderate =0 SD political orientation; conservative = + 1 SD political orientation; extreme conservatives = + 2 SD political orientation. These contrasts are presented in Table 5.

As evident in Fig. 4A, after the initial boost in acceptance following the attacks, there is little evidence for further growth in Muslim acceptance among extreme liberals and liberals. By contrast, among moderates (postattack wave +3), conservatives (all postattack waves), and extreme conservatives (all postattack waves), there is evidence for further growth over and above the initial boost to Muslim acceptance. As indicated in Fig. 4B, a clear separation in growth patterns are evident in the counterfactual contrasts at different levels of political orientation. Among extreme conservatives, the postattack contrast with the no-attack condition in the year of the attacks is $\delta_{k=0,\upsilon=2} = 0.278$ [0.172, 0.383]. At postattack wave +3, this contrast is not much changed $\delta_{k=3,\upsilon=2} = 0.262$ [0.155, 0.369]. Turning to extreme liberals, the postattack causal contrast is $\delta_{k=0,\upsilon=-1.92} = 0.290$ [0.198, 0.383] for the year of the attacks. At postattack wave +3, the interval for the causal contrast of attack and no-attack conditions crosses zero $\delta_{k=3,\upsilon=-1.92} = 0.066$ [− 0.027, 0.160]. Among extreme liberals, we infer that after three years, the expected difference between the attack exposure condition and the attack no-exposure condition was no longer reliable.

Table 4. Coefficients from the generalized estimating equation model.

Parameter	Coefficient	SE	CI_low	CI_high
(Intercept)	4.034	0.016	4.001	4.066
Attack1	0.284	0.022	0.241	0.327
Wave1	0.064	0.002	0.060	0.067
Wave2	0.126	0.003	0.121	0.132
Wave3	0.190	0.003	0.183	0.197
Pol.Orient_cZ	−0.336	0.017	−0.370	−0.302
Attack1:Wave1	−0.041	0.003	−0.046	−0.036
Attack1:Wave2	−0.080	0.004	−0.087	−0.073
Attack1:Wave3	−0.122	0.005	−0.131	−0.113
Attack1:Pol.Orient_cZ	−0.003	0.023	−0.049	0.042
Wave1:Pol.Orient_cZ	0.003	0.002	−0.001	0.007
Wave2:Pol.Orient_cZ	0.009	0.003	0.003	0.014
Wave3:Pol.Orient_cZ	0.015	0.004	0.008	0.022
Attack1:Wave1:Pol.Orient_cZ	0.019	0.003	0.014	0.024
Attack1:Wave2:Pol.Orient_cZ	0.035	0.004	0.028	0.042
Attack1:Wave3:Pol.Orient_cZ	0.053	0.005	0.044	0.062

Note that parameters with suffix _cZ are centered and z-score transformed.

Negative control validation

As presented in Table 6, we do not find any reliable differences in warmth ratings preattacks/postattacks for the three stigmatized negative control groups (i.e. elderly, mentally ill, and overweight; models adjust for preattack responses for each outcome). By contrast, we observe a large effect of the attacks on Muslim warmth ratings. We replicate these findings in the total NZAVS Time 10 sample (online Supplementary material 13).

Prepandemic/postpandemic Muslim warmth validation—NZAVS Time 11 (2019–2020)

We next investigate the effects of the COVID-19 pandemic on attitudes to Muslims. We do not find evidence for such effects in the Time 4 cohort; see Table 7. We replicate these findings in the total NZAVS Time 11 sample; see Table 8. Again, we do not observe the effects of the COVID-19 pandemic in the no-exposure condition. However, had we observed the pandemic hindered Muslim acceptance in the attack-exposure group, we would have inferred that the counterfactual imputation model might be too pessimistic about the growth of Muslim acceptance in the observed postattack condition.

Nor do we find evidence for such effects in the full NZAVS sample (Table 8).

Discussion

This study addresses a fundamental question in the social and political sciences: what are the long-term psychological effects of terrorism? We combine a counterfactual approach with 10 years of national-scale panel data to investigate the psychological effects of the 2019 Christchurch New Zealand terrorist attacks on the acceptance of the targeted Muslim community during the following three years.

Immediately following the attacks, Muslims experienced about five years of instantaneous growth in acceptance above what they would have experienced had the attacks not occurred. These effects were durable: there was no evidence for regression to a preattack level of Muslim acceptance. A previous regression discontinuity analysis suggested declining Muslim acceptance within 90 days after the March 2019 mosque attacks among political conservatives (20). Regression to preattack acceptance levels of average Muslim acceptance would have been consistent with transitory effects observed for satisfaction with the New Zealand government (24). However, we find that, even as the growth rate in acceptance slowed during the following three years, an absolute boost to Muslim acceptance was retained.

