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

Brendon McConnell* Kegon Teng Kok Tan† Mariyana Zapryanova‡

December 18, 2022

Abstract

We provide the first evidence of the impact of 9/11 on outcomes for Muslims in the US criminal justice system by bringing novel data that captures the religion of inmates to bear. 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 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 deeply 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, Director of Communications at the Georgia Board of Pardons and Paroles, for many useful conversations. We are grateful for useful comments and suggestions from seminar participants at Smith College Economics Brown Bag Seminar and the University of Wisconsin-Madison Empirical Micro Seminar.

1 Introduction

The threat of discrimination in the American criminal justice system is an ongoing debate that has, in recent years, intensified and broadened to cover policing, sentencing, bail, 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.

The focus of this paper is on disparities in parole outcomes, a key aspect of reentry that allows prison inmates to be released before their sentences are fully served. Parole decisions are of particular interest in that they are the last instance where the criminal justice system has power over the inmate for their current spell. As a result, the impact of changes in parole outcomes are followed almost immediately with post-release outcomes such as recidivism, without much opportunity for other factors within the criminal justice system to come into play.

In this paper we examine parole outcomes in the time period around a large, societal event – the 9/11 terrorist attacks. Whilst the attacks 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 (Singh, 2002; Panagopoulos, 2006; Kaufman, 2019; Brodeur and Wright, 2019). We thus conceptualize the attacks as an exogenous shock to the level of animosity towards Muslims.

We use administrative records from the Department of Corrections in the state of Georgia, where a five-member parole board reviews inmates' cases and decides whether or not to grant parole. Importantly, the data has information on self-reported religion, which allows us to consider the impact of such a shock to animosity towards Muslims in a regression-adjusted difference-in-differences (DD) framework. In doing so, our study is the *first* to investigate the impact of 9/11 on criminal justice outcomes for Muslims in the US.

We restrict our sample to Black, male, parole-eligible inmates. This is because virtually all Muslims in our data are Black males. Further, 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.

Given the detail of the administrative data we use, we are able to condition on a rich set of relevant control variables. We interpret our DD estimates as the causal effects of the 9/11 terrorist attacks on sentencing outcomes for Muslim inmates. In order to make such a causal statement, we require two core assumptions to hold. First, the parallel trends assumption and second, that the composition of the groups we study are stable across our two time periods.[1] We provide three pieces of evidence in support of the parallel trends assumption inherent in our DD approach – (i) we provide placebo DD evidence from the years prior to 9/11/2001, (ii) we present the raw trends in our data, and test for equality of trends and (iii) 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. Each piece of evidence approaches the topic of parallel trends from a different perspective. Each piece of evidence provides support for the claim 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.

In order to provide support for the stable group composition assumption, we present two separate analyses. First we consider the balance across a range of characteristic for both Muslim and non-Muslim inmates across the two periods we study. Our evidence is consistent with stability of group composition. Second, we implement a set of duration model analyses, presented in the Appendix, in order to assess if there was any re-ordering of when Muslim inmates came up for review for parole in the aftermath of 9/11. We find no such evidence.

We document a substantial change in the parole outcomes of Black Muslim inmates in the aftermath of 9/11. 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+ prior convictions) and first-time offenders. These estimates are all highly statistically significant, even with the power issues we face in our empirical analysis due to a severe treatment-control imbalance.

We show that the large impacts we document for the main sample persist for multiple years after 9/11, in contrast to other outcomes for Muslims that seem to exhibit only short term effects (for e.g., homicide rates against Muslims). In particular, we present analysis of recidivism rates,

[1] We require this second assumption as we use repeated cross-sectional data (Blundell and Dias, 2009).

providing suggestive evidence that Muslims were in fact at a higher risk of recidivism after 9/11. However, the increase in recidivism risk attenuates fairly quickly as we study inmates that were exposed not only to the environment outside of prison but also who were impacted by parole board choices post-9/11, although our effects are imprecisely estimated. The disparity in the dynamic patterns of parole outcomes and recidivism rates is is consistent with a permanent shift in religion-based parole decision making that does not match the recidivism outcomes of Muslims post 9/11.

Finally, we provide less precise but interesting heterogeneity analysis. Our estimated impact on parole outcomes vary significantly with characteristics that are correlated with ex-ante recidivism risk. We find that the negative effects on parole outcomes are *larger* for inmates who have a higher ex-ante recidivism risk. We interpret this as prima facie evidence that our baseline estimates may not be driven by a blanket taste-based discrimination mechanism, but rather is sensitive to underlying recidivism risks of inmates.

Our work contributes to the growing empirical literature on the influence of extraneous events on the administration of justice. Within this literature, attention has nearly exclusively focused on either judicial decision making (Philippe and Ouss, 2018; Brodeur and Wright, 2019; Light et al., 2019; Eren and Mocan, 2018; Shayo and Zussman, 2011) or prosecutorial behavior (Bielen and Grajzl, 2021; McConnell and Rasul, 2021).[2] We provide evidence that a large societal shock, such as the terrorist attacks of 9/11, impacts the decision making of parole boards as well.

Previous empirical work has focused on examining the effect of terrorism on various outcomes. We contribute to this literature by documenting the effects of the terrorist attacks of 9/11 on parole releases. Cornelissen and Jirjahn (2012), Davila and Mora (2005), Kaushal et al. (2007) examine the impacts of 9/11 on the labor market outcomes of Arabs and Muslims.[3] Our work also relates to studies examining the effect of terrorism on ethnic attitudes (Ratcliffe and von Hinke Kessler Scholder, 2015), self-identification (Mason and Matella, 2014), perception of risk (Abadie and Dermisi, 2008), and assimilation (Bisin et al., 2008; Elsayed and De Grip, 2018; Gould and Klor, 2016).

