WORKING PAPER · NO. 2023-11

Criminal Charges, Risk Assessment, and Violent Recidivism in Cases of Domestic Abuse

Dan A. Black, Jeffrey Grogger, Tom Kirchmaier, and Koen Sanders

JANUARY 2023

Becker Friedman Institute
FOR ECONOMICS AT UCHICAGO

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

Electronic copy available at: https://ssrn.com/abstract=4336508

Criminal Charges, Risk Assessment, and Violent Recidivism in Cases of Domestic Abuse*

Dan A. Black

University of Chicago, NORC, and Institute for the Study of Labor

Jeffrey Grogger

University of Chicago; Center for Economic Performance, London School of Economics; National Bureau of Economic Research; and Institute for the Study of Labor

Tom Kirchmaier

Copenhagen Business School and Center for Economic Performance, London School of Economics

Koen Sanders

Center for Economic Performance, London School of Economics

Abstract

Domestic abuse is a pervasive global problem. Here we analyze two approaches to reducing violent DA recidivism. One involves charging the perpetrator with a crime; the other provides protective services to the victim on the basis of a formal risk assessment carried out by the police. We use detailed administrative data to estimate the average effect of treatment on the treated using inverse propensity-score weighting (IPW). We then make use of causal forests to study heterogeneity in the estimated treatment effects. We find that pressing charges substantially reduces the likelihood of violent recidivism. The analysis also reveals substantial heterogeneity in the effect of pressing charges. In contrast, the risk-assessment process has no discernible effect.

Keywords: Domestic abuse, charges, risk assessment, propensity score weighting

*We thank Cole Frank, Smriti Ganapathi, and Merritt Smith for outstanding research assistance; ACC Chris Sykes, Roger Pegram, and Duncan Stokes from Greater Manchester Police for providing us with data and support; and numerous seminar and conference participants for helpful comments. This work was supported by an Innovation Fund Grant from the Kenneth C. Griffin Applied Economics Incubator at the University of Chicago. The views expressed herein are our own, and do not necessarily reflect those of the funder or Greater Manchester Police.

January 21, 2023

Electronic copy available at: https://ssrn.com/abstract=4336508

1. Introduction

Domestic abuse (DA) is a pervasive problem worldwide. According to the World Health Organization (2021), roughly one-third of women will experience physical or sexual violence by a partner at some point during their lives. In the U.S., one-third of female murder victims are killed by intimate partners (National Coalition Against Domestic Violence). The Crime Survey for England and Wales estimates that 1.6 million women, or roughly 5 percent of the entire female population, experienced domestic abuse in just the 12-month period prior to March 2020 (Grierson, 2020).

Domestic abuse has far-reaching economic consequences. It adversely affects the employment, earnings, and welfare dependency of victims (Bhuller et al., 2021). It harms the health of babies in utero at the time the abuse takes place (Aizer, 2011). It lowers the educational performance of affected children and that of their peers as well (Bhuller et al., 2021; Gutierrez and Molina, 2020).

Considering the prevalence and consequences of the problem, a key question for policy is, what can be done to reduce it? In this paper, we focus on two interventions initiated by the police. The first involves pressing criminal charges against the perpetrator. The second is a process of providing protective services on the basis of a systematic risk assessment made at the scene of the incident. We estimate how these two different interventions affect reported violent recidivism in domestic abuse cases.

Criminal charges may arise from an investigation carried out in response to a DA-related call for service, if police determine that a crime has taken place. However, officers exercise discretion in determining whether a crime has occurred and, if so, whether it warrants prosecution (Myhill, 2019; Her Majesty's Inspectorate of Constabulary, 2015). Officers may also arrest the perpetrator, but as we explain below, we have no useful data on arrests. In England, the setting for our study, the perpetrator need not be arrested in order to be charged.

Systematic risk assessment is used by law enforcement and social service agencies in a wide variety of settings. In the context of DA, 42 percent US police departments use it in some form (Police Executive Research Forum, 2015). It is also used for DA cases in Canada and several EU nations (Berk et al., 2005; European Institute for Gender Equality, 2019), although its use may be most advanced in England and Wales. Almost all police agencies there follow the Domestic Abuse, Stalking, Harassment, and Honor-Based (DASH) violence risk assessment model, and assign protective resources to victims whose cases are judged to be high-risk.

This paper makes several contributions. First, the literature has found mixed effects of arrests on DA recidivism. We contribute further evidence on how police interventions–in this case, charges rather than arrests–deter domestic abuse recidivism. Second, ours is the first large-scale study of the causal effect of a protective process that begins with risk assessment. Prior work has been based on smaller samples, and some studies have lacked a comparison

Electronic copy available at: https://ssrn.com/abstract=4336508

group (Robinson, 2006; Robinson and Tregidga, 2007; Messing et al., 2014; Whinney, 2015). Third, we systematically analyze heterogeneity in our estimated treatment effects. To do so we first estimate causal forests (Wager and Athey, 2018; Athey et al., 2019), then use them to learn a simple decision rule by which recidivism may be further reduced. More broadly, we contribute to the growing literature on the economics of domestic violence. Studies on this topic can be classified broadly into analyses of the effect of DA on economic outcomes (Aizer, 2011; Anderberg et al., 2016; Bhuller et al., 2021; Bindler and Ketel, 2022; Currie et al., 2022; Gutierrez and Molina, 2020); studies of the effect of the economic environment on DA (Aizer and Dal Bó, 2009; Aizer, 2010; Bhalotra et al., 2021; Card and Dahl, 2011; Guarnieri and Rainer, 2018; Hidrobo et al., 2016; Koppensteiner et al., 2020; Tur-Prats, 2019); and analyses of policy responses to DA (Amaral et al., 2022; Chin and Cunningham, 2019; Iyengar, 2009; Sviatschi and Trako, 2021).

The key obstacle to estimating the causal effects of the interventions is that they may be correlated with characteristics of the incident or its participants that predict high recidivism risk. That is, incidents that are charged or classified as high-risk may be likely to result in recidivism, even if there were no intervention. If so, the simple correlation between recidivism and the intervention may understate the causal effect of the intervention. The reason is that it confounds the effect of the intervention with the selection bias that stems from the process by which incidents are chosen for treatment.

To deal with this fundamental problem, we use inverse propensity score weighting (IPW). Separately for each intervention, we estimate the propensity score, or probability of treatment, as a function of characteristics of the DA incident, the perpetrator, the victim, and the dyad, that is, the victim-perpetrator pair. We then use the predicted propensity score to construct weights, which we apply to a regression of violent recidivism on the intervention indicator. Under two key assumptions this procedure identifies the average effect of treatment on the treated (ATT), that is, the causal effect of the intervention on the probability of violent recidivism among the treated incidents.

One of the key assumptions is common support. This requires the distribution of propensity scores for the intervention group to have the same domain as that of the comparison group. We show that common support holds for all but a small number of observations in our sample.

The other key assumption is conditional independence. This requires the probability of treatment, given observable predictors, to be independent of any unobservable factors that may influence the outcome. Put differently, once we have controlled for all the observable covariates at our disposal, conditional independence implies that treatment status is uncorrelated with any unobservables that may predict recidivism.

Conditional independence is a strong assumption. It implies that balancing the sample on observables eliminates selection bias. At a minimum, this requires the analyst to have access

Electronic copy available at: https://ssrn.com/abstract=4336508

to a sizeable number of predictors that strongly predict treatment status and the outcome.

On this ground, we are in a reasonably good position. Many of our predictors reflect past DA and criminal behavior of members of the dyad, which help predict both the interventions and violent recidivism. Other covariates include indicators that should help predict whether charges are filed, such as whether the victim was injured. For the risk-assessment intervention, the case for predictive covariates is even stronger, since we have the responses to the DASH questionnaires that officers use to grade the victim's risk status.

Beyond estimating the ATT, we analyze heterogeneity in the effects of the interventions. To to so we estimate a causal forest, which provides estimates of the treatment effect for each incident in the sample (Wager and Athey, 2018; Athey et al., 2019). We use those estimates to learn a decision rule, which potentially could lead to a reallocation of the interventions that further reduced violent recidivism.

We find that criminal charges reduce violent recidivism in DA cases. Estimates are similar across different approaches to propensity score estimation, and indicate that charges reduce violent recidivism by about 5 percentage points. That amounts to almost a 40 percent reduction relative to the mean recidivism rate. This estimate is larger than many estimates of the effect of arresting the perpetrator, but similar to estimates from Sherman and Berk (1984) and Amaral et al. (2022). We also identify substantial heterogeneity in the effect of charges as a function of observed covariates. At the same time, we find no evidence that the risk assessment process reduces violent recidivism.

In the next section we provide the reader with some background on the history of different police responses to domestic abuse. We also discuss current-day police treatment of domestic abuse cases in England and Wales. After that, we discuss our data, which was provided by a large English police force. A discussion of our methods follows in section 4, followed by results in Section 5. Section 6 concludes.

2. Background

2.1. Law Enforcement Approaches to Domestic Abuse

The most-studied police approach to domestic abuse is arresting the perpetrator, which in the US is generally a precursor to pressing charges in court. The first study of the effect of arrest on domestic abuse recidivism was the Minnesota Domestic Violence Experiment (MDVE), in which cases were randomly assigned to one of three treatment conditions: arrest the perpetrator, send him away, or provide counseling to the couple.[1] Recidivism rates based on police records over a 12-month follow-up period were 13 percent for the arrest group and 26 percent for the

[1]We use masculine pronouns for perpetrators and feminine pronouns for victims. Although there are male victims of DA, roughly 80 percent of the victims in our sample are female.

Electronic copy available at: https://ssrn.com/abstract=4336508

baseline group. Recidivism rates based on survey responses, which may capture incidents that were not reported to police, were 19 percent and 37 percent, respectively (Sherman and Berk 1984).[2]

In light of the findings from MDVE, many jurisditions began implementing mandatory or presumptive arrest policies (Fagan, 1995). The US National Institute of Justice commissioned five follow-up studies of mandatory arrest, known as the Spousal Abuse Replication Studies (SARP), all of which employed random assignment. Two of these studies did not find negative effects on recidivism (Dunford et al., 1990; Hirschel and Hutchison, 1992), whereas others yielded evidence of treatment effect heterogeneity. Three studies showed that arresting some perpetrators may have actually backfired in some cases, increasing recidivism by provoking a retaliatory response (Berk et al., 1992; Pate and Hamilton, 1992; Sherman et al., 1992).

In England, authorities began calling for a presumptive arrest policy roughly 20 years ago (Home Office, 2000), although to date there appear to be many deviations in practice (Her Majesty's Inspectorate of Constabulary, 2014). However, despite policy changes similar to those in the US, there has been little research into the matter. As far as we are aware, this study represents one of only two causal analyses of the effect of law enforcement approaches to DA recidivism in the UK. The other is a study of the effect of arrests, which was written contemporaneously with our paper (Amaral et al., 2022).

2.2. Risk Assessment for Domestic Abuse Recidivism

Since the mid-1990s, risk assessment has arisen as another police response to domestic abuse (Dutton and Kropp, 2000; Campbell et al., 2009; Ericson and Haggerty, 1997). Different assessment protocols differ with respect to the level of violence they seek to predict, but many share basic features. They typically involve a questionnaire that can be administered to the victim in an interview or be answered on the basis of an official records search (Messing and Thaller, 2013). Some involve an explicit scoring rule, whereas others ask the person administering the questionnaire to use the responses, together with their professional judgement, to classify the risk presented by the case (Kropp, 2004). Risk assessments in domestic abuse cases are conducted routinely in several European countries and in many parts of Canada and the US (Kropp, 2004; Berk et al., 2005; Roehl et al., 2005; Police Executive Research Forum, 2015; Turner et al., 2019; European Institute for Gender Equality, 2019).

A sizeable literature has addressed the predictive validity of roughly a half-dozen risk assessment instruments that have been used to predict DA recidivism in the US and Canada. Some studies analyze predictive validity for binary measures of recidivism, whereas others analyze predictions reflecting the severity of incidents of recidivism. Most of the most common

[2]These should be regarded at intent-to-treat estimates, since police were explicitly allowed to arrest the perpetrator if the situation warranted it, regardless of his assigned treatment status. Angrist (2006) found considerable departures from randomness in arrests made as opposed arrests assigned.

Electronic copy available at: https://ssrn.com/abstract=4336508

instruments exhibit at least moderate levels of predictive validity (Hilton and Harris, 2005; Roehl et al., 2005; Messing and Thaller, 2013; Jung and Buro, 2017; Svalin and Levander, 2020; van der Put et al., 2019; Graham et al., 2021).

In comparison, few studies have analyzed the effect of protective measures allocated on the basis of risk assessments. In the US, Messing et al. (2014) used a difference-in-differences approach to analyze the effect of protective resources allocated on the basis of a risk assessment instrument known at the Lethality Screen. Unfortunately, that study suffered from such high levels of attrition that it is difficult to draw conclusions from it.

The subject has received somewhat more study in England, where most police forces carry out risk assessment routinely using the DASH process (Robinson et al., 2016). DASH consists of two components: a questionnaire and a risk grade assigned by the responding officer. The questionnaire is reproduced in the Appendix. It asks about factors thought to predict recidivism, such as the victim's level of fear, whether she feels socially isolated, whether the victim is financially dependent on the perpetrator, and whether the perpetrator has used or threatened violence with the victim (or others) in the past, among other things. The officer is instructed to use the responses to these questions, and her professional judgement, to grade the victim's risk as high, medium, or standard. Victims designated as high-risk are provide with a customized package of services designed to keep them safe.

Three studies have sought to estimate the effect of this approach. Robinson (2006) and Robinson and Tregidga (2007) each followed DA victims at high risk of re-victimization. Neither study included a comparison group, making it difficult to draw conclusions. Whinney (2015) matched 539 high-risk DA victims to 539 lower-risk victims. He used a coarsened exact-matching approach, matching treatment and comparison groups on the basis of age, gender, the severity of the current incident, the severity of reported incidents within the prior 12 months, date of the incident, and location. He reported that the process had no effect on recidivism.

2.3. The Police Response to Domestic Abuse Calls in England

Most calls for service to the police, whether for domestic abuse or otherwise, originate by telephone (Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services, 2019). Nationwide, police handle a domestic-abuse related call roughly every 30 seconds (Her Majesty's Inspectorate of Constabulary, 2014). Calls are answered by call handlers, whose job is to ascertain basic information about the incident. This includes the location and identity of the caller, the incident, the suspect, the victim and any children; whether any injuries or weapons were involved; whether the suspect is present; and whether the victim is currently in danger (College of Policing, 2022). The call handler will then give the call a priority grade. If the call handler believes the incident to involve a domestic dispute, it will be flagged as such. The call handler may also provide the caller with advice on how to remain safe until the police arrive.

Electronic copy available at: https://ssrn.com/abstract=4336508

The call handler passes this information on to the dispatcher, whose job it is to locate a response officer and assign them to calls based on their priority grade. Typically, the dispatcher will request that any available unit in proximity to the incident respond to the call. The dispatcher also relays to the responding officer any information provided by the call handler.

When the responding officer appears on the scene of a domestic abuse case, her first responsibility is to protect any children, the victim, themselves, and the suspect, if present, from any further harm. Beyond that, the responding officer initiates safety planning for the victim. She also carries out risk assessment by means of the DASH protocol.

The first step is to administer the questionnaire, which consists of 27 items. The second is to assign the risk grade, which amounts to a prediction of future risk. The officer grades the case as standard-, medium-, or high-risk, where high risk implies that "[t]here are identifiable indicators of risk or serious harm. The potential event could happen at any time and the impact would be serious" (Richards, 2009). Officers are instructed to use both the victim's responses to the questionnaire and their own professional judgement in making their risk grade (Robinson et al., 2016).

Victims graded as high-risk are assigned to a Multi-Agency Risk Assessment Conference (MARAC), which may involve service providers from several organizations (Coordinated Action Against Domestic Abuse, 2012). The MARAC assesses the victim's needs and provides a package of services designed to keep her safe and provide her with time to consider her options (Her Majesty's Inspectorate of Constabulary, 2014). The MARAC targets services to suit the circumstances of the victim. Depending on her needs, they may include general safety planning; sanctuary housing; longer-term housing, possibly combined with relocation services; aid in obtaining restraining orders; liaison and advocacy with police, prosecutors, and the courts; physical and mental health services; and support with benefits, children, or immigration issues. In aggregate, nearly all victims are provided with safety support and roughly 40 percent receive housing assistance (SafeLives, 2015). We have no data on the specific services offered to specific victims, which implies that we are evaluating the effect of a protective process, rather than specific protective services.

Beyond their responsibilities for safekeeping and risk assessment, the responding officer initiates an investigation. In the course of the investigation, the officer may arrest the perpetrator, if grounds for arrest exist. Grounds for arrest under English law include protecting a child or vulnerable person; preventing the suspect from causing injury; or allowing for the prompt and effective investigation of the offense (gov.uk, 2012; Her Majesty's Inspectorate of Constabulary, 2014, p. 75). In many cases, these conditions may not be satisfied. Perpetrators are more likely to be arrested if they are present when the police arrive, if they have injured the victim, or if the victim is willing to make a formal statement to the police (Myhill, 2019; Richards and Harinam, 2020; Koppensteiner et al., 2020).

Electronic copy available at: https://ssrn.com/abstract=4336508

Under English law, charges may be initiated by the police regardless whether they make an arrest. Indeed roughly one in seven DA investigations involves a "voluntary attendance" on the part of the perpetrator, rather than an arrest (Office for National Statistics, 2021). Either way, if the police decide to pursue charges, they refer them to the Crown Prosecution Service (CPS).[3] The standard for pressing charges, for both police and prosecutor, is twofold. First, there must be "sufficient evidence to provide a realistic prospect of conviction," and second, it must be in the public interest to pursue prosecution (Her Majesty's Inspectorate of Constabulary, 2014, p. 98).

In the analysis to follow, we focus on the police decision to refer charges to CPS. Although other prosecution decisions (such as CPS accepting the charges) would also be interesting to study, our focus on interventions that are initiated by police leads us to study this first step in prosecuting the perpetrator.

3. Data

We analyze data provided by Greater Manchester Police (GMP), a police force serving a population roughly the same size as that of the city of Chicago. During the period we study, all calls for service were recorded as an incident in GMP's command-and-control database.[4] Once they had been investigated, incidents that were determined to be crimes were recorded in the crime database. In addition, all incidents that were determined to be domestic abuse were recorded in a separate DA database, whether they were classified as crimes or not.

The DA database is a key source of data for our study. It includes information on the date, time, and location of the incident, other characteristics of the incident, whether it was classified as a crime, whether the police pursued charges against the perpetrator, and if so, the charge referred to CPS.

Each incident in the DA database can be linked to victim and suspect databases, which provide basic information about the parties involved. Because incidents involving the same parties can be linked together, we can construct measures of recidivism as well as histories of past crimes and domestic abuse. The crime file can also be linked to information about the responding officer and the DASH reports. The crime, victim, suspect, and DASH databases are the sources of the data we analyze here. A different database, the custody database, contains information about arrests, bail conditions, and any pre-trial incarceration spells. Unfortunately, the crime and custody databases were never meant to be linked, and our attempts to do to so were not successful.

[3] Although the police generally may decide to file formal charges in less-serious cases, all decisions to file in domestic abuse cases are made by CPS (Crown Prosecution Service (2020), Appendix 1).

[4] In mid 2019, GMP adopted a completely new recording system. It is not backward compatible with the system that was in place during our sample period.

