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

Brendon McConnell* Kegon Teng Kok Tan† Mariyana Zapryanova‡

February 27, 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 that 9/11 led to large 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 after 9/11 with similar sized magnitudes. We also examine heterogeneity in the effects by recidivism risk and find 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 for useful comments and suggestions from seminar participants at Smith College Economics Brown Bag Seminar and the University of Wisconsin-Madison Empirical Micro Seminar, as well as conference participants at the 2022 Southern Economics Association Conference. We thank Steven Durlauf, Robert Kaestner, Stéphane Mechoulan, Mike Mueller-Smith, Imran Rasul, and Lucie Schmidt for helpful feedback on an earlier draft of the paper. 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 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. Examples include the 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 thus 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. In doing so, 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 focus on Black, male, parole-eligible inmates as the overwhelming majority of Muslim inmates were 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 9/11. At the extensive margin, we find a 12 percentage point (17%) reduction

[1]Religion is reported during the process of being admitted into the Georgia State Correctional System that begins once the convicted felons are transferred from the court to a diagnostic and classification prison. 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.

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 highly statistically significant. We also show that the large short-run impacts we document persist to the end of our sample period after 9/11, in contrast to other outcomes for Muslims that seem to exhibit large short term effects and substantial fade out (for 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 then shift our attention to a key post-parole outcome – recidivism. We consider two different aspects of recidivism in relation to outcomes for Muslims post-9/11. First, we replicate our event study analysis for one-year recidivism. We find no evidence of any changes in recidivism rates, in either direction, for Muslim releases. Such results are in stark contrast to the highly persistent changes in parole board behavior in the post-9/11 era. In our view, this contrast suggests that parole decisions were adjusted to mitigate potential increases in Muslim recidivism. If, instead, the parole decisions were born out of pure taste-based discrimination, penalizing well-behaved inmates who would have otherwise been released, we would expect sizeable declines in the recidivism rates of Muslims. It is worth noting, however, that the estimates are noisy.

Second, we construct a measure of ex-ante recidivism risk and conduct 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 recidivsm risk, the change in outcomes in the post-9/11 period is muted, whereas for those with high ex-ante recidivism risk, we find larger (negative) effects post-9/11. We interpret this as prima facie evidence that our short-run estimates are not driven by a blanket taste-based discrimination mechanism, but rather are sensitive to underlying recidivism risks 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; Danziger et al., 2011; Eren and Mocan, 2018; Light et al., 2019; 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

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

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 is 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 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). Examples of other economic variables that the literature has found to be affected by terrorism include labor market outcomes (Cornelissen and Jirjahn, 2012; Davila and Mora, 2005; Kaushal et al., 2007), housing market (Nowak and Sayago-Gomez, 2018), macroeconomy (Abadie and Gardeazabal, 2003; Blomberg et al., 2004), ethnic attitudes (Ratcliffe and von Hinke Kessler Scholder, 2015), self-identification (Mason and Matella, 2014), perception of risk (Abadie and Dermisi, 2008), assimilation (Bisin et al., 2008; Elsayed and De Grip, 2018; Gould and Klor, 2016), infant health (Bravo and Castello, 2021; Quintana-Domeque and Ródenas-Serrano, 2017), and mental health (Kim and Albert Kim, 2018).

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 (Weisburst, 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.

[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).

2 Parole in Georgia and Data

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.[4] The Georgia parole board consists of five-members appointed by the governor to seven-year terms subject to confirmation by the State Senate.[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 Parole Decisions Guidelines Grid System (hereafter the grid).[6]

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. The parole process in Georgia starts 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 the grid along with the prisoner's success 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 hearing examiner then prepares prisoner's parole file and writes a summary discussing its contents. When the parole file reaches the members of the parole board, it contains all records from the diagnostic prison, the personal history, social, and legal reports from the pre-parole investigation, the grid success score, severity level and recommendation, and the rater's summary.[7]

2.1 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

[4]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 B. 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.

[5]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).

[6]Refer to Table B1 for more details about the grid.

[7]The parole file also contains disciplinary reports for misconduct in prison, a list of all programs the person completed while in prison, any relevance correspondence from the prisoner, family members, or the victim and any statements by the judge, the district attorney, witnesses, police officers, or victims and community members.

prisoner's self-reported religion, and 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. 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 TPM 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 for 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.

3 Empirical Approach and Identification

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

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

where y_{it} is the parole board outcome of interest, $Post_t$ is an indicator that takes the value 1 for those reviewed by the parole board in the year following 9/11, and 0 for those reviewed in the year prior to 9/11, 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 prisoner intake from the court into the Georgia State Correctional System, thus ruling out endogenous recording of treatment status as a function of 9/11.

[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.

We condition on a rich set of covariates, the most important of which are a series of dummy variables for each of the 21 guideline cells used as part of the determination of parole in Georgia.[10] These guidelines combine a prisoner's probability to succeed to parole 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_m, 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 selfsame 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 (Forthcoming), 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 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 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. 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

[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.

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.

4 Results

We present our main results in Table 1.

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
Adjusted R^2	.283	.84	.549
Observations	4,832	4,832	4,832

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. A +/-365 day window around 9/11/2001 is used for estimation.

The DD coefficients we estimate are large, statistically significant, and economically meaningful. The evidence we present in Table 1 suggests that the outcomes of Muslim inmates in Georgia prisons 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 of 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 out of prison on parole. Finally, we present results for days in prison. In line with the parole results, we find Muslim inmates spend over half a year longer in prison if their file is reviewed by the parole board post-9/11, a 23% increase compared to the reference sub-sample of non-Muslim inmates with parole case 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).

