ESRC SDAI Application

1 Intro

Domestic abuse is a complex phenomenon affecting people from all walks of life. It is increasingly recognised as a major public policy concern in many countries, including the UK (European Parliament, 2018). While anyone can become a victim of domestic abuse, women are disproportionately affected, with more than 25% of women, and 15% of men in England and Wales reported to have experienced some form of domestic abuse since the age of 16 (Office for National Statistics, 2018). The current legal definition in the UK (Home Office, 2018) aims to capture the multifaceted nature of domestic abuse, by recognising that domestic abuse encompasses a wide range of behaviours, including emotional, sexual, and physical abuse, threatening, intimidating, coercive and controlling behaviour.

Domestic abuse has substantial mental health implications, with an estimated three-quarter of survivors experiencing posttraumatic stress disorder symptoms, and a significantly higher likelihood of reporting symptoms of anxiety and depression, compared to the general population (Ferrari et al., 2016). The long lasting impacts of domestic abuse are not limited to the direct target of abuse. Witnessing domestic abuse at home can have severe developmental impacts on children, including an increased risk of experiencing mental and physical health problems and encountering difficulties in interpersonal relationships in later life, worse educational attainment, and increased likelihood of engaging in criminal behaviours (Callaghan, Alexander, Sixsmith, & Fellin, 2015).

As the legal definition reflects, one characteristic of domestic abuse that differentiates it from other violent crimes is its repeated nature. Estimates show that on average, survivors live in the abusive relationship for 2.7 years, experiencing an estimated 50 cases of abuse before getting effective help (?, ?). The most reliable statistics on domestic abuse in the UK is the Crime Survey for England and Wales (CSEW; Office for National Statistics, 2018), a victimisation survey including a self-completion module on domestic abuse. According to the CSEW, of those respondents who experienced domestic abuse between April 2017 and March 2018, only 17% reported it to the police (Office for National Statistics, 2018). This extremely high level of underreporting is another characteristic that is specific to domestic abuse.

In the most extreme cases, domestic abuse can culminate in domestic homicide. In the period between April, 2017 and March, 2018, 70 people in England and Wales were killed by their current or former partner, 90% of these victims were women (Office for National Statistics, 2019), demonstrating that domestic abuse is a fundamentally gendered phenomenon. While the pervasive problem of underreporting poses a significant obstacle to deriving reliable estimates of the true extent of the problem, the economic cost of domestic abuse in England and Wales between April, 2016 and March, 2017 was estimated to be £66 billion (Rhys Oliver, Barnaby Alexander & Wlasny, 2019). The largest component of this cost is represented by the physical and emotional consequences of abuse, reflected in a reduced expected quality of life for survivors. In addition, lost economic output resulting from missed workdays and reduced productivity, as well as costs to the health care system also significantly contribute to the overall figure.

2 Data

The first dataset is provided by the West Midlands Police, and includes all recorded crimes and incidents from the period between January, 2010 and October, 2018. The West Midlands Police is the third largest police force in England, serving a population of 2.9 million *(ref)*. Crimes that have a domestic abuse marker indicate cases of domestic abuse that meet the criteria for notifiable offences in the UK, whereas domestic abuse incidents refer to cases that do not qualify as a crime. About 31% of all crimes and incidents have a domestic abuse marker. For each record in this dataset, we have information about the exact location and time of the incident or crime, the gender, age and ethnicity of the offender and victim, and the severity of the injury sustained by the victim, if any. The first and last occurrence of the offence, as well as the exact time of reporting is also recorded. Each person in this dataset has a unique person identifier, allowing us to follow people over time. *Sentence about access to this dataset.*

The second dataset is the Crime Survey for England and Wales (CSEW), which is an annual, cross-sectional representative survey collecting information on victims of crime across England and Wales, including 30,000–40,000 households every year *(ref)*. The survey has a self-completion module containing questions about the respondent's experiences of domestic abuse throughout their life (since the age of 16), and in the past 12 months in particular *(ref)*. Given the serious problem of underreporting uniquely characteristic to domestic abuse, the CSEW is the most reliable source of information on the prevalence of domestic abuse in England and Wales. since when?