Electronic copy available at: https://ssrn.com/abstract=4336508

In England and Wales, domestic abuse is broadly defined to include incidents between individuals age 16 years or older who are or have been intimate partners or family members (Brown 2020). This can include incidents between siblings, incidents between adult children and parents, or intimate-partner incidents involving current or past spouses or romantic partners. Since the great majority of domestic abuse incidents involve heterosexual intimate partners, we restrict attention to those calls.[5]

We seek to estimate whether criminal charges or the high-risk designation from the DASH process reduce domestic abuse recidivism involving violence with injury or a sex offense that is reported to police. Hereafter, we refer to this as violent recidivism so as to economize on language. We define violent recidivism at the level of the dyad. We code it as any reported DA incident involving violence with injury or a sex offense that occurs within one year of the current call from the same dyad, ignoring multiple calls on the same day.[6] Of course, many DA incidents are not reported to police; we are unable to analyze those incidents.[7]

In principle, recidivism could be defined differently, for example, on the basis of future calls for service. We focus on violent recidivism for several reasons. First, incidents involving violence are more serious than non-violent incidents, and more costly to both the victim and law enforcement. Second, changes in calls for service could either reflect changes in underlying behavior or changes in reporting. Changes in incidents involving violence are more likely to reflect changes in underlying behavior, since more-severe incidents of domestic abuse are more likely to be reported to police (Barrett et al., 2017). Finally, the DASH/MARAC process is intended to predict and prevent serious harm, not merely repeat calls for service (Richards, 2009; Her Majesty's Inspectorate of Constabulary, 2014).

We define our measure of violent recidivism with the goal of capturing a reasonable notion of serious harm, given the offense categories recorded in the GMP data. Violence with injury consists primarily of two classes of offenses: Wounding with Intent to do Grievous Bodily Harm (GBH) and Assault Occasioning Actual Bodily Harm (ABH). GBH includes injuries resulting in permanent disability; permanent, visible disfigurement; compound fractures; substantial loss of blood; lengthy treatment, or psychiatric injury. GBH clearly accords with any notion of serious harm; the maximum sentence for GBH is life imprisonment. ABH also includes injuries involving considerable harm, such as broken noses, broken fingers, loss of teeth, and shock, as well as lesser injuries such as grazes, swelling, and black eyes. It carries a maximum sentence of five years. It is important to note that violence with injury does not include

[5]Intimate partners include ex-partners, partners, wives, girlfriends, ex-wives, husbands, boyfriends, ex-husbands, and civil partners.

[6]Dyads are defined to involve the same two people, but not necessarily in the same roles as victim and perpetrator.

[7]Since our data come from GMP's command and control database, we may also miss any incidents that take place in another police jurisdiction.

Electronic copy available at: https://ssrn.com/abstract=4336508

Common Assault, which may entail slaps, punches, or other attacks that leave no visible mark or injury, and which carries a maximum sentence of six months (Home Office 2020).

As a measure of serious harm, ABH may seem too inclusive, whereas GBH seems too restrictive. A final statistical consideration led us to focus on violence with injury: GBH accounts for only about 1 percent of the domestic abuse cases in our sample. Focusing on such a statistically rare outcome would make it almost impossible to detect any effect of our interventions, even with a substantial sample.[8]

The variables that we use in our analysis, combined with the time periods over which the various components of our data are available, define our sample period. The call-for-service and crime data are available from April 2008 to July 2019. The DASH data are available from July 2013 to July 2019. Information linking victims and perpetrators is available beginning in April 2012. We include in the sample only those dyads whose first call for service took place after April 2014. This is to avoid truncating the two-year criminal and DA history measures that we use as predictors. To ensure that we have a full year to measure violent recidivism for all DA calls, we only include calls that occurred before July 2018. Figure 1 illustrates this timeline.

Figure 1: Timeline showing availability of data and sample period

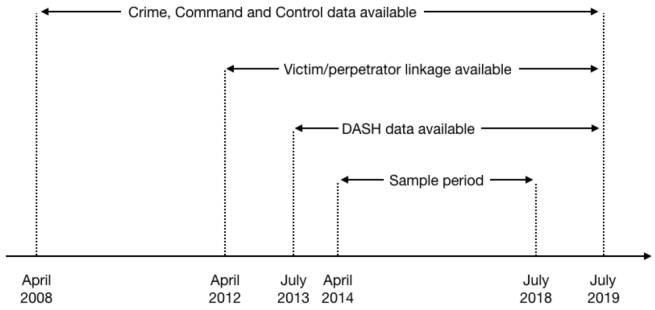


Table 1 presents cross-tabulations of our outcome with each of our treatment variables. Each cell presents the cell size, the row share, and the column share. Panel A cross-tabulates violent recidivism with charge status. The last row shows that violent recidivism occurs in 12.99 percent of the 154,102 DA incidents in our sample. The last column shows that charges are pressed in 11.66 percent of those cases.

Panel B cross-tabulates violent recidivism with high-risk status. As mentioned above, the responding officer administers the DASH questionnaire, then grades the case as standard, medium, or high risk. Since only high-risk cases are referred to MARAC services, we dichotomize this measure into two categories: high-risk and other. The last column shows that just under 9 percent of all incidents are graded as high-risk.

[8]Turner et al. (2019) seemingly worked with an earlier version of the Home Office classification system, which allowed them to distinguish finer categories of offenses.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 1: Recidivism rates by charge and high-risk status

A. Charges			
	Recidivism = no	Recidivism = yes	Total
Perpetrator Charged = no	118149 86.79 88.11	17988 13.21 89.89	136137 100 88.34
Perpetrator Charged = yes	15942 88.74 11.89	2023 11.26 10.11	17965 100 11.66
Total	134091 87.01 100	20011 12.99 100	154102 100 100

B. High-risk			
	Recidivism = no	Recidivism = yes	Total
High Risk = no	122539 87.36 91.38	17734 12.64 88.62	140273 100 91.03
High Risk = yes	11552 83.53 8.62	2277 16.47 11.38	13829 100 8.97
Total	134091 87.01 100	20011 12.99 100	154102 100 100

Incidents in which charges are pressed are a bit less likely to result in violent recidivism than incidents in which they are not. The recidivism rate for incidents where charges were pressed is 11.26 percent, compared to 13.21 percent for incidents that did not result in charges. In contrast, incidents graded as high-risk are more likely to result in violent recidivism. The recidivism rate was 16.47 percent for incidents graded as high-risk, compared to 12.64 percent for incidents graded as standard- or medium-risk.

If one took these numbers at face value, one would conclude that charges reduced violent recidivism by about two percentage points, whereas the DASH/MARAC process appears to have backfired, raising violent recidivism by nearly four percentage points. However, these simple comparisons are unlikely to reveal the causal effects of the interventions, since treated incidents and their participants may differ from untreated incidents in ways that may be correlated with both treatment status and the outcome. If perpetrators who are charged or graded as high-risk would be more likely to recidivate in the absence of the interventions, then one would expect simple comparisons such as these to understate any favorable effects they might have,

Electronic copy available at: https://ssrn.com/abstract=4336508

or even suggest perverse effects.

In order to control for such differences, we estimate treatment propensities using a number of predictor variables. These variables are constructed from the command-and-control and crime files and are summarized by treatment status in Table A1. They capture characteristics of both the incident and its participants. We discuss the variables below when we discuss the estimated propensity score models.

4. Methods

Our primary method for estimating the average effect of treatment on the treated (ATT) is inverse propensity score weighting (IPW). IPW provides a straightforward approach to estimation and allows us to deal with potential complications involving multiple treatments and persistent treatment effects. Our primary approach to treatment effect heterogeneity involves training a causal forest (Athey et al., 2019; Wager and Athey, 2018), which estimates conditional average treatment effects (CATE) for each incident in the sample. We then learn a decision tree from those estimates to understand how the CATEs differ for different values of the predictors.

4.1. Inverse propensity-score weighting to estimate the ATT

The ATT measures how charges affect perpetrators who are charged, on average, or how the MARAC process affects cases labeled as high-risk by the DASH protocol. Treating each intervention separately for now, let $D_i = 1$ if the ith incident is treated, that is, receives the intervention, and let $D_i = 0$ otherwise. Define potential outcomes $Y_i(D_i)$ as a function of treatment status, so $Y_i(1)$ represents the outcome for incident i if treated, and $Y_i(0)$ represents the outcome if not treated. The observed outcome is given by $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$, which shows that we only observe the potential outcome associated with the observed treatment status.[9]

Let X_i denote a vector of covariates that may predict treatment status and potentially confound the relationship between Y_i and D_i. Define the propensity score $p(X_i)$ as the probability of treatment conditional on the predictors, that is $p(X_i) = P(D_i = 1|X_i)$. Identification relies on the Common Support and Conditional Independence assumptions. Common support requires the domains of the propensity score to be the same for treatment and comparison groups, or equivalently, that

$$0 < P(D_i = 1|X_i) < 1$$

[9]Strictly speaking, we observe an unbalanced panel of dyads, since some dyads are involved in multiple incidents. Since our estimators make no explicit use of this fact, we use a single subscript to denote incidents, rather than more cumbersome double subscripts denoting dyads and incidents. However, in computing standard errors and hypothesis tests, we always make use of robust covariance matrix estimators that account for any dependence which may arise from the presence of multiple incidents per dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508

for all values of X_i. Conditional Independence requires potential outcomes to be independent of treatment, given the predictors, that is,

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i|X_i,$$

where $\perp\!\!\!\perp$ denotes statistical independence. With these two assumptions, Rosenbaum and Rubin (1983) showed that

$$\{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i|X_i => \{Y_i(0), Y_i(1)\} \perp\!\!\!\perp D_i|p(X_i)$$

That is, conditioning on the propensity score renders treatment status independent of potential outcomes. Under these assumptions, IPW weighting identifies the average effect of treatment on the treated Δ^{ATT} from the observed data, since one can show that

$$\Delta^{ATT} \equiv E[Y_i(1) - Y_i(0)|D_i = 1] \tag{1}$$

$$= E[D_i Y_i – (1 - D_i)\frac{p(X_i)}{1 - p(X_i)} Y_i]. \tag{2}$$

As discussed above, CIA is a strong assumption. It is untestable, since it involves both the observed and counterfactual potential outcomes. However, we can test a key property of our estimated propensity scores. Rosenbaum and Rubin (1983) show that the true propensity score balances the observables. Therefore, if an estimated propensity score fails to yield balance, it cannot be the true propensity score.

We estimate the propensity score in several different ways. First, we estimate a logistic regression model as a function of X_i, where the dependent variable is the treatment indicator. This yields a set of coefficients $\hat{\beta}$, predicted values $\hat{p}_i$, and ATT weights $w_i = D_i + (1 - D_i)(\hat{p}_i/(1 - \hat{p}_i))$. We then test for balance by regressing the treatment dummy D on the predictors X, weighting the observations by the estimated ATT weights. We test the null hypothesis that the coefficients in this regression are jointly zero.

If the joint F-statistic rejects the null, we expand the set of predictors used to estimate the propensity scores. We do this by adding squared and and first-order interaction terms for all variables with absolute t-statistics greater than 10.[10] We then estimate the expanded model and test for balance again. If necessary, one could continue iterating along these lines, using lower threshold values for the t-statistics, until the estimated propensity scores balanced the predictors.

The logic underlying this approach stems from the discussion above. Since any candidate propensity score that does not yield balance cannot be the true propensity score, we experiment with the functional form, in hopes that sufficient flexibility will achieve balance.

[10]For dichotomous predictors, we can only add the interaction terms.

Electronic copy available at: https://ssrn.com/abstract=4336508

Our second approach is to estimate the propensity scores using a random forest, which can be thought of as an alternative approach to finding the correct functional form. A random forest involves fitting a pre-specified number of regression or classification trees, where each tree is built from a random subsample of the data, and fit to a random subsample of the predictors. The trees are built recursively. At each node, all values of the selected predictors are scanned so as to split the node in the manner that maximizes the variance in the target variable, which is this case is the treatment indicator, between the resulting child nodes (Hastie et al., 2009). Predictions are made by averaging over trees which did not make use of the *i*th observation. Unlike logistic regression, random forests produce no coefficients. Their main virtue is that they provide a non-parametric procedure which, in principle, is capable of learning the correct functional form of the propensity score from the data.

Our third approach involves the Covariate Balancing Propensity Score (CBPS) of Imai and Ratkovic (2014). Rather than focusing on functional form, CBPS achieves balance using the baseline set of predictors. It posits a logistic model for the probability of treatment, but then estimates the parameters of that model by solving a set of equations that directly impose balance on the covariates.[11]

Once the propensity score is estimated, the ATT is estimated by means of a weighted regression of the dependent variable on the treatment dummy, using the estimated ATT weights. The ATT weights take high values for comparison observations with high propensity scores, that is, whose probability of treatment was high, despite not being treated. Presumably, such observations are more like those that actually received treatment than comparison observations with low propensity scores. Those comparison observations are down-weighted.

4.2. Causal forests to estimate heterogeneous treatment effects

For our main heterogeneity analysis, we estimate and analyze causal forests (Wager and Athey, 2018; Athey et al., 2019). Causal forests are an adaptation of the random forests discussed above. Just as random forests are an ensemble of a pre-specified number of regression trees, causal forests are composed of a number of causal trees.

Like regression trees, causal trees are grown by recursive partitioning over random samples of the data, but they differ from regression trees in a few important ways. First, the natural target variable, the idiosyncratic treatment effect, is unobserved. As a result, the causal trees estimated by the R package *grf* target a gradient-based approximation to the node-specific treatment effect Athey et al. (2019); Athey and Wager (2019). Second, the trees make use of what Athey and Imbens (2016) refer to as honest splitting. In honest splitting, one random subset of the data is used to set the splits, and another is used for estimation.

[11] The Appendix provides more details on estimation.

Electronic copy available at: https://ssrn.com/abstract=4336508

Finally, whereas random forests aggregate regression trees by averaging, causal forests aggregate the causal trees in a different manner. Rather than averaging, the causal trees are used to estimate a set of weights that show how close each observation is to the others, in the sense of how frequently they share the same terminal node of a tree. These weights are then used to estimate the CATE for each incident in the sample, defined as

$$\tau(x) = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]. \tag{3}$$

The CATEs are estimated by means of a weighted regression of the dependent variable on the treatment variable, where both have been pre-residualized with respect to X to deal with confounding. Athey et al. (2019) show that under conditional independence, common support, and some other regularity conditions, the resulting CATEs are consistent and asymptotically normal.[12]

The procedure also yields a set of "doubly-robust scores," which are equal to the estimated CATE plus a bias-correction term. These can be averaged to estimate the ATT. We can thus compare whether the causal forest yields an estimate similar to that produced by IPW weighting. We also analyze the doubly-robust scores to uncover heterogeneity in the estimated treatment effects.

5. Results

5.1. IPW estimates of the ATT

5.1.1. Estimating the propensity score

We begin by reporting estimated coefficients. Coefficients for charges appear in Appendix Table A2, and those for high-risk appear in Appendix Table A3. These tables report coefficients for the parametric models only; the non-parametric random forest procedure generates no coefficients. The models in columns (1) and (3) include the baseline set of predictors, which consists of the variables shown in Appendix Table A1.[13] Coefficients in the first column are from a logistic regression; those the third column are estimated via CBPS. The coefficients in the second columns are from a logistic regression that includes the expanded set of predictors chosen by one iteration of the procedure described above.[14]

The coefficients in columns (1) and (3) are directly comparable. Those in column (2) are not, due to the different functional form. For the most part, the coefficients from the logistic

[12]Consistency also requires that residualization be done on a leave-out basis. This means that an observation cannot be used to estimate the predicted value that is used to residualize it. More details about causal forests are provided in the Appendix.

[13]For the parametric procedures, for each categorical variable, we exclude one of the categories in Appendix Table A1.

[14]All standard errors, reported in parentheses, are clustered by dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508

regression in column (1) and the CBPS estimator are similar. The first row of Appendix Table A3 shows that incidents classified as crimes are more likely to be graded as high risk. Note that this variable cannot be included in the model for charges, since it is a perfect predictor. If an incident is not classified as a crime, then charges cannot be filed.

The next two rows of both tables show that injuries elevate the likelihood of treatment. Myhill (2019) similarly reports that police are more likely to make an arrest when an injury has occurred. Likewise, our findings are consistent with those from Whinney (2015), who found that victim injuries make a high risk grade more likely. Perpetrator injuries have a smaller effect on treatment status.[15]

The next several rows show the effects of alcohol and drug involvement among the parties to the incident. Perpetrator alcohol use raises the likelihood of being charged, but lowers the likelihood of a high risk grade. Perpetrator drug use raises the likelihood of both interventions, whereas victim drug use reduces the likelihood that charges are filed.

Incidents involving former (as opposed to current) partners are more likely to result charges but less likely to lead to a high risk grade. Incidents occurring at the victim's home are more likely to result in both interventions, whereas the role-switch indicator, which equals one if the dyad had a prior incident in which the parties played the opposite roles from those in the current incident, reduce the probability of both interventions. Weekend incidents are more likely to result in charges, but less likely to result in a high risk classification. Incidents that take place on a holiday are less likely to result in a high risk grade.

The next five variables summarize the victim's responses to the DASH questionnaires.[16] The standard DASH questionnaire contains 27 questions; GMP adds a 28th question for the officer, asking whether he/she gathered any other relevant information about the case.[17] For dyads with multiple incidents, the responses to the questions (as well as the officer's risk grade) may vary from incident to incident, reflecting the dynamics of the situation.

For each question, there were two possible responses, yes or no. Responses were often omitted as well. A bit of experimentation revealed that the sums of the number of yeses and the number of omissions (including the roughly 10 percent of incidents which consisted entirely of omitted responses) were highly predictive of a high risk grade. A higher number of yeses positively predicts both charges and a high risk grade, although there is a non-linearity above 13

[15]Since most of our predictors are dichotomous, their coefficients can be interpreted as the change in the log-odds of treatment as the predictor changes from zero to one. Thus the coefficients provide a meaningful quantitative metric of the importance of the predictors.

[16]The primary victim provides the answers to the DASH questionnaire. Standard procedure is to separate the parties before administering the protocol. In a small number of cases there are multiple victims, who are most often underage children. Our original datasets contain only data on the primary victim as recorded by the police. We do not have data on secondary victims.

[17]The DASH questionnaire is included in Richards (2009).

Electronic copy available at: https://ssrn.com/abstract=4336508

yeses. Given the total number of yeses, an incident with 14 or more yes responses is less likely to result in charges, but more likely to be graded as high risk. The latter finding is consistent with Whinney (2015). He indicates that early guidance suggested that 14 or more yeses should automatically result in a high risk grade.

The next three variables are dummy variables indicating whether the dyad had been involved in an incident graded as high-risk during the previous three, six, or 12 months. These variables are moderately predictive of charges, but they are highly predictive of a high risk grade on the current incident. This is again consistent with Whinney (2015), who indicated that any high-risk grade in the past year should lead to an automatic high risk grade at the current incident.

The next variables characterize the police officer who responds to the DA call.[18] Male officers are slightly more likely to press charges than female officers, whereas more experienced officers are less likely to press charges than less experienced officers. Officers who charge a higher share of incidents are unsurprisingly more likely to press charges in the current case, but are also less likely to grade it as high-risk. For officers with high shares of high risk grades, the opposite is true. A higher share of blank DASH questionnaires is associated with a higher likelihood of both interventions.[19]

The next several rows summarize the dyad's DA histories, including incidents, crimes, and incidents involving violence. We constructed indicators equal to one if such an event occurred during the past 3, 6, and 12 months. We also constructed variables reflecting the dyad's history over two years. For incidents and crimes, we constructed two indicators. One was equal to one if the dyad had one such incident, and the other was equal to one if they had more than one incident. For incidents involving violence, which is less common, the indicator equals one if there are any such incidents. Taken as a whole, these variables are moderately predictive of treatment status. They are somewhat more predictive of violent recidivism (Grogger et al., 2021).