4.1 Treatment Effect Heterogeneity

We next assess the extent to which our main DD estimates mask treatment effect heterogeneity. To do so, we split the sample into low and high severity offenses[12] and re-estimate our baseline specification on the two sub-samples. We present our findings in Table D1 and we find that the large decline in parole grants we document in Table 1 is driven by Muslim inmates convicted of low severity offenses. Muslim inmates with low severity offense who are reviewed for parole post-9/11 are 16 percentage points, or 20%, less likely to be granted parole. This translates into an average prison sentence that is 26% longer. For the sub-group of inmates convicted of high severity offenses, our DD estimate for parole is considerably smaller, and statistically indistinguishable from zero. The extensive margin effects of 9/11 seem to be concentrated among the low severity inmates, while intensive margin effects seem more similar across inmate types. This perhaps is surprising given Kuziemko (2013)'s findings that the parole board in Georgia exerts more discretion for high-severity offenses.

4.2 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:

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

[12]Following Kuziemko (2013) we choose the low severity offenses as offense levels 1-4, and high severity offenses as 5 and above.

[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.

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 we document contrasts with studies that have documented short-term impacts of 9/11 on hate crimes against Muslims (Gould and Klor, 2016) as well as on labor market outcomes (Kaushal et al., 2007).

5 Potential Mechanisms

The size and duration of our estimates begs the question: what drives these results? As a starting point, we performed similar analyses on disciplinary outcomes in prison that would be inputs to the parole process but found no statistically meaningful changes post 9/11 (see Table D3 in the appendix). It thus seemed unlikely that in-prison behavioral changes drove our findings. Delving deeper, we investigated a parole-board driven mechanism, namely discriminatory decisions against Muslim inmates post 9/11.

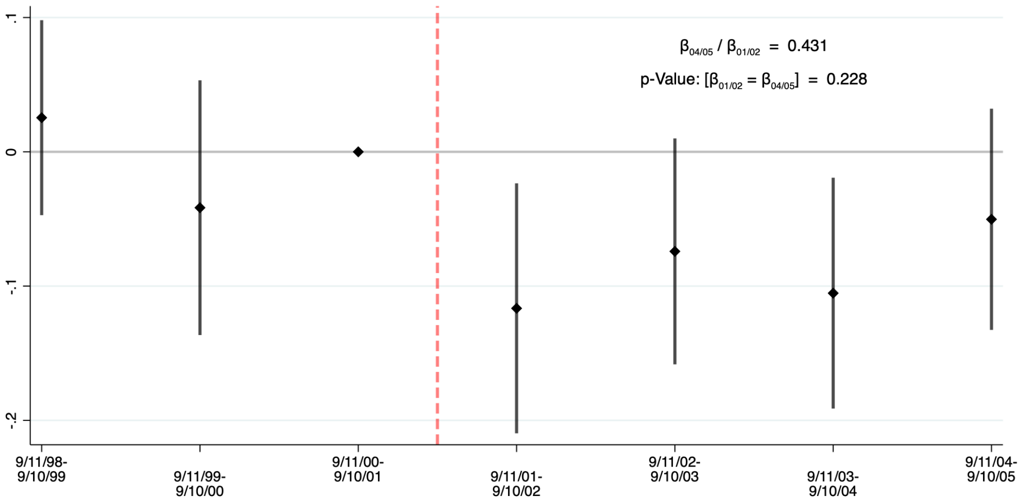
5.1 Do Muslim Releasees Change Recidivism Patterns Post-9/11?

In order to better understand our parole board findings, we considered recidivism, from two different, but related, aspects.[14] We first study how recidivism patterns evolve in the post-9/11 period for Muslim relative to non-Muslim former inmates. Second, we examine the extent to which the parole board decisions we document in Section 4 vary with ex-ante recidivism risk.

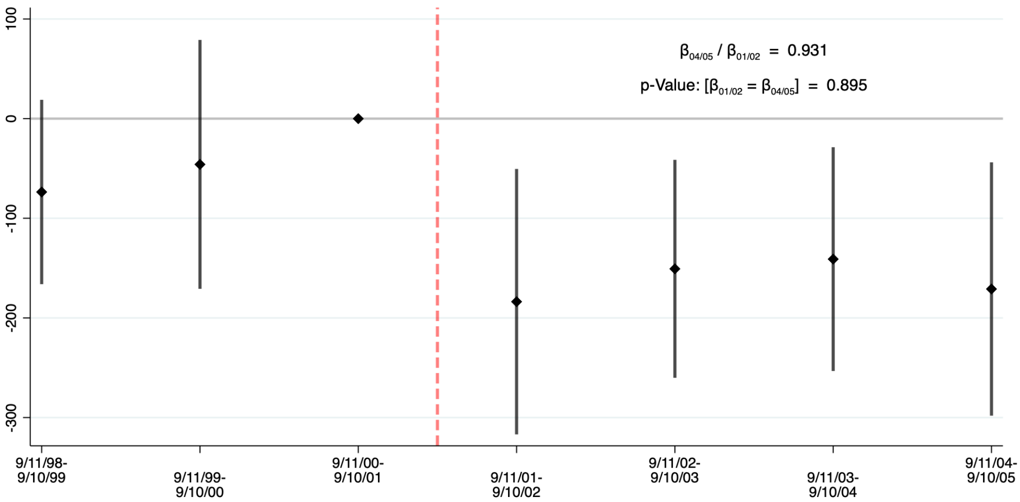
There are several reasons why recidivism patterns may have changed for Muslim releasees post-9/11, many of which relate to the canonical model of crime (Becker, 1968). Examples of dimensions which may have changed in differential ways for Muslim individuals include labor market opportunities, policing intensity, the geographical deployment of police, as well as changes in offending rates. In order to investigate if recidivism patterns for Muslim releasees did indeed change post-9/11, we estimate a dynamic DD specification akin to Equation (2), but

[14] We measure recidivism by return to prison in Georgia with or without a new sentence withing a year of prison release.

