Three Lions or Three Scapegoats: Racial Hate Crime in the Wake of the Euro 2020 Final in London

Christof Nägel, Mathijs Kros, Ryan Davenport

Christof Nägel, Insitute of Sociology & Social Psychology, University of Cologne,
naegel@wiso.uni-koeln.de
Mathijs Kros, Department of Sociology, Utrecht University,
m.kros@uu.nl
Ryan Davenport, University College London
ryan.davenport@ucl.ac.uk

Accepted for publication in: *Sociological Science*

Acknowledgments: We are grateful to Frank van Tubergen, Huyen Nguyen, Jan-Willem Simons, Eva Jaspers, Chloé Lavest, Lucas Drouhot, Jeffrey Mitchell, Malcom Fairbrother, and Alexandra Heyden for feedback on an earlier version of this article.

Reproducibility statement: A package to reproduce the results presented in this article is accessible at https://osf.io/hzuqk/.

Abstract

Does (under-)performance of athletes from stigmatized racial groups influence the incidence of racial hate crimes? We consider the case of the English national football team during the 2020 European Football Championship and analyze whether the performance of black players during the final at Wembley affected the number of racial hate crimes committed in London. The three English players who missed their penalties in the final are all black English players. Combining insights from (displaced) frustration-aggression and scapegoat theory, we argue that the frustration of losing the final resulted in violence directed at racial minority group members in London. Our findings show that the lost final triggered a 30 percent increase in racial hate crimes in the weeks following the event. The immediate impact was larger in boroughs with higher pre-event levels of racial hate crimes, indicating a galvanizing instead of a mobilizing exacerbation of this trigger event.

Keywords: hate crimes; intergroup relations; scapegoat theory; prejudice; Euro 2020

Does the performance of athletes from black and ethno-racial minority groups influence the incidence of racial hate crimes? We study this question in the case of the English national football team during final of the 2020 Union of European Football Associations (UEFA) European Football Championship, which took place at Wembley on July 11, 2021, and analyze whether losing the final affected the number of racial hate crimes in London. Additionally, we study the structural distribution of violence after the Euro final across boroughs within London. In particular, we test whether the event triggered an increase in racial hate crimes in boroughs with a history of racial violence, or rather mobilized boroughs that hitherto were not prone to racial violence. Following Legewie's argument that discriminatory behavior is not bound to arise from static conditions, this study contributes to a "small but growing line of research in criminology that uses events to better understand the mechanisms that trigger the diffusion of intergroup violence" (Legewie 2016:381).

The England football team advanced further in Euro 2020 than any other English team has in an international tournament since the 1966 World Cup final, which was England's last and only such triumph. The Three Lions' Euro 2020 run ended in the final, in which they lost to Italy following a penalty shootout. The Italian squad scored three penalties, whereas England scored only two. The three England players Marcus Rashford, Jadon Sancho and Bukayo Saka, who missed their penalties, were abused online with comments targeting their black skin color. Twitter UK stated that their automated tools identified and removed 1,622 racist Tweets during the final and the 24 hours that followed (Twitter UK 2021). The UK Football Policing Unit launched a hate crime investigation into these racist comments, which led to at least 11 arrests (NPCC 2021). The anger expressed in these tweets was not just directed at these three players, but at black people in general. As one person put it: "These [black people] have lost us the EUROs," which was followed by a call to "punish" black people (Newsable 2021). In one sentence, this captures the frustration of losing the final, demarcates black people as an out-group separate from "us," and faults and seeks to victimize a minority group for the loss.

There is growing empirical consensus that hate crimes spike after "trigger events," such as 9/11 and the election of Donald Trump (Disha, Cavendish, and King 2011; Edwards and Rushin 2018). In the United Kingdom and London in particular, previous studies have shown that hate crimes increased after the Brexit referendum (Devine 2021), 9/11, and the 7/7 bombings in London (Hanes and Machin 2014), and that British Muslims experienced more indirect and overt anti-Islamic discrimination in the wake of 9/11 (Sheridan 2006). We extend this line of research by considering the Euro 2020 final as a "trigger event" of racial hate crimes in London, thereby furthering our understanding of the type of events that may cause a surge in hate crimes.

Previous research has also shown that exposure to athletes who perform exceptionally well can reduce prejudice toward the athletes' minority group (Alrababa'H et al. 2021). In particular, the phenomenal success of football player Mohammed Salah at Liverpool F.C. has been shown to be related to less Islamophobic attitudes and fewer instances of hate crimes and speech amongst Liverpool fans. Others have shown that anti-immigrant attitudes are less pronounced in regions in Spain where local soccer clubs have a relatively high percentage of immigrant players (Lago and Lago-Peñas 2021). However, this kind of exposure does not have to be positive. Athletes can underperform or disappoint fans and supporters, which can instill strong sentiments of anxiety, anger, and rage.

To test our assumptions, we use temporally granular data on recorded racial hate crimes from the London Metropolitan Police Service (MPS) pertaining to all London boroughs. Methodologically, we combine a Regression Discontinuity in Time (RDiT) analysis with fixed-effects panel regressions. Although the former allows for robust causal inferences under relatively weak assumptions, the latter provides more flexibility to study heterogenous treatment effects among London boroughs. The remainder of this article is structured as follows: we will explain out theoretical rationale, lay out both our underlying data and applied methods in detail, present our results and robustness checks, and finally discuss our findings in light of extant research.

Theory

Scapegoating and Frustration-Aggression

It is a basic finding of research on aggression that frustrating and provocative events can trigger strong sentiments of anger, which have the potential to result in aggressive and violent behavior. Frustration arises if someone or something—an externality—interferes with someone's goal-directed behavior (Harris 1974). Aggressive behavior is a way to remove this interference (Dollard et al. 1939; Miller 1941). Although aggression can be directed at the source of the frustration, research also shows that aggression may be displaced and be redirected at other people who had nothing to do with whatever instilled the frustration in the first place (Krahé 2013; Marcus-Newhall et al. 2000).

Frustration-aggression theory and the notion of displaced aggression inspired early works on intergroup prejudice and violence (Allport 1979). Whereas individual frustrations lead people to blame other individuals for their problems, social identity theorists (Tajfel 1981) hypothesized that shared, group-level frustrations would predispose people to blame other groups (Dovidio 2013). Although displaced aggression typically describes situations between individuals, the scapegoating theory of prejudice argues that an out-group, typically marginalized, is blamed and often punished for another group's misfortune that is actually due to other causes (Glick 2005, 2008; Kudrnáč, Eger, and Hjerm 2017). Following Allport's (1979) original ideas, scapegoating can be seen as a way to maintain a positive image of their own group, by minimizing feelings of responsibility, guilt, and inferiority (Rothschild et al. 2012).

Furthermore, scapegoating not only becomes more likely in the case of shared, group-level frustrations, but particularly when there is a salient out-group of lower social standing (Dovidio 2013; Kudrnáč et al. 2017). Experimental research shows that majority group members systematically shift punishment for harm done to an in-group member to members of a minority group (Bauer et al. 2021). Intergroup aggression, like other forms of intergroup behavior, relies on salient racial identities. For instance, if it is not evident that the players are of a stigmatized, racial minority, their performance is unlikely to affect people's attitudes, sentiments, and behavior toward the racial group as a whole (Alrababa'H et al. 2021; Schiappa, Gregg, and Hewes 2005). Racial minority out-groups, conveniently available as scapegoats, can make interracial aggression a more likely response to collective frustration.

Of course, not all frustration is answered with aggression, but such a response becomes more likely if frustration is coupled with strong feelings of anger (Berkowitz 2008). The literature on hate crimes, or offenses in which victims are targeted because of their group membership, also highlights the importance of anger (Kros, Jaspers, and Van Tubergen 2023). Hate crimes are seen as acts born out of strong emotions and may be justified or instilled by ideologies and prejudicial attitudes that marginalize or derogate victims because of their race, ethnicity, religion, or sexual orientation (Krahé 2013). These hate crimes occur when "feelings of anger and rage dominate the individual's rational decision-making process" (Hanes and Machin 2014:250). Other research further emphasizes the importance of anger in intergroup aggression. For instance, Parrott and Peterson (2008) have shown that intergroup anger is a particularly strong predictor of antigay aggression.

Combining insights from (displaced) frustration-aggression and scapegoat theory, racial hate crimes can then be seen as a way to unfairly blame and punish minority group members for a frustrating event in order to alleviate anger and protect the positive image of the in-group (Glick 2005; Rothschild et al. 2012).

The Euro 2020 Final as a Trigger Event of Hate Crimes

From the previous discussion, we can distill several characteristics of events that can trigger spikes in racial hate crimes. We highlight two of these characteristics: a collectively frustrating and anger-inducing event and salient intergroup relations. We argue that Rashford, Sancho, and Saka missing their penalties during the Euro 2020 final against Italy aligns with both characteristics of a trigger event.

Firstly, the collectively frustrating nature of the event is easily understood. Discounting the wider performance and game, the public at large will see the inability of the England team to win on penalties as the principal reason for their loss in the final to the Italian national team. The loss thwarted a long-cherished goal of many Brits, and was therefore collectively frustrating. This is further accentuated by the important place that football has in the cultural sphere of the United Kingdom (Burdsey 2021). To illustrate, the final between England and Italy was the most-watched event on television in the United Kingdom in 24 years (UEFA 2021). It stands to reason that losing that nerve-wracking final after a grueling penalty shoot-out was immensely frustrating for supporters of the English team. This is especially true when considering the Three Lions' history of losing after penalties (Brischetto 2022). Coming short in a competition is often upsetting. In fact, research on the frustration-aggression hypothesis has suggested that failure in a competition can be particularly frustrating and can result in aggressive behavior (Burnstein and Worchel 1962). This also holds true when supporting a team people are passionate about, rather than playing sports themselves. Witnessing a favorite team lose can turn this passion into anger and even violence (Card and Dahl 2011). Furthermore, the sentiment expressed in the hate tweets after England lost the final also reflected a strong sense of intergroup anger (Newsable 2021). Losing the final can thus be seen as a collectively frustrating and anger-inducing event with a lot of cultural significance.

Secondly, the assigned ethnic identity of the three black players who missed their penalties became apparent in the aftermath of the final. Several famous football players of immigrant descent have mentioned that the inclusivity often induced during international sports tournaments, however brief, may also be contingent on their performance (Alrababa'H et al. 2021). Mesut Özil, a German football player of Turks descent, famously brought this point across by stating: “I am a German when we win and an immigrant when we lose” (Stanley-Becker 2018). This phenomenon does not occur for players who are part of their nation’s ethnic majority group. As soon as Rashford, Saka, and Sancho missed the penalties, their racial identity immediately became salient—despite them playing for the national team. Rather than three football players representing the national team and the Three Lions, they became scapegoats. Subsequently, the frustration of some British fans may have led to a generalization to all racial minorities, targeting them as the innocent victims of frustration-induced violence.

Taking into account the specific nature of the Euro 2020 final as a trigger event, and considering the previous research on (displaced) frustration-aggression, scapegoats, and hate crimes, we hypothesize that the number of racial hate crimes committed in London increased after England lost to Italy in the Euro 2020 final.

Distribution of Violence after Trigger Events

Although increases in hate crimes after trigger events have at times not been found to vary across locations (e.g., Disha et al. 2011), more recently scholars have argued that reactions to triggering events within communities might “depend on their previous beliefs about minority groups” (Frey 2020:687). Based on claims put forward by a Bayesian rationale (Mancosu and Ferrín Pereira 2021), Frey (2020) argues that comparable high-profile events should particularly affect those individuals with proimmigration attitudes, as those might update their priors in the aftermath. In support of this claim, he finds that massive sexual assaults committed on New Year’s Eve in Cologne, Germany by a large group of young males of North African and Middle Eastern appearance increased antirefugee violence particularly in those districts with relatively little prior hostility. Borrowing the terminology of Sniderman, Hagendoorn, and Prior (2004), we could expect a similar *mobilizing effect* in the case of the Euro 2020 final. That is, boroughs typically less inclined toward racial hate crimes might mobilize against ethnic minorities because of the collectively felt frustration after the final in order to retain a positive image of the in-group. This could occur as individuals update their pro–out-group attitudes in response to the shared frustration and find a suitable scapegoat. Given the history of underlying racism in British football fan culture (Burdsey 2021; Cleland and Cashmore 2014), the blow of seeing their team lose because three black players failed to score might have been enough to unleash violent negativity toward racial minorities that was not expressed previously. Following the idea of mobilizing trigger events, there should be less room for exacerbation in neighborhoods usually prone to racial tensions, as their priors were not updated but confirmed.

Contrary to this, and as put forward by Sniderman et al. (2004), a *galvanizing effect* would suggest trigger events to disproportionally affect those areas with preexisting local tensions between minority and majority group members. Attitudinal shifts toward minority groups following trigger events, such as terror attacks, have been noted to be particularly pronounced among individuals within the majority group with right-wing political attitudes (Jungkunz, Helbling, and Schwemmer 2019; Nägel and Lutter 2020). This observation supports the notion of galvanizing effects, indicating a reinforcement of preexisting beliefs within an anti-immigrant demographic. A galvanizing effect may increase the likelihood of violent outbursts after the event in boroughs with a history of ethnic hostility because it provides an opportunity to even more vehemently and violently express preexisting prejudices. This should be particularly strong in boroughs where a normalization of such acts has occurred through previous behavior. On top of that, preexisting hostility might go hand in hand with limited social control (Sampson, Raudenbush, and Earls 1997) to prevent or sanction racial hate crimes.