Calls and crimes over the last three months to two years are generally predictive of both charges and a high risk grade. Past violence is predictive as well, but more for high risk grades than for charges.

Male perpetrators are more likely to be charged than female perpetrators. The next variables reflect the perpetrator's DA and offending history over the previous two years. They are coded in the same way as the two-year histories for the dyad, described above. The perpetrator histories may reflect incidents both within and outside the dyad, although in fact, over a

[18]These are based on the characteristics of the officer who filled out the DASH report. We do not have information about any other officers who may have been present.

[19]These latter variables are calculated on a leave-out basis. They are based on averages, or sums, that exclude the current incident and any other incidents involving the same dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508

two-year horizon, the perpetrator histories mostly reflect events within the dyad. Despite this collinearity, perpetrator histories are strongly predictive of treatment status. This is particularly so for crimes and incidents involving violence over the two years prior to the current incident.[20]

Perpetrators who have violated a protection order or been accused of stalking during the two years prior to the current incident are likely to be charged. Violating protection orders has little effect on the high risk grade, although stalking raises its likelihood. Finally, the age distribution of victims and perpetrators predict treatment status as well.

Turning from the coefficient estimates to the predicted propensities of treatment, Figure 2 presents histograms of the estimated propensity scores by estimation method and treatment status. For each estimation method, we present two graphs. The first depicts histograms by treatment status for the full domain of the propensity score. The second truncates the distribution from below at 0.5, allowing for a closer look at the incidents whose estimated propensity of treatment is relatively high. We plot the histogram for the comparison observations on the axis extending upward from the origin, and that for the treatment observations on the axis extending downward. Note that the y-axis scales are quite different above and below the x-axis, owing to the many-fold greater number of comparison incidents.

Focus first on the propensity scores from the logistic regression with the baseline predictors, in the top row. Not surprisingly, we see a large number of comparison observations with low propensity scores. Relative to the comparison group, the treatment-group histogram is shifted somewhat to the right. For the relatively high-propensity incidents depicted on the right-hand graph, there are more similar absolute numbers of treatment and comparison observations. Put differently, common support is satisfied over nearly all the domain. The exception is at the farthest end of the right tail, where two treatment observations had estimated propensity scores that exceeded the highest value among the comparison group.

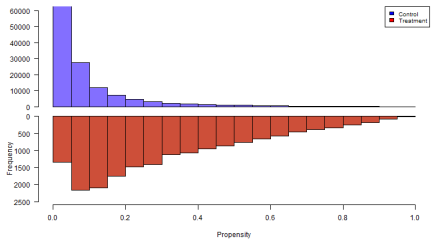
[20]We similarly constructed two-year histories of the victim. Because most events over a two-year horizon involve the same partner, they were highly collinear with history of the perpetrator (and the dyad). We dropped them to lessen the overall collinearity.

Electronic copy available at: https://ssrn.com/abstract=4336508

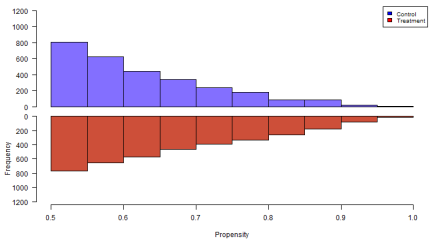
Figure 2: Histograms for estimated propensity scores for charges, by treatment status and estimation method

A. Logistic regression with baseline set of predictors

Unconditional

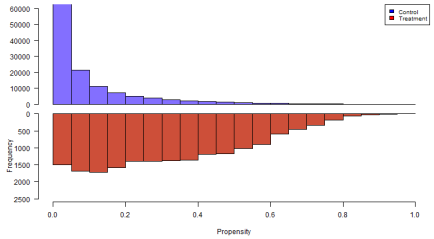


Conditional on $\widehat{p} > 0.5$

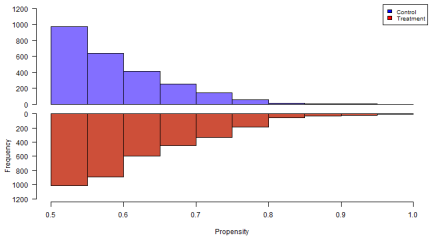


B. Logistic regression with expanded set of predictors

Unconditional

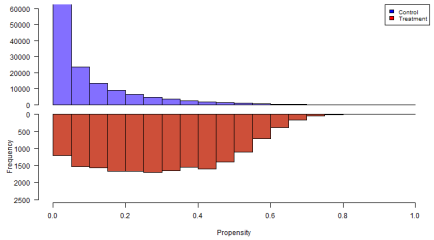


Conditional on $\widehat{p} > 0.5$

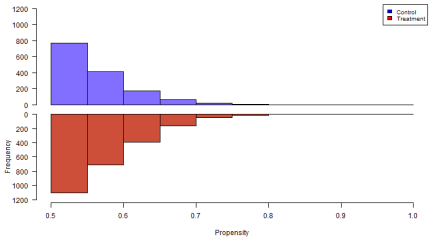


C. Random forest

Unconditional

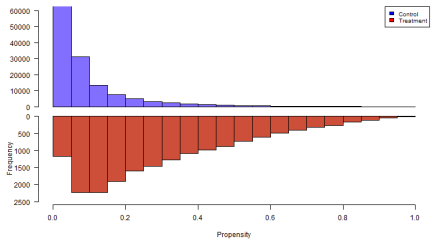


Conditional on $\widehat{p} > 0.5$

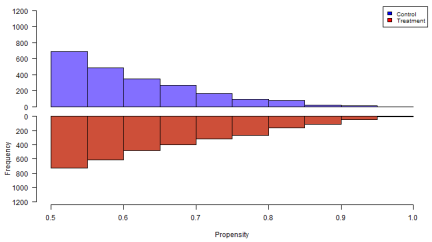


D. CBPS

Unconditional



Conditional on $\hat{p} > 0.5$



Note: Propensity scores estimated on the full sample, N=154,102.

The second row shows histograms from the logistic regression with the extended set of predictors. Compared to the more restrictive model above, this model places a few more comparison observations in the lowest bin, and more observations from both groups in the right

Electronic copy available at: https://ssrn.com/abstract=4336508

tail.

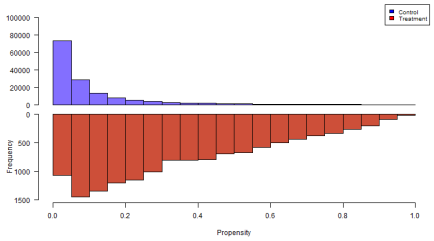
The next row shows histograms for propensity scores from the random forest, which are based on 2000 trees. Compared to the logistic regression plot, the histogram for the comparison group is fairly similar. That for the treatment group has more weight in the center of the distribution. The linear decline in frequencies is also evident in the truncated histogram on the right. The bottom row of the Figure shows the histogram for the CBPS estimates. It resembles the histogram for the baseline logistic model, except that it places a bit less weight in the bottom of the comparison group distribution.

Electronic copy available at: https://ssrn.com/abstract=4336508

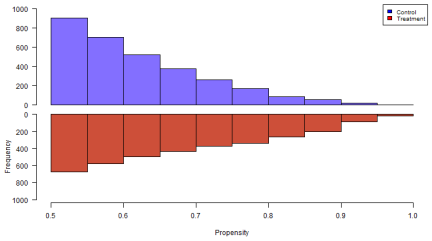
Figure 3: Histograms for estimated propensity scores for high-risk, by treatment status and estimation method

A. Logistic regression with baseline set of predictors

Unconditional

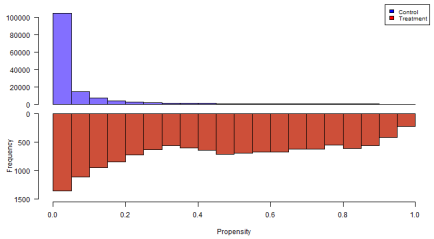


Conditional on $\widehat{p} > 0.5$

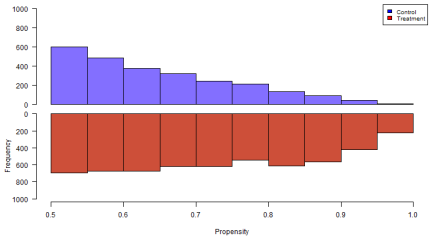


B. Logistic regression with expanded set of predictors

Unconditional

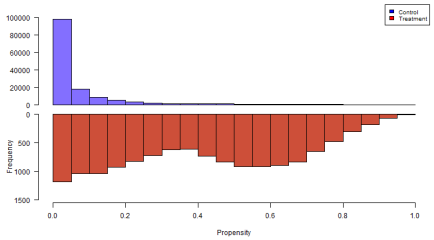


Conditional on $\widehat{p} > 0.5$

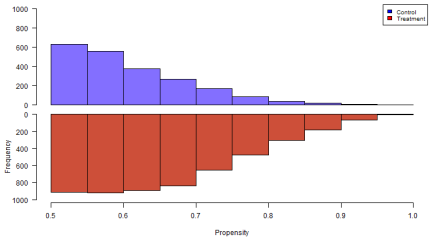


C. Random Forest

Unconditional

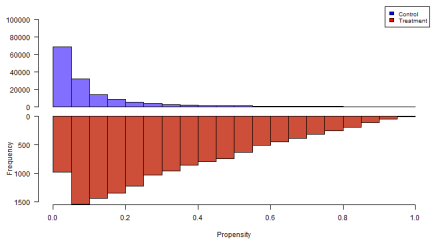


Conditional on $\widehat{p} > 0.5$

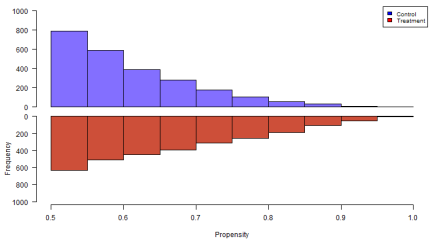


D. CBPS

Unconditional



Conditional on $\hat{\hat{p}} > 0.5$



Note: Propensity scores estimated on the full sample, N=154,102.

Figure 3 reports histograms for the propensity of a high risk grade by estimation method. Its layout is the same as in Figure 2. The baseline logit model places roughly 50,000 observations in the lowest bin of the histogram, and nearly 30,000 in the next-lowest. Nonetheless, both

Electronic copy available at: https://ssrn.com/abstract=4336508

distributions extend nearly to the point where $p(X) = 1$; only 36 treatment-group estimates exceeded the maximum comparison-group propensity score. The expanded logistic regression model places more observations from both groups in the lowest bin. Comparatively, the random forest yields a similar histogram for the comparison group, but yields a bi-modal histogram for the treatment group. The histogram from the CBPS model is most similar to that from the baseline logit model. These various propensity scores are the basis for the ATT weights which we use to test for balance.

5.1.2. Balance

Tests for balance are reported in Table 2. The top panel reports results for charges; the lower panel reports results for high-risk. The first row reports the F-statistic for the regression of the treatment dummy on all the variables used to estimate the propensity score. It is based on a covariance matrix that is clustered by dyad. The second and third rows present the numerator degrees of freedom and the p-value for the test of balance.

The first two columns report results from an unweighted regression. One uses only the baseline predictors, the other uses the expanded set of predictors. We present these as a benchmark. They show how imbalanced the treatment and comparison groups are without weighting, and help demonstrate the role that the expanded predictors play in helping to achieve balance. The remaining columns report results that adjust for imbalance using the ATT weights estimated from the different models.

Focusing first on the unweighted results for charges, the F-statistic in column (1) is 288.97, indicating that the baseline predictors are highly imbalanced between the treatment and comparison groups. Adding the additional predictors reduces overall imbalance a bit, resulting in an F-statistic of 140.8. Weights based on the logistic regression with baseline predictors improve balance a great deal, reducing the F-statistic from 288.97 to 1.84. Similarly, weights based on the expanded set of predictors reduce imbalance by three orders of magnitude, lowering the F-statistic from 140.8 to 0.273. With the expanded logit weights, we clearly fail to reject the null of balance between treatment and comparison groups. Weights based on the random forest do not fare as well, yielding an F-statistic of 12.65.[21] CBPS yields the smallest F of all, at 0.004. These weights also balance the predictors.

Results for high risk appear in Panel B. As above, the unweighted results show that the predictors are highly unbalanced across the treatment and comparison groups. Weighting improves balance greatly. The weights from the baseline logistic regression reduce the F-statistic from 256.23 to 1.406, whereas those from the expanded logit model reduce it from 236.8 to 0.874. With the weights from the expanded logit model, we do not reject the null that the predictors

[21] Goller et al. (2019) find that random forests may perform worse than parametric methods when the treatment condition is rare relative to the control condition.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 2: Balance Tests

A. Charges						
Weights from:	Unweighted		Logistic regression		Random Forest	CBPS
Predictors	Baseline	Expanded	Baseline	Expanded	Baseline	Baseline
Joint F for balance	228.969	140.818	1.843	0.273	12.973	0.004
Numerator df	79	146	79	146	79	79
p-value	0.000	0.000	0.000	1.000	0.000	1.000
B. High-risk						
Weights from:	Unweighted		Logistic regression		Random Forest	CBPS
Predictors	Baseline	Expanded	Baseline	Expanded	Baseline	Baseline
Joint F for balance	256.234	236.767	1.406	0.874	14.554	0.025
Numerator df	80	111	80	111	80	80
p-value	0.000	0.000	0.010	0.825	0.000	1.000

Notes: F-statistics are from a regression of the treatment variable on the indicated set of predictors, using ATT weights produced by the indicated statistical model. The other two numbers in each column are the numerator degrees of freedom and p-value associated with the F-statistic. The unweighted results are provided as a point of reference, showing how imbalanced the predictors are in the raw data.

are balanced. Once again, the weights from the random forest perform less well than those from the expanded logit. Weights from the CBPS estimator again yield the smallest F-statistic of all.

5.1.3. Estimates of the ATT

Table 3 presents estimated ATT's by the method used to estimate the propensity scores. These were estimated by a weighted regression of the violent recidivism outcome on the treatment dummy. Weights are given as above for the balance tests.[22]

The estimated treatment effects for charges appear in the top panel of the Table. All of the estimates are negative and statistically significant at the 1 percent level. They are fairly similar in magnitude as well, ranging from -0.048 to -0.059.

One might ask why the estimates are so similar, considering that only two of the four sets of propensity score weights yielded balanced predictors. We suspect that this stems from the fact that all of the weights greatly reduce imbalance, whether or not they formally fail to reject the null. In any case, there is no clear pattern between the balance statistics and the estimated ATT's. The two sets of weights that deliver the largest F-statistics yield both the smallest and largest ATT's.

[22]Here and below, we drop treatment observations with treatment propensities greater than the maximum propensity among the comparison group. Depending on the intervention and the estimation method, this ranges from zero to 81 incidents.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 3: Estimates of the average treatment effect on the treated (ATT)

A. Charges				
Weights from:	Logistic regression		Random Forest	CBPS
Predictors	Baseline	Expanded	Baseline	Baseline
Charges	-0.059	-0.05	-0.049	-0.052
	(0.005)	(0.004)	(0.003)	(0.004)
B. High-risk				
Weights from:	Logistic regression		Random Forest	CBPS
Predictors	Baseline	Expanded	Baseline	Baseline
High-risk	-0.006	-0.005	-0.004	-0.003
	(0.010)	(0.007)	(0.005)	(0.008)

Notes: ATT estimates are coefficients are from a weighted regression of the violent recidivism indicator on the treatment variable, using ATT weights from the indicated statistical model. Standard errors are clustered at the level of the dyad.

Compared to the mean rate of violent recidivism shown in Table 1, the estimates show that charging the perpetrator reduces the violent recidivism rate among those charged by roughly 37 to 45 percent. Comparing this to the estimated effect of arresting the perperator, it is a bit lower than the reduction of 50 percent from the MDVE, but much larger than the reduction of 4 percent from the re-analysis of the SARP experiments by Maxwell et al. (2002). Our results are also comparable to estimates from Amaral et al. (2022), who report proportionate reductions of about 50 percent.

The estimated treatment effects for the high-risk grade appear in the bottom panel. Although all the coefficients are negative, they are quite small, and none are significant. Nevertheless, the estimates are fairly precise, since we can reject the hypothesis that the effect of the DASH/MARAC process is even one-half as large as the effect of pressing charges. Despite the resources that are devoted to the risk assessment process, it appears to have no effect on violent recidivism, on average.

Before moving on, we note that the estimates in Table 3 are more negative, or less positive, than the estimates based on the raw data in Table 1. Adjusting for the predictors in our sample reduces the effect of charges from about -0.02 to about -0.05. It reduces the estimated effect of the high-risk intervention from about 0.04 to roughly zero. This is precisely the direction of bias one would expect if the probability of treatment were higher for perpetrators who were more likely to recidivate in the absence of treatment.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 4: Estimates of ATTs for multiple treatments

	ATT	SE	N
High-risk only	0.016	0.007	136099
Charge only	-0.045	0.004	140258
Charge and high-risk	-0.054	0.010	133680

Notes: ATT estimates are coefficients are from a weighted regression of the violent recidivism indicator on the given treatment variable, formed from interactions between the charge indicator and the high-risk indicator. Each row is from a separate regression that includes the relevant treatment group and the comparison group of non-charged, non-high-risk incidents. Sample sizes (N) are the sum of the number of incidents in those two groups. Separate propensity score models estimated by CBPS using predictors described in text. Standard errors are clustered at the level of the dyad.

5.1.4. Robustness and interpretation

Here we present several analyses that probe the robustness of our estimates and help us interpret them. One issue is that, to this point, we have treated charges and the high-risk intervention in isolation. However, as Table 1 shows, some incidents resulted in both a criminal charge and a high-risk classification. Here we allow for multiple treatments, and ask how doing so affects our conclusions.

We proceed by estimating the effects of interactions between our two binary treatment variables. That is, we consider three mutually exclusive interventions: being graded as high-risk only, being charged only, and being both charged and graded as high-risk. In each case, the comparison group consists of those incidents which resulted in neither a criminal charge nor a high-risk grade.

Since our focus is on the ATT, we need only compare each of the three intervention groups to the comparison group one-at-a-time (Lechner 2001; McCaffrey et al 2013). This simplifies estimation, since we require only three separate binary choice models to estimate the propensity scores. We estimate these models via CBPS, then estimate the ATT's by running three regressions. Each regression sample consists of the comparison group and the group that experienced the relevant intervention. The regressions include the relevant intervention indicator and are weighted by the corresponding ATT weights.

Estimated ATT's appear in Table 4.[23] [24] The estimated effect of a high-risk grade alone is

[23]The sample size in the last column is the sum of the number of comparison observations and the number of observations receiving the indicated treatment.

[24]Here and for the remainder of this section, we restrict attention to estimates that are based on the CBPS

Electronic copy available at: https://ssrn.com/abstract=4336508

now positive and significant. The effects of charge-only and the combined charge and high-risk intervention are both are negative and significant. Comparing these estimates to those in Table 3, it appears that the negative (though insignificant) effect there of high-risk stemmed from the combination of the positive high-risk-only effect and the negative effect of charges and high-risk combined. At the same time, charges have similar effects, regardless whether they are accompanied by a high-risk grade. Either way, the effect of charges in Table 4 is similar to that from the simpler model reported in Table 3.

Another question is whether our estimates could be biased due to misspecified dynamics. Forty-two percent of our sample dyads contribute multiple incidents to the sample. Several of the controls in our model, such as participants' criminal histories, evolve over time. If past treatment status also affect current treatment status, then our static approach may be misspecified.