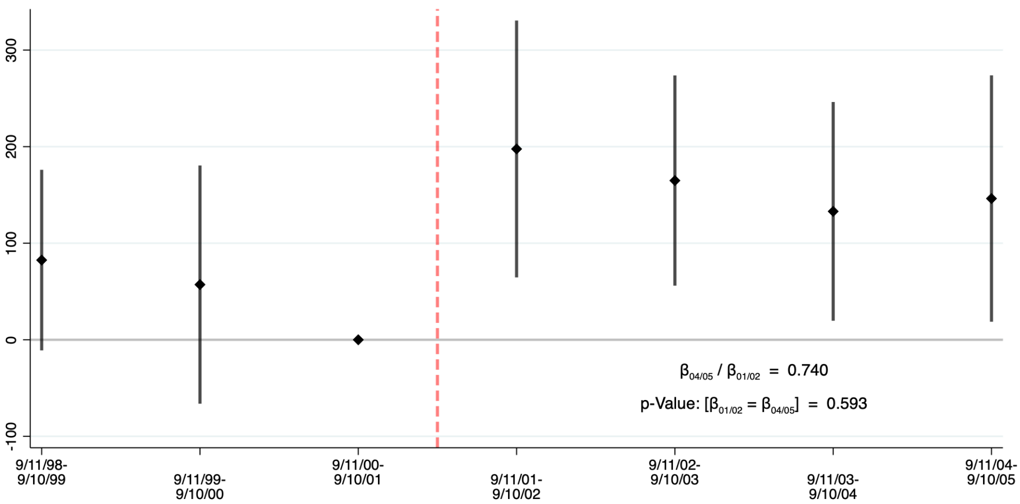
Figure 1: Dynamic Effects for Parole Outcomes



(a) Parole Granted



(b) Days Parole

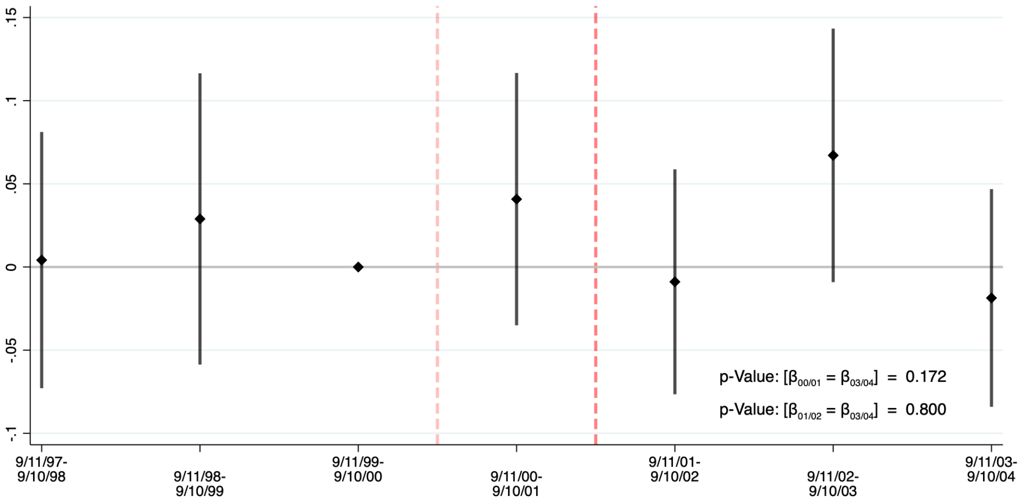


(c) Days Prison

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.

where the outcome variable is 1-year recidivism. We make a couple of other adjustments. First, due to the new outcome, our time variable is now the date of release from prison, rather than the day one's parole file is rated. Second, we restrict the sample to those inmates who were statutorily eligible for parole and who do not serve their full sentence behind bars.[15] We do so based on a key insight from Anwar and Fang (2015), who note that if the parole board can continually assess inmates, and even if the parole board sets different thresholds for different groups in a way synonymous with bias, those granted parole (strictly) within the parole release window of one third of their sentence and their full sentence should all have a predicted recidivism risk equal to their group-specific threshold. This is important, as it allows us to sidestep the inframarginality problem that plagues outcome-based tests – within group, *all* released individuals are marginal.

Figure 2: Recidivism



Notes: The solid, black 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 released from prison. The omitted time period is the year prior to 9/11/00, and comprises inmates released between 9/11/99 and 9/10/00. We shift the reference period back one year in order to allow those released in this period a full year to recidivate in the pre-9/11 environment. It is for this reason that we mark two vertical lines on the graph – those released between 9/11/00 and 9/10/01 are released pre-9/11 but may recidivate in both pre- and post-9/11 time periods, thus are a mixed treatment group.

We present the results from this dynamic DD specification graphically in Figure 2 and conclude that there were no statistically significant systematic changes to recidivism rates – either short- or longer-run – for Muslim individuals in our sample of marginal releasees.

While our standard errors are large, the small point estimates are consistent with a board that is balancing increased recidivism risk by Muslim parolees with harsher parole decisions (see Kuziemko (2013)). Consider three points. First, in our recividism analysis, we consider only inmates who are eligible for release, thus those who are released may be considered on the

[15]We drop individuals who served less than 33% or 100% of their sentence.

margin of release – those for whom the costs of incarceration and costs of release are balanced. Second, it seems reasonable to assert that incarceration costs did not change abruptly just after 9/11, and third, we show in Figure 2 that recidivism did not change differentially for Muslims post-9/11. Taken together, these results suggest that the board took into account recidivism risk for Muslims in their post-9/11 decisions. In contrast, if the board had acted purely out of animosity, penalizing low recidivism risk inmates with harsher parole decisions, we would expect to see a *decline* in recidivism for Muslim parolees post-9/11. Put differently, the harsher parole outcomes for Muslim inmates that we document in Section 4 may be a reasonable response of the board to external factors influencing Muslim individuals revidivism risk post-release – even with considerably longer time in prison for Muslim inmates, the point estimates for post-release recidivism are unchanged post-9/11.

5.2 Does the Parole Board Response to 9/11 Vary With Inmate Ex-Ante Recidivism Risk?

Building on the heterogeneity analysis we document in Section D.3 above, and the small point estimates for recidivism rates post 9/11, we hypothesize that an inmates underlying recidivism risk is a key factor driving the parole decision-making in the period around 9/11. We turn to Table D4, which 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.