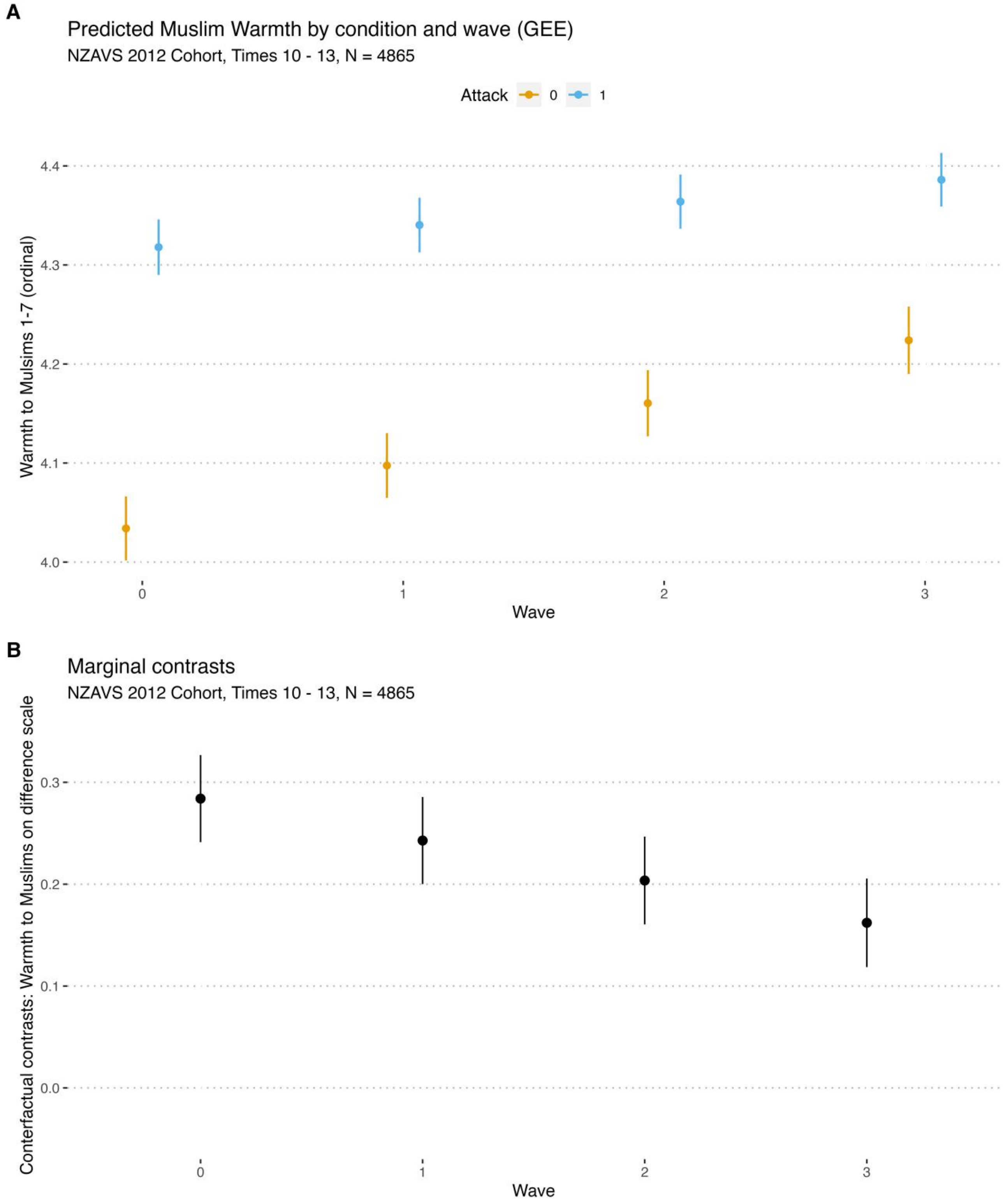

Fig. 3. A) Expected Muslim warmth across attack conditions and waves. The response range is 1–7. This graph focuses on the region reflecting the marginal (unconditional) response across the population over time. B) Marginal contrast for expected Muslim warmth between the attack and no-attack conditions by wave. We find an initial boost to marginal acceptance, which gradually declined.

Moreover, among political conservatives, the acceptance of Muslims continued to grow following the attacks, over and above the initial postattack boost to acceptance. Among political liberals, acceptance did not regress to preattack levels. Furthermore, among political conservatives, acceptance of Muslims steadily increased beyond the initial postattack surge. Among political liberals, while acceptance levels did not recede to preattack norms, they also did not exhibit further growth.