Aggregating these individual-level assessments to an ecological perspective and applying it to our particular case, we will test two opposing hypotheses. On the one hand, in line with a mobilizing effect, the spike in racial hate crimes following the Euro 2020 final could be more pronounced in boroughs with a limited history of interracial violence. On the other hand, in line with a galvanizing effect, the spike in racial hate crimes could be more pronounced in boroughs with a history of interracial violence.

Data and Methods

Data and Measures

We use official hate crime data from 2021 in London recorded by the Metropolitan Police Service. As per UK Home Office Counting Rules, the data set is a record of all such allegations reported to the MPS. These data contain accurate spatiotemporal markers indicating when (in minutes) and in what borough an offense was committed.[1] It is important to note that the data

[1] London has 32 borough councils alongside the City of London, a small area less than 3 km² in the city center. Law enforcement within the City of London is managed by the City of London Police, a separate police force distinct from the Metropolitan Police Service. Crime data from the City of London are not included in our data set because of this separation.

allow us to differentiate between the time the offense was recorded and the time it was committed based on the report of the victim. In this context, we are interested in the latter. The raw data set comprises 22,210 unique hate crime offenses. Hate crimes are defined by the UK Home Office as "…any criminal offence which is perceived, by the victim or any other person, to be motivated by hostility or prejudice toward someone based on a personal characteristic" (O'Neill 2017:2). Because there is no legal definition of hostility, according to the MPS, police officers apply an everyday understanding of the word that includes ill will, spite, contempt, prejudice, unfriendliness, antagonism, resentment, and dislike. Classifying an offense as a hate crime has real-world legal consequences. According to the Crime and Disorder Act 1988 and section 66 of the Sentencing Act 2020, prosecutors are allowed to apply an increased sentence to those convicted of hate crimes. The data we use do not include references to judicial outcomes. However, to date there is no better data source on hate crimes as the only alternative, the Crime Survey of England and Wales (CSEW), provides annual data only (Devine 2021).

Our main outcome variable is the number of "racial incidents." From the overall category of racial incidents, we carefully discard offense classes that should, in theory, remain unaffected. First of all, we filter out hate crimes that have occurred online because our theoretical argument focuses on offline forms of racist abuse. Furthermore, we omit racial hate crime–related burglaries, drug offenses, robberies, theft and handling, fraud, and forgery as well as sexual offenses. As such, our outcome exclusively includes "violence against persons," "criminal damage," and "other notifiable offenses." Because "other notifiable offenses" also include irrelevant offenses like stalking and blackmail as well as attacks on emergency workers, we excluded these subcategories as well.[2]

Other hate crime categories included in the data are antisemitic, disability-related, faith-related, homophobic, Islamic, and transgender hate crimes. These hate crimes are used as placebo checks as they should not have been affected by the English team losing the Euro 2020 according to our theoretical expectations.

Our "treatment" variable is coded as 1 for all days after the Euro 2020 final, which took place on July 11, 2021, and 0 for all days before the final. A "running variable" is also coded, which is simply a continuously running vector from 1 to 365 for all included days (see Methods section for details). We control for weekdays, as it is plausible to assume that various types of

[2] We do this to arrive at an outcome variable that clearly relates to our theoretical expectations. However, our main results are robust to either including or excluding all of the omitted offenses. This has to do with the fact that the discarded offenses only make up a small percentage of the complete data. In fact, "violence against persons" makes up about 93 percent of all recorded offenses (16,459 of 17,680 racial hate crimes) in the raw data, followed by "criminal damage" (679 of 17,680 racial hate crimes).

crime—including hate crime— surge during the weekend when more people are on the streets, thus increasing number of suitable targets (Cohen and Felson 1979). Also, in line with routine activity theory, we control for average daily temperature in Fahrenheit in London. Temperature, unlike rainfall or hours of sunshine, has been associated with higher levels of violent crime, as higher temperature will increase the number of people spending time outside, which also increases the risk of criminal victimization (Field 1992).

Methods

Regression Discontinuity in Time. Our main focus lies on an RDiT, which has been used in similar sociological and criminological studies to study the impact of high-profile events (Legewie 2016; Reny and Newman 2021; Zoorob 2020). It employs a completely data-driven approach to determine the time window for the analysis (i.e., temporal bandwidth), and can be subjected to the evaluation of critical assumptions. These assumptions are discussed in the Results section and online supplement.

Our RDiT design can formally be expressed as follows: let $T \in [0; 365]$, be the running variable (in days) and let the receipt of treatment be denoted by the dummy variable $D \in \{0, 1\}$, so that we have $D = 1$ if $T \geq c$ and $D = 0$ if $T < c$, where c is the date of the Euro 2020 final. We fit a weighted model at either side of c:

$$Y_i\,(0) = \alpha + D_i\tau + T_i\beta + \varepsilon_i \qquad (1)$$

for $D_i = 0$ and

$$Y_i\,(1) = \alpha + D_i\tau + T_i\beta + \varepsilon_i \qquad (2)$$

for $D_i = 1$

to calculate the local average treatment effect (LATE) as follows:

$$E[Y_i\,(1)|D_i{=}1] - E[Y_i\,(0)|D_i{=}0]. \qquad (3)$$

Note that the LATE is largely dependent on the definition of bandwidth h around the cutoff. Generally, the challenge is to find an optimal balance between precision and bias in defining h. Although larger bandwidths will yield more efficient estimates, the potential for bias is smaller for an h that is close to c. In our main models, we rely on the Imbens and Kalyanaraman (IK) method to obtain an asymptotically optimal bandwidth (in days) to estimate the LATE. The IK algorithm chooses the bandwidth for which the LATE, that is, the regression discontinuity (RD) point estimator, provides the smallest mean squared error (MSE). We further apply a triangular kernel over c that gives larger weights to days close to c. This shields against potential bias from collateral or unrelated events. For more details on the IK algorithm, see Imbens and Kalyanaraman (2012). The IK method is usually considered the most popular data-driven choice to determining the bandwidth (Thoemmes, Liao, and Jin 2017). However, to avoid model dependence, we also estimate RD models from a universe of 90 different bandwidths. As recommended in the econometrical literature (Lee and Card 2008), we cluster standard errors on the distinct values of T, which are the days before/after the event in all following RD specifications.[3] Following Hausman and Rapson (2018), we include a lagged term of the dependent variable to account for potential autoregression. For our first set of analyses, we run the models on a borough panel of the data set in which we have borough*days (=11,680) observations. This allows us to assess the causal effect of the Euro 2020 final on daily racial hate crimes within London boroughs. We extend these analyses by RD models run on the entire time series to estimate an overall effect in London.

Geographical heterogeneity of treatment effects. We will study the geographical heterogeneity in two different ways. First, we compute prior hostility by taking the mean of racial hate crimes before the Euro 2020 final in each borough. We then rerun the RD models for each borough and correlate the treatment effect with the previously constructed prior hostility measure. A negative correlation would suggest a mobilizing effect, and a positive correlation would indicate a galvanizing effect. We will also control for percentage of black people living in the boroughs with data we downloaded from the London Datastore. As a more formal second test, we run a fixed-effects regression model with standard errors clustered on the borough level. We construct a treatment dummy that takes the value 1 for the optimal bandwidth determined through the previous RD models, and 0 otherwise. We interact this dummy with the prior

[3] Clustering standard errors on the borough level yields similar results (not shown).

hostility measure. We will control for time-varying covariates such as seasonal trends with a time spline, weekday dummies, and average temperature. A significant negative product term would suggest a mobilizing effect, whereas a significant positive product term would be indicative of a galvanizing effect.

Results

Local-Linear RDiT Analyses

Table 1. Regression discontinuity models with and without controls.

	Racial hate crimes	
	Model 1: without controls	Model 2: with controls
LATE	0.53***	0.49***
Cluster-robust SE	0.14	0.13
Cluster-robust CI	[0.26, 0.79]	[0.23, 0.75]
N	11,680	11,680
Effective *N*	2,272	2,464
Bandwidth	35.93	38.81
Borough fixed effects	Yes	Yes
Controls	No	Yes

Notes: LATE is based on triangular kernel with optimal bandwidth.[4] ***$p < 0.001$. *N* relates to borough days.

Table 1 gives the point estimates for the LATE in both a model with and without controls. The base model suggests an average increase of 0.53 more daily hate crimes in the first five weeks within a borough ($p < 0.001$, Cluster-Robust CI [0.26, 0.79]) compared with the pre-event period. This suggests an average of more than half a hate crime more in each borough after the final. The model with controls suggests a similar LATE with only marginal efficiency gains. Because we are interested in the overall treatment effect in London, we converted the data to a time series format in which individual days are the units of observations

[4] The optimal bandwidth choice can lead to incorrect standard errors that can be fixed by bias correction methods (Cattaneo and Vazquez-Bare 2017). Applying these alternative algorithms in our case leads to smaller bandwidths (28 days), and a somewhat smaller point estimate ($\tau = 14.572$) that still is statistically significant with $p < 0.001$.

and reran the models.[5] Figure 1 provides the visual representation of the treatment effect of the Euro 2020 final on daily racial hate crimes in all of London. The effect is positive and significantly different from zero for $p < 0.001$ ($\tau = 16.64$, Cluster-Robust 95 percent CI [8.18, 25.39]). Compared with the pretreatment time period, which encompasses the 35 days before the event, there is an average increase of 16 daily hate crimes after the final compared with the pretreatment period in the posttreatment period (the following 35 days). This translates to an excess of 30 percent.

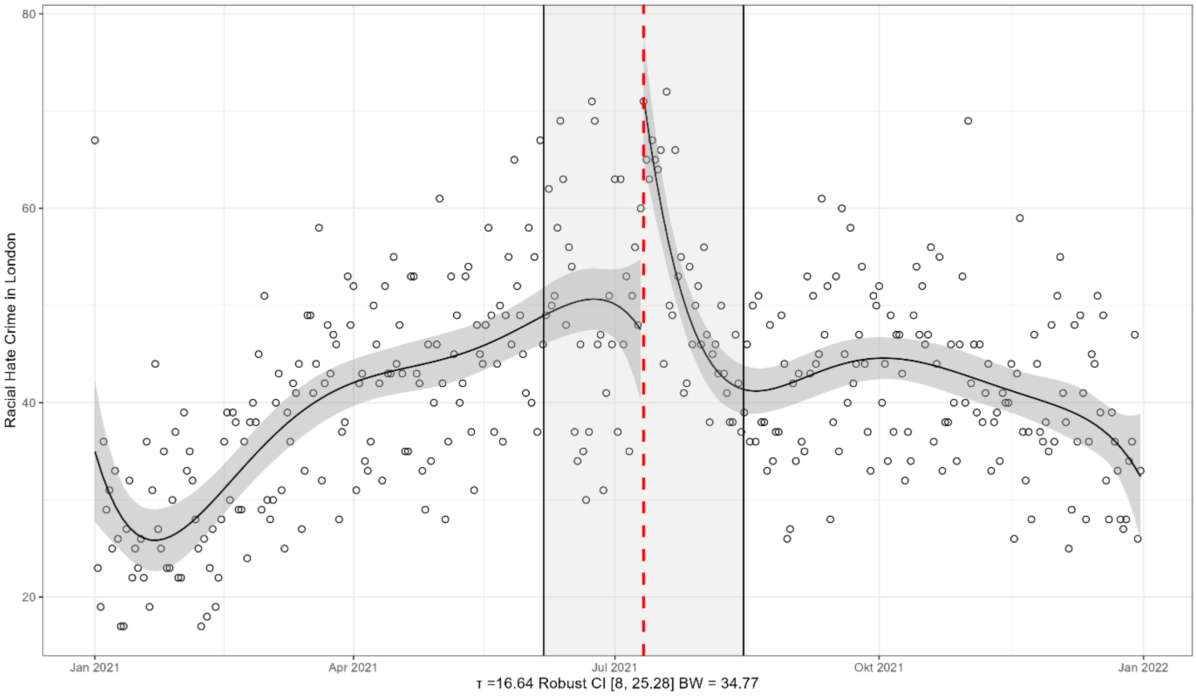

Figure 1. Discontinuity graphic for racial hate crimes in London. Red line indicates date of the final. X values are bound to the mean. Optimal bandwidth is indicated by the grey area.

[5] Results from subsequent RD models are virtually the same (although interpretation differs) in both panel format as well as daily data format.