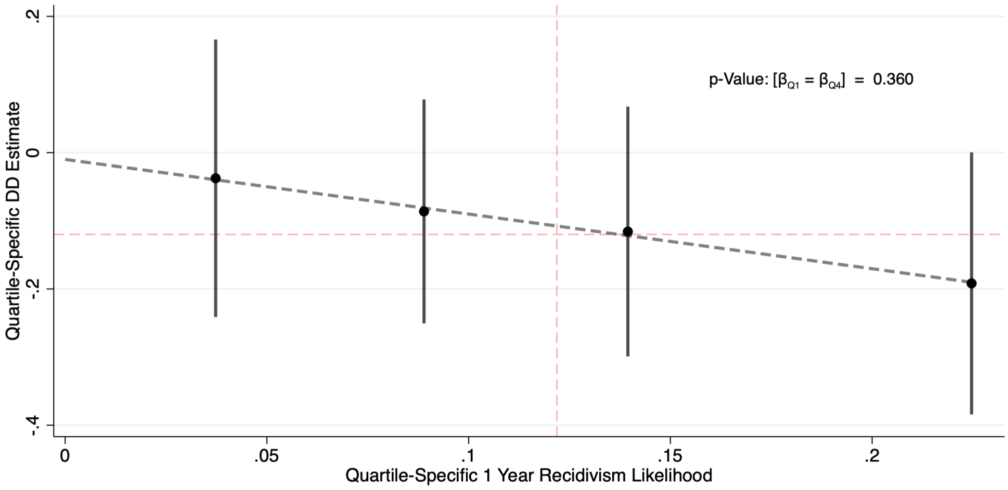
We first see that just over one in ten former inmates recidivates in the first year. What is more relevant for our previous results becomes apparent in Panel B – the recidivism risk for those convicted of a low-severity offense is more than double that of those convicted of high-severity offenses.[16] 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. We plot the quartile-specific DD estimates against quartile-specific baseline recidivism risk. The plots can be found in Figure 3.

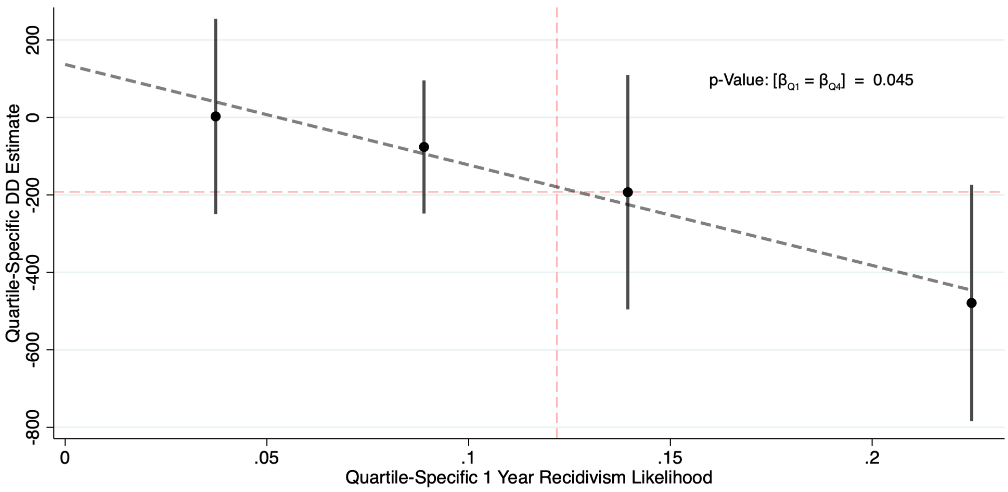
These graphs can shed light on the underlying mechanisms driving our core findings. If the parole board changes their behavior towards Muslim inmates in the post-9/11 period for statis-

[16] In order to characterize low-severity offenders, we estimate a probit model where the dependent variable is a low-severity offense indicator, and use a similar set of explanatory variables to those used in the main analysis. These inmates are more likely to be serving time for property and drugs crimes, are slightly older and are less likely to have children.

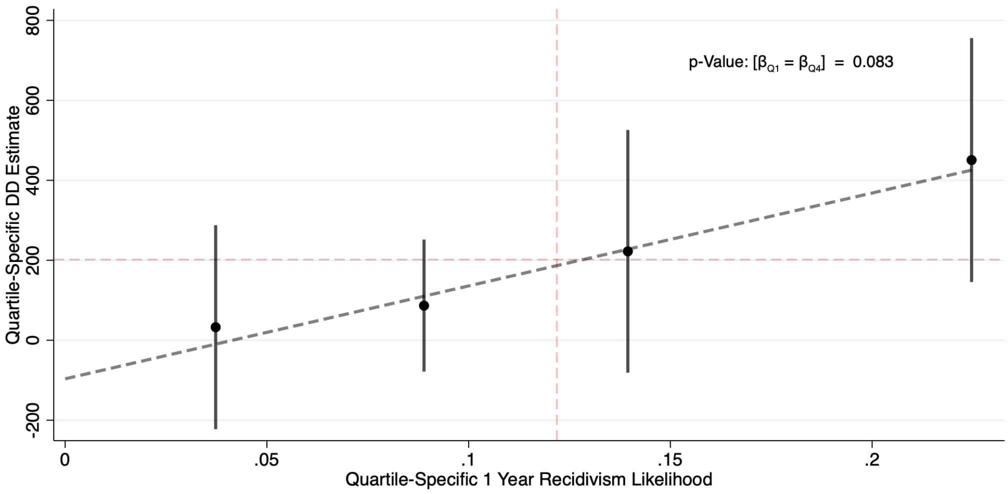
Figure 3: Sub-group DD Estimates Versus Sub-group Baseline Recidivism Risk



(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 was 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.

tical discrimination-based reasons, then the board will likely use known correlates of recidivism to inform such a decision. This should give rise to a negative (positive) relationship between the quartile-specific DD estimates and quartile-specific recidivism risk for DD estimates that are negative (positive) for the full sample, as presented in Table 1 – the higher the predicted recidivism risk, the larger the estimated effects for Muslim inmates rated and reviewed by the parole board post-9/11. If however the parole board changes their behavior due to a uniform post-9/11 animus directed towards Muslim inmates up for parole, then we should not expect to find a relationship between the predicted recidivism risk and the quartile-specific DD estimate. Instead we should find a level effect.