The two datasets have different merits, and combining them allows us to gain a deeper understanding the characteristics and dynamics of domestic abuse, and deliver policy-relevant insights. The CSEW is more likely to give a better estimate of the true prevalence of domestic abuse, and contains rich demographic, socio-economic and geographic (LSOA level) information on the respondent (survivor). However, owing to the fact that it is a victimisation survey, it does not contain any information on the perpetrators and has limited longitudinal information. In contrast, the crime dataset has information (age, gender, ethnicity) on both the victim and the offender, and records all reported incidents for the same victim-offender pair, allowing us to conduct a quantitative analysis of the dynamics of abusive relationships and identify the predictors of escalation. Information on the exact timing of the abuse and the time of reporting will help us identify high-risk times of the year and understand the factors influencing the decision to report. Information on alcohol involvement is more likely to be accurate in the crime dataset than in the self-completion survey. Furthermore, the crime dataset provides us with information on other criminal behaviours of the offender and victim. In all the analysis plans outlined below, we will not link the individuals in these datasets.

Research Programme

Our research programme involves four research projects exploring different aspects of domestic abuse. Drawing on the extensive, rich respondent information in the CSEW, and the unique, temporal nature of the WMP, we aim to complement, extend and deepen our understanding of domestic abuse. In outlining each of these research projects, for the sake of clarity, we will specify a set of core regression analyses we plan to conduct. We will preregister all of our analysis plans using the Open Science Framework (OSF) ahead of beginning research. We will also aim to complement all analyses with robustness checks and further exploratory regressions as the data allows (see, *Trendl Football*, for an example of this approach).

1. Understanding domestic abuse: survivor characteristics and police mis-recording

Short intro about aims, highlighting added value. In this research project, we will first use the CSEW to investigate the demographic and socio-economic characteristics of survivors of domestic abuse. Exploring the risk factors in domestic abuse victimisation is key in designing effectively targeted policy measures. We will also use the WMP data to explore the extent of police mis-recording of domestic abuse. Increasing victim's trust in the police through improving law enforcement response to reported incidents is key in encouraging victims to report domestic abuse and prevent further harm.

Detailed plan First, using the CSEW data, we will extend our understanding of the demographic and socio-economic factors predicting domestic abuse victimisation. The CSEW provides a broad range of information on the characteristics of the respondent, including age, sex, marital status, number of children, ethnicity, education, employment, income, benefit history, physical and mental health, frequency of going out, house and car ownership, self-reported well-being, and frequency of drug and alcohol use. Previous research has identified a number of risk factors predicting domestic abuse victimisation (being female, young, unemployed, separated or divorced, living in single parent household and earning in the lowest income bracket), based on descriptive statistics derived from the CSEW data. Building on what we know so far, we will explore the predictive power of a broader range of victim characteristics in statistical model, using a logistic regression approach. Investigating the explanatory power of these characteristics within one statistical model will provide us with a deeper understanding of the importance of these factors in predicting domestic abuse victimisation, compared to a purely descriptive statistics approach. We are also interested in how the predictive power of victim characteristics vary with the type of abuse suffered (physical abuse, threats, sexual abuse).

However, given the complex nature of domestic abuse, it is likely that there are important interactions between various victim characteristics in predicting the risk of domestic abuse victimisation (for example, we would expect that not having stable employment is a more significant risk factor of domestic abuse victimisation for those who have young children). To uncover the structure of these interactions without having to a priori specify them, as is required in a logistic regression model, we will also use a random forest classification algorithm, a machine learning method that has the ability to detect non-linear relationships between variables, allowing us to identify particularly vulnerable subgroups of the population. Insights from both methods will provide us with the first extensive characterisation of the risk factors predicting domestic abuse victimisation in England, using a national-level, large representative sample like the CSEW.

