# The Effects of Policing and Labor Market Interventions on Crime: An Econometric Analysis

Victoria Kaiser (née Endl-Geyer)

*Dissertation*

ifo Institute
Center for Labor and Demographic Economics

&

Ludwig Maximilian University Munich
Department of Economics

2024

# The Effects of Policing and Labor Market Interventions on Crime: An Econometric Analysis

Inaugural-Dissertation
zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Ludwig-Maximilians-Universität München

2024

vorgelegt von:

**Victoria Kaiser**

Referent: Prof. Helmut Rainer, Ph.D.
Korreferent: Prof. Ana Tur-Prats, Ph.D.
Promotionsabschlussberatung: 31.01.2024

Datum der mündlichen Prüfung: 10. Januar 2024
Berichterstatter: Prof. Helmut Rainer, Ph.D., Prof. Ana Tur-Prats, Ph.D. und Prof. Ingrid Hägele, Ph.D.

# Acknowledgments

I am thankful to my advisor Helmut Rainer for the opportunity to pursue a Ph.D. He literally always had an open door - not only for Ph.D. related advice but also for helping me find the perfect road bike (which I love). I learned a lot from working together on the second chapter of this dissertation. I would also like to thank Ana Tur-Prats and Ingrid Hägele for joining the examination committee.

I am grateful to Sofia Amaral. Through our joint work, I developed a passion for the topic of gender based violence and got to know many inspiring researchers in the field. I thank my co-authors Gordon Dahl, Kim Chaney, Timo Hener, Pablo Kolb, Nishith Prakash, and Abhilasha Sahay for collaboration and their genuine advice. I owe special thanks to Gordon, who hosted me at UC San Diego.

I received generous funding from ifo Institute, Leibniz Association, and Joachim Herz Foundation. I am grateful for the many opportunities to travel, present my research, and expand my network.

I am very glad to have received the support of colleagues (former and new) at ifo Institute who became friends. In particular, I want to thank Anna, Clara, Ele, Fabian, Katharina, Michael, Leander, Marc, Pablo, Patrick, Sarah, and Stella for our journey, marked by countless coffee breaks, shared challenges, Starkbierfest dances and laughter. I am grateful also to my long-time friends who were always there (in the form of encouraging messages, spontaneous visits, spa trips, hikes, spaghetti carbonara, beer and wine): Chrissi, Jan, Kathi, Kathrin, Simone, and Theresa.

The reason why I never doubted that it was the right decision to pursue my Ph.D. is the fact that I met my husband Michael at LMU. He is the greatest blessing of my university days and brought immeasurable joy and completeness to my life. I am eternally grateful for him and our love story.

The love and unconditional support of my family are the foundation of who I am today. My sister Gina brings warmth and sunshine to my world and I know that in her I have my closest ally. I could not imagine my journey without her by my side. My parents Gordana and Reinhold taught me that the pursuit of dreams and the courage to try are achievements in themselves. Their cheers, warm embraces, and faith in me - regardless of outcomes - have been a source of profound inspiration and empowered me to soar higher and dream bigger. I am aware that when they celebrate my success, it is because they are happy for me and not because the success itself is important to them. Their joy is a reflection of their love, and that makes each triumph all the more meaningful. *Von Herzen Danke!*

Victoria Kaiser

# Table of Contents

# List of Tables in Main Text

# List of Figures in Main Text

# Preface

Violence against women is a blatant human rights violation. The World Health Organization (WHO) estimates that, across the globe, 30% of women experience Intimate Partner Violence (IPV) or non-partner sexual violence at some point in their life (WHO, 2021). Homicides committed by partners remain one of the leading causes of mortality among women (UNODC, 2019). In addition to the grave humanitarian consequences, there is also a significant economic impact at the individual, community, and national levels. In fact, it is one of the costliest crime types as it comes not only with severe repercussions for the individual but also for society and the economy as a whole (Bindler, Ketel, and Hjalmarsson 2020; Chalfin 2015). The European Institute for Gender Equality (2014) estimates that the implied costs of Gender Based Violence (GBV) in an average European country amounts to €4.5 billion per year.

GBV affects institutions such as the police, specialized services, the health- and legal sector, and the allocation of their resources. It can result in physical and psychological injuries that can lead to missed work days, reduced productivity, and in some cases, even job loss. When workers are victims of violence, their ability to participate in working life is impaired, which can have a significant economic impact. Violence against girls and women can also prevent them from continuing their education and developing their skills. This not only limits their economic opportunities but also hinders the development of a country's human capital, which is critical to economic growth. Addressing GBV is thus essential to promoting a more equitable and economically productive society.

Against this background, it is important to investigate a) drivers of GBV and b) how to prevent it. In the first Chapter of this dissertation, I document unintended negative consequences of a minimum wage policy for IPV. Chapters 2 and 3 focus on how to reduce the incidence of GBV by exploring the role of law enforcement officers' actions. All three Chapters of this dissertation are self-contained and can be read independently. Each Chapter has its own appendix and all appendices are included after Chapter 3. The bibliography contains references of all Chapters and can be found at the end of this dissertation. The following paragraphs provide a summary of the three Chapters.

*Divided Impact: How Minimum Wage Policy Shapes Crime Patterns and Intimate Partner Violence.* Chapter 1 examines the causal effect of introducing a minimum wage on IPV in Germany. Since minimum wages impact labor income and employment, they may also affect non-labor-market behaviors such as crime. Importantly, low wages can be distributed very differently among social groups and economic sectors. In Germany, for example, women often work in low-paid occupations and industries and are significantly more likely to be part-time or marginal employees than men. Women's position and success in the labor market affects their bargaining power in the household: e.g., loss of income may make them (more likely to be) victims of intra-household violence.

I combine data from several sources on crime outcomes, employment, and wages to assemble a monthly labor market area panel over the period 2011-2015. I exploit the distance between a labor market area's wage level in 2014 and the uniform federal minimum wage in 2015 (the bite of the minimum wage) to identify the minimum wage's causal effect. Since the uniform minimum wage policy does not account for individual regions' economic conditions, this results in plausibly exogenous variation in the bite. The intensity of treatment will be higher in labor market areas with lower wage levels in 2014. I employ a generalized difference-in-differences approach. In particular, I compare the extent of crime in high- and low- bite labor market areas after controlling for labor market area fixed effects and potential sources of heterogeneity across seasons of the year.

My main finding is that the German federal minimum wage of 2015 increased the number of IPV cases. This effect cannot be attributed to a general increase in crime. Non-IPV crime declined over the same period. To assess the factors that might explain the increase in IPV, I draw on data on income and employment. I refer to the bargaining model specified by Aizer (2010) in which (some) men have preferences for violence, and partners negotiate over the extent of abuse and the distribution of household consumption. I establish that adverse labor market effects disproportionately affect marginally employed women. This results in a loss of bargaining power for women and more females being at risk of IPV.

The results presented in this paper point to unintended consequences of minimum wage policy. An interesting task for future research is to investigate long-term impacts. The conclusions of this study can also be applied to a country that is beginning to enforce a minimum wage that has already been officially implemented.

*Deterrence or Backlash? Arrests and the Dynamics of Domestic Violence.* Chapter 2, coauthored with Sofia Amaral, Timo Hener, and Helmut Rainer, asks how arrest affects the dynamics of Domestic Violence (DV). A key aspect of DV is that it is seldom a one-time occurrence, with women frequently experiencing repeat abuse by the same partner (Tjaden and Thoennes, 2000; Aizer and Dal Bo, 2009). Despite the prevalence and seriousness of DV, the question of how best to police this crime so as to break the cycle of DV is still largely unresolved. A highly controversial police response is to arrest suspects on the spot.

We use emergency call data for West Midlands, the second most populous county in England, and merge these records with data on whether a criminal investigation is opened by an investigative officer and, if so, whether offenders are charged with a crime. To create a linked panel of DV incidents over time, we exploit information on the precise geo-location of where the incident occurred. This takes advantage of the fact that most DV occurs at home and that most police interventions originate via a 999 emergency call (HM Inspectorate of Constabulary, 2014). The key benefit of this approach is that it allows us to track repeat DV even if there is not a formal criminal charge filed.

One challenge in the analysis of the effects of an arrest is that there are likely to be characteristics of a case that are observable to police officers, but not the researcher. For example, if cases which result in arrest are (unobservably) more serious and hence more likely to be positively related to the underlying risk of repeat violence, OLS will underestimate any potential deterrent effect or possibly yield a positive estimate even if there is no backlash. To address endogeneity, we exploit the conditional random assignment of police officers to 999 DV calls combined with heterogeneity in officers' propensity to arrest. Since it is difficult to precisely predict when and where demands for police resources will emerge, the availability of patrol officers who can be dispatched to a DV incident is as good as random after conditioning on time, geography, and the priority level assigned by the call handler. Since some officers are more likely to arrest than others, the average arrest propensity in other cases can be used as an instrument for arrest in the current case. As we show, the instrument is highly predictive of arrest in the current case, but uncorrelated with observable case characteristics.

Our main finding is that arrest significantly reduces the probability of a repeat DV emergency call within the ensuing 12 months. In sharp contrast, OLS finds a precisely estimated zero effect. The implication is that not accounting for selection bias would lead one to erroneously conclude that arrest has no effect on DV call trajectories. These findings argue against recent calls for a complete decriminalization of domestic violence. In our setting where the arrest rate is low, the optimal policy response is to arrest more suspected batterers if the objective is to reduce future abuse. We caution, however, that our results do not necessarily imply that arrest should occur in all cases and in all settings; for example, in countries where the arrest rate is high, the pendulum could well have swung too far in the other direction. Future research for other countries and in other contexts can help shed light on this issue.

*Confronting Prejudice: Uncovering Stereotypes Among Police Officers in India.* Chapter 3, coauthored with Sofia Amaral, Kim Chaney, Nishith Prakash, and Abhilasha Sahay, investigates the effects of prejudice confrontation on police officers' handling of GBV cases in India. Bias among police officers can deter people from reporting incidents of GBV (García-Moreno, Zimmerman, Morris-Gehring, Heise, Amin, Abrahams, Montoya, Bhate-Deosthali, Kilonzo, and Watts, 2015), and when GBV incidents are reported, they may be dismissed because police officers view such victims as dishonest, discriminate against them, or blame them for the incident. Addressing GBV police bias is, therefore, a necessary condition to ensure equal access to justice and equal participation of women and girls in society.

We cooperated with the police in Madhya Pradesh, a state in central India, and conducted a lab-in-the-field experiment. The participants reviewed two cases, including one that involved a woman reporting GBV, and one non-GBV case, and answered a computer-based survey on how they would handle the cases. We randomly assigned officers to a treatment condition where a high-ranking police officer confronted them with their bias in dealing with the GBV case after the survey was completed. Officers in the control condition received neutral feedback on the non-GBV case. One week later, all officers were asked to review a GBV case again. They additionally solved a computerized 'stereotyping reaction time task' in which pictures of potential victims were shown and the officers were asked to rapidly categorize descriptions that might apply to the victim or not, and which allows researcher to measure implicit bias. This design allows us to understand future responses to GBV crimes by police officers after a confrontation with prejudice in two dimensions. First, how do confronted officers change their behavior in handling GBV cases? Second, to what extent does the victim's appearance cause confronted officers to react in a biased manner?

Our main finding is that - while there is no statistically significant aggregate effect of the confrontation treatment - the confrontation treatment does indeed lead female and male officers to react differently: Female officers respond less stereotypically in handling a GBV case, which means they place greater emphasis on the victim's account. Men tend to react in the opposite way and side with the offender. A potential explanation for our findings is that females predominantly show a mild bias in handling a GBV case before treatment, while more than half of the males are strongly biased. Considering that policing is highly male-dominated, the average female officer perceives a bias in her work environment that is stronger than her own. The confrontation treatment thus de-biases female officers and makes them adjust their beliefs in a professional context. For male confronted officers, prejudice confrontation seems to trigger a backlash that is driven by strongly biased men. In the computerized negative stimuli reaction task, we find significantly greater use of GBV stereotypes by men after a confrontation, while there is no effect for female officers. This finding is consistent with our results concerning the GBV case handling. While men seem to push back after a confrontation, women do not change their inner beliefs.

Taken together, the intervention encouraged women to de-bias their actions, which means less stereotyping when working on GBV cases. Men, on the other hand, are less receptive to the confrontation feedback. Therefore, the task of future research is to carefully design interventions for strongly biased men in a country where norms do not indicate gender egalitarianism. On a cautionary note, our findings do not necessarily allow us to draw general conclusions about settings in other countries. Research has highlighted that the response to confrontation is influenced by different historical, social, and cultural factors.

# 1

# Divided Impact: How Minimum Wage Policy Shapes Crime Patterns and Intimate Partner Violence*

Abstract

This paper examines the causal effect of a minimum wage introduction on Intimate Partner Violence (IPV). I exploit regional variation in the bite of Germany's 2015 uniform federal minimum wage and show that moving from the 25th to the 75th quartile of the bite distribution leads to an 8% increase in IPV. Importantly, the effect cannot be attributed to a general increase in crime. Non-IPV crime declined by around 2% over the same period. Exploring possible mechanisms for the IPV increase, I establish that marginally employed women were exposed to the most adverse labor market effects, reducing their bargaining power within the household and exposing them to a greater risk of IPV.

*This paper has greatly benefited from helpful comments from Eleonora Guarnieri, Michael Kaiser, Helmut Rainer, and Fabian Siuda. All errors and omissions are my own.

## 1.1 Introduction

Many countries have a statutory minimum wage. Higher wage floors have been shown to lead to a significant increase in expected wages for low-wage workers and a small but statistically significant decrease in employment (Caliendo, Wittbrodt, and Schröder 2019; Campolieti, Fang, and Gunderson 2005; Neumark and Adams 2003). To the extent that minimum wages impact labor income and employment, they may also affect non-labor-market behaviors such as crime. Importantly, low wages can be distributed very differently among social groups and economic sectors. In Germany, for example, women often work in low-paid occupations and industries and are significantly more likely to be part-time or marginal employees than men.[1] Since changes in the labor market for women could affect their bargaining power in the household, loss of income may make them (more likely to be) victims of intra-household violence.

In this paper, I examine the causal effect of minimum wage policy on Intimate Partner Violence (IPV) and uncover opposing signs for the effect on IPV and violent crime overall.[2] To assess this relationship, I investigate the introduction of the nationwide minimum wage in Germany in 2015. When it was introduced, 15% of the labor force was affected, although there were large regional variations (Dustmann, Lindner, Schönberg, Umkehrer, and Vom Berge, 2022). I combine data from several sources to assemble a monthly labor market area panel over the period 2011-2015. The primary outcome variables, the number of IPV offenses and the number of violent crime cases are derived from extensive registry data at the case level provided by the Federal Criminal Police Office.

I exploit the distance between a labor market area's wage level in 2014 and the uniform federal minimum wage in 2015 (i.e. I refer to this distance as ”bite”) to identify the minimum wage's causal effect. Since the uniform minimum wage policy does not account for regions' economic conditions, this results in plausibly exogenous variation in the bite. The intensity of treatment will be higher in labor

[1] Source: `https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-2/niedriglohnquote.html`. Marginal Employment in Germany is defined as employment in which the regular pay does not exceed a maximum amount defined by law (in 2015: €450). In Germany, this is also referred to as a minijob. Under German law, employees in marginal employment are not subject to compulsory health, long-term care, and unemployment insurance in this employment (Bundesministerium für Wirtschaft, 2022).

[2] Violent crime includes the universe of crime cases with all victims that were offended against their legally protected personal rights. Not included are state protection crimes, traffic offenses, administrative offenses, offenses that do not fall within the remit of the police (e.g. financial and tax offenses), and offenses that are reported directly to the public prosecutor's office.

market areas with lower wage levels in 2014. I employ a generalized difference-in-differences approach. In particular, I compare the extent of crime in high- and low- bite labor market areas after controlling for labor market area fixed effects and potential sources of heterogeneity across seasons of the year.

My main finding is that moving from the labor market area with the least to the one with the most bite causes an increase in IPV offenses by 28%. This effect cannot be attributed to a general increase in crime. Non-IPV crime declined by 7% over the same period. My main specification assumes a linear effect of bite on crime. I relax this assumption by adding a specification in which treatment is measured using quartile indicators for the minimum wage bite in 2014. The results show that the effects found are borne almost equally by labor market areas in the third and fourth quartiles. Since the minimum wage has less effect in areas in the first and second quartiles, the result is consistent with economic intuition. A largely non-binding wage floor is unlikely to affect crime. The fact that the third and fourth quartile impacts are comparable suggests that the average impacts are not entirely due to differences between the most extreme quartiles.

To address potential threats to identification, I control for migration across labor market areas. The results show that selective migration is not a threat to the empirical design. Another concern could be that the surge in immigration in 2015 confounds the results of this paper. However, the effects on crime outcomes remain almost unchanged in terms of size and significance when controlling for the number of asylum requests. To test whether the observed positive effect for IPV follows from a change in victims' reporting behavior after the introduction of the minimum wage, I compare the effects for reported actual IPV offenses and reported attempts of IPV. The argument is that if the reporting threshold for IPV went down after the minimum wage introduction, less serious cases, such as attempted offenses, are increasingly reported. However, there is no significant effect for IPV attempts, suggesting that the increase in IPV cases is not due to a change in reporting behavior, but rather an increase in abuse.

I next turn to mechanisms. For the context of IPV, it is important to examine gender differences in labor market consequences of minimum wage policy. Aizer (2010) investigates the link between women's relative economic status and their vulnerability to domestic violence. The study proposes a model in which (some) men have preferences for violence, and partners negotiate over the extent of abuse and household consumption distribution. The main prediction of Aizer's model is that a decrease in a woman's relative wage worsens her bargaining power in the relationship, leading to an increase in violence as her outside options deteriorate and vice versa. In this paper, I find a negative effect of the minimum wage introduction on marginal employment. Compared to men, women are particularly affected in absolute terms as they accounted for the vast majority of these jobs in 2014. This finding is consistent with a loss of bargaining power for women and results in more women being at risk of IPV.

Taken together, the results presented in this paper point to unintended consequences of the minimum wage policy. Policymakers may have to consider that (enforcing) minimum wage policy can have adverse effects on IPV due to gender-specific labor market consequences. Many countries have introduced an official minimum wage, however, not all enforce them. The conclusions of this study can also be applied to a country that is starting to enforce a minimum wage that has been officially in place already.

The literature on the minimum wage effects on crime is sparse.[3] Fone, Sabia, and Cesur (2023) examine minimum wage increases in the US, using data from the Uniform Crime Reports, and find a positive relationship between minimum wage hikes and property crime arrests among 16-to-24-year-olds, while not observing any evidence that an increase in the minimum wage impacts violent crime arrests. The authors claim that job losses explain their findings as the main mechanism. Using data on annual crime rates for large cities in the United States, Fernandez, Holman, and Pepper (2014) find that living-wage ordinances are associated with reductions in property-related crime. Beauchamp and Chan (2014) focus on low-wage workers, using individual-level panel data over a similar period, and find that a rise in the minimum wage increased property and violent crimes among teenagers, but decreased crime for young adults. Agan and Makowsky (2018) examine whether the minimum wage is associated with adult recidivism among those who were previously incarcerated. The authors find that minimum wage increases go along with a decline in recidivism for property- and drug crime-related offenses. Finally, Hansen and Machin (2002) explore the introduction of a new national minimum wage law in the United Kingdom and conclude that jurisdictions with larger shares of low-wage workers saw crime reductions in comparison to areas where the minimum wage was less binding.

I extend this literature by establishing that a minimum wage introduction causes

[3]Most studies exploit variation in minimum wages over time and across different jurisdictions. The most common identification strategy is a difference-in-differences approach, with exceptions. Hashimoto (1987), for example, uses time series regressions and finds a positive relationship between a Federal minimum wage and property crimes for teenagers (ages 15-to-19).

IPV to increase while no such effect is observable for overall crime.[4] Spencer, Livingston, Woods-Jaeger, Rentmeester, Sroczynski, and Komro (2020) look at self-reported non-physical IPV in the US context and find no evidence for the minimum wage to have a differential effect on it. Further, this study explores the connection between the introduction of a minimum wage policy and crime in the context of Germany. Profound regional differences in wage levels and crime make Germany an ideal setting for this study. The maps in Figure 1.2 demonstrate this variation. Panels a) and b) plot the variation in overall violent crime and IPV rates and Panel c) shows significant variation in average wages across labor market areas in 2014.

The remainder of this chapter is structured as follows. The next section provides information on the minimum wage policy and IPV in Germany. Section 1.3 explains the data and the variables. Section 1.4 contains a description of the empirical approach. Section 1.5 reports results, validity checks, robustness tests, and a discussion on potential channels. Section 1.6 concludes.

## 1.2 Context

### 1.2.1 Minimum Wage Policy in Germany

Germany introduced a federal minimum wage of €8.50 on January 1, 2015. Alongside the Hartz reforms, which came into force in 2002, the introduction of the minimum wage is considered the most significant labor market reform in Germany in the last three decades and affected 15% of the labor force (Dustmann, Lindner, Schönberg, Umkehrer, and Vom Berge, 2022). Previously, some sectors (e.g. construction and electric trade) had already introduced specific minimum wages. Fitzenberger and Doerr (2016) provide an extensive overview of sector-specific minimum wages. The federal minimum wage is mandatory, with only a few exceptions. Excluded are persons under 18 years of age, trainees, specific interns, inmates, long-term unemployed in the first six months after re-entry into the labor market, and volunteers (Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt, 2018). Despite these exceptions, the minimum wage bound for roughly 4 million out of the 5.5 million employees who received a gross hourly wage less than €8.50 in 2014. A nationwide binding minimum wage was proposed after the German parliamentary

[4] I am aware of two master theses and drafts that are available online and somewhat related to this paper. Carlos Chavez looks at informal employment changes and domestic violence in connection with a minimum wage policy in Peru. Given that the draft is not entirely translated to English, my take is that this is work at an early stage (link to study). Ankita Banerjea explores a statutory wage for domestic workers and domestic violence in India (link to study).

elections in fall 2013. The new government under Chancellor Merkel endorsed a nationwide minimum wage in its coalition agreement. The government's proposal was passed by both houses of parliament in July 2014 and came into force in January 2015. The body responsible for the minimum wage is the Minimum Wage Commission, which deliberates at regular intervals on adjustments regarding its level. In the meantime, the minimum wage has been raised several times. In 2017, it was raised to €8.84, in 2019 to €9.19, in 2020 to €9.35, in 2021 to €9.60, and is at €12 since 2022.

### 1.2.2 Intimate Partner Violence in Germany

Police crime statistics (PCS) by the Federal Criminal Police Office are one of the most important data sources for describing and analyzing crime situations in Germany.[5] While there is no official crime category for IPV, since 2011, the PCS of the Federal Criminal Police Office has mapped the victim-suspect relationship in terms of relationship types (spouse, registered civil partnership, partners in non-marital partnerships, former partnerships). This is crucial to the present evaluation of IPV (BKA, 2022). From 2011 (120,878) to 2014 (130,726) I observe an 8% increase in cases of IPV in Germany. A comparison of the data used in this paper with official reports from the Federal Criminal Police Office is difficult as they continuously adjust the definition of IPV, including more and more categories in the field of non-physical violence such as threatening or stalking, while I am considering all crimes that happen between intimate partners- physical or non-physical. The criminal statistics analysis of data on IPV shows that violence in existing and former partnerships in Germany has become more and more significant also in recent years. An examination of the recorded number of victims over the last years shows an overall increase of 3.4% (2017: 138,893; 2018: 140,755; 2019: 141,792; 2020: 148,031; 2021: 143,604). The increased number of victims of IPV is mainly due to the continuous increase in the number of victims of intentional simple assault in recent years. IPV offenses are mainly directed against women (80.3%) (BKA, 2022).

Criminology (dark-field) research on violence provides important insights into the actual extent, prevalence, and relevant risk factors of violence and the conditions under which it occurs. While police crime statistics only register the bright field, i.e. all cases known to the law enforcement agencies, victim surveys representative of the population also allow for the consideration of experiences of violence in the

[5] Not included are state protection crimes, traffic offenses, administrative offenses, offenses that do not fall within the remit of the police (e.g. financial and tax offenses), and offenses that are reported directly to the public prosecutor's office.

dark field (European Union Agency for Fundamental Rights, 2014). Further data on women's exposure to violence is also provided by the representative study 'Living Situation, Safety and Health of Women in Germany' published in 2004 (BMFSFJ, 2014). According to this study, around 25% of women between the ages of 16 and 85 have experienced physical and/or sexual violence at least once in their lives.

## 1.3 Data Sources and Main Variables

The data set used for the analysis covers the time window from January 2011 to September 2015 and contains 401 counties in Germany for which crime statistics were uniformly reported to the German Federal Criminal Police Office. Counties are the smallest geographic unit in Germany for which wage data is publicly available. As the job market regions in Germany are larger than the area within a county (e.g. Bonin, Isphording, Krause-Pilatus, Lichter, Pestel, Rinne, et al. 2018), I aggregate them to 223 labor market areas. This forms the level at which the main analysis of this paper is conducted.[6] I combine several data sources to examine the impact of a minimum wage introduction on violent behavior- particularly in the family context.

### 1.3.1 Crime

The crime data is derived from the German Police Crime Statistics (PCS), provided by the Federal Criminal Police Office. It includes the universe of crime cases with all victims who were offended against their legally protected personal rights between 2011 and 2015. Not included are state protection crimes, traffic offenses, administrative offenses, offenses that do not fall within the remit of the police (e.g. financial and tax offenses), and offenses that are reported directly to the public prosecutor's office. In addition to the time and location (municipality level) of the crime, the data also includes the code for the type of crime, age and gender of the victim, information on some characteristics of the offense (attempted/completed, use of a firearm, lone perpetrator/act committed by a group), and information on the relationship between victim and suspect, which is crucial to analyzing IPV as there is no official crime category for it. However, as the data is not reported until after police procedures are completed, only cases from January 2011 to September 2015 are used for analysis to avoid problems with lags between the occurrence of the

[6] The definition of the Federal Office for Building and Regional Planning is used for the aggregation to labor market areas (German: Bundesamt für Bauwesen und Raumordnung (BBSR)). A robustness check on the county level is conducted in section 1.5.4.

crime and the time of reporting at the end of the observational period.[7] In Section 1.5 I conduct robustness tests that show that my results are not sensitive to different cut-offs. The sample used for analysis contains 3,909,225 reported offenses. Roughly 41% of victims are female, the average age is 34 years, and in 15% (597,533) of the cases the victim is or was in an intimate relationship with the offender. There are 7 categories of offenses for which Table 1.A.1 indicates the proportion of offenses among (ex-)partners. Intimate partner violence is defined in this paper as any crime being reported where the victim is or was in an intimate relationship with the offender. Roughly 83% of all IPV victims are female, about 63% of the victims are in a current intimate relationship with the offender (as opposed to 37% who are offended by former partners) and the majority of the cases are categorized as assault (about 71%). For an extensive overview see Figure 1.1. The level of overall crime and IPV varies significantly across labor market areas in Germany. Figure 1.2 shows in Panel a) the variation in the overall violent crime rate that ranges from 45 (per 10,000 inhabitants) per month in Odenwaldkreis to 198 in Neumünster and in Panel b) the variation in IPV rate that spans from 5 in Neumarkt i.d. Oberpfalz to 32 in Neumünster.

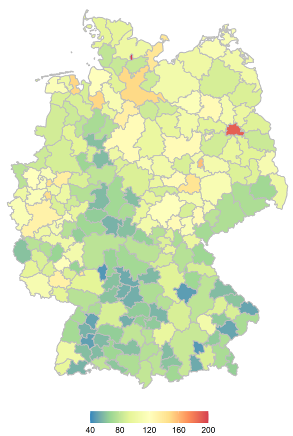
[7] 80% of the cases are reported within 3 months. Therefore, the last 3 months of the data set are excluded.


*Figure 1.1:* Analysis Sample: Intimate Partner Violence Cases

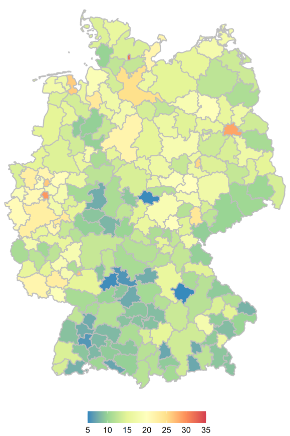
### 1.3.2 Minimum Wage Bite

The effect of the minimum wage depends on the level of the pre-reform average wage. A lower wage causes the minimum wage to have a stronger effect, and vice versa. If only a few people are directly affected by the new wage floor, potential consequences are unlikely to be economically significant.

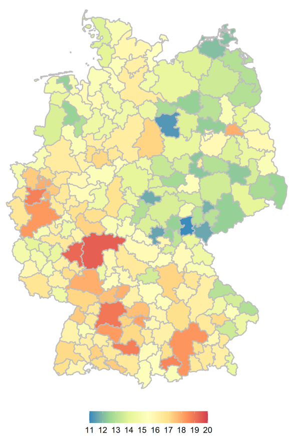
**Hourly Wages.** Data on the hours worked are a necessary part of the calculation of the hourly wage of the monthly earnings. This data is retrieved from the Earnings Structure Survey 2014 (German: Verdienststrukturerhebung, VSE), which is conducted by the Federal Statistical Office and is representative at the state level (Statistisches Bundesamt, 2014). The data provide the average weekly hours worked of full-time employees in April 2014 for all states. On average, employees work between 39 and 40 hours per week. At the upper end, 40.32 hours are worked in Saxony-Anhalt, while employees in Saarland work only 39.23 hours. I




*(a)* Geographical Variation in Crime Rates




*(b)* Geographical Variation in IPV Rates




*(c)* Implied Hourly Wages in 2014

Notes: Panel (a) shows the geographic variation in crime rates across German labor market areas. The rate is computed as the number of crimes per 10,000 inhabitants per month. Darker shades of blue indicate lower values (min =45) and lighter shades indicate higher values (max = 198). Panel (b) shows the geographic variation in Intimate Partner Violence (IPV) rates across German labor market areas (min=5, max=32). The rate is computed as the number of IPV cases per 10,000 inhabitants. Panel (c) plots the 2014 hourly median wage. It ranges from darker shades of blue indicated areas (min = €12.36) to red (max = €23.95).

