HCEO WORKING PAPER SERIES

How do Parole Boards Respond to Large, Societal Shocks? Evidence from the 9/11 Terrorist Attacks

Brendon McConnell
Kegon Teng Kok Tan
Mariyana Zapryanova

Working Paper

2023-010

03/2023; Revision 10/2023

HUMAN CAPITAL AND ECONOMIC OPPORTUNITY GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

How do Parole Boards Respond to Large, Societal Shocks? Evidence from the 9/11 Terrorist Attacks

Brendon McConnell* Kegon Teng Kok Tan† Mariyana Zapryanova‡

October 15, 2023

Abstract

We provide the first evidence of the impact of 9/11 on outcomes for Muslims in the US criminal justice system. We focus on parole outcomes of Muslim men in the state of Georgia, and find large post-9/11 declines in the likelihood of being granted parole and a subsequent 23% relative increase in prison time for Muslim inmates. We find that these impacts persisted for several years after 9/11 with similar sized magnitudes. We examine heterogeneity in the effects by recidivism risk and document suggestive evidence that the effects were larger for higher risk inmates.

JEL codes: D91, J15.
Keywords: parole board, discrimination, terrorist attacks.

*Department of Economics, University of Southampton. Email: brendon.mcconnell@gmail.com.
†Department of Economics, University of Rochester. Email: ttan8@ur.rochester.edu.
‡Department of Economics, Smith College. Email: mzapryanova@smith.edu.

We are grateful to Doctor Tim Carr at the Georgia Department of Corrections for sharing the administrative prison files with us. We are thankful to Steve Hayes at the Georgia Board of Pardons and Paroles for many useful conversations regarding the parole process in Georgia. We benefited from useful comments and suggestions from seminar participants at Smith College Economics Brown Bag Seminar, Washington and Lee University Economics Seminar, and the University of Wisconsin-Madison Empirical Micro Seminar, as well as conference participants at the 2022 Southern Economics Association Conference and the Spring 2023 NBER Law and Economics Meeting. We thank Monica Deza, Steven Durlauf, Peter Grajzl, Robert Kaestner, Stéphane Mechoulan, Mike Mueller-Smith, Imran Rasul, and Lucie Schmidt for their feedback. Priscilla Liu provided excellent research assistance.

1 Introduction

The degree and extent of discrimination in the American criminal justice system is a topic that has long been debated, and has, in recent years, intensified and broadened to cover policing, bail, sentencing, and reentry. The focus of the debate has centered on disparate treatment by race, yet there is evidence from both the US and other countries of disparities across other dimensions including ethnicity, gender and religion. Despite the multitude of studies, it is still unclear to what extent large disparities in outcomes arise due to discriminatory practices versus underlying unobserved differences across groups.

In this paper we focus on disparities in parole outcomes, a key aspect of prison reentry that allows prison inmates to be released before their sentences are fully served. Our interest lies in investigating religious-based disparities – specifically disparate parole outcomes for Muslim inmates. Our focal point is the time period around a large, societal event – the 9/11 terrorist attacks. Whilst the 9/11 attacks impacted a wide swathe of outcomes and led to changes in many dimensions of life in the US, including policing responses, the war in Afghanistan, and numerous psychological effects (Davis, 2007; Woods, 2011), the US Muslim population was particularly affected in terms of labor market (Davila and Mora, 2005; Kaushal et al., 2007), victimization (Singh, 2002), criminalization (Kaufman, 2019), discrimination (Sheridan, 2006), societal resentment and reservations (Panagopoulos, 2006), and assimilation (Gould and Klor, 2016). We conceptualize the attacks as an exogenous shock to the level of animosity towards Muslims.

We estimate the impact of such a shock to animosity towards Muslims in a regression-adjusted difference-in-differences (DD) framework, conditioning on a rich set of relevant control variables. We interpret our DD estimates as the causal effects of the parole board response to the 9/11 terrorist attacks on parole outcomes for Muslim inmates. We provide a battery of evidence in support of the key identifying assumptions—namely the parallel trends assumption and the stability of the composition of the groups before and after 9/11—that underpin this causal interpretation.

Our study is the first to investigate the impact of the criminal justice system response to 9/11 on Muslims in the US. We use administrative records from the Department of Corrections in the state of Georgia, which importantly contain information on self-reported religion.[1] We

[1]Religion is collected as a part of the process of being admitted into the Georgia State Correctional System and is reported once the convicted felon is transferred from the court to one of the state's diagnostic and classification prisons. For our core analysis, we restrict our sample to those sentenced prior to 9/11/01, thus ruling out endogenous reporting of Muslim religion status in response to the attacks.

focus on Black, male, parole-eligible inmates as the overwhelming majority of Muslim inmates in our data are Black.

We document a substantial *short-run* change in the parole outcomes of Muslim inmates in the aftermath of 9/11, detailing impacts for those who came up for parole within a one year window around the attacks. At the extensive margin, we find a 12 percentage point (17%) reduction in the probability of receiving parole. This translates to roughly 200 more days in prison, a 23% increase. To benchmark this effect, our DD estimate is of a similar magnitude to the difference in prison time between serial offenders (8 or more prior convictions) and first-time offenders. These estimates are all statistically significant. We also show that the large short-run impacts we document persist after 9/11 to the end of our sample period, in contrast to other outcomes for Muslims that seem to exhibit large short term effects and substantial fade out (e.g., homicide rates against Muslims (Gould and Klor, 2016)). Within an event study framework, we cannot reject equality of the short-run estimates (0-1 years post-9/11) with the medium-run estimates (3-4 years post-9/11).

We perform a decomposition analysis to show that observable characteristics of Muslim inmates relative to non-Muslim inmates are unable to explain the results, and our research design also suggests that unobservable individual inmate characteristics would likewise be unlikely to explain the results. We also check for changes in inmate behavior while in prison and whether raters who prepare the parole file reacted to 9/11 and find no statistically significant effects. In our view, it is likely that the effects on parole outcomes stem from the parole board's decision making rather than the profiles or in-prison behaviors of the Muslim inmates. Taken together, we believe unwarranted discrimination drove the disparity in parole outcomes.

While it is challenging and beyond the scope of this paper to parse the various potential sources of discrimination, we attempt to shed light on the nature of the discrimination by constructing a measure of ex-ante recidivism risk and conduct a heterogeneity analysis of our core parole board outcomes based on this risk measure. We document that parole outcomes for Muslim inmates post-9/11 worsen monotonically (and almost perfectly linearly) with ex-ante recidivism risk. For inmates with low predicted recidivism risk, the change in outcomes in the post-9/11 period is muted, whereas for those with high ex-ante recidivism risk, we find more pronounced effects post-9/11. We interpret this as evidence that our short-run estimates are unlikely to be driven by an unconditional discrimination mechanism, but rather are sensitive to the underlying recidivism risk of inmates.

Our work contributes to three key literatures. First, we contribute to the growing empirical literature on the influence of extraneous events and factors on the application of justice.[2] Within this literature, attention has nearly exclusively focused on the decision-making of judges (Brodeur and Wright, 2019; Eren and Mocan, 2018; Philippe and Ouss, 2018), prosecutors (Bielen and Grajzl, 2021; McConnell and Rasul, 2021), and juries (Bindler and Hjalmarsson, 2019). Ours is the first study that sheds light on impact of extraneous factors on the decision-making process of parole boards. The decisions of parole board members are not only different of judges or juries, where the degree of discretion and stakes of the decisions are quite different, but also are highly impactful given the wide use of discretionary parole boards in the US criminal justice system.[3]

Second, by providing the first evidence of how parole board decision making was impacted by the shock of the terrorist attacks of 9/11, we contribute to a body of empirical work that has studied the impact of terrorism on various societal outcomes. For criminal-justice outcomes, scholars have documented impacts on federal criminal sentences (McConnell and Rasul, 2021), asylum approvals (Brodeur and Wright, 2019), and civil cases (Shayo and Zussman, 2011).[4]

Finally, with our focus on disparate outcomes for Muslim inmates, we expand the scope of the literature studying bias in the US criminal justice system. Existing evidence from the US has largely focused on examining *racial bias* in the decision-making of police officers (Feigenberg and Miller, 2022; Goncalves and Mello, 2021), judges (Arnold et al., 2018), juries (Flanagan, 2018), prosecutors (Sloan, 2019), parole boards (Anwar and Fang, 2015; Mechoulan and Sahuguet, 2015), and parole officers (LaForest, 2022). Whilst some papers focus on religious bias in the context of the criminal justice system in India (Ash et al., 2021) and the Netherlands (Bielen and Grajzl, 2021), our paper is the first to focus on religion as a basis for discrimination rather than race or ethnicity in the US criminal justice system.

The rest of the paper is organized as follows. Section 2 outlines the parole process in Georgia and describes the data. Section 3 describes presents our empirical strategy. Section 4 presents our main results while Section 5 discusses potential mechanisms behind our results. Section 6 concludes.

[2]See Ludwig and Mullainathan (2021) for a review.

[3]In 2019, 413,985 individuals were released on parole, of which 178,074 were released via the discretion of a parole board (Oudekerk and Kaeble, 2019).

[4]Examples of other economic outcomes that the literature has found to be affected by terrorism include labor market (Cornelissen and Jirjahn, 2012; Glover, 2021; Kaushal et al., 2007), housing market (Lepage, 2023), macroeconomy (Abadie and Gardeazabal, 2003; Blomberg et al., 2004), and assimilation (Gould and Klor, 2016).

2 Parole in Georgia and Data

2.1 Institutional Setting

Sentenced felons in Georgia are transferred from the court to a diagnostic prison where they go through a battery of tests and diagnostic questionnaires before being assigned to a prison. Importantly, for our study prisoners self-report their religion during this diagnostic process. Once prisoners are assigned to prisons, the parole process is initiated with a pre-parole investigation. This investigation is conducted by a parole hearing examiner, a role we also refer to as rater, and comprises of the rater interviewing the prisoner and gathering information about the prisoner's personal information and criminal record. In addition, as part of this investigation, the rater uses Parole Decisions Guidelines Grid System (hereafter the grid)[5] along with the prisoner's risk score and current offense crime severity level, to determine the recommended prison time, also known as the grid recommendation. Once the pre-parole investigation is finished, the rater prepares prisoner's parole file and writes a summary discussing its contents. When the parole file reaches the members of the parole board, it usually contains all records from the diagnostic prison, pre-parole investigation reports, the grid risk score, severity level and recommendation, and the rater's summary of the content of the parole file.

The state of Georgia releases prisoners from prison using a discretionary parole system, where releases are granted on the assessment and full discretion of a parole board.[6] The Georgia parole board consists of five-members appointed by the governor to seven-year terms subject to confirmation by the State Senate.[7] Most parole-eligible inmates become statutorily eligible for parole release after serving one-third of their prison sentence (O.C.G.A. §42-9-45). Around this time, the parole file is sent, in a randomized order, to each of the five parole board members, one at a time. After reviewing the file, each member marks their decision on a ballot on the parole file. A parole decision is reached if three out of five board members vote the same way (O.C.G.A. §42-9-42).

[5]The board is required by law to make release decisions on the basis of the risk a person may pose to public safety if they were to be released from prison (O.C.G.A. §42-9-40). To assess this risk the board has adopted the grid, which we present in Table B1.

[6]Using data from Renaud (2019), we compare the parole system in Georgia to that of other states with discretionary parole systems on some key characteristics in Table B2. Overall, the parole system in Georgia is comparable to the national average, especially in terms of board size, but differs in terms of its use of guidelines for prison releases. In addition, Zapryanova (2020) shows that Georgia's prison population appears to be representative of that nationwide.

[7]It is important to note that there were no changes in the parole board composition during our main estimation sample period (Godfrey et al., 2022).