Police forces in England, and particularly the West Midlands Police, have previously been criticised for mis-recording violent cases of violent crime, including domestic abuse. Mis-recording these already heavily underreported crimes decrease victim's trust in the Criminal Justice System and their willingness to report subsequent incidents. *Why don't we use the aggregate number of police-recorded domestic abuse cases for each police force to do this? But if we don't particularly rely on the WMP data is there a point in this? They use CSEW and police data in this report, but they don't seem to compare the two*

Core outcomes

The first extensive investigation of the demographic and socio-economic predictors of domestic abuse victimisation.

Table 1: Understanding domestic abuse: survivor characteristics and police mis-recording, regressions

Research question	Dataset	Unit of analysis	Outcome variables	Explanatory variables	Model
What are the characteristics of domestic abuse victims?	CSEW	Individual (respondent-level)	Domestic abuse victimisation; Victimisation by type (physical, sexual abuse, threats)	demographic and socio-economic characteristics	Logistic regression/random forest
What is the extent of police mis-recording of domestic abuse?	ONS data				

2. Predicting serious harm and understanding the decision to report

Short intro about aims, highlighting added value. In this project, we will explore the victim and offender characteristics that predict serious harm and the factors determining the decision to report the abuse. Identifying high-risk victim-offender pairs is key in preventing serious harm. Understanding what affects the decision to report is particularly important, as underreporting represents the biggest obstacle to effectively tackling domestic abuse. Insights from these analyses can inform decisions about the optimal timing and target audience for domestic abuse awareness campaigns.

Detailed plan Drawing on the relative strengths of the CSEW and the WMP datasets, we will explore the relative importance of the risk factors predicting serious harm. Previous studies investigating this question mostly relied on police data and focused on identifying high-risk offenders. We will complement and extend this research by simultaneously considering the victim characteristics that are predictive of suffering serious injuries using the CSEW. Specifically, using a multinomial logistic approach, we will explore the extent to which financial independence and resilience (income bracket), mental health (self-reported well-being), drug and alcohol dependence (frequency of usage), social isolation (self-reported frequency of going out, living alone or with the perpetrator, family disputes), recent separation, and feeling frightened (increased home security, reporting feeling scared) are predictors of various levels of harm (threats, minor or serious injury) *causality issue?*. We will know the level of harm suffered, because the CSEW asks whether the respondent needed to see a doctor, nurse or other health worker, or needed to take time off work because of the abuse. Given the cross-sectional nature of the CSEW data, this analysis will be exploratory rather than causal.

Using the WMP data, we will also explore the risk factors associated with the perpetrator, using each perpetrator-victim pair as a unit of analysis. The strength of the WMP data is that it has a temporal dimension, and we can follow the same perpetrator-victim pair over time. Our detailed crime dataset will allow to investigate the extent to which the perpetrator's previous recorded self-harm and suicide attempts, previous violent and drug offences, breaches of court orders and separation from the victim will predict causing serious harm. From the detailed description of the incident, we will know if the perpetrator and the victim live together. We will explore this question using a multinomial logistic regression analysis with perpetrator-victim fixed effects, where we will use previously known information about the perpetrator to predict serious harm. Using a fixed-effects regression approach will allow us to disentangle the relative importance of our explanatory variables and perpetrator-victim specific factors in predicting serious harm.

We will also use both datasets to explore the predictors of the decision to report and escape the abuse. We are not aware of previous research exploring this question. Using the CSEW and logistic regression approach, we will investigate how the decision to report (or the decision to seek any form of external help) depends on the individual-level, time-invariant characteristics of the survivor, controlling for the severity and type of the abuse suffered. We are especially interested in exploring how trust in the Criminal Justice System (CJS), social ties (living with family, going out, length of living in the area, member of Neigh-

Table 2: Predicting serious harm and understanding the decision to report, regressions