Several approaches have been proposed recently for estimating dynamic treatment effects (Lewis and Syrgkanis 2021; Vandenberg and Vikstrom 2022). However, those methods typically assume that the data are observed at regular intervals, such as once per year. Those methods are not well suited for our setting, where new incidents generate new data, and those incidents do not occur at regular intervals.

We take two related approaches to this problem. First, we divide the sample into two subsamples, one containing only initial incidents, and the other containing all 2nd- and higher-order incidents. We then estimate models along the lines discussed so far, estimating separate propensity scores for the two subsamples. The virtue of this approach is that predictors at the initial incident are unaffected by treatment status at the initial incident, and current treatment status is unaffected by past treatment status. Thus the estimates from the initial incidents will be unaffected by any potentially misspecified dynamics. Differences in estimated ATT's between these two samples may stem either from dynamic misspecification in the sample of higher-order incidents, or simply from parameter heterogeneity between first and later incidents.

Second, to draw this distinction, we allow for first-order dynamics in the effect of treatment on the outcome. To do so, we define interventions by the interaction between treatment status at incident t and treatment status at incident $t-1$. That is, we estimate the effect of being charged at $t-1$ but not at t, of being charged at t but not at $t-1$, and the effect of being charged at both $t-1$ and t. In each case, the comparison group consists of incidents in which charges were not filed at either $t-1$ or t. This is analogous to the approach we took above to estimate the effect of multiple treatments, although now we define interventions in terms of being charged at two different incidents, rather than in terms of being charged and/or graded as high risk at the same incident. We carry out the same procedure separately for the high-risk intervention.

weights, since they achieved the best balance. Estimates based on the expanded logit weights, which also provided balance, were similar.

Electronic copy available at: https://ssrn.com/abstract=4336508

Because we are interested in the ATT of these interventions, once again we can estimate the propensity scores via binary choice models, comparing each of the intervention groups to the comparison group one at a time. Because we need to balance the sample with respect to the interventions at both incidents (i.e., at t and $t-1$), we specify the propensity score models differently here than we have so far. Here, we use characteristics of both incidents (for example, whether the victim was injured) in the choice model. We also hold characteristics of the participants (such as their prior criminal history) fixed at their values at the very first incident, in order to shut down the evolution of predictors over time. We estimate these propensity score models by CBPS.

Table 5: ATT estimates for dynamic treatment effects

	Treatment variable	
	Charges	High-risk
Treated at past incident only	-0.024	0.013
	(0.006)	(0.009)
	[73600]	[74827]
Treated at current incident only	-0.051	0.001
	(0.006)	(0.008)
	[74528]	[76654]
Treated at current and past incident	-0.070	0.029
	(0.010)	(0.018)
	[69543]	[73143]

Notes: ATT estimates are coefficients are from a weighted regression of the violent recidivism indicator on the given treatment variable, formed from interactions between the treatment indicators at the $t-1$st and tth incidents . Each cell is from a separate regression that includes the relevant treatment group and the comparison group of non-charged, non-high-risk incidents. Sample sizes [in brackets] are the sum of the number of incidents in those two groups. The sample is restricted to dyads with at least two incidents. Separate propensity score models estimated by CBPS using predictors described in text. Standard errors are clustered at the level of the dyad.

Estimated ATT's are reported in Table 5. The estimates for charges yield some evidence of dynamics, since being charged at the past incident has a negative and significant effect on recidivism following the current incident. This effect is smaller than the effect of charges at the current incident based on our previous static specification (Table 3). In contrast, the estimated ATT's involving charges at the current incident are similar to or larger than the estimates from Table 3. Neglecting the dynamics in the effect of charges seems to bias the estimated effect of

Electronic copy available at: https://ssrn.com/abstract=4336508

being charged toward zero. As for the high-risk intervention, all of the estimates are positive though insignificant.

The next question we address is whether our estimates reflect the effect of short-term incapacitation, which is an issue particularly for the effect of charges. Although our data includes no information on post-charge processing, we can ask whether our results stem from short-term detention around the time of the incident. Under English law, suspects who are arrested may be detained for at most 96 hours before being charged (gov.uk, 2022). To analyze whether such short-term detention explains our results, we redefine our outcome variable to equal 1 if there is an incident of violent recidivism between 4 and 365 days after the current incident. The ATT estimated from this outcome is reported in Table 6. It is identical to the estimate in the fourth column of Table 3 above. This suggests that short-term detention does not explain our results, since if it did, we would expect the estimate here to be smaller (in absolute value) than the estimate based on the full one-year follow-up. It is line with a report by Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services (2019) that the vast majority of perpetrators who are arrested are released on their own recognizance while the incident is being investigated.

Table 6: Estimates of the ATTs for different outcomes

	Treatment variable	
	Charges (1)	High-risk (2)
Violent Recidivism in 4-365 days	−0.051	−0.002
	(0.004)	(0.008)
Violent Recidivism in 3 months	−0.029	0.001
	(0.003)	(0.004)
Violent Recidivism in 6 months	−0.043	−0.001
	(0.004)	(0.006)
Violent Recidivism 6-12 months	−0.010	−0.002
	(0.002)	(0.004)

Notes: ATT estimates are coefficients are from a weighted regression of the violent recidivism indicator on the treatment variable. The dependent variable equals one if a repeat violent DA incident occurred between 4 and 365 days after the current incident. Propensity score models are the same as those in the fourth column of Table 3. Standard errors are clustered at the level of the dyad.

A natural question is, how much of the effect of charges is attributable to deterrence, and how much to incapacitation? Unfortunately, we cannot give a precise answer to this question. As mentioned above, we have no information on post-charge processing. Moreover, nationwide

Electronic copy available at: https://ssrn.com/abstract=4336508

data sources that do track penal outcomes, from which we might be able to glean some general insights, either report sentencing outcomes for only a few narrow classes of DA-related charges (Office for National Statistics, 2021), or do not separately report DA-related charges at all (Ministry of Justice, 2022).

Finally, we ask whether the effect of charges represents a true reduction in recidivism, or whether instead it represents a reduction in the willingness of the victim to report to the police. As noted above, more serious DA incidents are more likely to be reported (Barrett et al., 2017). Since we analyze DA incidents involving violence, there is reason to believe that the results reflect more than just changes in reporting.

To pursue this issue further, we make use of information about the informant, that is, the person who called the police to report the incident. The problem is that victims may experience retribution for calling the police, which could cause reported recidivism to fall, even as it causes actual recidivism to rise. Retribution is presumably less of an issue for incidents reported by third parties. Thus we estimate ATT's corresponding to different types of informants. If the effect of charges were more negative for incidents reported by the victim, it would raise concerns that our results reflected reporting changes rather than true recidivism changes.

Table 7: Estimates for the ATT on different informants

		Treatment variable	
	N	Charges	High-risk
	(1)	(2)	(3)
Victim	94,961	-0.048	0.007
		(0.005)	(0.008)
Other than victim	59,141	-0.059	-0.016
		(0.007)	(0.014)

Notes: ATT estimates are coefficients are from a weighted regression of the violent recidivism indicator on the treatment variable, using ATT weights from the indicated statistical model. For the first row, the sample consists of incidents reported by the victim. For the second row, the sample consists of incidents reported by a third party. Standard errors are clustered at the level of the dyad.

Table 7 displays the results, which are based on weights from the CBPS propensity scores. For charges, both estimates are negative and significant. The estimate for incidents reported by third parties is more negative than that for incidents reported by victims. This gives us some confidence that our estimates reflect real effects on recidivism, rather than reporting alone. The same pattern appears for the estimated effects of high risk, although they are insignificant.

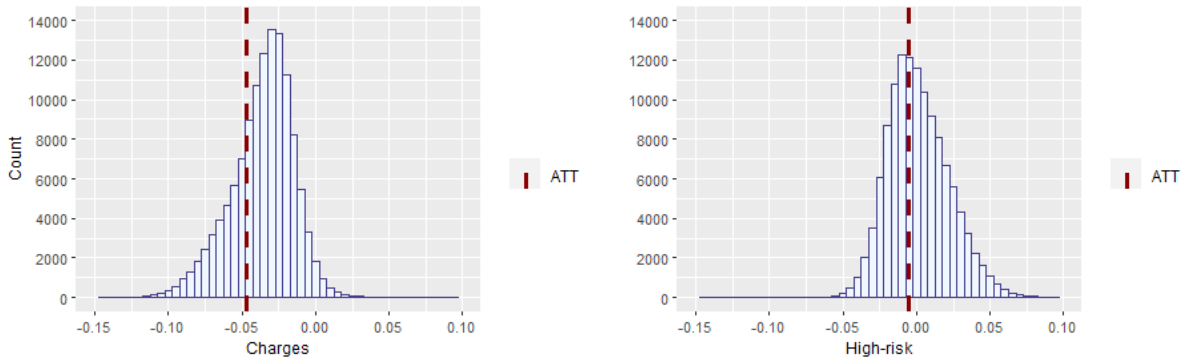
Electronic copy available at: https://ssrn.com/abstract=4336508

5.2. Heterogeneous treatment effects via causal forests

To analyze heterogeneity in the effects of the interventions, we first randomly split the sample into independent training and test samples. The training sample contained 80 percent of our data; the test sample included the remaining 20 percent. We took this approach because we wanted to reserve an independent test sample for specification checking and hypothesis testing below.[25]

A histogram for CATEs estimated from the training sample, which are based on causal forests of 10,000 trees, are shown in Figure 4.[26]

Figure 4: Histograms for estimated CATEs, by treatment variable



Notes: Estimates of $\hat{\tau}_i$, from causal forests composed of 10,000 causal trees. Estimated from 80-percent training sample.

Estimates of the ATTs, computed as the sample averages of the doubly-robust scores from the causal forests, are plotted as vertical lines in Figure 4. They are also reported in Panel A of Table 8. The ATTs estimated from the causal forests are similar to those obtained via IPW weighting in Table 3 above. Thus two different estimators yield similar estimates of this key parameter.

Although our goal is to analyze how the CATE's in Figure 4 are related to specific predictors, we first ask whether they are related to any of the predictors in general, or whether they reflect mostly noise. We present several tests, since a single test may not provide a comprehensive answer to the question. First, we construct and compare two sets of CATEs. We first applied the causal forest that was trained on the 80 percent training sample to predict CATEs on the 20 percent test sample. We refer to these as predicted CATEs. Second, we trained a separate causal forest to the test sample. We refer to these as estimated CATEs. We plot the

[25] We sampled by dyad to ensure that the training and test samples were independent.

[26] These estimates make use of cross-fitted propensity scores estimated via CBPS, which provide leave-out estimates that help reduce bias due to overfitting. The need for cross-fitting, and the algorithm used to carry it out, are discussed in the Appendix.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 8: Estimates from CATEs

	Treatment variable	
	Charge	High-risk
	(1)	(2)
Panel A: ATT		
ATT	−0.047	−0.005
	(0.004)	(0.009)
Panel B: Calibration checks		
$\tilde{D}$	0.949	1.978
	(0.086)	(3.319)
$\tilde{D} * (\hat{\tau} - \overline{\tau})$	0.983	0.649
	(0.183)	(0.213)

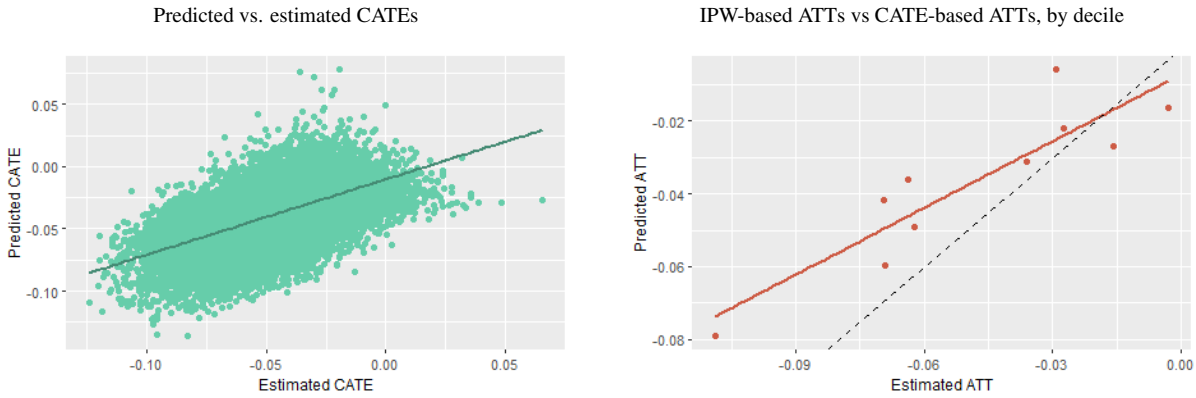
Notes: Panel A: ATT estimates are means of doubly-robust scores from causal forests depicted in Figure 4. Panel B: Coefficients from regression of residualized violent recidivism indicator on residualized treatment indicator ($\tilde{D}_i$) and interaction between residualized treatment indicator and centered CATE ($\hat{\tau}_i$). In both panels, standard errors are clustered at the level of the dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508

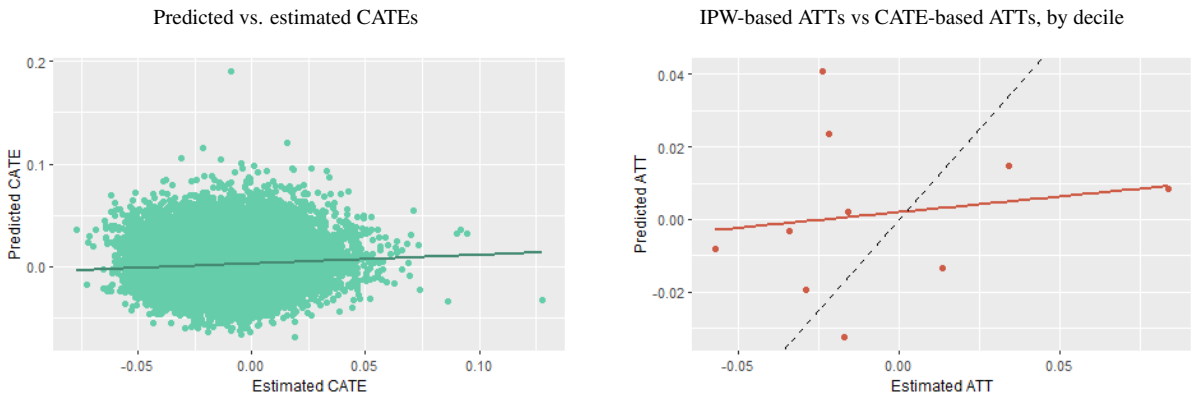
estimated CATEs against the predicted CATEs in the left panels of Figure 5, together with the best-fit regression lines.

Figure 5: Predicted vs estimated ATTs and CATEs, by treatment variable

A. Charges



B. High-risk



Notes: Left column: Scatterplots of predicted CATEs against estimated CATEs. Predicted CATEs are based on estimates from training sample, predicted on independent test sample. Estimated CATEs are from causal forests, composed on 10,000 trees, trained on test sample. Right column: Scatterplots of predicted against estimated ATTs, by decile of CATEs estimated from training sample and predicted onto test sample. Predicted ATTs are decile-specific means of predicted CATEs. Estimated ATTs are from decile-specific regressions of the violent recidivism indicator on the treatment dummy, using ATT weights estimated via CBPS. Also depicted are the best-fit line (solid red) and 45-degree line (dashed).

The scatter plot in the top panel of Figure 5 is clearly upward sloping. This indicates that the structure of the CATEs for charges is similar between the training and test samples, suggesting that the CATEs reflect at least some systematic heterogeneity. No such relationship is present in the scatterplot in the bottom panel of Figure 5, suggesting that the CATEs for the high-risk intervention may be mostly noise.

For the second exercise, we again make use of the predicted CATE's. We first sort the test sample into deciles based on the predicted CATE's, then average the predicted CATE's within

Electronic copy available at: https://ssrn.com/abstract=4336508

decile to estimate decile-specific ATT's. We then use the outcome data from the test sample to estimate decile-specific ATT's directly via IPW weighting.

The right panels of Figures 5 show scatterplots of the IPW-based ATTs against the CATE-based ATTs, along with best-fit lines and the 45-degree line. If the CATEs were all signal and no noise, the scatter plot would lie along the 45-degree line. For charges, the slope of the scatter plot is less than one, but still clearly positive. Groups of observations with higher predicted CATEs in fact tend to have higher ATTs. For the high-risk intervention, in contrast, the best-fit line is essentially flat.

Finally, we estimate a calibration regression that directly relates outcomes in the test sample to the predicted CATEs. According to Chernozhukov et al. (2020), the best linear projection of the true idiosyncratic treatment effects on the estimated CATE's can be obtained from the weighted regression of the residualized outcome variable $\tilde{Y}_i$ on the residualized treatment variable $\tilde{D}_i$ and an interaction between $\tilde{D}_i$ and the centered estimated CATE $\hat{\tau}_i - \bar{\tau}$, using weights $\frac{1}{\hat{p}^{(-i)}(1-\hat{p}^{(-i)})}$.

Estimated coefficients are standard errors are presented in Panel B of Table 8. Since the coefficient on $\tilde{D}_i$ is approximately one, we can conclude that the CATEs are well calibrated on average (Athey and Wager, 2019). This is consistent with the similarity of the estimated ATTs in Tables 3 and Panel A of Table 8. However, the coefficient on the interaction term is key for our purposes here. If our estimates of τ_i reflect true heterogeneity, then its coefficient should be approximately one. Conversely, if our estimates of τ_i reflect pure noise, then it should equal zero. For charges, we cannot reject the hypothesis that it equals one. For high-risk, we can reject the hypothesis that it equals one, but we can also reject the null that it equals zero.

To summarize, there appears to be treatment effect heterogeneity for charges which is correlated with our predictors. All three of our tests point in the same direction. For the high-risk intervention, the results are more mixed. The first two tests suggest that there is little heterogeneity in treatment effects that is correlated with the predictors. The calibration test is more optimistic.

We now ask how the heterogeneity in the estimated treatment effects relates to specific predictors in our data. There are many ways one can do this, and the literature provides examples of several different approaches. One involves regressing the estimated CATEs on subsets of predictors, either parametrically or non-parametrically; another involves characterizing extreme values (Athey and Wager, 2019; Chernozhukov et al., 2018; Davis and Heller, 2020; Knittel and Stolper, 2019; Knaus, 2022). Presumably, one should seek to characterize heterogeneity in a way that is useful for the question at hand.

In our case, we would like to ask how one could re-allocate the interventions across incidents so as to achieve greater reductions in violent recidivism. Specifically, we would like to learn a low-dimensional decision tree that could be used to guide investigative resources in the

Electronic copy available at: https://ssrn.com/abstract=4336508

field. We also need to be attentive to constraints imposed by the setting. One such constraint involves the budget for investigation and implementation of the interventions. The model for charges must satisfy an additional constraint. Since charges can only be filed against incidents that are classified as crimes, our decision tree for charges is grown only from such incidents.

In principle, one could learn a decision tree directly from the doubly-robust scores from the causal forest. Athey and Wager (2021) propose an exhaustive search algorithm for this purpose, and we initially attempted that method. However, we never obtained a decision tree with more than a single split. As a result, we modified our approach.

For each intervention, we constructed a binary indicator which equaled one for all observations whose doubly-robust score was less than the quantile of the score distribution which corresponded to the share of treated observations in our sample. Thus for charges, which were filed in 29.7 percent all cases classified as crimes, our binary indicator equals one if the efficient score for an observation was below the 29.7th percentile. For the other intervention, 8.97 percent of all incidents were classified as high-risk, so our binary indicator equals one for observations with doubly-robust scores less than the 8.97th percentile. These indicator variables equal one for the incidents with the largest favorable (i.e., negative) treatment effects for the corresponding interventions. If we could somehow learn decision trees that targeted precisely those incidents, then we would have a set of decision rules that minimized violent recidivism while charging the same share of crimes, and classifying the same share of incidents as high-risk, as those observed in our sample.