In fact, Figure 3 provides suggestive evidence against this level effect. Rather we document more negative parole-based DD estimates for higher recidivism risk groups, and larger DD estimates for time spent in prison. We interpret these findings as suggestive evidence that the parole board appears to be using their knowledge of the non-religious correlates of recidivism to exercise discretion over their decision-making when reviewing Muslim inmates in the post-9/11 environment. We argue that the results we present here are thus suggestive of statistical discrimination in the parole board decision-making.

As an extension, we speculate that parole boards are less likely to inaccurately sterotype Muslims post-9/11 for Muslim inmates with the lowest recidivism risk. In that case, the intercepts of the slopes in Figure 3 indicate any bias generated by non-risk factors, which continues to impact Muslim inmates with low or zero recidivism risk.

Table 2: 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	-.00987	137	–96.5
	(.118)	(149)	(147)
ZRR as Percentage of Baseline	8.22%	–71.2%	–47.8%

Notes: Predicted recidivism is based on 1 year recidivism probabilities for sample released 1/1/1999-9/10/2000. 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 2, 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 2, 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.

In sum, the evidence we present in Figure 3 and Table 2 suggests that the stark estimates we document for our full sample are unlikely to arise from a blanket change in taste-based discrimination of the parole board post-9/11. Rather, we present evidence that is consistent with a form of bias that varies by recidivism risk from the parole board, whereby high ex-ante recidivism risk Muslim inmates who came up for parole post-9/11 were treated harshly by the parole board, but low recidivism risk Muslim inmates were not. The findings we document here share a common thread with those from Section 5.1 – the parole board appear to be using knowledge of underlying recidivism risk of inmates to guide their response to the external shock of 9/11.

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. This is in contrast to recidivism outcomes, for which we find no statistically significant effects post 9/11. 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 sentences. Taken together with the dynamic effects, we suggest that risk-based factors likely played a sizeable role in the parole outcome responses.

Our work has policy implications for the optimal design of criminal justice systems. Our findings point to a potentially under-appreciated aspect of parole boards – their flexibility to respond to external, post-sentencing events. In a parole board based system, a judge sentences an individual to a prison term, which is then assessed at intervals by a parole board whose aim is to balance predicted recidivism risk with the costs of incarceration to decide upon when to optimally release the inmate. If there are large, external shocks, such as the one that we study in this work, that impact perceived recidivism risk upon release, the parole board could potentially change their decision making in response. This is very different from a purely judicial-based system, where there is little scope to respond to external shocks post-sentencing.

On the other hand, 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, such as potential religious discrimination in the labor market post-9/11. 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 both more efficient and more equitable, should thus consider the nuanced role of a parole board system.

References

Abadie, A. and S. Dermisi (2008): "Is terrorism eroding agglomeration economies in central business districts? Lessons from the office real estate market in downtown Chicago," *Journal of urban Economics*, 64, 451–463.

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.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *The Journal of Political Economy*, 76, 169–217.

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.

Bisin, A., E. Patacchini, T. Verdier, and Y. Zenou (2008): "Are Muslim immigrants different in terms of cultural integration?" *Journal of the European Economic Association*, 6, 445–456.

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.

Bravo, G. A. and J. V. Castello (2021): "Terrorist attacks, Islamophobia and newborns health," *Journal of health economics*, 79, 102510.

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

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

Danziger, S., J. Levav, and L. Avnaim-Pesso (2011): "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences*, 108, 6889–6892.

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.

Elsayed, A. and A. De Grip (2018): "Terrorism and the integration of Muslim immigrants," *Journal of Population Economics*, 31, 45–67.

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

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

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.

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.

Kim, D. and Y.-I. Albert Kim (2018): "Mental health cost of terrorism: Study of the Charlie Hebdo attack in Paris," *Health economics*, 27, e1–e14.

Kuziemko, I. (2013): "How should inmates be released from prison? An assessment of parole versus fixed-sentence regimes," *The Quarterly Journal of Economics*, 128, 371–424.

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

Light, M. T., M. Massoglia, and E. Dinsmore (2019): "How do criminal courts respond in times of crisis? Evidence from 9/11," *American journal of sociology*, 125, 485–533.

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

Mason, P. L. and A. Matella (2014): "Stigmatization and racial selection after September 11, 2001: self-identity among Arab and Islamic Americans," *IZA Journal of Migration*, 3, 1–21.

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.

Nowak, A. and J. Sayago-Gomez (2018): "Homeowner preferences after September 11th, a microdata approach," *Regional Science and Urban Economics*, 70, 330–351.

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.

Quintana-Domeque, C. and P. Ródenas-Serrano (2017): "The hidden costs of terrorism: The effects on health at birth," *Journal of Health Economics*, 56, 47–60.

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

Ratcliffe, A. and S. von Hinke Kessler Scholder (2015): "The London bombings and racial prejudice: Evidence from the housing and labor market," *Economic Inquiry*, 53, 276–293.

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. (Forthcoming): "Pre-test with caution: Event-study estimates after testing for parallel trends," *American Economic Review: Insights*.

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.

Weisburst, E. K. (2022): "Whose help is on the way? The importance of individual police officers in law enforcement outcomes," *Journal of Human Resources*, 0720–11019R2.