RDiT Assumption and Robustness Checks

The validity of our RD design requires that all other potentially relevant factors, apart from the treatment and outcome variables, must exhibit continuity at the specific point where the treatment and outcome discrepancies emerge. There are a number of potentially biasing factors to consider. The spike we see could be part of collateral events, a sudden jump in temperature, a seasonal pattern, random noise, or just model dependence. We test for these possibilities in a number of ways. First, we assess placebo effects on other outcomes that should theoretically not be affected by the Euro 2020 final. To do so, we analyze data on other hate crime categories. We run models with the exact same specification as shown below Figure 1 for these outcomes. Results can be seen in Figure A1 in the online supplement. None of these outcomes are significantly or substantially affected be the day of the event. Although controlling for average daily temperature should account for potential biases due to weather changes, we also explicitly run an RD model on the temperature covariate. We find that temperature rises discontinuously after the Euro 2020 final by 4.7 °F (τ_p = 4.76, Cluster-Robust CI [0.47, 9.05], p = 0.03); see Figure A2 in the online supplement for a visual display. Although we cannot exclude the possibility that average temperature might have partially driven the effect on racial hate crimes, controlling for temperature does not level out the treatment effect in any way. It is furthermore not unreasonable to assume that racial hate crimes follow a yearly, seasonal pattern. Seasonal patterns have been observed for all major crimes rates (McDowall, Loftin, and Pate 2012). Although this increase might also be a function of loosening the COVID lockdown restrictions over the first half of the year, a seasonal spike in the summer could lead us to erroneously interpret a nonlinear relationship as a discontinuity. We assessed this possibility by rerunning the exact same model as shown below Figure 1 for the years 2017 and 2018.[6] Results are presented in Figure A3 in the online supplement. Although a certain seasonal pattern is visible, the models reveal that, in 2017 and 2018, July 11 as a cutoff point implies a nonsignificant (negative) effect. Racial hate crimes therefore appear to generally follow a downward trend from that moment on, which even suggests a potential downward bias in our estimates on the 2021 data.

If we were to observe unexpected discontinuities at different points in time during 2021, this would cast doubt on the validity of our design. In the RD literature, it is common to test for the absence of discontinuities in the outcome after fake thresholds. We assess *all* possible

[6] Hate crime date were only available to us for these two additional years.

cutoffs in the entire data. Figure A4 in the online supplement graphs the effects of all possible "fake" cutoffs. We find effect sizes that are different from zero for $p \leq 0.05$ in only 2.25 percent of these placebo RD models. These effects likely represent random noise, given that they are scarce and much smaller in size than the LATE of our RD model with July 11 as the cutoff point. We also systematically assessed each effect that was significantly different from zero and checked whether there were any events that might have caused them. A small spike in September might be related to the fuel supply crisis, but this is speculative. We did not find any potential reasons for the significant decreases in the pretreatment period that are depicted in Figure A4.

As explained above, we use an automated bandwidth algorithm. However, other bandwidth choices could also be valid. Therefore, we assess the LATE from a universe of 90 potential bandwidth choices, while holding constant the model specification. The results can be seen in Figure A5 in the online supplement. Clearly, the LATE only becomes dependent on the bandwidth for very large analysis windows, bolstering our confidence that the positive effect is unlikely to be purely a result of our modelling choices. In the online supplement, we also provide different estimation strategies like a parametric RD that is based on a Bayes Information Criterion (BIC)-chosen polynomial (Figure A6) as well as an RD analysis based on an hourly aggregated data set (Figure A7). Both estimation strategies yield results that are in line with those from our main models.

The RD design is generally not well suited to study the precise duration of treatment effects in event studies, because observations close to c are given larger weights (either by a nonrectangular kernel in the local linear approach, or by the order of the polynomial in the parametric approach). To study the duration of the effect more precisely, we therefore include an Interrupted Time Series analysis in online supplement A2. This additional analysis suggests a significant increase in racial hate crimes for a week after the final, with inconsistent significant increases in the following subsequent week. After two weeks, there is no evidence for an increased number of racial hate crimes.

Finally, we also ran some additional robustness checks that are specific to our case study. We tested whether football games in Wembley generally affect the number of racial hate crimes and whether we see any effects of the COVID relaxations that occurred throughout the first half of 2021. We do this by running the exact same RD specification as shown above on those placebo cutoff dates. With the exception of the Euro 2020 final, football games at Wembley do not seem to influence the number of racial hate crimes committed in London (see

Figure A11). Reopening policies similarly do not appear to have changed the number of racial hate crimes (see Figure A12). Drawing from criminological strain theories (Agnew and Brezina 2019), the rise in racial hate crimes also might be part of a broader increase in offenses, potentially fueled by frustration among English fans following the team's loss in the final. Our data only contain offenses classified as hate crimes, making this alternative explanation hard to assess. However, the London Datastore provides open data on a broad range of crime types. Given that these data are only available on a monthly basis, we cannot replicate our discontinuity analysis. Still, we are able to visually juxtapose the racial hate crime spike in July (consistent with Figure 1) with other related crime offenses in Figure A13 in the online supplement. Although a few crime types reach a seasonal peak in July (such as public order offenses, violence against the person, and domestic abuse),[7] the relative spike in racial hate crimes appears to be quite unique. Even though this visual analysis implies that other types of crime may not have increased similarly to racial hate crimes after the final, there could still be smaller spikes around the cutoff date, which we are unable to see in monthly crime counts.

Geographical Heterogeneity of Treatment Effects

Figure 2(A) provides a London map depicting the LATE for an RD model (without additional controls) in each borough. Effect sizes range from around –1.5 to 2.5, suggesting considerable between-borough variation, but no evident spatial clustering. As Figure 2(B) suggests, there is a strong positive relation between the size of the LATE from RD models and the daily mean of racial hate crimes before the Euro 2020 final (r = 0.36, p = 0.04). This is indicative of a galvanizing heterogenous treatment effect: hate crimes were more numerous in boroughs that had history of interracial violence before the Euro 2020 final. The correlation between percentage of black people living in boroughs and the LATE is smaller than 0.1 and nonsignificant. Next, we ran a fixed-effects regression as specified above. Model 4 in Table 2 suggests that the interaction is positive and significantly different from zero, again supporting galvanizing effects. Essentially, the product term in model 4 suggests that in boroughs with an already higher-than-average number of racial hate crimes, there is an expected average increase

[7] Because most racial hate crimes in our data are incidents of "violence against the person" that are flagged to be racially motivated, the juxtaposition with all "violence against the person" offenses is a difficult comparison. This is due to the fact that racial incidents are also counted in the general count for "violence against the person" offenses. The data provided by the London Datastore do not allow differentiation between these offenses.

of about half an additional racial hate crime per day. In boroughs with a below-average prior-hostility measure, the average increase amounts to only one-tenth of a racial hate crime.[8] The spike in hate crimes following the Euro 2020 final thus appears to be stronger in boroughs with a history of racial violence.

These results are more robust to choosing other treatment dummies than the 35 days of the RD optimal bandwidth. Figure A14 in the online supplement shows how far the product term between prior hostility and the treatment dummy depends on the definition of the treatment by daily increments. Furthermore, results are not sensitive to using different ways to compute the prior hostility measure (e.g., the prior median instead of the prior mean in a given borough), log-transforming the outcome, or using nonlinear regression such as negative binomial regression for estimation (see Tables A1 and A2 in online supplement A3). All in all, there is evidence for heterogenous treatment effects clearly favoring the notion of galvanizing as opposed to mobilizing exacerbation.

Table 2. Fixed effects regression models

	Racial Hate Crimes			
	Model 1	Model 2	Model 3	Model 4
After Euro 2020 final (OBW)	0.33***	0.10*	0.33***	0.10*
	(0.04)	(0.04)	(0.04)	(0.04)
After Euro 2020 final (OBW) * Prior Hostility			0.38***	0.38***
			(0.07)	(0.07)
R^2	0.01	0.02	0.01	0.03
Adjusted R^2	0.00	0.02	0.00	0.03
F Statistic	79.43***	97.24***	53.60***	35.87***
Borough fixed effects	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes

Notes: * p<0.05; *** p<0.001, N = 11,680 in all models, Cluster-Robust Standard Errors in brackets, OBW = optimal bandwidth from RDiT Model 1 in table 1 (35 days after the final), interacted vectors are mean centred, additional controls include a linear time spline, average temperature, and weekday dummies
.

[8] Note that the RD estimates in Table 1 are not directly comparable to the slope estimates in Table 2, because the former apply larger weights to observations that are temporally close to the day of the event.

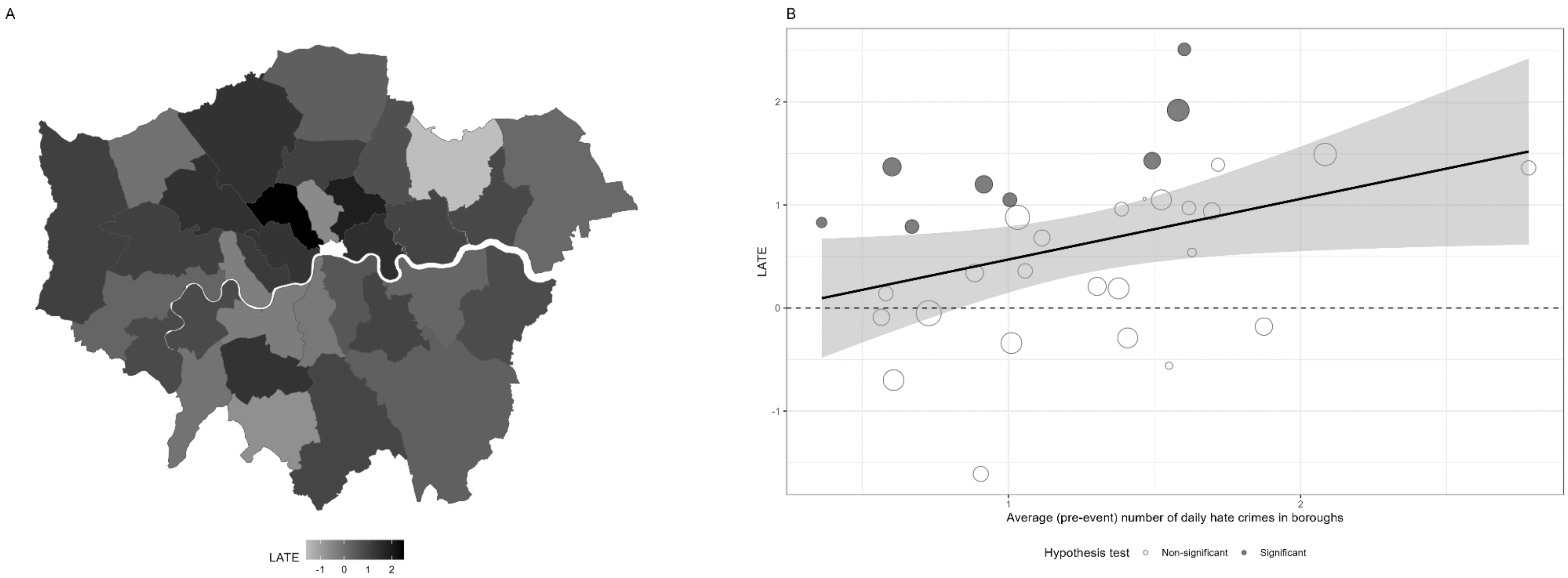

Figure 2. (A) Spatial distribution of LATEs from RD models in London boroughs. (B) Relation between average number of daily hate crimes in boroughs before the event and the treatment effect size in RD models. Circle size represents the percentage of black people living in the respective borough.

Discussion

We set out to study whether racial hate crimes in London spiked in the wake of the Euro 2020 final, which England lost after penalties. Given that the three English players who missed their penalties were black, we argued that the final was a particularly frustrating event that triggered strong sentiments of anger and aggression, with a salient racial component. We hypothesized that this event would result in scapegoating, in which racial minorities would be unfairly blamed, increasing the number of registered racial hate crimes in London in the aftermath.

In short, we find a surge in racial hate crimes in London of about 30 percent in the first five weeks after the final. The immediate increase in racial hate crimes is not found in any of the other hate crime categories and cannot be convincingly attributed to alternative explanations. The initial increase in racial hate crimes in the first weeks after the final is substantial and appears robust. We add to the growing literature on trigger events and hate crimes by showing that sport tournaments can also be trigger events. This is an important contribution, as these sports events have mainly positive attributions (Vallerand et al. 2008) unlike trigger events that have so far been studied, such as 9/11, the 7/7 bombings in London, and also the Brexit referendum (Devine 2021; Hanes and Machin 2014; Sheridan 2006).

We can further conclude that the performance of minority athletes can be a double-edged sword in relation to hate crimes. Previous research has shown that exposure to out-group athletes who perform exceptionally well, like Mohammed Salah at Liverpool F.C., can reduce prejudice and hate toward the athletes' minority group (Alrababa'H et al. 2021). However, our results demonstrate that underperformance of athletes may also trigger an increase in racial hate crimes. The fact that this might not be limited to the case of the Euro 2020 and London became painfully apparent after the more recent World Cup 2022 final in Lusail, Qatar. France lost the final against Argentina, after a penalty shootout. The French players Kingsley Coman and Aurélien Tchouaméni missed decisive penalties and were subsequently targeted with racist online abuse (Hill 2022). These sentiments largely align with the notion of scapegoating, and demonstrate the ubiquity of racialized resentment expressed after the (alleged) underperformance of minority football players.

Moreover, our study shows that racial hate crimes can already spike after one unfortunate match. Research on the "Salah effect" raised the issue that sustained exposure, across several football seasons, to a well-performing athlete might be required to reduce prejudice (Alrababa'H et al. 2021). Although it could still be true that sustained parasocial contact is necessary to bring about more *long-term* changes in hate and prejudice, our study is in line with the idea that a single significant sports moment, like losing a final, can already result in measurable changes in racial hate crimes. Still, the cultural significance of the Euro 2020 final for Great Britain cannot be overstated. This implies that the dramatic increase in racial hate crimes we see in our data might be less generalizable to other, less important sports events.