We learned decision trees from these favorable-treatment-effect indicators via exhaustive search. As potential decision variables, we started with the 10 most important variables for constructing the causal forests, where importance is measured as a depth-weighted index of the number of times each predictor was used for splitting in building the causal forest. Our rationale was that variables important for splits in the causal trees should be important sources of treatment effect heterogeneity generally.[27] To simplify the computational complexity of the problem, we restricted attention to decision trees of depth two. We also replaced the cardinal-valued number of yeses on the DASH form (one of the top-10 splitting variables for charges) with four dummy variables, constructed by dichotomizing the number of yeses at the values 4, 6, 7, and 14.[28] We constructed all possible depth-2 trees based on these decision variables, and chose the ones that minimized the impurity of the terminal leaves. The resulting decision tree for charges is reported in Table 9, and the tree for the high-risk intervention is reported in Table 10. Both of these decision trees are estimated from the training sample.

[27]From the top 10 variables, we dropped those pertaining to the attending officer, such as his/her years of experience, on the grounds that we desired a decision tree which would apply for any officer attending a call.

[28]The median number of yeses is roughly 4; 14 yeses are supposed to result in an automatic high-risk classification (Whinney, 2015). We chose the values 6 and 7 on the basis of preliminary trees estimated via recursive partitioning.

Electronic copy available at: https://ssrn.com/abstract=4336508

For charges, the first split divided the sample of criminal incidents according to whether the incident had more than seven yeses on the DASH form. This divided the sample of criminal incidents nearly in half. Among perpetrators with lower DASH scores, the tree next split according to whether the perpetrator had been accused of more than one crime in the past two years. For the cases with higher DASH scores, it split according to whether the perpetrator had more than one DA crime during the past two years.

Cases with the least serious combination of features, meaning an incident with a DASH score below seven and a perpetrator with one or fewer past crimes, accounted for 41.4 percent of the sample. Among that group, only 13.3 percent had doubly-robust scores in the bottom 29.7 percent. The next group, with a DASH score below seven more than one crime in the past two years, accounted for 8.8 percent of the sample. Among that group, 40.4 percent had doubly-robust scores in the bottom 29.7 percent. Among incidents with DASH scores of 7 or greater and a perpetrator accused of zero or one prior DA crimes, which made up 23.6 percent of the sample, 28.2 percent had doubly-robust scores in the favorable category. Finally, among the 26.2 percent of incidents with DASH scores of 7 or greater and a perpetrator accused of more than one prior DA crime, 51.1 percent had doubly-robust scores in that group.

Our next step was to estimate ATTs for the four terminal leaves. We did this by constructing dummy variables reflecting the definitions of the leaves, then regressing the doubly-robust scores on these dummies. The estimated ATTs are reported in the next-to-last column of the Table, with standard errors in the last column. The estimates align with the share of incidents below the 29.7th percentile, even though they were not guaranteed to do so, since the tree used only qualitative information to place the splits. The estimated ATTs vary by a factor of 10, ranging from -0.012 (0.003) for the group with the least serious combination of features, to -0.130 (0.019) for the with DASH score of 7 or higher and whose perpetrators had more than one crime in the last two years. If the estimates were asymptotically normal, they would all be significant at the 5 percent level.

However, it seems unlikely that the estimates in Table 9 would have standard limiting distributions. Although Wager and Athey (2019) provide conditions under which the estimated CATEs are asymptotically normal at a given value of x, the values of x over which we have estimated leaf-specific ATTs were selected by an automated specification search that was designed to maximize differences in a function of the estimated CATEs. Such estimates would presumably be afflicted by pre-test estimation bias.

To deal with this issue, we estimated and tested the tree reported in Table 9 on the independent test sample. To do so, we first estimated a separate causal forest on the test sample, in the same manner as we did for the training sample. We then coded the dichotomous indicator for whether each observation fell below the 29.7th percentile of the efficient scores. We next coded in the test sample the dummy variables that define the decision tree that was learned

Electronic copy available at: https://ssrn.com/abstract=4336508

from the training sample. Finally, we fit two regressions to those dummy variables. The dependent variable for the first was the favorable-treatment-effect indicator. This regression tells us the share of each leaf that falls in the lowest part of the doubly-robust score distribution. The dependent variable for the second is the doubly-robust score itself. This regression gives us the leaf-specific ATTs. The key point here is that the rules that define the leaves were chosen from the independent training sample. Thus established central limit theorems should apply to these test-sample regressions.

The test-sample estimates are presented in panel B of Table 9. The leaf-specific shares in the favorable-treatment-effect group here are quite close to their counterparts from the training sample. The rank-ordering of the estimated ATTs is the same as well, and indeed, the estimated ATTs are fairly close. The estimate for the group with the least serious combination of features is -0.018 (0.006), compared with -0.012 (0.003) in the training sample. The estimate for incidents with DASH scores of seven or more involving perpetrators with more than one crime in the recent past is -0.185 (0.044), compared to -0.130 (0.019). All of the estimates are significant at the 5 percent level.

To quantify the extent to which the training-sample decision tree replicates to the test sample, we assigned two sets of estimated ATTs to the test-sample observations. The first were those estimated from the training sample, from Panel A of Table 9. The second were those estimated from the test-sample, from Panel B. The correlation between the two measures was 0.927. We conclude that our approach has succeeded in finding groups, identifiable based on the values of their predictors, whose ATTs are different from one another, and which replicate across independent samples.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 9: Decision tree for charges

A. Training data					
Decision variable	Sample size	Perc. of sample	Perc. favorable TE	ATT	SE
DASH, number yes					
Less than 7	24627	50.2	18.0		
Perp. accused of over 1 crime					
No	20324	41.4	13.3	-0.012	0.003
Yes	4303	8.8	40.4	-0.088	0.015
7 or more	24465	49.8	40.3		
Perp. accused of over 1 DA crime					
No	11597	23.6	28.2	-0.029	0.010
Yes	12868	26.2	51.1	-0.130	0.019
B. Test data					
Decision variable	Sample size	Perc. of sample	Perc. favorable TE	ATT	SE
DASH, number yes					
Less than 7	6020	49.4	17.9		
Perp. accused of over 1 crime					
No	4912	40.3	13.2	-0.018	0.006
Yes	1108	9.1	38.7	-0.044	0.028
7 or more	6172	50.6	40.1		
Perp. accused of over 1 DA crime					
No	2819	23.1	26.6	-0.042	0.018
Yes	3353	27.5	51.4	-0.185	0.044

Electronic copy available at: https://ssrn.com/abstract=4336508

To produce the decision tree for the high-risk intervention, we proceeded similarly. Results are reported in Table 10. The splits are different than those for charges. One leaf is very large, containing 82 percent of the sample, only 3.3 percent of which falls in the favorable-treatment-effect group. At the other end of the spectrum, the tree identifies a small group, with DASH scores greater than 15, of whom 52.3 percent fall into the favorable-treatment-effect category. This leaf has a sizeable negative ATT, although it is not significant.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table 10: Decision tree for high-risk

A. Training data					
Decision variable	Sample size	Perc. of sample	Perc. favorable TE	ATT	SE
DASH, number yes					
Less than 12	109542	88.9	5.3		
Any HR in past 12 mo					
Yes	100990	82.0	3.3	0.003	0.002
No	8552	6.9	28.7	0.080	0.040
12 or more	13679	11.1	37.8		
DASH, number yes					
Between 12-15	8351	6.8	28.5	-0.054	0.039
15 or more	5328	4.3	52.3	-0.220	0.179
B. Test data					
Decision variable	Sample size	Perc. of sample	Perc. favorable TE	ATT	SE
DASH, number yes					
Less than 12	9716	79.7	6.1		
Any HR in past 12 mo					
Yes	8758	71.8	3.3	-0.005	0.005
No	958	7.9	31.1	-0.059	0.058
12 or more	2476	20.3	20.2		
DASH, number yes					
Between 12-15	1450	11.9	17.2	0.002	0.051
15 or more	1026	8.4	24.4	-0.044	0.169

Electronic copy available at: https://ssrn.com/abstract=4336508

However, in this case, applying the same decision rules to the test sample yields a rather different decision tree. The sizes of the leaves, their shares of observations with favorable treatment effects, and the leaf-specific treatment effects are quite different. Furthermore, none of the estimated ATTs in the test-sample tree are significant at conventional levels. When we assigned the training- and test-sample ATTs to the test-sample observations, the correlation between the two was 0.075.

Whereas the test-sample tree for charges largely replicated its training-sample counterpart, that is not the case for the high-risk intervention. This seems consistent with the tests above. Those tests strongly pointed toward heterogeneity in the effects of charges, but were more mixed regarding heterogeneity in the effect of the high-risk intervention. We conclude that either there is no heterogeneity in the effect of the high-risk intervention, or that any heterogeneity is not correlated with the predictors in our sample.

Since the decision trees are based on causal forests, which are built from random samples of the data and the predictors, it is important to assess whether the decision trees learned from the causal forests are robust. To assess robustness, we built five different causal forests for each of our two treatment variables, then trained decision trees to each. The structure of these auxiliary decision trees was similar to that shown in Tables 9 and 10. All but one used the same splitting variables, in the same order. We also quantified whether the ATTs estimated from the training sample replicated to the test sample. For the decision trees for charges, the correlations ran from 0.912 to 0.929. For those for the high-risk intervention, they ran from 0.136 to 0.248. All of the training-sample decision trees for charges were similar, and they replicated similarly well to the test data. The training-sample decision trees for high-risk were also similar, but all of them failed to replicate to the test samples.[29]

Finally, we illustrate how the heterogeneous effects of filing charges might be used to guide policy. Under the current regime, 29.7 percent of crimes are charged, and violent recidivism is reduced by roughly -0.050. Consider an alternative regime, whereby two groups of perpetrators were charged: one with a DASH score less than seven and more than one crime on their record over the past two years, and the other with a DASH score of seven or more and more than one DA crime over the last two years. That policy would result in more charges, since those groups together make up 35 percent of the population. At the same time, it would reduce violent recidivism by -0.119 (0.017) percentage points.[30] Put differently, in return for charging 18 percent more crimes, the reduction in violent recidivism would more than double.[31]

[29] As a further check on robustness, we estimated causal forests in which we replaced the cross-fit CBPS propensity scores with random-forest propensity scores, which are the default in the R *grf* package. Across the six causal forests for charges, the correlations between the CATEs estimated via CBPS and those estimated via random forests were all at least 0.99. For the six causal forests estimated for high risk, they ran from 0.78 to 0.99.

[30] This is a weighted average of the estimates from these two groups.

[31] This calculation is based on the training sample. The corresponding estimate from the test sample is -0.150

Electronic copy available at: https://ssrn.com/abstract=4336508

To be sure, this represents an upper bound on the tradeoff between charges and violent recidivism. The reason is that perpetrators of domestic abuse crimes cannot be charged solely on the basis of their recent criminal history or their current DASH score. Under English law, offenders may only be charged if there is "sufficient evidence to provide a realistic prospect of conviction against each suspect on each charge" (Her Majesty's Inspectorate of Constabulary, 2014, p. 98). Nonetheless, observable heterogeneity in the effects of criminal charges could be used to prioritize resources for investigations, with the idea of building stronger cases against perpetrators for whom the effects of being charged would be greater. The substantial difference between the smallest and largest ATTs in Table 9 suggests that there is scope to further reduce violent recidivism by means of such a policy.

6. Conclusion

Domestic abuse is a ubiquitous problem. We study two interventions to reduce violent recidivism in DA cases. The first is charging the perpetrator with an offense. The second is providing protective services to victims assessed to be at high risk of serious recidivism.

Our method statistically equates perpetrators who were charged with those who were not. We equated these two groups on the basis of several dozen characteristics of the incident, the participants, their domestic-abuse and criminal histories, the police officer to responded to the call, and their risk assessment scores. Although many of these characteristics are highly predictive of treatment, we can not equate on the basis of any unobservable characteristics. This is an inherent limitation of our approach.

Regarding criminal charges, two different techniques yielded similar estimates of the average effect of treatment on the treated. Both IPW weighting and the causal forest indicated that charges reduce the likelihood of violent recidivism by about 5 percentage points. Relative to the violent recidivism rate in the sample, that amounts to a reduction of almost 40 percent. That magnitude is much larger than the estimated effect of arrests from the SARP studies, but is comparable to estimates from Berk and Sherman (1984) and Amaral et al (2022).

In contrast, we found no evidence that the DASH/MARAC process reduced violent recidivism. Both IPW weighting and the causal forest yielded ATTs that were small and insignificant. Since we have no data on the specific protective services that were provided by this process, we cannot rule out the possibility that some of those services could offer effective protection. However, on average, the process does not reduce violent recidivism, despite the resources devoted to it.

We also found evidence of heterogeneity in the effects of criminal charges. One group with a fairly serious criminal history had an ATT that was nearly 10 times larger than another group

(0.041).

Electronic copy available at: https://ssrn.com/abstract=4336508

with a much less serious record. This suggests that it may be possible to target investigative resources in such a way as to protect a greater number of victims from repeat domestic violence.

References

Aizer, A., 2010. The gender wage gap and domestic violence. American Economic Review 100, 1847–1859. doi:10.1257/aer.100.4.1847.

Aizer, A., 2011. Poverty, violence, and health: The impact of domestic violence during pregnancy on newborn health. Journal of Human Resources 46, 518–538. doi:10.3368/jhr.46.3.518.

Aizer, A., Dal Bó, P., 2009. Love, hate and murder: Commitment devices in violent relationships. Journal of Public Economics 93, 412–428. URL: http://dx.doi.org/10.1016/j.jpubeco.2008.09.011, doi:10.1016/j.jpubeco.2008.09.011.

Amaral, S., Dahl, G.B., Endl-Gayer, V., Hener, T., Rainer, H., 2022. Deterrence or backlash? Arrests and the dynamics of domestic violence. Technical Report. URL: https://sofiamaral.weebly.com/research.html.

Anderberg, D., Rainer, H., Wadsworth, J., Wilson, T., 2016. Unemployment and Domestic Violence: Theory and Evidence. Economic Journal 126, 1947–1979. doi:10.1111/ecoj.12246.

Angrist, J.D., 2006. Instrumental variables methods in experimental criminological research: What, why and how. Journal of Experimental Criminology 2, 23–44. doi:10.1007/s11292-005-5126-x.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Annals of Statistics 47, 1148–1178.

Athey, S., Wager, S., 2019. Estimating Treatment Effects with Causal Forests: An Application. Observational studies 5, 1–10.

Athey, S., Wager, S., 2021. Policy learning with observational data. Econometrica 89, 133–161.

Barrett, B.J., Peirone, A., Cheung, C.H., Habibov, N., 2017. Pathways to Police Contact for Spousal Violence Survivors: The Role of Individual and Neighborhood Factors in Survivors' Reporting Behaviors. Journal of Interpersonal Violence 36, 1–31. doi:10.1177/0886260517729400.

Berk, R.A., Campbell, A., Klap, R., Western, B., 1992. A Bayesian Analysis of the Colorado Springs Spouse Abuse Experiment. The Journal of Criminal Law and Criminology 83, 170. doi:10.2307/1143828.

Berk, R.A., He, Y., Sorenson, S.B., 2005. Developing a practical forecasting screener for domestic violence incidents. Evaluation Review 29, 358–383. doi:10.1177/0193841X05275333.

Bhalotra, S., Britto, D.G., Pinotti, P., Sampaio, B., 2021. Job Displacement, Unemployment Benefits and Domestic Violence.

Bhuller, M., Dahl, G.B., Løken, K.V., Mogstad, M., 2021. Consequences of Domestic Violence for Victims and their Families.

Bindler, A., Ketel, N., 2022. Scaring or scarring? Labor market effects of criminal victimization. Journal of Labor Economics 4, 939–205.

Campbell, J.C., Webster, D.W., Glass, N., 2009. The Danger Assessment. Journal of Interpersonal Violence 24, 653–674. doi:10.1177/0886260508317180.

Card, D., Dahl, G.B., 2011. Family violence and football: The effect of unexpected emotional cues on violent behavior. Quarterly Journal of Economics 126, 103–143. doi:10.1093/qje/qjr001.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. Econometrics journal 21, C1–C68. doi:doi:10.1111/ectj.12097.

Chernozhukov, V., Demirer, M., Duflo, E., Fernandez-Val, I., 2020. Generic machine learning inference on

Electronic copy available at: https://ssrn.com/abstract=4336508

heterogeneous treatment effects in randomized experiments, wiht an application to immunization in India. Technical Report. URL: https://arxiv.org/abs/1712.04802.

Chin, Y.M., Cunningham, S., 2019. Revisiting the effect of warrantless domestic violence arrest laws on intimate partner homicides. Journal of Public Economics 1, 1–10.

College of Policing, 2022. Call handler and frong counter staff response to a domestic abuse incident: authorized professional practice. Technical Report. URL: https://www.app.college.police.uk/app-content/major-investigation-and-public-protection/domestic-abuse/call-handler-and-front-counter-staff-response/#checklist-information-gathering.

Coordinated Action Against Domestic Abuse, 2012. A place of greater safety. Technical Report. Coordinated Action Against Domestic Abuse. URL: www.caada.org.uk/commissioning.

Crown Prosecution Service, 2020. Charging (The Director's Guidance) sixth edition, December 2020. Technical Report. URL: https://www.cps.gov.uk/legal-guidance/charging-directors-guidance-sixth-edition-december-2020.

Currie, J., Mueller-Smith, M., Rossin-Slater, M., 2022. Violence while in utero: The impact of assaults during pregnancy on birth outcomes. Review of economics and statistics 104, 525–540.

Davis, J.M., Heller, S.B., 2020. Rethinking the benefits of youth employment programs: The heterogeneous effects of summer jobs. Review of economics and statistics 4, 664–677.

Dunford, F.W., HUIZINGA, D., ELLIOTT, D.S., 1990. The Role of Arrest in Domestic Assault: the Omaha Police Experiment. Criminology 28, 183–206. doi:10.1111/j.1745-9125.1990.tb01323.x.

Dutton, D.G., Kropp, P.R., 2000. A review of domestic violence risk assessments. Trauma, Violence, and Abuse 1, 171–181.

Ericson, R., Haggerty, K.D., 1997. Policing the Risk Society. Clarendon Press, Oxford.

European Institute for Gender Equality, 2019. Risk assessment and management of intimate partner violence in the EU. Technical Report. European Institute for Gender Equality. Vilnius, Latvia.

Fagan, J., 1995. The Criminalization of Domestic Violence: Promises and Limits. Technical Report. National Institute of Justice. Washington, DC.

Goller, D., Lechner, M., Moczall, A., Wolff, J., 2019. Does the estimation of the propensity score by machine learning improve matching estimation? The case of Germany's programs for long term unemployed. Technical Report. URL: https://www.econstor.eu/bitstream/10419/207352/1/dp12526.pdf.

gov.uk, 2012. PACE Code G 2012. Technical Report. URL: https://www.gov.uk/government/publications/pace-code-g-2012.

gov.uk, 2022. Being arrested: your rights. Technical Report. URL: https://www.gov.uk/arrested-your-rights/how-long-you-can-be-held-in-custody.

Graham, L.M., Sahay, K.M., Rizo, C.F., Messing, J.T., Macy, R.J., 2021. The Validity and Reliability of Available Intimate Partner Homicide and Reassault Risk Assessment Tools: A Systematic Review. Trauma, Violence, and Abuse 22, 18–40. doi:10.1177/1524838018821952.

grf-labs, 2022. Software package. Technical Report. URL: https://github.com/grf-labs/grf/blob/e001f08daac6378aef664734e544fa2234a567da/r-package/grf/R/forest_summary.R.