*Figure 1.2:* Geographical Variation in Wages and Crime Rates

employ earnings data from the 'INKAR' database which is maintained by the Federal Ministry of the Interior, Construction and Homeland, and is constructed from annual reports from the various states (BBSR Bonn, 2019). It provides county-level data on median earnings that I combine with the hours worked from the VSE to construct counties' implied average hourly wages. Median earnings (German: Monatsentgelte der Vollzeitbeschäftigten, English: monthly labor earnings for full-time employees) measures the median level of gross labor earnings for full-time employees reported to the Federal Agency for Employment (German: Bundesagentur für Arbeit). The county hourly wages are then collapsed (using population weights) to the average labor market area hourly wages. Average hourly wages based on median income tend to be highest in the area of Dingolfing-Landau (€23.95; i.e. the location of BMW's largest factory in Europe) while they are lowest in the labor market area Vorpommern-Rügen(€12.36; i.e. an island in the north-east). In Section 1.5.4, I present a robustness check based on an alternative measure of hours worked, namely the 'annual volume of work' published by the INKAR database. The annual work volume (German: geleistete Arbeitsstunden (Jahreswert) der Arbeitnehmer, English: hours worked (annual value) by employees) comprises the actual hours worked by all persons in employment who pursue an economic activity as employees or self-employed persons. The good thing about this statistic is the fact that, like the earnings data, it varies at the county level. The disadvantage is that it includes not only full-time employees, implying lower average hours worked (on average individuals work between 27 and 28 hours a week) and therefore higher average hourly wages and a lower level of minimum wage bite. Against this background, the robustness test shows that the results imply the same pattern but yield smaller point estimates.

**Kaitz Ratio.** The main explanatory variable in the analysis is the bite of the minimum wage, measured by the so-called Kaitz ratio.[8] In the Kaitz index, the minimum wage is expressed as a share of the respective median wage. A value of 50% thus means that the minimum wage is half the median wage and larger values imply stronger effects of the minimum wage. The Kaitz ratio is given by $\frac{8.50}{w_{i,2014}}$. Note that the non-linearity of the transformation implies an increasing marginal bite as wages fall. Figure 1.A.1 shows a histogram of observed Kaitz ratios in the data. The range is about 0.35 to 0.69 with a mean (median) of 0.49 (0.47) and a standard deviation of 0.08. A necessary condition for the Kaitz ratio to be a valid measure of the minimum wage bite is that it (rank) correlates with the share of individuals who earned less than €8.50 in 2014. If wage dispersion is greater in counties with higher average wages, the Kaitz ratio may be low, indicating little bite, when in fact there are significant numbers earning below the minimum wage. While I cannot empirically test this assumption, a comparison of states with low implicit bites with those that Bellmann, Bossler, Gerner, and Hübler (2015) find to have a lower share of firms affected by the minimum wage suggests that they are correlated.

### 1.3.3 Time-Varying Controls

Data on population, refugees, and migration is derived from the INKAR database which is maintained by the Federal Ministry of the Interior, Construction and Homeland, and is constructed from annual reports from the various states (BBSR Bonn, 2019). It is provided at the county level and is aggregated to labor market areas for

[8] An alternative measure of the minimum wage would be the proportion of employees in a county earning less than €8.50 in 2014. However, this measure also requires additional assumptions, as it does not indicate the gap between the wage of a worker's wage in 2014 and the minimum wage. The proportion would be equally affected for workers earning €8.40 or €5. Caliendo, Fedorets, Preuss, Schröder, and Wittbrodt (2018) find comparable employment effects when they use both measures of the minimum wage in their analysis of the 2015 minimum wage introduction in Germany. A third concept would be the wage gap (German: Lohnlücke) that refers to the average absolute difference in hourly wages from the statutory minimum wage of € 8.50 for hourly wages below the minimum wage, with the difference taking the value zero for wages of at least € 8.50. The concept of the wage gap is thus based on the population of all employees and quantifies the minimum wage relevance in terms of both the number of employees and the level of the minimum wage to the starting level of wages (Bonin, Isphording, Krause-Pilatus, Lichter, Pestel, Rinne, et al., 2018). However, for both measures individual earning data would be needed that in turn could not be matched with the anonymized crime data. Further, it is established in the literature that in the IPV context, it is not the actual wage that determines one's outside option but the potential wage. This is because the woman's bargaining power is determined by her earnings at her so-called threat point and not by the earnings at the bargaining equilibrium, where she does not necessarily earn the same as at the threat point. According to Pollak (2005), a married woman who is not employed (zero wages) at the cooperative equilibrium but who would work in the event of the dissolution of the marriage, can also have bargaining power. Implementing the average bite in the same labor market area would be consistent with this argument.

the main analysis. I use employment data published by the Federal Employment Agency (German: Bundesagentur für Arbeit).

## 1.4 Empirical Approach

In order to identify the causal effect of minimum wage policy on criminal behavior, I exploit within-labor market area variation over time. To be precise, I compare crimes in a given month in areas that are more affected (i.e. with a higher minimum wage bite) to areas that are less affected, conditional on the season of the year. Since the crime data is count data, I use a Poisson pseudomaximum likelihood (PPML) model with multi-way factors fixed effects to estimate the impact of the introduction of a minimum wage on violent crime:

$$E[Crime_{a,t}|Bite_a, \alpha_a, \delta_t] = exp(\beta_1[t > 2014]_t + \beta_2[t > 2014]_t * Bite_a + \alpha_a + \delta_s + \epsilon_{a,t}), \tag{1.1}$$

where $Crime_{a,t}$ is the number of violent crime cases in labor market area a, in month t. The time-invariant variable $Bite_a$ is the measure of treatment intensity, i.e., a measure of minimum wage bite in labor market area a. It is based on the Kaitz ratio, which is calculated from the average hourly wage level in 2014 and the nominal minimum wage of €8.50. The coefficient of interest, $\beta_2$, multiplies the interaction between the post-reform indicator and the continuous measure of the minimum wage bite. Population size is used as an exposure variable to fit the Poisson model to the different sizes of the labor market areas studied (see e.g. Andres, Fabel, and Rainer 2023; Lindo, Schaller, and Hansen 2018). This specification assumes a linear effect of the minimum wage bite on crime. I relax this assumption by implementing a specification in which treatment is measured by quartile indicators of the Kaitz-ratio in 2014. Thus, three interaction terms of indicators for quartiles 2, 3, and 4 with the post-reform dummy are estimated. The first quartile, i.e. the 25% of labor market areas where the minimum wage bite is lowest, being the omitted group. All specifications include labor market area and season of the year fixed effects. For an extended specification, I expand the baseline model by adding in area-by-season fixed effects that account for systemic changes in the degree of violent behavior over the year for each region. The standard errors are clustered at the labor market area level, allowing arbitrary auto-correlation in the error terms within an area (Abadie, Athey, Imbens, and Wooldridge, 2020). In a robustness check, I relax the assumption of the Poisson regression model that the variance is equal to the mean of the outcome and verify that my results are almost identical when estimating a conditional fixed effects negative binomial model. I implement Equation 1.1 using the ppmlhdfe command in Stata (Correia, Guimarães, and Zylkin, 2020). The percentage effects of the coefficient of interest can be calculated as $(e^{\beta} - 1) * 100$ (Halvorsen and Palmquist, 1980).

The identification of the causal effect in this setting requires the existence of a valid counterfactual scenario with parallel trends. In the present context, this means that crime rates would have to trend the same in all labor market areas, regardless of the 2014 wage level. This premise would be violated if, for example, in areas where the minimum wage threshold was high crime rates were already changing differently before 2015. However, a close examination of the trends in crime rates suggests that this is not the case and that the causal effect of the minimum wage reform on crime can be determined. Crucial to this argument is the fact that it is a uniform minimum wage of €8.50 for the whole of Germany. This precludes the concern that the level of the minimum wage is endogenously determined by the economic characteristics of specific areas. Therefore, the relationship between a labor market area's wage in 2014 and the minimum wage is plausibly exogenous. Figure 1.A.2 assesses the plausibility of trends in crime rates independent of 2014 wage levels.

## 1.5 Results

This section first reports the results from the preferred specifications. In the next step, I address threats to identification. A discussion of margins of heterogeneity, robustness checks, and mechanisms conclude the section.

### 1.5.1 Benchmark Results

Table 1.1 reports estimates corresponding to Equation 1.1. The coefficient of interest multiplies the interaction term reported in the first row of the Table. The causal effect of the minimum wage on overall crime is small and imprecise (Columns 1 and 2). However, if offenses among intimate partners are excluded, there is a negative and precise effect on criminal behavior (Columns 3 and 4).

The largest effect is found to be positive and highly significant for IPV (Columns 5 and 6). In Columns (1), (3), and (5) only fixed effects for labor market areas and the season of the year are included, whereas Columns (2), (4) and (6) implement labor market area by season of the year fixed effects. The difference in size and precision between the estimates in the two specifications is negligible across all outcomes.

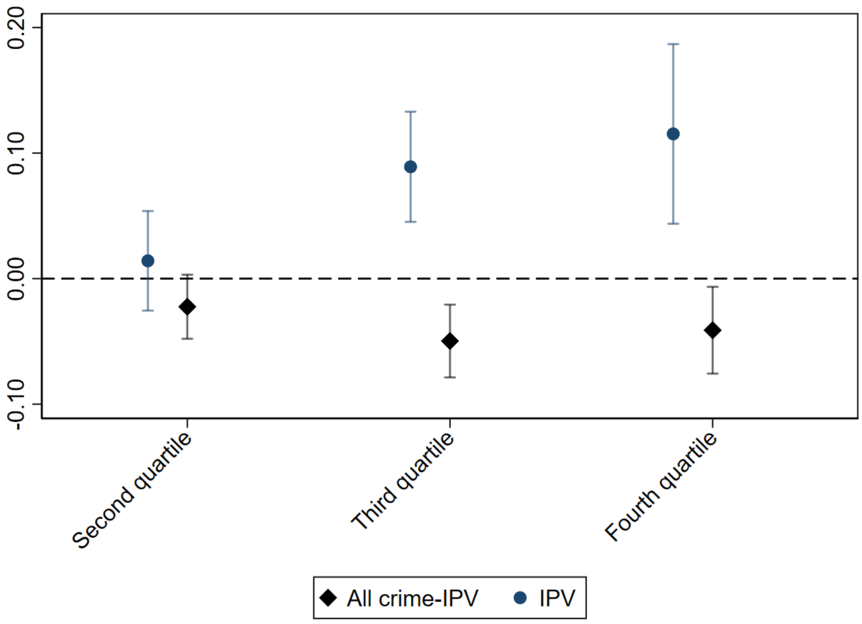
The preferred specification for non-IPV crime and IPV is presented in Columns (3) and (5), respectively. It implies a 19.72% decrease in non-IPV crime and an 85.41% increase in IPV for a one point difference in the Kaitz ratio. Yet, there are no two labor market areas whose Kaitz ratios are one point apart; in fact, the maximum distance is 0.33. This distance implies an effect of minimum wage bite of -6.51% on non-IPV crime and +28.34% on IPV offenses. Moving from the 25th to the 75th percentile of the Kaitz-ratio's distribution results in a change of 0.09; this implies an effect equal to -1.77% for non-IPV and +7.73% for IPV.

*Table 1.1:* The Effect of Minimum Wage Policy on Crime

| | Crime | | Crime w.o IPV | | IPV | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| post x mw bite | -0.073 | -0.109 | -0.180** | -0.220*** | 0.620*** | 0.612*** |
| | (0.071) | (0.071) | (0.084) | (0.085) | (0.157) | (0.156) |
| *Effect size [in %]* | -7.57 | -11.52 | -19.72 | -24.61 | +85.89 | +84.41 |
| *25th-75th percentile effect [in %]* | -0.68 | -1.04 | -1.77 | -2.21 | +7.73 | +7.59 |
| post 2015 | -0.078** | -0.062* | -0.043 | -0.025 | -0.317*** | -0.314*** |
| | (0.034) | (0.034) | (0.039) | (0.040) | (0.071) | (0.070) |
| Labor market area FE | yes | no | yes | no | yes | no |
| Season FE | yes | no | yes | no | yes | no |
| Area by season FE | no | yes | no | yes | no | yes |
| Dep.var mean | 304 | 304 | 255 | 255 | 49 | 49 |
| N | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Columns 1 and 2); non-IPV crime, i.e. crime excluding intimate partner violence (Columns 3 and 4); and Intimate Partner Violence (IPV) offenses (Columns 5 and 6). 'post 2015' is an indicator for the year 2015. 'Post x mw bite' is an interaction term of this indicator and the Kaitz ratio at the labor market level. The Kaitz ratio is constructed from median income. The specification in Columns (1), (3), and (5) includes labor market area and season of the year fixed effects. In Columns (2), (4) and (6) they are replaced by labor market area-by-season fixed effects. Standard errors are clustered at the labor market area level and reported in parentheses.
*p<.10, **p<.05, ***p<.01

One potential concern is that the linear specification presented above is not able to capture the nature of the causal effect of minimum wages on criminal behavior. Figure 1.3 presents estimates that are based on grouping labor market areas into quartiles based on their 2014 minimum wage bite. The first quartile corresponds to the 25% of labor market areas with the highest implied average wage in 2014. These areas represent the base category, and all estimates should be interpreted relative to this group. Figure 1.3 suggests that the effect size is proportional to labor market areas' bite quartiles. The effect is borne almost equally by labor market areas in the third and fourth quartiles. Since the minimum wage has less effect in areas in the first and second quartiles, the result is consistent with economic intuition. A largely non-binding wage floor is unlikely to affect crime. The fact that the third and fourth quartile impacts are comparable suggests that the average impacts are not entirely due to differences between the extreme quartiles.




Notes: This Figure shows the estimates using a specification similar to the Equation 1.1, in which labor market areas are divided into quartiles of the minimum wage bite. The Figure thus shows the interaction term between the period after 2015 and the quartile indicators. The 25% of labor market areas with the highest average implied wages in 2014 represent the first quartile and are the base category.

*Figure 1.3:* Quartile Effects of Minimum Wage Bite

### 1.5.2 Potential Threats to Identification

In this section, I consider the possibility that the results do not reflect changes due to the minimum wage introduction, but are driven by simultaneous events and their effects.

**Selective migration.** Selective migration might confound my estimates. For example, employment-driven migration of low-wage workers from labor market areas with low wage levels to areas with higher levels could create a downward bias, if low-wage workers are more likely to commit crimes than high-wage workers. To mitigate such omitted-variables bias, I control for inward and outward migration. As migration data is only available at the county level, I compute three extreme scenarios: Columns (1), (4) and (7) of Table 1.A.3 assume that all immigrants come from outside the labor market area while all emigrates migrate to counties in the same labor market area. This scenario presents an extremely attractive labor market area, where nobody chooses to leave. A very unattractive labor market area is then sketched in Columns (2), (5), and (8), where the emigration rate is included as a control, assuming that all emigrates leave the labor market area and there are no immigrants. Finally, in Columns (3), (6), and (9) both, the immigration rate and the emigration rate are added as controls, drawing a scenario in which half of the migration happens within and half across labor market areas. The estimates of these 3 different scenarios for overall crime, non-IPV crime, and IPV are very comparable to those in the main specifications presented in Table 1.1. Selective migration is unlikely to be a threat to the validity of the empirical design.

**Refugees.** Previously, I have presented evidence that a minimum wage policy increases the likelihood of IPV. However, it is possible that the German refugee 'crisis' in 2015 affected labor market areas where the minimum wage was more binding in a more extreme way[9]. If more refugees went to lower-income areas and the probability of IPV was higher among the new residents, it could explain the increase in IPV. However, the geographic allocation of refugees was close to random. This hypothesis is tested in Column 1 of Table 1.A.2. The minimum wage bite has no statistically significant effect on the number of asylum requests in a labor market area. Since the estimate is not statistically significant but not equal to zero, the number of asylum requests is included as a control for the main regressions in Columns (2)-(4) as an additional test. The effect sizes of the crime outcomes remain almost unchanged in terms of size and significance.

**Reporting behavior for IPV.** Given that I draw on reported crime data, an important question is whether the observed positive effect for IPV is driven by an increase in incidence or merely an increase in reporting. Although there is no direct test to solve this interpretation problem, there is an intuitive approach to assess it. I compare the effect of the minimum wage policy on reported actual offenses

[9]The refugee crisis in Germany 2015/2016 is the term used to describe the situation for the state and society that arose in connection with the entry of more than one million refugees, migrants and other people seeking protection in Germany in 2015 and 2016. It is part of the Europe-wide refugee crisis and reached its peak in the fall of 2015. The year 2015 was particularly characterized by a high influx of people seeking protection. In 2015, 476,649 asylum applications (initial and subsequent applications) were registered (compared to 202,834 in 2014). This represents an increase of 135% compared to the previous year (Bundesamt für Migration und Flüchtlinge, 2016).

and reported attempts of IPV. The argument is that if the reporting threshold for IPV goes down after the minimum wage introduction, less serious cases, such as attempted offenses, are increasingly reported. However, this is not the case. Table 1.A.4 reports in Column (1) the main effect, which is split into completed IPV and IPV attempts in Columns (2) and (3), respectively. The estimate for IPV attempts is not statistically significant while the effect for completed IPV cases is very comparable in size and significance to the main effect. This suggests that the increase in IPV cases is not driven by a change in reporting behavior, but rather an increase in abuse.

### 1.5.3 Heterogeneity

This section addresses two aspects of effect heterogeneity. First, are there significant differences in labor market areas in the former GDR (East Germany)? Second, do the effects differ between urban and rural labor market areas?

**East vs West.** Decades after reunification, socioeconomic differences between West and East Germany persist. Those states that used to form the GDR (East Germany) tend to have higher levels of unemployment, lower wages and income. East Germany is geographically smaller and is comprised of 52 labor market areas (as opposed to 171 in the West). To determine whether there is heterogeneity in the effects of the minimum wage reform between West and East Germany, I estimate Equation 1.1 and introduce an interaction effect for labor market areas in East Germany. In this type of specification, the interaction term of the Kaitz-ratio with the post-2015 indicator measures the minimum wage policy effect for labor market areas in West Germany; the additional interaction with East Germany measures the difference in the effects between West and East Germany. Results are displayed in Panel (A) of Table 1.2. Column (2) shows that the negative effect on non-IPV crime (Table 1.1) is driven by the West, while there is actually a small but positive effect for East Germany. For IPV, however, there is no significant difference between East and West. Note that the minimum wage bite has a different distribution in West Germany than in East Germany and the 25th-75th percentile effect is comparable in size to the main estimate in Table 1.1 (full minimum wage bite range: 0.19, 25th-75th percentile: 0.05).

**Rural vs Urban.** Policymakers are concerned about the asymmetric development of rural and urban areas (Bauer, Rulff, and Tamminga, 2019). A larger effect on crime in rural labor market areas could contribute to a widening of the gap in economic performance between urban and rural areas. The INKAR database classifies counties along the urban-rural spectrum, measuring how many inhabitants live

*Table 1.2:* Geographical Variation in the Minimum Wage's Effects

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: West vs East Germany** | | | |
| post x mw bite | -0.169 | -0.376** | 0.996*** |
| | (0.137) | (0.150) | (0.240) |
| *Effect size for West [in %]* | -18.41 | -45.64 | +170.74 |
| *25th-75th percentile effect for West [in %]* | -0.92 | -2.28 | +8.54 |
| post x bite x east | 0.408** | 0.510*** | -0.079 |
| | (0.165) | (0.186) | (0.312) |
| post 2015 | -0.032 | 0.046 | -0.476*** |
| | (0.062) | (0.068) | (0.104) |
| **Panel B: Urban vs rural labor market areas** | | | |
| post x mw bite | -0.158** | -0.282*** | 0.602*** |
| | (0.075) | (0.086) | (0.225) |
| *Effect size for Urban [in %]* | -17.12 | -32.58 | +82.58 |
| *25th-75th percentile effect for Urban [in %]* | -0.86 | -1.63 | +4.12 |
| post x bite x rural | 0.265* | 0.346* | -0.144 |
| | (0.158) | (0.181) | (0.337) |
| post 2015 | -0.039 | 0.004 | -0.313*** |
| | (0.035) | (0.040) | (0.098) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| Dep.var mean West | 304 | 254 | 50 |
| Dep.var mean Urban | 483 | 405 | 79 |
| N | 12,711 | 12,711 | 12,711 |

Notes: This Table reports geographical variation in the effect of minimum wage bite on overall crime (Column 1), non-IPV crime (Column 2), and Intimate Partner Violence (IPV). In Panel A, the effect is broken by West vs. East Germany. Panel B analyzes the effect in rural vs urban labor market areas. 'Rural' is a dummy variable equal to 1 if the labor market area is classified as rural, and 0 if it is classified as urban. All specifications contain the same set of fixed effects and level of clustering as the main specification that can be found in Columns (1), (3), and (5) of Table 1.1.
*p<.10, **p<.05, ***p<.01

in areas with less than 150 residents per square kilometer. I aggregate the data to the labor market area and define 'rural' as being above the median in the distribution including areas with low population density. In the previous paragraph's spirit, I test for heterogeneous effects by interacting the minimum wage bite in Equation 1.1 with an indicator for rural labor market areas. Panel B of Table 1.2 presents the results. The negative effect for non-IPV crime shown in Table 1.1 is driven by urban areas while there is a small positive effect for rural areas. Again, for IPV, there is no significantly different effect of the minimum wage introduction in rural labor market areas. The effect for rather urban labor market areas is comparable to the effect from Table 1.1, while the interaction effect is small and imprecisely estimated.

### 1.5.4 Robustness

I perform several sensitivity tests to assess the robustness of the findings. Overall, the findings demonstrate that the main results are robust to alternative specifications and estimations, indicating that minimum wage policy does indeed lead to an increase in IPV and a small reduction in other violent crime.

**Alternative econometric specifications.** I begin with addressing potential overdispersion in the data. A poisson regression model assumes that the variance is equal to the mean of the outcome. However, the variance may be larger for greater values of the outcome and therefore larger than its mean. This leads to an underestimation of standard errors and false positive estimates. I test for a potential overdispersion in the data, plotting the standard deviation for the IPV quintiles in the year 2014 (Figure 1.A.3). The data shows a weakly positive relationship between the standard deviation and the number of IPV crimes, driven by outliers in metropolitan areas at the very right of the Figure. Table 1.A.5 reports regression results, excluding labor market areas in the top 5% of the IPV distribution in the sample. The effect of the minimum wage policy on IPV is smaller but still comparable in size and equally precisely estimated. Since the standard deviation and number of IPV cases continue to correlate even after excluding outliers, I also run negative binomial regressions. These models rely on less restrictive assumptions about the relationship between variance and mean. The results are presented in Table 1.A.6 and are virtually identical to my baseline estimates from Table 1.1. Finally, Table 1.A.7 demonstrates that the crime pattern is not sensitive to alterations in the *sample*, although the estimates for non-IPV crime are getting smaller and less precise the more the sample is restricted.

**County level analysis.** The baseline model is estimated using a labor market area panel. Labor market areas in turn consist of different counties. Each county is in exactly one labor market area. The reason for aggregating the data to labor market areas is the accurate measurement of the minimum wage bite. The wage floor might vary in neighboring counties, however, the labor market opportunities and effects of the minimum wage on employment decisions (that matter for criminal behavior) should be the same for counties that are, for example, surrounding a bigger city most people commute to for work. However, one could argue that when using aggregate instead of individual-level data the smallest geographic unit should be used. Crime data is available at the municipality level, while counties are the smallest geographic unit in Germany for which wage data is publicly available. Therefore, I conduct the same analysis described in Section 1.4 on the county level. Results are reported in Table 1.A.8. The patterns match the baseline model: A small negative effect -albeit not precisely estimated- for non-IPV crime and a positive and highly significant effect for IPV that is very comparable to the main estimate in Table 1.1 (25th to 75th percentile effect: 6.16%).

**Alternative measure for hours worked.** In the main analysis of the effect of minimum wage policy on crime, hourly wages are calculated from the median earnings from the INKAR database and hours worked from the Earnings Structure Survey 2014 (German: Verdienststrukturerhebung, VSE). While both, median earnings and hours worked consistently refer to full-time employees, hours worked are only available at the state level while median income varies at the county level. The robustness check is based on an alternative measure of hours worked that varies at the county level. This variable is called 'annual volume of work' and is published by the INKAR database. The Annual work volume (German: geleistete Arbeitsstunden (Jahreswert) der Arbeitnehmer, English: hours worked (annual value) by employees) measures the hours worked by all employed and self-employed individuals. In order to construct a measure of minimum wage bite, I then calculate hourly wages assuming four weeks per month. Monthly earnings are divided by weekly hours times four. The advantage of this approach is that, like the earnings data, hours worked vary at the county level. However, it also includes part-time workers, implying lower average hours worked (on average individuals work between 27 and 28 hours a week) and therefore higher average hourly wages and a lower level of minimum wage bite. The results imply the same pattern but yield smaller effect sizes that turn insignificant for non-IPV crime (see Table 1.A.9).

### 1.5.5 Mechanisms

The primary mechanisms hypothesized through which an introduction of a minimum wage may have impacted crime are the minimum wage's labor market effects. To empirically explore these effects, I draw on data on income and employment.

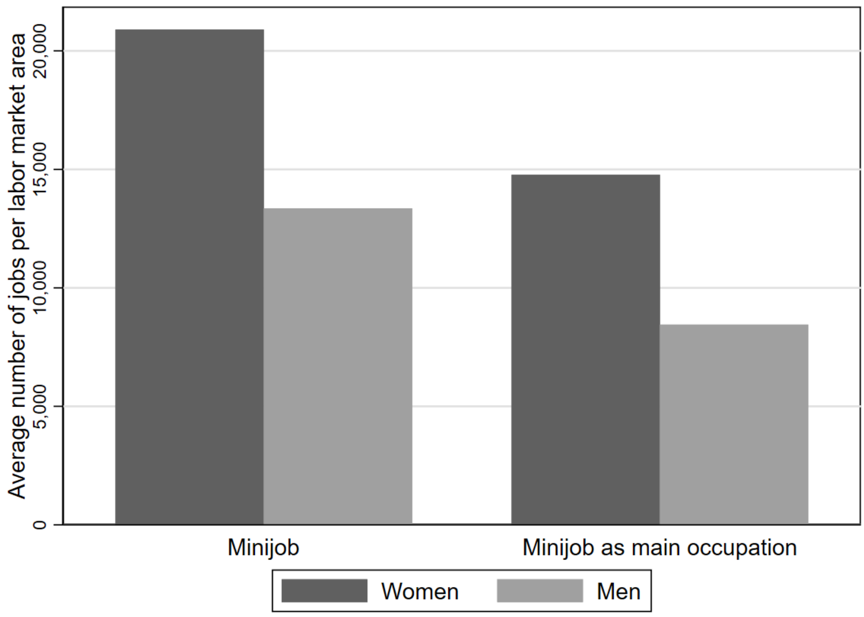
Research has shown that criminal behavior is negatively related to wages (Gould, Weinberg, and Mustard 2002; Yang 2017). The median income increased from €2,791 in 2014 to €2,851 in 2015 (2%) in Germany and may have contributed to the small decrease in non-IPV violent crime.

However, for the context of IPV, it is important to examine the potentially different labor market consequences of the minimum wage policy by gender. Aizer (2010) investigates the relationship between women's relative economic status and their vulnerability to domestic violence. She specifies a model in which (some) men have preferences for violence and partners bargain over the extent of abuse and the distribution of consumption in the household. The main prediction of the model is that an increase in a woman's relative wage comes with a higher bargaining power in the relationship and leads to a decrease in the level of violence as her outside option improves. Aizer (2010) presents empirical evidence that decreases in the wage gap reduce violence against women in the US, which is consistent with a household bargaining model. A study in the UK by Anderberg, Rainer, Wadsworth, and Wilson (2016) finds that while female unemployment increases the risk of domestic abuse, unemployment among males reduces it.[10]

Bonin, Isphording, Krause-Pilatus, Lichter, Pestel, Rinne, et al. (2018) have shown that the minimum wage introduction in Germany led to small aggregate effects for employment. There is, however, a more sizeable negative effect for marginal employment as the main occupation (aka minijobs[11]) and women are over-represented in this job type. This picture emerges also in my data. Figure 1.4

[10]The literature cited in the main text relates to findings for well-developed countries. A number of papers have studied the impact of women's labor market participation and the risk of domestic violence in the context of developing countries. Bhalotra, Kambhampati, Rawlings, and Siddique (2018) look at business cycle variations across 31 developing countries to examine how IPV is influenced by improved employment prospects for women. They come to the conclusion that an increase in female unemployment rates is associated with a decrease in the likelihood that women will be victims of IPV. One of the key results of the studies by Cools and Kotsadam (2017), Erten and Keskin (2021a), Guarnieri and Rainer (2021) and Tur-Prats (2021) is a positive association between female employment and IPV. These findings are consistent with male backlash.

[11]Marginal employment is defined as employment in which the regular pay does not exceed a maximum amount defined by law (in 2015: €450). In Germany, this is also referred to as a minijob. Under German law, employees in marginal employment are not subject to compulsory health, long-term care, and unemployment insurance in this employment (Bundesministerium für Wirtschaft, 2022). I will refer to employment subject to social insurance as 'regular employment'.




Notes: This Figure shows the average number of overall minijobs and minijobs as main occupation by gender per labor market area in year 2014.

*Figure 1.4:* Minijobs by Gender

illustrates that men represent only about a third of the people who only work in marginal employment.

Table 1.3 shows the short-term employment effects until September 2015. Column (1) presents a small negative effect for the overall number of employees, Column (2) a more sizeable negative estimate for the number of the so-called minijobs, which particularly decline if they represent a person's main occupation (Column 3). These effects are in line with the range of the outcomes for minijobs found by Bonin, Isphording, Krause-Pilatus, Lichter, Pestel, Rinne, et al. (2018). Women are particularly affected in absolute terms as they accounted for the vast majority of these jobs in 2014. Column (4) of Table 1.3 shows that a change from the 25th to the 75th percentile of the minimum wage bite distribution corresponds to 2.64% fewer minijobs for women, that is on average 391 fewer minijobs per labor market area. The decrease in minijobs for men amounts to only 176 per labor market area. Given that there is no corresponding increase in regular employment (see Table 1.A.10), this constitutes a net increase in unemployment. Viewed through the lens of Aizer's bargaining model, ceteris paribus, relationships in which women lose their jobs are at higher risk for IPV while where the man loses his job the risk of IPV is reduced. If one counts the number of affected men against the one of women, there are on average 215 (391-176) more women at risk per labor market area. Table 1.1 shows that on average 3.79 more IPV cases can be expected due to the minimum wage in a labor market area (7.73% of the mean of 49). If four out of the 215 people at higher IPV risk face violence, this corresponds to a rate of about 1.8%. This depicts a higher rate than the rate of 0.5% at which IPV shows up in the data for females of working age in 2014 (125,345 IPV cases/25,498,397 adult women under age 65).[12] The observed rate of IPV in the crime data, however, understates the incidence of IPV among the additional women at risk. The rate of IPV in crime data is an average over all socio-economic groups. The additional women at risk are women who recently lost a minijob; these tend to be less affluent households. The rate of IPV in such circumstances is presumably higher than the population-wide average (Lofstrom and Raphael, 2016).

[12]Source for population by age: https://www.inkar.de.