That said, the chosen bandwidth algorithm produced a relatively short optimal bandwidth of five weeks to estimate the treatment effect, and excess hate crimes clearly decreased before that. A simple look at Figure 1 suggests a visible drop after about two weeks. Indeed, the discontinuity in our models appears to be driven by about eight days after the cutoff. This is in line with the interrupted time series analysis in online supplement A2 and suggests a short-lived effect. Comparing these results with those from other studies, the EURO 2020 final's effect on hate crimes appears to be not as persistent as Brexit's effect (~15 days) on hate crimes, but more durable than the effect of some domestic terror attacks such as the London Bridge attack or the Manchester Arena attack (Piatkowska and Stults 2022)**.** This is also consistent with other research on hate crimes in the United Kingdom, which has shown that trigger events only (Devine 2021) or mostly (Hanes and Machin 2014) result in increases in hate crimes immediately after the event. Broadly speaking, this indicates that trigger events may instill frustration, anger, and aggression, but these emotions also subside after some time. However, for the specific case of the Euro 2020, final other factors might have contributed to the curbing of the surge in hate crimes. The online hate speech directed at the three English players also prompted swift and widespread condemnation. Team captain Harry Kane, coach Gareth Southgate, Prince William, then–prime minister Boris Johnson, and other public figures all denounced the racist comments and voiced their solidarity with the black players (Sullivan 2021). Murals were painted in support of Rashford, Sancho, and Saka (Eurosport 2021). Furthermore, white supporters might have been influenced by positive vicarious contact. This is a form of indirect contact in which people observe coethnics interacting positively with out-group members (Mazziotta, Mummendey, and Wright 2011). Observing Harry Kane support his teammates, Rashford, Sancho, and Saka, may have alleviated some of the prejudice and frustration amongst white football fans. Perhaps such positive vicarious contact and the public outcry made it clear to would-be perpetrators of racial hate crimes that their frustration was misdirected and out-group aggression was not condoned or accepted. Then again, previous research established that even the positive, inclusive sentiment of an overarching national identity and accompanied feelings of togetherness, spurred on by international sports tournaments, may themselves be rather short-lived and shallow (Burdsey 2021). Accordingly, the visible decrease of the effect might simply be due to potential perpetrators moving on from the frustrating event that was the Euro 2020 final.

When investigating the spatial heterogeneity of the treatment effect, our analyses uncovered a noteworthy trend: boroughs characterized by higher levels of preexisting hostility exhibited a greater surge in racial hate crimes following the event. This pattern aligns with a galvanizing effect rather than a mobilizing one.[9] Interestingly, although this finding confirms Frey's (2020) notion that trigger events do not have a uniform impact, it conflicts with his observations following the New Year's Eve sexual assaults in Germa

ny. In that context, districts with lower preexisting hostility experienced a greater increase in antirefugee violence after the focal event. We posit that these divergent empirical findings may be attributed to the impact of *threatening* events as opposed to merely *frustrating* ones. An event that highlights the potential danger associated with an out-group might be more likely to negatively affect beliefs of those who held positive attitudes toward these out-groups before. Although the Euro 2020 final was a collective cultural disappointment, it might not have been enough to mobilize individuals with prior positive attitudes. The New Year's Eve event in Germany, on the other hand, indeed might have "changed everything" (Hewitt 2016), especially within those German districts with little prior antirefugee hostility.[10]

Considering the galvanizing effects in our data, we believe that a history of high levels of racial hate crimes might be indicative of a normalization of such behavior within those boroughs. The Euro 2020 final, acting as a trigger, could have prompted individuals to express their prejudices more openly and violently, believing that their actions align with the prevailing sentiment within their communities. This is also in line with the notion that the online racist comments after the game might have created a window of opportunity for people with preexisting racist attitudes to express their prejudice. A similar phenomenon has been observed after jihadist terror attacks (Czymara et al. 2023). Online hate speech might have destigmatized offline hate crime, especially in those boroughs where those sentiments existed before the final. Again, this is in line with findings from other contexts, in which racially connotated events or

[9] It is also important to note that, although 24 boroughs indicated a positive LATE, a positive and statistically significant effect was observed in only eight of London's 32 boroughs (including Brent where Wembley stadium is located).

[10] Note that Schwitter and Liebe (2023) argue that the causal effect in Frey (2020) might be due to a policy change in the data collection. In a rejoinder, Frey (2023) reanalyzes these data that were not affected by this change and finds that although results generally remain comparable, they are smaller in size than in the original study .

policies might mobilize those individuals who already possess negative racial attitudes (Flores 2017; Spörlein and Schlueter 2021). Finally, in areas where racial hate crimes are more prevalent, there may be a lack of effective social controls or mechanisms to counteract such behaviors (Sampson et al. 1997). This absence of deterrents allows for the escalation of hate crimes after the event, as individuals feel emboldened to act on their prejudices without fear of significant consequences. If similar clustering of racial hate crimes is to be observed after comparable events, preventive efforts, like increased police presence, could then focus on neighborhoods that are generally prone to racial hate crimes.

Finally, our study has limitations to be kept in mind. First, our analyses rely on administrative data provided by the MPS, so the validity of our results is contingent on the accuracy of the underlying information. Second and relatedly, our analyses pertain to recorded hate crimes. It is therefore possible that the spike in racial hate crimes we witness from July 11, 2021 onward is due to an increase in reports of such crimes, rather than an increase in the actual number of racial hate crimes. Similarly, the hateful online comments against the English players and the subsequent arrests have contributed to a wider public discourse on racial hate crimes. We therefore cannot exclude the possibility that a larger number of crimes were coded by police officials as hate crimes because of this discourse. Third, we do not know whether the increase in racial hate crimes we observe might have been part of a general surge in violent crime and public order offenses after the Euro 2020 final. This general surge in violent behavior could be attributed to coping with the inability to achieve positively valued goals, such as losing a final (Agnew and Brezina 2019). For this, it is also important to keep in mind that the Euro 2020 final, on July 11, 2021, was the first large public event that took place in quite some time and was made possible by early relaxations of COVID regulations. Bottled-up frustration from the pandemic and the excitement of the start of the summer, coupled with alcohol consumption, could have also attributed to the increase in racial hate crimes. However, it is hard to imagine that any "heat of the moment" effects lasted for more than a few hours, and we do believe it is telling that we did not see spikes in any of the other hate crimes categories. On top of that, the COVID relaxations themselves appeared not to have any effect on racial hate crimes. Finally, although our visual analysis of monthly crime trends shows some slight increases for selected crimes in July, we find little evidence for a similarly sudden increase in other crime types than incidents coded as racial hate crimes. Fourth, an alternative explanation for our findings that cannot be completely ruled out is the observed increase in temperature that coincided with the lost final. Rising temperature goes hand in hand with more people on the street. This brings together both suitable targets and motivated offenders. Coupled with the potential lack of capable guardians, the spike could be due to a change of routine activities (Cohen and Felson 1979; Field 1992). Yet we would again argue that for this explanation to have any merit, we should also see an increase in other hate crime categories. Additionally, controlling for temperature did not level out the treatment effect in our models. Finally, as implied above, analyzing only one case study is a limitation and our findings might not travel easily to other idiosyncratic contexts. The 2022 incident in Qatar during the World Championship's final, as well as a comparable case in the German Under-21 team,[11] suggest the ubiquity of events that align with the notion of scapegoating. However, the context of the EURO 2020 final in London is extremely specific. Future research might analyze how these events generally affect offline crimes against minorities.

Still, our findings suggest that performance of minority players can be a double-edged sword. Although good performances, sustained over longer periods of time, can reduce prejudice, one moment of underperformance—even at the end of an immensely successful tournament—can lead to increases in racial hate crimes. Minority football players are Three Lions when England wins and scapegoats when things go sour. The increase in racial hate crimes we discovered after the Euro 2020 final is notable in size, and, importantly, reinforced intergroup conflict in those boroughs with greater racial hostility.

References

Agnew, Robert, and Timothy Brezina. 2019. "General Strain Theory." Pp. 145–60 in *Handbook on Crime and Deviance, Handbooks of Sociology and Social Research*, edited by Marvin D. Krohn, Nicole Hendrix, Gina Penly Hall, and Alan J. Lizotte. New York, NY: Springer International Publishing.

Allport, Gordon W. 1979. *The Nature of Prejudice*. Unabridged, 25th anniversary ed. Reading, MA: Addison-Wesley Pub. Co.

Alrababa'H, Ala', William Marble, Salma Mousa, and Alexandra A. Siegel. 2021. "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes." *American Political Science Review* 115(4):1111–28.

Bauer, Michal, Jana Cahlíková, Julie Chytilová, Gérard Roland, and Tomas Zelinsky. 2021. *Shifting Punishment on Minorities: Experimental Evidence of Scapegoating*. Cambridge, MA: National Bureau of Economic Research.

[11] A game against Israel in the European Under-21 Championship ended in a draw, after Youssoufa Moukoko and Jessic Ngankam, two players of German-Cameroonian descent, missed their penalties. Afterward the players were the target of vile racist social media attacks: https://tinyurl.com/8caaa86z.

Berkowitz, Leonard. 2008. “On the Consideration of Automatic as Well as Controlled Psychological Processes in Aggression.” *Aggressive Behavior* 34(2):117–29.

Brischetto, Patrick. 2022. “England Penalty Shootout History: Three Lions’ Record at World Cups and Other Major Tournaments Ending on Penalties.” *The Sporting News*, December 12.

Burdsey, Daniel. 2021. *Racism and English Football: For Club and Country*. Abingdon New York, NY: Routledge.

Burnstein, Eugene, and Philip Worchel. 1962. “Arbitrariness of Frustration and Its Consequences for Aggression in a Social Situation.” *Journal of Personality* 30(4):528–40.

Card, David, and Gordon B. Dahl. 2011. “Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior.” *The Quarterly Journal of Economics* 126(1):103–43.

Cattaneo, Matias D., and Gonzalo Vazquez-Bare. 2017. “The Choice of Neighborhood in Regression Discontinuity Designs.” *Observational Studies* 3(2):134–46.

Cleland, Jamie, and Ellis Cashmore. 2014. “Fans, Racism and British Football in the Twenty-First Century: The Existence of a ‘Colour-Blind’ Ideology.” *Journal of Ethnic and Migration Studies* 40(4):638–54.

Cohen, Lawrence E., and Marcus Felson. 1979. “Social Change and Crime Rate Trends: A Routine Activity Approach.” *American Sociological Review* 44(4):588.

Czymara, Christian S., Stephan Dochow-Sondershaus, Lucas G. Drouhot, Müge Simsek, and Christoph Spörlein. 2023. “Catalyst of Hate? Ethnic Insulting on YouTube in the Aftermath of Terror Attacks in France, Germany and the United Kingdom 2014–2017.” *Journal of Ethnic and Migration Studies* 49(2):535–53.

Devine, Daniel. 2021. “Discrete Events and Hate Crimes: The Causal Role of the Brexit Referendum.” *Social Science Quarterly* 102(1):374–86.

Disha, Ilir, James C. Cavendish, and Ryan D. King. 2011. “Historical Events and Spaces of Hate: Hate Crimes against Arabs and Muslims in Post-9/11 America.” *Social Problems* 58(1):21–46.

Dollard, John, Neal E. Miller, Leonard W. Doob, O. H. Mowrer, and Robert R. Sears. 1939. *Frustration and Aggression.* New Haven, CT: Yale University Press.

Dovidio, John F. 2013. “Bridging Intragroup Processes and Intergroup Relations: Needing the Twain to Meet.” *British Journal of Social Psychology* 52(1):1–24.

Edwards, Griffin Sims, and Stephen Rushin. 2018. “The Effect of President Trump’s Election on Hate Crimes.” *SSRN,* first published on January 18, 2018, as doi:10.2139/ssrn.3102652.

Eurosport. 2021. “Football News - Twitter Reveals Scale of Racism in the UK and beyond after England's Euro 2020 Final Defeat.” *Eurosport,* August 10.

Field, Simon. 1992. "The Effect of Temperature on Crime." *The British Journal of Criminology* 32(3):340–51.

Flores, René D. 2017. "Do Anti-Immigrant Laws Shape Public Sentiment? A Study of Arizona's SB 1070 Using Twitter Data." *American Journal of Sociology* 123(2):333–84.

Frey, Arun. 2020. "'Cologne Changed Everything'—The Effect of Threatening Events on the Frequency and Distribution of Intergroup Conflict in Germany." *European Sociological Review* 36(5):684–99.

Frey, Arun. 2023. "On the Effect of the New Year's Eve Sexual Assaults on Anti-Refugee Violence: A Rejoinder to Schwitter and Liebe (2023)." *European Sociological Review* 39(6):1011–17.

Glick, Peter. 2005. "Choice of Scapegoats." Pp. 244–61 in *On the Nature of Prejudice: Fifty Years after Allport*, edited by John F. Dovidio, Peter Glick, and Laurie A. Rudman. Hoboken, NJ: Blackwell Publishing.

Glick, Peter. 2008. "When Neighbors Blame Neighbors: Scapegoating and the Breakdown of Ethnic Relations." Pp. 121–46 in *Explaining the Breakdown of Ethnic Relations: Why Neighbors Kill*, edited by Victoria M. Esses and Richard A. Vernon. Hoboken, NJ: Wiley.

Hanes, Emma, and Stephen Machin. 2014. "Hate Crime in the Wake of Terror Attacks: Evidence from 7/7 and 9/11." *Journal of Contemporary Criminal Justice* 30(3):247–67.