Grierson, J., 2020. Fifth of crimes involved domestic abuse in first England and Wales lockdown. URL: https://www.theguardian.com/society/2020/nov/25/fifth-of-crimes-involved-domestic-abuse-in-first-england-and-wales-lockdown.

Grogger, J., Gupta, S., Ivandic, R., Kirchmaier, T., 2021. Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases. Journal of Empirical Legal Studies 18, 90–130.

Guarnieri, E., Rainer, H., 2018. Female Empowerment and Male Backlash. URL: http://tertilt.vwl.uni-mannheim.de/conferences/FamilyConference/EleonoraGuarnieri.pdf.

Electronic copy available at: https://ssrn.com/abstract=4336508

Gutierrez, I.A., Molina, O., 2020. Does domestic violence jeopardize the learning environment of peers within the school? Peer effects of exposure to domestic violence in urban Peru. doi:10.1016/j.econedurev.2021.102147.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Her Majesty's Inspectorate of Constabulary, 2014. Everyone's business: Improving the police response to domestic abuse. Technical Report. Her Majesty's Inspectorate of Constabulary. London.

Her Majesty's Inspectorate of Constabulary, 2015. Increasingly everyone's business: a progress report on the police response to domestic abuse. Technical Report. Home Office.

Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services, 2019. The police response to domestic abuse: An update report. Technical Report. Her Majesty's Inspectorate of Constabulary and Fire and Rescue Services. URL: www.justiceinspectorates.gov.uk/hmicfrs.

Hidrobo, M., Peterman, A., Heise, L., 2016. The effect of cash, vouchers, and food transfers on intimate partner violence: Evidence from a randomized experiment in Northern Ecuador. American Economic Journal: Applied Economics 8, 284–303. doi:10.1257/app.20150048.

Hilton, N.Z., Harris, G.T., 2005. Predicting Wife Assault: A Critical Review and Implications for Policy and Practice. Trauma, Violence, & Abuse 6, 3–23. doi:10.1177/1524838004272463.

Hirschel, J.D., Hutchison, I.W., 1992. Female spouse abuse and the police response: the Charlotte, North Carolina experiment. Journal of Criminal Law and Criminology 83, 73–120.

Home Office, 2000. Domestic violence: revised circular to the police.

Imai, K., Ratkovic, M., 2014. Covariate balancing propensity score. Journal of the Royal Statistical Society. Series B: Statistical Methodology 76, 243–263. doi:10.1111/rssb.12027.

Iyengar, R., 2009. Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws. Journal of Public Economics 1-2, 85–98.

Jung, S., Buro, K., 2017. Appraising Risk for Intimate Partner Violence in a Police Context. Criminal Justice and Behavior 44, 240–260. doi:10.1177/0093854816667974.

Knaus, M.C., 2022. Double machine learning based program evaluation under unconfoundedness. Econometrics journal 0, 1–26. doi:10.1093/ectj/utac015.

Knittel, C., Stolper, S., 2019. Using machine learning to target treatment: the case of household energy use. Technical Report. URL: https://www.nber.org/system/files/working_papers/w26531/w26531.pdf/.

Koppensteiner, M., Matheson, J., Plugor, R., 2020. Public service access and domestic violence: Lessons from a randomized controlled trial. Technical Report. URL: http://www.koppensteiner.info/.

Kropp, P.R., 2004. Some questions regarding spousal assault risk assessment. Violence Against Women 10, 676–697. doi:10.1177/1077801204265019.

Maxwell, C.D., GARNER, J.H., FAGAN, J.A., 2002. The Preventive Effects of Arrest on Intimate Partner Violence: Research, Policy and Theory. Criminology and Public Policy 2, 51–80. doi:10.1111/j.1745-9133.2002.tb00107.x.

Messing, J.T., Campbell, J., Wilson, J.S., Brown, S., Patchell, B., University, A.S., Work, S.o.S., 2014. Police Departments' Use of the Lethality Assessment Program: A Quasi-Experimental Evaluation. Technical Report. Report to the National Institute of Justice. URL: https://www.ncjrs.gov/pdffiles1/nij/grants/247456.pdf.

Messing, J.T., Thaller, J., 2013. The Average Predictive Validity of Intimate Partner Violence Risk Assessment Instruments. Journal of Interpersonal Violence 28, 1537–1558. URL: https://doi.org/10.1177/0886260512468250, doi:10.1177/0886260512468250.

Ministry of Justice, 2022. Criminal Justice System and Offenders Criminal History. Technical Report. URL:

Electronic copy available at: https://ssrn.com/abstract=4336508

https://moj-analytical-services.github.io/criminal_history_sankey/index.html.

Myhill, A., 2019. Renegotiating domestic violence: police attitudes and decisions concerning arrest. Policing and Society 29, 52–68. URL: https://doi.org/10.1080/10439463.2017.1356299, doi:10.1080/10439463.2017.1356299.

National Coalition Against Domestic Violence, . Domestic Violence. URL: https://assets.speakcdn.com/assets/2497/domestic{_}violence2.pdf.

Office for National Statistics, 2021. Domestic abuse and the criminal justice system, England and Wales: November 2020. Technical Report. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseandthecriminaljusticesystemenglandandwales/november2021.

Pate, A.M., Hamilton, E.E., 1992. Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment. American Sociological Review 57, 691. doi:10.2307/2095922.

Police Executive Research Forum, 2015. Police Improve Response to Domestic Violence,But Abuse Often Remains the Hidden Crime'. Subect to Debate 29.

van der Put, C.E., Gubbels, J., Assink, M., 2019. Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. Aggression and Violent Behavior 47, 100–116. doi:10.1016/j.avb.2019.03.008.

Richards, L., 2009. Domestic Abuse, Stalking and Harassment and Honour Based Violence Risk Identification Checklist. URL: http://www.dashriskchecklist.co.uk/uploads/V-DASH2010-2015.pdf.

Richards, O., Harinam, V., 2020. Tracking Police Arrests of Intimate Partner Domestic Abuse Suspects in London: a Situational Factors Analysis. Cambridge Journal of Evidence-Based Policing 4, 103–113. doi:10.1007/s41887-020-00047-y.

Robinson, A.L., 2006. Reducing Repeat Victimization Among High-Risk Victims of Domestic Violence. Violence Against Women 12, 761–788.

Robinson, A.L., Myhill, A., Wire, J., Roberts, J., Tilley, N., 2016. Risk-led policing of domestic abuse and the DASH risk model. Technical Report. College of Policing. URL: http://www.college.police.uk/News/College-news/Documents/Risk-led{_}policing{_}of{_}domestic{_}abuse{_}and{_}the{_}DASH{_}risk{_}model.pdf.

Robinson, A.L., Tregidga, J., 2007. The perceptions of high-risk victims of domestic violence to a coordinated community response in Cardiff, Wales. Violence Against Women 13, 1130–1148. doi:10.1177/1077801207307797.

Roehl, J.P., O'Sullivan, C.P., Webster, D.S., Campbell, J.P., 2005. Intimate Partner Violence Risk Assessment Validation Study: The RAVE Study Practitioner Summary and Recommendations: Validation of Tools for Assessing Risk from Violent Intimate Partners. National Institute of Justice , 20.

Rosenbaum, P.R., Rubin, D.B., 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects A. Biometrika 70, 41–55.

SafeLives, 2015. Insights IDVA national dataset 2013-14. Technical Report. SafeLives.

Sherman, L.W., Berk, R.A., 1984. The Specific Deterrent Effects of Arrest for Domestic Assault Author (s): Lawrence W . Sherman and Richard A . Berk Published by : American Sociological Association Stable URL : http://www.jstor.org/stable/2095575 ARREST FOR DOMESTIC ASSAULT *. American Sociological Review 49, 261–272. URL: https://www.jstor.org/stable/2095575.

Sherman, L.W., Schmidt, J.D., Rogan, D.P., Smith, D.A., Gartin, P.R., Cohn, E.G., Collins, J., Bacich, A.R., 1992. The Variable Effects of Arrest on Criminal Careers: The Milwaukee Domestic Violence Experiment. The Journal of Criminal Law and Criminology 83, 137. doi:10.2307/1143827.

Svalin, K., Levander, S., 2020. The Predictive Validity of Intimate Partner Violence Risk Assessments Conducted

Electronic copy available at: https://ssrn.com/abstract=4336508

by Practitioners in Different Settings: a Review of the Literature. Journal of Police and Criminal Psychology 35, 115–130. doi:10.1007/s11896-019-09343-4.

Sviatschi, M.M., Trako, I., 2021. Gender Violence, Enforcement, and Human Capital: Evidence from Women's Justice Centers in Peru.

Tur-Prats, A., 2019. Family types and intimate partner violence: A historical perspective. Review of Economics and Statistics 101, 878–891. doi:10.1162/rest_a_00784.

Turner, E., Medina, J., Brown, G., 2019. Dashing Hopes? the Predictive Accuracy of Domestic Abuse Risk Assessment by Police. The British Journal of Criminology 59, 1013–1034. doi:10.1093/bjc/azy074.

Wager, S., Athey, S., 2018. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. Journal of the American Statistical Association 113, 1228–1242. doi:10.1080/01621459.2017.1319839, arXiv:1510.04342.

Whinney, A., 2015. A descriptive analysis of Multi-Agency Risk Assessment Conferences (MARACs) for reducing the future harm of domestic abuse in Suffolk. Ph.D. thesis.

Electronic copy available at: https://ssrn.com/abstract=4336508

Appendix

1. DASH questionnaire

CURRENT SITUATION THE CONTEXT AND DETAIL OF WHAT IS HAPPENING IS VERY IMPORTANT. THE QUESTIONS HIGHLIGHTED IN BOLD ARE HIGH RISK FACTORS. TICK THE RELEVANT BOX AND ADD COMMENT WHERE NECESSARY TO EXPAND.	YES ☑	NO ☑
1. Has the current incident resulted in injury? (please state what and whether this is the first injury)	☐	☐
2. Are you very frightened? Comment:	☐	☐
3. What are you afraid of? Is it further injury or violence? (Please give an indication of what you think (name of abuser(s)….. might do and to whom) Kill: Self ☐ Children ☐ Other (please specify) ☐ Further injury and violence: Self ☐ Children ☐ Other (please specify) ☐ Other (please clarify): Self ☐ Children ☐ Other (please specify) ☐	☐	☐
4. Do you feel isolated from family/ friends i.e. does (name of abuser(s)…..) try to stop you from seeing friends/family/Dr or others?	☐	☐
5. Are you feeling depressed or having suicidal thoughts?	☐	☐
6. Have you separated or tried to separate from (name of abuser(s)….) within the past year?	☐	☐
7. Is there conflict over child contact? (please state what)	☐	☐
8. Does (…..) constantly text, call, contact, follow, stalk or harass you? (Please expand to identify what and whether you believe that this is done deliberately to intimidate you? Consider the context and behaviour of what is being done. Ask 11 additional stalking questions*)	☐	☐
CHILDREN/DEPENDENTS (If no children/dependants, please go to the next section)	YES	NO
9. Are you currently pregnant or have you recently had a baby in the past 18 months?	☐	☐
10. Are there any children, step-children that aren't (…..) in the household? Or are there other dependants in the household (i.e. older relative)?	☐	☐
11. Has (…..) ever hurt the children/dependants?	☐	☐
12. Has (…..) ever threatened to hurt or kill the children/dependants?	☐	☐
DOMESTIC VIOLENCE HISTORY	YES	NO
13. Is the abuse happening more often?	☐	☐
14. Is the abuse getting worse?	☐	☐
15. Does (…….) try to control everything you do and/or are they excessively jealous? (In terms of relationships, who you see, being 'policed at home', telling you what to wear for example. Consider honour based violence and stalking and specify the behaviour)	☐	☐
16. Has (…..) ever used weapons or objects to hurt you?	☐	☐
17. Has (…..) ever threatened to kill you or someone else and you believed them?	☐	☐

©Laura Richards (2009). Please do not reproduce without permission. For enquiries about training staff in the use of the DASH, S-DASH or SN-DASH (2009) Risk Identification Checklists, please contact laura@laurarichards.co.uk

Electronic copy available at: https://ssrn.com/abstract=4336508

18. Has (…..) ever attempted to strangle/choke/suffocate/drown you?	☐	☐
19. Does (….) do or say things of a sexual nature that makes you feel bad or that physically hurt you or someone else? (Please specify who and what)	☐	☐
20. Is there any other person that has threatened you or that you are afraid of? (If yes, consider extended family if honour based violence. Please specify who. Ask 10 additional HBV questions*)	☐	☐
21. Do you know if (…..) has hurt anyone else ? (children/siblings/elderly relative/stranger, for example. Consider HBV. Please specify who and what) Children ☐ Another family member ☐ Someone from a previous relationship ☐ Other (please specify) ☐	☐	☐
22. Has (…..) ever mistreated an animal or the family pet?	☐	☐
ABUSER(S)	YES	NO
23. Are there any financial issues? For example, are you dependent on (…..) for money/have they recently lost their job/other financial issues?	☐	☐
24. Has (…..) had problems in the past year with drugs (prescription or other), alcohol or mental health leading to problems in leading a normal life? (Please specify what) Drugs ☐ Alcohol ☐ Mental Health ☐	☐	☐
25. Has (…..) ever threatened or attempted suicide?	☐	☐
26. Has (…..) ever breached bail/an injunction and/or any agreement for when they can see you and/or the children? (Please specify what) Bail conditions ☐ Non Molestation/Occupation Order ☐ Child Contact arrangements ☐ Forced Marriage Protection Order ☐ Other ☐	☐	☐
27. Do you know if (……..) has ever been in trouble with the police or has a criminal history? (If yes, please specify) DV ☐ Sexual violence ☐ Other violence ☐ Other ☐	☐	☐
Other relevant information (from victim or officer) which may alter risk levels. Describe: (consider for example victim's vulnerability - disability, mental health, alcohol/substance misuse and/or the abuser's occupation/interests-does this give unique access to weapons i.e. ex-military, police, pest control) or is there serial offending?		
Is there anything else you would like to add to this?		

In **all** cases an initial risk classification is required:

RISK TO VICTIM:		
STANDARD ☐	MEDIUM ☐	HIGH ☐

©Laura Richards (2009). Please do not reproduce without permission. For enquiries about training staff in the use of the DASH, S-DASH or SN-DASH (2009) Risk Identification Checklists, please contact laura@laurarichards.co.uk

Electronic copy available at: https://ssrn.com/abstract=4336508

2. Details on estimation methods

2.1. IPW weighting

We initially assume that the propensity score is a logistic function of K predictors X_i and a commensurable vector of unknown parameters β, that is,

$$p(X_i) = \frac{exp(X_i\beta)}{1 + exp(X_i\beta)}. \tag{4}$$

We take three approaches to estimating $p(X_i)$. Our first approach iterates between specification, estimation, and balance-checking. Starting with a baseline set of predictors, we estimate β via logistic regression, a parametric procedure. This involves maximizing the logistic log-likelihood, or equivalently, setting its gradient $g(\beta)$ to zero, where

$$g(\beta) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - p(X_i)]X_i. \tag{5}$$

and n is the number of observations. This yields coefficient estimates $\hat{\beta}$, predicted values $\hat{p}_i = exp(X_i\hat{\beta})/[1 + exp(X_i\hat{\beta})]$, and estimated ATT weights $w_i = D_i + (1 - D_i)(\hat{p}_i/(1 - \hat{p}_i))$.

We then test for balance by regressing the treatment dummy D on the predictors X, weighting the observations by the estimated ATT weights. We test the null hypothesis that the coefficients in this regression are jointly zero. If the joint F-statistic rejects the null, we expand the set of predictors used to estimate the propensity scores. We do this by adding squared and and first-order interaction terms for all variables with absolute t-statistics greater than 10.[32] We then estimate the expanded model and test for balance again. If necessary, one could continue iterating along these lines, using lower threshold values for the t-statistics, until the estimated propensity scores balanced the predictors.

Our second approach is to estimate the propensity scores using a random forest. A random forest involves fitting a pre-specified number of regression or classification trees, where each tree is built from a random subsample of the data, and fit to a random subsample of the predictors. The trees are built recursively. At each node, all values of the selected predictors are scanned so as to split the node in the manner that maximizes the variance in the target variable, which is this case is the treatment indicator, between the resulting child nodes (Hastie et al., 2009). Predictions are averages over trees that did not make use of the *i*th observation.

Our third approach involves the Covariate Balancing Propensity Score (CBPS) of Imai and Ratkovic (2014). This approach is again parametric, but rather than focusing on functional form, it achieves balance using the baseline set of predictors. It does this by estimating the coefficients β in such a way as to target balance directly.

[32] For dichotomous predictors, we can only add the interaction terms.

Electronic copy available at: https://ssrn.com/abstract=4336508

To do so, CBPS assumes the logistic model in equation (4) for the probability of treatment. However, instead of solving the gradient of the logistic likelihood in equation (5), it estimates the coefficients by solving a set of balance conditions. These are given by

$$q(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{n}{n_1}[D_i - (1 - D_i)\frac{p(X_i)}{1 - p(X_i)}]X_i, \tag{6}$$

where n_1 is the number of observations in the treatment sample. Predicted values are based on the logistic probabilities in (4).[33]

2.2. Causal forests

This section draws heavily on Chernozhukov et al. (2018), Wager and Athey (2018), and particularly Athey et al. (2019). A causal forest provides an estimate of the conditional average treatment effect (CATE) at some $X_i = x$, which is given by

$$\tau(x) = E[Y_i(1)|X_i = x] - E[Y_i(0)|X_i = x]. \tag{7}$$

Like other estimators based on double or debiased machine learning, causal forests make use of two key techniques: double residualization and leave-out estimation. To deal with confounding, both the outcome and the treatment variable are residualized with respect to the covariates. This is conceptually similar to the familiar double-residual linear regression from the Frisch-Waugh-Lovell theorem, except that the residuals come from non-linear estimates of the conditional means of Y_i and D_i. The conditional means are typically estimated via machine-learning methods, whose adaptability means that they tend to overfit, leading to bias. The bias can be eliminated by the use of leave-out estimation, whereby no observation is used to estimate its own conditional means (Chernozhukov et al., 2018). This general approach can also be used to estimate the ATT, allowing us to to assess the robustness of our estimates based on IPW weighting to an alternative estimation technique.

For consistency with the other estimates in the paper, we use CBPS to produce leave-out estimates of $E(D_i|X_i)$ via 5-fold cross-fitting (Chernozhukov et al., 2018). To do this, we randomly assigned dyads with equal probability to one of five folds. We estimated $E(D_i|X_i)$ using the first four folds, then predicted $E(D_i|X_i)$ for the fifth fold based on those estimates. We then followed the same procedure to obtain predicted values for each of the five folds. To construct leave-out estimates of $E(Y_i|X_i)$, we used the default random forests from the R *grf* package, obtaining leave-out estimates by using predictions from trees to which the dyad to which the *i*th observation belongs did not contribute.

[33] One can also estimate the parameters via GMM, minimizing a quadratic form in the gradient and balance conditions jointly. That approach yielded estimated ATT's that were nearly identical to those reported here. We thank Erik Sverdrup for providing us with his code to solve the balance equations.