In our setting, parole board members are potentially adjusting the perceived risk of recidivism for Muslims, for reasons consistent with those documented by the literature on the impact of 9/11 for a wide swathe of outcomes, including labor market outcomes. If so, our finding regarding parole board discrimination against Muslims, even if statistical, would indicate that

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

[3]For examination of the macroeconomic consequences of terrorism, see for example Abadie and Gardeazabal (2003); Blomberg et al. (2004).

the prison system is ratifying broader social discrimination and stigmatization. We contribute to the vast literature on in-group bias in the criminal justice system by examining the existence of religious bias. Existing research 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), and parole boards (Anwar and Fang, 2015; Mechoulan and Sahuguet, 2015). Unlike these studies, we focus on religion as a basis for discrimination rather than race or ethnicity.[4]

Finally our work speaks to a potentially under-appreciated aspect of parole boards as a way to implement parole – their adaptability to 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. In Georgia, this intermittent assessment starts once the inmate has served one third of their full sentence. If there are large, external shocks, such as the one that we study, that impact recidivism risk upon release, the parole board can internalize the new environment and respond accordingly. This is very different to a purely judicial-based system, where once a judge hands down a sentence, the inmate will typically serve a portion of this sentence with pre-set guidelines that are not influenced by external changes.

2 The Parole Process in Georgia

The state of Georgia releases prisoners from prison using a discretionary parole system, where release is granted following a decision by a parole board that grants or withholds parole based on its assessment of individual cases.[5] The parole process in Georgia starts when a parole-eligible individual is transferred to a Georgia Department of Corrections Diagnostic Prison, where the pre-parole investigation begins. The pre-parole investigation is conducted by a parole hearing examiner, a role we also refer to as rater. This investigation comprises of the rater interviewing the prisoner and gathering information about the prisoner's personal information and criminal record. Importantly, for our study prisoners self-report their religion during this interview process.

The parole board in Georgia is required by law to make release decisions on the basis of the

[4]Religious bias has been examined in the context of the criminal justice system in India (Ash et al., 2021) and the Netherlands (Bielen and Grajzl, 2021), for example.

[5]Using data from Renaud (2019), we compare the parole process 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 parole size and overall rating of the parole system. In addition, Zapryanova (2020) shows that Georgia's prison population appears to be representative of that nationwide.

risk a person may pose to public safety if they were to be released from confinement (O.C.G.A. §42-9-40). The board has adopted the Parole Decisions Guidelines Grid System (hereafter the grid) as an assessment of this risk. Once the pre-parole investigation is finished, the rater uses the grid along with the prisoner's risk to re-offend score and current offense crime severity level, to determine the recommended prison time, also known as the grid recommendation.[6]

The hearing examiner then writes a summary discussing the contents of the parole file. When the parole file reaches the members of the parole board, it usually contains the pre-parole investigation interviews, grid rating, rater summary, prison disciplinary record, participation in self-improvement activities, work performance, and any other correspondence and statements by relevant actors, such as the judge and victims.

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 and a parole decision is reached if three out of five board members vote the same way (O.C.G.A. §42-9-42). At the time of consideration, the board establishes a Tentative Parole Month (TPM) in the future or will deny parole entirely. The TPM is not a release date, but rather a date on which the board will review the person's parole file and determine whether to set a final parole release date. 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).

3 Data and Empirical Framework

3.1 Data and Sample Selection Criteria

Data 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, 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 the parole board for a vote.[7]

Sample Selection 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

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

[7] 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.

restriction in Figure E2. 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 received 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. This decision allows us to directly focus on our key treatment variable – Muslim religion status.

3.2 Empirical Framework

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

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

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 for our estimation sample because it was collected during the pre-parole investigation, 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 guideline cells used as part of the determination of parole in Georgia.[8] These guidelines combine a prisoner's risk to re-offend 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 IQ, indicators for the most serious offense committed, social class categories, 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, and present the related p-value as well as randomization inference p-values. This is not standard in difference-in-differences, yet we proceed in this manner due to the overwhelming treatment/control imbalance we face with our data, and the inherent power

[8]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.

issues this creates for us. 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.[9]

3.2.1 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. 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 holding. 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 are stable across the pre and post periods (Blundell and Dias, 2009). We do so in two ways. 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.

[9] 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.

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.

Confident that our core identifying assumptions are satisfied, we proceed with our main analysis.

3.2.2 Sample Imbalance and Power

One of the main challenges we face when working with this sample is the imbalance between our control individuals (non-Muslim inmates) and their treated counterparts (Muslim inmates). The ratio of non-Muslim to Muslim inmates is 20.8 to 1. From the optimal sample size and power calculation literature (Duflo et al., 2007; McConnell and Vera-Hernández, 2015), we know that power is maximized when the sample proportion of treated, p, is equal to 0.5, with the sample size required to achieve a given power being proportional to $1/(p(1-p))$. In our case $p = .046$, which means we require a sample 5.7 times larger[10] to achieve the same power as the balanced treatment/control case.

We use the Shiny R dashboard, *Power_Panel*, that accompanies Schochet (2022) to calculate, for each of our dependent variables, the required sample size we would need in order to achieve a power of .8. We present the results of these calculations in Table C3. The key message we convey with these calculations is that our actual sample size is always below the sample size we require to achieve a power of .8, given our estimated DD coefficients and related parameters.

For this reason, we present two sets of p-values with our key results. The first are the regular p-values associated with our Eicker-Huber-White standard errors. The second set are randomization inference p-values, based on 10,000 permutations. For each permutation, we randomly assign our treatment variable – an indicator variable for Muslim religion status – and run our baseline DD regression specification.

4 Parole Board Decisions and Prisoner Outcomes

4.1 Core Results

We present our main results in Table 1.