Research question	Dataset	Unit of analysis	Outcome variables	Explanatory variables	Model
What victim characteristics predict serious harm?	CSEW	Individual (respondent-level)	Level of harm resulting from domestic abuse victimisation (no physical harm, minor, serious)	financial resources, mental health, dependence, social isolation, separation, feeling scared	Multinomial logistic regression
What perpetrator characteristics predict serious harm?	WMP	Perpetrator-victim pairs	Violent offences resulting in injury	previous self-harm and suicide attempts, previous violent and drug offences, breaching court orders, separation	Multinomial logistic regression with perpetrator-victim fixed effects
What time-invariant (victim) characteristics predict the decision to report?	CSEW	Individual (respondent-level)	Reported abuse to the police	severity/type of abuse, trust in CJS, social ties, financial status, children, mental health, dependencies	Logistic regression
What time-varying characteristics predict the decision to leave?	WMP	Perpetrator-victim pairs	Reported ongoing abuse to the police	time of year, specific holidays, length of abuse, severity of last occurrence	Logistic regression

bourhood Watch), financial independence (own income, employment status, car ownership), number of children, mental health, and alcohol and drug dependency affect the decision to report. Using the WMP data, we will explore time-varying factors affecting the decision to leave the abuser. Due to the high levels of underreporting, it is hard to distinguish between time periods when the number of domestic abuse incidents increase, and when the victims have a higher propensity to finally report the ongoing abuse. Here we are mainly interested in the latter case (and will explore the first case in more detail in the last research project), therefore we will focus on the subset of incidents where the victim for the first time reported a long history of abuse. We are interested in whether these reports are more likely to occur after certain days of the year (e.g., birthday of the victim or perpetrator, Christmas, Halloween, Easter, etc.), potentially reflecting cases where the abuse has suddenly escalated, or became known to outsiders.

Core outcomes

A deeper understanding of the victim and perpetrator characteristics that predict serious harm, and the time-invariant and time-varying factors affecting the decision to report the abuse to the police.

3. The long-lasting effects of domestic abuse

Short intro about aims, highlighting added value. Domestic abuse has long-lasting adverse effects on victims and those close to them. In this research project, our is aim to explore some of these consequences on the direct victims of domestic abuse and the children who live in the same household. Gaining a deeper understanding of the tangible, far-reaching consequences of domestic abuse will help to design survivor support programmes and quantify the societal harm caused by it.

Detailed plan First, we will explore the causal effect of past experiences of domestic and child abuse on present socio-economic outcomes. First, using the CSEW, we will explore how experiencing abuse in childhood affects educational, employment and health outcomes in adulthood. The 2016 CSEW included a module asking respondents about whether they experienced abuse as a child or witnessed domestic abuse. Descriptive statistics of this dataset revealed that those who reported to have experienced abuse as a child were significantly more likely to report long-term health problems, and live in a single parent household. However, due to the possibility of various confounders (e.g., those growing up in economically deprived areas are more likely to experience abuse, but economic deprivation also affects educational outcomes), these statistics do not tell us the causal effect of childhood abuse on socio-economic outcomes in adulthood. To overcome this difficulty in establishing causal relationships, we will use propensity score matching, a statistical approach that allows for causal inferences to be drawn about the effect of childhood abuse on educational, employment, and mental and physical health outcomes, and the probability of experiencing domestic abuse in adulthood. In a similar vein, using propensity score matching, we will explore how past experiences of domestic abuse affect current health outcomes, by comparing people who have experienced domestic abuse in the past, but not in the last 12 months, with people who have never experienced domestic abuse, but otherwise have similar socio-economic characteristics.

Second, using both the CSEW and WMP, we will explore how children's behavioural problems are affected by witnessing or experiencing domestic abuse at home. Using data from the 10-15 year old questionnaire of the CSEW, which can be linked to the adult questionnaire, we can examine the effect of domestic and child abuse on children living in such households. The children's questionnaire records information on whether the child had been physically or verbally abused by someone (and if that person lives in the household), as well as their experiences with bullying, carrying knives, and information on gang membership, school truancy, learning difficulties and health, including drug use and drinking behaviour. This allows us to explore the causal effect of past and present domestic abuse in the household and child abuse on these outcomes through propensity score matching. *Estimating the effect of indirect exposure to domestic abuse through a previously abused parent?*