2.2 Data and Sample Selection Criteria

Our data are sourced from rich administrative internal records of the Georgia Department of Corrections (GDC), and include a record of all prisoners admitted in prison in Georgia from 1980 to 2008. We observe detailed information on the prisoner demographic characteristics, including prisoner's self-reported religion at admission to prison, and on the parole board decision-making process. In addition, we observe the date on which each prisoner was rated by the grid. This date is the earliest date on which the parole file is complete and passed to the parole board for a vote.[8]

We make several sample restrictions that we visually present in schematic Figure A1. First, we exclude prisoners who were rated within 180 days prior to 9/11 in order to avoid contaminating our control group with prisoners who might have been rated prior to 9/11 but whose parole file was ultimately reviewed by the board after 9/11. We check the sensitivity of our results to this data restriction in Figure D2. Second, we base our sample on prisoners rated by the Georgia parole board within a 365 day window around 9/11. Third, we ensure that all defendants in our sample have been sentenced prior to 9/11 and have been released or have a release date set up after 9/11. Note that some of these prisoners were rated by the grid pre 9/11 while others ended up being rated post 9/11. To maintain comparability of both groups we restrict the sample further so that those defendants rated before 9/11, they were sentenced at least 365 days before 9/11. Finally, we restrict our sample to Black male parole-eligible inmates with non-missing admission, release, sentence and rate dates. We do this for two reasons. First, virtually all Muslims in our data are Black males.[9] Second, there is a large literature documenting racial and gender disparities in criminal justice outcomes. By restricting our sample to Black male inmates, we are able to focus solely on our key treatment variable – Muslim religion status – without complications from sampling variation leading to different proportions of other protected characteristics across both treatment and time that could drive differences in potential confounders.

[8] We do not observe the exact date on which the parole board makes a decision. However, we use the rate date as the earliest date on which the parole file is ready to be reviewed by the board.

[9] Of individuals sentenced in the decade running up to the 9/11 attacks, Black male inmates represent 94.3% of the Muslim inmate population, compared to 56.3% of the non-Muslim inmate population.

3 Empirical Approach and Identification

Our empirical approach takes the form of a difference-in-differences (DD) specification as follows:

$$y_{it} = \alpha_1 Post_t + \alpha_2 Muslim_i + \beta(Post_t \times Muslim_i) + \gamma X_i^{'} + \pi_t + \epsilon_{it} \ , \tag{1}$$

where y_{it} is the parole outcome of interest for prisoner i rated at time t, $Post_t$ is an indicator that takes the value 1 for prisoners with parole files prepared and ready to be reviewed by the board after 9/11, and 0 otherwise, and $Muslim_i$ is an indicator for Muslim religion status. Importantly for our setting, the religion of the inmates was recorded prior to 9/11 because it was collected during the prisoners' intake from the court into the Georgia State Correctional System, thus ruling out endogenous recording of treatment status as a function of 9/11.

We condition on a rich set of covariates, the most important of which are a series of dummy variables for each of the 21 guidelines cells used as part of the determination of recommended prison time in Georgia.[10] These guidelines combine a prisoner's parole risk score and the severity level of the crime that placed the inmate behind bars. We additionally control for the sentence length received from the sentencing judge, dummies for having children, being married, age at sentencing deciles, education categories, quartiles of Culture Fair IQ test score, indicators for the most serious offense committed, socio-economic status, and dummies for number of prior convictions. The fixed effects, π_t, capture any rating month-specific unobservables. The error term is ϵ_{it}. We use Eicker-Huber-White standard errors throughout. We note that there are no district or area fixed effects, and no parole board fixed effects – the files of all inmates are reviewed by the same parole board.[11]

3.1 Identification

The key identifying assumption underpinning our empirical approach is that Muslim and non-Muslim inmates experience common trends in parole board outcomes. Taking into account the recent critique to canonical pre-trends testing made by Roth (2022), we provide a battery of evidence using multiple approaches in support of parallel trends in our setting.

We first implement a set of placebo DD regressions. We shift all key dates one year back in time, and re-estimate Equation 1, with the sole difference that now the $Post_t$ term takes value zero for the period 11 September 1999–10 September 2000, and one for the period 11

[10]See Table B1 for the guideline cell grid in operation during the period of study. Note that in our estimation sample we do not observe any prisoners in the highest severity category.

[11]Our main specification is robust to inclusion of rater fixed effects that account for any time-invariant heterogeneity of the ways raters prepare the parole files.

September 2000–10 September 2001. We present the results in Table C1 and find no significant placebo DD estimates. Next, we provide graphical evidence of the lack of existence of pre-trends by presenting the raw, underlying data for three years prior to our estimation sample – the calendar years of 1998-2000. We cannot reject the null of equality of trends in any case based on Figure C1. Finally, we implement the honest difference-in-differences approach of Rambachan and Roth (2022), in order to create worst-case treatment effect bounds for potential violations of the parallel trends assumption, based on pre-trends. We discuss these results in Section C.1.3. Taken together, the evidence we present here is strongly supportive of parallel trends in parole board outcomes for non-Muslim and Muslim inmates in the period prior to 9/11.

Given that we are using repeat cross-sectional data for our empirical analysis, we also provide evidence for a second identifying assumption – that the composition of the two groups is stable across the pre and post periods (Blundell and Dias, 2009). We do so in two ways. First, in Table C3 we present the results of a series of balance tests. Column (3) and Column(6) show the p-values for a null of no difference in means across the two periods, for non-Muslim and Muslim inmates respectively. We cannot reject the null in 20 out of 22 cases. Non-Muslim inmates have slightly more prior convictions post-9/11 and Muslim inmates are less likely to be married in the post-9/11 sample. Column (7) presents the p-value of the difference-in-differences across the control variables. These p-values never fall below .05. We interpret these results as supportive of the assumption of group composition stability. Second, we test whether there is any strategic reordering of when Muslim and non-Muslim inmates appear before the parole board in the aftermath of 9/11 by implementing a series of duration model regressions, where the key duration variable is the time from prison admission to rate date. We present these in Table C4 and find no evidence that this is the case, either in the raw durations or once we condition on our main set of control variables.

One potential threat to our identification strategy could be that we fail to observe Muslim-status over time, in a setting where religious conversions are potentially common (Boddie and Funk, 2012; Hamm, 2007; Kusha, 2016). If 9/11 lead to changes in both the number of conversions and the composition of who converts, this could result in selection bias. That means that even if 9/11 had no causal effect on parole board decision making, differential composition across the treatment and control groups in the pre and post periods could lead us to detect changing outcomes using our DD approach. We address this concern in Appendix D.8, where we show that our results are highly robust to different measures of Muslim exposure in prison, which serves as a proxy for conversion rates.[12]

4 Results

Table 1: Parole Board Decisions and Prisoner Outcomes

	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	-.12**	–192**	202**
	(.0566)	(81.2)	(81)
$\overline{Y}_{0,PRE}$	.71	961	869
(Post-9/11×Muslim) / $\overline{Y}_{0,PRE}$	-.169**	-.2**	.232**
	(.0797)	(.0845)	(.0932)
Adjusted R^2	.283	.84	.549
Number Muslim Inmates	222	222	222
Observations	4,832	4,832	4,832

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

We present our main results in Table 1. The DD coefficients we estimate are large, statistically significant, and economically meaningful. We find that the parole outcomes of Muslim inmates in Georgia were negatively impacted as a consequence of 9/11. Post 9/11, Muslim inmates are 12 percentage points less likely to be granted parole. Expressed in terms of the pre-9/11, non-Muslim average (which we denote by $\overline{Y}_{0,PRE}$), this is a 17% reduction in the probability of receiving parole. The effect is statistically significant at the 5% significance level. This lower likelihood of parole translates in 194 fewer days on parole. Given that the average parole period for non-Muslim inmates in the pre-9/11 period is 961 days (or 2.6 years), our DD estimates amounts to a 20% reduction in time on parole. Finally, we present results for days in prison. In line with the parole time results, we find Muslim inmates spend over half a year longer in prison if their file is reviewed by the parole board after 9/11, a 23% increase compared to the reference sub-sample of non-Muslim inmates reviewed prior to 9/11. To understand the magnitude of this effect, we benchmark our DD estimates with the estimates from other control variables. Our estimate of the increase in days in prison for Muslim inmates whose cases were reviewed post-9/11 (202 days) is of the same magnitude as the average conditional difference in time served between a serial offender (eight or more prior convictions) and a first-time offender (192 additional prison days).

[12]Last names may also provide clues about religious status but in our data, the names are static for any given prison spell, and most Muslims in our data have common Black American names rather than Muslim sounding names.

4.1 Longer-Run Effects

We next assess the extent to which our main DD estimates persist. To do so, we expand our sample by considering years from 1998 to 2005. We maintain similar sampling rules for defining the extended sample.[13] We estimate a dynamic version of Equation 1 as follows:

$$y_{it} = \alpha Muslim_i + \sum_{\substack{t=1998, \\ t\neq 2000}}^{2004} \beta_t \left(Period_t \times Muslim_i\right) + X_i^{'}\gamma + \theta_t + \pi_t + \epsilon_{it}, \tag{2}$$

where $Period_t$ denotes a year that starts on 9/11 of a given calendar year, t, and runs until 9/10 of the following calendar year, and θ_t are period fixed effects. We present the resulting estimates in the form of event study graphs in Figure 1.

For all three outcomes, we document a striking persistence of the short-run effects we detail in Section 4, namely the large declines in parole grants and days on parole and the corresponding increase in days in prison. For days paroled and days in prison, the long-run effects (i.e., estimates for the year 2004/2005) are statistically significantly different from zero at conventional levels (the associated p-values are respectively .027 and .059). For all outcomes, we cannot reject the null that the short- and long-run effects are equal. The persistence of the effects against Muslims we document contrasts with studies that have documented short-term impacts of 9/11 on hate crimes (Gould and Klor, 2016) and labor market outcomes (Kaushal et al., 2007).

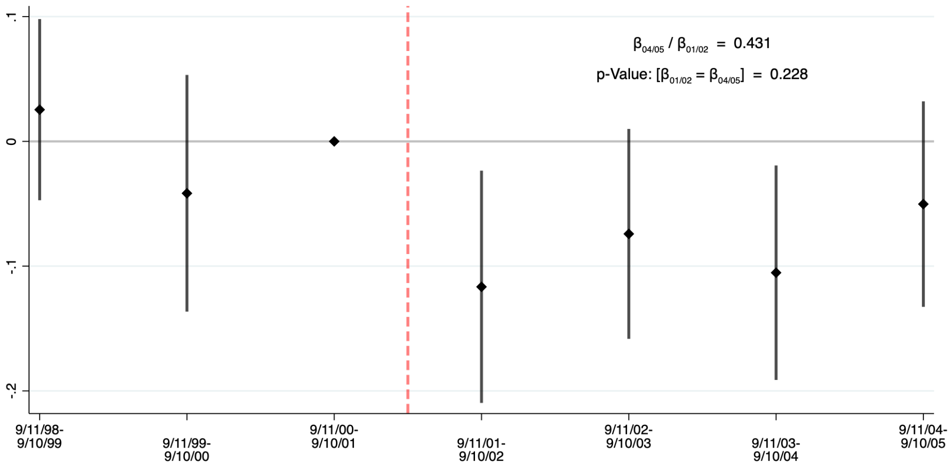
5 Discussion

The size and persistence over time of our estimates begs the question: what drives these results? To begin, we perform a Juhn, Murphy, and Pierce style decomposition and show that observable characteristics are unable to explain the disparity generated by 9/11 (see Figure 2). While disparities do not always entail discrimination if there are characteristics that may be observable to the parole board but not to the econometrician, the decomposition and the setting provide strong support for discrimination underlying the results.