Harris, Mary B. 1974. "Mediators between Frustration and Aggression in a Field Experiment." *Journal of Experimental Social Psychology* 10(6):561–71.

Hausman, Catherine, and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10(1):533–52.

Hewitt, Gavin. 2016. "Cologne Attacks' Profound Impact on Europe." *BBC News*, January 11.

Hill, Glynn. 2022. "Three French Players Face Racist Online Abuse after World Cup Loss." *The Washington Post*, December 20.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *The Review of Economic Studies* 79(3):933–59.

Jungkunz, Sebastian, Marc Helbling, and Carsten Schwemmer. 2019. "Xenophobia before and after the Paris 2015 Attacks: Evidence from a Natural Experiment." *Ethnicities* 19(2):271–91.

Krahé, Barbara. 2013. *The Social Psychology of Aggression*. 2nd ed. New York, NY: Psychology Press.

Kros, Mathijs, Eva Jaspers, and Frank Van Tubergen. 2023. “Longitudinal Changes in Interracial Hate Crimes in the USA, 1990–2014: Does Racial Composition Matter?” *Journal of International Migration and Integration* 24(2):547–66.

Kudrnáč, Aleš, Maureen A. Eger, and Mikael Hjerm. 2017. “Scapegoating Immigrants in Times of Personal and Collective Crises: Results from a Czech Panel Study.” *International Migration Review* 58(2):573–92.

Lago, Ignacio, and Carlos Lago-Peñas. 2021. “The Glories of Immigration: How Football Wins Shape Opinion on Immigration.” *Migration Studies* 9(3):466–89.

Lee, David S., and David Card. 2008. “Regression Discontinuity Inference with Specification Error.” *Journal of Econometrics* 142(2):655–74.

Legewie, Joscha. 2016. “Racial Profiling and Use of Force in Police Stops: How Local Events Trigger Periods of Increased Discrimination.” *American Journal of Sociology* 122(2):379–424.

Mancosu, Moreno, and Mònica Ferrín Pereira. 2021. “Terrorist Attacks, Stereotyping, and Attitudes toward Immigrants: The Case of the Manchester Bombing.” *Social Science Quarterly* 102(1):420–32.

Marcus-Newhall, Amy, William C. Pedersen, Mike Carlson, and Norman Miller. 2000. “Displaced Aggression Is Alive and Well: A Meta-Analytic Review.” *Journal of Personality and Social Psychology* 78(4):670–89.

Mazziotta, Agostino, Amélie Mummendey, and Stephen C. Wright. 2011. “Vicarious Intergroup Contact Effects: Applying Social-Cognitive Theory to Intergroup Contact Research.” *Group Processes & Intergroup Relations* 14(2):255–74.

McDowall, David, Colin Loftin, and Matthew Pate. 2012. “Seasonal Cycles in Crime, and Their Variability.” *Journal of Quantitative Criminology* 28(3):389–410.

Miller, N. E. 1941. “I. The Frustration-Aggression Hypothesis.” *Psychological Review* 48(4):337–42.

Nägel, Christof, and Mark Lutter. 2020. “The Christmas Market Attack in Berlin and Attitudes toward Refugees: A Natural Experiment with Data from the European Social Survey.” *European Journal for Security Research* 5(2):199–221.

Newsable. 2021. “UEFA Euro 2020 Final: English Footballers Victim of Racial Abuse Following Missed Penalties against Italy.” *Newsable,* July 12.

NPCC. 2021. “Hate Crime Investigation Following Euro 2020 Final Leads to 11 Arrests.” *NPCC,* August 5.

O’Neill, Aofie. 2017. “Hate Crime, England and Wales, 2016/17.” *Home Office: Statistical Bulletin* 17(17).

Parrott, Dominic J., and John L. Peterson. 2008. “What Motivates Hate Crimes Based on Sexual Orientation? Mediating Effects of Anger on Antigay Aggression.” *Aggressive Behavior* 34(3):306–18.

Piatkowska, Sylwia J., and Brian J. Stults. 2022. “Brexit, Terrorist Attacks, and Hate Crime: A Longitudinal Analysis.” *Social Problems* 69(4):968–96.

Reny, Tyler T., and Benjamin J. Newman. 2021. “The Opinion-Mobilizing Effect of Social Protest against Police Violence: Evidence from the 2020 George Floyd Protests.” *American Political Science Review* 115(4):1499–507.

Rothschild, Zachary K., Mark J. Landau, Daniel Sullivan, and Lucas A. Keefer. 2012. “A Dual-Motive Model of Scapegoating: Displacing Blame to Reduce Guilt or Increase Control.” *Journal of Personality and Social Psychology* 102(6):1148–63.

Sampson, Robert J., Stephen W. Raudenbush, and Felton Earls. 1997. “Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy.” *Science* 277(5328):918–24.

Schiappa, Edward, Peter B. Gregg, and Dean E. Hewes. 2005. “The Parasocial Contact Hypothesis.” *Communication Monographs* 72(1):92–115.

Schwitter, Nicole, and Ulf Liebe. 2023. “Not Cologne but the Data Collection (Might Have) Changed Everything: A Cautionary Tale on Ignoring Changes in Data Recording in Sociological Research.” *European Sociological Review* 39(6):1005–10.

Sheridan, Lorraine P. 2006. “Islamophobia Pre– and Post–September 11th, 2001.” *Journal of Interpersonal Violence* 21(3):317–36.

Sniderman, Paul M., Louk Hagendoorn, and Markus Prior. 2004. “Predisposing Factors and Situational Triggers: Exclusionary Reactions to Immigrant Minorities.” *American Political Science Review* 98(1):35–49.

Spörlein, Christoph, and Elmar Schlueter. 2021. “Ethnic Insults in YouTube Comments: Social Contagion and Selection Effects during the German ‘Refugee Crisis.’” *European Sociological Review* 37(3):411–28.

Stanley-Becker, Isaac. 2018. “Mesut Özil: World Cup Player Says He Is ‘a German When We Win’ but ‘an Immigrant When We Lose,’ Quits National Team.” *The Washington Post*, July 23.

Sullivan, Becky. 2021. “Three Black Soccer Players Are Facing Racist Abuse after England’s Euro 2020 Defeat.” *NPR*, July 12.

Tajfel, Henri. 1981. *Human Groups and Social Categories: Studies in Social Psychology*. Cambridge, UK: Cambridge University Press.

Thoemmes, Felix, Wang Liao, and Ze Jin. 2017. “The Analysis of the Regression-Discontinuity Design in R.” *Journal of Educational and Behavioral Statistics* 42(3):341–60.

Twitter UK. 2021. “Combatting Online Racist Abuse: An Update Following the Euros.” *Twitter UK*, August 10.

UEFA. 2021. “UEFA EURO 2020 Impresses with 5.2 Billion Cumulative Global Live Audience.” *UEFA*, September 2.

Vallerand, Robert J., Nikos Ntoumanis, Frederick L. Philippe, Geneviève L. Lavigne, Noémie Carbonneau, Arielle Bonneville, Camille Lagacé-Labonté, and Gabrielle Maliha. 2008. "On Passion and Sports Fans: A Look at Football." *Journal of Sports Sciences* 26(12):1279–93.

Zoorob, M. 2020. "Do Police Brutality Stories Reduce 911 Calls? Reassessing an Important Criminological Finding." *American Sociological Review* 85(1):176–83.

APPENDIX A1

A1. RDD Assumption checks

A1.1. Effects on placebo outcomes

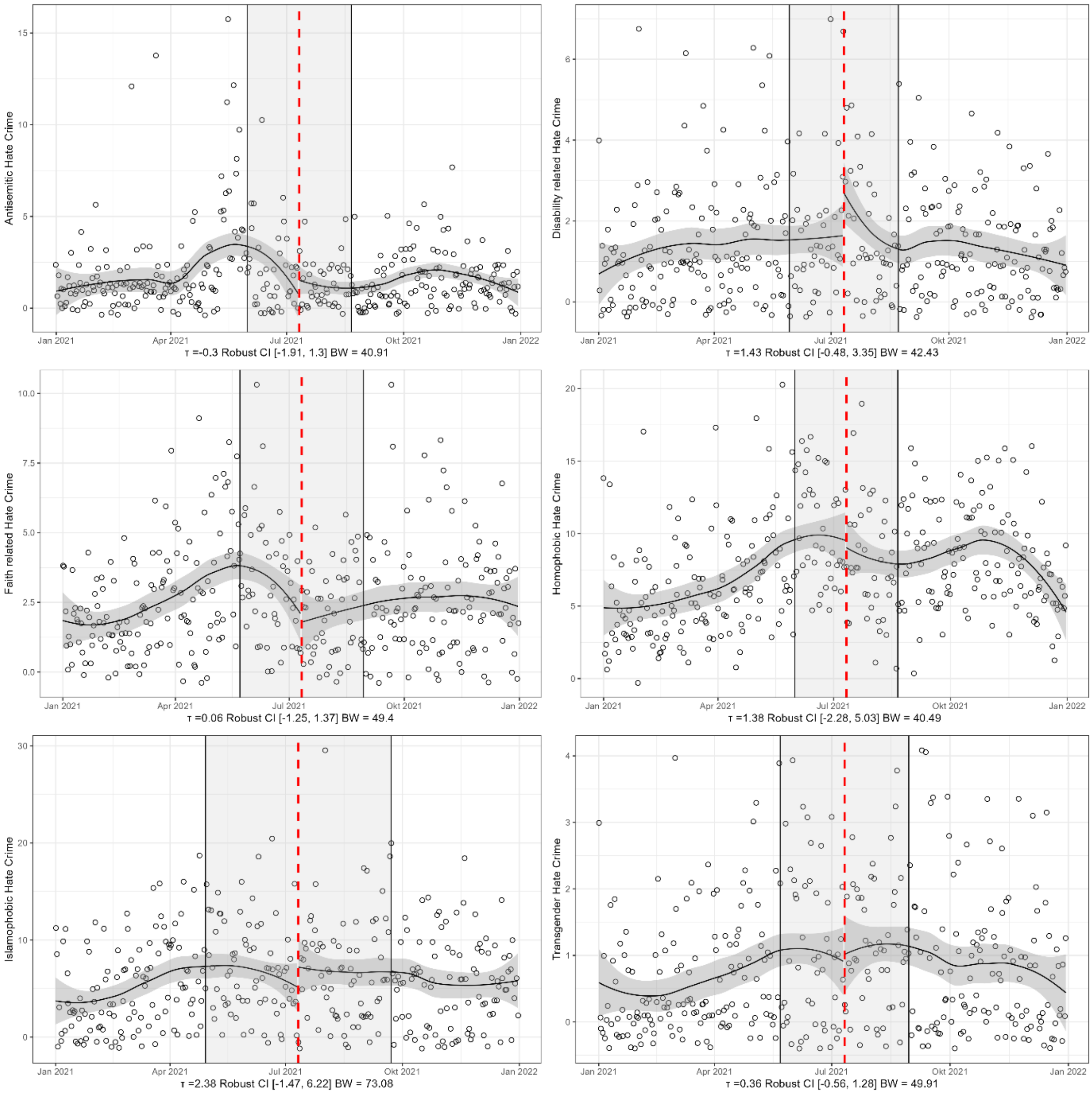

Figure A1. Placebo effects on other hate crimes

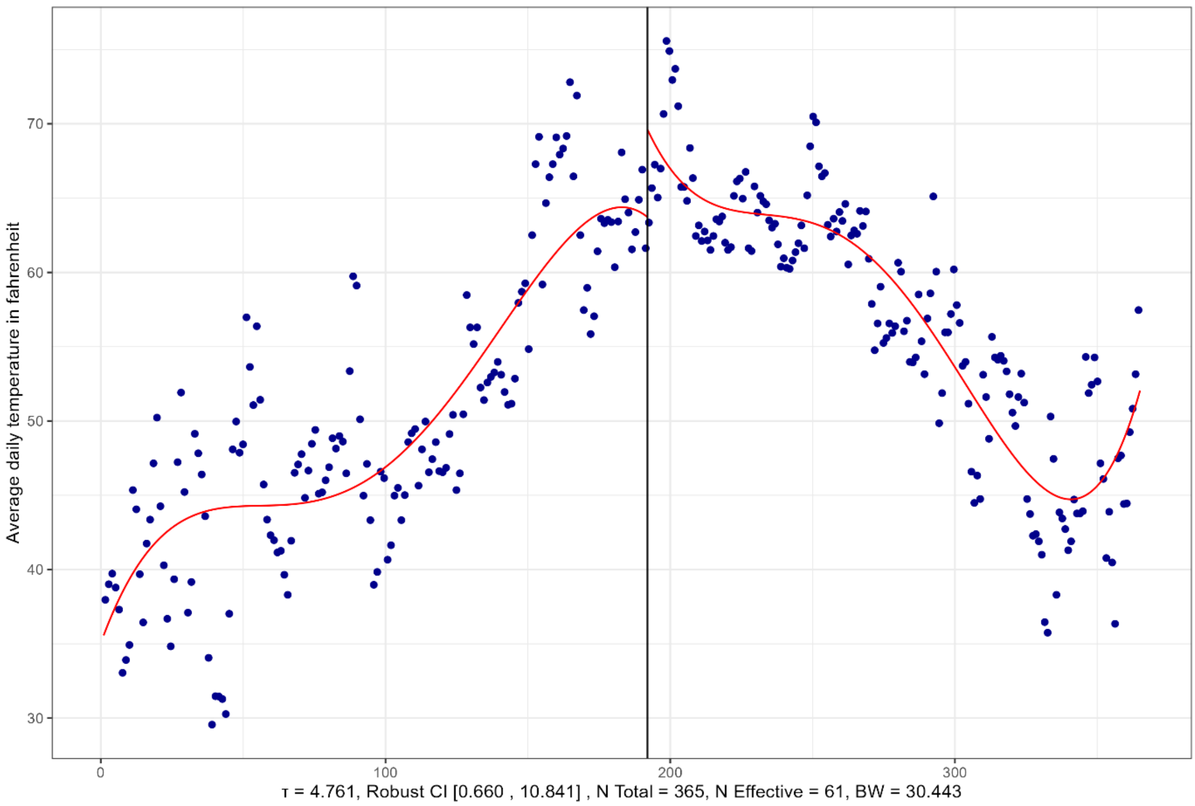

Figure A2. RD with average daily temperature as outcome.