[10] $[1/(.046(1-.046))] \,/\, [1/(.5*(1-.5))]] = 5.7$

Table 1: Parole Board Decisions and Prisoner Outcomes

	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	-.124	–194	204
	(.0565)	(80.7)	(80.5)
	[.0288]**	[.0161]**	[.0112]**
	{.0076}***	{.0036}***	{.0027}***
$\overline{Y}_{0,PRE}$	.71	961	869
Adjusted R^2	.293	.842	.555
Observations	4,832	4,832	4,832

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 9/11/2001 is used for estimation.

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 display on the third to last line in the table and denoted by $\overline{Y}_{0,PRE}$), this is a 17% reduction in the probability of receiving parole. The effect is statistically significant based on both regular and randomization inference approaches.

This lower likelihood of parole translates in 194 fewer days of parole. As one can see from the table, the average parole period for non-Muslim inmates in the pre-9/11 period is 961 days or 2.6 years. Thus, 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 reviews prior to 9/11.

We can benchmark the 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 (i) the effect of moving from grid cell 1,1 to grid cell 3,3 (199 additional prison days) or (ii) the (conditional) difference in time served between serial offender (eight or more prior convictions) and first-time offenders (192 additional prison days).

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 – on average, we document these men spent almost seven additional months of their lives behind bars. In Section **??**, we attempt to understand the reason for our core results.

Before doing so, we first conduct sensitivity analyses in order to probe our results. In Figure

E1 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 across the inclusion of the controls. In Figure E2 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 E2 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. Viewing the first panel is again instructive to see this.

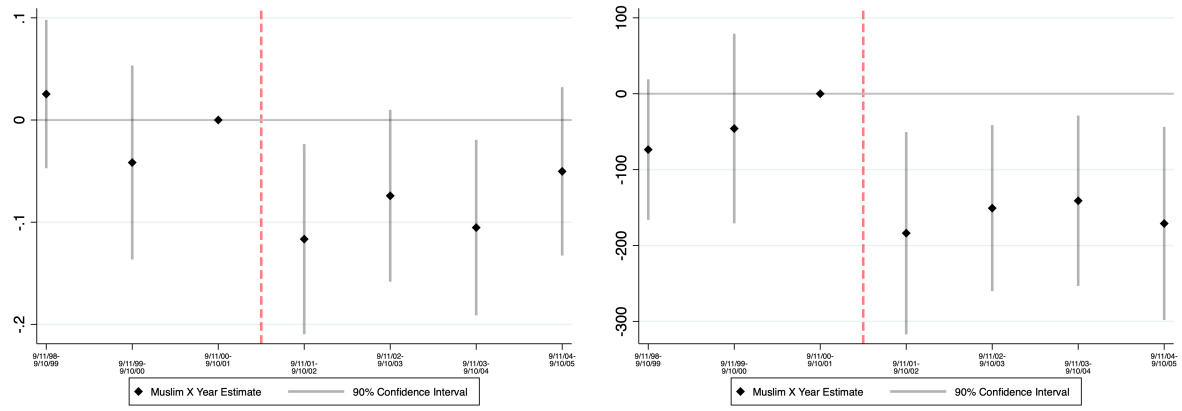
4.2 Dynamic Effects

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 maintained similar sampling rules for defining the extended sample. We present our findings in Figure 1.

The large decline in parole grants and days on parole, and the corresponding increase in days in prison that we document in Table 1 remain statistically significant and of similar magnitude through 2005. This is in contrast to other studies that have documented short-term impacts of 9/11 on hate crimes against Muslims as well as on labor market outcomes (Kaushal et al., 2007). Furthermore, we can check if recidivism risk in our data changed in a way consistent with parole outcomes.

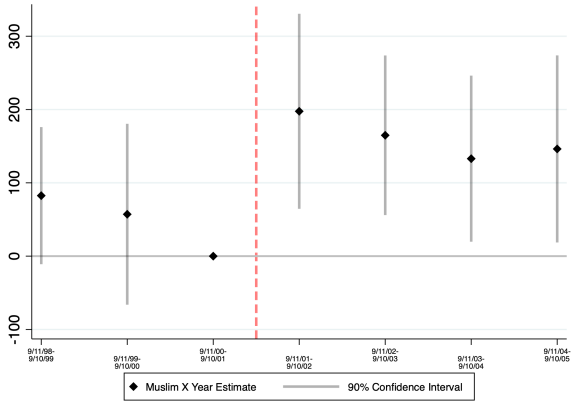
In the following analysis, we present suggestive, yet inconclusive, evidence that 9/11 may have impacted recidivism risk directly – once again treatment-control imbalance-driven power issues hamper our ability to say anything with greater statistical precision. We consider four groups of inmates, where group membership is based on the timing of the inmates being rated and reviewed by the board, released from prison, and being on non-custodial parole supervision in the community. We measure recidivism for up to one year post-release. The first temporal group – our base category – are rated and reviewed by the board, and then released from prison at least one year prior to 9/11. This means this groups one year post-release period is entirely prior to 9/11. The second temporal group are rated and reviewed by the board and released from prison before 9/11. However, they spend part of their one-year parole supervision in the community post 9/11. The third temporal group are rated and reviewed by the board before 9/11, and hence their parole decisions are not impacted by 9/11. However, they are released and spend their one-year parole supervision in the community after 9/11. In the fourth and

Figure 1: Dynamic Effects for Parole Outcomes



(a) Parole Granted

(b) Days Parole



(c) Days Prison

Notes: The lines of best fit are based on OLS, where each point is inverse weighted by the variance of the quartile-specific DD estimate.

final group, all key actions occur post-9/11.

By exploiting these different in the timing of parole review, prison release, and non-custodial supervision in the community treatment groups, we can show how 9/11 affected recidivism risk, both through external factors such as discriminatory labor market conditions (group 2, 3, and 4), as well as through parole decisions (group 4). The distinction between groups 2 and 3 come from the importance of being exposed to the 9/11 event whilst released. Note that both groups 2 and 3 would fall into our control group in the main specification.