Electronic copy available at: https://ssrn.com/abstract=4336508

Let the residualized outcome and treatment variables be given by $\tilde{Y}_i = Y_i - \hat{Y}_i^{(-i)}$ and $\tilde{D}_i = D_i - \hat{p}_i^{(-i)}$, respectively, where $\hat{Y}_i^{(-i)}$ and $\hat{p}_i^{(-i)}$ denote the leave-out predictions of $E[Y_i|X_i]$ and $E(D_i|X_i) = P(D_i = 1|X_i)$.[34] The CATE is estimated by means of a weighted regression, which takes the form

$$\hat{\tau}(x) = [\sum_{i=1}^{n} \alpha_i(x)(\tilde{D}_i - \tilde{D}_\alpha)^2]^{-1} \sum_{i=1}^{n} \alpha_i(x)(\tilde{D}_i - \tilde{D}_\alpha)(\tilde{Y}_i - \tilde{Y}_\alpha), \tag{8}$$

where $\alpha_i(x)$ are weights, and $\tilde{D}_\alpha$ and $\tilde{Y}_\alpha$ are the weighted means of $\tilde{D}_i$ and $\tilde{Y}_i$.

The weights are learned from a causal forest, which is an ensemble of causal trees. Causal trees are constructed in a manner analogous to regression or classification trees, the main difference being that the natural target variable, the idiosyncratic CATE, is unobservable. Instead, the causal tree algorithm targets a variable that is proportional to the moment condition used to estimate a regression of the residualized outcome on the residualized treatment variable within the node.

Consider a particular split in a particular tree. The target variable at the parent node P is given by

$$v_i = A_P^{-1}(\tilde{D}_i - \tilde{D}_P)(\tilde{Y}_i - \tilde{Y}_P - (\tilde{D}_i - \tilde{D}_P)\hat{\beta}_P), \tag{9}$$

where $\tilde{D}_P$ and $\tilde{Y}_P$ are the averages of $\tilde{D}_i$ and $\tilde{Y}_i$ in the parent node, $A_P = \frac{1}{n_P} \sum_{\{X_i \epsilon P\}} (\tilde{D}_i - \tilde{D}_P)^2$, n_P is the number of observations in the parent node, and $\hat{\beta}_P$ is the coefficient from the regression of $\tilde{Y}_i - \tilde{Y}_P$ on $\tilde{D}_i - \tilde{D}_P$ in the parent node. Equation (9) is essentially the contribution of the *ith* observation to the moment condition used to calculate $\hat{\beta}_P$. The tree-building algorithm will split the parent node into two child nodes so as to maximize the variance of the v_i between those nodes, indirectly targeting heterogeneity in treatment effects. Once the parent node is split, the algorithm updates the node-specific quantities in equation (9) and splits the child nodes in the same manner. This proceeds until a certain pre-specified minimum number of treatment and control observations remain in each terminal leaf.

The causal forest consists of some pre-determined number of causal trees. Once the forest is built, the weights $\alpha_i(x)$ are constructed as the number of times that the ith observation falls into the same leaf as x, divided by the number of trees in the forest. They thus provide a measure of how close the ith observation is to x, similar to other matching methods. The difference is that causal forests learn the matching metric from the data, rather than requiring it to be pre-specified.

Despite the apparent potential for overfitting, Athey et al. (2019) provide conditions under which the estimated CATEs are consistent and asymptotically normal, with standard errors that

[34] Considering our clustered data, the approach described above ensures that these quantities are estimated without any data from the dyad to which the ith observation belongs.

Electronic copy available at: https://ssrn.com/abstract=4336508

can be estimated by conventional clustering techniques. Those conditions include conditional independence and common support, as above, plus conditions on the data and the sampling and splitting techniques used to build the causal forest.[35]

Once we have an estimated CATE for each observation in the sample, we construct doubly-robust scores, given by (grf-labs, 2022)

$$\hat{\theta}_i = \frac{D_i}{F_1}[\hat{\tau}_i(X_i) + (Y_i - \hat{Y}^{(-i)}(1))] - \frac{(1 - D_i)}{F_2}\frac{\hat{p}^{(-i)}(X_i)}{1 - \hat{p}^{(-i)}(X_i)}(Y_i - \hat{Y}^{(-i)}(0)), \tag{10}$$

where $\hat{Y}^{(-i)}(D_i)$ is a leave-out estimate of $E[Y_i(D_i)|X_i]$, $F_1 = n/n_1$ and $F_2 = n[\sum_{\{D_i=1\}} \frac{\hat{p}^{(-i)}(X_i)}{1-\hat{p}^{(-i)}}]^{-1}$.[36] The scores amount to an estimate of the CATE (for the treatment group) plus a bias-correction term. Given the weights on the bias-correction terms, the sample average of these scores should provide a consistent and asymptotically normal estimate of the ATT under conditions provided by Athey et al. (2019). We also use them to characterize heterogeneity in the treatment effects.

[35]Consistency and asymptotic normality require "honest" splitting. This means that, for each tree, a subsample of the data of size $S < n$ is drawn without replacement, where S scales appropriately with n. That subsample is then further split into two halves, the first of which is used for setting the splits, and the second of which contributes to estimating the weights. So-called out-of-bag predictions, based on trees in which the *ith* observation contributed neither to the splits nor the weights, are used for the leave-out estimates needed to residualize the outcome and treatment variable.

[36]Knaus (2022) provides an equivalent expression, except that he uses F_1, rather than F_2, to normalize the second term in equation (10).

Electronic copy available at: https://ssrn.com/abstract=4336508

3. Appendix tables

Table A1: Predictor means by treatment status

	Charged		High-risk		
Variable	No	Yes	No	Yes	Total
Current incident is labeled as a crime	0.318	1.000	0.356	0.819	0.398
Victim injured	0.075	0.319	0.082	0.317	0.103
Perp. injured	0.022	0.069	0.023	0.072	0.028
Victim influenced by alcohol	0.210	0.209	0.211	0.205	0.210
Perp. influenced by alcohol	0.310	0.408	0.320	0.336	0.322
Victim influenced by drugs	0.019	0.024	0.018	0.037	0.020
Victim ethnicity: Non-White	0.078	0.060	0.075	0.083	0.076
Perp. influenced by drugs	0.062	0.150	0.063	0.164	0.072
Perp. ethnicity: Non-White	0.117	0.123	0.113	0.163	0.118
Former partners	0.467	0.568	0.472	0.554	0.479
Partner status missing	0.068	0.063	0.068	0.070	0.068
At victim's home	0.592	0.588	0.590	0.608	0.591
Location info missing	0.070	0.074	0.072	0.061	0.071
Role switch	0.248	0.180	0.241	0.232	0.240
Weekend incident	0.342	0.350	0.347	0.303	0.343
Holiday incident	0.022	0.022	0.022	0.018	0.022
Initial Grade = 1	0.329	0.404	0.337	0.347	0.338
Initial grade greater than 2	0.031	0.031	0.031	0.035	0.031
Initial grade missing	0.048	0.020	0.046	0.032	0.045
DASH, number yes	4.250	9.320	4.246	10.875	4.841
DASH, number omitted	5.665	2.306	5.478	3.202	5.274
DASH Q. 28, omitted	0.174	0.282	0.168	0.376	0.186
DASH Q. 28, yes	0.153	0.056	0.148	0.072	0.141
DASH, 14 or more yes	0.041	0.213	0.031	0.371	0.061
Any high-risk incidents, past 3 mo (dyad)	0.044	0.094	0.033	0.215	0.050
Any high-risk incidents, past 6 mo (dyad)	0.062	0.133	0.049	0.293	0.071
Any high-risk incidents, past 12 mo (dyad)	0.082	0.173	0.066	0.364	0.092
Officer male	0.743	0.725	0.746	0.684	0.741
Officer experience	10.257	9.886	10.257	9.770	10.214
Officer visited dyad more than once	0.026	0.035	0.026	0.044	0.027
Number of reports in sample (leave-out)	93.371	93.278	93.975	87.123	93.360
Officer's charge share (leave-out)	0.115	0.131	0.115	0.129	0.117
Officer's high-risk share (leave-out)	0.088	0.099	0.086	0.122	0.090
Officer's share blank DASH (leave-out)	0.098	0.096	0.097	0.098	0.097
Officer's average DASH Q's omitted (leave-out)	2.849	2.802	2.845	2.830	2.844
Dyad in any DA calls, past 3 mo	0.329	0.411	0.323	0.494	0.339
Dyad in any DA calls, past 6 mo	0.427	0.537	0.422	0.621	0.440

Electronic copy available at: https://ssrn.com/abstract=4336508

Table A1: Predictor means by treatment status *(continued)*

	Charged		High-risk		
Variable	No	Yes	No	Yes	Total
Dyad in any DA calls, past 12 mo	0.517	0.642	0.514	0.716	0.532
Dyad in one DA call, past 2 yrs	0.193	0.184	0.195	0.165	0.192
Dyad in over 1 DA call, past 2 yrs	0.397	0.529	0.394	0.606	0.413
Dyad in any DA crimes, past 3 mo	0.153	0.254	0.148	0.332	0.165
Dyad in any DA crimes, past 6 mo	0.216	0.357	0.211	0.450	0.233
Dyad in any DA crimes, past 12 mo	0.283	0.451	0.277	0.552	0.302
Dyad in one DA crime, past 2 yrs	0.191	0.228	0.191	0.236	0.195
Dyad in over 1 DA crime, past 2 yrs	0.153	0.295	0.149	0.381	0.170
Dyad in any DA calls involving violence, past 3 mo	0.053	0.076	0.048	0.128	0.055
Dyad in any DA calls involving violence, past 6 mo	0.083	0.124	0.077	0.197	0.087
Dyad in any DA calls involving violence, past 12 mo	0.120	0.185	0.113	0.279	0.128
Dyad in one DA call involving violence, past 2 yrs	0.163	0.250	0.156	0.349	0.173
Perp. male	0.818	0.954	0.821	0.960	0.834
Perp. in 1 DA incident, past 2 yrs	0.195	0.172	0.197	0.147	0.193
Perp. in over 1 DA incident, past 2 yrs	0.406	0.597	0.406	0.661	0.429
Perp. accused of 1 DA crime, past 2 yrs	0.190	0.234	0.192	0.229	0.195
Perp. accused of over 1 DA crime, past 2 yrs	0.160	0.345	0.158	0.430	0.182
Perp. accused of 1 crime, past 2 yrs	0.150	0.197	0.152	0.189	0.155
Perp. accused of over 1 crime, past 2 yrs	0.159	0.367	0.163	0.392	0.183
Perp. accused of DA violence, past 2 yrs	0.159	0.280	0.153	0.376	0.173
Perp. accused of violence with injury, past 2 yrs	0.104	0.222	0.104	0.265	0.118
Perp. accused of violence without injury, past 2 yrs	0.098	0.211	0.099	0.232	0.111
Perp. accused of violating protection order, past 2 yrs	0.103	0.230	0.108	0.220	0.118
Perp. accused of stalking, past 2 yrs	0.044	0.159	0.047	0.158	0.057
Victim age less than 20	0.045	0.048	0.044	0.053	0.045
Victim age 20-24	0.165	0.195	0.167	0.188	0.169
Victim age 30-34	0.181	0.187	0.181	0.192	0.182
Victim age 35-39	0.136	0.129	0.136	0.129	0.135
Victim age 40-44	0.100	0.089	0.101	0.083	0.099
Victim age 45-49	0.077	0.065	0.077	0.061	0.075
Victim age 50-54	0.046	0.034	0.046	0.031	0.045
Victim age 55-59	0.023	0.015	0.022	0.015	0.022
Victim age over =60	0.018	0.010	0.018	0.009	0.017
Perp. age less than 20	0.029	0.022	0.028	0.025	0.028
Perp. age 20-24	0.133	0.145	0.135	0.136	0.135
Perp. age 30-34	0.188	0.206	0.188	0.209	0.190
Perp. age 35-39	0.146	0.142	0.145	0.152	0.145
Perp. age 40-44	0.112	0.099	0.111	0.101	0.110
Perp. age 45-49	0.090	0.078	0.090	0.077	0.088

Electronic copy available at: https://ssrn.com/abstract=4336508

Table A1: Predictor means by treatment status *(continued)*

	Charged		High-risk		
Variable	No	Yes	No	Yes	Total
Perp. age 50-54	0.055	0.048	0.055	0.047	0.054
Perp. age 55-59	0.026	0.020	0.026	0.021	0.025
Perp. age over =60	0.021	0.011	0.021	0.014	0.020
Incident informant is the victim	0.617	0.607	0.621	0.571	0.616

Electronic copy available at: https://ssrn.com/abstract=4336508

Table A2: Coefficient estimates for propensity of charged

	Propensity score model		
	Logistic baseline (1)	Logistic expanded (2)	CBPS (3)
Victim injured	1.320	1.961	1.233
	(0.024)	(0.114)	(0.021)
Perp. injured	0.248	0.195	0.163
	(0.043)	(0.042)	(0.039)
Victim influenced by alcohol	-0.226	-0.203	-0.256
	(0.028)	(0.028)	(0.025)
Perp. influenced by alcohol	0.444	0.761	0.420
	(0.023)	(0.103)	(0.020)
Victim influenced by drugs	-0.354	-0.329	-0.297
	(0.065)	(0.064)	(0.059)
Victim ethnicity × Non-White	-0.083	-0.079	-0.109
	(0.040)	(0.041)	(0.033)
Perp. influenced by drugs	0.191	0.199	0.126
	(0.030)	(0.030)	(0.029)
Perp. ethnicity × Non-White	0.085	0.076	0.087
	(0.030)	(0.030)	(0.027)
Former partners	0.227	0.473	0.227
	(0.021)	(0.102)	(0.018)
Partner status missing	0.116	0.092	0.107
	(0.040)	(0.040)	(0.035)
At victim's home	0.042	0.039	0.066
	(0.020)	(0.020)	(0.018)
Location info missing	0.140	0.139	0.163
	(0.037)	(0.037)	(0.033)
Role switch	-0.416	-0.325	-0.405
	(0.027)	(0.109)	(0.026)
Weekend incident	0.070	0.068	0.062
	(0.019)	(0.019)	(0.017)
Holiday incident	-0.000	-0.003	0.015
	(0.063)	(0.063)	(0.058)
Initial Grade = 1	0.416	0.310	0.435
	(0.020)	(0.103)	(0.018)

Electronic copy available at: https://ssrn.com/abstract=4336508

Initial grade greater than 2	-0.014	-0.037	-0.083
	(0.053)	(0.052)	(0.048)
Initial grade missing	-0.667	-0.567	-0.705
	(0.061)	(0.312)	(0.057)
DASH, number yes	0.199	0.436	0.178
	(0.003)	(0.017)	(0.003)
DASH, number omitted	0.009	0.019	0.005
	(0.002)	(0.002)	(0.002)
DASH Q. 28, omitted	-0.135	-0.142	-0.135
	(0.022)	(0.022)	(0.020)
DASH Q. 28, yes	-0.040	0.126	-0.056
	(0.061)	(0.063)	(0.053)
DASH, 14 or more yes	-0.632	-1.508	-0.569
	(0.035)	(0.341)	(0.033)
Any high-risk incidents, past 3 mo (dyad)	-0.019	-0.013	0.019
	(0.067)	(0.065)	(0.072)
Any high-risk incidents, past 6 mo (dyad)	-0.008	-0.003	-0.024
	(0.078)	(0.076)	(0.084)
Any high-risk incidents, past 12 mo (dyad)	0.042	0.046	0.023
	(0.057)	(0.055)	(0.060)
Officer male	0.064	0.065	0.072
	(0.021)	(0.021)	(0.017)
Officer experience	0.001	0.001	0.002
	(0.002)	(0.002)	(0.001)
Officer visited dyad more than once	0.010	0.010	0.043
	(0.051)	(0.051)	(0.052)
Number of reports in sample (leave-out)	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Officer's charge share (leave-out)	2.332	2.364	2.500
	(0.132)	(0.598)	(0.119)
Officer's high-risk share (leave-out)	-0.826	-0.811	-0.666
	(0.130)	(0.129)	(0.125)
Officer's share blank DASH (leave-out)	0.499	0.489	0.449
	(0.074)	(0.074)	(0.065)
Officer's average DASH Q's omitted (leave-out)	0.005	0.002	0.007
	(0.003)	(0.003)	(0.003)
Dyad in any DA calls, past 3 mo	-0.184	-0.173	-0.165
	(0.038)	(0.038)	(0.041)
Dyad in any DA calls, past 6 mo	0.004	-0.000	-0.057

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.046)	(0.046)	(0.044)
Dyad in any DA calls, past 12 mo	0.040	0.042	0.061
	(0.050)	(0.050)	(0.044)
Dyad in one DA call, past 2 yrs	0.047	0.029	0.052
	(0.052)	(0.052)	(0.044)
Dyad in over 1 DA call, past 2 yrs	0.069	0.036	0.096
	(0.061)	(0.062)	(0.052)
Dyad in any DA crimes, past 3 mo	0.146	0.126	0.162
	(0.050)	(0.049)	(0.052)
Dyad in any DA crimes, past 6 mo	0.081	0.078	0.059
	(0.057)	(0.057)	(0.054)
Dyad in any DA crimes, past 12 mo	0.098	0.103	0.088
	(0.057)	(0.057)	(0.052)
Dyad in one DA crime, past 2 yrs	0.024	0.011	0.067
	(0.058)	(0.058)	(0.054)
Dyad in over 1 DA crime, past 2 yrs	0.247	0.218	0.274
	(0.072)	(0.072)	(0.070)
Dyad in any DA calls involving violence, past 3 mo	-0.146	-0.143	-0.159
	(0.064)	(0.062)	(0.061)
Dyad in any DA calls involving violence, past 6 mo	-0.015	-0.006	0.002
	(0.066)	(0.065)	(0.070)
Dyad in any DA calls involving violence, past 12 mo	-0.064	-0.073	-0.022
	(0.060)	(0.059)	(0.060)
Dyad in one DA call involving violence, past 2 yrs	0.038	0.042	-0.008
	(0.060)	(0.059)	(0.056)
Perp. male	0.751	1.231	0.778
	(0.042)	(0.134)	(0.029)
Perp. in 1 DA incident, past 2 yrs	0.089	0.084	0.078
	(0.042)	(0.042)	(0.030)
Perp. in over 1 DA incident, past 2 yrs	0.104	0.112	0.093
	(0.049)	(0.049)	(0.040)
Perp. accused of 1 DA crime, past 2 yrs	-0.016	-0.011	-0.027
	(0.045)	(0.045)	(0.037)
Perp. accused of over 1 DA crime, past 2 yrs	-0.140	-0.133	-0.161
	(0.059)	(0.058)	(0.052)
Perp. accused of 1 crime, past 2 yrs	0.297	0.312	0.303
	(0.029)	(0.030)	(0.025)
Perp. accused of over 1 crime, past 2 yrs	0.473	0.904	0.473
	(0.037)	(0.131)	(0.034)