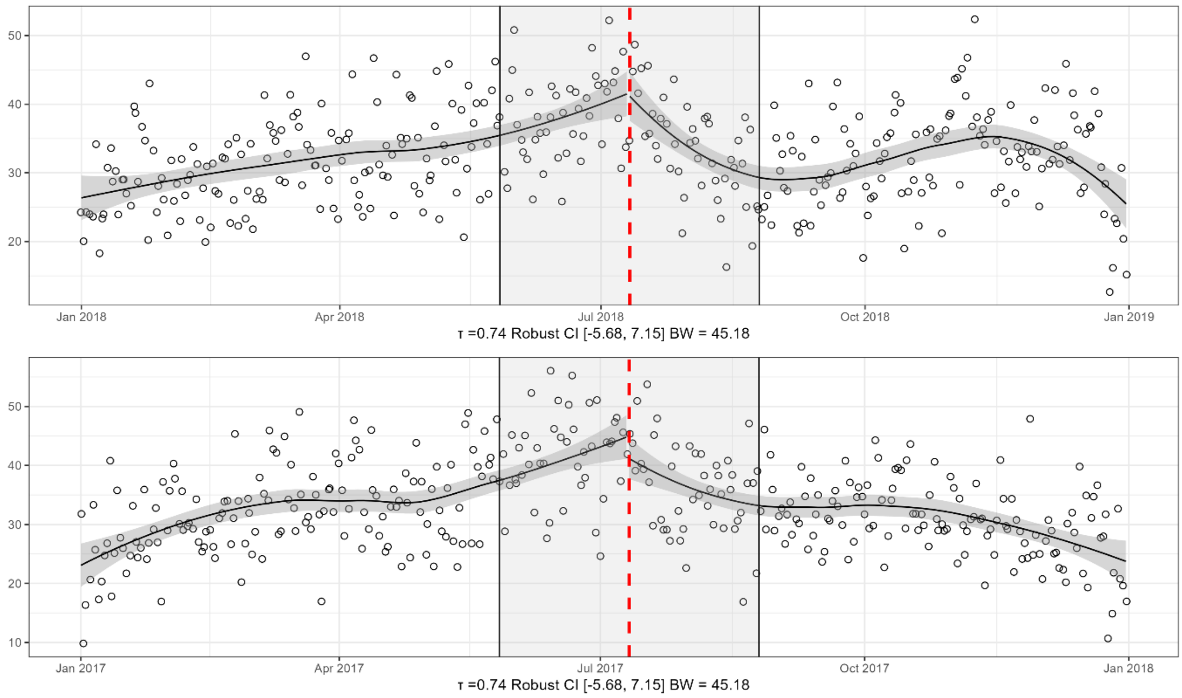

Figure A3. RD models for 2017 and 2018.

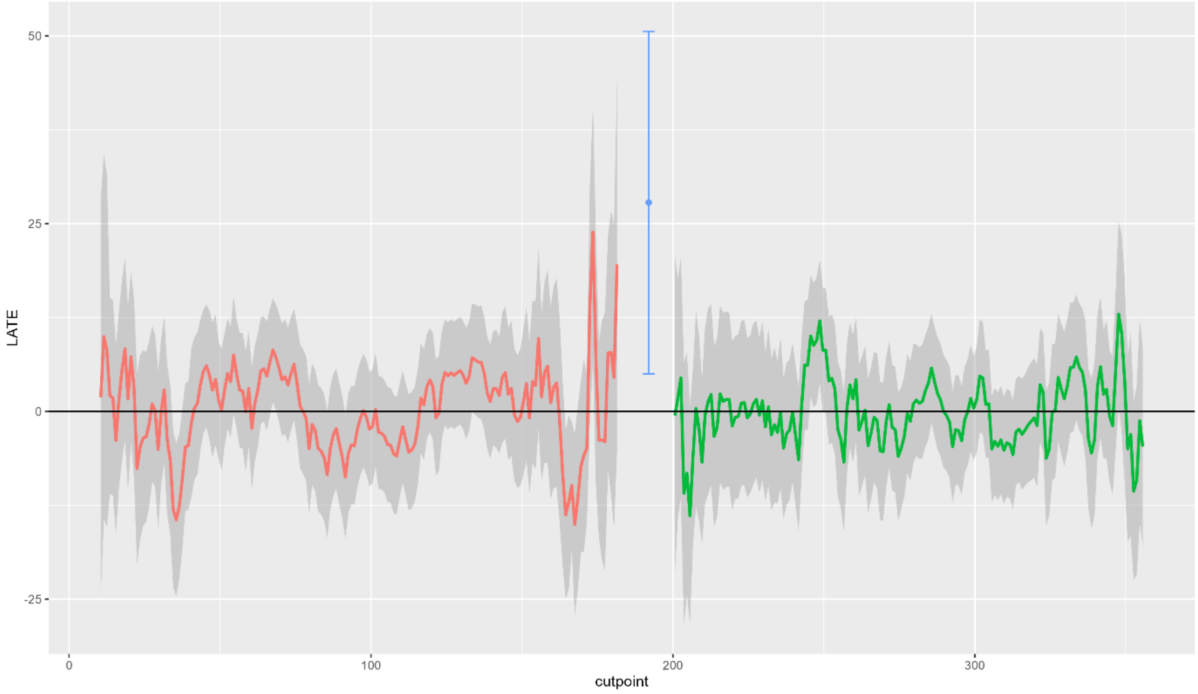

Figure A4. Fake cut-offs.

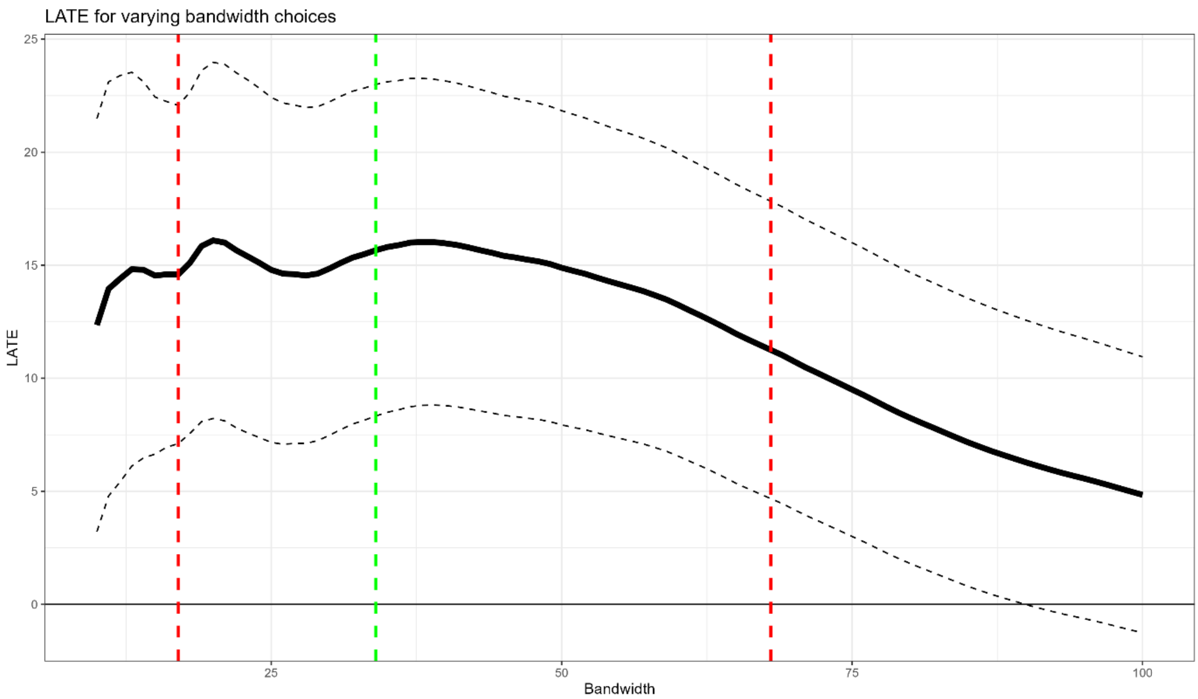

Fig A5. LATEs for varying bandwidth choices. We start at bw = 10 for reasons of statistical power. Dashed green line is the optimal bandwidth chosen in the main models. Dashed red lines indicate double and half that bandwidth.

Parametric RDiT analysis

Following (Hausman and Rapson 2018), we extend the local linear RDiT approach by a traditional parametric approach which involves using the entire available data and estimating the discontinuity by choosing the appropriate polynomial (Angrist and Pischke 2009). One drawback of this design is that the estimates are highly sensitive to the degree of the polynomial and the choice is usually quite arbitrary (Gelman and Imbens 2019). The best practice is to choose the polynomial order based on the smallest Bayesian Information Criterion (BIC) of the respective models (Hausman and Rapson 2018). In our case, this means estimating a model with a polynomial of order 4. Again, high order polynomials lead to poor coverage of confidence intervals (Gelman and Imbens 2019). For the sake of transparency, we also ran regressions with smaller and higher order polynomials. Results are graphically displayed in the appendix in Figure A5. The discontinuity estimate is not significant in the linear and quadratic model. However, for all models with higher order polynomials, the effect is positive and significantly different from zero. While the local linear RD clearly outperforms the parametric model in this application, it is reassuring to know that the effect is not dependent on the particular approach taken. In the case of time-varying treatment effects, the parametric models might not be able to estimate the initial local impact (in time) to the threshold (Hausman and Rapson 2018). We therefore assess the duration of the effect by applying an interrupted time series analysis in appendix A2.

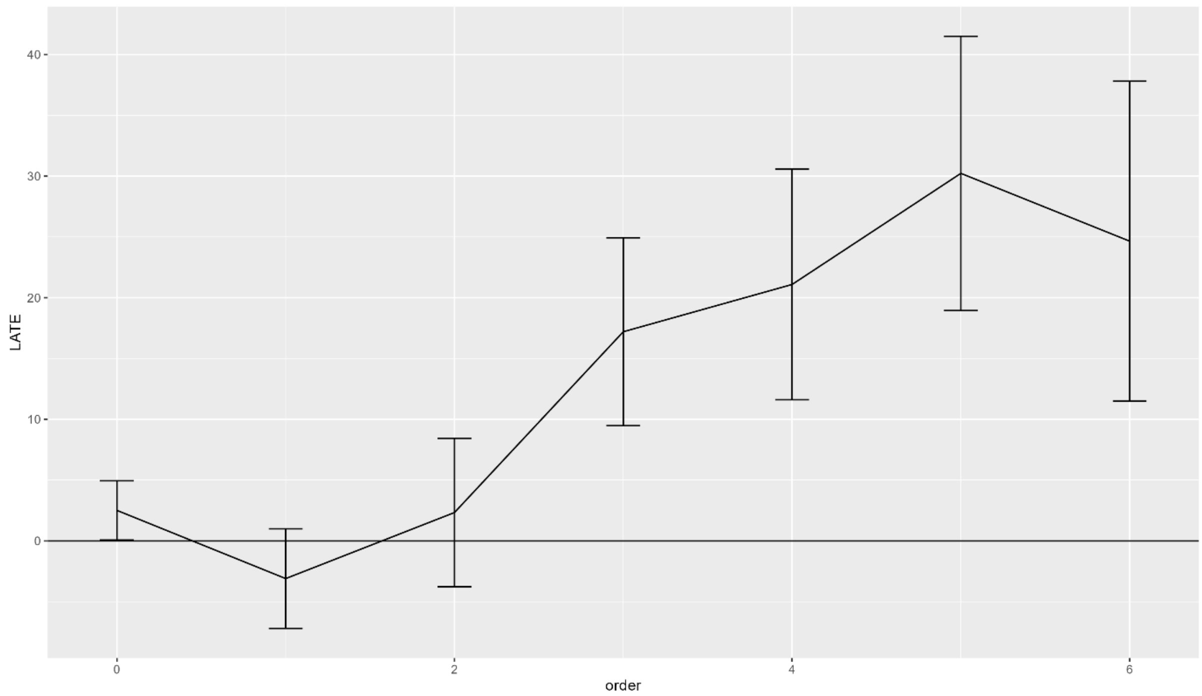

Figure A6. LATE dependent on chosen polynomial order.

RDiT hourly analyses

Since the data we were provided by the Met police is extremely granular, we are also able to assess the LATE in an hourly analysis. The Euro 2021 final ended at 22:53 British Summer Time (BST). We thus used 23:00 BST as the cut-off in the hourly RDiT analyses. Results can be seen in Figure A7. The analysis suggests an increase of 0.78 racial hate crimes on average in the post treatment period per hour compared to the pre-treatment period (Robust CI [0.19, 1.14]).