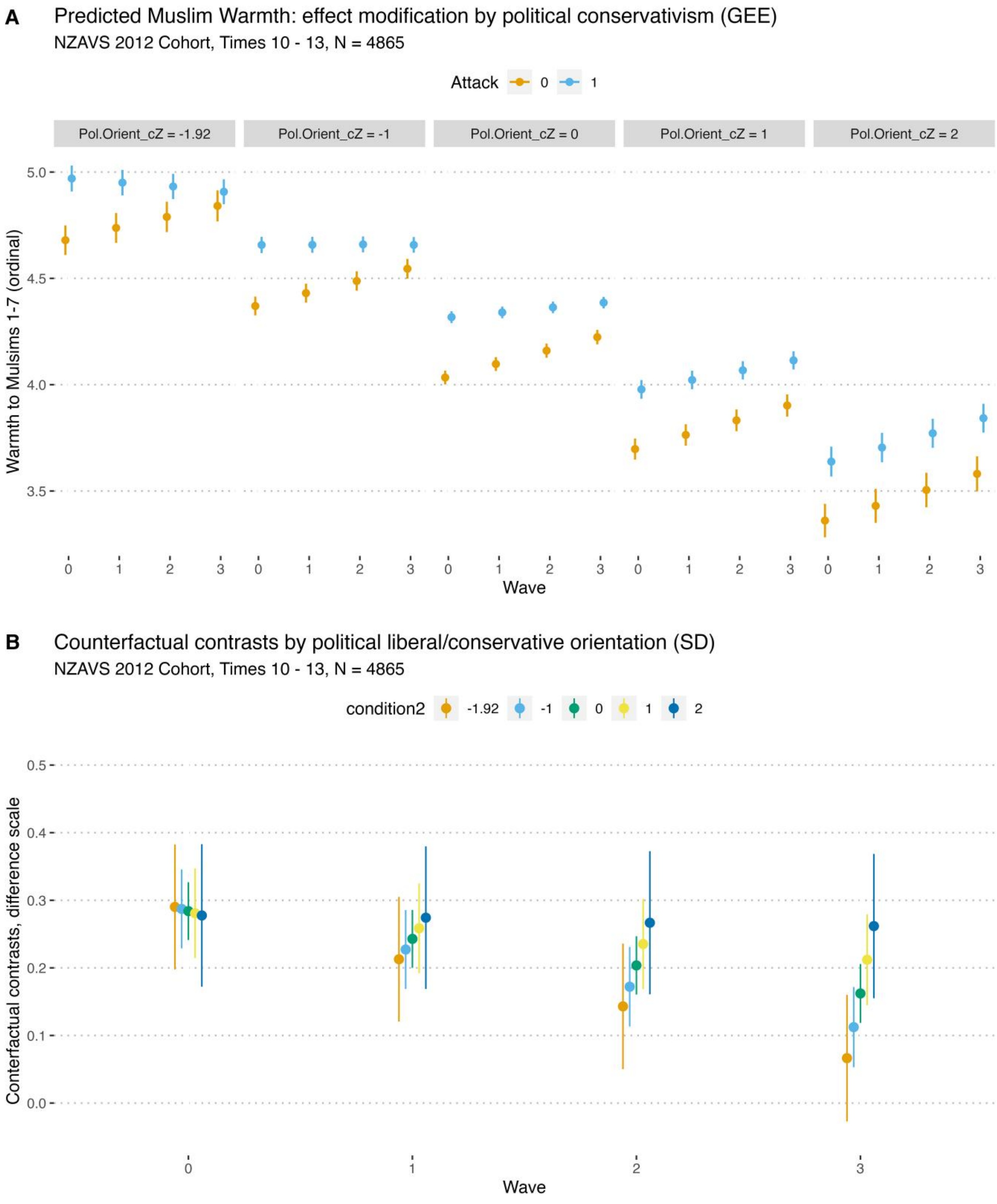

Fig. 4. A) Heterogeneity in expected Muslim warmth over time by preattack political orientation. The response range is 1–7. This graph focuses on the region reflecting the predicted average response. This region, stratified by levels of political orientation, extends from politically conservative to politically liberal responses. The graph highlights inferred differences over time in conditional average responses across these strata of political orientations. In absolute terms, acceptance of Muslims remained higher among political liberals throughout all three waves. B) Causal contrasts over time in Muslim warmth by preattack political orientation. The Y-axis illustrates conditional contrasts between the attack and no-attack conditions on Muslim Warmth by political orientation strata. The effect estimates and confidence intervals specific to strata of political orientation are presented using different colors (navy blue for extreme conservatives and orange for extreme liberals). The X-axis denotes the NZAVS postattack waves, showing the temporal progression of these conditional contrasts. We observe a consistent postattack boost in Muslim acceptance among political conservatives (navy blue) across all waves. However, this boost appears to wane among political liberals (orange). In relative terms, the long-run effect of the attacks on Muslim warmth was stronger in political conservatives.

Table 5. Estimands for conditional causal contrasts.