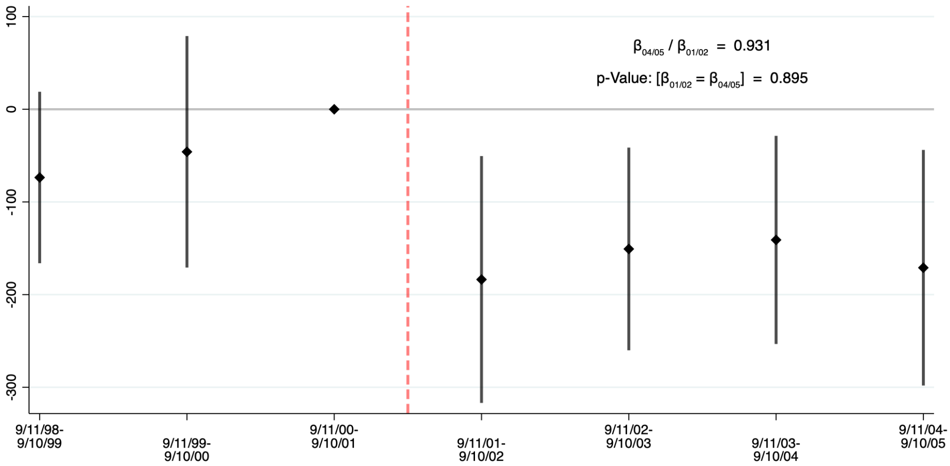
Aside from inmate characteristics, we also investigate whether the raters could have played a role in our main results. While raters have somewhat limited scopes for impacting the parole outcome as their job is primarily to compile information for the board, we check to see if controlling for rater effects could change our estimates. We show that the inclusion of rater

[13] Specifically, we implement the same sample selection procedure that we do for our core sample, outlined in Appendix A, for each year in our extended sample. This does mean that for later years, we may include inmates who are sentenced post-9/11.

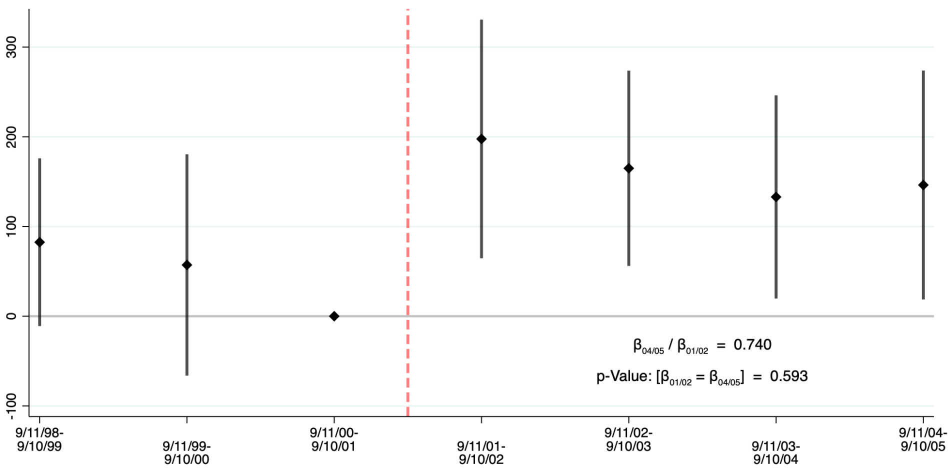
Figure 1: Dynamic Effects for Parole Outcomes



(a) Parole Granted



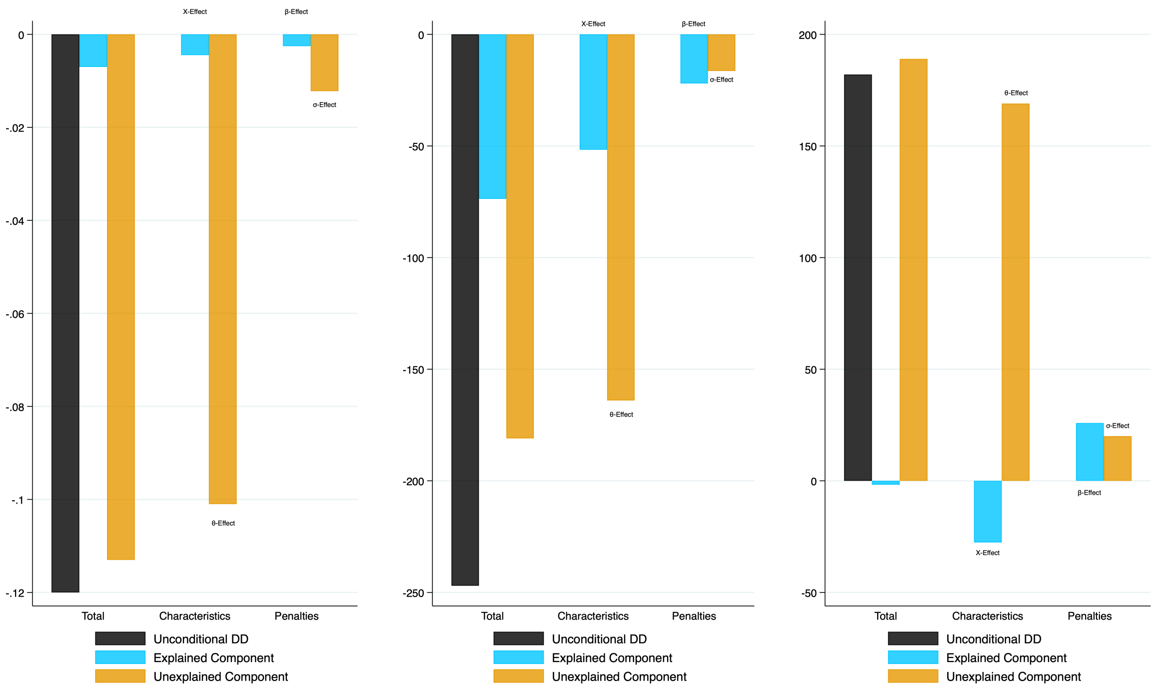
(b) Days Parole



(c) Days Prison

Notes: The lines indicate the 90% confidence intervals of the point estimates, represented as diamonds. The time variable, displayed on the x-axis, shows the date range of when the inmate was rated. The omitted time period is the year prior to 9/11/01, and comprises inmates rated between 9/11/00 and 9/10/01. Regression specifications include the follow control variables: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+.

Figure 2: Decomposition Analysis

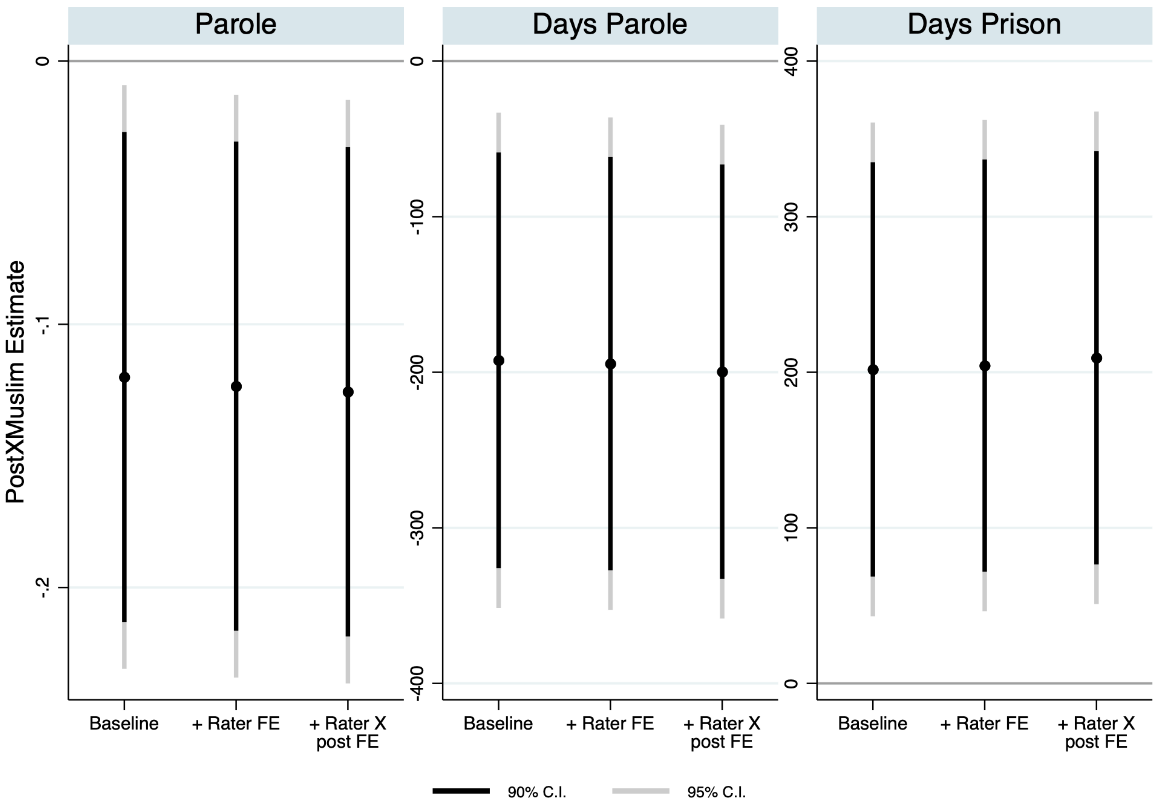


(a) Parole Granted (b) Days Parole (c) Days Prison

Notes: The unconditional DD estimates for parole outcomes are decomposed into explained and unexplained components. The components are further decomposed into differences due to inmate characteristics (quantity) versus differences due to the coefficients capturing the impact of said characteristics on parole outcomes (price). Characteristics include the following control variables: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+.

fixed effects allowing for changes to the rater effect post 9/11 do not affect our results (see Figure 3). Similarly, the timing for when a Muslim inmate's file is prepared does not seem to have changed significantly after 9/11, suggesting that raters are not strategically reordering their rating dates to penalize Muslims.

Figure 3: Rater sensitivity analysis



The results so far suggest that the new disparity in parole outcomes generated by 9/11 is likely driven by discrimination on the part of the parole board. Inmate characteristics do not appear to explain the results and other in-prison processes also do not seem to have changed.

To shed light on the kind of animus that might underlie parole board choices, we check if our results vary by recidivism risk. Table D3 presents the one-year recidivism risk for different groups of former inmates who were released at least one year prior to 9/11. This choice of sample means that all released prisoners spend their full non-custodial parole supervision in the community prior to 9/11. We first see that just over one in ten former inmates recidivates in the first year. What is more relevant for our previous results becomes apparent in Panel B – the recidivism risk for those convicted of a low-severity offense is more than double that of those convicted of high-severity offenses. In Panel C, we present the recidivism risk for quartiles of predicted recidivism risk. The monotonic increase across quartiles is by design.

Using these predicted recidivism risk quartiles, we conduct a further set of heterogeneity analysis – we estimate a triple difference version of our baseline specification, where the third difference is recidivism risk quartile.[14] We plot the quartile-specific DD estimates against quartile-specific baseline recidivism risk. The plots can be found in Figure 4. The significant heterogeneity by recidivism risk factors suggest that the board likely used known correlates of recidivism to inform their choices.

