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 (?, ?). 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 (?, ?). The current legal definition in the UK (?, ?) 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 (?, ?). 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 (?, ?).

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, victims 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; ?, ?), 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 (?, ?). 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 (?, ?), 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 (?, ?). 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.

Recognising the severity of this widespread societal problem led the UK government to develop a strategic plan to tackle domestic abuse with a funding of £100m for the period between 2016 and 2020. The action plan focuses on a range of areas to reduce the prevalence of domestic abuse, including increasing awareness and willingness to report through education and information campaigns, introducing legal measures to increase victim safety (e.g., Domestic Violence Protection Order, Domestic Violence Disclosure Scheme, intervention programmes), and improve the responses of support services and health care professionals (?, ?).

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 (victim). 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. We do not intend to link the two datasets.

Research Programme

In outlining the research program we specify X core regression analyses, which we will preregister ahead of beigging research on the OSF. These will be complemented with the robustness checks and further exploratory regressions (see, Trendl Football, for an example of this approach).

1. Understanding domestic abuse: victim characteristics, extent of domestic abuse and police mis-recording

Short intro about aims, highlighting added value. In our first research project, we will be using the CSEW and WMP datasets to investigate three main research questions. First, we will use the CSEW and WMP to identify the demographic and socio-economic characteristics of the victims of domestic abuse, and the WMP to characterise perpetrators of reported cases. In addition, we aim to explore the predictors of the decision to report and the level of harm suffered. Exploring the characteristics of groups vulnerable to domestic abuse, the true extent of the problem of underreporting and the factors affecting the likelihood of reporting is crucial for designing effective interventions and prevent potential cases of domestic homicide. Second, we will estimate the true extent of domestic abuse victimisation. Given the multifaceted nature of domestic abuse and the problem of underreporting, it is very hard to reliably estimate the extent of the problem and the societal costs of domestic abuse. We wish to do so by combining the strengths of the two datasets (that is, the CSEW provides a more reliable measure of the true rate of victimisation across different offence types, while the WMP provides unique information on whether various types of offences were domestic abuse related). Finally, we will estimate the extent of police mis-recording domestic abuse. Improving the police's response to reported incidents is key in encouraging victims to report domestic abuse. By contrasting estimated levels from reported incidents within the West Midlands from the two datasets, we can identify any reporting gaps between the two sources of data that might arise from police mis-recording.

What we know so far

Aim 1. CSEW : 7.9% women and 4.2% men reported to have experienced domestic abuse in the past year; women between 20-24 are most likely to have experienced domestic abuse; for men, age doesn't represent a significant risk factor; divorced or separated status are significant risk factors across both sexes; long-term illness or disability is a significant predictor across both sexes; single parent households are the most vulnerable; women in the lowest household income bracket are the most likely to have experienced domestic abuse, the same is true for men but the pattern is less pronounced; only 17% of people who have experienced domestic abuse in the past year have reported to the police, there is no significant difference in this proportion across genders.

Aim 2. The prevalence of DA is usually based on the CSEW, so I think the second analysis can add value.

Aim 3. HMICFRS

Detailed plan

Aim 1. First, using the CSEW data, we will conduct a statistical analysis to provide a deeper understanding of the demographic and socio-economic factors and their interactions in predicting domestic abuse victimisation. We will first use a logistic regression approach to explore the factors predicting individual-level domestic abuse victimisation, the decision to report, the type of abuse, and the level of harm suffered. The CSEW provides a wealth 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, house and car ownership, well-being, drug and alcohol use, and we will evaluate the importance of each of these variables and their interaction in predicting our outcome variables. *Could we also use a e.g. a random forest classification algorithm and look at variable importance?* Understanding what affects the decision to report is particularly important since underreporting represents the biggest obstacle to effectively tackling domestic abuse. These analyses will represent the first extensive quantitative investigation of the individual-level predictors of domestic abuse victimisation in England, using a national-level, large representative sample like the CSEW.

Second, we can validate some of these findings using the WMP data. This dataset contains information about the sex and age of victims. We can contrast the sex-age victimisation numbers from the WMP with

corresponding population estimates to validate the findings of the CSEW*(we were initially talking about using victims of burglary, but maybe this is better?)*. We will also use the WMP data on the offender's sex and age to provide a characterisation of domestic abuse perpetrators.

Aim 2. While a unique strength of the CSEW is that it provides reliable estimates of the proportion of the population who have been a victim of a certain crime, it cannot capture the true extent of domestic abuse victimisation for several reasons. First, because it does not collect detailed information on repeat domestic abuse victimisation (an inherent characteristic of domestic abuse cases), but only on the number of incidents involving physical violence perpetrated by household members. In addition, victims who suffer frequent, ongoing abuse are less likely to accurately remember the exact number of incidents. Second, because it only focuses on certain sub-categories of domestic abuse (physical violence, sexual assault, stalking), while ignores others (criminal damage, verbal abuse, financial abuse, burglary). In addition, it does not collect information on whether different offences against the same victim were perpetrated by the same offender. In contrast, the WMP data allows us to tell if a reported case was domestic abuse related, regardless of the actual offence committed. In addition, it allows us to identify various incidents where the same victim-offender pair were involved. We can estimate the co-morbidity of self-reported victimisation by crime type using the CSEW. We can contrast this with the estimated co-morbidity of reported victimisation from the police data. *is this silly? then we can maybe estimate what percentage of true victimisation is domestic-abuse related? the £66 billion cost estimate is based on the CSEW*

Aim 3. Police forces in England, including 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. By contrasting the extrapolated estimates from the CSEW (which is representative sample) with WMP data on the true number of reported cases, we can estimate the true extent of the mis-recording of domestic abuse cases by the WMP. Given the repeated cross-sectional nature of the CSEW, and the longitudinal nature of the WMP, we can quantify the proportion of mis-recorded incidents over time.