We also restrict the sample to those inmates who were statutory eligible for parole and who do not serve their full sentence behind bars.[11] 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

[11]We dropped individuals who served less than 33% or 100% 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.

We report our estimates of the differential impact of 9/11 on recidivism of Muslims versus non-Muslims in Table D.1.

Table 2: Recidivism of Those Released at the Margin

	(1)	(2)	(3)	(4)
	Recidivism Within:			
	3 Months	6 Months	9 Months	12 Months
Temporal Group 2×Muslim	.0161	.0562	.0447	.0906
	(.0201)	(.0439)	(.0538)	(.0613)
Temporal Group 3×Muslim	-.00529	-.00099	-.0216	.0502
	(.00834)	(.0284)	(.0404)	(.0531)
Temporal Group 4×Muslim	.0103	-.0176	-.0107	.0146
	(.0156)	(.0228)	(.0388)	(.047)
$\overline{Y}_{0,PRE}$	.00608	.0308	.0625	.106
Adjusted R^2	.00135	.00544	.0153	.0225
Observations	7,257	7,257	7,257	7,257

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The base period for the temporal difference is all those rated after September 11, 1998 and subsequently released between September 11, 1998 and September 10, 2000, thereby allowing a full year for potential recidivism prior to 9/11/01.

While the estimates are noisy, they are consistent with potential increases in recidivism risk for Muslims after 9/11, particularly for the inmates who were already released when 9/11 occurred (group 2 and 3). We therefore cannot rule out the possibility that the parole board, post-9/11, had a statistical-based justification for making harsher parole decisions for Muslim inmates in the initial response.

However, the estimates decline substantially for group 4, indicating a wedge between observed recidivism risk farther out from 9/11 and parole decisions. Taken together, this is consistent with a persistent bias that is harder to justify with statistical differences in recidivism risk as we move into the subsequent years following the terrorist attacks.

4.3 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 3.

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

Table 3: 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	-.166	–188	184	-.0268	–163	217
	(.0646)	(88.1)	(88)	(.101)	(177)	(178)
	[.0102]**	[.033]**	[.0361]**	[.791]	[.357]	[.224]
	{.002}***	{.0059}***	{.0053}***	{.395}	{.182}	{.121}
$\overline{Y}_{0,PRE}$	.794	1056	671	.482	704	1404
Adjusted R^2	.168	.89	.372	.4	.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, regular p-values in brackets and randomization inference p-values, based on 10,000 permutations, in braces. 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 9/11/2000 is used for estimation.

The large decline in parole grants we document in Table 1 is driven by Muslim inmates with low severity offenses. Muslim inmates with low severity offense who are reviewed for parole post-9/11 are 17 percentage points, or 21%, less likely to be granted parole. This translates into an average prison sentence that is 27% 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 not surprising given that the parole board in Georgia appears to exert more discretion for high-severity offenses (Kuziemko, 2013).

Building on the heterogeneity effects above, we hypothesize that underlying recidivism risk is a key factor driving the parole decision responses. We turn to Table **??**, which presents the one-year recidivism risk for different groups of former inmates who were released at least one year prior to 9/11. The choice of sample means that the released prisoners spend their full non-custodial parole supervision in the community pre 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.[13] In Panel C, we present the recidivism risk for quartiles of predicted recidivism risk. The monotonic increase across quartiles is by design.

[13]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.

Table 4: 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 Quintile		
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

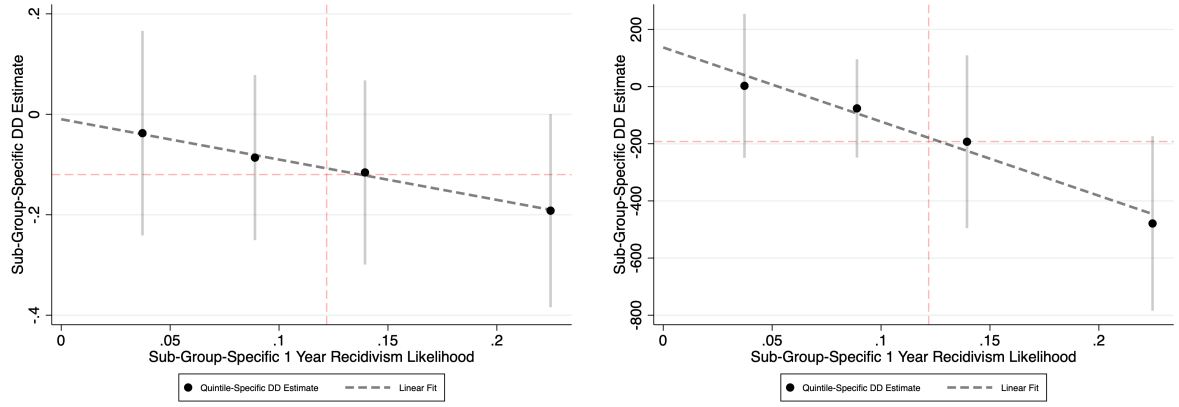
Using these predicted recidivism risk quartiles, we conduct a further set of heterogeneity analysis – we re-estimate our baseline specification separately for each quartile, and then plot the quartile-specific DD estimates against quartile-specific baseline recidivism risk. The plots can be found in Figure 2.

If the parole board changes their behavior towards Muslim inmates in the post-9/11 period for statistical 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 post-9/11 animus directed towards Muslim inmates up for parole equally for Muslims in all predicted recidivism risk quartiles, then we should not expect to find a relationship between the predicted recidivism risk and the quartile-specific DD estimate. Instead this should be a level effect.

In fact, Figure 2 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 that the parole board appears to be using their knowledge of the non-religious correlates of recidivism, x, 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.