*Table 1.3:* Mechanisms

| | Employees | Minijober | Minijob as main occupation | Women: Minijob as main occupation | Men: Minijob as main occupation |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| post x mw bite | -0.060$^{***}$ | -0.164$^{***}$ | -0.216$^{***}$ | -0.257$^{***}$ | -0.208$^{***}$ |
| | (0.010) | (0.023) | (0.026) | (0.024) | (0.039) |
| *Effect size [in %]* | -6.18 | -17.82 | -24.11 | -29.30 | -23.12 |
| *25th-75th percentile effect [in %]* | -0.56 | -1.60 | -2.17 | -2.64 | -2.08 |
| post 2015 | 0.043$^{***}$ | 0.060$^{***}$ | 0.052$^{***}$ | 0.047$^{***}$ | 0.090$^{***}$ |
| | (0.005) | (0.011) | (0.012) | (0.011) | (0.018) |
| Labor market area FE | yes | yes | yes | yes | yes |
| Season FE | yes | yes | yes | yes | yes |
| Dep.var mean | 80,995 | 34,253 | 23,228 | 14,777 | 8,452 |
| N | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on the overall number of employed persons (Column 1), number of minijobs (Column 2), number of minijobs as main occupation (Column 3), number of female minijober (as main occupation) (Column 4), and number of male minijober (as main occupation) in Column (5). Results are shown for the same specification presented in the main Table 1, Columns (1), (3), and (5). Refer also to the table notes of Table 1.1. In Germany, a so-called minijob is defined as employment in which the regular pay does not exceed a maximum amount defined by law (in 2015: €450).
*p<.10, **p<.05, ***p<.01

## 1.6 Conclusion

This paper has two central objectives. The first is to examine the causal effect of the minimum wage introduction on Intimate Partner Violence (IPV). The second is to investigate empirically the mechanisms that might explain it. To achieve these, I combine data from several sources on crime outcomes, employment, and wages to assemble a monthly labor market area panel over the period 2011-2015. I employ a generalized difference-in-differences approach, exploiting the regional variation in the distance between a labor market area's wage level in 2014 and the uniform federal minimum wage in 2015 (i.e. the bite of the minimum wage).

My main finding is that the German federal minimum wage of 2015 increased the number of IPV cases, while it led to a small decrease in non-IPV violent crime. Moving from the 25th to the 75th percentile of the minimum wage bite distribution, the minimum wage reform resulted in roughly 8% of 2014 levels more IPV cases per labor market area. At the same time, non-IPV crime went down by about 2% in a labor market area. I decompose the effects along margins of interest to policymakers and show that the effect for non-IPV crime is driven by urban areas in West Germany. For IPV, there are no significant heterogeneous effects found. Several robustness checks corroborate these results.

To assess the factors that might explain the increase in IPV, I draw on data on income and employment. I refer to the bargaining model specified by Aizer (2010) in which (some) men have preferences for violence, and partners negotiate over the extent of abuse and household consumption distribution. I establish that adverse labor market effects are disproportionately in play for marginally employed women. This is consistent with a loss of bargaining power for women and results in more females being at risk of IPV.

The results presented in this paper point to unintended consequences of minimum wage policy in the first months after its implementation. An interesting task for future research is to investigate long-term impacts. The conclusions of this study can also be applied to a country that is beginning to enforce a minimum wage that has already been officially implemented.

# 2

# Deterrence or Backlash? Arrests and the Dynamics of Domestic Violence*

Abstract

There is a vigorous debate on whether arrests for domestic violence (DV) will deter future abuse or create a retaliatory backlash. We study how arrests affect the dynamics of DV using administrative data for over 124,000 DV emergency calls (999 calls) for West Midlands, the second most populous county in England. We take advantage of conditional random assignment of officers to a case by call handlers, combined with systematic differences across police officers in their propensity to arrest suspected batterers. We find that an arrest reduces future DV calls in the ensuing year by 51%. This reduction is not driven by reduced reporting due to fear of retaliation, but instead a decline in repeat victimization. We reach this conclusion based on a threshold reporting model and its testable implications regarding (i) the severity of repeat DV calls and (ii) victim versus third-party reporting. Exploring mechanisms, we find that arrest virtually eliminates the large spike in re-victimization which occurs in the 48 hours after a call, consistent with arrest facilitating a cooling off period during a volatile, at-risk time. In the longer run, we estimate a sizeable deterrence effect. Substantiating this, arrest increases the probability an offender is charged with a crime. Our findings argue against recent calls for a decriminalization of domestic violence and suggest the optimal police response is to lower the threshold for arrest.

---

*This Chapter is based on joint work with Sofia Amaral (World Bank), Gordon Dahl (UC San Diego), Timo Hener (Aarhus University) and Helmut Rainer (ifo Institute and LMU Munich). A version of this Chapter has been published in the NBER and CESifo working paper series (NBER Working Paper No. 30855; CESifo Working Paper No. 10205). I thank the West Midlands Police for generously providing access to their data and expertise, without which this project would not have been possible.

## 2.1 Introduction

Domestic violence (DV) is a pervasive threat to the well-being of women worldwide, with one third of women reporting some form of physical or sexual violence from a partner during their lifetime (WHO, 2021). A key aspect of DV is that it is seldom a one-time occurrence, with women frequently experiencing repeat abuse by the same partner (Tjaden and Thoennes, 2000; Aizer and Dal Bo, 2009). Despite the prevalence and seriousness of DV, the question of how best to police this crime so as to break the cycle of DV is still largely unresolved. A highly controversial police response is to arrest suspects on the spot.

The nature of the controversy is multifaceted. Proponents of arrest contend that in addition to temporarily incapacitating offenders, it deters men from future abuse by signaling a high cost for repeat incidents (Berk, 1993). Opponents raise concerns about backlash effects (Schmidt and Sherman, 1993; Goodmark, 2018), arguing that while arrest offers immediate relief, it triggers an escalation of DV in the longer term as abusers retaliate against their partners. Both sides also make arguments related to whether victims will report future DV, with proponents claiming that arrest empowers women to do so and opponents saying that arrest discourages future calls for police help.

Against the backdrop of these debates, this paper asks how arrest affects the dynamics of DV. Estimating the consequences of arrest on future victimization is challenging for two reasons. The first is the scarcity of large-scale datasets which include information on DV incidents, police responses, and repeat victimization. In most datasets, DV is not directly identified as a crime category, but must be inferred based on the characteristics of the incident. Moreover, researchers often only observe cases where criminal charges are filed, but in DV cases, the victim is often reluctant to press charges. And finally, it is difficult to link repeat victimization to a prior incident; even in datasets which track identities, they are usually only recorded if there is a documented criminal charge.

The second estimation challenge is that arrest is endogenous, leading to selection bias. The problem is that there are likely to be characteristics of a case which are observable to police officers, but not the researcher. For example, if cases which result in arrest are (unobservably) more serious and hence more likely to be positively related to the underlying risk of repeat violence, OLS will underestimate any potential deterrent effect or possibly yield a positive estimate even if there is no backlash.

We address both the data and endogeneity challenges in the context of emergency

999 calls in the United Kingdom (similar to 911 calls in the US). On the data front, we observe the universe of all 999 emergency calls for West Midlands, the second most populous county in England, over 10 years. Due to the priority placed on domestic abuse in the UK in recent years, DV is singled out as a separate category by call handlers. We follow an incident from the time the call is placed until the first-response police officers arrive and complete their on-scene intervention (including whether they arrest a suspected offender on the spot). We merge these records with data on whether a criminal investigation is opened by an investigative officer and, if so, whether offenders are charged with a crime.

To create a linked panel of DV incidents over time, we exploit information on the precise geo-location of where the incident occurred. This takes advantage of the fact that most DV occurs at home and that most police intervention originates via a 999 emergency call (HM Inspectorate of Constabulary, 2014). The key benefit of this approach is that it allows us to track repeat DV even if there is not a formal criminal charge filed. For this reason, we are likely to pick up a much higher fraction of repeat cases compared to other panel datasets which only have information related to formal charges. We provide several checks showing that the scope for misclassification errors using this measure (e.g., residential moves after being exposed to DV) is limited.[1]

To address endogeneity, we exploit the conditional random assignment of police officers to 999 DV calls combined with heterogeneity in officers' propensity to arrest.[2] Since it is difficult to precisely predict when and where demands for police resources will emerge, the availability of patrol officers which can be dispatched to a DV incident is as good as random after conditioning on time, geography, and the priority level assigned by the call handler. Since some officers are more likely to arrest than others, the average arrest propensity in other cases can be used as an instrument for arrest in the current case. As multiple patrol officers can be dispatched to an incident (there are usually 2 officers in a patrol car), we use the

[1]For the subsample of cases which result in a criminal investigation (and hence for which we observe victim ID), we show that an arrest does not impact whether an observation is identified as a repeat DV incident using geo-location. We also show that our geo-location based results are robust to focusing on areas with single family homes as opposed to multifamily units.

[2]Our IV strategy is inspired by the quasi-random assignment of criminal cases to judges to investigate the consequences of juvenile incarceration (Aizer and Doyle Jr, 2015), adult incarceration (Kling, 2006; Bhuller, Dahl, Løken, and Mogstad, 2020), pre-trial detention (Dobbie, Goldin, and Yang, 2018), and electronic monitoring (Di Tella and Schargrodsky, 2013). In contexts other than crime, quasi-random assignment of caseworkers or judges has been used to examine the effects of disability insurance receipt (Maestas, Mullen, and Strand, 2013; Dahl, Kostøl, and Mogstad, 2014; French and Song, 2014; Autor, Kostol, Mogstad, and Setzler, 2019), child protection (Doyle Jr, 2007, 2008), and consumer bankruptcy (Dobbie and Song, 2015).

weighted average arrest rate of dispatched team members to DV calls.[3] As we show, the instrument is highly predictive of arrest in the current case, but uncorrelated with observable case characteristics. Tests support the monotonicity assumption. To address concerns about the exclusion restriction, we show that the IV estimate for arrest is unlikely to be biased by other actions taken by responding officers on the scene or the characteristics of responding officers.

Our main finding is that arrest reduces the probability of a repeat DV emergency call within the ensuing 12 months by 51%. In sharp contrast, OLS finds a precisely estimated zero effect. The implication is that not accounting for selection bias would lead one to erroneously conclude that arrest has no affect on DV call trajectories.

Whether the reduction in repeat DV calls after an arrest is good or bad from a policy perspective depends on whether it is driven by a reduction in incidence or merely a reduction in reporting. To disentangle these explanations for our main result, we use a simple threshold reporting model. In the model, women who experience backlash from an arrest raise the level of abuse they are willing to tolerate before reporting in the future. In contrast, women who are empowered due to the deterrent effect of arrest will report future abuse at a lower threshold. Using a measure for the severity of repeat DV calls, we find that reporting thresholds drop on average after an arrest: there is a large reduction in severe DV calls and an increase in less severe DV calls. This compositional shift is statistically significant. Viewed through the lens of the model, it implies the decline in future DV calls is not driven by a change in reporting behavior, but rather a reduction in abuse. As a second exercise, we use differences in victim versus third-party reports of repeat DV calls. The idea behind this comparison is that any reluctance to report a repeat case due to fears of retaliation should be substantially larger for the victim than a third party. Yet we find, if anything, the opposite: a larger (but statistically insignificant) reduction in third party versus victim reporting. This is also consistent with a drop in actual abuse.

We next turn to mechanisms. In the absence of an arrest, we estimate that 23% of compliers will experience repeat DV within 48 hours. An arrest prevents virtually all of this re-victimization. This is consistent with arrest facilitating a cooling-off period, which could happen if the offender is removed from the scene to be processed or held in temporary custody. Additional reductions occur over the following year, consistent with a longer-term deterrence effect. Consistent with a deterrence effect,

[3] We use the initially dispatched officers to construct the instrument, as additional officers may be called to the scene if backup is needed or if the initially dispatched officers are not able to respond.

we estimate that an arrest leads to a higher probability of being formally charged with a crime. For non-arrested compliers, the probability of being formally charged with a crime is just 2%, while for arrested compliers, this probability rises to 12%. These findings go against the claim that arrest has a weak dosage of legal sanctions and therefore will be ineffective.

Taken together, our findings provide a cautionary tale for recent policy debates calling for a decriminalization of DV. Goodmark (2018) argues that the criminal legal system has been ineffective in deterring DV and has had detrimental consequences for victims, offenders, and their communities. Arrest in particular has been singled out as having negligible effects on DV recidivism. Indeed, our OLS estimates show no association between arrest and future violence. However, our causal estimates indicate a strong deterrent effect of arrest on domestic assault for compliers. Our results argue against calls to decriminalize DV and suggest instead that the optimal police response is to deal more strictly with batterers and lower the police threshold for arrests.

In the existing literature, there is sparse causal evidence on the effect arrest has on repeat victimization. Much of what we know comes from the 1981 Minneapolis Domestic Violence Experiment (Sherman and Berk, 1984) and its replications in Omaha (Dunford, Huizinga, and Elliott, 1990), Charlotte (Hirschel and Hutchison III, 1992), Milwaukee (Sherman, Schmidt, Rogan, and Smith, 1992), Dade County (Pate and Hamilton, 1992), and Colorado Springs (Berk, Campbell, Klap, and Western, 1992). In these social experiments, the research design called for one of three treatments to be randomly enforced by patrol officers encountering a DV situation: arrest, separate, or advise. Despite being highly innovative at the time, the significance of these trials is limited in that they produced inconclusive findings,[4] had to contend with small samples, and suffered from flawed designs. A major concern was non-compliance with random assignment: patrol officers frequently deviated from the treatment called for by random assignment, and their delivered treatments had a substantial behavioral component (Angrist, 2006). These studies highlight some of the challenges involved in using randomized control trials for the empirical analysis of crime (Pinotti, 2020). More recently, a related literature examines how the staggered introduction of arrest laws in the US affected state-level intimate partner homicide rates. Iyengar (2009) finds an increase in homicides after a mandatory arrest while Chin and Cunningham (2019) revisit the issue and find no effect for

[4] In Omaha, Charlotte, and Milwaukee, offenders assigned to the arrest group showed higher levels of repeat offending while in the other three jurisdictions, a modest reduction in recidivism was found among batterers assigned to arrest (Schmidt and Sherman, 1993).

mandatory arrest laws but a reduction in homicides for discretionary arrest laws.

Our paper is also related to a small but growing number of papers which study the role of law enforcement more generally in the prevention of DV. Miller and Segal (2019) provide evidence that when female representation increases among police officers in an area, DV crimes are reported at higher rates, intimate partner homicide falls, and non-fatal domestic abuse declines. Aizer and Dal Bo (2009) find that the adoption of no-drop policies in the US, which require prosecution even if DV victims request charges be dropped, resulted in an increase in reporting of DV. Golestani, Owens, and Raissian (2021) show that victims in cases assigned to specialized DV courts are less likely to be re-victimized and more likely to cooperate with police and prosecutors. In a contemporary working paper, Black, Grogger, Sanders, and Kirchmaier (2022) use inverse propensity-score weighting to study the role of criminal charges against offenders and protective services for victims, finding that the former substantially reduces the risk of DV recidivism. Grogger, Gupta, Ivandic, and Kirchmaier (2021) apply machine learning methods to forecast recidivism in DV cases, laying out how their approach can be used to prioritize DV emergency calls.

While our focus is on how DV can be prevented, there is an important line of work examining its determinants. Bhalotra, GC Britto, Pinotti, and Sampaio (2021) estimate the impacts of job loss and unemployment benefits while work by Aizer (2010), Tur-Prats (2021), and Erten and Keskin (2021b) has focused on local labor market conditions. Cultural factors have also been linked to DV (e.g., Alesina, Brioschi, and La Ferrara (2021), González and Rodríguez-Planas (2020), and Tur-Prats (2019)). Our paper also relates to a small literature that examines the effects of DV on victims and their children. Aizer (2011) and Currie, Mueller-Smith, and Rossin-Slater (2022) find that domestic violence during pregnancy significantly increases the incidence of negative birth outcomes.

The remainder of this chapter is organized as follows. In Section 2 we discuss our setting and research design and in Section 3 we assess our instrument. Section 4 presents the main results on how arrest affects repeat DV calls. Section 5 tests for changes in incidence versus reporting behavior. Section 6 explores mechanisms, followed by further robustness checks and heterogeneity in Section 7. The final section concludes.

## 2.2 Research Design

In this section we outline our research design. We begin by briefly describing our setting, followed by an explanation of how DV emergency calls are handled. We then discuss how the conditional random assignment of police officers with different arrest propensities can be used to estimate the effect of arrests on repeat DV.

### 2.2.1 Domestic Violence in West Midlands County

Our setting is the policing of domestic violence which occurs in West Midlands County in England. The prevalence of DV in the UK mirrors that of other high-income countries. The World Health Organization estimates that the lifetime incidence of intimate partner violence is 24% in the UK; for comparison, it is 22% in Europe more broadly and 26% in the US (WHO, 2021).

West Midlands is the second largest county in England, with a population of 3 million. The county includes the cities of Birmingham, Coventry, and Wolverhampton. The area is ethnically diverse, with 79% Whites, 11% Asians, 3% Blacks, and 7% Other (Office for National Statistics, 2020). Using survey data, the estimated annual intimate partner victimization rate in the police force area is 6.6%, a percentage which is somewhat higher compared to England as a whole (5.8%).[5]

There are several features of DV which distinguish it from other types of crimes. First, most DV occurs at home (90% of cases in England according to HM Inspectorate of Constabulary (2014)). Second, repeat victimization is common (49% within a year in our data). Moreover, the majority of victims facing repeat episodes of DV are women (89% according to Walby and Allen (2004)). At its most severe, DV can result in mental health breakdowns, bodily injury, or even death (WHO, 2002). In the UK, 44% of female homicide victims are killed by their partners or ex-partners (Office for National Statistics, 2013). These patterns for England in general, and the West Midlands in particular, mirror those for other countries (WHO, 2002, 2021).

### 2.2.2 Handling of DV Emergency Calls

In England, 999 is the official telephone number which allows citizens to contact the police for emergency assistance (similar to a 911 call in the US). Figure 2.1 outlines how 999 emergency calls for DV work their way through the system, beginning

[5] Based on own calculations using British Crime Survey data. The reported numbers are average victimization rates (for both physical and non-physical DV) over the survey years 2004/05 to 2009/10.

with the call itself and ending with the resolution of the case.

Call handlers are the first step in the process. In the West Midlands, 999 calls arrive in one of two main control rooms of a centralized call center. Due to the seriousness of DV offenses, the government requires call handlers to identify and separately classify DV calls. DV includes incidents of threatening behavior, violence, or abuse between intimate partners. DV calls make up 9% of all emergency calls. The call handler records the identity of the caller (victim 33% vs. third party 67%), confirms the GPS location of the incident,[6] and grades the priority level of the incident. This call grade priority ranking determines how quickly officers are expected to respond based on the severity and immediacy of the situation; the goal is to respond within 15 minutes for priority 1 (immediate response) and within 60 minutes for priority 2 (priority/early response). 60% of DV calls are given a priority ranking of 1 and 31% are given a priority ranking of 2. The fact that few DV calls are given a priority ranking of 3 or below is consistent with the mandate that police officers need to defuse DV situations before they escalate.

In the next step, the information collected by call handlers is transferred electronically to the dispatcher. The role of dispatchers is to coordinate first response police teams (i.e., police officers driving in the same patrol car) in the field. To that end, they monitor an electronic map with the real-time locations of all police vehicles within each catchment area. Based on the severity ranking of the call, and the availability of response teams and their proximity to the incident's location, the dispatcher assigns a team of first response officers to the incident.

First response teams are not specialized, but rather respond to all types of emergency 999 calls. The primary purpose of the first response team in DV cases is to protect victims and prevent further harm. To achieve this goal, police officers can separate the offender from the victim, provide advice, and collect information. Officers make a short-term safety plan with the victim, often making them aware of support organizations and safe places in their communities. The most drastic immediate action that can be taken at the scene is to arrest the perpetrator. Arrest is a discretionary action which occurs if there are reasonable grounds to suspect a crime has been committed, if the victim's safety is at risk, or if it will help facilitate the collecting of information. Our database records the most severe action taken by the first response team, with (in descending order of severity) arrest occurring 3.1% of the time, a recommended criminal investigation 42.0% of the time, advice/warning

[6] Location is recorded with an accuracy of ten by ten meters using the Ordnance Survey National Grid reference system (a 12 digit point reference). The determination of geo-location is a semi-automated process, with call handlers entering address information when the incident occurs at a different location compared to the origination of the cell phone call.

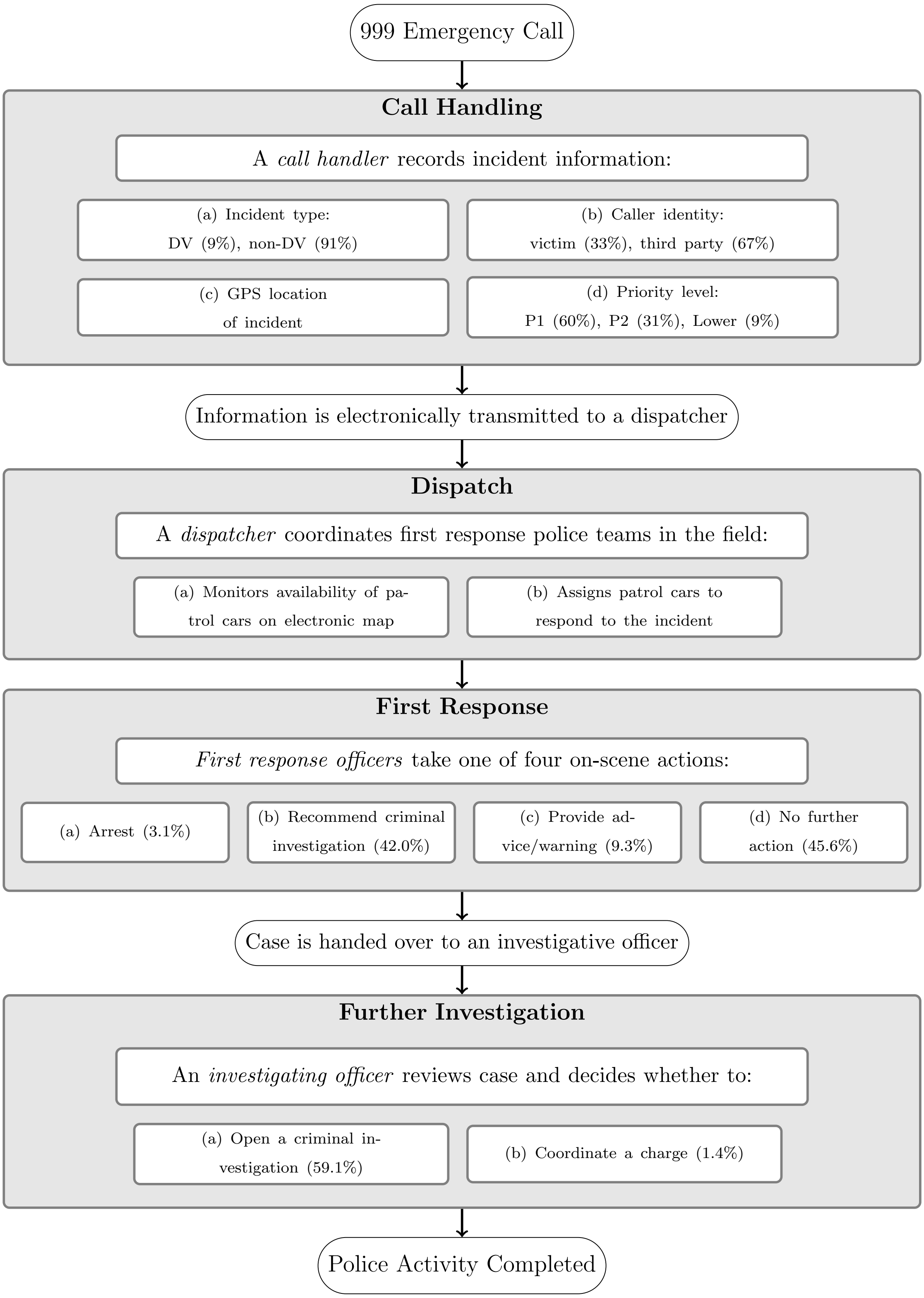



*Figure 2.1:* Police Handling of DV Emergency Calls

9.3% of the time, and no further action 45.6% of the time. Since only the most severe action is recorded in the dataset, an arrest could also include a recommended criminal investigation or advice.

After completing their on-scene work, first response officers transfer the responsibility of the case to an investigative officer who specializes in domestic violence. The investigative officer decides whether to open a formal investigation, which prompts an entry in a national crime database with unique person identifiers for the victim and perpetrator. Police can search this database when there is a future incident, and eligible parties can make information requests according to the Domestic Violence Disclosure Scheme (also known as Clare's Law). Criminal investigations occur in 59.1% of incidents. During the investigation, further evidence can be gathered and information shared with a prosecutor. The prosecutor works with the investigative officer to determine whether the offender should be formally charged with a crime and summoned to court (1.4% of incidents).[7] During this process, suspects can be taken into custody for a maximum of 24 hours, with a possibility of extended custody up to 96 hours for more serious crimes.[8]

Our analysis is possible due to detailed data which allows us to observe what happens at each stage of the process depicted in Figure 2.1. We combine information from five different police registries for (i) incidents, (ii) officer deployments, (iii) investigations, (iv) charges, and (v) police personnel records.

### 2.2.3 Conditional Random Assignment

The way DV emergency calls are handled leads to conditional random assignment. Dispatchers receive only limited information from the call handlers, making dispatch decisions based on the priority grading of the call and the availability of nearby officers on patrol. While we do not directly observe the availability of officers, we control for predictable staffing needs using a rich set of time and geographic variables. Specifically, we use variables for year, calendar month, day of week, time of day, and bank holidays as well as ward (146 geographic areas which are subsets of police officer catchment areas). We also control for call grade, as higher ranked calls jump to the top of the dispatch queue. After controlling for these variables flexibly, precisely when and where residual demands for police resources will emerge

[7]The charging decision usually lies with the prosecutor. However, for summary offenses (those with a maximum punishment of 6 months in prison), investigative officers can directly charge the offender.

[8]The custody decision is made by custody officers and not directly by the first response or investigative officers.

should be as good as random, something we verify empirically.

Several features of the dispatch process make it unlikely that dispatchers selectively assign response teams to incidents after conditioning on this information. First, the guiding principle in DV incidents is to ensure a speedy police response, with the first available officers being sent to the scene. In our baseline specification, we restrict our sample to DV cases with a call grade of 1 or 2 (91% of all DV calls), as in these cases the mandate is to arrive on the scene as soon as possible. Calls with a grade of 1 are classified as "immediate response" and include calls where there is "A danger to life/use or immediate threat of use of violence/serious injury to a person." The directive is for the police officers to arrive within 15 minutes of the call; we observe a mean response time of 8 minutes. Calls with a grade of 2 are classified as "priority response/early response" and include calls where there is a "concern for someone's safety." The directive is to arrive within 60 minutes; we observe a mean response time of 26 minutes.[9]

Second, the limited information available to dispatchers argues against non-random assignment. On their electronic map with patrol cars' real-time locations, dispatchers only observe officers' identification numbers without information on their names or background. Dispatchers also never have any direct contact with the individual who calls for help, ruling out that they have extra information on specific victim needs.

A first response team is usually 2 police officers in a patrol car. We measure the team arrest propensity as the total number of DV arrests by all officers in the team divided by the total number of DV cases handled by all officers on a team. In other words, we use the weighted average of individual officers' arrest propensities, where the weights are proportional to the number of cases handled by an officer. To construct the instrument, we use all DV cases an officer has been involved in (minus the current case and any cases at the same address), including both past and future cases, and not just those cases which appear in our estimation sample.[10]

In the construction of our instrument for team arrest propensity, we only consider police officers who are initially dispatched to a DV incident before any other officers arrive, leaving out those who are called to the scene later for reinforcement. Doing so ensures that information about the case collected from officers who are the first to

[9]Potential harm to individuals is not the only principle used to assign call grades. For example, call grade 1 includes cases where the crime is in progress and call grade 2 includes cases where evidence is likely to be lost if there is a delay.

[10]We use both past and future cases to increase precision, similar to Bhuller, Dahl, Løken, and Mogstad (2020). Due to random assignment, using future cases from other addresses does not invalidate the instrument even if past cases influence future officer arrest behavior.

arrive at the scene does not lead to selective reinforcements. We also use the initially dispatched officers to construct the instrument, even if they are not able to respond and a different team is assigned in their place. On average there are 2.6 officers initially dispatched to the scene. When using the instrument, we always condition flexibly on call grade, time, and geography fixed effects. We further require at least 400 total cases for a team so as to minimize any mechanical bias; the rationale for requiring a large number of cases is that arrest occurs rarely (3% of cases). As we will show, results are robust to higher or lower thresholds.

Table 2.A.1 details the sample sizes for various components of our analysis. Our baseline estimation sample includes 124,216 observations, comprised of cases classified as DV by call handlers between 2011-2016, with at least 400 DV cases for the dispatched team, and with a call grade of 1 or 2. The construction of our instrument uses 631,834 officer-case level observations between 2010-2019.

Table 2.A.2 tests whether first response teams in our baseline sample are randomly assigned to DV calls. We start by showing which characteristics predict whether an arrest will be made. The first Column regresses arrest in the current case on predetermined characteristics, controlling flexibly for call grade, time, and geography. The estimates reveal that a prior DV case, a prior arrest, and a prior criminal charge all strongly predict whether an arrest will be made in the current case. Who initiates the call, the gender of the call handler, and the experience of the call handler have no statistically significant effect. The joint F statistic for all of these variables is highly significant (p-value $< 0.001$).

We next show that despite the predictive power of these case characteristics, they are uncorrelated with our instrument. The second Column regresses team arrest propensity on the same set of characteristics, controlling flexibly for call grade, time, and geography. The estimates are all close to zero and much smaller than those in the first Column. They are also statistically insignificant, both individually and as a group (p-value=.478). This provides empirical support for first response teams being conditionally randomly assigned to DV calls. Note that for this Table, we multiplied the arrest and team propensity variables by 100 so that the Table would be more readable (i.e., to avoid estimates of 0.000). We do not multiply these variables by 100 anywhere else in the paper.

### 2.2.4 Regression Model

We are interested in estimating the effect of arrests on future victimization, which we model with the regression

$$Future\ DV_i = \beta_0 + \beta_1 A_i + X_i'\delta + \epsilon_i \tag{2.1}$$

where the outcome variable $Future\ DV_i$ measures whether domestic violence occurs again in the future. $A_i$ is an indicator for whether an arrest is made in the current case, $X_i'$ is a vector of flexible controls for call grade, time, and geography, and $\epsilon_i$ is an error term. In our preferred specification, we include fixed effects for call grade (1 versus 2), year, calendar month, day of week, time of day (6 hour intervals), and bank holidays, each interacted with ward fixed effects.

For our measure of repeat victimization, $Future\ DV_i$, we use the officers' coding of future DV, even if the future case was not initially classified as DV by the call handler. Note, however, that when constructing our instrument we only use calls initially classified as DV by the call handler to avoid the possible endogeneity of officers' classifications.

Our research design recognizes that arrests are unlikely to be random. More serious incidents of domestic violence are likely to both increase the probability of arrest and lead to future victimization. Failure to account for the endogeneity of arrest may therefore lead to an underestimate of any deterrent effect or could yield a positive estimate even in the absence of backlash. We therefore use the average arrest propensity of the first response team as an instrument for arrest in the current case. The intuition for our team arrest propensity instrument is that a suspected batterer will more likely be arrested if the police officers handling the current case have a higher average arrest rate in other DV cases they handle.