Estimand	Contrast	Comparison	Std. error	Conf. low	Conf. high	Attack	Wave	Pol_Orient_SD
$\delta_{k=0,v=-1.92}$	1–0	0.290	0.047	0.198	0.383	0	0	−1.92
$\delta_{k=0,v=-1}$	1–0	0.287	0.030	0.229	0.346	0	0	−1
$\delta_{k=0,v=0}$	1–0	0.284	0.022	0.241	0.327	0	0	0
$\delta_{k=0,v=+1}$	1–0	0.281	0.034	0.214	0.347	0	0	1
$\delta_{k=0,v=+2}$	1–0	0.278	0.054	0.172	0.383	0	0	2
$\delta_{k=1,v=-1.92}$	1–0	0.213	0.047	0.121	0.305	0	1	−1.92
$\delta_{k=1,v=-1}$	1–0	0.227	0.030	0.169	0.285	0	1	−1
$\delta_{k=1,v=0}$	1–0	0.243	0.022	0.200	0.286	0	1	0
$\delta_{k=1,v=+1}$	1–0	0.259	0.034	0.192	0.325	0	1	1
$\delta_{k=1,v=+2}$	1–0	0.274	0.054	0.169	0.380	0	1	2
$\delta_{k=2,v=-1.92}$	1–0	0.143	0.047	0.050	0.236	0	2	−1.92
$\delta_{k=2,v=-1}$	1–0	0.172	0.030	0.113	0.231	0	2	−1
$\delta_{k=2,v=0}$	1–0	0.204	0.022	0.161	0.247	0	2	0
$\delta_{k=2,v=+1}$	1–0	0.235	0.034	0.169	0.302	0	2	1
$\delta_{k=2,v=+2}$	1–0	0.267	0.054	0.161	0.373	0	2	2
$\delta_{k=3,v=-1.92}$	1–0	0.066	0.048	−0.027	0.160	0	3	−1.92
$\delta_{k=3,v=-1}$	1–0	0.112	0.030	0.053	0.172	0	3	−1
$\delta_{k=3,v=0}$	1–0	0.162	0.022	0.119	0.206	0	3	0
$\delta_{k=3,v=+1}$	1–0	0.212	0.034	0.145	0.279	0	3	1
$\delta_{k=3,v=+2}$	1–0	0.262	0.054	0.155	0.369	0	3	2

Open questions

The explanation for why acceptance effects systematically varied by political orientation is speculative. In absolute terms, liberals exhibited significantly higher acceptance of Muslims at the preattack baseline. Political conservatives, on the other hand, had more room for an upward shift in acceptance. Therefore, the observed differences by political orientation might stem from the initial disparities in acceptance levels across the political spectrum. Another possibility is that political conservatives became more concerned about social desirability in their expression of Muslim attitudes following the attack due to changing norms about the acceptance of Muslims in New Zealand following such a violent and immoral incident. In this scenario, the attacks could have induced directed measurement bias in estimating prejudice in the postattack condition. However, for this explanation to hold weight, such bias would need to have persisted over several years. Perhaps the simplest explanation is that the terrorist's violent actions prompted those on the right to distance themselves from actions they took to be immoral. In support of this conjecture, psychological research suggests that public support for a cause tends to decline in response to extreme actions because such actions are perceived as immoral (64, 65). We note that the increase in conservative acceptance is not an artifact of our imputation approach because the associations are also found in *observed* responses (online Supplementary material S10). Nor is it likely to have been an artifact in the 2012 NZAVS Cohort. As shown in online Supplementary material S10, acceptance also grew disproportionately among conservatives in the full NZAVS Time 10 cohort (N = 47,948). Because the causes for the gradually increasing acceptance of Muslims in the years before the attacks are presently unknown, we cannot explain this finding.

Similarly, why liberals were stalled in further growth cannot be definitively explained. Of course, explanations might differ for the different strata of political orientation. Speculating, further increases in acceptance among liberals might be challenging because liberals tend to be less religious than conservatives, and anti-Muslim prejudice may be affected by attitudes to religion (57). Future work is needed to better understand what motivated the changes in Muslim acceptance among people across the political spectrum.

Another open question is whether subjective warmth ratings reflect accepting or prejudicial behaviors. Perhaps the most apparent open question is whether the long-term minority acceptance effect observed in New Zealand following the 2019 Christchurch mosque attacks generalizes to other countries.

Importance of findings

Although much remains to be discovered about social attitudes in the wake of terrorist attacks, the New Zealand example is important because it clarifies the minimal boundary of human possibility. The minority acceptance effects that follow far-right terrorist attacks need not diminish over time.

Our study is also interesting for its methods. Quasi-experimental approaches have a long history within the social sciences. In 1950, Stouffer wrote, "Though we cannot always design neat experiments when we want to, we can at least keep the experimental model in front of our eyes and behave cautiously" [(37), p. 356] The counterfactual framework for causal inference has become commonplace in the health sciences (38). However, the counterfactual framework for causal inference remains relatively unfamiliar among observational social scientists (16, 39). We hope our study contributes to a more explicit use of counterfactual reasoning in those social sciences that have yet to embrace causal

Table 6. Effects of the attacks on warmth to negative control groups and warmth to Muslims reveals selective effects on warmth to Muslims.

	Warmth to elderly	Warmth to mentally ill	Warmth to overweight	Warmth to muslims
Attack	−0.015 [−0.084, 0.054]	0.013 [−0.060, 0.086]	0.008 [−0.064, 0.080]	0.280 [0.207, 0.354]
Attack (standardized)	−0.006 [−0.032, 0.020]	0.004 [−0.020, 0.029]	0.003 [−0.020, 0.025]	0.083 [0.061, 0.105]
Num. Obs.	4155	4252	4264	4258

Table 7. Pre-/post-COVID pandemic analysis of Muslim warmth reveals no differences (Time 4 cohort).