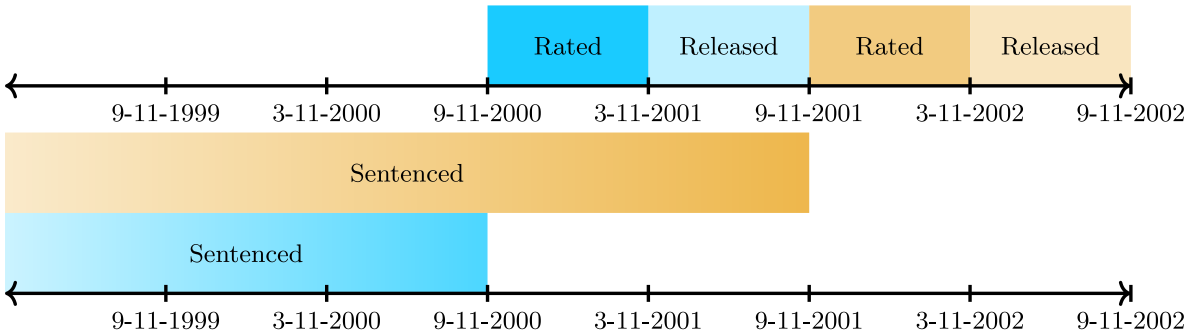
Woods, J. (2011): "The 9/11 effect: Toward a social science of the terrorist threat," *The Social Science Journal*, 48, 213–233.

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)

	Success score group (success score range)		
Crime severity level	Excellent (14-20)	Average (9-13)	Poor (0-8)
I	10	16	22
II	12	18	24
III	14	20	26
IV	16	22	28
V	34	40	52
VI	52	62	78
VII	72	84	102
VIII	65% prison sentence	75% prison sentence	90% prison sentence

Notes: This table shows the Parole Board Guidelines (Grid) used in Georgia during our sample period. The Grid specifies the recommended prison time (in months) based on the crime severity level and success scores. Details on the calculation of the success scores and the classification of the crime severity level can be found on the Georgia's Board of Pardons and Paroles 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 both (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. This "triangulation" approach is powerful – if we look at the same issue from three distinct perspectives, and in all three cases return the same verdict, we can be more confident of our conclusion than had we considered only a single viewpoint.

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

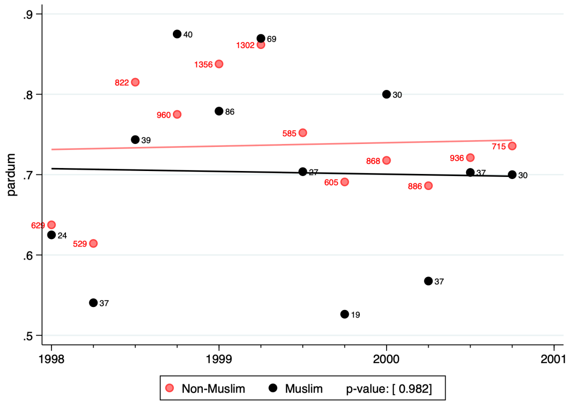
	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	.0416	49.1	–59.4
	(.0552)	(74.9)	(74.1)
$\overline{Y}_{0,PRE}$	.714	1011	882
Adjusted R^2	.319	.873	.606
Observations	5,031	5,031	5,031

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. A +/-365 day window around 9/11/2000 is used for estimation.

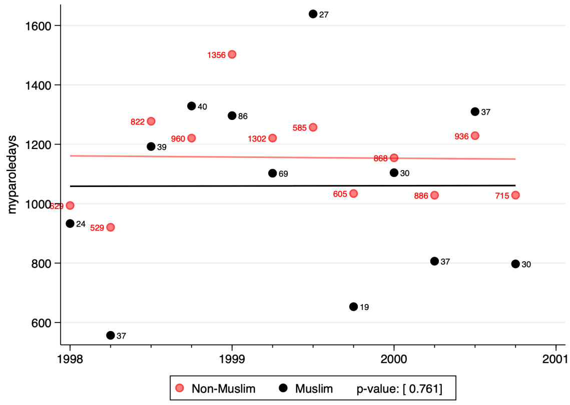
C.1.2 Trends in the Raw Data

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.

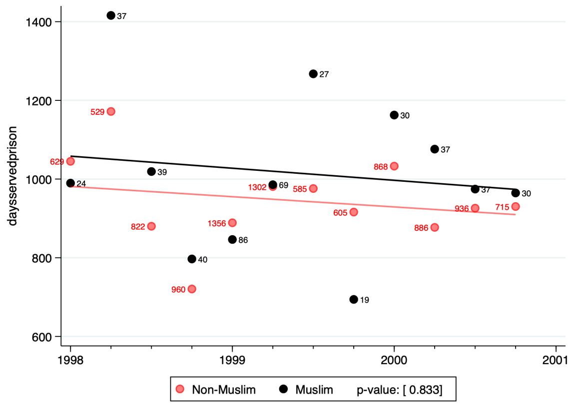
Figure C1: Raw Pre-Trends



(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.

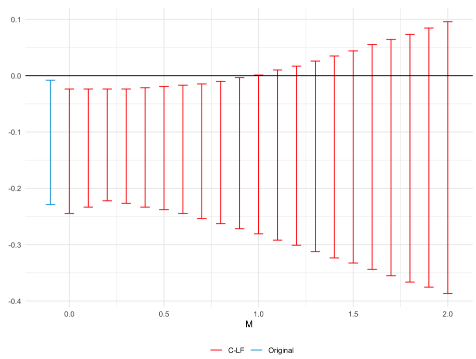
The graphical outputs from the Rambachan and Roth (2022) approach, where we use the Relative Magnitude approach for bounding, are presented in Figure C2. For all three outcomes, the "breakdown value" of $\overline{M}$ – the factor of the pre-trends at which the bounds on the estimated treatment effect 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

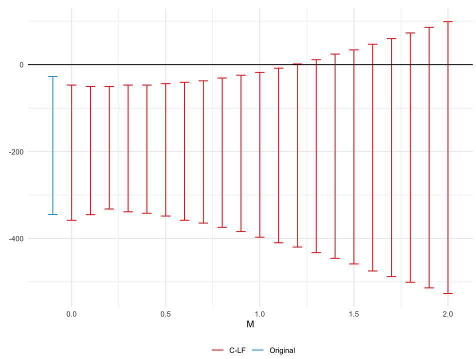
	(1)	(2)	(3)
	Parole	Days Parole	Days Prison
Period$_1 \times$Muslim	-.0381	–43.2	50.5
	(.0563)	(75.7)	(74.9)
	[.499]	[.568]	[.5]
Period$_3 \times$Muslim	-.118	–186	199
	(.0563)	(81)	(80.7)
	[.0354]**	[.0215]**	[.0139]**
$\overline{Y}_{0,PRE}$	.71	960	868
Adjusted R^2	.295	.855	.577
Observations	7,460	7,460	7,460

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses, regular p-values in brackets and randomization inference p-values, based on 10,000 permutations, in braces. A +/-365 day window around 9/11/2000 is used for estimation.