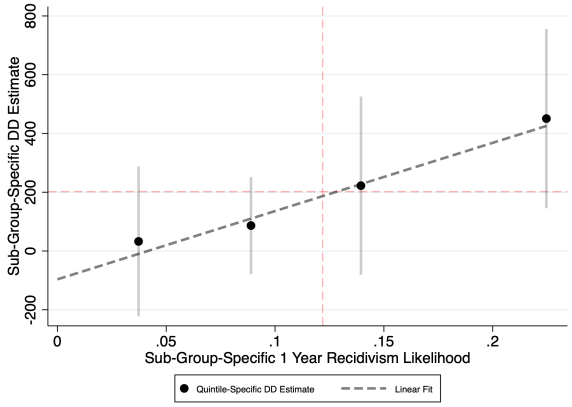
As an extension, we stipulate that the prediction error of the parole board is constant only for inmates with low or zero recidivism risk. The idea is that parole boards are less likely to commit prediction error due to inaccurate sterotypes of Muslims post-9/11 for the "best" Muslim

Figure 2: Sub-group DD Estimates Versus Sub-group Baseline Recidivism Risk



(a) Parole Granted

(b) Days Parole



(c) Days Prison

Notes: The lines of best fit are based on OLS (thick, dashed line) and median regression (thing, red line), where each point was inverse weighted by the variance of the quartile-specific DD estimate.

inmates with low or zero recidivism risk, even if they may commit prediction error that varies with non-religion factors for other Muslim inmates. In that case, the intercepts of the slopes in Fig 2 indicate any bias generated by non-risk factors, which continues to impact Muslim inmates with low or zero recidivism risk.

We present the results of this exercise in Table 5, where we show estimates of both the mean and the median projections of the zero recidivism risk DD estimate. If we are willing to extrapolate from our estimates to the point at which predicted recidivism risk is zero (ZRR), we find that both the mean and median ZRR projections for parole grants (Table 5, Column (1) are essentially zero. These two points are represented by the intercepts of the grey dashed (mean) and red solid (median) lines in Figure 2a. For parole supervision length we find a ZRR that is far smaller than our baseline DD estimate, although we document a divergence between mean and median estimates here that we do not find for the other outcomes. For prison days our ZRR projections

Table 5: Extrapolating Sub-Group DD estimates to Zero Recidivism Risk

	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Baseline DD Estimate	-.12	–192	202
Terciles of Recidivism Risk			
Zero Recidivism Risk (ZRR) Projection	.0495	38.3	-.542
ZRR as Percentage of Baseline	–41.2%	–19.9%	-.269%
Quartiles of Recidivism Risk			
Zero Recidivism Risk (ZRR) Projection	-.00987	137	–96.5
ZRR as Percentage of Baseline	8.22%	–71.2%	–47.8%
Quintiles of Recidivism Risk			
Zero Recidivism Risk (ZRR) Projection	.0617	148	–132
ZRR as Percentage of Baseline	–51.4%	–77.2%	–65.5%

Notes: Predicted recidivism is based on 1 year recidivism probabilities for sample released 1/1/1999-9/10/2000. The linear extrapolation for the mean is based on OLS, and for the median based on a median regression using a Parzen kernel and Chamberlain's bandwidth. In both cases we inverse-weight by the variance of the quintile-specific DD estimate.

are both negative.

In sum, the evidence we present in Figure 2 and Table 5 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.

5 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.

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 find that both taste-based bias and risk-based factors likely played a 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 susceptibility 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. 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 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 P. Hull (2022): "Measuring racial discrimination in bail decisions," *American Economic Review*, 112, 2992–3038.

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.

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.

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 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.

Duflo, E., R. Glennerster, and M. Kremer (2007): "Using Randomization in Development Economics Research: A Toolkit," Elsevier, vol. 4 of *Handbook of Development Economics*, chap. 61, 3895 – 3962.

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.

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.

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.

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.

McConnell, B. and M. Vera-Hernández (2015): "Going beyond simple sample size calculations: a practitioner's guide," .

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

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.

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*.

Schochet, P. Z. (2022): "Statistical Power for Estimating Treatment Effects Using Difference-in-Differences and Comparative Interrupted Time Series Estimators With Variation in Treatment Timing," *Journal of Educational and Behavioral Statistics*, 47, 367–405.

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

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.

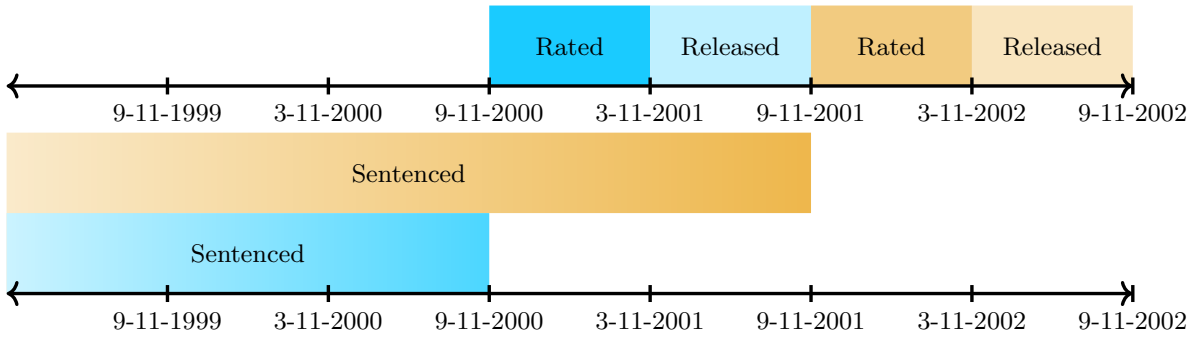
Woods, J. (2011): "The 9/11 effect: Toward a social science of the terrorist threat," *The Social Science Journal*, 48, 213–233.