Parameter	Coefficient	SE	CI	CI_low	CI_high	t	df_error	P
(Intercept)	4.323	0.025	0.95	4.274	4.372	171.800	3913	0.000
Pre–post	0.016	0.059	0.95	−0.099	0.131	0.271	3913	0.787

inference methods for observational data (40). Here, we contribute to this effort by demonstrating how information-rich panel data may be leveraged to impute counterfactual nonexposure trajectories that are otherwise inaccessible after population-wide exposures. Using psychologically rich information from a national panel study, we identify counterfactual outcomes under the assumption that many years of longitudinal measures repeated on individuals enable reliable imputation of missing counterfactual outcomes.

Although our method might apply in other contexts, we urge caution when estimating counterfactual outcomes when there are population-wide missing observations. In such cases, where preexposure and postexposure panel data are unavailable, or where the evolution over time of such data is inadequately represented by linear models, researchers will likely be unable to define credible counterfactual contrasts, at least not without strong assumptions. On the other hand, where historical trajectories align well with the assumptions inherent to linear models and where individual-level data are accessible for predicting future states, the method we have outlined here could prove valuable for addressing causal questions in the aftermath of population-wide exposures such as pandemics, natural disasters, or conflicts. In any case, causal estimation should not be viewed as a process where one-size-fits-all. Rather, each causal question warrants a custom-made approach, tailor-fitted to its unique circumstances. We advocate for a comprehensive and bespoke approach to each causal question in conjunction with the contribution of subject matter experts. Moreover, as demonstrated in this study, performing sensitivity analyses wherever feasible to test causal assumptions is crucial for understanding the robustness and validity of findings (online Supplementary material S6–S11). Most importantly, we caution researchers against reporting longitudinal correlational data. In observational settings, correlations may only estimate causal effects without bias accidentally.

Again, assumptions play a critical role in causal inference. Our approach uses panel data to generate the mission-critical "what-if" scenarios required to compute causal contrasts. Importantly, two significant events—the 2019 mosque attacks and the 2020 COVID-19 pandemic—unfolded in quick succession. The distress and empathy resulting from the attacks might have increased Muslim acceptance during the pandemic. In a world without the attacks, the enhanced acceptance of Muslims during COVID might not exist. If the attacks had not happened, increased Muslim acceptance during COVID might not have been sustained. Because the contrast condition would have presented lower acceptance than we simulated, this could mean our study *underestimates* the true effect of the attacks. Thus, despite our efforts to be precise, the real-world impact of the attacks might be slightly larger than our findings show. Unfortunately, the attacks occurred, and we cannot know. Regardless, researchers should always strive to report more than just naive trends in observational data. If we did not account for increasing preattack acceptance trajectories, we could potentially exaggerate the influence of the attacks. To avoid this, we have clearly stated our assumptions and methods and verified their robustness using sensitivity analyses. Overall, we think this strategy provides a more dependable approach than reporting naively observed trends over time. Again, it is crucial to remember that basic longitudinal correlations do not clarify causal questions.

George Box famously wrote that all models are wrong but some are useful (66, p. 792). Less famously, Box also wrote: "Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad." (66, p. 792). To naively report correlations and trends is to ignore the tigers. However, methods for causal inference are not magic wands that convert data into blueprints of reality. Rather, they are analytic tools that enable researchers to improve their understanding of causal phenomena. They are essential because they help researchers to remove biases from statistical associations obtained from the applications of models to data. Methods for causal inference also alert researchers to where the applications of models to data might go wrong. Beyond the interest of our findings, we hope this study motivates more social and behavioral scientists to attempt causal inference in their research. When doing so, we recommend stating one's assumptions as clearly as possible, performing sensitivity analyses, and reporting the implications for inference if one's assumptions are wrong.

Finally, we hope this study will interest a general audience beyond the social and behavioral sciences. Despite a positive trend in Muslim acceptance before the attacks, Muslims remained the least accepted minority in New Zealand (5). Our study identified challenges to further growth in Muslim acceptance among liberals. Additionally, we observed that although acceptance among conservatives is rising, it remains low overall. By establishing clear benchmarks of the changes in Muslim acceptance after the 2019 Christchurch mosque attacks, we hope to provide useful guidance to those working towards fostering a more inclusive society for the Muslim minority in New Zealand.

Despite the ongoing challenges that New Zealand encounters in fostering acceptance towards its Muslim minority, we should not overlook the broader, encouraging narrative that has unfolded. In the aftermath of the attacks, New Zealand witnessed a substantial and lasting increase in the acceptance of Muslims. This rise was especially pronounced among political conservatives—the very demographic the far-right terrorist aimed to instigate against Muslims. This unexpected yet enduring surge in acceptance among conservatives, sustained over at least three

Table 8. Pre-/post-COVID pandemic analysis of Muslim warmth reveals no differences (full NZAVS sample).

Parameter	Coefficient	SE	CI	CI_low	CI_high	t	df_error	P
(Intercept)	4.407	0.008	0.95	4.390	4.423	528.650	41416	0.000
Pre–post	0.001	0.015	0.95	−0.028	0.031	0.076	41416	0.939

years, is intriguing and important. Contrary to the intent of the terror act, which sought to incite discord, the terror bred greater unity, sparking the most durable acceptance in the group the terrorist aimed to influence the most. Although nothing can ever compensate for the lives tragically lost, we hope the knowledge that the attacks led to deeper, enduring acceptance, particularly among those it was intended to sway, offers solace and hope.