5.1 Discrimination

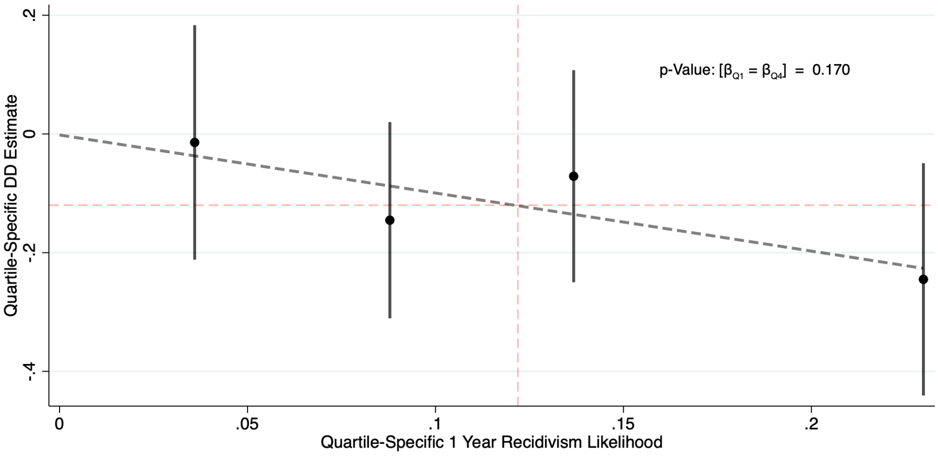
Understanding the type of discrimination at play is challenging and difficult. There are three major categories of discrimination that have been discussed in the literature. The first two are widely known in economics, and consist of taste-based (or preference-based) discrimination, as well as statistical-based discrimination (Lang and Spitzer, 2020). The third, that has had much more visibility in the other social sciences, is institutional discrimination (Small and Pager, 2020). These are established protocol, perhaps arising due to a combination of historical taste or statistical discrimination, but manifest without the need for human agency to actively discriminate. An important commonality, across all types of discrimination, is how notoriously difficult it is to identify and quantify discrimination. On top of the ever-pervasive omitted variables problem that plagues any comparison between groups (Heckman, 1998), further issues such as the well-known inframarginality problem (Ross and Yinger, 2002) are also formidable challenges.

In our setting, the institutional component is unlikely to be at play as the exogenous shock of 9/11 did not immediately alter institutional details of the parole decision making process. Increases in religious disparities of parole outcomes could be due to more prejudicial views regarding Muslim inmates by parole board members after 9/11. Such views would depend on the parole board members being able to observe the religious status of inmates (Rose, 2022), which they do in the rating files that they receive. Taste discrimination may also be heterogenous with respect to other characteristics of Muslim inmates, and persistent, hence consistent with our heterogeneity by recidivism risk results (Brock et al. 2012, Arnold et al. 2018).

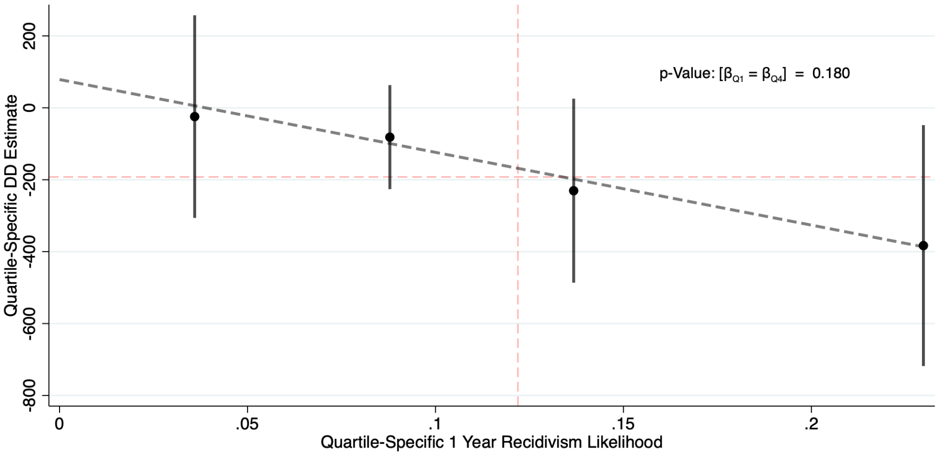
On the other hand, a statistically-based explanation would posit that parole board members update their beliefs regarding the relative recidivism risks of Muslim inmates after 9/11. The parole board members could have raised their expectations of recidivism risk for Muslim inmates

[14] We construct predicted recidivism risk based a regression of 1 year recidivism on a rich set of individual and offense characteristics for the sample released 9/11/1998-9/10/2000. We then predict ex-ante recidivism risk for our baseline sample of interest using their characteristics and the previously estimated coefficients.

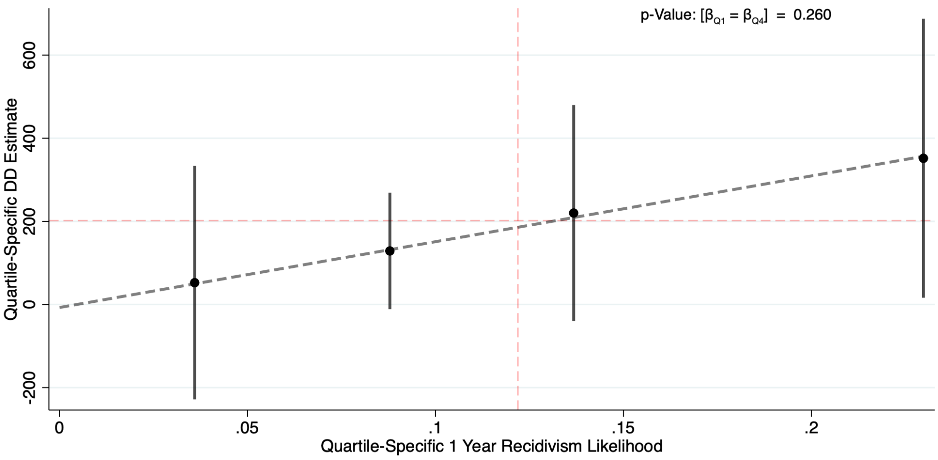
Figure 4: Sub-group DD Estimates Versus Sub-group Baseline Recidivism Risk



(a) Parole Granted



(b) Days Parole



(c) Days Prison

Notes: Each point represents the quartile-specific DD estimate, and the solid lines the 90% confidence intervals. The lines of best fit are based on OLS (thick, dashed line), where each point is inverse weighted by the variance of the quartile-specific DD estimate. Within the graph, a p-value is presented based on a test of equality of parameter estimates for the first and fourth quartiles. Regression specifications include the follow control variables: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+.

if they believed that the broader social climate for Muslims post-9/11 worsened, making it more likely for Muslim inmates to recidivate.[15] This includes potentially social discriminatory factors such as anti-Muslim sentiment in the labor market, policing rates, harshness of judges, and others, all contributing to heightened recidivism risk. If the conditions Muslim inmates face upon release are persistently worse than before 9/11, this would be consistent with our results. We search for evidence that such concerns apply to our sample by looking at labor market outcomes in the county of release, and an analysis of the names of Muslims in our sample to see how salient their Muslim identity might be in a post-9/11 world. Our results show that both dimensions are insignificant.

6 Conclusion

Using administrative data from the state of Georgia, and a difference-in-differences approach, we provide the first evidence of how Muslims in the criminal justice system were affected by the terrorist attacks of 9/11. Outcomes worsen for some Muslims reviewed for parole in the aftermath of 9/11 – these inmates are 17% less likely to be granted parole, and consequently spend 200 additional days in prison, an average – a 23% increase from baseline. The effects on parole outcomes are strongly persistent up to 2005, 4 years after the attacks. We further document that the Muslim inmates reviewed post-9/11 with higher ex-ante recidivism risk experience the largest falls in parole likelihood, and greatest increase in prison time. Taken together with the dynamic effects, we suggest that risk-based factors likely played a sizeable role in the responses of the parole board.

Our work has policy implications for the optimal design of criminal justice systems. If there are large, external shocks, such as the one that we study in this work, the parole board could potentially discriminate in their decision making in response, even though individual characteristics of inmates did not change. This is very different from a purely judicial-based system, where there is little scope to respond to external shocks post-sentencing.

It is worth noting that even if the board made choices based on real increases in recidivism risk among Muslim inmates, these increases in risk may be driven by broader discrimination that in turn may be undesirable. Our paper highlights the open question of whether the criminal justice system should reflect and buttress potential societal discrimination as a part of their duty. Any debate regarding how to reform the criminal justice system, in order to make it

[15]We note that even if discrimination were statistical-based, this is arguably an unjustifiable penalty on the Muslim inmates as it reflects broader social discrimination in the decision making as a factor of consideration for parole.

both more efficient and more equitable, should thus consider the nuanced role of a parole board system, given the response that we document in this paper.

References

Abadie, A. and J. Gardeazabal (2003): "The economic costs of conflict: A case study of the Basque Country," *American economic review*, 93, 113–132.

Anwar, S. and H. Fang (2015): "Testing for racial prejudice in the parole board release process: Theory and evidence," *The Journal of Legal Studies*, 44, 1–37.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial bias in bail decisions," *The Quarterly Journal of Economics*, 133, 1885–1932.

Ash, E., S. Asher, A. Bhowmick, D. L. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi (2021): "Measuring gender and religious bias in the indian judiciary," *Center for Law & Economics Working Paper Series*, 2021.

Bielen, S. and P. Grajzl (2021): "Prosecution or Persecution? Extraneous Events and Prosecutorial Decisions," *Journal of Empirical Legal Studies*, 18, 765–800.

Bindler, A. and R. Hjalmarsson (2019): "Path dependency in jury decision making," *Journal of the European Economic Association*, 17, 1971–2017.

Blomberg, S. B., G. D. Hess, and A. Orphanides (2004): "The macroeconomic consequences of terrorism," *Journal of monetary economics*, 51, 1007–1032.

Blundell, R. and M. C. Dias (2009): "Alternative approaches to evaluation in empirical microeconomics," *Journal of Human Resources*, 44, 565–640.

Boddie, S. C. and C. Funk (2012): "Religion in prisons: A 50-state survey of prison chaplains," in *Pew Forum*, Pew Research Center Washington, DC.

Brodeur, A. and T. Wright (2019): "Terrorism, immigration and asylum approval," *Journal of Economic Behavior & Organization*, 168, 119–131.

Cai, J. and S.-Y. Wang (2022): "Improving management through worker evaluations: Evidence from auto manufacturing," *The Quarterly Journal of Economics*, 137, 2459–2497.

Cornelissen, T. and U. Jirjahn (2012): "September 11th and the earnings of Muslims in Germany—The moderating role of education and firm size," *Journal of Economic Behavior & Organization*, 81, 490–504.

Davila, A. and M. T. Mora (2005): "Changes in the earnings of Arab men in the US between 2000 and 2002," *Journal of Population Economics*, 18, 587–601.

Davis, D. W. (2007): *Negative liberty: Public opinion and the terrorist attacks on America*, Russell Sage Foundation.

Eren, O. and N. Mocan (2018): "Emotional judges and unlucky juveniles," *American Economic Journal: Applied Economics*, 10, 171–205.

Feigenberg, B. and C. Miller (2022): "Would eliminating racial disparities in motor vehicle searches have efficiency costs?" *The Quarterly Journal of Economics*, 137, 49–113.

Flanagan, F. X. (2018): "Race, gender, and juries: Evidence from North Carolina," *The Journal of Law and Economics*, 61, 189–214.

Glover, D. (2021): "Job search and intermediation under discrimination: Evidence from terrorist attacks in France," *Working paper*.

Godfrey, J., K. T. K. Tan, and M. Zapryanova (2022): "The Effect of Parole Board Racial Composition on Prisoner Outcomes," *Working paper*.

Goncalves, F. and S. Mello (2021): "A few bad apples? Racial bias in policing," *American Economic Review*, 111, 1406–41.