We will complement this analysis using the WMP, by estimating how living in a household with domestic and child abuse affects the likelihood of engaging in various types of criminal behaviours (property-related vs violent crime). We will do so by identifying what percentage of young offenders are victims of child abuse, and have a home address that is associated with domestic abuse incidents. Given the problem of underreporting, this will give us a lower bound for the estimate of the proportion of children with behavioural problems who live in abusive households. *Am I missing something here?*

Table 3: The long-lasting effects of domestic abuse, regressions

Research question	Dataset	Unit of analysis	Outcome variables	Explanatory variables	Model
How does past childhood abuse affect socio-economic outcomes in adulthood?	CSEW	Individual (respondent-level)	Educational attainment, employment status and health-related outcomes, including domestic abuse victimisation	Degree and type of childhood abuse suffered	Propensity score matching
How does experiences of past (but not recent) domestic abuse affect socio-economic outcomes in adulthood?	CSEW	Individual (respondent-level)	Current health and employment outcomes	Degree and type of domestic abuse suffered	Propensity score matching
How does experiences of domestic abuse at home and experiences of child abuse affect behavioural outcomes in childhood?	CSEW	Individual (respondent-level)	Bullying, involvement with gangs, school truancy, carrying knives, drugs and alcohol usage, learning difficulties, health	Degree and type of domestic abuse witnessed/child abuse suffered	Propensity score matching
What proportion of young offenders come from abusive households?	WMP	Young offenders			

Core outcomes The first UK-based quantitative exploration of the causal effect of domestic abuse and childhood abuse on socio-economic outcomes.

Environmental factors

Short intro about aims, highlighting added value. In this research project, we will be using the CSEW and WMP datasets, as well as external sources of data to investigate some of the environmental predictors of domestic abuse. We will first explore how neighbourhood-level characteristics affect domestic abuse victimisation and the willingness to report. Understanding how neighbourhood-level characteristics affect the prevalence of domestic abuse and reporting behaviour is crucial for the most effective distribution of resources to help victims and targeted awareness campaigns.

We will also explore how time-varying, exogenous factors affect domestic abuse. We will first explore the effect of local events that increase alcohol-consumption, since our previous study have found a significant relationship between alcohol-related domestic abuse and football. Second, we will explore the relationship between financial stress and the prevalence of domestic abuse. First, we will explore the association between gambling accessibility in the local area, and reported numbers of domestic abuse. Second, we will examine how the roll-out of the Universal Credit system affected levels of domestic abuse across the West Midlands. Apart from delivering important policy-relevant insights, these investigations will also aim to inform the local police about the optimal allocation of police resources.

Detailed plan Using the CSEW, we will first explore the association between the environmental characteristics of the neighbourhood, the prevalence of domestic abuse and willingness to report. Previous literature has suggested various pathways, including reduced collective efficacy of neighbourhood, and normalisation and acceptance of violence through social norms to explain the link between neighbourhood characteristics and domestic abuse, but the question has not been explored in a UK context yet. Using a logistic regression, we will explore how the interviewer's perception of the street (signs of rubbish, vandalism, and the general condition of houses, member of Neighbourhood Watch) as well as the respondent's connection and attitude towards the neighbourhood (length of time living in the local area, noisy neighbours, rubbish lying around, teenagers hanging around, vandalism, drunken and anti-social behaviour, drug trafficking, abandoned cars, speeding traffic, police presence in the area, worries about crime levels) predict domestic abuse victimisation, after controlling for the socio-economic characteristics of the respondent. In investigating this question further, we will explore how deprivation in a neighbourhood (LSOA level, measured by the Multiple Indices of Deprivation) affects the prevalence of domestic abuse (estimated from the CSEW), and the willingness to report, using a spatial regression framework. Using this statistical approach will alleviate the problem of spatial dependency in our dataset, and will allow us to derive a precise estimate of the effect. These analyses will extend our understanding of the neighbourhood-level predictors of domestic abuse victimisation, and complement our individual-level approach outlined in the first project to provide a comprehensive exploration of the factors affecting domestic abuse victimisation and willingness to report.