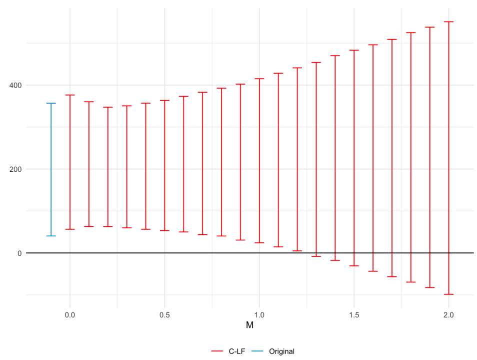
Figure C2: Honest Difference-in-Differences



(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	2317		112	110		
Education:							
$\leq$ High School	.649	.673	[.081]	.688	.591	[.135]	[.066]
High School	.249	.235	[.274]	.196	.245	[.381]	[.271]
Some College	.0868	.0777	[.261]	.0982	.155	[.209]	[.148]
College	.0153	.0134	[.590]	.0179	.00909	[.573]	[.664]
I.Q. Score	94	95.1	[.070]	96	96.1	[.984]	[.697]
	(22.7)	(19.9)		(18.9)	(21.6)		
Has Children	.674	.67	[.799]	.705	.618	[.171]	[.197]
Married	.111	.103	[.376]	.179	.0818	[.032]	[.052]
Prior Convictions	1.35	1.49	[.000]	1.63	1.62	[.970]	[.434]
	(1.28)	(1.31)		(1.41)	(1.33)		
Age at Sentencing	30.9	30.8	[.738]	29.5	28.8	[.559]	[.627]
	(9.59)	(9.58)		(8.77)	(7.65)		
Severity Level	2.87	2.8	[.137]	2.94	2.7	[.274]	[.454]
	(1.69)	(1.59)		(1.61)	(1.62)		
Sentence Length	1977	2055	[.109]	2052	2049	[.989]	[.716]
	(1512)	(1759)		(1524)	(1697)		

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
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. A +/-365 day window around 9/11/2001 is used for estimation.

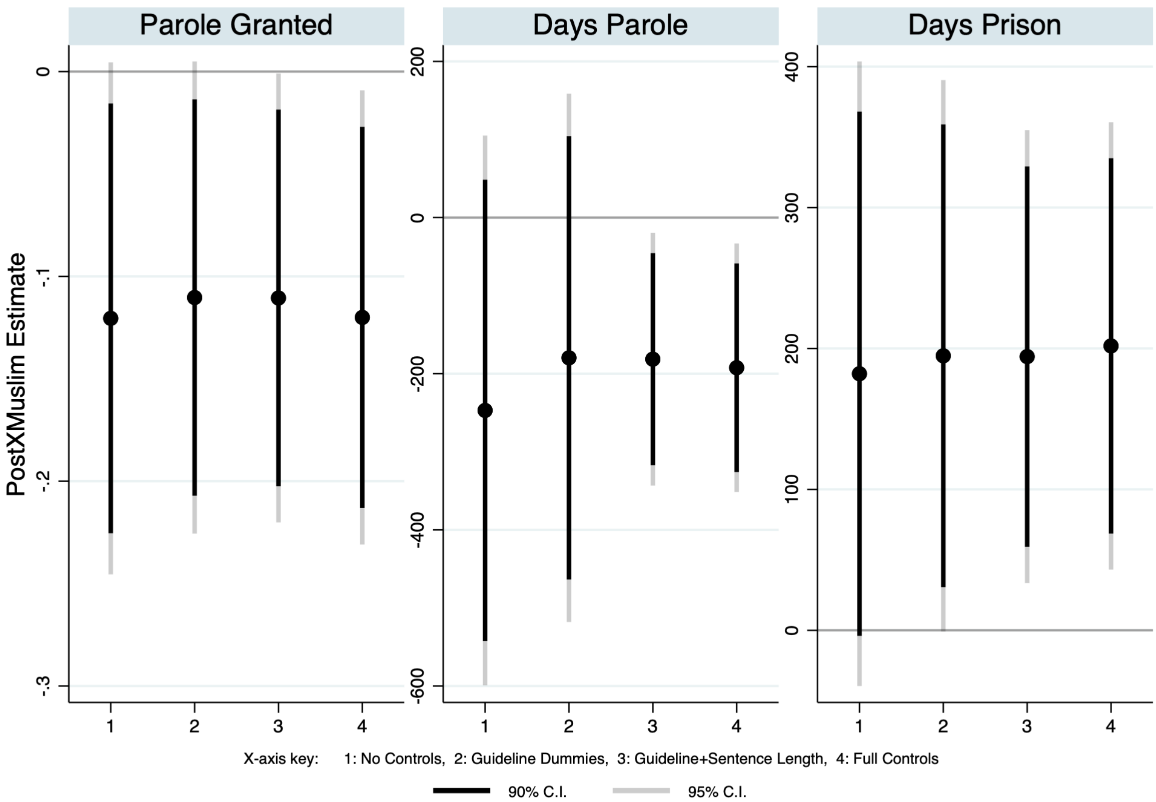
D Robustness and Ancillary Results

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 is 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.