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

Note: 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 website at https://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

Note: 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

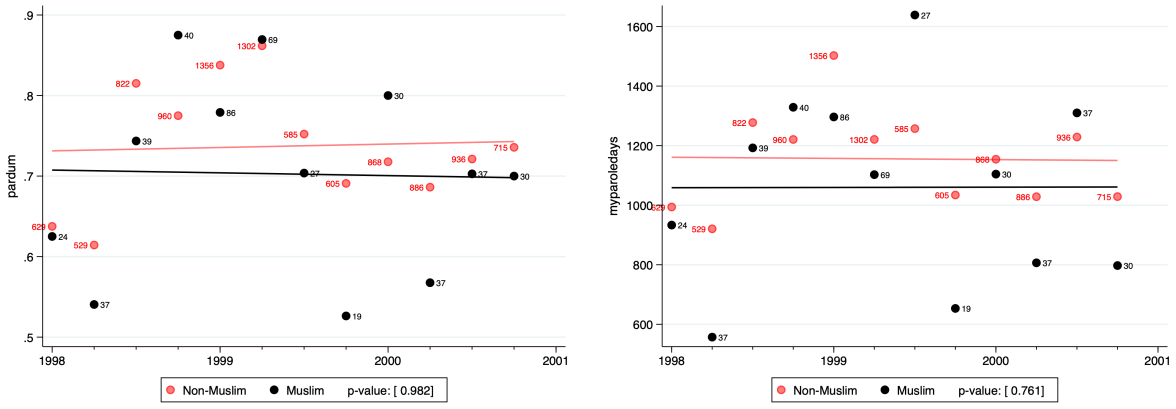
	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Post-9/11×Muslim	.0428	41.2	–53.9
	(.0545)	(74.6)	(73.9)
	[.432]	[.581]	[.465]
	{.218}	{.283}	{.225}
$\overline{Y}_{0,PRE}$	.714	1011	882
Adjusted R^2	.323	.875	.611
Observations	5,031	5,031	5,031

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.

C.1.2 Trends in the Raw Data

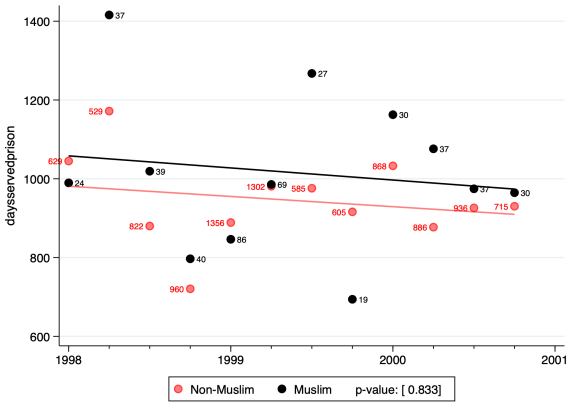
Next, in Figure C1, we provide graphical evidence of the existence of pre-trends by presenting the raw, underlying data for three years prior to our estimation sample – the calendar years of 1998-2000. We present the p-values from a test of equality of trends. We cannot reject this null of equality of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

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

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.

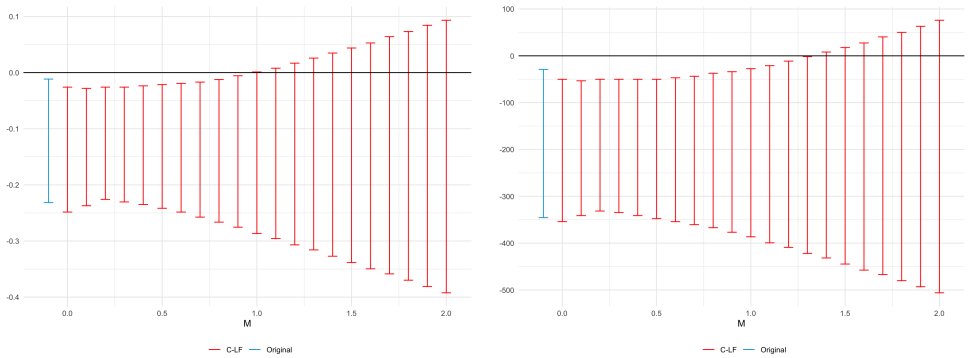
The graphical outputs from the Rambachan and Roth (2022) approach, where we use the Relative Magnitude approach for bounding, are presented in Figure C2. For all three outcomes, the "breakdown value" of $\overline{M}$ – the factor of the pre-trends at which the bounds on the estimated treatment effect 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
$\text{Period}_1 \times \text{Muslim}$	-.0381	–43.2	50.5
	(.0563)	(75.7)	(74.9)
	[.499]	[.568]	[.5]
$\text{Period}_3 \times \text{Muslim}$	-.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

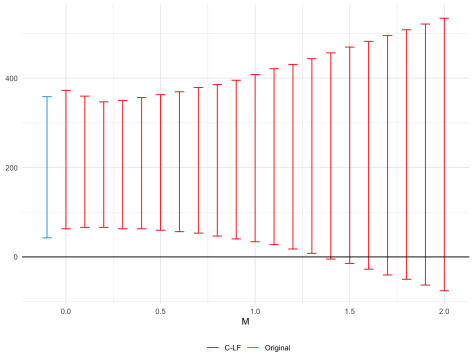
Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses, regular p-values in brackets and randomization inference p-values, based on 10,000 permutations, in braces. A +/-365 day window around 9/11/2000 is used for estimation.

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	.152	-.162	-.201	-.299
	(.142)	(.134)	(.153)	(.233)	(.12)	(.143)	(.129)	(.22)
	[.519]	[.315]	[.34]	[.959]	[.205]	[.259]	[.119]	[.175]
$\overline{Y}_{0,PRE}$	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 Additional Recidivism Results

D.1 Did R Change?