The first stage regression can be written as

$$A_i = \alpha_0 + \alpha_1 Z_{j(i)} + X_i'\gamma + \eta_i \tag{2.2}$$

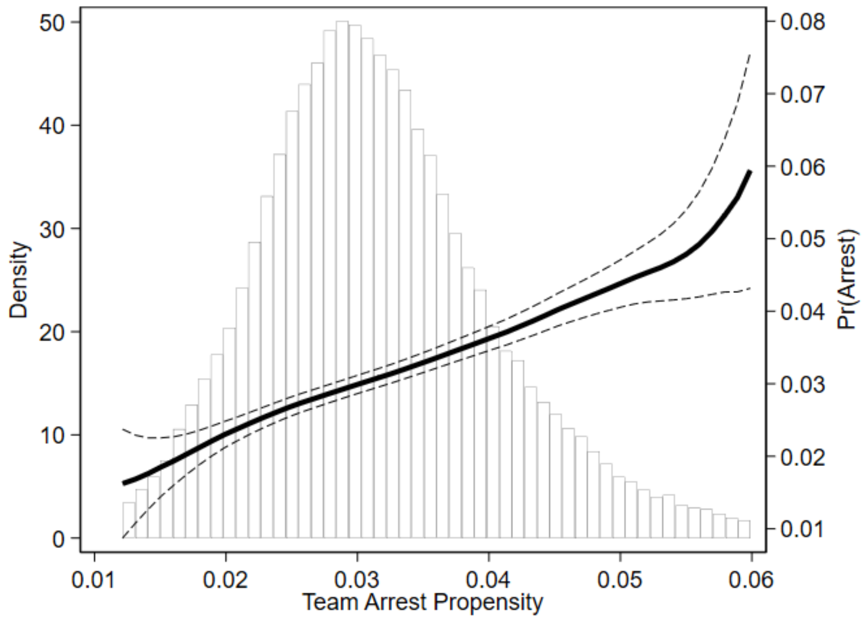
where the scalar $Z_{j(i)}$ denotes the arrest propensity of the first response team $j$ assigned to case $i$, as defined in Section 2.2.3. We use instrumental variable regression based on equations (1) and (2) to estimate the causal effect of arrest on future victimization. We cluster standard errors at the level of the dispatched officer on a team with the most domestic violence cases.

## 2.3 Assessing the Instrument

### 2.3.1 Relevance

Key to our design is that not only are first response teams conditionally randomly assigned, but also that they differ in their arrest propensities. The histogram in Figure 2.2 displays the distribution of arrest propensities. In constructing this histogram, we first regress out the conditioning variables for call grade, time, and geography to match the variation used in our analysis. The mean of the instrument is 0.030 with a standard deviation of 0.012. The Figure reveals substantial heterogeneity in arrest propensities, with the 1-99 percentile range spanning arrest rates of 0.012 to 0.060.[11]

[11] A natural question is why some police officers are more prone to arrest. A simple regression reveals that teams with fewer female officers are more likely to arrest, while average officer age (which proxies for experience) has no impact. Later, we show our findings are robust when also including these average team characteristics in our regressions. Importantly, other factors which we do not measure account for the overwhelming share (99%) of the residual variation in arrest propensities (after regressing out call grade, time, and geography).




Notes: The probability of an arrest is plotted on the right-hand axis against team arrest propensity on the x-axis. Plotted values are based on mean-standardized residuals from a regression of arrest on the baseline set of call grade, time, and geography controls used in Table 2.1 Column (4). The solid line shows a local linear regression of arrest on team arrest propensity, with dashed lines indicating 95% confidence intervals. The histogram shows the density of team arrest propensities along the left-hand axis (top and bottom 1% excluded).

*Figure 2.2:* First Stage Graph of Arrest on Team Arrest Propensity

Figure 2.2 also graphs the probability the suspected offender will be arrested in the current case as a function of the arrest propensity of the first response team. The solid line plots estimates from a local linear regression, and hence represents a flexible analog of the first stage. The predicted line is monotonically increasing in the instrument and close to linear. First stage estimates based on equation (2) are reported in the bottom panel of Table 2.1, and are all highly significant. The Kleibergen-Paap Wald F statistics reveal that our instrument is not weak. Column (4) includes our preferred set of conditioning variables, and indicates that being assigned a first response team with a 1 percentage point higher arrest rate increases the probability of being arrested by 0.722 percentage points.

Note that the first stage estimate need not mechanically equal one as the number of cases per team goes to infinity for several reasons: (i) the sample of cases used to calculate the instrument is not the same as the estimation sample, (ii) there are covariates (i.e., call grade, time, and geography controls), and (iii) the instrument is calculated by weighting the arrest propensities of different officers in a team, where teams are not held constant over time. Hence, there is no reason to expect a coefficient of one in our setting.

### 2.3.2 Validity

For the instrument to be valid, arrest propensities must be conditionally independent of case characteristics that affect the likelihood of repeat DV. In Section 2.2.3, we argued the assignment of DV emergency calls to first response teams should be random after conditioning flexibly on call grade, time, and geography, and tested for balance on predetermined observable characteristics in Table 2.A.2. As a second test, we add in these same predetermined characteristics to the first stage. If teams are randomly assigned, these characteristics should be uncorrelated with the instrument, and hence not significantly change the estimate. As expected, the first stage coefficient does not change appreciably (0.721 versus 0.722).

### 2.3.3 Monotonicity

If the effect of an arrest is homogeneous across DV incidents, then conditional random assignment and exclusion are enough for IV to capture the causal effect of arrest. If effects are heterogeneous, then the instrument also needs to satisfy monotonicity. With monotonicity, IV can be interpreted as the local average treatment effect (LATE) for compliers – i.e., the arrest effect for cases which would have a different arrest outcome if they had been assigned a first response team with a higher or lower arrest propensity. These compliers are particularly relevant for policy, as any changes in arrest guidelines for police officers are likely to be targeted towards these marginal DV cases.

In our setting, monotonicity requires that if an arrest is made by a lenient officer team (i.e., a team with a low arrest propensity), then it would also have been made by a stricter team, and vice versa when an arrest does not occur. One testable implication is that the first stage estimates should be the same sign for any subsample of the data. To implement this test, we construct the instrument using the entire sample of cases, but estimate the first stage on specific subsamples: by incident order, by DV hotspot area, and by time of day. Panel A of Table 2.A.3 reveals that the estimated first stage coefficients are all positive and statistically significant, as expected if monotonicity holds.

Panel B of Table 2.A.3 conducts the ”reverse instrument” test for monotonicity proposed by Bhuller, Dahl, Løken, and Mogstad (2020). The test is based on the implication that response teams which have higher arrest propensities for one case type (e.g., first time callers) should also have higher arrest propensities for other case types (e.g., higher order callers). To test this, we break the data into the same subsamples as in panel A, but redefine the instrument for each subsample to be the arrest propensity of the team for cases outside the subsample. Consistent with the monotonicity requirement that officer teams which are more arrest prone in one case type also being more arrest prone in other case types, the first stage estimates are all positive.

### 2.3.4 Exclusion

The exclusion restriction for the IV estimate requires that a first response team with a higher arrest propensity only affects repeat DV by increasing the probability of an arrest. After discussing our main results, we present tests related to two concerns about the exclusion restriction: (i) that teams with a higher arrest propensity affect recidivism not only through the arrest channel but also through the other actions they take as part of their on-scene police work and (ii) that a team's arrest propensity correlates with other team characteristics that affect the probability of repeat victimization directly. These tests indicate that the exclusion restriction is likely to hold.

## 2.4 Effect of Arrest on Repeat DV Calls

We now examine how arrest affects repeat emergency DV calls. In our data, we do not observe the gender of the victim, so our analysis concerns victims of both genders. However, since most victims of DV are female,[12] for simplicity we sometimes refer to the victim as a woman and the offender as a man.

Column (1) of Table 2.1 starts by reporting results using OLS, where the outcome variable is a repeat 999 DV call within the ensuing 12 months. For comparability with our preferred IV specification, the OLS regression includes the same set of call grade, time, and geography controls. The OLS estimate of the effect of arrest on a repeat DV call is a precisely estimated zero.

The IV estimates in Table 2.1 stand in sharp contrast to OLS. Regardless of how we control for call grade, time, and geography across Columns (2)-(4), there is a roughly 50 percentage point reduction in repeat DV calls after an arrest. Our preferred specification is found in Column (4), which includes the most flexible set

[12] According to Walby and Allen (2004), women are the victim in DV 89% of the time in the UK.

of controls for call grade, time, and geography. Focusing on our baseline estimate, if a suspected batterer is arrested on the scene, the probability of a repeat emergency call for DV falls by 48.8 percentage points. To help interpret this magnitude, the Table reports the estimated control complier mean of the outcome (i.e., the mean for compliers where there is not an arrest), which is 96.2%.[13] Hence, our IV estimate represents a sizable 51% reduction in the probability of a repeat call for compliers. As we will show in Section 2.7.3, there is a similarly sized 50% reduction in the *number* of repeat calls.

The estimated control complier mean for having a repeat call within the following year is substantially larger compared to the overall mean of 49.2%, but not unexpectedly so. The reason is that a large fraction of DV emergency calls are relatively minor in terms of not meriting an arrest from any response team. We calculate that never takers (cases where no police officer would make an arrest) comprise 94.3% of our sample. The cases where arrest could be considered a commensurate response are those which are serious and likely to result in repeat victimization. But not all police officer teams choose to arrest in these cases, as teams with low arrest propensities may believe that arrests almost always do more harm than good in terms of the severity of repeat victimization, for example. In these cases without an arrest, the probability of revictimization will be very high.

[13] We estimate the control complier mean following the Technical Appendix to Dahl, Kostøl, and Mogstad (2014). We use the 1st and 99th percentiles of team arrest propensity to define the least and most stringent teams, combined with the estimated coefficients from a linear first stage, to calculate the fraction of compliers, always takers, and never takers. Further details are found in Appendix B of this paper.

*Table 2.1:* The Effect of Arrest on Repeat Emergency Calls for Domestic Violence

| | Dependent variable: **Repeat call for DV** | | | |
|---|---|---|---|---|
| | OLS | IV | | |
| | (1) | (2) | (3) | (4) |
| Arrest | 0.001 | -0.517*** | -0.488*** | -0.488*** |
| | (0.008) | (0.170) | (0.187) | (0.187) |
| Call grade, time, ward F.E.'s | yes | yes | yes | yes |
| Ward x time F.E.'s | yes | no | yes | yes |
| Ward x call grade F.E.'s | yes | no | no | yes |
| Mean of dep. var. | 0.492 | | | |
| Control complier mean | 0.962 | | | |
| First stage | | 0.772*** | 0.723*** | 0.722*** |
| | | (0.068) | (0.070) | (0.070) |
| Reduced Form | | -0.400*** | -0.353*** | -0.352*** |
| | | (0.131) | (0.136) | (0.136) |
| Kleibergen-Paap Wald F statistic | | 128 | 108 | 107 |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: Call grade fixed effects are for call grade 1 versus 2. Time fixed effects include year, calendar month, day of week, time of day (6 hour intervals), and bank holidays. Ward fixed effects correspond to 146 geographical regions. The baseline specification in Column (4) adds in complete interactions of the ward and time fixed effects as well as interactions between ward and call grade fixed effects. The first stage is a regression of arrest on the team arrest propensity instrument and the reduced form is a regression of repeat call for DV on the team arrest propensity instrument, controlling for the relevant fixed effects. Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

This argument is further supported by the estimated characteristics of compliers in our sample. We estimate the fraction of compliers to be 4%. In Table 2.A.5 we characterize compliers by their past DV history.[14] Compared to the overall population, compliers are more likely to be serial offenders (Column 1) who have committed more serious crimes (Column 4). These complier cases appear to part of

[14]Our characterization of compliers adapts the binary-instrument methodology proposed by Marbach and Hangartner (2020) to a setting with a continuous instrument. See Appendix B for details.

increasing pattern of DV, with these offenders having evaded arrest at a higher rate in the past (Column 2).

One possible explanation for the divergence between the IV and OLS estimates in Table 2.1 is that the effect of arrest is heterogeneous, differing for compliers versus the entire population. To explore this possibility, we characterize compliers by the four characteristics describing past DV history appearing in Table 2.A.2: whether there was a prior DV case, a prior arrest, a prior investigation, and a prior charge.[15] The OLS estimate for this complier reweighted sample is 0.040 (s.e.=0.014), suggesting that the difference in estimates is not driven by heterogeneous effects, at least based on the observable characteristics available to us.

An alternative explanation for the divergence between the OLS and IV estimates is selection bias. The OLS estimate is likely to be biased upwards, as cases which have characteristics associated with higher recidivism probabilities are also plausibly more likely to result in arrest. The bias arises because these case characteristics are observed by police officers, but not by the researcher. Our IV estimates reveal that OLS would lead to the mistaken conclusion that arrest has no effect on repeat DV calls, when in fact, arrest results in a halving of repeat calls.

These results provide compelling evidence that repeat emergency calls fall by a large amount after an arrest. But this arrest effect can only be interpreted as breaking the cycle of domestic violence if it is not driven by changes in reporting behavior. DV is notoriously underreported, and arrest might affect the probability of reporting. If arrest encourages DV reporting, our estimates represent a lower bound of the arrest effect on repeat DV abuse. However, it is also possible that arrest discourages victims from reporting future DV incidents, in which case, the arrest effect we estimate could reflect a change in reporting behavior rather than a reduction in incidence. Given the importance of figuring out whether arrest changes DV incidence versus reporting behavior, we defer further robustness checks, heterogeneity analyses, and tests of mechanisms until after the next section where we tackle this issue.

[15] We first split the sample into subgroups based on combinations of these prior characteristics, recognizing that not all combinations are possible (e.g., there must be a prior case to have a prior arrest). We combine prior cases where there was an arrest but no investigation with prior cases where there was an arrest and an investigation, as the former is rare. This yields 6 different subgroups. We then estimate the first stage separately by subgroup so that we can calculate the fraction of compliers in each subgroup. Finally, we reweight the OLS estimation sample so that the proportion of compliers in each subgroup matches the share of the estimation sample for that subgroup.

## 2.5 A Reduction in Incidence or Reporting?

Whether the drop in repeat DV calls after an arrest is good or bad from a policy perspective depends on whether it is driven by a reduction in incidence or merely a change in reporting behavior. If there is a drop in abusive incidents after an arrest, victims are better off. If instead there is a drop in reporting without a reduction in actual abuse, victims are worse off because they are not getting help from police to resolve emergency situations, which in turn could embolden men to commit even more abuse.

In this section, we provide two tests based on the composition of repeat DV calls for whether the reduction reflects a drop in incidence versus a change in reporting behavior. Both tests point to a reduction in incidence as the primary explanation.

### 2.5.1 Test 1: Severity of Repeat DV Calls

Our first test is based on a simple threshold model for reporting behavior, with testable implications regarding the severity of repeat DV calls. The model builds on those used by Dahl and Knepper (2021) and Boone and Van Ours (2006) in the contexts of workplace sexual harassment and workplace safety, respectively.

In our threshold model, a woman will call the police if the immediate benefit of police intervention exceeds the retaliatory costs of reporting. Let $\theta$ be the retaliatory abuse a woman expects to occur if she calls the police. We model the benefit $\mu$ of calling the police to be a function of the level of abuse $\alpha$ perpetrated by a man against his partner, where we assume $\partial\mu(\alpha)/\partial\alpha > 0$. The idea is that in the heat of the moment, intervention by a police officer reduces harmful violence, and more so the more serious the abuse. Since benefits are increasing, we can define the threshold level of abuse above which a woman will report as the level of $\alpha = \bar{\alpha}$ that equates benefits to costs.

Now consider the impact of an arrest, ceteris paribus. On the one hand, an arrest could disempower victims from reporting DV in the future. This might happen if arrest causes backlash. In this case, the cost of calling the police in the future rises and hence the reporting threshold $\bar{\alpha}$ will be higher. The intuition is that women experiencing backlash after an arrest are more likely to remain silent, only reporting when abuse is so severe that the immediate benefit of police intervention outweighs the higher expected level of retaliation. On the other hand, an arrest could empower women to report by signaling to them that something is done about DV. In this scenario, the cost of calling the police decreases, and there will be a reduction in the reporting threshold, with women tolerating less abuse before calling the police. The testable implication of this model is that the composition of repeat DV calls after an arrest should be more severe on average if women are disempowered from reporting, whereas repeat calls should be less severe on average if empowerment dominates.

Extending the model to account for the optimal responses of men only reinforces these predictions. Men's optimal response in a world where an arrest disempowers victims from reporting is to increase the amount and severity of DV they commit. With empowerment, men will instead optimally decrease both the amount and severity of DV.[16] The model could also easily be extended to allow for a direct deterrence effect, where the perpetrator faces a higher cost of repeat DV after an arrest.[17] This would additionally contribute to decreasing the incidence of DV.

We use the call handler's grading of repeat DV calls to measure severity. We define severe repeat cases as those with a call grade of 1 (61% of repeat DV calls), i.e., those where there is a "danger to life/use or immediate threat of use of violence/serious injury to a person." We define less severe repeat cases as those with a call grade of 2 (29% of repeat calls), i.e., those where there is a "concern for someone's safety" plus call grades lower than 2 (10% of repeat calls), where the case is even less serious.

The first Column of Table 2.2 reports our baseline estimate for comparison purposes. Columns (2) and (3) decompose the baseline estimate by severity. Each Column is a separate regression which uses the entire sample and the same instrument, but Column (2) uses repeat calls which are severe as the outcome, while Column (3) uses repeat calls which are less severe. By construction, the estimates in Columns (2) and (3) will sum to the estimate in Column (1).

As Column (3) shows, there is a 55.2 percentage point drop in severe DV calls after an arrest. Relative to the estimated control complier mean of 83.0%, this is a sizeable 67% reduction. In contrast, Column (2) reveals a 6.4 percentage point *increase* in the number of less severe DV calls after an arrest. Relative to the

[16] One model that would yield this prediction is as follows. Consider two partners who play a sequential game with perfect information. After an initial DV incident, the male partner moves first and decides whether to repeat DV and, if so, chooses the severity $\alpha$ of it. The female partner observes these choices, and if subjected to repeat DV, decides whether to report according to the threshold model we have outlined. The utility function of the husband is of a piecewise form: $u = \phi$ if he decides not to repeat DV, where $\phi$ measure the utility he obtains from a violence-free relationship; $u = v(\alpha)$ if he repeats DV and the female partner does not report it ($\alpha < \bar{\alpha}$), where $v(\alpha)$ measures the utility he derives from it as a function of severity $\alpha$; and $u = v(\alpha) - c$ if he repeats DV and the wife reports it ($\alpha \geq \bar{\alpha}$), where $c$ is the cost to the husband of being reported to the police. Assuming that $v(\cdot)$ is a single-peaked function with a maximum at some $\alpha^*$, comparative statics on the sub-game perfect equilibrium of this two-stage game give rise to the predictions described in the text.

[17] This could be achieved by allowing the cost parameter $c$ in footnote 16 to be higher after a prior arrest.

control complier mean of 13.2%, this is a 48% rise. These compositional effects are consistent with a drop in victim's reporting thresholds and an accompanying decline in the level of abuse. Backlash would have predicted a compositional shift to more severe cases being reported, which is exactly the opposite of what we find. A formal test rejects the null hypothesis of backlash versus the alternative of empowerment (p-value=0.036).[18] Since both the volume and severity of repeat calls fall after an arrest, we conclude that the true incidence of repeat DV falls after an arrest.

### 2.5.2 Test 2: Who Reports Repeat DV Calls

Our second test is based on the composition of who reports repeat DV. Using a similar threshold model, if victims are more subject to backlash than neighbors, then the composition of future calls after an arrest should be skewed towards those being reported by third parties rather than the victim. The intuition is that victims facing backlash will remain silent for fear of retaliation, while such concerns should be smaller for neighbors as they do not live with the abuser and can report anonymously. In contrast, if victims are empowered relative to their neighbors, the composition of future calls should tilt towards victims.

In Columns (4) and (5) we test these predictions by decomposing the baseline estimate into who reports. Column (4) shows that repeat DV calls by a victim fall by 9.9 percentage points after an arrest, which is 36% drop relative to the estimated control complier mean. Column (5) reveals a proportionately larger decrease in repeat DV calls by a third party. There is a 39.0 percentage point reduction, which is a 57% drop relative to the control complier mean.

The fact that repeat calls by third parties fall by 57%, while victim calls fall by only 36%, is the opposite of what backlash would predict. Instead, the compositional shift is consistent with empowerment, with actual abuse falling and women reporting abuse at a lower threshold. Since the overall and relative volume of victim repeat calls fall after an arrest, this is consistent with a decline in true incidence after an arrest. However, we cannot reject the model of backlash at conventional significance levels (p-value = 0.354).[19]

[18] We conduct a one-sided test of the null hypothesis of backlash versus empowerment by restricting the sample to cases with a repeat call, defining the outcome to be a dummy variable for a severe case, and using our preferred IV specification. The estimated effect of an arrest is -0.452 (s.e.=0.251).

[19] Using a similar test as in footnote 18, the estimated effect of an arrest on the outcome of a third party report is -0.105 (s.e.=0.280).

*Table 2.2:* Testing for a Reduction in Incidence versus Reporting

| | Dependent variable: | | | | |
|---|---|---|---|---|---|
| | **Repeat call for DV** | **Low severity repeat call** | **High severity repeat call** | **Victim-initiated repeat call** | **Third party-initiated repeat call** |
| | (1) | (2) | (3) | (4) | (5) |
| Arrest | -0.488*** | 0.064 | -0.552*** | -0.099 | -0.390** |
| | (0.187) | (0.150) | (0.173) | (0.157) | (0.190) |
| Mean of dep. var. | 0.492 | 0.192 | 0.300 | 0.182 | 0.311 |
| Control complier mean | 0.962 | 0.132 | 0.830 | 0.275 | 0.687 |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: Column (1) repeats the baseline specification from Table 2.1 Column (4). Columns (2) and (3) decompose the baseline estimate by the severity of repeat calls. Low severity repeat calls are defined as those with a call priority grade of 2 or lower. High severity repeat calls are those with a call priority grade of 1. Columns (4) and (5) decompose the baseline estimate by the identity of the caller. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

## 2.6 Mechanisms

In this section, we investigate factors which could explain the results we document, finding that (i) arrest has a short-term incapacitation effect as well as a longer-term deterrent effect and (ii) arrest leads to consequential legal sanctions for the offender.

### 2.6.1 Short-Term and Longer-Term Effects of Arrest

In the first three Columns of Table 2.3, we explore the short-term effects of arrest. The time window we focus on in Column (1) is the first 96 hours after an arrest. Two pieces of evidence highlight the importance of arrests in the short-run. First, the control complier mean for a repeat DV call within 96 hours is 25 percentage points. Second, and strikingly, if a suspected batterer is arrested on the scene, the probability of a repeat DV call within 96 hours falls by 20 percentage points, which amounts to a 79% reduction.

One possibility is that this sharp drop is due to an incapacitation effect. In Columns (2) and (3) we further explore this by splitting the 96-hour time window into hours 1-48 and hours 49-96, respectively. To get some insight into a potential incapacitation effect, hours 1-48 are of particular interest, since this is when offenders are typically under investigation and are potentially placed in temporary custody. Comparing the IV estimates in Columns (2) and (3), we see that almost the entire short-term effect of arrest is explained by a sharp drop in repeat DV calls in hours 1-48. We interpret this as evidence that an important part of the arrest effect is driven by short-term incapacitation. Given that DV appears to be disproportionately common in the first few days after an initial emergency call, this suggests the ”cooling off” period provided by an arrest is highly effective.

Columns (4) to (6) of Table 2.3 investigate the longer-term effects of arrest. As the dependent variable, Column (4) uses a dummy for whether there is at least one repeat DV case within 12 months, excluding the first 96 hours. An arrest leads to a statistically significant 45 percentage point decrease in repeat calls during this time period. Note that the arrest effects in Columns (1) and (4) do not need to sum to our main estimate appearing in Table 2.1. The reason is that repeat victimization can occur in more than one time period. In Columns (5) and (6), we focus instead on whether repeat cases are reported in months 1-6 (again omitting the first 96 hours) and months 7-12, respectively. For both time windows, we find that arrest reduces the probability of a repeat call by roughly 50% of the control complier mean, though the estimate for repeats in months 7-12 is not statistically different from zero.

Thus, besides causing a strong short-term incapacitation effect, arrest appears to persistently reduce repeat calls both 1-6 and 7-12 months subsequent to the initial incident.

### 2.6.2 Criminal Sanctions

Sceptics of arrest argue that it will not have a deterrent effect due to its weak "dosage" of punishment, with offenders and victims learning that the consequences of an arrest are not very serious.[20] Advocates for a deterrence effect argue that arrest makes salient the seriousness of the crime, in part due to a higher probability of criminal sanctions.

If a DV case is formally investigated, the investigative officer coordinates with the prosecutor to determine whether the offender should be formally charged with a crime and summoned to court. Non-arrested compliers have an estimated 2% probability of being charged. As Column (7) in Table 2.3 shows, an arrest increases this probability by a statistically significant 10.4 percentage points, which is a five-fold increase. Criminal charges can lead to a conviction with the possibility of jail time, community service, or parole, although we do not observe these outcomes in our data.[21] Based on the large increase in criminal charges, combined with the persistent reduction in repeat calls shown in Columns (4)-(6) of Table 2.3, we conclude that arrest has a sizeable deterrence effect.

[20]Related, one of the most commonly cited reasons for not reporting DV is the perception that the police do not take cases seriously enough or have no means of providing help. See National Research Council (1998) and Fleury, Sullivan, Bybee, and Davidson II (1998).

[21]Having a criminal record is consequential in the UK, as employers are are allowed to ask about recent convictions on job applications. Punishment for future crimes is also influenced by prior criminal sanctions.

*Table 2.3:* Mechanisms

| | Dependent variable: **Repeat call for DV in the specified time frame** | | | | | | Dependent variable: **Criminal Charge** |
|---|---|---|---|---|---|---|---|
| | within 96 hours | in hours 1-48 | in hours 49-96 | within 12 months (excl. hours 1-96) | in months 1-6 (excl. hours 1-96) | in months 6-12 | filed by a prosecutor |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Arrest | -0.198* | -0.197** | -0.025 | -0.450** | -0.323* | -0.244 | 0.104** |
| | (0.104) | (0.099) | (0.055) | (0.184) | (0.185) | (0.178) | (0.053) |
| Mean of dep. var. | 0.069 | 0.054 | 0.018 | 0.469 | 0.364 | 0.273 | 0.014 |
| Control complier mean | 0.251 | 0.233 | 0.043 | 0.909 | 0.684 | 0.507 | 0.020 |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

## 2.7 Measurement, Exclusion Restriction, Robustness, and Heterogeneity

We now return to the task of probing (i) the robustness of our outcome measure, (ii) the exclusion restriction, (iii) specification checks of our main results, and (iv) heterogeneous effects of arrests.

### 2.7.1 Using Geo-Coordinates to Measure Repeat Calls

We measure repeat DV emergency calls based on whether a new DV case with the same geo-coordinates is reported within 12 months of the initial call. The measure is potentially prone to misclassification errors. We might over-assign repeat cases if several households live at a given geo-location, or under-assign repeat cases if victims move away from it. In this section, we assess whether these two types of misclassification error are likely to bias our estimates.

To minimize the chances of over-assigning repeat cases, we exploit features of the built environment in the sample area. High-rise buildings with multiple apartments in the same geo-location can lead to DV cases being reported by different victims. This problem should be less of a concern if we focus on areas largely made up of detached houses, which is the most common type of accommodation in the region we study. In Column (1) of Table 2.A.6, we restrict our sample to wards with at least 80% detached houses. Although the sample shrinks by a third, the arrest effect remains largely unchanged compared to our main estimate in Table 2.1. In Column (2), following a similar argument, we exclude the city center of Birmingham - the largest city in the county - by excluding areas within a 3 km radius of the city center, where mostly larger buildings are located. The estimate for arrests remains close to the baseline estimate, consistent with over-assignment of repeat cases not confounding our results.

A second potential bias comes from under-assigning repeat cases. A particular concern is that victims of DV might be more likely to make residential moves after an arrest, but continue to be victimized in their new location. To assess this concern, we check whether the accuracy of our geo-coded repeat variable is affected by an arrest. To do so, we proceed in two steps. We first draw on the subsample of our data where we observe unique victim IDs (the 59% of cases where a criminal investigation occurs), and construct a dummy for re-victimization based on their victim ID. We then compare this variable to our geo-coded re-victimization measure for the same subsample. This allows us to work out how many victim-ID based repeat cases we miss with our geo-based definition. Overall, 91% of the no-repeats based on our geo-coded measure turn out to also be no-repeats based on the victim-ID measure, while 9% of our geo-coded non-repeats miss that the victim is actually re-victimized. Importantly, the false negative rate is not significantly different across cases with and without a prior arrest, implying that the misclassification error is not correlated with our treatment variable. Table 2.A.7 summarizes this sensitivity check.

### 2.7.2 Exclusion Restriction

The IV estimate requires the exclusion restriction that first response teams with higher arrest propensities only affect repeat DV calls by increasing the probability of an arrest. In this section, we provide a series of exclusion tests related to (i) other actions taken by responding officers as part of their on-scene police work and (ii) characteristics of responding officers.

The on-scene actions by first response teams are multidimensional in nature. The first priority of responding officers is to make the victim safe, which may mean arresting a suspect if it is considered a necessary and proportionate response. In addition, responding officers are also tasked with taking steps to build a case for a potential evidence-led prosecution. This includes, *inter alia*, collecting evidence, convincing victims to cooperate, and communicating with other involved parties (e.g., witnesses). In this setting, a violation of the exclusion restriction would arise if officers with higher arrest propensities are better at building evidence-led cases. This issue can be addressed by augmenting our IV model to include a measure for the quality of a team's on-scene work as an additional endogenous regressor plus an extra instrument for it. We proxy for the quality of a team's on-scene work using whether the case ends up being formally investigated. Indeed, how well responding officers carry out their on-scene tasks is a critical input into whether an investigative officer opens a formal investigation and enters the case in the national crime database (College of Policing, 2022).

We construct an instrument for formal investigations that is similar to our instrument for arrest: the propensity of a first response team's cases to be formally investigated. There is independent variation in formal investigations and arrests. Formal investigations happen much more often (59% of the time versus 3% for arrest), and an arrest is only followed by a formal investigation 67% of the time. As Column 2 in Table 2.A.8 shows, the instrument for formal investigation strongly predicts whether an investigation will happen. Table 2.A.4 reports IV estimates, with Column 1 repeating our main specification for arrests and Column 2 adding in formal investigations. Formal investigations reduce repeat DV calls by a modest, and statistically insignificant, 5 percentage points. More importantly, our main arrest effect is virtually unchanged. Thus, our arrest finding does not appear to be driven by how effective police are at doing their job.