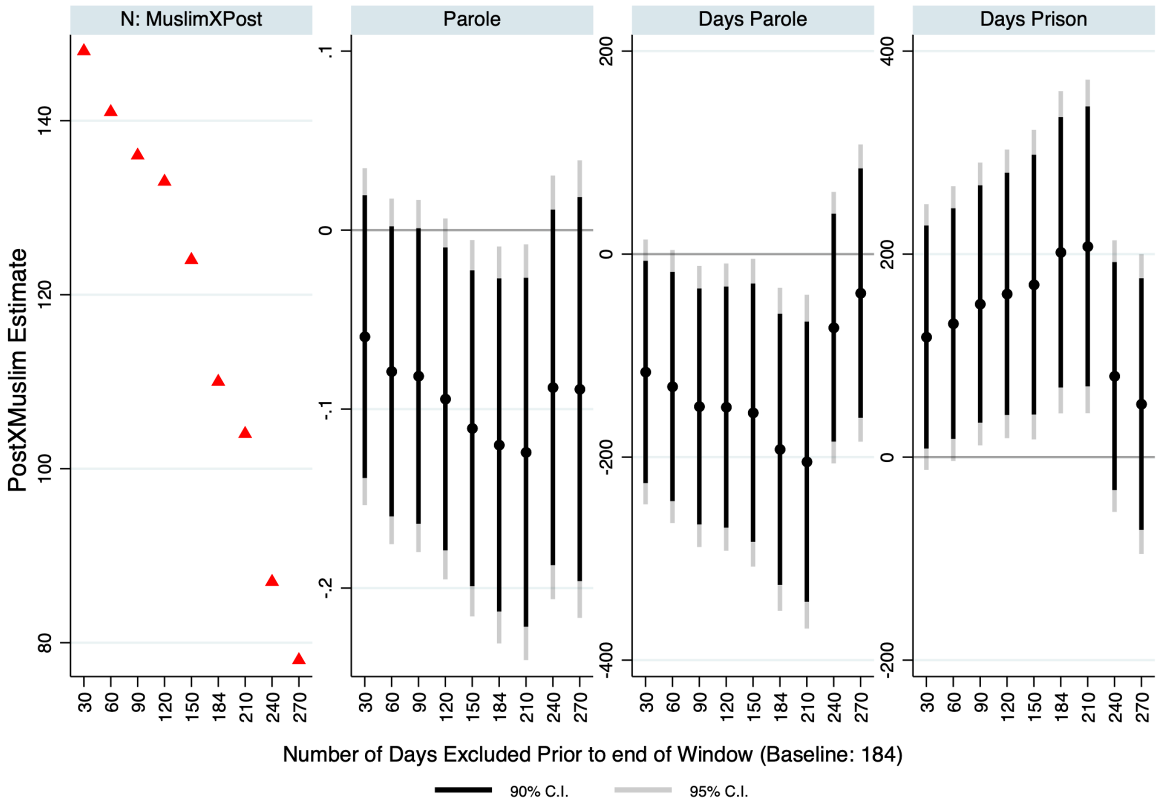
Figure D1: The DD Estimate is Stable as we Sequentially Include Controls



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 will not see a case on the rate date. Our baseline is to specify a 6 month exclusion window, the cost of which is to effectively halves 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.

Figure D2: The DD Estimate is Stable Based on Sample Selection Decision Regarding Exclusion Window



D.3 Short Run Impacts – Treatment Effect Heterogeneity

D.3.1 Heterogeneity by Offense Severity

Table D1: Parole Board Decisions and Prisoner Outcomes by Offense Severity

	(1)	(2)	(3)	(4)	(5)	(6)
	Low-Severity Offenses			**High-Severity Offenses**		
	Parole Granted	Days Parole	Days Prison	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	-.161**	–181**	177**	-.0316	–177	229
	(.0647)	(88.5)	(88.3)	(.103)	(177)	(180)
$\overline{Y}_{0,PRE}$	.794	1056	671	.482	704	1404
(Post-9/11×Muslim) / $\overline{Y}_{0,PRE}$	-.203**	-.171**	.264**	-.0656	-.252	.163
	(.0815)	(.0838)	(.131)	(.213)	(.252)	(.128)
Adjusted R^2	.161	.888	.355	.385	.765	.644
Observations	3,619	3,619	3,619	1,213	1,213	1,213

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. Low-Severity Offenses are those with an offense severity level of 1-4. High Severity Offenses are those with an offense severity level of 5 and above. The offense severity level is one of the two inputs that form the parole board grid. A +/-365 day window around 9/11/2001 is used for estimation.

D.3.2 Heterogeneity by Grid Score Group

Table D2: Parole Board Decisions and Prisoner Outcomes by Grid Score Group

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Excellent			**Average**			**Poor**		
	Parole Granted	Days Parole	Days Prison	Parole Granted	Days Parole	Days Prison	Parole Granted	Days Parole	Days Prison
(Post-9/11× Muslim)	-.162	–103	145	-.0626	–165	162	-.177**	–192*	206*
	(.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× Muslim) / $\overline{Y}_{0,PRE}$	-.219	-.106	.161	-.0843	-.169	.197	-.271**	-.206*	.226*
	(.207)	(.24)	(.27)	(.106)	(.126)	(.146)	(.135)	(.122)	(.124)
Adjusted R^2	.384	.775	.601	.291	.879	.528	.254	.82	.547
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. A +/-365 day window around 9/11/2001 is used for estimation.

D.4 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 D3. We find no evidence of any statistically meaningful changes in disciplinary outcomes for Muslim inmates in the post-9/11 period, thus suggesting that disciplinary infractions whilst incarcerated was not a key mediator for our main results.

Table D3: Disciplinary Outcomes

	(1)	(2)	(3)	(4)
	Reports	**Charges**		
	Total	Violent	Non-Violent	Total
Post-9/11×Muslim	1.11	.288	1	1.29
	(.98)	(.241)	(1.29)	(1.46)
$\overline{Y}_{0,PRE}$	3.47	.468	4.39	4.85
Adjusted R^2	.181	.0753	.142	.141
Observations	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. A +/-365 day window around 9/11/2001 is used for estimation.

D.5 Ex-Ante Recidivism Risk

Table D4: Sub-Group Recidivism Risk

	(1)	(2)
	Recidivism Risk	Sub-Group Size
A.) Full Sample	0.122	15082
B.) Offense Severity		
0	0.147	10854
1	0.056	4228
C.) Predicted Recidivism Quartile		
1	0.037	3700
2	0.089	3699
3	0.139	3700
4	0.225	3699

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