Gould, E. D. and E. F. Klor (2016): "The long-run effect of 9/11: Terrorism, backlash, and the assimilation of Muslim immigrants in the West," *The Economic Journal*, 126, 2064–2114.

Hamm, M. S. (2007): "Terrorist Recruitment in American Correctional Institutions: An Exploratory Study of Non-traditional Faith Groups; Final Report," Indiana State University, Department of Criminology, https://www.ojp.gov/pdffiles1/nij/grants/220957.pdf.

Kaufman, S. B. (2019): "The criminalization of Muslims in the United States, 2016," *Qualitative Sociology*, 42, 521–542.

Kaushal, N., R. Kaestner, and C. Reimers (2007): "Labor market effects of September 11th on Arab and Muslim residents of the United States," *Journal of Human Resources*, 42, 275–308.

Kusha, H. R. (2016): *Islam in American prisons: black Muslims' challenge to American penology*, Routledge.

LaForest, M. (2022): "Racial Bias, Gender Bias, and the Effects of Parole Officers on Reentry," *Working Paper*.

Lepage, L.-P. (2023): "Discrimination and sorting in the real estate market: Evidence from terrorist attacks and mosques," *European Economic Review*, 153, 104386.

Ludwig, J. and S. Mullainathan (2021): "Fragile algorithms and fallible decision-makers: lessons from the justice system," *Journal of Economic Perspectives*, 35, 71–96.

McConnell, B. and I. Rasul (2021): "Contagious Animosity in the Field: Evidence from the Federal Criminal Justice System," *Journal of Labor Economics*, 39, 739–785.

Mechoulan, S. and N. Sahuguet (2015): "Assessing racial disparities in parole release," *The Journal of Legal Studies*, 44, 39–74.

Oudekerk, B. and D. Kaeble (2019): "Probation and parole in the United States, 2019," *Washington, DC: US Department of Justice.*

Panagopoulos, C. (2006): "The Polls-Trends: Arab and Muslim Americans and Islam in the aftermath of 9/11," *Public Opinion Quarterly*, 70, 608–624.

Philippe, A. and A. Ouss (2018): ""No hatred or malice, fear or affection": Media and sentencing," *Journal of Political Economy*, 126, 2134–2178.

Rambachan, A. and J. Roth (2022): "A More Credible Approach to Parallel Trends," Tech. rep., Working Paper.

Renaud, J. (2019): "Grading the parole release systems of all 50 states," *Prison Policy Initiative*, https://www.prisonpolicy.org/reports/parole_grades_table.html.

Roth, J. (2022): "Pretest with Caution: Event-Study Estimates after Testing for Parallel Trends," *American Economic Review: Insights*, 4, 305–22.

Shayo, M. and A. Zussman (2011): "Judicial ingroup bias in the shadow of terrorism," *The Quarterly journal of economics*, 126, 1447–1484.

Sheridan, L. P. (2006): "Islamophobia pre–and post–September 11th, 2001," *Journal of interpersonal violence*, 21, 317–336.

Singh, A. (2002): *" We are Not the Enemy": Hate Crimes Against Arabs, Muslims, and Those Perceived to be Arab Or Muslim After September 11*, vol. 14, Human Rights Watch.

Sloan, C. (2019): "Racial bias by prosecutors: Evidence from random assignment," in *ICCJ 2019: International Conference on Criminal Justice June*, 25–26.

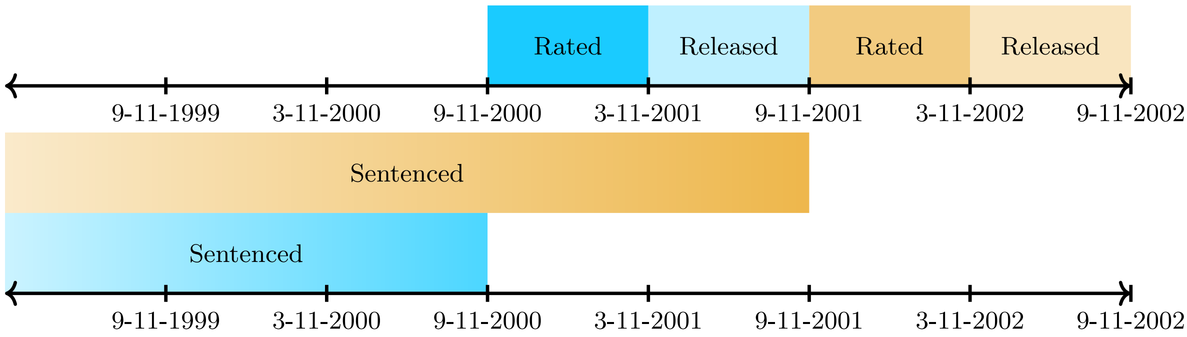
Woods, J. (2011): "The 9/11 effect: Toward a social science of the terrorist threat," *The Social Science Journal*, 48, 213–233.

Zapryanova, M. (2020): "The effects of time in prison and time on parole on recidivism," *The Journal of Law and Economics*, 63, 699–727.

Appendix

A Sample Selection Schematic

Figure A1: Sample Restrictions



B The Georgia Parole Board Process

Table B1: Parole Decision Guidelines (Grid)

	Recommended Prison Time (in months) by Risk Group		
Crime Severity Level	Low Risk (14-20)	Medium Risk (9-13)	High Risk (0-8)
1	10	16	22
2	12	18	24
3	14	20	26
4	16	22	28
5	34	40	52
6	52	62	78
7	72	84	102
8	65% prison sentence	75% prison sentence	90% prison sentence

Notes: This table shows the Parole Board Guidelines (the grid) used in Georgia during our sample period. The grid specifies the recommended prison time (in months) based on the crime severity level and risk scores (in parenthesis), grouped in three risk groups (low, medium, and high). Details on the calculation of the risk scores and the classification of the crime severity level can be found on the Georgia's Board of Pardons and Paroles website at pap.georgia.gov/parole-consideration/parole-consideration-eligibility-guidelines.

Table B2: Discretionary Parole Board Systems in US and Georgia

Parole System Characteristic	US				Georgia
	Mean	Std. dev.	Min	Max	
Has discretionary parole for new offenses	1	0	1	1	1
Would mandate face-to-face hearings	0.561	0.446	0	1	0
Would provide method to challenge incorrect information	0.273	0.452	0	1	0
Prohibits input from prosecutors	0.621	0.434	0	1	0.5
Prohibits input from crime survivors	0.727	0.282	0	1	1
Would allow input from applicant, family, community, employers, prison admin	0.545	0.289	0	1	0
Employs presumptive parole policies	0.197	0.248	0	0.5	0
Does not deny parole for subjective reasons	0.727	0.282	0	1	0.5
Would mandate yearly reviews	0.409	0.404	0	1	0
Would provide case managers to assist individuals	0.227	0.397	0	1	0
Would provide individuals with access to all records	0.333	0.389	0	1	0
Would incorporate parole guidelines	0.333	0.27	0	1	1
Would require parole board to file yearly report to an oversight committee	0.394	0.496	0	1	0
Would have meaningful appeal process	0.53	0.432	0	1	0
Prison Policy Initiative overall score	38	29	0	83	42
Number of members on the Parole Board	7	2.25	3	13	5

Notes: Data on the number of members on the Parole Board in each state was collected from each state's Parole Board website. We were not able to retrieve information on the size of the Parole Boards in Alaska and Maryland. All other data comes from Renaud (2019). Overall score is a weighted average of each of the characteristics calculated by Prison Policy Initiative. All other characteristics are graded on the scale 0-0.5-1, where 0 stands for no, 0.5 for partially, and 1 for yes. The states included in the US average are states that offer discretionary parole, namely, Alabama, Alaska, Arkansas, Colorado, Connecticut, Hawaii, Idaho, Iowa, Kentucky, Louisiana, Maryland, Massachusetts, Michigan, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New York, North Dakota, Oklahoma, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, West Virginia, and Wyoming.

C Identifying Assumptions

In this section, we present supportive evidence for (i.) the common trends assumption and (ii.) stability of group composition over time. These are the two core identifying assumptions of a repeat cross section difference-in-differences approach.

C.1 Parallel Trends

We provide three pieces of evidence in support of the parallel trends assumption inherent in our DD approach. Each piece of evidence approaches the topic of parallel trends from a different perspective. Each piece of evidence provides support that the parallel trends assumption holds for our empirical specification in the sample period under consideration.

C.1.1 Placebo DDs

We first implement a set of placebo DD regressions. We shift all key dates one year back in time, and re-estimate Equation 1, with the sole difference that now the $Post_t$ term takes value zero for the period 11 September 1999–10 September 2000, and one for the period 11 September 2000–10 September 2001. We present the results in Table C1. Given the absence of any significant placebo DD parameters, we consider the placebos as the first piece of evidence in support of the parallel trends assumption.

Table C1: Parole Board Decisions and Prisoner Outcomes

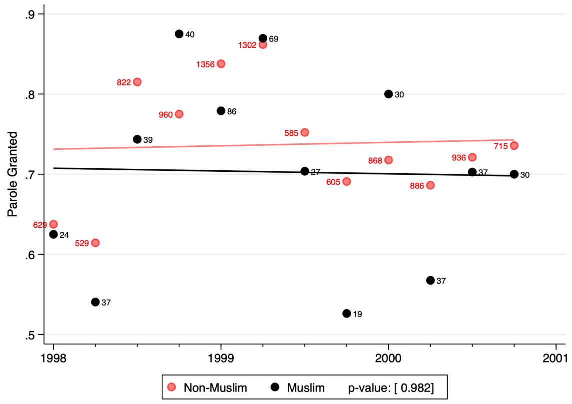
	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	.0416	49.1	–59.4
	(.0552)	(74.9)	(74.1)
$\overline{Y}_{0,PRE}$	.714	1011	882
(Post-9/11×Muslim) / $\overline{Y}_{0,PRE}$	.0583	.0486	-.0673
	(.0773)	(.0741)	(.084)
Adjusted R^2	.319	.873	.606
Number Muslim Inmates	192	192	192
Observations	5,031	5,031	5,031

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2000 is used for estimation.

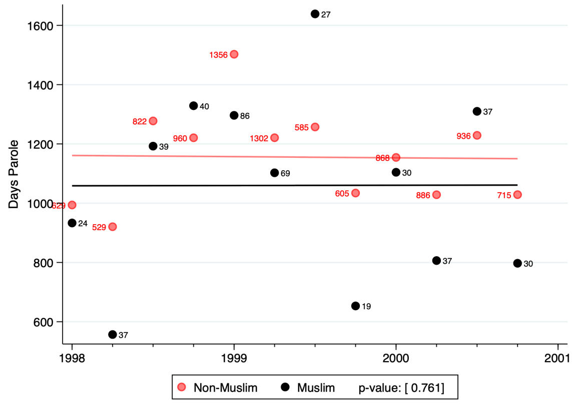
C.1.2 Trends in the Raw Data

Next, in Figure C1, we provide graphical evidence of the existence of pre-trends by presenting the raw, underlying data for three years prior to our estimation sample – the calendar years of 1998-2000. We present the p-values from a test of equality of trends. We cannot reject this null of equality of trends in any case – the smallest p-value is .76.