Notes

[a] To identify mediated effects would require counterfactual contrasts for every element of the attack–response–sequence of interest, as well as for the confounders of any mediators/outcome association. The relevant data are seldom available. Moreover, even with complete data, if the exposure were to affect the confounders of the mediator and outcome association then natural mediated effects are not identified. For example, if the media's response to the attacks was affected by the government's response to attacks, and if both the media's response and the government's response were mediators of the attack/acceptance effect, then the natural indirect effect of the attacks on prejudice through the media response would not be identified. More fundamentally, we might wonder whether the concept of a "media response" and "government response" denote well-defined interventions; yet without well-defined interventions, it is unclear how to interpret evidence for causality (17). However, using the framework of causal inference under multiple versions of treatment, given K variations of treatment A, each variation k is associated with a distinct outcome Y_k. Assuming no confounding exists for the relationship between k and Y with measured confounders L, the causal impact of A—represented by an aggregated measure of all K treatments—on each Y_k can be consistently estimated (18, 19). As indicated in the Method section, we interpret "attacks" to be such a coarsened indicator of the multiple versions of treatment that constitute the complex attack–sequence.

[b] β_2 represents an $n \times L$ matrix for all covariates in L'.

[c] Obtain standard errors through bootstrapping, the delta method, sandwich estimators, or in Bayesian inference, work directly with predicted posterior distributions.

[d] Numerous algorithms exist for imputing missing responses (52–54). We opted for Bayesian multilevel models from our familiarity with such models.

[e] For instance, consider a situation where the outcome is "recovery from illness" and a post-treatment variable is "employment status," which could be affected by the outcome. If those who recover (outcome) are more likely to be employed (post-treatment covariate) and also more likely to provide follow-up data (less likely to have missing outcome data), including "employment status" in the imputation model can introduce bias. This is because the imputation model would, in effect, be using the outcome (recovery from illness) to predict itself through the post-treatment covariate (employment status); on the problem of healthy worker bias, see Robins (60).

[f] Imputation approaches condition on observed outcomes during imputation, and so avoid imputation problems by regression using surrogates [on this problem see Ogburn et al. (61)]. We additionally caution researchers against imputation by surrogates.

Ethics

The NZAVS was approved by the University of Auckland Human Participants Ethics Committee on 2017 September 5 until 2021 June 3 (Reference Number: 014889). All participants granted informed written consent and the University of Auckland Human Participants Ethics Committee approved all procedures.

Acknowledgments

The authors thank the anonymous reviewers for their valuable suggestions.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

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

Author Contributions

J.A.B., M.U.A., K.Y., and C.G.S. conceived the study. J.A.B. developed the counterfactual and statistical approach, did the analysis and the graphs, and wrote the first draft with M.U.A. M.U.A. and K.Y. developed the background psychological research on terrorism and prejudice. C.G.S. developed and managed New Zealand Attitudes and Values Study panel data collection (from 2009 to the present). All authors provided critical feedback and contributed to this manuscript.

Preprints

A preprint of this article is published at https://psyarxiv.com/8tfxm/, DOI: 10.31234/osf.io/8tfxm.

Data Availability

The data described in this study are part of the New Zealand Attitudes and Values Study (NZAVS). Full copies of the NZAVS data files are held by all members of the NZAVS management team and advisory board. A de-identified dataset containing the variables analyzed in this manuscript is available upon request from the corresponding author, or any member of the NZAVS advisory board for the purposes of replication or checking of any published study using NZAVS data. For NZAVS data inquires contact C.G.S.: c.sibley@auckland.ac.nz. The code for this analysis can be found at https://github.com/go-bayes/attacks.

References

1 Royal Commission of Inquiry into the terrorist attack on Christchurch masjidain on 15 March 2019. Executive Summary, 2022. [accessed 2023 Jan 3]. https://christchurchattack.royalcommission.nz/the-report/executive-summary-2/executive-summary/.

2 Sulaiman-Hill RC, *et al.* 2021. Psychosocial impacts on the Christchurch Muslim community following the 15 March terrorist attacks: a mixed-methods study protocol. *BMJ Open*. 11(10): e055413.

3 Cellan-Jones R. 2019. Facebook: New Zealand attack video viewed 4,000 times. https://www.bbc.com/news/business-47620519.

4 Shaver JH, Sibley CG, Osborne D, Bulbulia J. 2017. News exposure predicts anti-Muslim prejudice. *PLoS ONE*. 12(3):e0174606.