One directly observable part of the work that first response officers do is how much time they spend on the scene. We use this variable for a second test of the exclusion restriction. A plausible interpretation of extra minutes spent is that officers are more diligent, take the needed time to finish their tasks, and spend more time with victims and offenders. As above, the concern is that officers with higher arrest propensities could influence recividism not only through arrest but also by how much time they spend on the sence. In Column 3 we include time on the scene as an additional endogenous variable, instrumenting it with a team's average time spent on the scene in other cases.[22] While the instrument is strongly predictive (Table 2.A.8, Column 4), time spent on the scene does not affect repeat DV calls, nor does it change the estimated effect of arrest (Table 2.A.4, Column 3).

In our dataset, response teams are recorded as taking one of four actions (in order of decreasing severity): arrest, recommend an investigation, provide advice, or do nothing. So far, we have been using arrest as our key independent variable and instrumenting for it. As a third test of the exclusion restriction, in Column 4 of Table 2.A.4 we include variables for these other actions, instrumenting for them using officer team propensities along each of these margins. While the data only records the most severe intervention, it typically includes the less severe actions as well. Therefore we code the action of arrest as also including recommend investigation and provide advice, and the action of recommend investigation as also including provide advice.[23] The arrest estimate remains virtually unchanged, while the coefficients for recommend investigation and advice are statistically insignificant. In Column 5, we add in all of the endogenous regressors found in Columns 2-4 simultaneously, and the results remain unchanged.

As a fourth, and final, test of the exclusion restriction, we probe whether team characteristics could cause a violation of the exclusion restriction. A prime example is the gender composition of response teams, as female officers are thought to increase the quality of DV-related police work (Miller and Segal, 2019). Experience could also

[22]The time on scene variable is missing in 30.8% of cases. It appears to be missing at random, as the team's leave-out average time spent on the scene regressed on whether the variable is missing yields a precise zero (controlling for call grade, time, and ward fixed effects). We therefore replace missings with the mean time spent in non-missing cases, and add an indicator for missing in the regression.

[23]If we alternatively code up arrest, recommend investigation, and provide advice as three non-overlapping actions, the results are similar.

matter. This source of bias can be eliminated by controlling for team characteristics. We observe both the gender and age of officers. Thus, in Column 6 of Table 2.A.4, we control for the fraction of females on the team and the average age of officers. This does little to change the estimated effect of arrest. As an alternative to this test, in Table 2.A.9 we instead control for the additional team propensities used as instruments in Table 2.A.4. This also does not appreciably affect our arrest estimate.

### 2.7.3 Alternative Specifications

In Table 2.A.10, we probe the robustness of our result to alternative specifications. Column (1) changes the definition of the outcome variable, focusing now on the intensive instead of the extensive margin. The new outcome is the *number* of repeat DV calls within 12 months, and we use the same set of controls as our baseline specification in Column (4) of Table 2.1. The point estimate indicates a statistically significant drop of 1.5 calls after an arrest. Relative to the estimated control complier mean of 2.9 calls, this is roughly a 50% reduction. This intensive margin result closely mirrors the magnitude of the effect for the extensive margin.

The next two Columns explore robustness to changing how the instrument is constructed. Our estimation sample is limited to current cases with a call grade of 1 or 2. The rationale is that call grades of 1 or 2 must be dealt with immediately, with the closest available officers sent to the scene and hence leaving little room for non-random officer assignment. However, to construct the instrument, we do not require a randomized set of cases for validity. Therefore in our baseline specification, we use all call grades to construct the instrument. In Column (2), we report estimates where we limit the construction of the instrument to calls with a priority grade of 1 or 2. The estimate is virtually identical to our baseline estimate. In Column (3), we define the instrument using the traditional leave-one-out measure which only excludes the current case (instead of excluding all calls from the same geo-location, whether it is the current case or any other case). This does not appreciably change the estimate.

In our main specification, we include any DV emergency call for which the assigned response team has handled at least 400 other DV cases. Columns (4) and (5) investigate what happens if we instead require response teams to have handled at least 300 cases or at least 500 cases, respectively. When applying these alternative thresholds, the results remain qualitatively similar.

### 2.7.4 Heterogeneous Effects

In Table 2.A.11 we explore whether there are heterogeneous effects by case characteristics. In Column (1), we present IV results that distinguish between emergency DV calls with and without a prior DV investigation. We construct a dummy variable for whether a call has been preceded by a formally investigated DV case in the past 12 months and interact it with our arrest variable. There is some evidence that the effect is larger for calls with a prior investigated case, but the difference is not statistically significant. Column (2) similarly divides the sample into two groups, but now based on the predicted probability of an arrest estimated using the case characteristics appearing in Table 2.A.2. The arrest effect is fairly similar across calls with high versus low arrest propensities.

The next two Columns show IV estimates broken down by the characteristics of response teams. In Column (3), we distinguish between all-male response teams and teams with at least one female officer. The arrest effect is similar for both types of response teams. Column (4) breaks the sample down into calls handled by response teams whose average age (a proxy for experience) is above or below the mean. Although the statistically significant point estimate for younger teams is larger than the non-significant estimate for older teams, the two are not statistically different from each other.

## 2.8 Conclusion

Our findings provide compelling evidence that arrest helps break the cycle of domestic violence. Using a rich dataset and quasi-random variation, we find that an arrest reduces future DV calls in the following year by 51%. We provide empirical evidence that the reduction in calls is not driven by a change in reporting behavior due to a fear of retaliation, but rather a decline in repeat victimization. In terms of mechanisms, we find that arrest virtually eliminates the large spike in re-victimization which occurs in the 48 hours after an emergency call, suggesting arrest facilitates a cooling off period during a volatile, at-risk time. Arrests also result in a 5-fold increase in the probability an offender will by charged with a crime, consistent with the longer-run deterrence effect we document.

These findings argue against recent calls for a complete decriminalization of domestic violence. In our setting where the arrest rate is low, the optimal policy response is to arrest more suspected batterers if the objective is to reduce future abuse. We caution, however, that our results do not necessarily imply that arrest should occur in all cases and in all settings; for example, in countries where the arrest rate is high, the pendulum could well have swung too far in the other direction. Future research for other countries and in other contexts can help shed light on this issue.

# 3

# Confronting Prejudice: Uncovering Stereotypes Among Police Officers in India*

Abstract

This paper examines the effects of prejudice confrontation on police officers' handling of gender based violence cases (GBV) in India. In collaboration with the Madhya Pradesh Police, we ran a lab-in-the-field experiment where officers were randomly confronted with their mishandling of a GBV case. We find that the confrontation leads female officers to place more emphasis on the victim's account when handling a GBV case. Men, on the other hand, tend to have the counter-reaction. We propose that an explanation for this finding could be the fact that female police officers perceive a bias in their male-dominated work environment that is more severe than their own. They thus de-bias their case handling and adjust their professional beliefs. For male confronted officers, we observe behavior consistent with backlash that is driven by strongly biased men.

*This Chapter is based on joint work with Sofia Amaral (World Bank), Kim Chaney (University of Connecticut), Nishith Prakash (Northeastern University) and Abhilasha Sahay (World Bank). I appreciate the collaboration of the Madhya Pradesh Police. The views described in this paper do not necessarily reflect those of the Madhya Pradesh Police. This study received bioethics approval from the University of Connecticut under the protocol X21-0091 approved May 21, 2021. I thank Vishakha Wadhwani and Asmi Khushi for their fieldwork and their superb work as research assistants. I appreciate the valuable feedback from Prashant Bharadwaj, Michael Kaiser, Paul Niehaus and Helmut Rainer.

## 3.1 Introduction

Gender-based violence (GBV) remains widespread in most parts of the world with low-to-middle income countries being the most affected (Sardinha, Maheu-Giroux, Stöckl, Meyer, and García-Moreno, 2022).[1] Police response is a critical component in reducing GBV. Bias among police officers can deter people from reporting incidents of GBV (García-Moreno, Zimmerman, Morris-Gehring, Heise, Amin, Abrahams, Montoya, Bhate-Deosthali, Kilonzo, and Watts, 2015), and when GBV incidents are reported, they may be dismissed because police officers view such victims as dishonest, discriminate against them, or blame them for the incident. Addressing GBV police bias is, therefore, a necessary condition to ensure equal access to justice and equal participation of women and girls in society.

Confronting people with their prejudices is an effective tool for reducing bias and changing behavior among perpetrators in contexts of racism and sexism (Alesina, Carlana, La Ferrara, and Pinotti, 2018; Mallett and Wagner, 2011).[2] After a prejudice confrontation, people strive to be more egalitarian and self-regulate their bias (Chaney and Sanchez, 2018). However, this research has overwhelmingly focused on prejudice in high-income countries where cultural norms signal egalitarianism (Jayachandran, 2015). When people receive feedback that threatens their self-image, they may also engage in defensive strategies such as denying, downplaying, or dismissing the feedback (e.g. Howell, Redford, Pogge, and Ratliff, 2017). This defensiveness could increase peoples' bias.

Against this background, this study examines a novel intervention of prejudice confrontation in the context of GBV and Indian police officers. We investigate prejudice confrontation and its effects on individuals' bias in a setting where norms of egalitarianism are weak or nonexistent. To the best of our knowledge, the present research is the first that looks at prejudice confrontations in the workplace, and in the context of state actors' GBV victim-blaming stereotypes. We cooperated with Madhya Pradesh Police (India) to design and implement a randomized controlled

[1] Gender-based violence can include sexual, physical, mental, and economic harm, threats of violence, coercion, and manipulation. While GBV can affect people of all genders, rates of GBV against women and girls are particularly high (`https://ncadv.org/STATISTICS`). GBV can take place in private or in public and takes many forms that can include intimate partner violence, sexual violence, child marriage, female genital mutilation, and so-called "honor crimes" (Degener and Koster-Dreese, 1995). Reliable cross-country estimates are not available for most forms of GBV, yet, (Sardinha, Maheu-Giroux, Stöckl, Meyer, and García-Moreno, 2022) estimate that worldwide 27 percent of women are victims of intimate-partner violence.

[2] Prejudice confrontation are defined as verbal challenges directed at those who commit a blatant, subtle, or unspoken act of discrimination (Chaney, Young, and Sanchez, 2015; Czopp, Monteith, and Mark, 2006).

trial (RCT) in the form of a lab-in-the-field experiment.[3] Participating officers reviewed two cases, including one that involved a woman reporting GBV, and one non-GBV case, and answered a computer-based survey on how they would handle the cases. Officers were randomly assigned to a treatment condition where a high-ranking police officer confronted them about their bias in dealing with victims of GBV after the officers completed the survey. The confrontation consisted of a private in-person conversation with the senior officer. Officers in the control condition received neutral feedback on the non-GBV case. One week later, all officers reviewed a GBV case again and solved a computerized stereotyping reaction time task in which pictures of potential victims were shown and they were asked to rapidly categorize descriptions that might apply to the victim or not.

This design allows us to understand future responses to GBV crimes by police officers after a confrontation with prejudice in two dimensions. First, how do confronted officers change their behavior in handling GBV cases? Second, to what extent does the victim's appearance cause confronted officers to react in a biased manner? Notably, we contend that female police officers are more likely than male officers to make egalitarian changes in their behavior after being confronted with their own stereotypes concerning GBV victims because of their shared gender identity with the predominantly female victims of GBV. Yet, as women make up only a small percentage of police officers in India (13%), they may feel pressured to conform to prevailing workplace norms, which could ex-ante lead to a stereotypical response[4]. Therefore, a second objective of this study was to examine differences in the impact of confronting prejudice between male and female police officers.

Our main finding is that - while there is no statistically significant aggregate effect of the confrontation treatment - the confrontation treatment does indeed lead female and male officers to react differently: Female officers respond less stereotypically in handling a GBV case, which means they place greater emphasis on the victim's account. Men tend to react in the opposite way. Females prioritize the victim's statement by 23 percentage points more than the control group, which roughly corresponds to a 27% change. Consistently, we observe a negative effect on the prioritization of the defendant's statement for female officers. Confronted females also have an 8 percentage points higher probability of pursuing a GBV complaint than the control group. For male officers, there is no statistically significant effect for dropping a complaint. Yet, they seem to push back in the further course of the

[3]The study is pre-registered at AEA RCT Registry. ID: AEARCTR-0008611.

[4]National Crime Records Bureau (Ministry of Home Affairs). 2018. Crimes in India 2018 Statistic. Volume I. Government of India.

investigation. After a confrontation, they put less priority on the victim's statement and more on the offender's statement. We find no effect on outcomes, that reflect officers' general disposition toward GBV, such as the belief in the truthfulness of GBV complaints and registering a GBV complaint in the first place. These findings are robust to a battery of robustness checks, including multiple hypotheses testing and alternative specifications.

A potential explanation for this finding is that female officers predominantly (72%) show mild bias in handling a GBV case before treatment. Yet, more than half of the male officers (51%) exhibit a strong bias.[5] Considering that policing is highly male-dominated, the average female officer thus perceives a bias in her work environment that is more severe than her own. The confrontation treatment leads female officers to de-bias their case handling and adjust their beliefs in a professional context. Male officers show behavior consistent with backlash that is driven by strongly biased officers. When we exclude strongly biased male officers from the analysis, we find that for our main outcome (prioritization of the victim's statement), the negative effect for male officers effectively vanishes while the positive effect for females persists.

We additionally conduct a computerized so-called 'negative stimuli reaction task', in which we assess the extent to which the victim's appearance causes officers to react in a biased manner. A similar pattern of results emerges. For male officers, we find a significantly greater use of GBV stereotypes after a confrontation. We do not find such an effect for female officers. This finding is consistent with our results concerning the GBV case handling. For men, prejudice confrontation seems to trigger a backlash. Women do not change their inner beliefs, but de-bias their case handling after being encouraged to do so by the confrontation feedback.

Taken together, the intervention has had the effect of encouraging women to de-bias their actions, which means less stereotyping when working on GBV cases. Men, on the other hand, are less receptive to the feedback. Policymakers may have to consider that a prejudice confrontation may affect female and male officers differently and result in adverse responses.

In the existing literature on prejudice confrontations, it has been demonstrated that they can reduce bias and change behavior. For example, White Americans confronted for using negative Black stereotypes are less likely to use them 1-week later (Chaney and Sanchez, 2018). Similarly, men confronted for relying on female

[5] An example of a strong bias would be the filing of a report against the victim. We consider it a mild bias if, for example, the officer places an investigation against the victim among the top three (of 5) important investigative parts.

gender role stereotypes engaged in compensatory behavior (Mallett and Wagner, 2011) and reported more favorable attitudes towards women compared to men who were not confronted. Outside of lab experiments, e.g. Alesina, Carlana, La Ferrara, and Pinotti (2018) found that confronting teachers' biases against immigrants led to an increase in grades teachers assigned to immigrants. This research has predominantly focused on high-income countries. With this paper, we add to the literature by investigating prejudice confrontations in the context of police officers in India.

The paper also relates to empirical research that investigates the drivers of GBV. Scholars have linked GBV to social norms (Green, Wilke, and Cooper, 2020; Bandiera, Buehren, Burgess, Goldstein, Gulesci, Rasul, and Sulaiman, 2020), cultural factors (Guarnieri and Rainer, 2021; Tur-Prats, 2019), labor market conditions (Aizer, 2010; Anderberg, Rainer, Wadsworth, and Wilson, 2016), liquidity constraints (Hidrobo, Peterman, and Heise, 2016), divorce legislation (Stevenson and Wolfers, 2006), and emotional cues (Card and Dahl, 2011). However, research regarding the prevention of GBV is limited. In fact, there is little evidence focusing on the role of state actors. The police and courts are arguably a necessary institution to address GBV, as their actions determine women's willingness to turn to formal support services (Palermo, Bleck, and Peterman, 2014), sanctioning, and affect deterrence (Amaral, Dahl, Hener, Kaiser, and Rainer, 2023). Our work contributes to this strand of the literature by examining how addressing police bias can change police officers' discrimination and behavior toward violence against women.

The remainder of this chapter is structured as follows. The next section provides information about the police force we collaborated with, the participants of the experiment, randomization, and the experimental protocol. In Section 3.3 we discuss our empirical approach and balance tests. Section 3.4 reports our main results and robustness checks. Section 3.5 concludes.

## 3.2 Experimental Design

In this section, we describe the context, the participants of the experiment, and randomization. We then outline the procedure of the two stage lab-in-the-field experiment.

### 3.2.1 Context

The experiment was conducted in the state of Madhya Pradesh in India. This state ranks second (out of 29) in sexual assault homicides and third in dowry deaths.[6] The police play a critical role in addressing and combating GBV, particularly when considering the limited availability of specialized services to tackle it. According to the 'IPF Citizen Satisfaction Survey on SMART Policing' in India, the police in Madhya Pradesh have received notably low ratings. In the specific category 'Fair, unbiased and lawful policing' the state has been positioned at the 24th rank (out of 29), while in terms of 'Integrity and corruption free service' it has reached the 27th rank.[7] The civil police in Madhya Pradesh consists of roughly 9,000 police officers in the ranks of Sub-inspector (SI) and Assistant Sub-inspector (ASI). These officers typically handle crime reports at the early stages of investigations. About 13% of them are female.[8] The police is a very hierarchical organization and senior officers are held in high esteem by their junior colleagues. Thus, a confrontation with one's own bias by a senior in a police environment can be particularly effective.

### 3.2.2 Participants

A total of 323 officers were recruited to complete a two-part experiment.[9] Officers were invited from the districts of Bhopal, Harda, Hoshangabad, Raisen, Sehore, and Vidisha in Madhya Pradesh, India. Officers were recruited via a two-stage process. First, each station received a communication to inform about the study and provided a list of officers eligible to participate. Second, eligible officers were invited directly via a phone call. Only officers who were available to attend sessions were ultimately

[6] National Crime Records Bureau (Ministry of Home Affairs). 2018. Crimes in India 2018 Statistic. Volume I. Government of India.

[7] IPF smart policing survey 2021.

[8] Bureau of Police Research and Development (Ministry of Home Affairs). Data on Police Organization 2021.

[9] The recruited officers were either of the rank Sub-inspector (SI) or Assistant Sub-inspector (ASI). These types of officers were selected since they would typically handle crimes - including GBV reports- at the early stages of investigations.

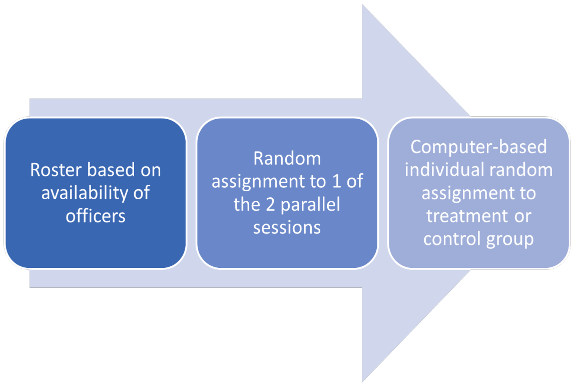



*Figure 3.1:* Randomization of Participants

part of the sample. Our approach involved inviting on average two officers per station in order to minimize spillover concerns between treated and control officers.[10] The participants ultimately consisted of 58 (20%) female and 239 (80%) male officers in an age range of 24 to 61 years (mean age: 42). Participants were invited for lunch on site and officers' time spent on the research was considered on-duty time.

### 3.2.3 Randomization

The experiment took place in November and December 2021. The selection of police officers to participate in the experimental days was contingent upon their scheduling availability. Two parallel sessions were held per day. The invited police officers were randomly divided into two groups on site (this division was achieved by instructing them to form a line and alternate between saying the numbers 1 and 2). The formation of these two groups was for purely organizational reasons since the computer rooms (see next) were not large enough for one large group. Subsequently, the lab-in-the-field experiment took place with these two groups, each occupying a separate computer room. The process of individually assigning participants to either the treatment or control group was integrated into the survey software and executed automatically on the computers where the officers sat (see Figure 3.1).

### 3.2.4 Procedure

**Baseline.** The experiment relies on a between subjects design. First, after the participants had given their consent, they reviewed one neutral, non-GBV, case fol-

[10]We approached 360 officers directly via phone, thus, we achieved a response rate of 90 percent. Officers' unavailability was mostly due to being on duty, or unexpected local events that were of higher priority. For this reason, we expect that there is minimal difference between officers taking part in the experiment and those who did not.

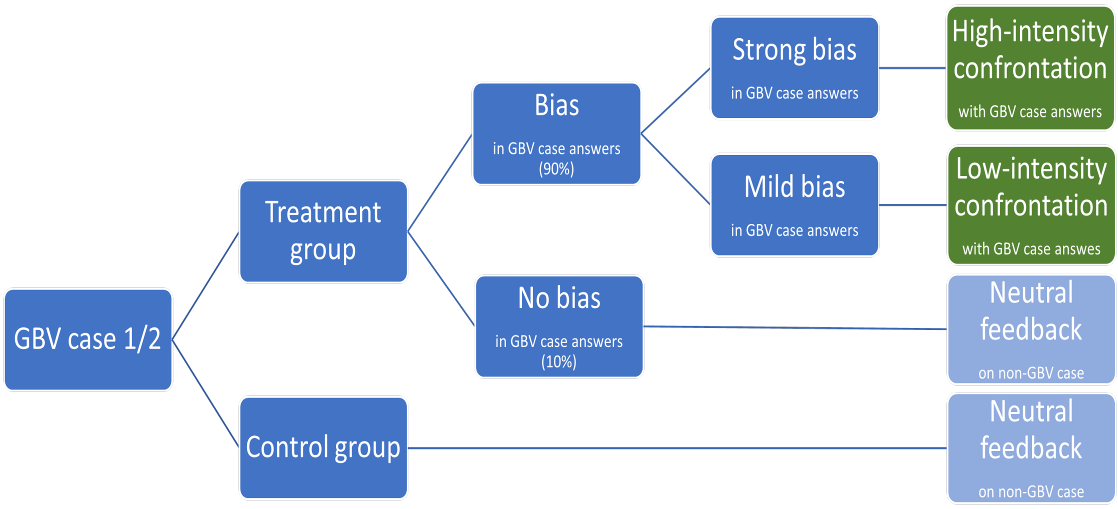



*Figure 3.2:* Procedure at Baseline

lowed by one GBV case in a computer-based survey format. Participants completed several questions involving the cases. Starting with whether they think a crime has been committed, and then how they would proceed in the investigative step by step. There were yes and no questions included, and questions where the officers were asked to rank parts of the investigation or select answer options. Based on these case-related questions, we constructed our key outcome variables: Whether the crime is detected, whether the officers decide to officially register the complaint, whether they follow up or drop the complaint, and whether they rank parts of the investigation that result from the victim's or offender's statement to be most important. The neutral case involved an instance of property crime. The GBV case was randomly selected from two. We included two GBV cases to ensure the generalizability of findings following a confrontation. Thus, while one GBV case involved domestic violence, the second included a case involving dowry. The random assignment of one of the two GBV cases was built into the survey software and automatically executed on the computers where the officers sat. All cases were developed by the research team to mimic real cases reported to the police in this region and included multiple people and claims in order to mirror the complicated, multi-party nature of real cases. They were extensively pilot-tested in separate samples of officers.

Second, after the officers completed these two cases, they notified the field staff that they had completed the initial computer-based survey and waited to enter a 1-1 session with a senior officer in a separate room.[11] When called into the room with the senior officer, participants received feedback based on their randomly assigned condition. Officers in the control condition and those in the treatment condition who demonstrated no bias (10%) in their handling of the GBV case received neutral feedback unrelated to GBV (see Figure 3.2).[12] In contrast, participants who were randomly assigned to the treatment condition and demonstrated bias in their handling of the GBV case received confrontation feedback. This confrontation was adapted slightly depending on which of the two GBV cases they completed and also depended on the extent of the bias shown in their answers.[13] Senior officers were instructed to deliver all feedback and confrontations in a neutral and non-aggressive fashion, focusing on an educational approach. Participants were given a chance to respond to this feedback, though the interaction time was cut off by field staff after two minutes. The feedback and response were audio recorded, and a member of the research team was present in the room during this interaction.

Third, after the feedback session, participants were directed back to a computer to complete a brief survey which included measures of negative feelings toward oneself or the senior officer ('negative self- and other-directed affect'; Chaney and Sanchez 2018; Czopp, Monteith, and Mark 2006), as well as measures of their respect for the senior officer, how they believe they performed, and demographic informa-

[11]We recruited four male senior officers to deliver the treatment. This choice was motivated by two factors. First, the experiment was intended to mimic officer's interactions while on duty. Second, in our context, due to historical factors, the police is a reserved institution. As a result, having external parties or researchers conduct the confrontation would be very challenging and likely ineffective. Since the police is a very hierarchical organization, senior officers are highly respected by their junior peers. Senior officers were recruited in the same fashion as the SI and ASI officers (via formal letter and phone calls), and also by proximity to the location of the experiment. Senior officers were not the direct managers and supervisors of the SI and ASI officers who participated in the experiment. This allows us to mitigate any potential backlash since officers operate in different jurisdictions.

[12]In a neutral feedback session, the senior officer informed the participants that in general their way of handling the cases was fine but that they 'did not pay proper attention and time to the case while answering' and '[...]when you solve a case on the computer, it is important to be careful while responding'.

[13]An example of a strong bias would be the filing of a report against the victim. We consider it a mild bias if, for example, the officer places an investigation against the victim among the top three (of 5) important investigative parts. If officers received the high-intensity confrontation, the senior officer stated '[...] it looks like you sided with the defendant's mother even though somebody else is the victim who came up with the complaint of domestic abuse. It seems like your actions were influenced by biased beliefs against women in society. Police officers cannot do their job based on their assumptions [...]'. Participants who demonstrated more mild bias during their response to the GBV case received the low-intensity confrontation, during which the senior officer stated: '[...] during the investigation you prioritized the statements of the defendant and his friends [...]. It seems that in this case your action was influenced by the prevailing beliefs against women [...]. Police officers cannot do their job based on their perceptions.'

tion. Participants were then dismissed from the session and informed they would return in 1 week.

**Endline.** One week later, participants returned to complete a computer-based survey and a measure of automatic stereotyping. The attrition rate was 8%, which we consider low and did not result in a selected sample.[14] The survey included a novel GBV case (again, developed by the research team) with questions regarding how participants would handle this case. After, participants completed a measure of reflection about their handling of GBV cases, perceived truth of GBV complaints, interest in future training, and several secondary outcomes (perceived norms about GBV case handling, attitude towards women, and empathy).

After completing the survey section, participants completed a computerized stereotyping reaction time task. In this task, participants were presented with various pictures of male and female faces in the center of the computer screen, below which was a sentence representing a report to the police, such as 'This woman reported molestation by a group of boys at the bus stop near her college'.[15] After the image of the face and complaint was shown for 8 seconds, nine words appeared on the screen above the image, and participants had the task of categorizing these words as 'Applies to this person' or 'Does not apply to this person' by pressing the corresponding key. These categories were presented at the top of the screen as a reminder for participants. Each word appeared on the screen for 2.5 seconds or until a response was recorded. The nine words were randomly pulled from a list of six victim-blaming stereotype words (e.g., liar, at fault, manipulative) and six innocent words (e.g., innocent, good, moral). Note, participants first completed two practice trials and then completed 40 test trials. Of the 40 trials, 8 included male targets reporting non-GBV complaints, 16 women targets reporting non-GBV complaints, and 16 women targets reporting GBV complaints. The two outcomes constructed based on this stereotyping task are counts of clicks that demonstrate a denial of stereotypes in the GBV and non-GBV context, respectively (i.e. how often innocent words were selected for victim images).

[14] The reason for absence was work-related unavailability and the characteristics of officers that dropped out do not differ from those of the entire sample.

[15] We modified a stereotype inhibition task (adapted from Chaney and Sanchez 2018). The task was administered on Inquisit software by Millisecond. Images were taken from the Chicago Face Database-India.

## 3.3 Empirical Approach

The effect of interest is the causal impact of prejudice confrontation on police bias. We therefore estimate the following linear model:

$$Y_{i,s,o} = \beta_1 + \beta_2 Treatment_i + \mathbf{X}_i + \gamma_s + \alpha_o + \epsilon_{i,s,o} \tag{3.1}$$

where Y is an outcome of interest (e.g., priority given to victim statement) for officer $i$. $\mathbf{X}_i$ is a vector of individual level controls including gender and posting of the officer. We also include baseline session ($\gamma_s$) and senior officer ($\alpha_o$) fixed effects to account for the fact that each session ran independently and minimal disruptions could influence the outcomes of interest. The main explanatory variable *Treatment* identifies the experimental condition for officers randomly assigned to the confrontation group (i.e. confrontation with GBV case handling) in comparison to those that did not. We cluster standard errors at the officer group (session) level.

We check for balance on an array of baseline characteristics. Among the 11 indicators evaluated, there are no systematic differences between the treatment and control group officers for the majority of indicators - both within and across gender groups (Table 3.A.1). On examining baseline characteristics among male officers in the treatment and control group, we find an imbalance in 2 of the 11 variables: Officers in the treatment group prioritize the victim's statement less often than the control group and take longer to respond in the non-GBV case (Column 5). The baseline characteristics of female officers in the treatment and control group are balanced, except one (Column 6). Treated females prioritize the victim's statement in one of the GBV cases slightly less than females in the control group. In Column (7), we look at the difference between male and female officers in the treatment group and find a significant difference in officers' age and response time in the non-GBV case. Female officers in the treated group were significantly younger and males had a longer response time in the non-GBV case.

Our results are robust to the inclusion of the above baseline controls that show statistically significant differences between treatment and control groups (Table 3.A.6). Our preferred specification controls for officer characteristics, i.e., officer age and whether the officer was posted in the capital city of Bhopal. Further, we conduct a randomization inference test, closely following Heß (2017). If the confrontation had not affected any officer's case handling at all, then the observed outcome for our main variable (i.e. prioritizing the victim's statement) would have occurred with a probability of below 0.2%, as measured by the p-value implied by Figure 3.A.1. We interpret this as corroborating evidence that the confrontation did in fact influence officers' case handling.

## 3.4 Results

This section first focuses on our main results. In the next step, we show robustness checks.

### 3.4.1 Primary Effects

First, we estimate the impact of the intervention on six key outcomes that capture the handling of a GBV case one week after the confrontation. The first four outcomes are variables that are constructed from survey questions on how the officers would handle the presented GBV case. They include binary variables on whether the officer recognizes that a crime has been committed, decides not to pursue the complaint (after having it registered), prioritizes the victim's statement over other aspects of the case and the offender's statement, or prioritizes the offender's statement in the case processing. In the other two outcomes, we look at officers' general disposition toward GBV. From the questions on how they would handle the presented cases, we derive the dummy variable whether they are likely to file a complaint against the accused at all. From a case-independent survey that the officers completed at the end of the experiment, the outcome of how many out of 10 rape reports officers believe to be true is constructed. Second, in a computerized stereotyping reaction time task, we analyze to what extent participants are biased based on the appearance of victims. The two outcomes derived from this task are counts of keyboard clicks that represent a denial of victim stereotypes (i.e. select innocent words for the victim image) in the context of GBV and non-GBV, respectively.