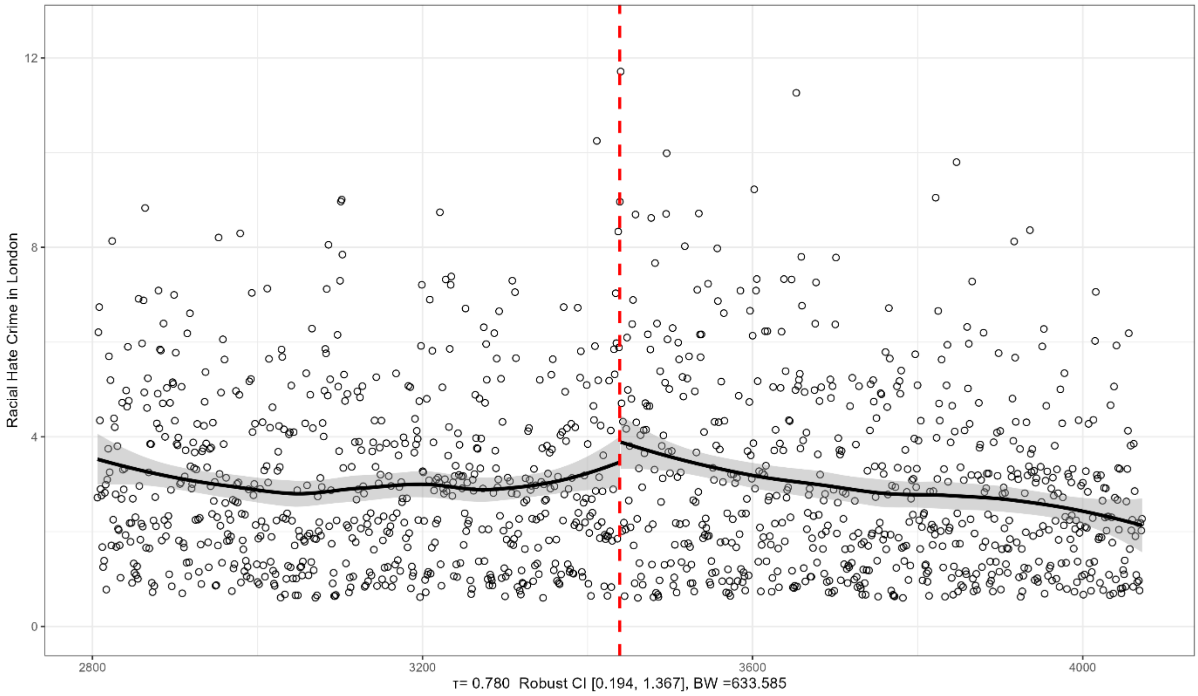

Figure A7. Discontinuity estimate for a RD model with optimal bandwidth in an hourly analysis.

APPENDIX A2

A2. Interrupted Time Series Analysis

We start our analyses by running a preliminary Ordinary Least Squares (OLS) model to assess the potential for temporal autocorrelation in the complete daily time series. Here, we regress daily racial hate crimes on a treatment dummy, a simple linear time indicator, and a trend variable (an interaction of the former two variables). First, we run a Durbin-Watson-test which indicates presence of autocorrelation at the first lag (DW = 0.17, $p < 0.001$). Secondly, we graph the residuals from a preliminary OLS regression to check for serially correlated errors. As can be seen in Figure A8, there is no clear pattern that would indicate temporal autocorrelation. Since these two tests come to dissimilar conclusions, we finally assessed both an autocorrelation function plot and a partial autocorrelation function plot. In the first plot in Figure A9, we see an exponential decay of significant lags. The partial autocorrelation plot furthermore suggests one significant lag before dropping to zero, indicating an AR (autoregression) 1 process. We also tested alternative model specifications of autoregression and/or moving averages in likelihood-ratio tests which did not provide a better model fit.

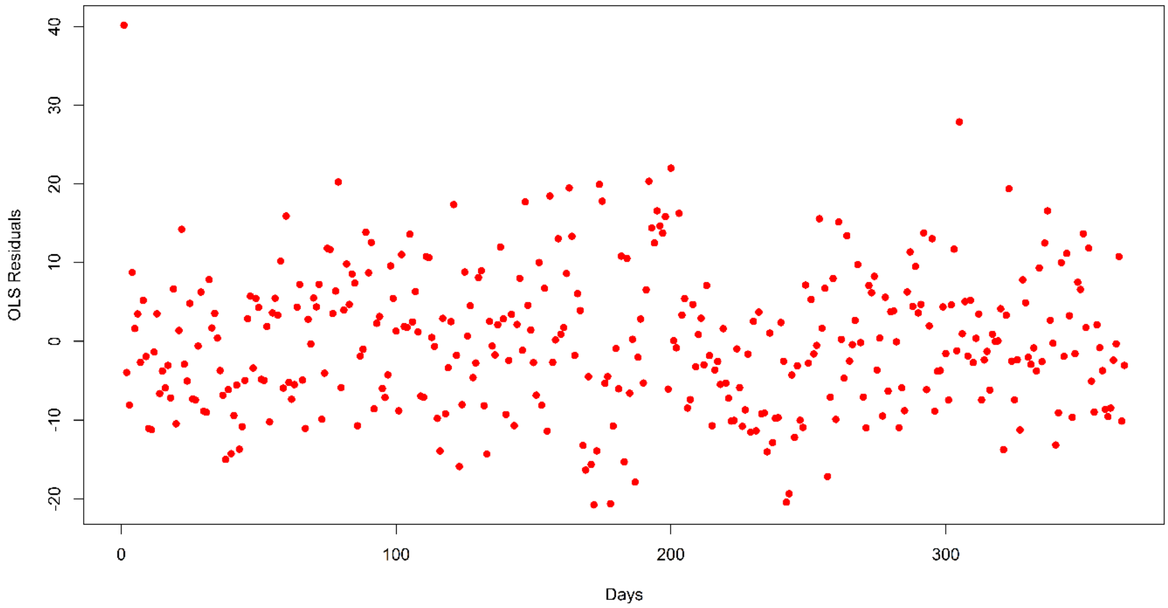

Figure A8. Simple OLS model residuals plotted against weekly time intervals.

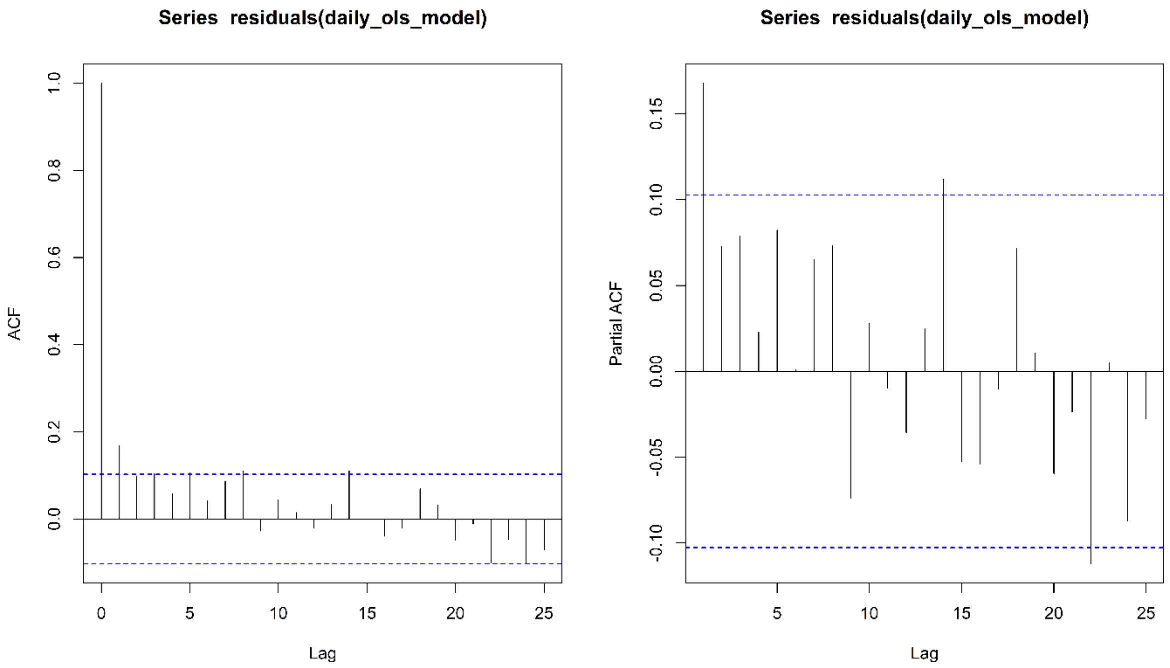

Figure A9. Autocorrelation Function (ACF) plot and Partial Autocorrelation (PACF) plot

We then performed Augmented-Dickey-Fuller (ADF) tests for stationarity (constant mean and variance over time). The overall time series is non-stationary (Dickey-Fuller = -2.48, $p = 0.37$). We account for this by using first differencing of the outcome in a SARIMA (Seasonal Autoregressive Integrated Moving Average) model with an AR (1), MA (0) process and a seasonal pattern of s = 7. This weekly pattern allows to model the weekly plummet in racial hate crimes on Sundays.[12]

Our goal is to arrive at a more precise estimation of the duration of the treatment effect. We therefore follow Piatkowska and Stults (2022) and include 20-day lags of the triggering event variable.[13] These models allow to study the effect on each of the 20 days in the wake of the EURO 2020 final. Similar to the models in the paper, we add weekday dummies as well as daily average temperature as additional covariates. Figure A10 plots the daily lags from models with (grey dots) and without (black dots) these additional covariates. Racial hate crimes appear to be significantly and substantially elevated by roughly 25 additional counts per day for 7 days (including the day of the event). After this period, the following day dummy is not significantly

[12] Note that simpler models (such as ARMA (Autoregressive Moving Average) models with AR (1) and MA (0) processes lead to very similar point estimates as those presented in figure A10. The same applies to different seasonal patterns such as s = 30 or s = 14

[13] Note that the study by Piatkwoski use 15-day lags. Our findings are largely non-sensitive to the number of daily dummies.

different from zero, which is likely to be attributed to the fact that Sundays generally have the lowest count for hate crimes in our data. After this day, the treatment dummies are not consecutively statistically significant with only two outliers (day 8 after the final and day 11 after the final).

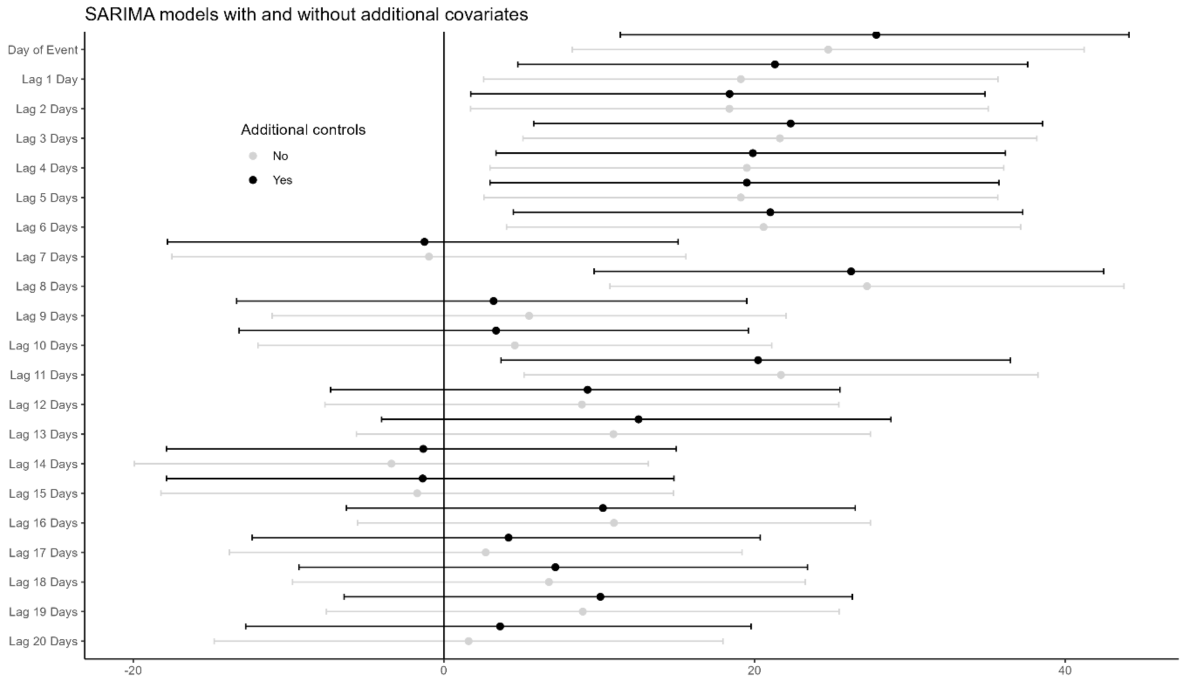

Figure A10. Daily treatment dummies for SARIMA models with and without additional covariates.