We start by adopting a definition of discrimination similar to Arnold et al. (2022):

$$E[D|m = 1, R = r] - E[D|m = 0, R = r], \tag{3}$$

which corresponds to our estimates for the differential impact of 9/11 for Muslims versus non-Muslims, under the assumption that the true risk of recidivism R did not change due to 9/11. In that case, any wedge generated by 9/11 must be coming from unjustified parole board responses through c, λ, or β. We acknowledge that this might be a strong assumption to make, given the likely widespread impact of 9/11 on Muslims including conditions in a post-9/11 world that are related to recidivism risk. However, it is challenging to recover the degree to which recidivism risk changed due to 9/11, and likely beyond the scope of this paper.

In the following analysis, we present suggestive, yet inconclusive, evidence that 9/11 may have impacted recidivism risk directly – once again treatment-control imbalance-driven power issues hamper our ability to say anything with greater statistical precision. We consider four groups of inmates, where group membership is based on the timing of the inmates being rated and reviewed by the board, released from prison, and being on non-custodial parole supervision in the community. We measure recidivism for up to one year post-release. The first temporal group – our base category – are rated and reviewed by the board, and then released from prison at least one year prior to 9/11. This means this groups one year post-release period is entirely prior to 9/11. The second temporal group are rated and reviewed by the board and released from prison before 9/11. However, they spend part of their one-year parole supervision in the community post 9/11. The third temporal group are rated and reviewed by the board before 9/11, and hence their parole decisions are not impacted by 9/11. However, they are released and spend their one-year parole supervision in the community after 9/11. In the fourth and final group, all key actions occur post-9/11.

By exploiting these different in the timing of parole review, prison release, and non-custodial supervision in the community treatment groups, we can show how 9/11 affected recidivism risk, both through external factors such as discriminatory labor market conditions (group 2, 3, and 4), as well as through parole decisions (group 4). The distinction between groups 2 and 3 come

from the importance of being exposed to the 9/11 event whilst released. Note that both groups 2 and 3 would fall into our control group in the main specification.

We also restrict the sample to those inmates who were statutory eligible for parole and who do not serve their full sentence behind bars.[14] 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.

We report our estimates of the differential impact of 9/11 on recidivism of Muslims versus non-Muslims in Table D.1.

Table D1: Recidivism of Those Released at the Margin

	(1)	(2)	(3)	(4)
	Recidivism Within:			
	3 Months	6 Months	9 Months	12 Months
Temporal Group 2×Muslim	.0161	.0562	.0447	.0906
	(.0201)	(.0439)	(.0538)	(.0613)
Temporal Group 3×Muslim	-.00529	-.00099	-.0216	.0502
	(.00834)	(.0284)	(.0404)	(.0531)
Temporal Group 4×Muslim	.0103	-.0176	-.0107	.0146
	(.0156)	(.0228)	(.0388)	(.047)
$\overline{Y}_{0,PRE}$	.00608	.0308	.0625	.106
Adjusted R^2	.00135	.00544	.0153	.0225
Observations	7,257	7,257	7,257	7,257

Notes: *** denotes significance at 1%, ** at 5%, and * at 10%. Eicker-Huber-White standard errors in parentheses. The base period for the temporal difference is all those rated after September 11, 1998 and subsequently released between September 11, 1998 and September 10, 2000, thereby allowing a full year for potential recidivism prior to 9/11/01.

While the estimates are noisy, they are consistent with potential increases in recidivism risk for Muslims after 9/11, particularly for the inmates who were already released when 9/11 occurred (group 2 and 3). We therefore cannot rule out the possibility that the parole board, post-9/11, had a statistical-based justification for making harsher parole decisions for Muslim inmates.

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

E Additional Outcomes

E.1 Sample Size Calculations

Table E1: Sample Size Calculations Using Our DD Estimate as the MDE

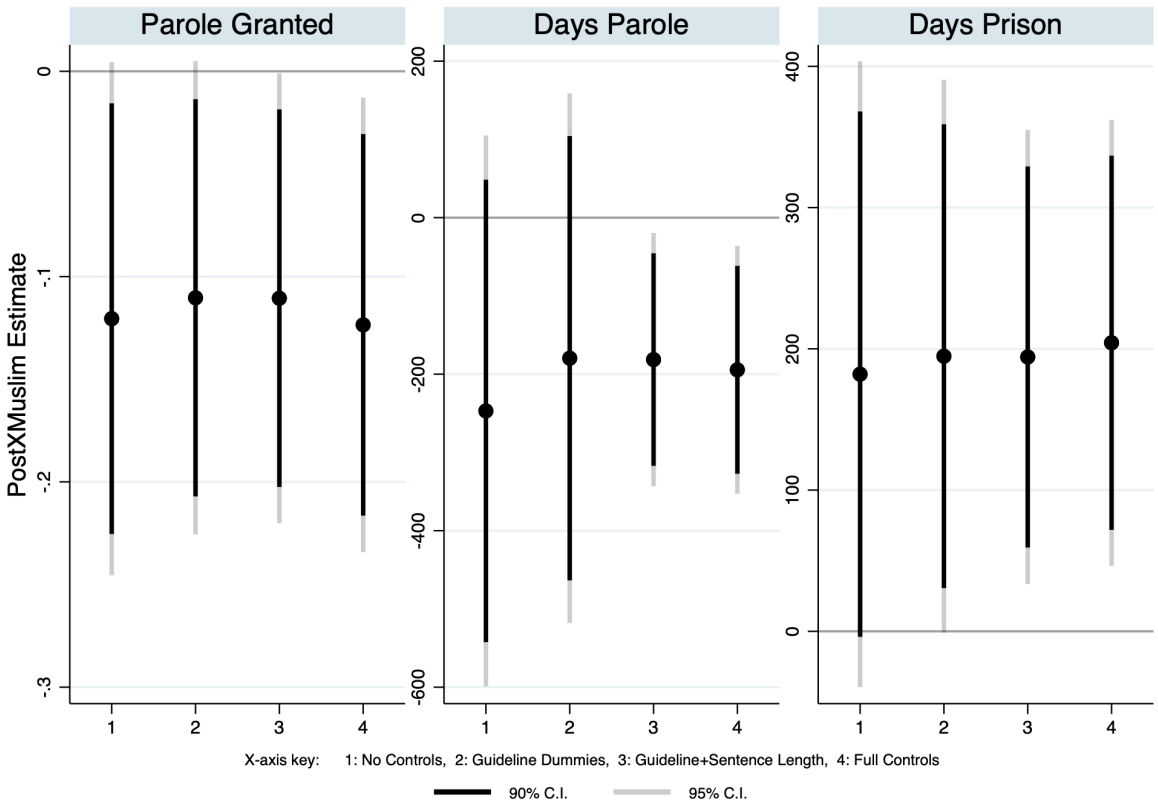
	(1)	(2)	(3)
	Parole Granted	Days Parole	Days Prison
Standardized DD Estimate	.2765	.1456	.2575
p	.04594	.04594	.04594
R^2_{YX}	.3038	.8437	.5608
R^2_{TX}	.01776	.01776	.01776
Actual Sample Size	4832	4832	4832
Required Sample Size	7404	5684	5200
Required Sample Size if p=.5	1268	974	890