Over 95% of the officers recognize that a crime has occurred after reviewing the GBV case and there is no impact of the intervention on this outcome (see Table 3.A.2). For outcomes concerning the further procedure with the GBV case, we find no impact of the confrontation feedback for the full sample (Panel A, Table 3.1). However, this average effect hides offsetting effects by gender (Panel B, Table 3.1). Treated female officers have an 8 percentage points higher probability of pursuing a GBV complaint than the control group (Columns 1 and 2). Next, we turn to our main outcome. Treated females prioritize the victim's statement by 23 percentage points more than the control group, which roughly corresponds to a 27% change. Consistently, we observe a negative effect on the prioritization of the offender's statement for female officers (Columns 5 and 6). For male officers, there is no statistically significant effect for dropping a complaint. However, they seem to push

*Table 3.1:* The Effects of Confrontation on the Handling of GBV Cases

| | Drop complaint | | Prio on victim's statement | | Prio on offender's statement | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A: Without gender heterogeneity** | | | | | | |
| GBV treatment | 0.011 | 0.013 | -0.058 | -0.055 | 0.047 | 0.043 |
| | (0.030) | (0.030) | (0.050) | (0.049) | (0.041) | (0.041) |
| | [0.377] | [0.652] | [0.164] | [0.260] | [0.304] | [0.313] |
| Controls | no | yes | no | yes | no | yes |
| Control Group Mean | 0.081 | 0.081 | 0.844 | 0.844 | 0.075 | 0.075 |
| Rsq | 0.052 | 0.069 | 0.038 | 0.053 | 0.064 | 0.072 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |
| **Panel B: With gender heterogeneity** | | | | | | |
| GBV treatment | 0.034 | 0.035 | -0.129** | -0.127** | 0.096* | 0.092* |
| | (0.034) | (0.034) | (0.055) | (0.055) | (0.045) | (0.046) |
| | [0.197] | [0.197] | [0.022] | [0.022] | [0.048] | [0.048] |
| treatment x female | -0.115* | -0.113* | 0.361*** | 0.360*** | -0.246*** | -0.246*** |
| | (0.057) | (0.055) | (0.072) | (0.075) | (0.050) | (0.052) |
| | [0.094] | [0.094] | [0.001] | [0.001] | [0.001] | [0.001] |
| female officer | -0.022 | 0.015 | -0.092 | -0.141** | 0.113** | 0.126*** |
| | (0.048) | (0.050) | (0.066) | (0.063) | (0.041) | (0.040) |
| Controls | no | yes | no | yes | no | yes |
| Control Group Mean | 0.081 | 0.081 | 0.844 | 0.844 | 0.075 | 0.075 |
| Rsq | 0.065 | 0.076 | 0.073 | 0.084 | 0.091 | 0.098 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

Notes: This Table reports estimates of the effect of the intervention on the likelihood of dropping a GBV complaint (Columns 1 and 2), the likelihood of prioritizing the victim's statement (Columns 3 and 4), and prioritizing the offender's statement (Columns 5 and 6). All three outcomes are dummies and take the value 1 if the complaint is not pursued, the priority is given to the victim's statement, and the priority is given to the offender's statement, respectively. 'GBV treatment' is an indicator for being confronted with one's own bias in the GBV case handling. 'treatment x female' is an interaction term of this indicator and the gender dummy (=1 if female). The specification in Columns (1), (3), and (5) includes senior officer and session fixed effects. In Columns (2), (4), and (6) we add officer characteristics as controls (officer's age and a dummy for posting is in the capital Bhopal). Standard errors are clustered at the session level and are reported in parentheses. Bootstrapped p values of multiple hypotheses tests are shown in square brackets. We use the Stata command mhtreg for multiple hypotheses testing that provides a procedure that asymptotically controls family-wise error rate and is asymptotically balanced (Steinmayer, 2020). It is based on List, Shaikh, and Xu (2019) but modified to be used in a multivariate regression setting.
*p<.10, **p<.05, ***p<.01

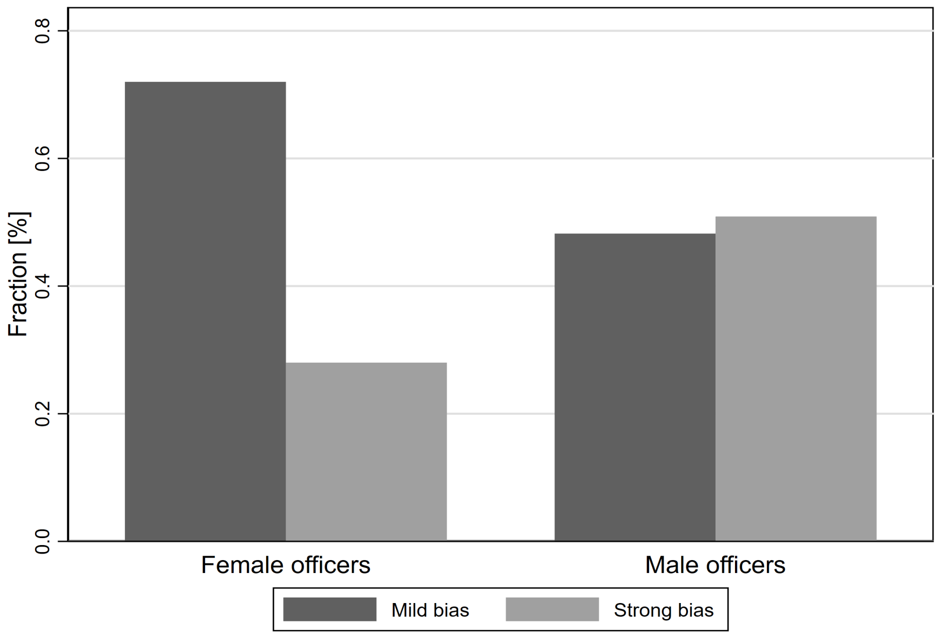
back after a confrontation, putting less priority on the victim's statement and more on the offender's statement. In Columns (1), (3), and (5) of Table 3.1 we implement fixed effects for the senior officer that confronted the officer and the experimental session. Controls for officer characteristics, i.e. the officer's age and posting, are then added in Columns (2), (4), and (6). The results for the two different specifications are almost identical in terms of effect sizes and precision.

For the second set of results, which concerns the officers' general disposition toward GBV, we do not find any statistically significant impact of the confrontation. There is no difference between the treatment and control group regarding the belief in the truth of rape reports (Columns 1 and 2 of Table 3.A.3) and the likelihood of registering a GBV complaint (Columns 3 and 4). Notably, in most instances, GBV cases are registered by officers, i.e. 93 percent of cases in the control group were registered. This is likely due to the fear of administrative sanctions or penalties that officials may face if they do not comply with the requirement to register any complaint. In the same fashion as in the previous outcome table, we show a specification with senior officer and session fixed effects first (Columns 1 and 3) and add officer characteristics as controls in Columns (3) and (4). Our preferred specification includes officer characteristics controls and is used in the further course of the paper.

Next, we report results for pre-specified outcomes for which standard errors are too large to be able to draw any conclusions. The participants completed a survey on how they felt after the feedback session with the senior officer at baseline. We do not observe any statistically significant effect of the confrontation on feeling guilty or other self-reflection variables, except for 'regret' where the confrontation leads to a decrease for males (see Table 3.A.7). At the end of the experiment (endline), the officers completed another short survey on victim-blaming (in how many out 10 GBV cases they think it is the woman's fault) and their perceived norms (to what extent they think their friends, family and partner think GBV is the woman's fault). There is also no statistically significant treatment effect for those outcomes (see Table 3.A.8).

We now turn to discussing possible mechanisms for the observed result patterns. While the majority of female officers (72%) show a mild bias in handling a GBV case before treatment, for more than half of the male officers (51%) a strong bias comes to light (Figure 3.3). Given that policing is highly male-dominated, the average female officer perceives a bias in her work environment that is more severe than her own.[16] The confrontation treatment thus de-biases female officers and makes them adjust their professional beliefs. Male officers seemingly behave in a way consistent with a backlash, which is driven by strongly biased men. In Table 3.A.4, we exclude male officers that showed a strong bias at baseline and find that for our main outcome (prioritization of victim's statement), the negative effect for males becomes substantially smaller and statistically insignificant while the positive effect for females persists.

[16] Among ASI and SI officers in Madhya Pradesh, 13% are female. (Bureau of Police Research and Development (Ministry of Home Affairs). Data on Police Organization 2021).




Notes: This Figure shows the fraction of female and male officers that show a mild or a strong bias in the handling of the GBV case at baseline, respectively. An example of a strong bias would be the filing of a report against the victim. We consider it a mild bias if, for example, the officer places an investigation against the victim among the top three (of 5) important investigative parts.

*Figure 3.3:* The Extent of Bias by Gender of Officer

A similar result pattern emerges in the negative stimuli reaction task where we evaluate to what extent participants are biased based on the appearance of victims. Participants solve a computerized stereotyping reaction time task in which pictures of potential victims are shown and they are asked to rapidly categorize descriptions (e.g. 'innocent', 'at fault') that might apply to the victim or not via key presses. We find that treated male participants show less denial of and more agreement with GBV stereotypes in GBV trials. This indicates significantly greater (15%) use of GBV stereotypes after a confrontation (Column 1 of Table 3.2). For female officers, the estimate ('treatment' + 'treatment x female') is not statistically different from zero. As expected, there is no impact of the confrontation on bias in gender-neutral cases (Column 2 of Table 3.2). Men's counter-reaction after a confrontation can explain this pattern, too (see above). Women do not change their inner beliefs, but de-bias their case handling after being encouraged to do so by the confrontation feedback.

*Table 3.2:* The Effects of Confrontation on Stereotypes Based on Victim's Appearance

| | Neg stimuli-GBV cases | Neg stimuli-other cases |
|---|---|---|
| | (1) | (2) |
| GBV treatment | -6.914** | 0.676 |
| | (2.512) | (1.242) |
| treatment x female | 3.157 | -5.373 |
| | (4.051) | (3.180) |
| female officer | 1.725 | 1.987 |
| | (3.303) | (2.300) |
| Controls | yes | yes |
| Control Group Mean | 44.828 | 45.867 |
| Rsq | 0.235 | 0.253 |
| N | 233 | 233 |

Notes: This Table reports estimates of a computerized negative stimuli reaction task. The outcome is the number of key presses that represent a denial of a victim stereotype in a GBV case (Column 1) and a non-GBV case (Column 2). The specification is the same as in our main Table 3.1 but adds a control for the officer's total number of key presses. The sample is smaller due to technical issues on some of the intervention days (47 (20%) females and 186 (80%) males participated). Refer also to the table notes of Table 3.1.
*p<.10, **p<.05, ***p<.01

### 3.4.2 Robustness Checks

We find our primary estimates to be robust to an array of sensitivity checks. Overall, the findings demonstrate that the main results are robust to alternative specifications and estimations, indicating that the confrontation does indeed lead to female officers responding less stereotypically in handling a GBV case and male officers responding in the opposite direction. We start with controlling for all baseline characteristics that we found to be statistically significant in our balance checks (Table 3.A.1) and find that our main results are qualitatively very similar but less precisely estimated (Table 3.A.6). Next, Table 3.A.5 presents evidence that our main results are robust to excluding all fixed effects and additional controls and implementing heteroskedasticity-robust standard errors (not clustered). The estimates are very comparable in size and precision. Further, we control for multiple hypotheses testing and compute bootstrapped p-values for our main outcomes (Table 3.1, 3.2, and 3.A.3). Our findings are statistically significant at conventional levels. We implement a procedure that asymptotically controls the family-wise error rate and is asymptotically balanced (Steinmayer, 2020). It is based on the approach of List, Shaikh, and Xu (2019) but modified to be used in a multivariate regression setting.

## 3.5 Conclusion

This paper has two central objectives. The first is to investigate how officers change their behavior in handling GBV cases after a prejudice confrontation. The second is to examine how treated and control officers differ in the degree of bias based on victim appearance. To achieve these objectives, we cooperated with the Madhya Pradesh police in India and conducted a lab-in-the-field experiment. The participants reviewed two cases, including one that involved a woman reporting GBV, and one non-GBV case, and answered a computer-based survey on how they would handle the cases. We randomly assigned officers to a treatment condition where a high-ranking police officer confronted them with their bias in dealing with the GBV case after the survey was completed. Officers in the control condition received neutral feedback on the non-GBV case. One week later, all officers were asked to review a GBV case again, solve a computerized stereotyping reaction time task in which pictures of potential victims were shown, and were asked to rapidly categorize descriptions that might apply to the victim or not. To estimate the causal effects of the intervention on biased responses to GBV cases and victims, we use an OLS estimation that exploits variation in being confronted or not.

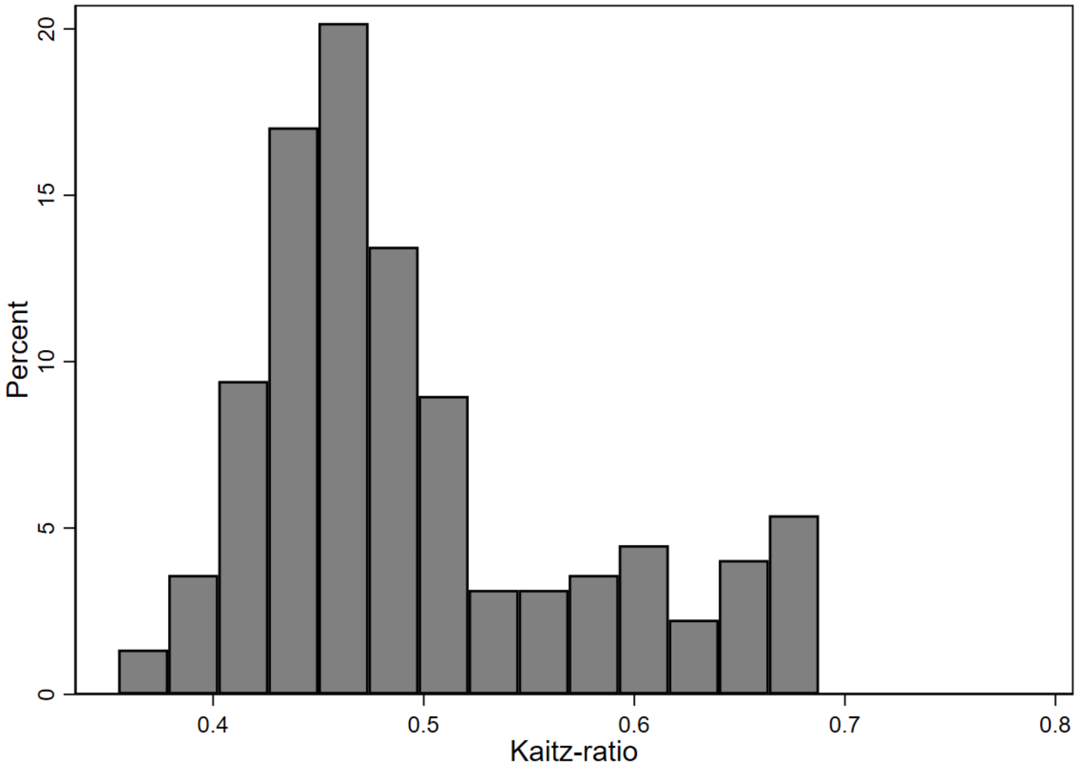
Our main finding is that the confrontation does lead to female officers responding less stereotypically in handling a GBV case while men tend to have a counter-reaction. Females significantly prioritize the victim's statement more than the control group and give less priority to the offender's statement. Confronted females also have a higher probability of pursuing a GBV complaint (instead of dropping it) than the control group. For male officers, there is no statistically significant effect for dropping a complaint. However, it seems to come to a backlash in the further course of the investigation: They put less priority on the victim's statement and more on the offender's statement. We find no effect on outcomes, that represent officers' general disposition toward GBV, such as the belief in the truthfulness of GBV complaints and registering a GBV complaint in the first place. A potential explanation for our findings is that females predominantly show a mild bias in handling a GBV case before treatment, while more than half of the males are strongly biased. Considering that policing is highly male-dominated, the average female officer perceives a bias in her work environment that is stronger than her own. The confrontation treatment thus de-biases female officers and makes them adjust their beliefs in a professional context. For male confronted officers, it seems to come to a backlash that is driven by strongly biased men.

In the computerized negative stimuli reaction task, we find significantly greater use of GBV stereotypes for men after a confrontation, while there is no effect for female officers. This finding is consistent with our results concerning the GBV case handling. While men seem to push back after a confrontation, women do not change their inner beliefs.

Taken together, the intervention encouraged women to de-bias their actions, which means less stereotyping when working on GBV cases. Men, on the other hand, are less receptive to the feedback. Therefore, the task of future research is to carefully design interventions for strongly biased men in a country where norms do not indicate egalitarianism. On a cautionary note, our findings do not necessarily allow us to draw general conclusions about settings in other countries. Research has highlighted that the response to confrontation is influenced by different historical, social, and cultural factors.

# Appendices

# Appendix to Chapter 1: Additional Tables and Figures



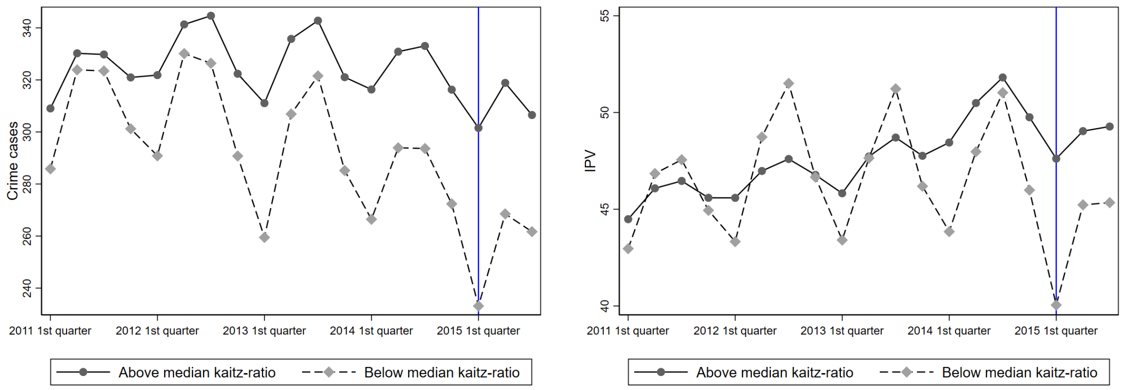

Notes: This Figure plots the distribution of observed Kaitz ratios of all German labor market areas in 2014. The distribution is based on wages implied by median earnings.

*Figure 1.A.1:* Kaitz Ratios in 2014

*Table 1.A.1:* Analysis Sample: Crime Types

| **Offense category** | **Cases** | **of which are among intimate partners** | **in %** |
|---|---:|---:|---:|
| Homicide | 12,751 | 1,990 | 15.61 |
| Offences against sexual self-determination | 149,334 | 11,010 | 7.37 |
| Brute force offenses and crimes against personal freedom | 218,372 | 4,234 | 1.94 |
| Assault | 2,516,548 | 424,987 | 16.89 |
| Offences against personal freedom | 902,894 | 155,051 | 17.17 |
| Resistance | 101,000 | 236 | 0.23 |
| Other | 8,326 | 25 | 0.3 |
| **Total** | **3,919,225** | **597,533** | **15.25** |

Notes: This Table shows the crime data used for analysis by crime type. Reading example: There are 12,751 homicide cases in the sample, of which 1,990 (or 15.61%) were committed by intimate partners.




*(a)* Pre-Trends in Crime

*(b)* Pre-Trends in Intimate Partner Violence (IPV)

Notes: This Figure shows the average crime and IPV cases over the sample period. The introduction of the minimum wage in 2015 is marked by a vertical line. Labor market areas are placed into two groups depending on the value of the prevailing Kaitz-ratio in 2014 (below or above median). Crime counts are residualized from the population size of the respective labor market area.

*Figure 1.A.2:* Trends in Crime by Minimum Wage Bite

*Table 1.A.2:* Refugee Crisis and the Effect of Minimum Wage Policy on Crime

| | Asylum requests | Crime | Crime w.o IPV | IPV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| post x mw bite | 0.404 | -0.073 | -0.180** | 0.627*** |
| | (0.678) | (0.070) | (0.085) | (0.156) |
| post 2015 | 1.198*** | -0.078** | -0.039 | -0.344*** |
| | (0.394) | (0.034) | (0.040) | (0.070) |
| asylum request rate | | 0.284 | -4.775 | 23.094*** |
| | | (5.362) | (7.142) | (7.877) |
| Labor market area FE | yes | yes | yes | yes |
| Season FE | yes | yes | yes | yes |
| Dep.var mean | 46667 | 304 | 255 | 49 |
| N | 12,711 | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on the number of asylum requests (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (Column 4). 'post 2015' is an indicator for the year 2015. 'Post x mw bite' is an interaction term of this indicator and the Kaitz-ratio at the labor market level. The asylum request rate is measured in number of requests per 100 inhabitants. All specifications contain the same set of fixed effects and level of clustering as the main specification that can be found in Columns (1), (3), and (5) of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.4:* The Effect of Minimum Wage Policy on Completed IPV Offenses vs. IPV Attempts

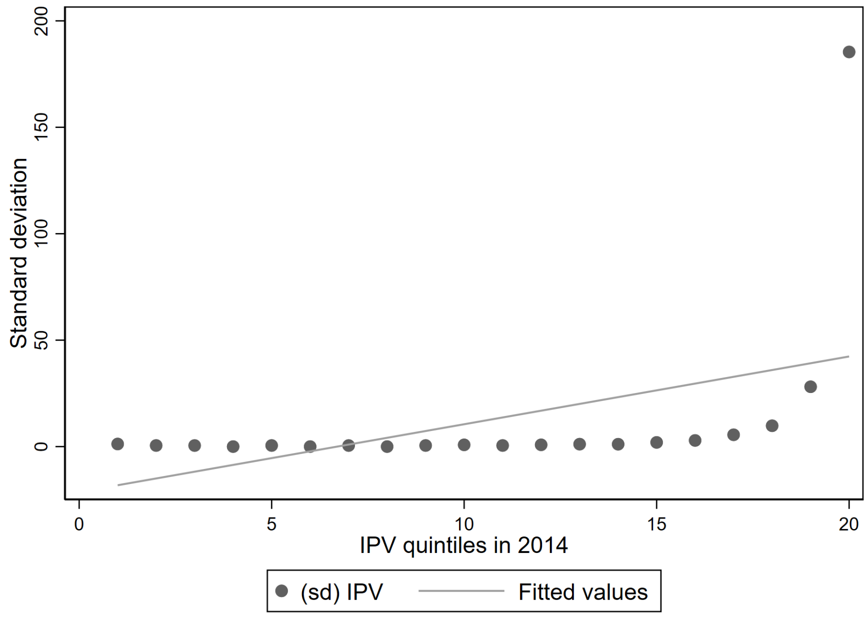
| | IPV | IPV completed | IPV attempt |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | 0.620*** | 0.636*** | 0.192 |
| | (0.157) | (0.155) | (0.409) |
| post 2015 | -0.317*** | -0.325*** | -0.112 |
| | (0.071) | (0.070) | (0.185) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 49 | 47 | 2 |
| N | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on all Intimate Partner Violence (IPV) cases (Column 1), actual completed IPV offenses (Column 2) and IPV attempts (Column 3). 'post 2015' is an indicator for the year 2015. 'Post x mw bite' is an interaction term of this indicator and the Kaitz-ratio at the labor market level. All specifications contain the same set of fixed effects and level of clustering as the main specification that can be found in Columns (1), (3), and (5) of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.3:* Selective Migration and the Effect of Minimum Wage Policy on Crime

| | Crime | | | Crime w.o IPV | | | IPV | | |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| post x mw bite | -0.059 | -0.068 | -0.076 | -0.158* | -0.173** | -0.161* | 0.594*** | 0.617*** | 0.512*** |
| | (0.074) | (0.074) | (0.069) | (0.089) | (0.088) | (0.084) | (0.155) | (0.156) | (0.151) |
| post 2015 | -0.070** | -0.072** | -0.074** | -0.031 | -0.036 | -0.032 | -0.335*** | -0.321*** | -0.363*** |
| | (0.035) | (0.035) | (0.037) | (0.042) | (0.041) | (0.044) | (0.070) | (0.070) | (0.071) |
| migration-inflow | -0.010*** | | 0.008 | -0.015*** | | -0.012 | 0.020*** | | 0.117*** |
| | (0.003) | | (0.015) | (0.004) | | (0.018) | (0.007) | | (0.017) |
| migration-outflow | | -0.014*** | -0.024 | | -0.018*** | -0.004 | | 0.008 | -0.129*** |
| | | (0.004) | (0.020) | | (0.005) | (0.023) | | (0.007) | (0.022) |
| Labor market area FE | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| Season FE | yes | yes | yes | yes | yes | yes | yes | yes | yes |
| Dep.var mean | 304 | 304 | 304 | 255 | 255 | 255 | 49 | 49 | 49 |
| N | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Columns 1, 2, 3), crime excluding intimate partner violence (Columns 4, 5, 6), and Intimate Partner Violence (IPV) offenses (Columns 7, 8, 9). 'post 2015' is an indicator for the year 2015. 'Post x mw bite' is an interaction term of this indicator and the Kaitz ratio at the labor market level. Migration in and outflow are measured as immigrants and emigrates per 100 inhabitants, respectively. All specifications contain the same set of fixed effects and level of clustering as the main specification that can be found in Columns (1), (3), and (5) of Table 1.1.
*p<.10, **p<.05, ***p<.01




Notes: This Figure plots the standard deviation of Intimate Partner Violence (IPV) crime per quintiles of IPV cases (monthly data) in 2014.

*Figure 1.A.3:* Potential Overdispersion in IPV Data

*Table 1.A.5:* Excluding Outliers in IPV

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | -0.130 | -0.225** | 0.521*** |
| | (0.084) | (0.099) | (0.178) |
| *Effect size in [%]* | -13.88 | -25.23 | +68.37 |
| *25th-75th percentile effect in [%]* | -1.32 | -2.40 | +6.50 |
| post 2015 | -0.045 | -0.017 | -0.255*** |
| | (0.042) | (0.048) | (0.083) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| N | 12,084 | 12,084 | 12,084 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (Column 3). Results are shown for the same specification presented in the main Table (1.1), Columns (1), (3), and (5), but excluding labor market areas in the top 5% of the IPV distribution in the sample. Refer also to the table notes of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.6:* Robustness: Negative Binomial Model

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | -0.111** | -0.210*** | 0.567*** |
| | (0.055) | (0.065) | (0.073) |
| *Effect size in [%]* | -11.74 | -23.37 | +76.30 |
| *25th-75th percentile effect in [%]* | -1.06 | -2.10 | +6.87 |
| post 2015 | -0.034 | -0.004 | -0.262*** |
| | (0.027) | (0.031) | (0.035) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| N | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (Column 3) from a conditional fixed effects negative binomial model, implemented using the 'nbreg' command in Stata. Labor market area and season of the year fixed effects are implemented. Standard errors are clustered at the labor market area level and reported in parentheses.
*p<.10, **p<.05, ***p<.01

*Table 1.A.7:* Robustness: Sample Cuts

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Crime data included until August 2015** | | | |
| post x mw bite | 0.004 | -0.099* | 0.673*** |
| | (0.054) | (0.060) | (0.088) |
| post 2015 | -0.098*** | -0.065** | -0.328*** |
| | (0.027) | (0.029) | (0.041) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| N | 12,488 | 12,488 | 12,488 |
| **Panel B: Crime data included until July 2015** | | | |
| post x mw bite | 0.021 | -0.072 | 0.637*** |
| | (0.060) | (0.072) | (0.159) |
| post 2015 | -0.107*** | -0.078** | -0.314*** |
| | (0.029) | (0.034) | (0.072) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| N | 12,265 | 12,265 | 12,265 |
| **Panel C: Crime data included until June 2015** | | | |
| post x mw bite | 0.068 | -0.019 | 0.653*** |
| | (0.059) | (0.070) | (0.162) |
| post 2015 | -0.132*** | -0.105*** | -0.330*** |
| | (0.029) | (0.033) | (0.073) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep. var mean | 304 | 255 | 49 |
| N | 12,042 | 12,042 | 12,042 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (Column 3). Results are shown for the same specification presented in the main Table (1.1), Columns (1), (3), and (5), but excluding the month of September in Panel A, the months August and September in Panel B, and July to September in Panel C. Refer also to the table notes of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.8:* The Effect of Minimum Wage Policy on Crime: County-Level Analysis

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | 0.026 | -0.060 | 0.571*** |
| | (0.055) | (0.061) | (0.114) |
| *Effect size [in %]* | +2.63 | -6.18 | +77.00 |
| *25th-75th percentile effect [in %]* | +0.21 | -0.49 | +6.16 |
| post 2015 | -0.113*** | -0.089*** | -0.278*** |
| | (0.027) | (0.029) | (0.052) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 169 | 142 | 27 |
| N | 22,857 | 22,857 | 22,857 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (3). Results are shown for the same specification presented in the main Table (1.1), Columns (1), (3), and (5), but on county instead of labor market area. Refer also to the table notes of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.9:* Robustness: Alternative Measure for Hours Worked

| | Crime | Crime w.o IPV | IPV |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | -0.075 | -0.128 | 0.298** |
| | (0.097) | (0.104) | (0.134) |
| *Effect size [in %]* | +7.79 | -13.66 | +34.72 |
| *25th-75th percentile effect [in %]* | +0.39 | -0.68 | +1.74 |
| post 2015 | -0.087*** | -0.082** | -0.136*** |
| | (0.032) | (0.034) | (0.045) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 304 | 255 | 49 |
| N | 11,970 | 11,970 | 11,970 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall crime (Column 1), crime excluding intimate partner violence (Column 2), and Intimate Partner Violence (IPV) offenses (Column 3). Results are shown for the same specification presented in the main Table (1.1), Columns (1), (3), and (5), using an alternative measure for hours worked and therefore the minimum wage bite. Hours worked are retrieved from the INKAR database. The state Thüringen seems to have not reported the data uniformly and is excluded therefore from the analysis (corresponds to 13 labor market areas). Also refer to the table notes of Table 1.1.
*p<.10, **p<.05, ***p<.01

*Table 1.A.10:* The Effect of Minimum Wage Policy on Employees Subject to Social Insurance

| | Employees subject to social insurance | ...Men | ...Women |
|---|---|---|---|
| | (1) | (2) | (3) |
| post x mw bite | -0.056*** | -0.050*** | -0.064*** |
| | (0.010) | (0.012) | (0.014) |
| *Effect size [in %]* | -5.76 | -5.13 | -6.61 |
| *25th-75th percentile effect [in %]* | -0.52 | -0.46 | -0.59 |
| post 2015 | 0.050*** | 0.055*** | 0.046*** |
| | (0.005) | (0.006) | (0.007) |
| Labor market area FE | yes | yes | yes |
| Season FE | yes | yes | yes |
| Dep.var mean | 134,597 | 62,409 | 72,188 |
| N | 12,711 | 12,711 | 12,711 |