5 Sibley CG, *et al.* 2020. Prejudice toward Muslims in New Zealand: insights from the New Zealand attitudes and values study. *NZ J Psychol*. 49(1):48–72.
6 Greaves LM, *et al.* 2020. Comparative study of attitudes to religious groups in New Zealand reveals Muslim-specific prejudice. *Kōtuitui*. 15(2):260–279.
7 Troian J, Arciszewski T, Apostolidis T. 2019. The dynamics of public opinion following terror attacks: evidence for a decrease in equalitarian values from Internet Search Volume Indices. *Cyberpsychology*. 13(3). https://doi.org/10.5817/CP2019-3-4
8 Wollebæk D, Enjolras B, Steen-Johnsen K, Ødegård G. 2012. After Utøya: how a high-trust society reacts to terror–trust and civic engagement in the aftermath of July 22. *Polit Sci Polit*. 45(1):32–37.
9 Jakobsson N, Blom S. 2014. Did the 2011 terror attacks in Norway change citizens' attitudes toward immigrants? *Int J Public Opin Res*. 26(4):475–486.
10 Cohu M, Maisonneuve C, Testé B. 2016. The "Charlie-Hebdo" effect: repercussions of the January 2015 terrorist attacks in France on prejudice toward immigrants and North-Africans, social dominance orientation, and attachment to the principle of laïcité. *Rev Int Psychol Soc*. 29(1):50–58.
11 Legewie J. 2013. Terrorist events and attitudes toward immigrants: a natural experiment. *Am J Sociol*. 118(5):1199–1245.
12 Echebarria-Echabe A, Fernandez-Guede E. 2006. Effects of terrorism on attitudes and ideological orientation. *Eur J Soc Psychol*. 36(2):259–265.
13 Boomgaarden HG, De Vreese CH. 2007. Dramatic real-world events and public opinion dynamics: Media coverage and its impact on public reactions to an assassination. *Int J Public Opin Res*. 19(3):354–366.
14 Noelle-Neumann E. 2002. Terror in America: assessments of the attacks and their impact in Germany. *Int J Public Opin Res*. 14(1): 93–98.
15 Traugott M, *et al.* 2002. How Americans responded: a study of public reactions to 9/11/01. *Polit Sci Polit*. 35(3):511–516.
16 Van der Weele T. 2015. *Explanation in causal inference: methods for mediation and interaction*. New York: Oxford University Press.
17 Hernán MA, Sauer BC, Hernández-D iaz S, Platt R, Shrier I. 2016. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 79:70–75.
18 Van der Weele TJ. 2009. Concerning the consistency assumption in causal inference. *Epidemiology*. 20(6):880–883.
19 Van der Weele TJ, Shpitser I. 2013. On the definition of a confounder. *Ann Stat*. 41(1):196.
20 Shanaah S, *et al.* 2021. Hate begets warmth? The impact of an anti-muslim terrorist attack on public attitudes toward Muslims. *Terror Polit Violence*. 35(1):1–19.
21 Skitka LJ. 2005. Patriotism or nationalism? Understanding post-September 11, 2001, flag-display behavior. *J Appl Soc Psychol*. 35(10):1995–2011. https://doi.org/10.1111/j.1559-1816.2005.tb02206.x
22 Jones JM. 2022. Who had the highest gallup presidential job approval rating? https://news.gallup.com/poll/271628/highest-gallup-presidential-job-approvalrating.aspx.
23 Pew Research Center, Bush and Public Opinion. 2008. https://www.pewresearch.org/politics/2008/12/18/bush-and-public-opinion/.
24 Satherley N, Yogeeswaran K, Osborne D, Shanaah S, Sibley CG. 2021. Investigating the effects of right-wing terrorism on government satisfaction: a time course analysis of the 2019 christchurch terror attack. *Stud Confl Terror*. 1–14. https://www.tandfonline.com/doi/full/10.1080/1057610X.2021.1913819
25 Eger MA, Olzak S. 2022. The polarizing effect of anti-immigrant violence on radical right sympathies in Germany. *Int Migr Rev*. 27: 01979183221126461. https://doi.org/10.1177/01979183221126461
26 Igarashi A. 2021. Hate begets hate: anti-refugee violence increases anti-refugee attitudes in Germany. *Ethn Racial Stud*. 44(11): 1914–1934. https://doi.org/10.1080/01419870.2020.1802499
27 Westreich D, Greenland S. 2013. The Table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol*. 177(4):292–298.
28 Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J Am Stat Assoc*. 91(434): 444–455.
29 Muñoz J, Falcó-Gimeno A, Hernández E. 2020. Unexpected event during survey design: promise and pitfalls for causal inference. *Polit Anal*. 28(2):186–206.
30 Sibley CG, *et al.* 2020. Effects of the COVID-19 pandemic and nationwide lockdown on trust, attitudes toward government, and well-being. *Am Psychol*. 75(5):618–630.
31 Neyman JS. 1923. On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (tlanslated and edited by dm dabrowska and tp speed, statistical science (1990), 5, 465–480). *Ann Agric Sci*. 10:1–51.
32 Rubin DB. 1976. Inference and missing data. *Biometrika*. 63(3): 581–592. https://doi.org/10.1093/biomet/63.3.581
33 Edwards JK, Cole SR, Westreich D. 2015. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. *Int J Epidemiol*. 44(4):1452–1459.
34 Westreich D, *et al.* 2015. Imputation approaches for potential outcomes in causal inference. *Int J Epidemiol*. 44(5):1731–1737.
35 Morgan SL, Winship C. 2014. *Counterfactuals and causal inference: methods and principles for social research*. 2nd ed. Cambridge: Cambridge University Press.
36 Robins JM, Hernán MA. 2020. *Causal inference: what if*. Boca Raton (FL): Chapman & Hall/CRC.
37 Stouffer SA. 1950. Some observations on study design. *Am J Sociol*. 55(4):355–361. https://doi.org/10.1086/220558
38 Lash TL, Rothman KJ, VanderWeele TJ, Haneuse S. 2020. *Modern epidemiology*. Philadelphia: Wolters Kluwer.
39 Rohrer JM. 2018. Thinking clearly about correlations and causation: graphical causal models for observational data. *Adv Methods Pract Psychol Sci*. 1(1):27–42.
40 Bulbulia JA. 2022. A workflow for causal inference in cross-cultural psychology. *Religion Brain Behav*. 1–16. https://doi.org/10.1080/2153599X.2022.2070245
41 VanderWeele TJ, Mathur MB, Chen Y. 2020. Outcome-wide longitudinal designs for causal inference: a new template for empirical studies. *Stat Sci*. 35(3):437–466.
42 Breskin A, *et al.* 2020. G-computation for policy-relevant effects of interventions on time-to-event outcomes. *Int J Epidemiol*. 49(6):2021–2029. https://doi.org/10.1093/ije/dyaa156
43 Chatton A, *et al.* 2020 Jun. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Sci Rep*. 10(1):9219. https://doi.org/10.1038/s41598-020-65917-x
44 Hernán MA, Robins JM. 2023. *Causal inference: what if?* Boca Raton: Chapman & Hall/CRC.
45 VanderWeele TJ, Hernan MA. 2013. Causal inference under multiple versions of treatment. *J Causal Inference*. 1(1):1–20.
46 Sibley CG. 2021. Sampling procedure and sample details for the New Zealand Attitudes and Values Study. *OSF*. https://osf.io/75snb/