Notes: Required Sample Sizes based on achieving a power of .80 and a significance level of .05. Sample size calculations were performed using the Shiny R Dashboard Program Power_Panel, supplied by Peter Z. Schochet, which accompanies Schochet (2022). As inputs, we use the Standardized DD Estimate for the minimum detectable effect (MDE), p the proportion of the sample treated (that is Muslim), as well as two R^2 measures – R^2_{YX} and R^2_{TX} – that respectively capture the explanatory power of our covariates on the outcomes and the DD term.

E.2 The Impact of Controls on the DD Estimates

In Figure E1, 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.

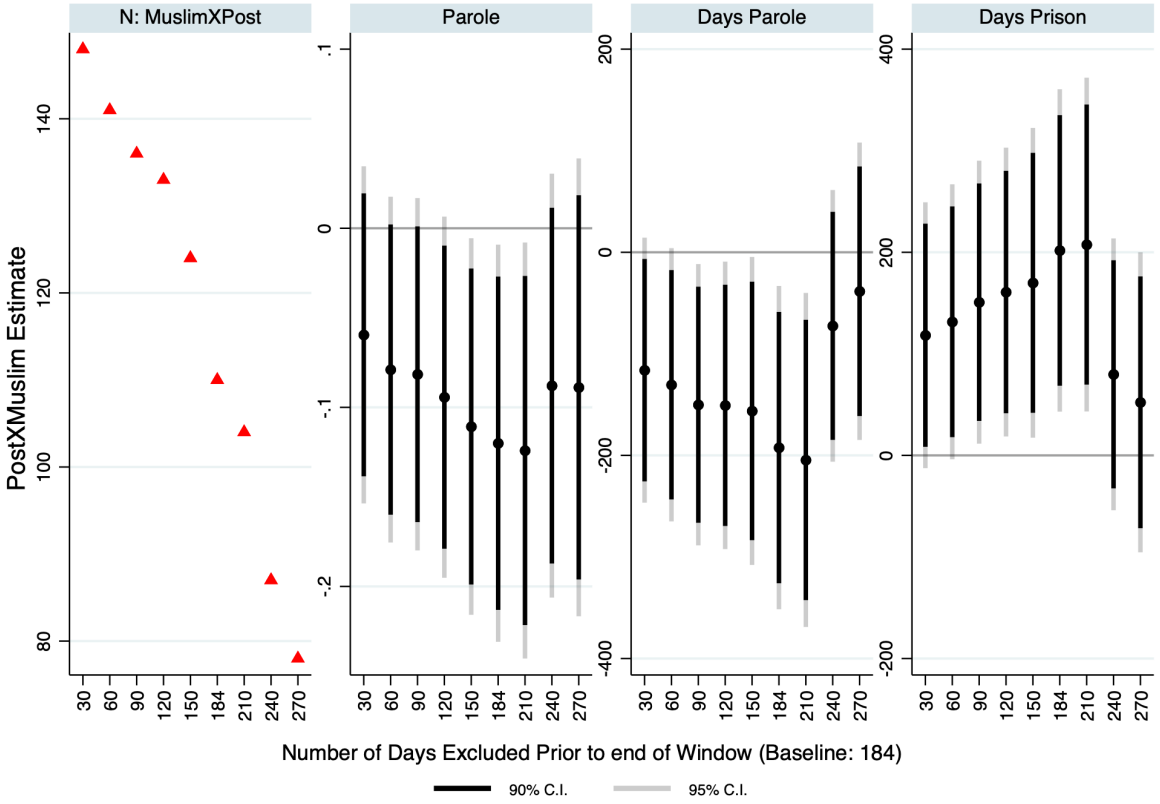
Figure E1: The DD Estimate is Stable as we Sequentially Include Controls



E.3 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 E2: The DD Estimate is Stable Based on Sample Selection Decision Regarding Exclusion Window



E.4 Randomization Inference Monte Carlo Simulations

Table E2: Monte Carlo Simulations Confirm That RI p-values are Half the Size of the Regular p-values

	(1)	(2)	(3)
Proportion Muslim	p-value	RI p-value	Ratio RI-to-regular
.046	.17	.082	.47
.05	.15	.073	.47
.1	.049	.023	.46
.25	.004	.0018	.45
.4	.0016	.00075	.48
.5	.00091	.00048	.53

Notes: Monte Carlo results based on 1,000 runs. For each Monte Carlo run, the RI p-values are based on 1,000 runs.