Notes: This Table reports estimates of the effect of the minimum wage introduction on overall number of employees subject to social insurance (Column 1), number of male employees subject to social insurance (Column 2), and number of female employees subject to social insurance (Column 3). Results are shown for the same specification presented in the main Table (1.1), Columns (1), (3), and (5). Refer to the table notes of Table 1.1.
*p<.10, **p<.05, ***p<.01

# Appendix A to Chapter 2: Additional Tables

*Table 2.A.1:* Sample Sizes

| | |
|---|---|
| **Estimation sample** | |
| Domestic violence cases classified by call handlers (2011-2016) | 184,468 |
| Nonmissing dispatch time | 174,130 |
| At least 400 DV cases in dispatched team | 136,649 |
| Call grade is 1 or 2 (baseline estimation sample) | 124,216 |
| **Instrument construction sample** | |
| Officer-case level observations in call handler defined DV cases (2010-2019) | 631,834 |

*Table 2.A.2:* Testing Random Assignment of First Response Teams

| | Dependent variable: | |
|---|---|---|
| | **Arrest x 100** | **Team arrest propensity x 100** |
| | (1) | (2) |
| **Past DV history**: | | |
| Case in past 12 months | 0.484** | -0.015 |
| | (0.172) | (0.011) |
| Arrest in past 12 months | 1.777*** | 0.016 |
| | (0.325) | (0.014) |
| Formal investigation in past 12 months | 0.163 | 0.015 |
| | (0.184) | (0.011) |
| Criminal charge in past 12 months | 2.043*** | -0.014 |
| | (0.305) | (0.013) |
| **Case characteristics**: | | |
| Caller identity (=1 victim) | 0.161 | 0.013 |
| | (0.110) | (0.008) |
| Gender of call handler (=1 female) | -0.126 | 0.006 |
| | (0.126) | (0.007) |
| Call handler experience (years) | 0.008 | 0.000 |
| | (0.008) | (0.001) |
| Mean of dep. var. | 3.120 | 3.166 |
| Joint F-statistic | 22.936 | 0.946 |
| [p-value] | [0.000] | [0.478] |
| Observations | 124,216 | 124,216 |

Notes: OLS regressions controlling for the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.3:* Testing the Monotonicity Assumption

| | Dependent variable: **Arrest** | | | | | |
|---|---|---|---|---|---|---|
| | Prior DV call | | DV hotspot | | Time of day | |
| | Yes | No | Yes | No | Day | Night |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| **Panel A: Baseline instrument** | | | | | | |
| Team arrest propensity | 0.656*** | 0.799*** | 0.799*** | 0.656*** | 0.543*** | 0.853*** |
| | (0.103) | (0.083) | (0.090) | (0.085) | (0.097) | (0.083) |
| Mean of dep. var. | 0.037 | 0.026 | 0.030 | 0.032 | 0.024 | 0.037 |
| Observations | 58,139 | 66,049 | 52,922 | 71,294 | 53,164 | 71,036 |
| **Panel B: Reverse sample instrument** | | | | | | |
| Reverse team arrest propensity | 0.570*** | 0.674*** | 0.682*** | 0.441*** | 0.424*** | 0.319*** |
| | (0.105) | (0.082) | (0.088) | (0.083) | (0.083) | (0.078) |
| Mean of dep. var. | 0.038 | 0.027 | 0.031 | 0.032 | 0.024 | 0.037 |
| Observations | 52,865 | 59,680 | 47,554 | 65,640 | 48,907 | 65,697 |

Notes: First stage estimates regressing arrest on team arrest propensity/reverse team arrest propensity for different subsamples, controlling for the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Prior DV call refers to whether an emergency DV call was made in the previous 12 months. DV hotspot is defined as wards where the fraction of DV calls relative to the population is above the 75th percentile. For time of day, day is defined as 6 am to 6 pm and night as 6 pm to 6 am. Panel A uses the instrument constructed using the entire baseline sample. Panel B uses reverse sample instruments constructed using cases in the opposite subsample. For example, the reverse sample instrument in Column (1) is based on cases with no prior DV call. Note that sample sizes are smaller in panel B, as we require at least 400 cases per team in each subsample. Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.4:* Testing the Exclusion Restriction

| | Dependent variable: **Repeat call for DV** | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Arrest | -0.488*** | -0.487*** | -0.487*** | -0.486*** | -0.480*** | -0.518*** |
| | (0.187) | (0.186) | (0.183) | (0.187) | (0.182) | (0.191) |
| Formal investigation | | -0.049 | | | -0.040 | |
| | | (0.034) | | | (0.035) | |
| Time on scene | | | 0.002 | | 0.009 | |
| | | | (0.022) | | (0.022) | |
| Recommend investigation | | | | -0.044 | -0.031 | |
| | | | | (0.061) | (0.061) | |
| Advice | | | | -0.001 | -0.000 | |
| | | | | (0.053) | (0.053) | |
| Instrument: Team arrest propensity | yes | yes | yes | yes | yes | yes |
| Instrument: FI propensity | no | yes | no | no | yes | no |
| Instrument: Time on scene propensity | no | no | yes | no | yes | no |
| Instrument: Recommend FI propensity | no | no | no | yes | yes | no |
| Instrument: Advice propensity | no | no | no | yes | yes | no |
| Control: Team characteristics | no | no | no | no | no | yes |
| Mean of dep. var. | | | 0.492 | | | |
| Control complier mean | | | 0.962 | | | |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: The first Columns add in additional endogenous variables and instruments. The corresponding first stages are found in Appendix Table 2.A.8. Formal investigation is a dummy for whether a case is formally investigated. Time on the scene is measured in hundreds of minutes, and is based on the difference between when the first officer arrives and the last officer leaves. Recommend investigation and advice are dummy variables for whether the response team's actions include recommending a criminal investigation and providing advice, respectively. Instruments for the variables in Column (2) to (4) are constructed analogously to how we construct the team arrest propensity instrument for arrests. The team characteristics in Column (6) include the fraction of females on a team (mean=0.20) and the average age of the team (mean=36.9). All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.5:* Characterization of Compliers

| | Case in past 12 months | Arrest in past 12 months | Formal investigation in past 12 months | Criminal charge in past 12 months |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Population mean | 0.468 | 0.062 | 0.376 | 0.058 |
| Complier mean | 0.602 | 0.045 | 0.404 | 0.142 |
| Bootstrap std. err. | [0.117] | [0.062] | [0.114] | [0.052] |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: For details on these calculations see Appendix B.

*Table 2.A.6:* Testing for Bias due to Misclassification Errors

| | Dependent Variable: **Repeat call for DV** | |
|---|---|---|
| | Only areas with min. 80% of HH's in detached houses | Excluding city center of Birmingham (3km radius) |
| | (1) | (2) |
| Arrest | -0.468* | -0.441** |
| | (0.263) | (0.194) |
| Mean of dep. var. | 0.479 | 0.491 |
| Control complier mean | 0.931 | 0.914 |
| Observations | 81,953 | 118,559 |

Notes: In Column (1), the sample is restricted to areas (wards) where at least 80% of the households live in detached houses. Data on the dwelling types by ward in West Midlands come from NOMIS labour market statistics. In Column (2), the city center of Birmingham (defined as the 3km radius around St. Philips Cathedral) is excluded. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.7:* Testing for Differential Accuracy of Geo-Coded Location as a Function of Arrest

| | **Based on official victim ID** | | |
|---|---|---|---|
| | No repeat identified | Repeat identified | Total |
| **Based on geo-coordinates** | | | |
| Full Estimation Sample | | | |
| No repeat identified % | 90.93 | 9.07 | 100.00 |
| Arrest Subsample | | | |
| No repeat identified % | 91.61 | 8.39 | 100.00 |
| No Arrest Subsample | | | |
| No repeat identified % | 90.90 | 9.10 | 100.00 |
| p-value for mean difference | 0.37 | | |

Notes: In the paper, we match victims over time based on the geo-coordinates of the incident location. For the subsample of the data where the investigative officer opens an investigation, we can use official victim IDs to track individuals over time. This Table constructs an alternative measure for a repeat DV call using official victim ID and compares it to our measure using geo-coordinates for the same subsample. While the Table reveals that we miss 9% of repeat cases in this subsample, whether a repeat case is missing is not significantly related to whether there was an arrest.

*Table 2.A.8:* First Stages for the Extended IV Models

| | **Arrest** | **Formal investigation** | **Arrest** | **Time on scene** | **Arrest** | **Recommend investigation** | **Advice** |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Team arrest propensity | 0.728*** | 0.359** | 0.737*** | -0.049 | 0.728*** | 1.237*** | 1.174*** |
| | (0.070) | (0.142) | (0.069) | (0.147) | (0.072) | (0.134) | (0.134) |
| FI propensity | 0.018* | 0.942*** | | | | | |
| | (0.010) | (0.031) | | | | | |
| Time on scene propensity | | | 0.025*** | 1.048*** | | | |
| | | | (0.008) | (0.025) | | | |
| Recommend FI propensity | | | | | 0.004 | 1.086*** | 1.124*** |
| | | | | | (0.010) | (0.031) | (0.029) |
| Advice propensity | | | | | -0.005 | 0.286*** | 1.280*** |
| | | | | | (0.017) | (0.047) | (0.050) |
| Kleibergen-Paap Wald F statistic | 52 | | 58 | | | 48 | |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

| | **Arrest** | **Formal investigation** | **Recommend investigation** | **Advice** | **Time on scene** |
|---|---|---|---|---|---|
| | (8) | (9) | (10) | (11) | (12) |
| Team arrest propensity | 0.733*** | 0.294** | -0.133 | 1.258*** | 1.184*** |
| | (0.073) | (0.146) | (0.154) | (0.132) | (0.134) |
| FI propensity | 0.014 | 0.984*** | -0.082** | -0.116*** | -0.041 |
| | (0.011) | (0.034) | (0.034) | (0.032) | (0.030) |
| Time on scene propensity | 0.024*** | -0.061** | 1.087*** | -0.004 | 0.011 |
| | (0.008) | (0.024) | (0.026) | (0.024) | (0.023) |
| Recommend FI propensity | -0.006 | -0.008 | -0.009 | 1.134*** | 1.139*** |
| | (0.011) | (0.036) | (0.035) | (0.033) | (0.031) |
| Advice propensity | 0.009 | 0.226*** | 0.251*** | 0.267*** | 1.279*** |
| | (0.017) | (0.056) | (0.056) | (0.048) | (0.051) |
| Kleibergen-Paap Wald F statistic | | | 42 | | |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: This Table displays first stages corresponding to Table 2.A.4. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.9:* Controlling for Propensities of Actions other than Arrest

| | Dependent Variable: **Repeat call for DV** | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Arrest | -0.488*** | -0.511*** | -0.487*** | -0.561*** | -0.551*** |
| | (0.187) | (0.187) | (0.183) | (0.190) | (0.188) |
| Instrumented: Team arrest propensity | yes | yes | yes | yes | yes |
| Control: FI propensity | no | yes | no | no | yes |
| Control: Time on scene propensity | no | no | yes | no | yes |
| Control: Recommend FI propensity | no | no | no | yes | yes |
| Control: Advice propensity | no | no | no | yes | yes |
| Mean of dep. var. | | | 0.492 | | |
| Control complier mean | | | 0.962 | | |
| Observations | 124,216 | 124,216 | 124,216 | 124,216 | 124,216 |

Notes: Formal investigation (FI) propensity is the propensity of a team's cases to be formally investigated. Time on scene propensity is a team's average time spend on the scene in other cases. Recommend investigation propensity is the team's average propensity in other DV cases to recommend a criminal investigation, and advice propensity is similarly defined. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.10:* Testing Robustness to Alternative Specifications

| | Dependent Variable: **Repeat call for DV** | | | | |
|---|---|---|---|---|---|
| | Intensive margin | IV: call grade 1 and 2 | IV: Traditional l-o-o | Team DV cases: 300 | Team DV cases: 500 |
| | (1) | (2) | (3) | (4) | (5) |
| Arrest | -1.516* | -0.514*** | -0.474** | -0.406** | -0.441** |
| | (0.813) | (0.198) | (0.187) | (0.184) | (0.192) |
| Mean of dep. var. | 1.337 | 0.493 | 0.492 | 0.491 | 0.494 |
| Control complier mean | 2.867 | 0.921 | 0.948 | 0.880 | 0.915 |
| Observations | 123,063 | 118,788 | 124,229 | 134,534 | 111,437 |

Notes: Column (1) uses the intensive margin of the number of repeat DV cases in the following 12 months as the outcome variable. For this Column, we trim the data to exclude the top 1% of the outcome variable. The remaining Columns all use the baseline outcome. Our main specification uses high priority calls (call grade 1 or 2) to define the sample, but uses calls of any grade to construct the team arrest propensity instrument; in Column (2) we construct the instrument using only high priority calls. Our main specification only uses DV cases from other geo-locations (whether the current case or any other case) to construct the instrument; in Column (3) we use a standard leave-one-out instrument that only excludes the current case. Our main specification requires the first response team to handle at least 400 DV cases; in Columns (4) and (5) we require the teams to handle at least 300 or at least 500 DV cases, respectively. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

*Table 2.A.11:* Heterogeneous Effects of Arrest

| | Dependent Variable: **Repeat call for DV** | | | |
|---|---|---|---|---|
| | Formally investigated DV case in past 12 months | Pr(Arrest\|X) is high | At least one female officer on response team | Response team with above mean age |
| | (1) | (2) | (3) | (4) |
| Yes | -0.744** | -0.626** | -0.552** | -0.322 |
| | (0.308) | (0.307) | (0.242) | (0.230) |
| No | -0.431** | -0.513*** | -0.474** | -0.677*** |
| | (0.185) | (0.190) | (0.222) | (0.230) |
| Observations | 124,216 | 124,216 | 124,216 | 124,211 |

Notes: This Table shows heterogeneity analyses based on IV regressions which interact both the arrest variable and the team arrest propensity instrument with identifiers for the specified groups. In Column (2), we differentiate between cases with arrest predictions above or below the 75th percentile, where the predictions are based on the variables appearing in Table 2.A.2. All specifications include the baseline set of call grade, time, and geography variables used in Table 2.1 Column (4). Standard errors are reported in parentheses and are clustered at the level of the dispatched officer on a team with the most domestic violence cases.
*p<.10, **p<.05, ***p<.01

# Appendix B to Chapter 2: Estimating Control Complier Means and Complier Characteristics

## Estimating the Shares of Compliance Types in the Sample

Our estimation of the fraction of compliers ($\pi_c$), always-takers ($\pi_a$) and never-takers ($\pi_n$) follows the methodology outlined in Dahl, Kostøl, and Mogstad (2014). We refer the reader to the Technical Appendix in Dahl, Kostøl, and Mogstad (2014) and repeat only the central estimation steps here. Let $\underline{z}$ and $\overline{z}$ denote the minimum (least-likely arresting officers) and maximum (most-likely arresting officers) values of the instrument, assumed to be the values for the bottom 1 percentile and top 1 percentile of arrest propensity. Let $\hat{\alpha}_0$ and $\hat{\alpha}_1$ denote the estimated coefficients from our first stage regression in equation (2.2). We calculate $\pi_c$ from $\hat{\alpha}_1(\overline{z} - \underline{z})$, $\pi_a$ from $\hat{\alpha}_0 + \hat{\alpha}_1\underline{z}$, and $\pi_n$ from $1 - \hat{\alpha}_0 - \hat{\alpha}_1\overline{z}$. The estimated shares of the three compliance types in our sample are $\pi_c = 0.040$, $\pi_a = 0.017$, and $\pi_n = 0.943$.

## Estimating Control Complier Means

The control complier mean (CCM) informs us about how likely repeat victimization would be if a call had not resulted in an arrest. We therefore need to estimate the sample analog of $E(DV_{i,t+1}(0)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z}))$. To do so, consider DV calls that do not result in an arrest ($A_{i,t} = 0$). Victims whose calls are handled by response teams with $Z_{j(i)} = \overline{z}$ are never-takers, and their conditional expectation of a repeat call is:

$$E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \overline{z}) = E(DV_{i,t+1}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 0) \tag{3.2}$$

Those victims whose calls are handled by response teams with $Z_{j(i)} = \underline{z}$ are a mixture of never-takers and compliers:

$$\begin{aligned} E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \underline{z}) = & \frac{\pi_n}{\pi_n + \pi_c} E(DV_{i,t+1}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 0) \\ & + \frac{\pi_c}{\pi_n + \pi_c} E(DV_{i,t+1}(0)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})). \end{aligned} \tag{3.3}$$

By combining these two equations, we obtain:

$$\begin{aligned} E(DV_{i,t+1}(0)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) = & \frac{\pi_n + \pi_c}{\pi_c} E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \underline{z}) \\ & - \frac{\pi_n}{\pi_c} E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \overline{z}). \end{aligned} \tag{3.4}$$

Above, we have already laid out how to estimate $\pi_c$ and $\pi_n$. To obtain the two remaining quantities in equation (5), we need an estimate of the relationship between $DV_{i,t+1}$ and $Z_{j(i)}$ conditional on $A_{i,t} = 0$. We use a linear probability model to estimate

$$DV_{i,t+1} = \gamma_0 + \gamma_1 Z_{j(i)} + \gamma_2 A_{i,t} + X'_{i,t} + \epsilon_{i,t}. \tag{3.5}$$

We then calculate $E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \underline{z})$ from $\hat{\gamma}_0 + \hat{\gamma}_1\underline{z}$ and $E(DV_{i,t+1}|A_{i,t} = 0, Z_{j(i)} = \overline{z})$ from $\hat{\gamma}_0 + \hat{\gamma}_1\overline{z}$.

An alternative way of obtaining an estimate for the control complier mean is to calculate the difference between the treated complier mean (TCM) and our IV estimate for the arrest effect (Katz, Kling, and Liebman, 2001). For this, we need to estimate the sample analog of $E(DV_{i,t+1}(1)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z}))$, which by similar arguments as used above is given by:

$$\begin{aligned} E(DV_{i,t+1}(1)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) = {} & \frac{\pi_a + \pi_c}{\pi_c} E(DV_{i,t+1}|A_{i,t} = 1, Z_{j(i)} = \overline{z}) \\ & - \frac{\pi_a}{\pi_c} E(DV_{i,t+1}|A_{i,t} = 1, Z_{j(i)} = \underline{z}), \end{aligned} \tag{3.6}$$

where $\pi_a$ and $\pi_c$ are the shares of always-takers and compliers, respectively. The control complier mean is then

$$E(DV_{i,t+1}(0)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) = E(DV_{i,t+1}(1)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) - \hat{\beta}_1, \tag{3.7}$$

where $\hat{\beta}_1$ is our IV estimate for the arrest effect. Estimating $E(DV_{i,t+1}|A_{i,t} = 1, Z_{j(i)} = \overline{z})$ and $E(DV_{i,t+1}|A_{i,t} = 1, Z_{j(i)} = \underline{z})$ in equation (7) with a linear probability model and substituting the estimatess into equation (8), we obtain control complier means that are virtually identical to the ones reported in the paper. Our conclusion of a high value of $E(DV_{i,t+1}(0)|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z}))$ is therefore robust to the choice of model.

## Estimating Characteristics of Compliers

Our characterization of compliers adapts the binary-instrument methodology proposed by Marbach and Hangartner (2020) to a setting with a continuous instrument. Let $X_{i,t}$ be a covariate. By arguments we have established above, the covariate means for never-takers and always-takers can be estimated and are respectively given by:

$$E(X_{i,t}|A_{i,t} = 0, Z_{j(i)} = \overline{z}) = E(X_{i,t}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 0) \tag{3.8}$$

and

$$E(X_{i,t}|A_{i,t}=1, Z_{j(i)}=\underline{z}) = E(X_{i,t}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 1). \tag{3.9}$$

We calculate $E(X_{i,t}|A_{i,t} = 0, Z_{j(i)} = \overline{z})$ from $\hat{\zeta}_0 + \hat{\zeta}_1\overline{z}$, where $\hat{\zeta}_0$ and $\hat{\zeta}_1$ are OLS estimates of the relationship between $X_{i,t}$ and $Z_{j(i)}$ conditional on $A_{i,t} = 0$. Similarly, we calculate $E(X_{i,t}|A_{i,t} = 1, Z_{j(i)} = \underline{z})$ from $\hat{\eta}_0 + \hat{\eta}_1\underline{z}$, where $\hat{\eta}_0$ and $\hat{\eta}_1$ are OLS estimates of the relationship between $X_{i,t}$ and $Z_{j(i)}$ conditional on $A_{i,t} = 1$.

Turning to compliers, we note that by the law of intereated expectations, the population mean of $X_{i,t}$ can be decomposed into the never-taker, always-taker, and complier means, weighted by the share of each group:

$$\begin{aligned} E(X_{i,t}) = {} & \pi_n E(X_{i,t}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 0) + \pi_a E(X_{i,t}|A_{i,t}(\overline{z}) = A_{i,t}(\underline{z}) = 1) \\ & + \pi_c E(X_{i,t}|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) \end{aligned} \tag{3.10}$$

Combining equations (9) to (11), we calculate covariate means for compliers according to:

$$\begin{aligned} E(X_{i,t}|A_{i,t}(\overline{z}) > A_{i,t}(\underline{z})) = {} & \frac{1}{\pi_c} E(X_{i,t}) - \frac{\pi_n}{\pi_c} E(X_{i,t}|A_{i,t} = 0, Z_{j(i)} = \overline{z}) \\ & - \frac{\pi_a}{\pi_c} E(X_{i,t}|A_{i,t} = 1, Z_{j(i)} = \underline{z}). \end{aligned} \tag{3.11}$$

# Appendix to Chapter 3: Additional Tables and Figures

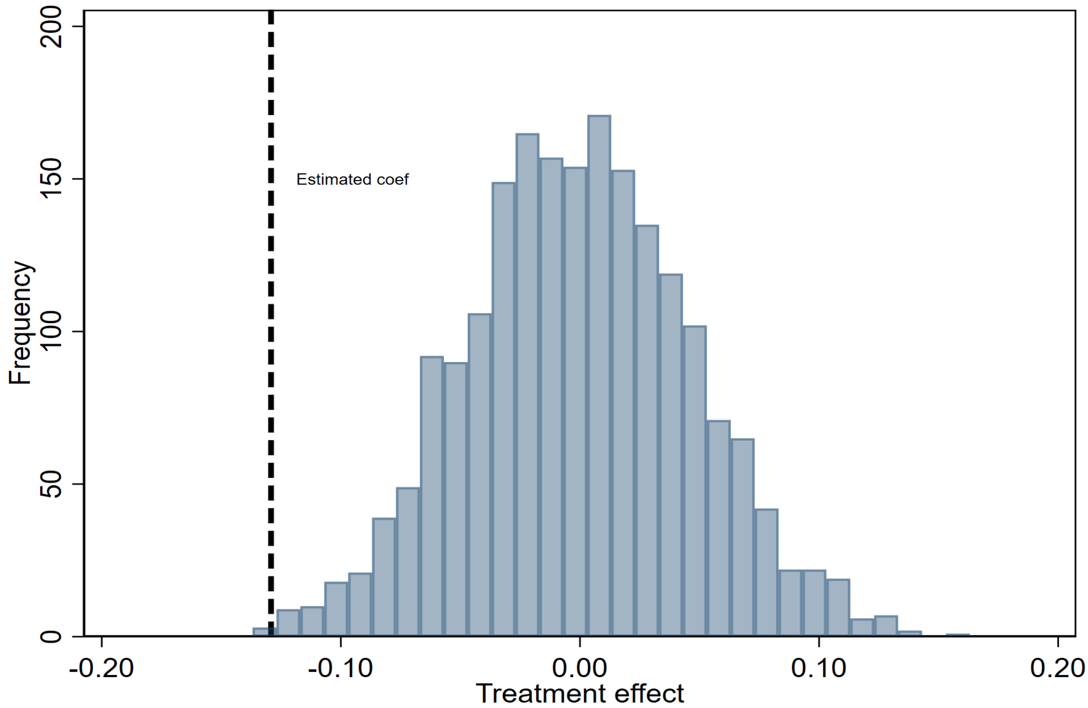



*Figure 3.A.1:* Randomization Test

Notes: This Figure shows the result of a randomization test (`ritest`, (Heß, 2017)). If the confrontation treatment had not affected any officer's case-handling at all, then the observed outcome for our main variable (i.e. prioritizing the victim's statement) would have occurred with a probability of below 0.2%, as measured by the p-value (0.002).

*Table 3.A.1:* Balancing Tests

| Variable | (1) Mean treated men | (2) Mean control men | (3) Mean treated women | (4) Mean control women | (5) Diff treated vs control men | (6) Diff treated vs control women | (7) Diff treated men vs treated women |
|---|---|---|---|---|---|---|---|
| Officer age | 45.152 | 44.543 | 31.440 | 31.909 | -0.540 | -0.554 | -9.383*** |
| | (9.687) | (9.414) | (7.030) | (8.431) | (1.321) | (2.161) | (2.626) |
| Officer posting in Bhopal (=1) | 0.420 | 0.394 | 0.320 | 0.273 | 0.119 | 0.032 | -0.119 |
| | (0.496) | (0.491) | (0.476) | (0.452) | (0.085) | (0.161) | (0.254) |
| GBV case 1: prio on victim (=1) | 0.109 | 0.324 | 0.222 | 0.368 | -0.214*** | 0.144 | 0.105 |
| | (0.315) | (0.471) | (0.441) | (0.496) | (0.072) | (0.294) | (0.127) |
| GBV case 1: response time [seconds] | 222.184 | 215.766 | 201.661 | 184.330 | -7.182 | 29.721 | 1.759 |
| | (90.360) | (95.273) | (98.161) | (128.191) | (23.536) | (27.710) | (45.451) |
| GBV case 2: prio on victim (=1) | 0.351 | 0.464 | 0.438 | 0.500 | -0.053 | -0.302* | -0.022 |
| | (0.481) | (0.503) | (0.512) | (0.519) | (0.118) | (0.145) | (0.208) |
| GBV case 2: response time [seconds] | 231.783 | 201.703 | 171.268 | 172.088 | 32.871 | -0.590 | -79.033 |
| | (117.034) | (105.139) | (60.482) | (59.732) | (35.745) | (22.505) | (45.244) |
| Non-GBV case q1: investigation of offender (=1) | 0.643 | 0.661 | 0.640 | 0.788 | 0.036 | -0.125 | -0.072 |
| | (0.481) | (0.475) | (0.490) | (0.415) | (0.070) | (0.195) | (0.102) |
| Non-GBV case q1: response time [seconds] | 335.614 | 283.763 | 259.262 | 330.310 | 35.977* | -48.319 | -91.789* |
| | (182.296) | (136.544) | (123.890) | (237.295) | (20.416) | (56.253) | (50.696) |
| Non-GBV case q2: prio on victim (=1) | 0.938 | 0.921 | 0.920 | 0.970 | 0.034 | -0.057 | -0.077 |
| | (0.243) | (0.270) | (0.277) | (0.174) | (0.049) | (0.066) | (0.092) |
| Non-GBV case q2: response time [seconds] | 71.293 | 65.774 | 53.542 | 56.433 | 5.369 | -7.202 | -19.587** |
| | (44.921) | (40.579) | (23.140) | (40.938) | (4.113) | (6.626) | (7.780) |
| Senior officer tone during callout | 2.080 | 2.142 | 2.240 | 2.242 | 0.004 | 0.000 | -0.022 |
| | (0.304) | (0.431) | (0.523) | (0.561) | (0.004) | (0.000) | (0.024) |
| Observations | 112 | 127 | 25 | 33 | 239 | 58 | 137 |

Notes: This Table compares the treatment and control group in terms of their officer characteristics and baseline outcomes-both within gender (Columns 5 and 6) and across gender dimensions (Column 7). The displayed differences are regression adjusted using baseline session and senior officer as controls and cluster at baseline session level. The values in Columns (1)-(4) are simple means. 'Officer age' is the participant's age in years, 'Officer posting in Bhopal' is a dummy that takes the value 1 if the officers serves in the capital Bhopal, 'GBV case 1 or 2: prio on victim' is a dummy that takes the value 1 if participants prioritized the victim's statement (ranked as most important) in the GBV case 1 or 2, 'GBV case 1 or 2: response time' is the time it took the participants to answer to the survey question on prioritizing statements measured in seconds, 'Non-GBV case q1: investigation of offender' is a dummy that takes the value 1 if the officer ranks the investigation the offender first (instead of the victim or other people involved) in the non-GBV case, 'Non-GBV case q1: response time' measures the response time for ranking the investigation parts in the non-GBV case, 'Non-GBV case q2: prio on victim' is a dummy that takes the value 1 if the victim's statement is ranked as most important, 'Non-GBV case q2: response time' measures the officers' time to respond to the Non-GBV case q2 question, and 'Senior officer tone during callout' is a categorial variable that was coded by the research team during the feedback session with the senior officer and indicates whether the tone of the senior officer was 1:rebuking 2:neutral 3:explanatory 4:reading from script. *p<.10, **p<.05, ***p<.01

*Table 3.A.2:* The Effects of Confrontation on the Recognition of Crime

| | Considered as crime |
|---|---|
| | (1) |
| **Panel A: Without gender heterogeneity** | |
| GBV treatment | 0.014 |
| | (0.028) |
| Controls | yes |
| Control Group Mean | 0.949 |
| Rsq | 0.050 |
| N | 293 |
| **Panel B: With gender heterogeneity** | |
| GBV treatment | -0.002 |
| | (0.036) |
| treatment x female | 0.081 |
| | (0.054) |
| female officer | -0.011 |
| | (0.046) |
| Controls | yes |
| Control Group Mean | 0.949 |
| Rsq | 0.058 |
| N | 293 |

Notes: This Table reports estimates of the effect of the intervention on whether the officer recognizes that a crime has committed. 'GBV treatment' is an indicator for being confronted with one's own bias in the GBV case handling. 'treatment x female' is an interaction term of this indicator and the gender dummy (=1 if female). The specification includes senior officer and session fixed effects, and officer characteristics as controls (officer's age and a dummy for posting is in the capital Bhopal). Standard errors are clustered at the session level and are reported in parentheses. The sample is reduced to 293 officers, as four of them did not answer the corresponding survey question.
*p<.10, **p<.05, ***p<.01