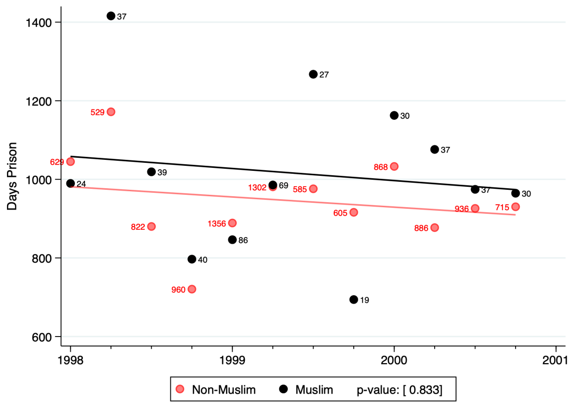
Figure C1: Raw Pre-Trends



(a) Parole



(b) Days Parole



(c) Days Prison

Notes: The p-value presented in the legend of each graph is based on a test of equality of trends between Muslim and non-Muslim inmates at the individual level using pooled data, with Eicker-Huber-White standard errors.

C.1.3 Honest Difference-in-Differences

Finally, we implement the honest difference-in-differences approach of Rambachan and Roth (2022), in order to create worst-case treatment effect bounds for potential violations of the parallel trends assumption, based on pre-trends.

In order to operationalize this approach, we use data on those who come before the parole board between September 11, 1999 and September 10, 2002 , and create 3 periods: 1. An initial period of those up for parole between September 11, 1999 and September 10, 2000 – the year prior to the pre-period used in the main analysis, 2. the pre-period of those inmates reviewed between September 11, 2000 and September 10, 2001 and 3. the post-period of September 11, 2001 and September 10, 2002. We then implement a continuous treatment and binary treatment version of our core DD model, but based on the extended data and a 3 period approach, as follows:

$$y_{it} = \alpha_0 Muslim_i + \sum_{j=1,\neq 2}^{3} \alpha_j Period_j + \sum_{j=1,\neq 2}^{3} (\beta_j Period_j \times Muslim_i) + X_i^{'}\gamma + \pi_m + \epsilon_{it} \ , \quad (3)$$

The coefficients presented in Table C2 below, and accompanying variance-covariance matrices are the required inputs into the R package (HonestDiD) that implements the Rambachan and Roth (2022) approach.

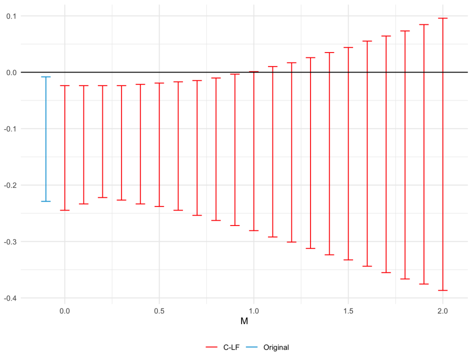
The graphical outputs from the Rambachan and Roth (2022) approach, where we use the Relative Magnitude approach for bounding, are presented in Figure C2. For all three outcomes, the "breakdown value" of $\overline{M}$ – the factor of the pre-trends at which the bounds on the estimated treatment effect overlap with zero – exceeds 1. This means that even if post 9/11 violations of parallel trends were as large as any pre-period violations, the confidence set for the treatment effects would not include zero.

Table C2: The Inputs For the Honest DD Approach Highlight The Large Ratio Between Placebo and Actual Treatment Effects From a Pooled Estimation

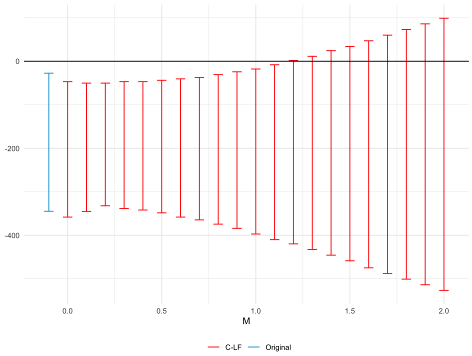
	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
$Period_1 \times Muslim$	-.0381	–43.2	50.5
	(.0563)	(75.7)	(74.9)
$Period_3 \times Muslim$	-.118	–186	199
	(.0563)	(81)	(80.7)
$\overline{Y}_{0,PRE}$	.71	960	868
Adjusted R^2	.295	.855	.577
Observations	7,460	7,460	7,460

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+.

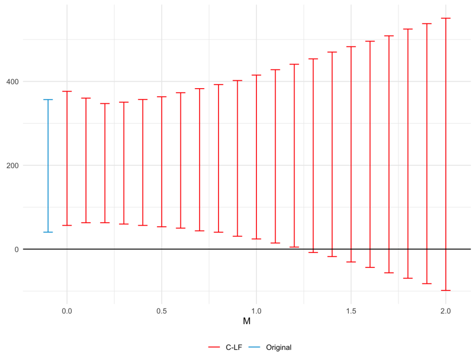
Figure C2: Honest Difference-in-Differences



(a) Parole



(b) Days Parole



(c) Days Prison

Notes: The blue band ("Original") is the 90% confidence interval of the DD treatment effect estimate ($Period_3 \times Muslim$ in Table C2). The red bands ("C-LF") are the robust 90% confidence intervals for the Rambachan and Roth (2022) Relative Magnitude-based bounds. These vary with the x-axis – $\overline{M}$ – which designates factors of the maximum pre-treatment violation of parallel trends. Thus a confidence interval that does not intersect 0 when $\overline{M} = 1$ informs us that when we allow any parallel trend violations in the post-period to be as large as the maximum pre-treatment violation, the 90% confidence intervals for the bounded treatment effect do not include zero.

C.2 Stable Group Composition

Given that we are using repeat cross-sectional data for our empirical analysis, we also provide evidence for a second identifying assumption – that the composition of the two groups are stable across the pre and post periods (Blundell and Dias, 2009). We do so in two ways.

C.2.1 Balance

In Table C3 we first present the results of a series of balance tests. Column (3) and Column(6) show the p-values for a null of no difference in means across the two periods, for non-Muslim and Muslim inmates respectively. We cannot reject the null in 20 out of 22 cases. Non-Muslim inmates have slightly more prior convictions post-9/11 and Muslim inmates are less likely to be married in the post-9/11 sample. Column (7) presents the p-value of the difference-in-differences across the control variables. These p-values never fall below .05. We interpret these results as supportive of the assumption of group composition stability.

Table C3: Balance Tests

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Non-Muslim			**Muslim**			
	Pre-9/11	Post-9/11	p-value: Difference	Pre-9/11	Post-9/11	p-value: Difference	p-value: DD
Sample Size	2293	2316		112	110		
Education:							
$\leq$ High School	.649	.674	[.077]	.688	.591	[.135]	[.065]
High School	.249	.235	[.278]	.196	.245	[.381]	[.272]
Some College	.0868	.0777	[.263]	.0982	.155	[.209]	[.148]
College	.0153	.0134	[.592]	.0179	.00909	[.573]	[.664]
I.Q. Score	94	95.2	[.066]	96	96.1	[.984]	[.693]
	(22.7)	(19.8)		(18.9)	(21.6)		
Has Children	.674	.67	[.791]	.705	.618	[.171]	[.197]
Married	.111	.103	[.379]	.179	.0818	[.032]	[.052]
Prior Convictions	1.35	1.49	[.000]	1.63	1.62	[.970]	[.436]
	(1.28)	(1.31)		(1.41)	(1.33)		
Age at Sentencing	30.9	30.8	[.734]	29.5	28.8	[.559]	[.628]
	(9.59)	(9.59)		(8.77)	(7.65)		
Severity Level	2.87	2.8	[.139]	2.94	2.7	[.274]	[.454]
	(1.69)	(1.59)		(1.61)	(1.62)		
Sentence Length	1977	2054	[.111]	2052	2049	[.989]	[.718]
	(1512)	(1759)		(1524)	(1697)		
Joint Test			[.001]			[.273]	[.240]

Notes: Means and standard deviations (in parentheses for continuous covariates) are shown. p-values are based on OLS regressions with Eicker-Huber-White standard errors.

C.2.2 Duration Analysis

Secondly, we implement a series of duration model regressions, where the key duration variable is the time from prison admission to rate date. We present these in Table C4. The point of this analysis is to ensure there is no strategic reordering of when Muslim and non-Muslim inmates appear before the parole board in the aftermath of 9/11. There is no evidence that this is the case, either in the raw durations or once we condition on our main set of control variables.

Table C4: Duration Analysis – (Rate Date - Admission Date)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Unconditional				**Full Covariate Set**			
	OLS	Cox	Gompertz	Weibull	OLS	Cox	Gompertz	Weibull
Post-9/11×Muslim	.0913	-.134	-.146	-.0119	.147	-.173	-.221	-.251
	(.142)	(.134)	(.153)	(.233)	(.127)	(.14)	(.135)	(.222)
$\overline{Y}_{0,PRE}$ (days)	304	304	304	304	304	304	304	304
Muslim Inmates	222	222	222	222	222	222	222	222
Observations	4,826	4,826	4,826	4,826	4,826	4,826	4,826	4,826

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses, regular p-values in brackets. The outcome variable in all cases is the duration from prison admission date until rate date – the start of the parole process. The exceptions to this are columns 1 and 5, where the outcome variable is the nautral log of duration. Logs were taken to deal with the extreme right skew of the data. For these proportional hazrd models we present coefficients and not hazard rates. For the Cox proportional hazard model, the Gompertz and the Weibull based models, a negative coefficient means a lower hazard rate, and thus a longer duration. For the Log Logistic and Weibull based models, we specify gamma frailty. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

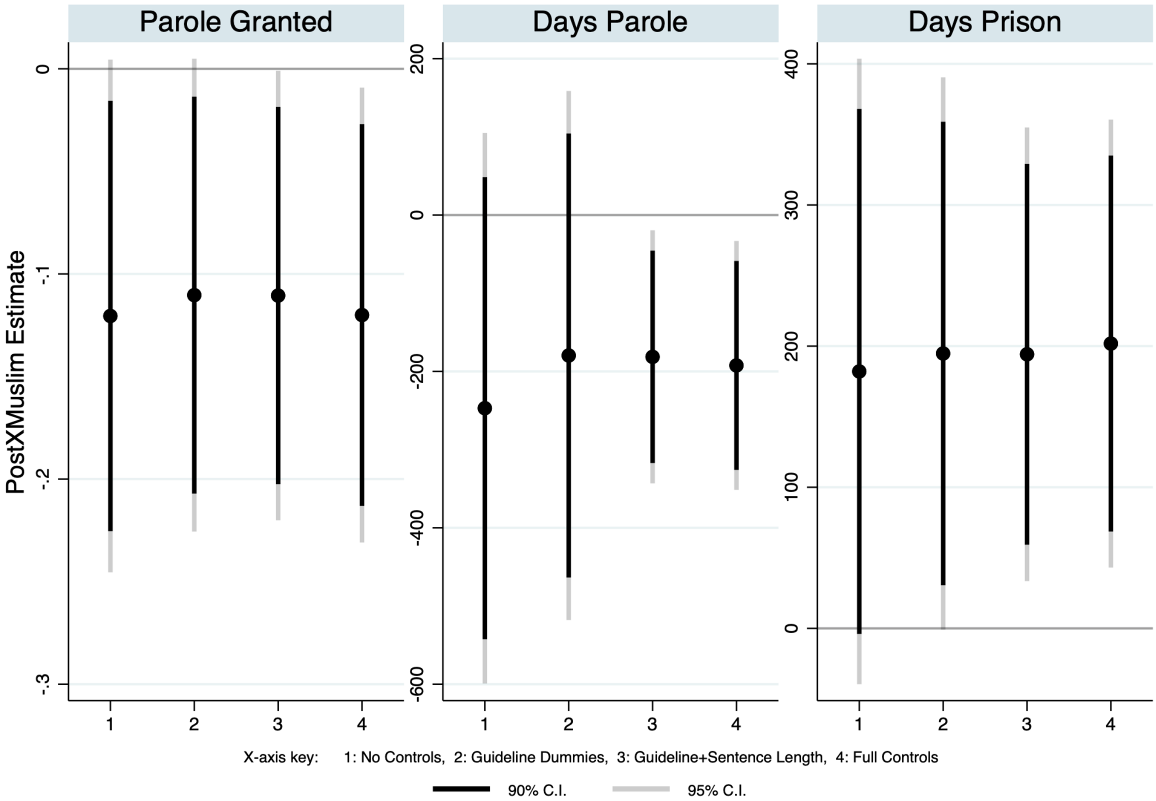
D Robustness and Ancillary Results

In order to probe our results, we conduct several sets of sensitivity analyses. In Figure D1 we explore the sensitivity of our results to the inclusion of key classes of control variables. We start with the unconditional DD estimates, include parole grid dummies, then sentence length, and finally all additional control variables. The DD estimates are stable to the inclusion of the controls. In Figure D2 we consider the sensitivity of our estimates to the width of the exclusion window that we specify, in order to ensure those individuals rated pre-9/11 are seen by the parole board pre-9/11. The wider the exclusion window, the smaller the sample size. The first panel of Figure D2 quantifies this intuition. The results are broadly stable until we force the exclusion window to be 8 months or wider, due primarily to the loss of sample size.