47 Richardson TS, Robins JM. 2013. Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality. *Working Paper 128*, Center for the Statistics and the Social Sciences, University of Washington Series. Citeseer.

48 Westreich D, Cole SR. 2010 Mar. Invited commentary: positivity in practice. *Am J Epidemiol*. 171(6):674–677; discussion 678–681.

49 R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

50 Bürkner P-C. 2017. brms: an R package for Bayesian multilevel models using Stan. *J Stat Softw*. 80:1–28.

51 Stan Development Team. 2020. RStan: the R interface to Stan. https://mc-stan.org/users/interfaces/rstan

52 Austin PC, White IR, Lee DS, van Buuren S. 2021. Missing data in clinical research: a tutorial on multiple imputation. *Can J Cardiol*. 37(9):1322–1331. https://doi.org/10.1016/j.cjca.2020.11.010

53 Ding P, Li F. 2018. Causal inference: a missing data perspective. *Stat Sci*. 33(2):214–237. https://doi.org/10.1214/18-STS645

54 Murray JS. 2018. Multiple imputation: a review of practical and theoretical findings. *Stat Sci*. 33(2):142–159. https://doi.org/10.1214/18-STS644.full

55 Arel-Bundock V. 2022. marginaleffects: marginal effects, marginal means, predictions, and contrasts. https://github.com/vincentarelbundock/marginaleffects.

56 Satherley N, *et al*. 2015. Demographic and psychological predictors of panel attrition: evidence from the New Zealand attitudes and values study. *PLoS ONE*. 10(3):e0121950.

57 Shaver JH, Troughton G, Sibley CG, Bulbulia JA. 2016. Religion and the unmaking of prejudice toward muslims: evidence from a large national sample. *PLoS ONE*. 11(3):e0150209.

58 Bürkner P-C. 2021. Bayesian item response modeling in R with brms and Stan. *J Stat Softw*. 100(5):1–54.

59 Zhang J, Dashti SG, Carlin JB, Lee KJ, Moreno-Betancur M. 2023. Should multiple imputation be stratified by exposure group when estimating causal effects via outcome regression in observational studies? *BMC Med Res Methodol*. 23(1):42. https://doi.org/10.1186/s12874-023-01843-6

60 Robins J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 7(9):1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6

61 Ogburn EL, Rudolph KE, Morello-Frosch R, Khan A, Casey JA. 2021. A warning about using predicted values from regression models for epidemiologic inquiry. *Am J Epidemiol*. 190(6):1142–1147.

62 McNeish D, Stapleton LM, Silverman RD. 2017. On the unnecessary ubiquity of hierarchical linear modeling. *Psychol Methods*. 22(1):114–140.

63 Lüdecke D. 2021. ggeffects: create tidy data frames of marginal effects for ggplot from model outputs. https://strengejacke.github.io/ggeffects/.

64 Feinberg M, Willer R, Kovacheff C. 2020. The activist's dilemma: extreme protest actions reduce popular support for social movements. *J Pers Soc Psychol*. 119:1086–1111.

65 Mitts T, Phillips G, Walter BF. 2022. Studying the impact of ISIS propaganda campaigns. *J Polit*. 84(2):1220–1225. https://doi.org/10.1086/716281

66 Box GEP. 1976 Dec. Science and statistics. *J Am Stat Assoc*. 71(356):791–799. https://doi.org/10.1080/01621459.1976.10480949