APPENDIX A3

A3. Additional robustness checks

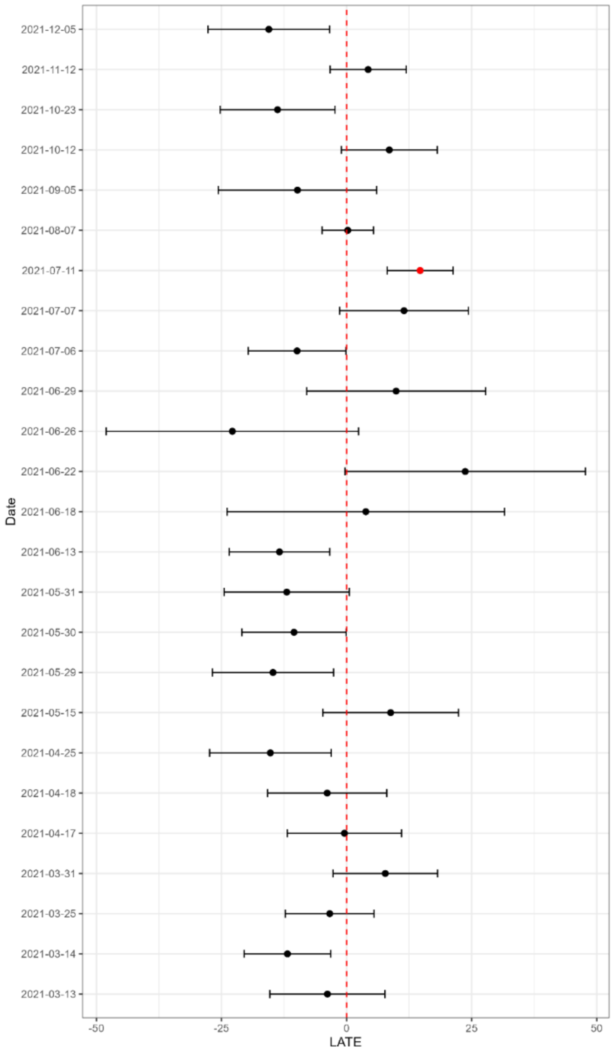

Fig A11. Placebo RD estimates on all games in Wembley in 2021

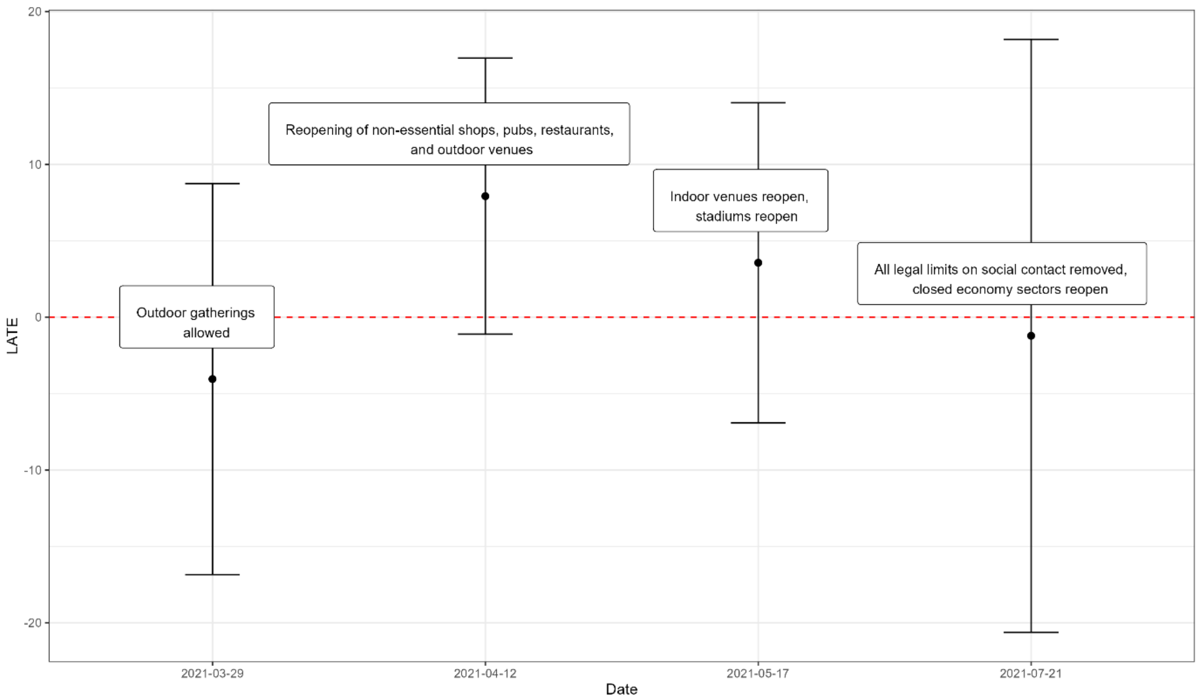

Fig A12. Placebo RD estimates on all COVID-19-restriction relaxations.

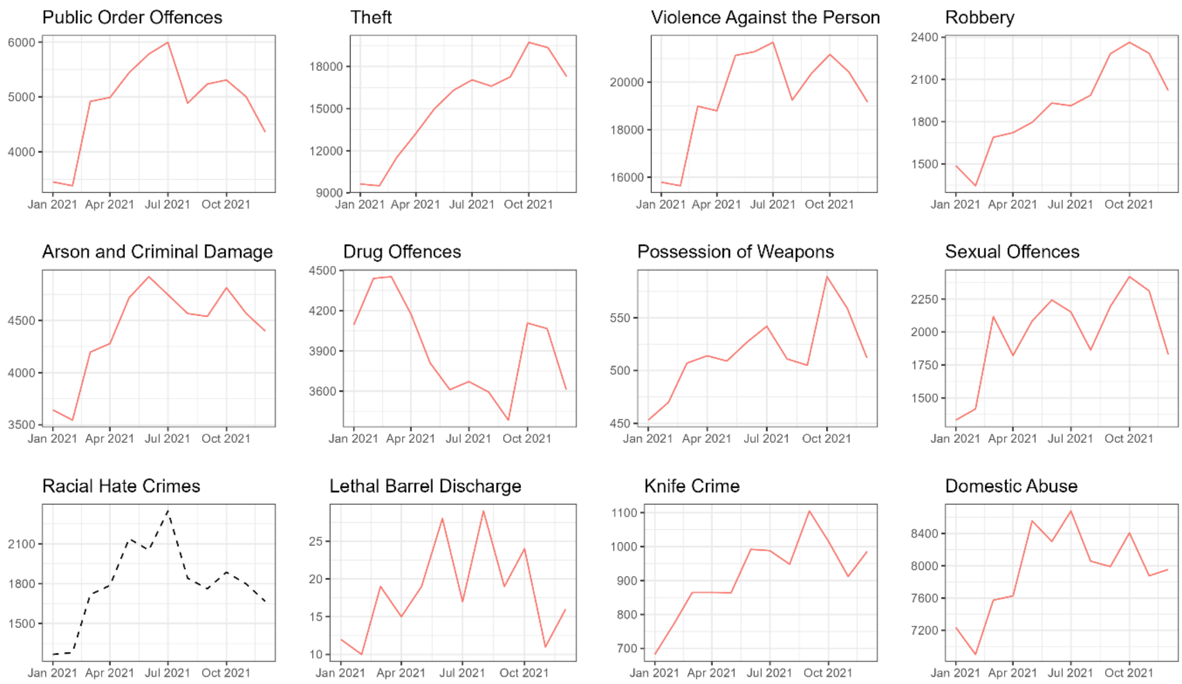

Fig A13. Monthly recorded crime trends throughout the year 2021. Y-axis in each plot is set by the default option in the software program to visualise relative spikes.

APPENDIX A4

A4. Spatially heterogenous treatment effects

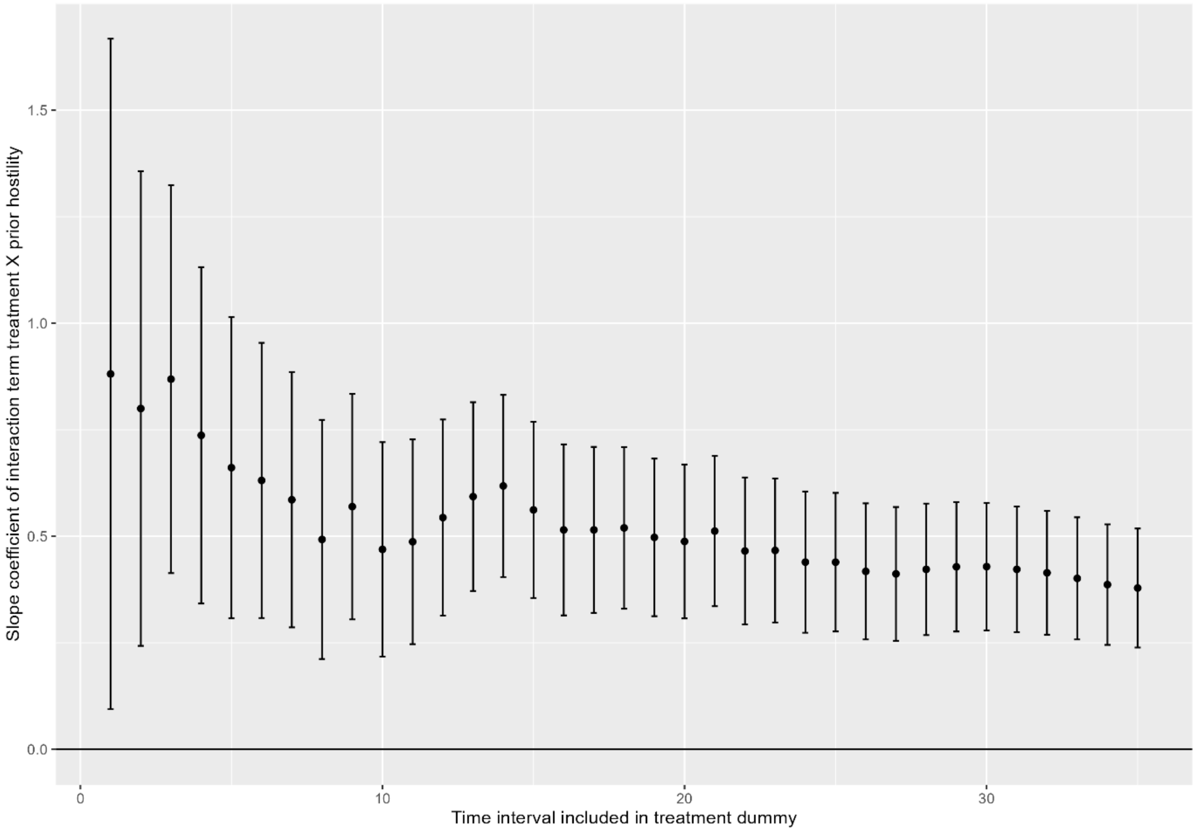

Fig A14. Interaction effect slope coefficient for choosing different lengths for the treatment dummies.

Table A1. Fixed effects regression models with logged dependent variable

	Racial Hate Crimes			
	Model 1	Model 2	Model 3	Model 4
After Euro 2020 final (OBW)	0.13***	0.03	0.13***	0.03
	(0.02)	(0.02)	(0.01)	(0.02)
After Euro 2020 final (OBW) * Prior Hostility			0.09**	0.09**
			(0.03)	(0.03)
R^2	0.01	0.02	0.01	0.03
Adjusted R^2	0.00	0.02	0.00	0.02
F Statistic	66.93***	95.21***	38.12***	33.08***
Borough fixed effects	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes

Notes: * p<0.05; ** p<0.01; *** p<0.001, N = 11,680 in all models, Cluster-Robust Standard Errors in brackets, OBW = optimal bandwidth from RDiT Model 1 in table 1 (35 days after the final), interacted vectors are mean centred, additional controls include a linear time spline, average temperature, and weekday dummies

Table A2. Negative Binomial regressions

	Racial Hate Crimes			
	Model 1	Model 2	Model 3	Model 4
After Euro 2020 final (OBW)	1.26***	1.06	1.24***	1.04
	(0.04)	(0.03)	(0.07)	(0.03)
After Euro 2020 final (OBW) * Prior Hostility			1.21***	1.21***
			(0.07)	(0.05)
Observations	11,680	11,680	11,680	11,680
AIC	35386.8	35189.9	35376.5	35179.0
BIC	35408.8	35271.0	35406.0	35267.3
RMSE	1.29	1.28	1.29	1.28
Borough fixed effects	Yes	Yes	Yes	Yes
Additional Controls	No	Yes	No	Yes

Notes: *** p<0.001, Displayed coefficients are incidence rate ratios, N = 11,680 in all models , Cluster-Robust Standard Errors in brackets, OBW = optimal bandwidth from RDiT Model 1 in table 1 (35 days after the final), interacted vectors are mean centred, additional controls include a linear time spline, average temperature, and weekday dummies

REFERENCES

Angrist, J. D., and J. S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37(3):447–56. doi: 10.1080/07350015.2017.1366909.

Hausman, Catherine, and David S. Rapson. 2018. "Regression Discontinuity in Time: Considerations for Empirical Applications." *Annual Review of Resource Economics* 10(1):533–52. doi: 10.1146/annurev-resource-121517-033306.

Piatkowska, Sylwia J., and Brian J. Stults. 2022. "Brexit, Terrorist Attacks, and Hate Crime: A Longitudinal Analysis." *Social Problems* 69(4):968–96. doi: 10.1093/socpro/spab005.