D.1 The Impact of Controls on the DD Estimates

In Figure D1, we progressively include more covariates starting from specification 1 (no control variables) and finishing with specification 4 (full set of controls, and our baseline specification). The coefficients are extremely stable across specifications.

Figure D1: Robustness of the DD Estimates to the Sequential Inclusion of Controls

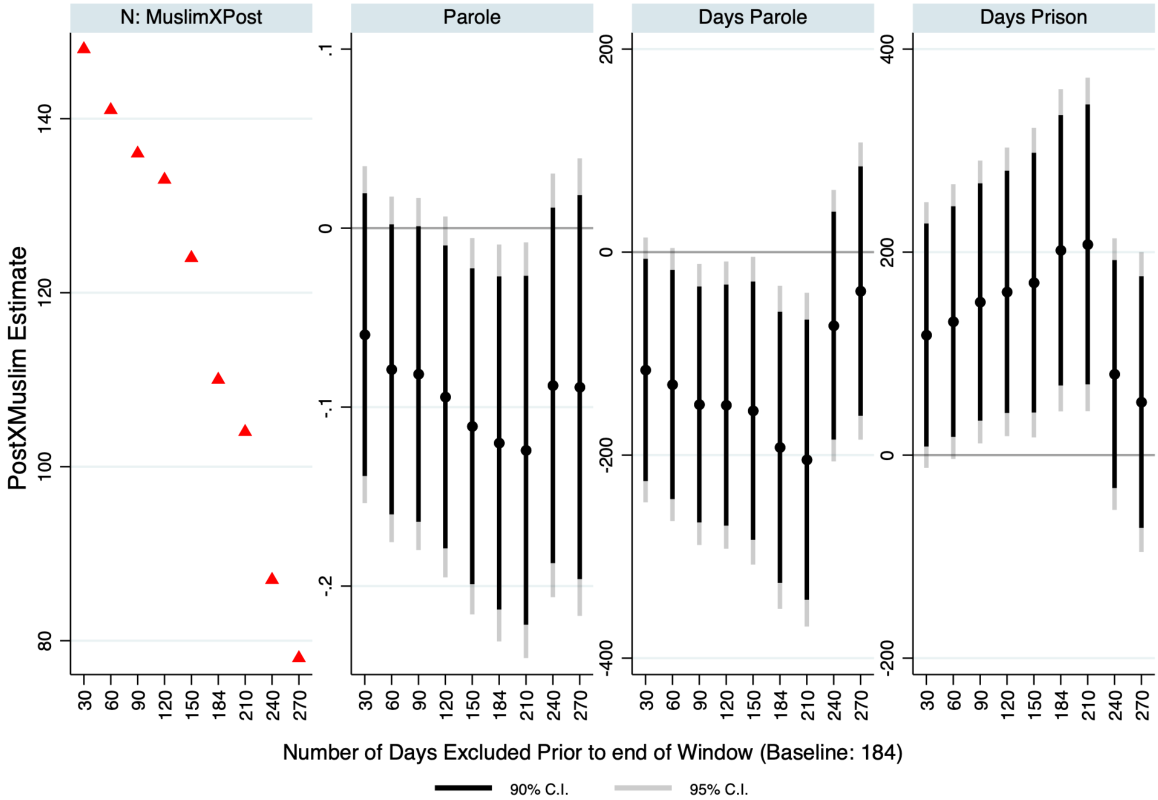


Notes: We present four, numbered specifications. Specification 1 includes no other controls, and thus yields unconditional DD estimates. Specification 2 includes a set of Parole Decision Guideline cell dummies. Specification 3 additionally includes as a control sentence length in days. Specification 4, our baseline specification includes sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

D.2 Exclusion Window Sensitivity and the DD Estimates

We focus on cases that have a rate date within a +/-1 year window around 9/11. We enforce a buffer/exclusion window towards the end of the window, as the parole board is unlikely to see a case on the rate date. Our baseline is to specify a 6 month exclusion window, the cost of which is to effectively halve our sample size. The benefit is that we can be fairly certain that all cases with rate dates in the pre-period are indeed seen by the parole board in the pre-period. For cases that are seen after, we will misallocate post cases as pre cases. The consequence of this will be to attenuate our treatment effect. The estimates show precisely this – with too short an exclusion window we attentuate our treatment effect – many of the cases with rate dates close to, but before, 9/11/01 are likely reviewed by the majority of the parole board in the post-period. Parameter estimates are broadly stable and monotonically decrease for the parole outcomes and increase for the prison outcome. This pattern breaks when we exclude too large a proportion of our sample – Excluding more than 240 of the possible 365 days reduces the sample size sufficiently that parameter estimates become erratic.

Figure D2: Robustness of the DD Estimates to Different Lengths of the Exclusion Window



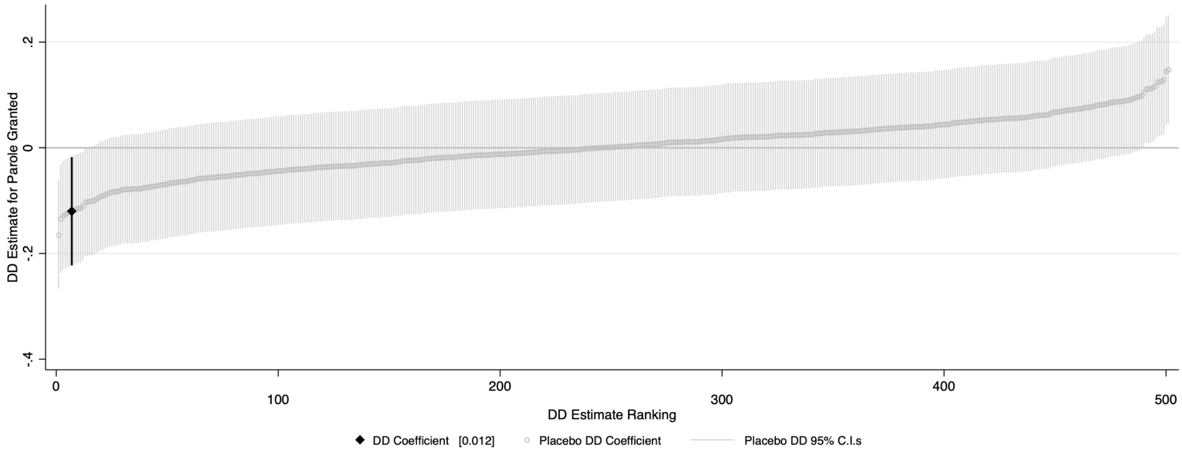
Notes: Panel A presents the number of Muslim inmates in the post-9/11 period for each exclusion window. Panels B, C and D display the point estimate, 95% and 90% confidence intervals for Post-9/11×Muslim based on the exclusion restriction range shown on the x-axis. 184 days is our baseline choice. The DD regression specifications that yield parameter estimates in Panels B, C and D include the following control variables: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

D.3 Placebo Treatment Results

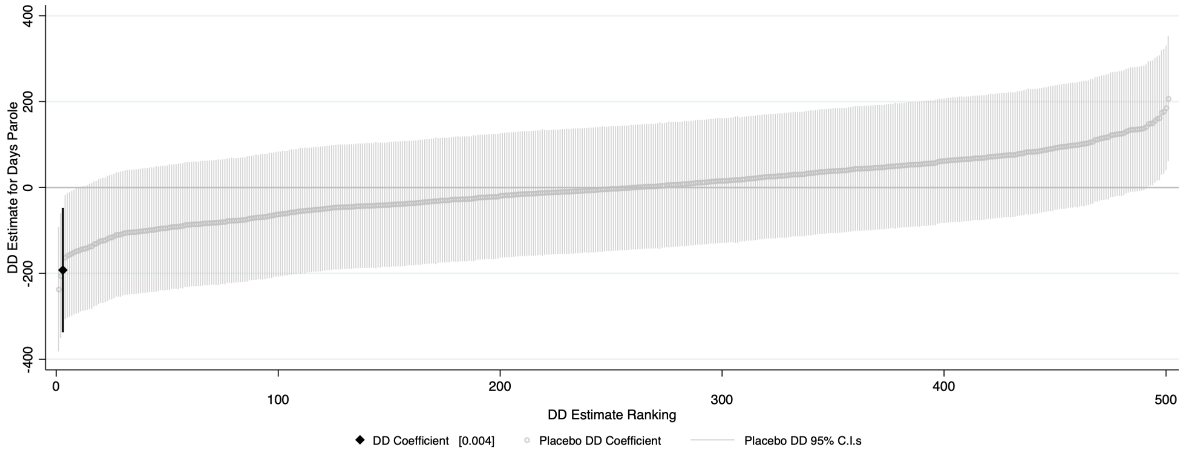
In order to gauge the magnitude of our findings, we conduct a placebo experiment involving random assignment of treatment status (Cai and Wang, 2022). We run 500 placebo experiments, where we randomly assign Muslim status across inmates, and then conduct our baseline DD analysis. Random assignment of Muslim status is conducted to reflect the proportion of the sample who are Muslim and non-Muslim. It is worth noting that this exercise follows the same approach as one would use to conduct randomization inference (RI) in this context, so we display the RI p-value in the legend of each graph.

The graphs below are striking, and make clear the (statistical) significance of our findings.

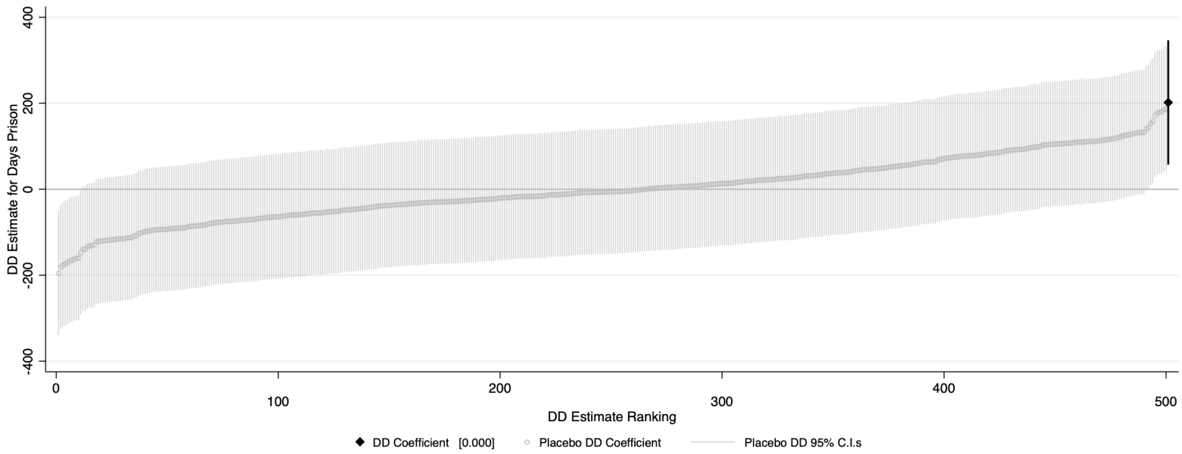
Figure D3: Placebo Treatment Results



(a) Parole Granted



(b) Days Parole



(c) Days Prison

Notes: Each graph presents the coefficient estimates and 95% confidence intervals from 501 regressions – one regression based on true Muslim status, and 500 placebo regressions where Muslim status is randomly assigned across inmates. Random assignment of Muslim status is conducted to reflect the proportion of the sample who are Muslim and non-Muslim. Given that the placebo regressions form the basis of how one would conduct randomization inference, we display randomization inference *p*-value in brackets at the bottom of each graphs. This is one way to reflect the finding that for our 3 outcomes (Parole Granted, Days Parole, Days Prison) only 6,1 and 6 of the 500 placebo estimates respectively exceed our true DD estimate.