*Table 3.A.3:* The Effects of Confrontation on the Handling of GBV Cases 2

| | Truth of rape complaints | | Register complaint | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| **Panel A: Without gender heterogeneity** | | | | |
| GBV treatment | -0.654 | -0.686 | 0.034 | 0.038 |
| | (0.454) | (0.448) | (0.032) | (0.029) |
| | [0.132] | [0.132] | [0.305] | [0.305] |
| Controls | no | yes | no | yes |
| Control Group Mean | 5.325 | 5.325 | 0.931 | 0.931 |
| Rsq | 0.058 | 0.080 | 0.066 | 0.092 |
| N | 297 | 297 | 297 | 297 |
| **Panel B: With gender heterogeneity** | | | | |
| GBV treatment | -0.677 | -0.706 | 0.044 | 0.048 |
| | (0.403) | (0.411) | (0.039) | (0.036) |
| | [0.095] | [0.095] | [0.285] | [0.285] |
| treatment x female | 0.131 | 0.112 | -0.050 | -0.050 |
| | (0.813) | (0.870) | (0.040) | (0.038) |
| | [0.700] | [0.700] | [0.285] | [0.285] |
| female officer | 0.877** | 0.517 | 0.067** | 0.036 |
| | (0.364) | (0.501) | (0.023) | (0.026) |
| Controls | no | yes | no | yes |
| Control Group Mean | 5.325 | 5.325 | 0.931 | 0.931 |
| Rsq | 0.071 | 0.084 | 0.074 | 0.094 |
| N | 297 | 297 | 297 | 297 |

Notes: This Table reports estimates of the effect of the intervention on the belief in the truthfulness of rape complaints (Columns 1 and 2) and the likelihood of registering a GBV complaint (dummy=1 if the complaint is registered) against the accused (Columns 3 and 4). The truthfulness of rape complaints is measured as the number of complaints (out of 10) that are considered false. 'GBV treatment' is an indicator for being confronted with one's own bias in the GBV case handling. 'treatment x female' is an interaction term of this indicator and the gender dummy (=1 if female). The specification in Columns (1) and (3) includes senior officer and session fixed effects. In Columns (2) and (4) we add officer characteristics as controls (officer's age and a dummy for posting is in the capital Bhopal). Standard errors are clustered at the session level and are reported in parentheses. Bootstrapped p-values of multiple hypothesis tests are shown in square brackets. We use the Stata command mhtreg for multiple hypotheses testing that provides a procedure that asymptotically controls family-wise error rate and is asymptotically balanced (Steinmayer, 2020). It is based on List, Shaikh, and Xu (2019) but modified to be used in a multivariate regression setting.
*p<.10, **p<.05, ***p<.01

*Table 3.A.4:* The Effect of Confrontation on the Handling of GBV Cases: Falsification Test

| | Prio on victim's statement |
|---|---|
| | (1) |
| GBV treatment | -0.085 |
| | (0.063) |
| treatment x female | 0.287*** |
| | (0.082) |
| female officer | -0.154** |
| | (0.063) |
| Controls | yes |
| Control Group Mean | 0.844 |
| Rsq | 0.093 |
| N | 239 |

Notes: This Table reports estimates of the effect of the intervention on the likelihood of prioritizing the victim's statement. The outcome takes the value 1 if the priority is given to the victim's statement. Male officers who showed a strong bias at baseline are removed from the sample. 'GBV treatment' is an indicator for being confronted with one's own bias in the GBV case handling. 'treatment x female' is an interaction term of this indicator and the gender dummy (=1 if female). The specification includes senior officer and session fixed effects and officer characteristics as controls (officer's age and a dummy for posting is in the capital Bhopal). Standard errors are clustered at the session level and are reported in parentheses. *p<.10, **p<.05, ***p<.01

*Table 3.A.5:* Robustness: The Effect of Confrontation on Prioritizing Victim Statement

| | Drop complaint | Prio on victim's statement | Prio on offender's statement |
|---|---|---|---|
| | (1) | (2) | (3) |
| GBV treatment | 0.047 | -0.135** | 0.088** |
| | (0.041) | (0.053) | (0.039) |
| | [0.249] | [0.012] | [0.027] |
| treatment x female | -0.108* | 0.347*** | -0.239*** |
| | (0.059) | (0.089) | (0.074) |
| | [0.072] | [0.000] | [0.001] |
| female officer | -0.026 | -0.070 | 0.096 |
| | (0.049) | (0.078) | (0.066) |
| Controls | no | no | no |
| Control Group Mean | 0.081 | 0.844 | 0.075 |
| Rsq | 0.017 | 0.045 | 0.031 |
| N | 297 | 297 | 297 |

Notes: This Table reports estimates of the effect of the intervention on the likelihood of dropping a GBV complaint (Column 1), the likelihood of prioritizing the victim's statement (Column 2) and prioritizing the offender's statement (Column 3). All three outcomes are dummies and take the value 1 if the complaint is not pursued, the priority is given to the victim's statement, and the priority is given to the offender's statement, respectively. 'GBV treatment' is an indicator for being confronted with one's own bias in the GBV case handling. 'treatment x female' is an interaction term of this indicator and the gender dummy (=1 if female). The specification does not include additional controls and robust standard errors are implemented. P-values obtained by bootstrapping with replacement with 999 replications are reported in square brackets.
*p<.10, **p<.05, ***p<.01

*Table 3.A.6:* Robustness: The Effect of Confrontation on Prioritizing Victim Statement-Including Baseline Controls

| | Drop complaint | Prio on victim's statement | Prio on offender's statement |
|---|---|---|---|
| | (1) | (2) | (3) |
| **Panel A: Including Baseline non-GBV case controls** | | | |
| GBV treatment | 0.036 | -0.133** | 0.097* |
| | (0.033) | (0.055) | (0.048) |
| treatment x female | -0.119** | 0.378*** | -0.258*** |
| | (0.052) | (0.082) | (0.062) |
| female officer | 0.021 | -0.151** | 0.131** |
| | (0.050) | (0.070) | (0.048) |
| Control Group Mean | 0.081 | 0.844 | 0.075 |
| Rsq | 0.079 | 0.089 | 0.104 |
| N | 297 | 297 | 297 |
| **Panel B: Including Baseline non-GBV case + GBV case 1 controls** | | | |
| GBV treatment | 0.016 | -0.105* | 0.089* |
| | (0.041) | (0.057) | (0.047) |
| treatment x female | -0.112* | 0.295*** | -0.183** |
| | (0.063) | (0.079) | (0.075) |
| female officer | 0.001 | -0.073 | 0.072 |
| | (0.055) | (0.072) | (0.066) |
| Control Group Mean | 0.081 | 0.844 | 0.075 |
| Rsq | 0.143 | 0.121 | 0.177 |
| N | 154 | 154 | 154 |
| **Panel C: Including Baseline non-GBV case + GBV case 2 controls** | | | |
| GBV treatment | 0.091* | -0.178 | 0.087 |
| | (0.049) | (0.105) | (0.081) |
| treatment x female | -0.100 | 0.436** | -0.336** |
| | (0.085) | (0.151) | (0.116) |
| female officer | -0.012 | -0.230* | 0.242** |
| | (0.058) | (0.118) | (0.100) |
| Control Group Mean | 0.081 | 0.844 | 0.075 |
| Rsq | 0.177 | 0.180 | 0.184 |
| N | 143 | 143 | 143 |

Notes: This Table presents estimates for the same specification as in Panel B of Table 3.1, but includes Baseline outcomes that appeared statistically significant in our balance checks (Table 3.A.1). In Panel A, we control for the following variables: officer's age, posting, and response time for ranking parts of the investigation in the non-GBV case. In Panel B, we control for the same set of variables but add the variable 'GBV case 1: prio on victim' (whether the officer prioritizes the victim's statement in the GBV case). The sample is smaller once this variable is added, as only half of the participants handled this case at baseline (every participant handles one of two GBV cases). In Panel C, the Baseline controls on the non-GBV case from Panel A and the variable 'GBV case 2: prio on victim' (whether the officer prioritizes the victim's statement in the GBV case) are included. Refer to the table notes of Table 3.1 and 3.A.1.
*p<.10, **p<.05, ***p<.01

*Table 3.A.7:* The Effects of Confrontation on Guilt and Self-Reflection

| | Guilty | Angry | Disappointed | Regretful | Ashamed | Annoyed |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| GBV treatment | -0.353 | 0.197 | -0.133 | -0.569** | 0.164 | -0.155 |
| | (0.259) | (0.213) | (0.159) | (0.232) | (0.186) | (0.217) |
| treatment x female | -0.133 | -0.604 | 0.081 | 0.057 | -0.155 | -0.695 |
| | (0.610) | (0.726) | (0.355) | (0.512) | (0.329) | (0.401) |
| female officer | -0.073 | 0.611 | -0.241 | -0.044 | 0.012 | 0.200 |
| | (0.398) | (0.369) | (0.365) | (0.443) | (0.196) | (0.403) |
| Controls | yes | yes | yes | yes | yes | yes |
| Control Group Mean | 0.214 | 2.225 | 1.975 | 2.819 | 1.525 | 2.206 |
| Rsq | 0.077 | 0.131 | 0.068 | 0.106 | 0.061 | 0.107 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

Notes: This Table shows our main specification (Panel B of Table 3.1) and the outcomes feeling 'guilty' (Column 1), 'angry at myself' (Column 3), 'regretful' (Column 4), 'ashamed' (Column 5), and 'annoyed' (Column 6). 'Guilty' is coded up as an index, specified by Chaney, Sanchez, Alt, and Shih (2021); the remaining variables are categorical variables (scale: 1 Not at all, 2 Doesn't apply to a great extent, 3 Applies to some extent, 4 Neither nor, 5 Somewhat applies, 6 Applies to a great extent, 7 Very much). Refer also to the table notes of our main Table 3.1. *p<.10, **p<.05, ***p<.01

*Table 3.A.8:* The Effects of Confrontation on Victim Blaming and Perceived Bias

| | Victim blaming husband beats | Victim blaming harassment | Victim blaming rape | Bias of friends | Bias of family | Bias of partner |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| GBV treatment | -0.407 | -0.384 | 0.104 | -0.168 | -0.087 | -0.015 |
| | (0.265) | (0.339) | (0.273) | (0.167) | (0.138) | (0.128) |
| treatment x female | 0.508 | 0.073 | -0.489 | 0.307 | 0.294 | 0.225 |
| | (0.884) | (0.733) | (0.729) | (0.372) | (0.272) | (0.290) |
| female officer | 0.532 | 0.650 | 0.441 | -0.282 | -0.226 | 0.045 |
| | (0.707) | (0.576) | (0.633) | (0.194) | (0.190) | (0.178) |
| Controls | yes | yes | yes | yes | yes | yes |
| Control Group Mean | 3.675 | 2.587 | 5.544 | 2.094 | 2.094 | 1.744 |
| Rsq | 0.107 | 0.102 | 0.105 | 0.124 | 0.093 | 0.090 |
| N | 297 | 297 | 297 | 297 | 297 | 297 |

Notes: This Table shows our main specification (Panel B of Table 3.1) and the outcomes 'victim blaming when husband beats up his wife' (Column 1), 'victim blaming when a woman is harassed' (Column 3), 'victim blaming when a woman is raped' (Column 4), where the officer is asked to indicate in how many (out of 10) cases she thinks it is the women's fault, respectively. Columns (4)-(6) show the officers' perceived bias among friends, family, and their partner, respectively, where participants were asked to what extent their friends, family, and partner think that gender based violence is the women's fault (scale: 1 0%, 2 About 25%, 3 About 50%, 4 About 75%, 5 100%). Refer also to the table notes of our main Table 3.1.
*p<.10, **p<.05, ***p<.01

# Bibliography


Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020): "Sampling-Based versus Design-Based Uncertainty in Regression Analysis," *Econometrica*, 88(1), 265–296.

Agan, A. Y., and M. D. Makowsky (2018): "The minimum wage, EITC, and criminal recidivism," Discussion paper, National Bureau of Economic Research.

Aizer, A. (2010): "The gender wage gap and domestic violence," *American Economic Review*, 100(4), 1847–1859.

——— (2011): "Poverty, violence, and health the impact of domestic violence during pregnancy on newborn health," *Journal of Human Resources*, 46(3), 518–538.

Aizer, A., and P. Dal Bo (2009): "Love, hate and murder: Commitment devices in violent relationships," *Journal of Public Economics*, 93(3-4), 412–428.

Aizer, A., and J. J. Doyle Jr (2015): "Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges," *Quarterly Journal of Economics*, 130(2), 759–803.

Alesina, A., B. Brioschi, and E. La Ferrara (2021): "Violence against women: a cross-cultural analysis for Africa," *Economica*, 88(349), 70–104.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018): "Revealing stereotypes: Evidence from immigrants in schools," *National Bureau of Economic Research*.

Amaral, S., G. B. Dahl, T. Hener, V. Kaiser, and H. Rainer (2023): "Deterrence or backlash? Arrests and the dynamics of domestic violence," Discussion paper, National Bureau of Economic Research.

Anderberg, D., H. Rainer, J. Wadsworth, and T. Wilson (2016): "Unemployment and domestic violence: Theory and evidence," *The Economic Journal*, 126(597), 1947–1979.

Andres, L., M. Fabel, and H. Rainer (2023): "How much violence does football hooliganism cause?," *Journal of Public Economics*, 225, 104970.

Angrist, J. D. (2006): "Instrumental variables methods in experimental criminological research: What, why and how," *Journal of Experimental Criminology*, 2(1), 23–44.

Autor, D., A. Kostol, M. Mogstad, and B. Setzler (2019): "Disability Benefits, Consumption Insurance, and Household Labor Supply," *American Economic Review*, 109(7), 2613–54.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020): "Women's empowerment in action: evidence from a randomized control trial in Africa," *American Economic Journal: Applied Economics*, 12(1), 210–59.

Bauer, T. K., C. Rulff, and M. M. Tamminga (2019): *Berlin calling-Internal migration in Germany*, no. 823. Ruhr Economic Papers.

BBSR Bonn (2019): "Indikatoren und Karten zur Raum- und Stadtentwicklung (INKAR)," *https://www.inkar.de/*.

Beauchamp, A., and S. Chan (2014): "The minimum wage and crime," *The BE Journal of Economic Analysis & Policy*, 14(3), 1213–1235.

Bellmann, L., M. Bossler, H.-D. Gerner, and O. Hübler (2015): "IAB-betriebspanel: reichweite des Mindestlohns in deutschen Betrieben," Discussion paper, IAB-Kurzbericht.

Berk, R. A. (1993): "What the Scientific Evidence Shows: On Average, We Can Do No Better Than Arrest," in *Current Controversies on Familiy Violence*, ed. by R. J. Gelles, and D. R. Loseke, pp. 323–336. Sage Publications.

Berk, R. A., A. Campbell, R. Klap, and B. Western (1992): "A Bayesian analysis of the Colorado Springs spouse abuse experiment," *Journal of Criminal Law and Criminology*, 83(1), 170–200.

Bhalotra, S., D. GC Britto, P. Pinotti, and B. Sampaio (2021): "Job displacement, unemployment benefits and domestic violence," *CEPR Discussion Paper No. DP16350*.

Bhalotra, S. R., U. S. Kambhampati, S. Rawlings, and Z. Siddique (2018): "Intimate partner violence and the business cycle," *IZA Discussion Paper*.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020): "Incarceration, recidivism, and employment," *Journal of Political Economy*, 128(4), 1269–1324.

Bindler, A., N. Ketel, and R. Hjalmarsson (2020): "Costs of victimization," *Handbook of labor, human resources and population economics*, pp. 1–31.

BKA (2022): "Partnerschaftsgewalt - Kriminalstatistische Auswertung Berichtsjahr 2021," `https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/Lagebilder/Partnerschaftsgewalt/partnerschaftsgewalt_node.html`, Accessed: 2023-09-16.

Black, D. A., J. Grogger, K. Sanders, and T. Kirchmaier (2022): “Criminal charges, risk assessment, and violent recidivism in cases of domestic abuse,” *Working Paper*.

BMFSFJ (2014): “Lebenssituation, Sicherheit und Gesundheit von Frauen in Deutschland. Eine repräsentative Untersuchung zu Gewalt gegen Frauen in Deutschland,” `https://www.bmfsfj.de/bmfsfj/studie-lebenssituation/sicherheit-und-gesundheit-von-frauen-in-deutschland-80694`, Accessed: 2023-09-16.

Bonin, H., I. E. Isphording, A. Krause-Pilatus, A. Lichter, N. Pestel, U. Rinne, et al. (2018): “Auswirkungen des gesetzlichen Mindestlohns auf Beschäftigung, Arbeitszeit und Arbeitslosigkeit,” Discussion paper, Institute of Labor Economics (IZA).

Boone, J., and J. C. Van Ours (2006): “Are recessions good for workplace safety?,” *Journal of Health Economics*, 25(6), 1069–1093.

Bundesamt für Migration und Flüchtlinge (2016): “Migrationsbericht 2015,” `https://www.bamf.de/SharedDocs/Anlagen/DE/Forschung/Migrationsberichte/migrationsbericht-2015.html?nn=403964`, Accessed: 2023-09-16.

Bundesministerium für Wirtschaft (2022): “Migrationsbericht 2015,” `https://www.bmas.de/DE/Soziales/Sozialversicherung/Geringfuegige-Beschaeftigung/mini-jobs.html`, Accessed: 2023-09-16.

Caliendo, M., A. Fedorets, M. Preuss, C. Schröder, and L. Wittbrodt (2018): “The short-run employment effects of the German minimum wage reform,” *Labour Economics*, 53, 46–62.

Caliendo, M., L. Wittbrodt, and C. Schröder (2019): “The causal effects of the minimum wage introduction in Germany–an overview,” *German Economic Review*, 20(3), 257–292.

Campolieti, M., T. Fang, and M. Gunderson (2005): “Minimum wage impacts on youth employment transitions, 1993–1999,” *Canadian Journal of Economics/Revue canadienne d’économique*, 38(1), 81–104.

Card, D., and G. B. Dahl (2011): “Family violence and football: The effect of unexpected emotional cues on violent behavior,” *Quarterly Journal of Economics*, 126(1), 103–143.

Chalfin, A. (2015): “Economic costs of crime,” *The encyclopedia of crime and punishment*, pp. 1–12.

Chaney, K. E., and D. T. Sanchez (2018): “The endurance of interpersonal confrontations as a prejudice reduction strategy,” *Personality and Social Psychology Bulletin*, 44(3), 418–429.

Chaney, K. E., D. T. Sanchez, N. P. Alt, and M. J. Shih (2021): "The breadth of confrontations as a prejudice reduction strategy," *Social Psychological and Personality Science*, 12(3), 314–322.

Chaney, K. E., D. M. Young, and D. T. Sanchez (2015): "Confrontation's health outcomes and promotion of egalitarianism (C-HOPE) framework.," *Translational Issues in Psychological Science*, 1(4), 363.

Chin, Y.-M., and S. Cunningham (2019): "Revisiting the effect of warrantless domestic violence arrest laws on intimate partner homicides," *Journal of Public Economics*, 179, 104072.

College of Policing (2022): "First response: Authorised Professional Practice," `https://www.college.police.uk/app/major-investigation-and-public-protection/domestic-abuse/first-response`, Accessed: 2022-11-03.

Cools, S., and A. Kotsadam (2017): "Resources and intimate partner violence in Sub-Saharan Africa," *World Development*, 95, 211–230.

Correia, S., P. Guimarães, and T. Zylkin (2020): "Fast Poisson estimation with high-dimensional fixed effects," *The Stata Journal*, 20(1), 95–115.

Currie, J., M. Mueller-Smith, and M. Rossin-Slater (2022): "Violence while in utero: The impact of assaults during pregnancy on birth outcomes," *Review of Economics and Statistics*, 104(3), 525–540.

Czopp, A. M., M. J. Monteith, and A. Y. Mark (2006): "Standing up for a change: Reducing bias through interpersonal confrontation.," *Journal of personality and social psychology*, 90(5), 784.

Dahl, G. B., and M. M. Knepper (2021): "Why is workplace sexual harassment underreported? The value of outside options amid the threat of retaliation," Working Paper 29248, National Bureau of Economic Research.

Dahl, G. B., A. R. Kostøl, and M. Mogstad (2014): "Family welfare cultures," *Quarterly Journal of Economics*, 129(4), 1711–1752.

Degener, T., and Y. Koster-Dreese (1995): "Declaration on the Elimination of Violence Against Women: by General Assembly Resolution 48/104 of 20 December 1993," in *Human Rights and Disabled Persons*, pp. 416–422. Brill Nijhoff.

Di Tella, R., and E. Schargrodsky (2013): "Criminal recidivism after prison and electronic monitoring," *Journal of Political Economy*, 121(1), 28–73.

Dobbie, W., J. Goldin, and C. S. Yang (2018): "The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges," *American Economic Review*, 108(2), 201–40.

DOBBIE, W., AND J. SONG (2015): "Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection," *American Economic Review*, 105(3), 1272–1311.

DOYLE JR, J. J. (2007): "Child protection and child outcomes: Measuring the effects of foster care," *American Economic Review*, 97(5), 1583–1610.

——— (2008): "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care," *Journal of Political Economy*, 116(4), 746–770.

DUNFORD, F. W., D. HUIZINGA, AND D. S. ELLIOTT (1990): "The role of arrest in domestic assault: The Omaha police experiment," *Criminology*, 28(2), 183–206.

DUSTMANN, C., A. LINDNER, U. SCHÖNBERG, M. UMKEHRER, AND P. VOM BERGE (2022): "Reallocation effects of the minimum wage," *The Quarterly Journal of Economics*, 137(1), 267–328.

ERTEN, B., AND P. KESKIN (2021a): "Female employment and intimate partner violence: evidence from Syrian refugee inflows to Turkey," *Journal of Development Economics*, 150, 102607.

——— (2021b): "Trade-Offs? The Impact of WTO Accession on Intimate Partner Violence in Cambodia," *The Review of Economics and Statistics*, 12, 1–40.

EUROPEAN INSTITUTE FOR GENDER EQUALITY (2014): "Estimating the costs of gender-based violence in the European Union: Report," `https://eige.europa.eu/publications-resources/publications/estimating-costs-gender-based-violence-european-union-report`, Accessed: 2023-09-09.

EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS (2014): "Violence against women: an EU-wide survey. Main results," .

FERNANDEZ, J., T. HOLMAN, AND J. V. PEPPER (2014): "The impact of living-wage ordinances on urban crime," *Industrial Relations: A Journal of Economy and Society*, 53(3), 478–500.

FITZENBERGER, B., AND A. DOERR (2016): "Konzeptionelle Lehren aus der ersten Evaluationsrunde der Branchenmindestlöhne in Deutschland," *Journal for Labour Market Research*, 49(4), 329–347.

FLEURY, R. E., C. M. SULLIVAN, D. I. BYBEE, AND W. S. DAVIDSON II (1998): "Why don't they just call the cops?: Reasons for differential police contact among women with abusive partners," *Violence and Victims*, 13(4), 333–346.

FONE, Z. S., J. J. SABIA, AND R. CESUR (2023): "The unintended effects of minimum wage increases on crime," *Journal of Public Economics*, 219, 104780.

FRENCH, E., AND J. SONG (2014): "The effect of disability insurance receipt on labor supply," *American Economic Journal: Economic Policy*, 6(2), 291–337.

García-Moreno, C., C. Zimmerman, A. Morris-Gehring, L. Heise, A. Amin, N. Abrahams, O. Montoya, P. Bhate-Deosthali, N. Kilonzo, and C. Watts (2015): "Addressing violence against women: a call to action," *The Lancet*, 385(9978), 1685–1695.

Golestani, A., E. Owens, and K. Raissian (2021): "Specialization in Criminal Courts: Decision Making, Recidivism, and Re-victimization in Domestic Violence Courts in Tennessee," *Unpublished manuscript*.

González, L., and N. Rodríguez-Planas (2020): "Gender norms and intimate partner violence," *Journal of Economic Behavior & Organization*, 178, 223–248.

Goodmark, L. (2018): *Decriminalizing domestic violence*. University of California Press.

Gould, E. D., B. A. Weinberg, and D. B. Mustard (2002): "Crime rates and local labor market opportunities in the United States: 1979–1997," *Review of Economics and statistics*, 84(1), 45–61.

Green, D. P., A. M. Wilke, and J. Cooper (2020): "Countering violence against women by encouraging disclosure: A mass media experiment in rural Uganda," *Comparative Political Studies*, 53(14), 2283–2320.

Grogger, J., S. Gupta, R. Ivandic, and T. Kirchmaier (2021): "Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases," *Journal of Empirical Legal Studies*, 18(1), 90–130.

Guarnieri, E., and H. Rainer (2021): "Colonialism and female empowerment: A two-sided legacy," *Journal of Development Economics*, 151, 102666.

Halvorsen, R., and R. Palmquist (1980): "The interpretation of dummy variables in semilogarithmic equations," *American Economic Review*, 70(3), 474–75.

Hansen, K., and S. Machin (2002): "Spatial crime patterns and the introduction of the UK minimum wage," *Oxford Bulletin of Economics and Statistics*, 64, 677–697.

Hashimoto, M. (1987): "The minimum wage law and youth crimes: time-series evidence," *The Journal of Law and Economics*, 30(2), 443–464.

Hess, S. (2017): "Randomization inference with Stata: A guide and software," *The Stata Journal*, 17(3), 630–651.

Hidrobo, M., A. Peterman, and L. Heise (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(3), 284–303.

Hirschel, J. D., and I. W. Hutchison III (1992): "Female spouse abuse and the police response: The Charlotte, North Carolina experiment," *Journal of Criminal Law and Criminology*, 83(1), 73–119.

HM Inspectorate of Constabulary (2014): "Everyone's business: Improving the police response to domestic abuse," https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/2014/04/improving-the-police-response-to-domestic-abuse.pdf, Accessed: 2022-05-17.

Howell, J. L., L. Redford, G. Pogge, and K. A. Ratliff (2017): "Defensive responding to IAT feedback," *Social Cognition*, 35(5), 520–562.

Iyengar, R. (2009): "Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws," *Journal of Public Economics*, 93(1-2), 85–98.

Jayachandran, S. (2015): "The roots of gender inequality in developing countries," *Annual Review of Economics*, 7(1), 63–88.

Katz, L. F., J. R. Kling, and J. B. Liebman (2001): "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *Quarterly Journal of Economics*, 116(2), 607–654.

Kling, J. R. (2006): "Incarceration length, employment, and earnings," *American Economic Review*, 96(3), 863–876.

Lindo, J. M., J. Schaller, and B. Hansen (2018): "Caution! Men not at work: Gender-specific labor market conditions and child maltreatment," *Journal of Public Economics*, 163, 77–98.

List, J. A., A. M. Shaikh, and Y. Xu (2019): "Multiple hypothesis testing in experimental economics," *Experimental Economics*, 22, 773–793.

Lofstrom, M., and S. Raphael (2016): "Crime, the criminal justice system, and socioeconomic inequality," *Journal of Economic Perspectives*, 30(2), 103–126.

Maestas, N., K. J. Mullen, and A. Strand (2013): "Does disability insurance receipt discourage work? Using examiner assignment to estimate causal effects of SSDI receipt," *American Economic Review*, 103(5), 1797–1829.

Mallett, R. K., and D. E. Wagner (2011): "The unexpectedly positive consequences of confronting sexism," *Journal of Experimental Social Psychology*, 47(1), 215–220.

Marbach, M., and D. Hangartner (2020): "Profiling compliers and noncompliers for instrumental-variable analysis," *Political Analysis*, 28(3), 435–444.

Miller, A. R., and C. Segal (2019): "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence," *The Review of Economic Studies*, 86(5), 2220–2247.

National Research Council (1998): *Violence in families: Assessing prevention and treatment programs*. National Academies Press.

Neumark, D., and S. Adams (2003): "Do living wage ordinances reduce urban poverty?," *Journal of Human Resources*, 38(3), 490–521.

Office for National Statistics (2013): "Homicide," https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/compendium/focusonviolentcrimeandsexualoffences/yearendingmarch2016/homicide#how-are-victims-and-suspects-related, Accessed: 2021-09-29.

——— (2020): "England and Wales 2011 Census," https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest#ethnic-groups-by-area, Accessed: 2022-06-24.

Palermo, T., J. Bleck, and A. Peterman (2014): "Tip of the iceberg: reporting and gender-based violence in developing countries," *American journal of epidemiology*, 179(5), 602–612.

Pate, A. M., and E. E. Hamilton (1992): "Formal and Informal Deterrents to Domestic Violence: The Dade County Spouse Assault Experiment," *American Sociological Review*, 57(5), 691–697.

Pinotti, P. (2020): "The credibility revolution in the empirical analysis of crime," *Italian Economic Journal*, 6(2), 207–220.

Pollak, R. A. (2005): "Bargaining power in marriage: Earnings, wage rates and household production," National Bureau of Economic Research Cambridge, Mass., USA.

Sardinha, L., M. Maheu-Giroux, H. Stöckl, S. R. Meyer, and C. García-Moreno (2022): "Global, regional, and national prevalence estimates of physical or sexual, or both, intimate partner violence against women in 2018," *The Lancet*, 399(10327), 803–813.

Schmidt, J. D., and L. W. Sherman (1993): "Does arrest deter domestic violence?," *American Behavioral Scientist*, 36(5), 601–609.

Sherman, L. W., and R. A. Berk (1984): "The specific deterrent effects of arrest for domestic assault," *American Sociological Review*, pp. 261–272.

Sherman, L. W., J. D. Schmidt, D. P. Rogan, and D. A. Smith (1992): "The variable effects of arrest on criminal careers: The Milwaukee domestic violence experiment," *Journal of Criminal Law and Criminology*, 83(1), 137–169.

Spencer, R. A., M. D. Livingston, B. Woods-Jaeger, S. T. Rentmeester, N. Sroczynski, and K. A. Komro (2020): "The impact of temporary assistance for needy families, minimum wage, and Earned Income Tax Credit on Women's well-being and intimate partner violence victimization," *Social Science & Medicine*, 266, 113355.

Statistisches Bundesamt (2014): "Verdienststrukturerhebung (VSE)," Wiesbaden.

Steinmayer, A. (2020): "MHTREG: Stata module for multiple hypothesis correction," *Statistical Software Components S458853, Boston College Department of Economics*.

Stevenson, B., and J. Wolfers (2006): "Bargaining in the shadow of the law: Divorce laws and family distress," *Quarterly Journal of Economics*, 121(1), 267–288.

Tjaden, P. G., and N. Thoennes (2000): *Full report of the prevalence, incidence, and consequences of violence against women: Findings from the National Violence Against Women Survey*. US Department of Justice, Office of Justice Programs, National Institute of Justice.

Tur-Prats, A. (2019): "Family types and intimate partner violence: A historical perspective," *Review of Economics and Statistics*, 101(5), 878–891.

——— (2021): "Unemployment and intimate partner violence: A cultural approach," *Journal of Economic Behavior & Organization*, 185, 27–49.

UNODC (2019): "Global study on homicide," `https://www.unodc.org/documents/data-and-analysis/gsh/Booklet_5.pdf`, Accessed: 2023-09-09.

Walby, S., and J. Allen (2004): *Domestic violence, sexual assault and stalking: Findings from the British Crime Survey*. Home Office.

WHO (2002): "World Report on Violence and Health," Discussion paper.

——— (2021): "Violence against women prevalence estimates, 2018: Global, regional and national prevalence estimates for intimate partner violence against women and global and regional prevalence estimates for non-partner sexual violence against women," Discussion paper.

Yang, C. S. (2017): "Local labor markets and criminal recidivism," *Journal of Public Economics*, 147, 16–29.