We are also interested in exploring how time-varying external factors, including alcohol consumption, and financial stress (measured by gambling expenditure and benefit dependency) affect the propensity to engage in abusive behaviours. In investigating this question, we will benefit from a dataset containing credit card and current account spending of x customers across the UK, to which we will have access as part of a data sharing agreement with a very large UK bank. This dataset will allow us to quantify spending on alcohol over time by LSOA. By using the WMP and credit card data, we will identify time-specific changes in alcohol consumption in parts of the West Midlands by identifying externals events that may affect it (e.g., local festivals, weather, bank holidays, sport tournaments), and investigate their effect on reported alcohol-related domestic abuse (from the WMP). Previously, we have found a 60% increase in alcohol-related domestic abuse when the England national football team won, highlighting the profound effect exogenous events can have on the propensity for violence.

The recent reforms of the UK benefit system, in the form of the introduction of the Universal Credit (UC) had been criticised widely, due to the temporal financial strain it imposes on the least financially resilient people in society. Drawing on previous findings about the link between financial stress and domestic abuse, we will explore what effect the roll-out of the UC across the West Midlands had on the reported number of domestic abuse cases. The fact that the exact date of the roll-out varied across the seven metropolitan

boroughs within the county allows for a more precise estimation of the effect of UC on the reported number of domestic abuse incidents in the West Midlands. To further explore the effect of financial shocks on the propensity for violence, we focus on gambling. Gambling is increasingly recognised as a serious health concern across the UK, and can have an adverse impact on family finances. We will use our credit card and WMP datasets to assess the link between changes in spending on gambling and domestic abuse.

Table 4: Environmental factors affecting domestic abuse, regressions

Research question	**Dataset**	**Unit of analysis**	**Outcome variables**	**Explanatory variables**	**Model**
How do environmental characteristics of the neighbourhood predict domestic abuse?	CSEW	Individual (respondent-level)	domestic abuse victimisation/willingness to report	Perception of neighbourhood, socio-economic characteristics	Logistic regression
How does neighbourhood deprivation predict domestic abuse?	IMD, CSEW	LSOA-level	domestic abuse victimisation/willingness to report (proportion of overall population)	Various measures of deprivation	Spatial Logistic regression
How do temporal changes in alcohol consumption and income affect domestic abuse?	WMP, Credit card data	LSOA, day/week-level	Number of reported domestic abuse cases	Spending on alcohol and gambling, changes in benefit receipts	Poisson/Negative binomial regression

Core outcomes The first extensive analysis of the neighbourhood-level predictors of domestic abuse within the UK. An exploration of the exogenous, time-varying factors, including alcohol consumption and financial stress (gambling, universal credit) affecting the reported number of domestic abuse cases in the West Midlands.

References

Callaghan, J. E. M., Alexander, J. H., Sixsmith, J., & Fellin, L. C. (2015, dec). Beyond "Witnessing". *Journal of Interpersonal Violence*, 088626051561894.

European Parliament. (2018). *Violence against women in the EU: State of play* (Tech. Rep.).

Ferrari, G., Agnew-Davies, R., Bailey, J., Howard, L., Howarth, E., Peters, T. J., ... Feder, G. S. (2016). Domestic violence and mental health: a cross-sectional survey of women seeking help from domestic violence support services. *Global Health Action*, *9*. doi: 10.3402/gha.v9.29890

Home Office. (2018). *GOV.UK: Domestic abuse.* Retrieved 2019-04-30, from `https://www.gov.uk/guidance/domestic-abuse-how-to-get-help{#}domestic-abuse-in-a-relationship-recognise-it`

Office for National Statistics. (2018). *Domestic abuse in England and Wales: year ending March 2018* (Tech. Rep.). Office for National Statistics.

Office for National Statistics. (2019). *Homicide in England and Wales: year ending March 2018.*

Rhys Oliver, Barnaby Alexander, S. R., & Wlasny, M. (2019). *The economic and social costs of domestic abuse* (Tech. Rep.). Retrieved from `https://assets.publishing.service.gov.uk/government/uploads/system/.../horr107.pdf{%}0A{%}0A`