D.4 Short Run Impacts – Treatment Effect Heterogeneity

Table D1: Parole Board Decisions and Prisoner Outcomes by Grid Score Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Low Risk			**Medium Risk**			**High Risk**		
	Parole Granted	Days Parole	Days Prison	Parole Granted	Days Parole	Days Prison	Parole Granted	Days Parole	Days Prison
(Post-9/11×	-.162	–103	145	-.0626	–165	162	- .177**	–192*	206*
Muslim)	(.154)	(234)	(243)	(.0788)	(123)	(120)	(.0882)	(114)	(113)
$\overline{Y}_{0,PRE}$	.741	973	899	.742	978	820	.655	935	912
(Post-9/11×	-.219	-.106	.161	-.0843	-.169	.197	- .271**	-.206*	.226*
Muslim) / $\overline{Y}_{0,PRE}$	(.207)	(.24)	(.27)	(.106)	(.126)	(.146)	(.135)	(.122)	(.124)
Adjusted R^2	.384	.775	.601	.291	.879	.528	.254	.82	.547
Number Muslim Inmates	26	26	26	89	89	89	107	107	107
Observations	844	844	844	2,056	2,056	2,056	1,932	1,932	1,932

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

D.5 Disciplinary Outcomes

In order to interpret our results as the causal effect of 9/11 on parole board decision making, it is important for us to consider margins along which other actors may influence parole outcomes. Due to our empirical design, we rule out the role of sentences, by focusing only on cases where sentencing occurs prior to 9/11. But what if inmates change their behavior post-9/11, or prison guards apply disciplinary rules in a disparate manner post-9/11? If this were the case, then such behavioral changes would contribute to our estimated treatment effects.

We implement our main DD strategy using data on infractions that resulted in both reports (incident was recorded, but did not result in a charge) and charges, and present the results in Table D2. We find no evidence of any statistically significant changes in disciplinary outcomes for Muslim inmates in the post-9/11 period, thus suggesting that disciplinary infractions whilst incarcerated could not be a key mediator for our main results.

Table D2: Disciplinary Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Disciplinary Count				**Any Disciplinary (Binarized)**			
	Reports	**Charges**			**Reports**	**Charges**		
	Total	Violent	Non-Violent	Total	Total	Violent	Non-Violent	Total
Post-9/11×Muslim	1.21	.3	1.19	1.5	.0339	.0361	.0157	.034
	(.973)	(.238)	(1.26)	(1.44)	(.0555)	(.0625)	(.0566)	(.0555)
$\overline{Y}_{0,PRE}$	3.4	.441	4.23	4.69	.629	.215	.61	.63
(Post-9/11×Muslim) / $\overline{Y}_{0,PRE}$	.354	.679	.282	.319	.0539	.168	.0257	.054
	(.286)	(.539)	(.298)	(.307)	(.0883)	(.291)	(.0928)	(.0881)
Adjusted R^2	.21	.109	.194	.191	.156	.099	.154	.157
Number Muslim Inmates	222	222	222	222	222	222	222	222
Observations	4,832	4,832	4,832	4,832	4,832	4,832	4,832	4,832

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The following controls are included in all regressions: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.

D.6 Sub-Group Ex-Ante Recidivism Risk

Table D3: Sub-Group Recidivism Risk

	(1)	(2)
	Recidivism Risk	Sub-Group Size
A.) Full Sample	0.122	15075
B.) Offense Severity		
0	0.147	10848
1	0.056	4227
C.) Predicted Recidivism Quartile		
1	0.036	3700
2	0.088	3699
3	0.137	3700
4	0.230	3699

Notes: Predicted recidivism is based on 1 year recidivism probabilities for sample released 9/11/1998-9/10/2000

D.7 Extrapolating Sub-Group DD estimates to Zero Recidivism Risk

As an extension to our analysis in Section ??, we speculate that parole boards are less likely to inaccurately stereotype Muslims with the zero or very low recidivism risk post 9/11. In that case, the intercepts of the slopes in Figure 4 indicate any bias generated by non-risk factors, which continues to impact Muslim inmates with zero recidivism risk.

Table D4: Extrapolating Sub-Group DD estimates to Zero Recidivism Risk

	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Baseline DD Estimate	-.12	–192	202
Quartiles of Recidivism Risk			
Zero Recidivism Risk (ZRR) Projection	-.00172	78.6	–7.28
	(.115)	(149)	(147)
ZRR as Percentage of Baseline	1.43%	–40.9%	–3.61%

Notes: Predicted recidivism is based on 1 year recidivism probabilities for sample released 1/1/1999-9/10/2000, based on a regression specification with the following controls: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. The linear extrapolation is based on OLS, where we inverse-weight by the variance of the quartile-specific DD estimate. Standard errors based on a bootstrap procedure, with 10000 iterations, are in parentheses.

We present the results of this exercise in Table D4, where we show estimates of the zero recidivism risk DD estimate. We bootstrap the standard errors to account for the fact that both dimensions of the graph are based on estimates – the one-year recidivism likelihood on the x-axis, and the DDD estimates on the y-axis. If we are willing to extrapolate from our estimates to the point at which predicted recidivism risk is zero (ZRR), we find that the ZRR projection for parole grants (Table D4, Column (1) is essentially zero. For parole supervision length we find a ZRR that is far smaller than our baseline DD estimate, and for prison days our ZRR projection is negative.

D.8 Differential Conversions to Islam as a Threat to Identification

Recent research suggests that prison conversions are common among Muslim inmates (Boddie and Funk, 2012; Hamm, 2007; Kusha, 2016), which our measure for religion would not capture. This could be a potential threat to our identification strategy, as our results could be driven by differential rates of in-prison conversion to Islam in the post-9/11. If fewer conversions occur post-9/11, and Muslim converts in our control group pre-9/11 were discriminated against, then parole outcomes may improve in the post-9/11 control group where there may have been fewer converts, negatively biasing our DD estimates for parole outcomes.

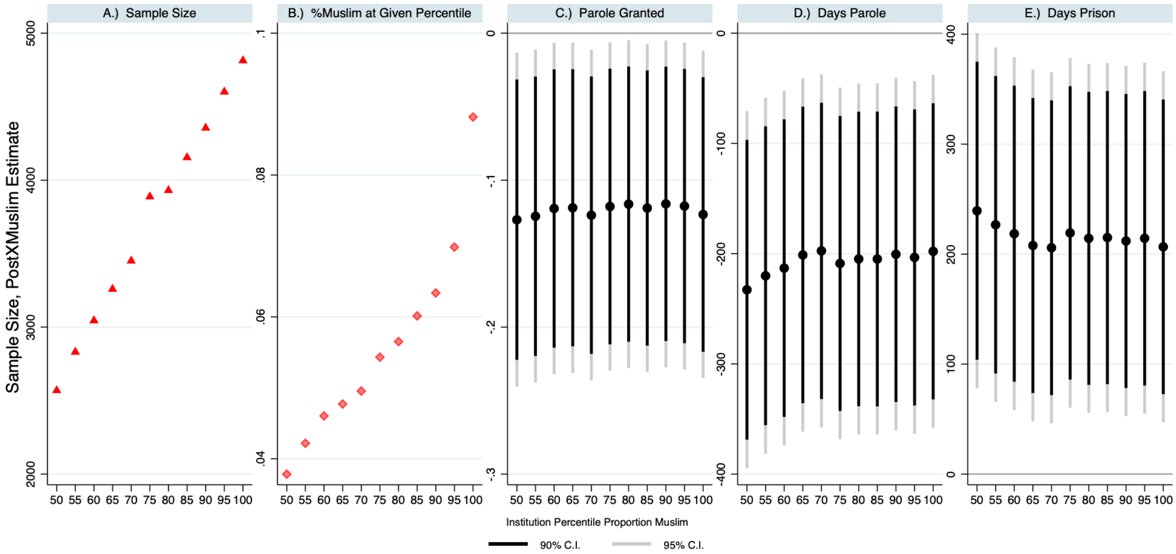
If in the wake of 9/11, some would-be converters to Islam do not convert, then we will have a differential composition of converters in our non-Muslim-at-entry control group. Recall, we observe religious affiliation at the point of entry into prison, but we do not observe in-prison conversion.

In order to address this data shortcoming we take the following approach. First, we assume that if an inmate who is non-Muslim at entry decides to convert to Islam, the inmate is more likely to do so at a high Muslim-at-entry concentration prison than a low concentration prison. This assumption is based on several considerations – a higher Muslim concentration prison will both (i) provide more opportunities for social interactions that spur a conversion and (ii) provide a more supportive network for practice of the religion (Hamm, 2007).

Accordingly, we construct the proportion of Muslim inmates at each prison facility from the stock of inmates in the 9/11 period[16] and then create (inmate-weighted) percentiles of the proportion of Muslim inmates. We create these percentiles for the non-Muslim-at-entry control group. We next run a series of regressions for our three main outcome variables, sequentially removing non-Muslim-at-entry inmates in the top p% of prison facilities by proportion of Muslim inmates, reporting the results of this sensitivity exercise in Figure D4 below. The first panel of the figure displays the sample size for the regression, once we remove non-Muslim-at-entry inmates from the top p% of prisons. The lower the percentile, the more we reduce the sample. The second panel displays the average proportion of % Muslim at the given percentile. The remaining three panels display DD estimates and respective confidence intervals for our three main outcomes. We can remove non-Muslim-at-entry inmates at the top 50% of Muslim concentration prisons, reducing our working sample size by half, and see no statistically or economically meaningful change in our parameter estimates. We take these results as strong evidence that differential conversion to Islam in the post-9/11 period is not driving our core findings.

[16]The mean proportion of Muslim inmates for our core sample is .042. The maximum is .103.

Figure D4: DD Estimates by Muslim Prison Composition



Notes: Panel A presents the total sample size when using data less than percentile p. Panel B presents the value of %Muslim inmates in the institution for the given percentile p. Panels C, D and E display the point estimate, 95% and 90% confidence intervals for Post-9/11×Muslim when we exclude the top p percentiles of institutions by %Muslim inmates. Thus, for example, the estimates corresponding to percentile 80 are based on our core estimation sample where prisons with the top 20% of proportion Muslim inmates have been removed. The DD regression specifications that yield parameter estimates in Panels C, D and E include the following control variables: sentence length in days, dummies for Parole Decision Guideline cell, rate date month, education level, IQ quintiles, having children, being married, age at sentencing deciles, major offense categories, social class groupings and number of prior convictions, where the maximum category is 8+. A +/-365 day window around 9/11/2001 is used for estimation.