Electronic copy available at: https://ssrn.com/abstract=4336508

Perp. accused of DA violence, past 2 yrs	-0.048	-0.039	-0.043
	(0.047)	(0.046)	(0.041)
Perp. accused of violence with injury, past 2 yrs	0.130	0.138	0.117
	(0.032)	(0.032)	(0.032)
Perp. accused of violence without injury, past 2 yrs	0.119	0.114	0.078
	(0.028)	(0.028)	(0.030)
Perp. accused of violating protection order, past 2 yrs	0.266	0.254	0.228
	(0.028)	(0.028)	(0.027)
Perp. accused of stalking, past 2 yrs	0.718	1.200	0.612
	(0.033)	(0.202)	(0.031)
Victim age less than 20	0.018	0.028	0.010
	(0.054)	(0.054)	(0.045)
Victim age 20-24	0.013	0.016	0.007
	(0.030)	(0.030)	(0.026)
Victim age 30-34	0.032	0.026	0.028
	(0.029)	(0.029)	(0.026)
Victim age 35-39	0.044	0.041	0.052
	(0.034)	(0.034)	(0.029)
Victim age 40-44	0.105	0.098	0.116
	(0.040)	(0.040)	(0.035)
Victim age 45-49	0.093	0.071	0.101
	(0.047)	(0.047)	(0.042)
Victim age 50-54	0.057	0.023	0.099
	(0.059)	(0.059)	(0.052)
Victim age 55-59	0.154	0.102	0.175
	(0.081)	(0.082)	(0.078)
Victim age over =60	0.203	0.161	0.212
	(0.102)	(0.103)	(0.077)
Perp. age less than 20	-0.262	-0.253	-0.270
	(0.071)	(0.071)	(0.067)
Perp. age 20-24	0.025	0.030	0.018
	(0.033)	(0.033)	(0.030)
Perp. age 30-34	-0.054	-0.053	-0.074
	(0.029)	(0.029)	(0.026)
Perp. age 35-39	-0.131	-0.131	-0.137
	(0.034)	(0.033)	(0.029)
Perp. age 40-44	-0.160	-0.172	-0.180
	(0.039)	(0.039)	(0.034)

Electronic copy available at: https://ssrn.com/abstract=4336508

Perp. age 45-49	-0.137	-0.145	-0.165
	(0.044)	(0.044)	(0.038)
Perp. age 50-54	-0.093	-0.093	-0.131
	(0.053)	(0.053)	(0.048)
Perp. age 55-59	-0.209	-0.202	-0.249
	(0.074)	(0.074)	(0.061)
Perp. age over =60	-0.439	-0.446	-0.458
	(0.097)	(0.097)	(0.067)
Incident informant is the victim	-0.048	-0.067	-0.047
	(0.019)	(0.019)	(0.017)
Victim injured_X_Perp. influenced by alcohol		-0.268	
		(0.045)	
Victim injured_X_Former partners		-0.041	
		(0.046)	
Victim injured_X_Role switch		0.102	
		(0.057)	
Victim injured_X_Initial Grade = 1		-0.042	
		(0.046)	
Victim injured_X_Initial grade missing		-0.721	
		(0.173)	
Victim injured_X_DASH, number yes		-0.065	
		(0.006)	
Victim injured_X_DASH, 14 or more yes		0.222	
		(0.079)	
Victim injured_X_Officer's charge share (leave-out)		-0.430	
		(0.275)	
Victim injured_X_Perp. male		0.135	
		(0.093)	
Victim injured_X_Perp. accused of over 1 crime, past 2 yrs		-0.159	
		(0.056)	
Victim injured_X_Perp. accused of stalking, past 2 yrs		-0.324	
		(0.098)	
Perp. influenced by alcohol_X_Former partners		0.037	
		(0.041)	
Perp. influenced by alcohol_X_Role switch		-0.023	
		(0.051)	
Perp. influenced by alcohol_X_Initial Grade = 1		0.168	

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.040)
Perp. influenced by alcohol_X_Initial grade missing	-0.200
	(0.152)
Perp. influenced by alcohol_X_DASH, number yes	-0.005
	(0.005)
Perp. influenced by alcohol_X_DASH, 14 or more yes	-0.025
	(0.073)
Perp. influenced by alcohol_X_Officer's charge share (leave-out)	0.085
	(0.263)
Perp. influenced by alcohol_X_Perp. male	-0.246
	(0.086)
Perp. influenced by alcohol_X_Perp. accused of over 1 crime, past 2 yrs	-0.206
	(0.047)
Perp. influenced by alcohol_X_Perp. accused of stalking, past 2 yrs	-0.049
	(0.067)
Former partners_X_Role switch	0.024
	(0.051)
Former partners_X_Initial Grade = 1	0.174
	(0.041)
Former partners_X_Initial grade missing	-0.250
	(0.132)
Former partners_X_DASH, number yes	-0.013
	(0.005)
Former partners_X_DASH, 14 or more yes	-0.054
	(0.073)
Former partners_X_Officer's charge share (leave-out)	-0.079
	(0.254)
Former partners_X_Perp. male	-0.291
	(0.088)
Former partners_X_Perp. accused of over 1 crime, past 2 yrs	0.190
	(0.048)
Former partners_X_Perp. accused of stalking, past 2 yrs	0.038
	(0.069)

Electronic copy available at: https://ssrn.com/abstract=4336508

Role switch_X_Initial Grade = 1	0.041
	(0.052)
Role switch_X_Initial grade missing	0.002
	(0.162)
Role switch_X_DASH, number yes	-0.004
	(0.007)
Role switch_X_DASH, 14 or more yes	-0.070
	(0.088)
Role switch_X_Officer's charge share (leave-out)	0.384
	(0.313)
Role switch_X_Perp. male	-0.167
	(0.084)
Role switch_X_Perp. accused of over 1 crime, past 2 yrs	0.078
	(0.056)
Role switch_X_Perp. accused of stalking, past 2 yrs	0.010
	(0.074)
Initial Grade = 1_X_DASH, number yes	0.014
	(0.006)
Initial Grade = 1_X_DASH, 14 or more yes	-0.016
	(0.075)
Initial Grade = 1_X_Officer's charge share (leave-out)	0.163
	(0.267)
Initial Grade = 1_X_Perp. male	-0.094
	(0.088)
Initial Grade = 1_X_Perp. accused of over 1 crime, past 2 yrs	-0.235
	(0.048)
Initial Grade = 1_X_Perp. accused of stalking, past 2 yrs	-0.282
	(0.071)
Initial grade missing_X_DASH, number yes	0.005
	(0.017)
Initial grade missing_X_DASH, 14 or more yes	-0.201
	(0.218)
Initial grade missing_X_Officer's charge share (leave-out)	-0.064
	(0.494)

Electronic copy available at: https://ssrn.com/abstract=4336508

Initial grade missing_X_Perp. male	0.158
	(0.268)
Initial grade missing_X_Perp. accused of over 1 crime, past 2 yrs	0.159
	(0.145)
Initial grade missing_X_Perp. accused of stalking, past 2 yrs	0.168
	(0.166)
DASH, number yes_squared	-0.012
	(0.001)
DASH, number yes_X_DASH, 14 or more yes	0.078
	(0.020)
DASH, number yes_X_Officer's charge share (leave-out)	-0.107
	(0.032)
DASH, number yes_X_Perp. male	-0.015
	(0.011)
DASH, number yes_X_Perp. accused of over 1 crime, past 2 yrs	-0.025
	(0.006)
DASH, number yes_X_Perp. accused of stalking, past 2 yrs	-0.077
	(0.008)
DASH, 14 or more yes_X_Officer's charge share (leave-out)	1.063
	(0.449)
DASH, 14 or more yes_X_Perp. male	0.274
	(0.196)
DASH, 14 or more yes_X_Perp. accused of over 1 crime, past 2 yrs	0.113
	(0.079)
DASH, 14 or more yes_X_Perp. accused of stalking, past 2 yrs	0.272
	(0.108)
Officer's charge share (leave-out)_squared	1.800
	(0.319)
Officer's charge share (leave-out)_X_Perp. male	-0.519
	(0.472)
Officer's charge share (leave-out)_X_Perp. accused of over 1 crime, past 2 yrs	-0.094

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.287)
Officer's charge share (leave-out)_X_Perp. accused of stalking, past 2 yrs	-0.041
	(0.421)
Perp. male_X_Perp. accused of over 1 crime, past 2 yrs	-0.171
	(0.109)
Perp. male_X_Perp. accused of stalking, past 2 yrs	0.137
	(0.168)
Perp. accused of over 1 crime, past 2 yrs_X_Perp. accused of stalking, past 2 yrs	0.047
	(0.077)

Notes: Estimated from full sample, N=154,102. Standard errors are clustered at the level of the dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508

Table A3: Coefficient estimates for propensity of high-risk

	Propensity score model		
	Logistic baseline (1)	Logistic expanded (2)	CBPS (3)
Current incident is labeled as a crime	1.341	1.141	1.333
	(0.028)	(0.142)	(0.031)
Victim injured	0.695	1.956	0.781
	(0.029)	(0.161)	(0.033)
Perp. injured	0.274	0.261	0.210
	(0.049)	(0.050)	(0.053)
Victim influenced by alcohol	0.010	0.010	0.081
	(0.035)	(0.035)	(0.036)
Perp. influenced by alcohol	-0.089	-0.086	-0.050
	(0.030)	(0.030)	(0.035)
Victim influenced by drugs	0.062	0.024	-0.112
	(0.070)	(0.071)	(0.106)
Victim ethnicity × Non-White	0.177	0.183	0.210
	(0.045)	(0.045)	(0.054)
Perp. influenced by drugs	0.223	0.217	0.222
	(0.036)	(0.036)	(0.045)
Perp. ethnicity × Non-White	0.353	0.348	0.294
	(0.034)	(0.034)	(0.047)
Former partners	-0.108	-0.113	-0.083
	(0.026)	(0.026)	(0.031)
Partner status missing	-0.096	-0.100	-0.107
	(0.048)	(0.048)	(0.056)
At victim's home	0.127	0.128	0.120
	(0.024)	(0.024)	(0.029)
Location info missing	-0.043	-0.048	-0.122
	(0.049)	(0.049)	(0.064)
Role switch	-0.135	-0.145	-0.157
	(0.031)	(0.031)	(0.041)
Weekend incident	-0.111	-0.106	-0.107
	(0.025)	(0.025)	(0.028)
Holiday incident	-0.223	-0.206	-0.166
	(0.083)	(0.084)	(0.076)
Initial Grade = 1	0.097	0.112	0.127

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.026)	(0.026)	(0.032)
Initial grade greater than 2	-0.088	-0.094	-0.123
	(0.063)	(0.063)	(0.088)
Initial grade missing	-0.346	-0.378	-0.421
	(0.062)	(0.062)	(0.071)
DASH, number yes	0.224	0.207	0.212
	(0.004)	(0.017)	(0.006)
DASH, number omitted	0.049	0.002	0.052
	(0.002)	(0.019)	(0.003)
DASH Q. 28, omitted	0.196	0.177	0.154
	(0.026)	(0.026)	(0.031)
DASH Q. 28, yes	-0.053	0.058	-0.114
	(0.064)	(0.071)	(0.063)
DASH, 14 or more yes	0.355	0.303	0.312
	(0.040)	(0.049)	(0.044)
Any high-risk incidents, past 3 mo (dyad)	0.227	0.291	0.276
	(0.067)	(0.066)	(0.086)
Any high-risk incidents, past 6 mo (dyad)	0.173	0.168	0.056
	(0.078)	(0.077)	(0.095)
Any high-risk incidents, past 12 mo (dyad)	1.541	1.957	1.468
	(0.059)	(0.149)	(0.076)
Officer male	-0.029	-0.029	-0.009
	(0.025)	(0.025)	(0.030)
Officer experience	0.005	0.005	0.003
	(0.002)	(0.002)	(0.003)
Officer visited dyad more than once	0.042	0.060	0.086
	(0.060)	(0.059)	(0.073)
Number of reports in sample (leave-out)	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)
Officer's charge share (leave-out)	-1.118	-1.150	-0.911
	(0.169)	(0.172)	(0.188)
Officer's high-risk share (leave-out)	3.321	3.796	3.335
	(0.140)	(0.655)	(0.164)
Officer's share blank DASH (leave-out)	0.398	0.410	0.353
	(0.089)	(0.090)	(0.104)
Officer's average DASH Q's omitted (leave-out)	-0.008	-0.007	-0.004
	(0.004)	(0.004)	(0.005)
Dyad in any DA calls, past 3 mo	-0.036	-0.045	0.020
	(0.047)	(0.047)	(0.062)

Electronic copy available at: https://ssrn.com/abstract=4336508

Dyad in any DA calls, past 6 mo	0.047	0.052	-0.032
	(0.059)	(0.059)	(0.063)
Dyad in any DA calls, past 12 mo	0.055	0.049	0.050
	(0.067)	(0.067)	(0.061)
Dyad in one DA call, past 2 yrs	-0.086	-0.095	-0.118
	(0.068)	(0.069)	(0.068)
Dyad in over 1 DA call, past 2 yrs	-0.074	-0.090	-0.114
	(0.080)	(0.081)	(0.085)
Dyad in any DA crimes, past 3 mo	0.025	0.002	-0.159
	(0.060)	(0.060)	(0.075)
Dyad in any DA crimes, past 6 mo	0.024	0.033	0.182
	(0.070)	(0.069)	(0.079)
Dyad in any DA crimes, past 12 mo	-0.098	-0.080	-0.075
	(0.072)	(0.073)	(0.080)
Dyad in one DA crime, past 2 yrs	0.103	0.096	0.132
	(0.074)	(0.075)	(0.082)
Dyad in over 1 DA crime, past 2 yrs	0.041	0.025	0.070
	(0.091)	(0.091)	(0.104)
Dyad in any DA calls involving violence, past 3 mo	0.054	0.082	0.057
	(0.069)	(0.069)	(0.099)
Dyad in any DA calls involving violence, past 6 mo	0.005	-0.013	-0.019
	(0.073)	(0.073)	(0.101)
Dyad in any DA calls involving violence, past 12 mo	0.003	-0.012	-0.027
	(0.069)	(0.069)	(0.089)
Dyad in one DA call involving violence, past 2 yrs	-0.004	-0.003	-0.028
	(0.072)	(0.072)	(0.080)
Perp. male	0.748	0.607	0.802
	(0.054)	(0.162)	(0.053)
Perp. in 1 DA incident, past 2 yrs	0.022	0.025	0.060
	(0.054)	(0.055)	(0.045)
Perp. in over 1 DA incident, past 2 yrs	0.044	0.045	0.115
	(0.063)	(0.064)	(0.057)
Perp. accused of 1 DA crime, past 2 yrs	0.137	0.134	0.022
	(0.058)	(0.058)	(0.058)
Perp. accused of over 1 DA crime, past 2 yrs	0.304	0.297	0.217
	(0.073)	(0.073)	(0.078)
Perp. accused of 1 crime, past 2 yrs	0.070	0.070	0.019
	(0.036)	(0.036)	(0.040)
Perp. accused of over 1 crime, past 2 yrs	0.178	0.175	0.094

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.045)	(0.045)	(0.055)
Perp. accused of DA violence, past 2 yrs	0.106	0.100	0.151
	(0.057)	(0.057)	(0.063)
Perp. accused of violence with injury, past 2 yrs	0.070	0.085	0.143
	(0.038)	(0.038)	(0.048)
Perp. accused of violence without injury, past 2 yrs	0.046	0.041	0.000
	(0.035)	(0.035)	(0.043)
Perp. accused of violating protection order, past 2 yrs	-0.007	-0.015	0.015
	(0.035)	(0.035)	(0.045)
Perp. accused of stalking, past 2 yrs	0.182	0.177	0.174
	(0.040)	(0.040)	(0.057)
Victim age less than 20	0.138	0.142	-0.030
	(0.065)	(0.065)	(0.091)
Victim age 20-24	-0.042	-0.042	-0.098
	(0.037)	(0.037)	(0.049)
Victim age 30-34	-0.056	-0.056	-0.057
	(0.036)	(0.036)	(0.046)
Victim age 35-39	-0.130	-0.128	-0.054
	(0.042)	(0.042)	(0.050)
Victim age 40-44	-0.135	-0.139	-0.130
	(0.050)	(0.050)	(0.059)
Victim age 45-49	-0.201	-0.186	-0.124
	(0.058)	(0.059)	(0.063)
Victim age 50-54	-0.210	-0.202	-0.198
	(0.074)	(0.074)	(0.085)
Victim age 55-59	-0.071	-0.054	-0.048
	(0.099)	(0.100)	(0.116)
Victim age over =60	-0.057	-0.027	-0.038
	(0.127)	(0.127)	(0.119)
Perp. age less than 20	-0.157	-0.162	-0.038
	(0.085)	(0.085)	(0.126)
Perp. age 20-24	-0.043	-0.041	0.004
	(0.041)	(0.042)	(0.050)
Perp. age 30-34	0.045	0.042	-0.037
	(0.036)	(0.036)	(0.042)
Perp. age 35-39	0.076	0.078	0.022
	(0.041)	(0.041)	(0.049)
Perp. age 40-44	0.076	0.073	0.002

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.048)	(0.048)	(0.054)
Perp. age 45-49	0.111	0.106	0.051
	(0.054)	(0.054)	(0.062)
Perp. age 50-54	0.145	0.135	0.091
	(0.065)	(0.066)	(0.070)
Perp. age 55-59	0.178	0.158	0.071
	(0.088)	(0.089)	(0.088)
Perp. age over =60	0.261	0.251	0.136
	(0.108)	(0.109)	(0.104)
Incident informant is the victim	-0.338	-0.584	-0.294
	(0.024)	(0.125)	(0.027)
Current incident is labeled as a crime_X_Victim injured		-0.911	
		(0.100)	
Current incident is labeled as a crime_X_DASH, number yes		-0.000	
		(0.007)	
Current incident is labeled as a crime_X_DASH, number omitted		0.012	
		(0.004)	
Current incident is labeled as a crime_X_Any high-risk incidents, past 12 mo (dyad)		0.377	
		(0.059)	
Current incident is labeled as a crime_X_Officer's high-risk share (leave-out)		0.098	
		(0.310)	
Current incident is labeled as a crime_X_Perp. male		0.010	
		(0.124)	
Current incident is labeled as a crime_X_Incident informant is the victim		0.058	
		(0.056)	
Victim injured_X_DASH, number yes		0.001	
		(0.007)	
Victim injured_X_DASH, number omitted		0.009	
		(0.004)	
Victim injured_X_Any high-risk incidents, past 12 mo (dyad)		0.240	
		(0.080)	
Victim injured_X_Officer's high-risk share (leave-out)		0.329	

Electronic copy available at: https://ssrn.com/abstract=4336508

	(0.291)
Victim injured_X_Perp. male	-0.491
	(0.113)
Victim injured_X_Incident informant is the victim	-0.118
	(0.056)
DASH, number yes_squared	0.000
	(0.001)
DASH, number yes_X_DASH, number omitted	0.004
	(0.001)
DASH, number yes_X_Any high-risk incidents, past 12 mo (dyad)	-0.115
	(0.007)
DASH, number yes_X_Officer's high-risk share (leave-out)	0.043
	(0.032)
DASH, number yes_X_Perp. male	0.032
	(0.012)
DASH, number yes_X_Incident informant is the victim	0.004
	(0.006)
DASH, number omitted_squared	0.001
	(0.001)
DASH, number omitted_X_Any high-risk incidents, past 12 mo (dyad)	-0.021
	(0.004)
DASH, number omitted_X_Officer's high-risk share (leave-out)	0.035
	(0.016)
DASH, number omitted_X_Perp. male	0.015
	(0.006)
DASH, number omitted_X_Incident informant is the victim	-0.000
	(0.003)
Any high-risk incidents, past 12 mo (dyad)_X_Officer's high-risk share (leave-out)	0.599
	(0.335)
Any high-risk incidents, past 12 mo (dyad)_X_Perp. male	0.176
	(0.114)

Electronic copy available at: https://ssrn.com/abstract=4336508

Any high-risk incidents, past 12 mo (dyad)_X_Incident informant is the victim	0.198
	(0.055)
Officer's high-risk share (leave-out)_squared	0.025
	(0.348)
Officer's high-risk share (leave-out)_X_Perp. male	-1.340
	(0.460)
Officer's high-risk share (leave-out)_X_Incident informant is the victim	-0.145
	(0.254)
Perp. male_X_Incident informant is the victim	0.187
	(0.104)

Notes: Estimated from full sample, N=154,102. Standard errors are clustered at the level of the dyad.

Electronic copy available at: https://ssrn.com/abstract=4336508