Core outcomes A deeper understanding of the socio-economic predictors of domestic abuse victimisation, the decision to report and the level of harm suffered. The first estimation of the true extent of domestic abuse-related victimisation across all offence types. An estimation of the extent of police (WMP) mis-recording of domestic abuse over time.

2.1 The time course of abuse and the predicting serious harm

Aim:

The CSEW respondents indicated the level of harm, and we can use the CSEW to estimate the cross-sectional association between the most serious harm and other comorbid factors like alcohol, other crimes,

Regression 3: type of abuse visit hospital mental health

Separation is one of the biggest predictors of abuse resulting in serious injury. *Subject to whether we can get data on the DVPOs (Domestic Violence Protection Order) and DVPNs (Domestic Violence Protection Notice) issued in the West Midlands in the relevant period, we could explore the offender characteristics that predict the likelihood of sustained perpetration after separation, and assess the overall efficacy of these new preventive measures. If not, we can still use info on location + breaches of non-molestation orders?*

But cross-sectional data present a particular problem for making casual claims: For example, incidents involving drugs or alcohol intoxication: some individuals may end up in an abusive relationship because of an existing dependency, or a dependency may develop as a counterproductive strategy of coping with the abusive relationship.

The scope of the WMP dataset across time lets us take these questions further, by looking at the order in which events occur and at patterns across many years. We can examine whether a recorded victim of domestic abuse was previously recorded as a victim in other crimes, and if they were subsequently a victim of other criminal behaviours. Identifying risk factors is crucial for police services who want to identify individuals at risk of being future victims, so that they can target interventions and give the best advice possible to reduce an individual's risk. In addition, we can measure the relationship between experiencing domestic abuse, and the likelihood of being recorded as a suspect in other crimes. It is possible that some criminal behaviour is a precursor to experiencing domestic abuse whilst others follow it.

Should you be identifying the couples most at risk of serious harm from time-invariant properties, or should you be following couples over time and using time-varying properties to spot when to intervene?

The WMP dataset will further allow us to follow victim-offender pairs over time, and explore how the severity and nature of reported abuse changes over time. In the WMP data we can follow X victim-perp pairs for X years, with each pair appearing X times on average in the data.

We will also be able to identify high-risk days of the year (e.g., birthdays, Christmas, Halloween) by the characteristics of the victim-offender pair and the pattern and duration of abuse. Insights from these analyses can inform decisions about the optimal timing and target audience for domestic abuse awareness campaigns.

Regression 4: offense classification[proxy for harm] . type of abuse .

Future victim behaviours. Future perpetrator behaviours

serious event at t stuff at t-1 + perpetrator previous crimes and a model with pair fixed effects

Murders are very rare, so we are adopting a "near miss" airline / nuclear industry strategy of predicting the more prevalent but less serious incidents.

Say pair fixed effects control for all non-time varying characteristics. With and without pair fixed effects let's use see how much state vs trait matters for prediction.

2.2 Collateral damage

Aim:

CSEW regression: child gang membership any domestic abuse or reported DA [2x2 table of abuse/no abuse by gang / no gang] Also estimate, given gang membership what fraction are DA—which links ot WMP analysis below

It also contains information about multiple members of the household. This allows us to examine whether young individuals living in a household with a victim of domestic abuse are more likely to be victims or perpetrators of other types of crime. By using other questions in the CSEW, we can also measure whether these children are more likely to perpetrate crimes, and to be involved in gang related activities.

explore the lasting effect of domestic abuse on the victims and their children (do we see kids having behavioural problems? No control group unfortunately)- CSEW question about gang membership - are

those who report having a child gang member more likely to be victims of domestic abuse? Following people over time (Dynamic topic modelling for victims)- what's the temporal order?

Match on name, location, and date of birth in WMP data to see effect on children

WMP regression:

take every juvinile in a gang and try to match all to domestic abuse cases—-estimates of what fraction of gang members have co-morbid domestic abuse

What fraction of reported child abuse has comorbid DA; what fraction of DA has comorbid child abuse

Using school census we can go futher and estimate frequencies in a 2x2 child abuse by domestic abuse table—providing we can match children to mothers.

2.3 Environmental factors

Aim:

CSEW analysis

how the house looks within the neighbourhood/litter on the streets, own perception of litter, foot patrol

Unit of analysis is large geographical area in a year Time span 2005-

DA or not

Given DA: report or not

Fixed effects for region and year

matched in socieoconomic

WMP analysis

Unit of analysis is LSOA (MSOA?) per month or per quarter Time span - 2010-2020

Number of DA cases

The CSEW allows us to identify the demographic and socio-economic characteristics of victims who are the least likely to report, and therefore should be in the focus of support services. We can validate and contrast these findings with insights from a spatial regression analysis of the crime data, using low-level geographical area information from the census.

Extend regression model X to include region fixed effects?

A sentence about what earlier investigations have found about particular risk factors, including socio-economic deprivation.

Australia study on domestic abuse and proximity to fixed-odds betting terminals, and the one on offlicences.

anything that affects propensity to drink/financial stress

neighbourhood characteristics (deprivation), fixed odds betting terminals, alcohol outlets, events (football, festivals)

Universal credit roll-out in the West Midlands

Universal credit roll-out in the Black Country

food bank dependence? data on housing e.g. evictions? homelessness.

Mention we can construct some of these variables by access to "a very large UK bank" which whom we have a data sharing agreement. We cna build them from the row-per-transactino data for spending on current account and credit cards.