## Unique and new big-data perspective

Our proposal aims to bring a data science approach to the study of domestic abuss, by using large-scale data from West Midlands Police (serving a population of 2.9 million people), Lloyds Banking Group (the UK’s largest retail bank with 22 million customers), and the Crime survey for England and Wales (CSEW). Our approach is new and ambitious, and has the potential to derive new insights bout domestic abuse in the UK and complement the existing body of knowledge on the topic, which has been predominantly generated via qualitative research methods, such as case studies and interviews.

We were dismayed by the reviews. Our approach contrasts with the qualitative approach and it is from this background that the reviewers appear to have come. Although the reviewers identified that there is a need for quantitative research, in a field dominated by qualitative work, they subsequently focused their reviews on qualitative models and theory. Reviewer 199347314 even suggests that a data driven approach is a negative aspect of the proposal. Something we found particularly incongruous given that this is a proposal for the Secondary Data Analysis Initiative. Furthermore, recent results from a related ESRC grant have shown that existing models such as DASH have little to no explanatory power when tested empirically, hence demonstrating the desperate need for analyses that are quantitative and data driven

It is not possible to provide a complete review of the literature in a six page proposal whilst also clearly outlining our analysis strategy, especially given the scale of the project. We are fully aware of the studies by Walby and Towers, and we agree they are important seminal works highlighting the limitations of the CSEW. However, our proposal builds upon the qualitative findings and analysis of the CSEW, by integrating far more data, from more sources, and using more advanced data analytic techniques.

## Research team experience

Several reviewers made the point that our research team does not have extensive experience in domestic abuse research. While we might not have a track record in domestic abuse research, we would like to clarify that our proposal was written in full knowledge of the relevant literature (which we chose not to detail so there was sufficient space to outline our empirical strategy). Our research team is fully qualified to carry out the proposed research programme, as is demonstrated by numerous, high-impact publications in other, policy-relevant research topics, from estimating the harm of gambling in the UK using bank transaction data (under review in *Nature*), [peer effects in police misconduct](https://rdcu.be/bEAue), [nudging customers to increase their credit card debt repayment](https://www.fca.org.uk/publication/occasional-papers/occasional-paper-45.pdf), [and using Big Data to test environmental effects on behaviour and personality](https://www.ncbi.nlm.nih.gov/pubmed/31046588#:~:targetText=Individual%2DLevel%20Analyses%20of%20the%20Impact%20of%20Parasite%20Stress%20on,Openness%20Only%20for%20Older%20Individuals.&targetText=The%20parasite%20stress%20hypothesis%20predicts,and%20extraversion%2C%20but%20higher%20conscientiousness.).  
In particular, in the gambling paper, we use Lloyds Banking Group transaction data from millions of individuals to demonstrate that gambling is associated with increased financial distress, reduced lifestyle and health outcomes, higher rates of future unemployment and disability, and higher mortality rates. This view is unique and the first ever to demonstrate such associations—again complementing the traditional qualitative case-study and small survey based approach.

We will also establish an advisory board for the project, comprising a senior police officer from West Midlands Police, a senior member of the WMP Police and Crime Commissioner’s Office, an academic with experience in domestic abuse, Prof Vanessa Munro, a lawyer based here at Warwick with extensive experience of domestic abuse, and a member of the charity sector (NAME CHARITY) and a survivor of domestic abuse (NOMINATED BY THE CHARITY).

## Feasibility

One reviewer raised concerns about the feasibility of the project regarding access to the data. We would like to clarify that we already hold the necessary data. In addition, we have support from Deputy Chief constable Louisa Rolfe at West Midlands Police, who is also the National Police Chiefs’ Council lead for domestic abuse. Named postdoc Trendl works here at Warwick and is also an employee of Lloyds Banking Group in their behavioural and data science team. We have director level support from Lloyds Banking Group for the project. Applying data at this scale to the study of domestic abuse has never been done before.

One reviewer raised concerns about the feasibility of the project regarding access to the data. We would like to clarify that we already hold the necessary data on our servers here at Warwick. In addition, we have support from Deputy Chief constable Louisa Rolfe at West Midlands Police, who is also the National Police Chiefs’ Council lead for domestic abuse. Named postdoc Trendl works as a postdoc here at Warwick and is also an employee of Lloyds Banking Group in their behavioural and data science team. We have director level support from Lloyds Banking Group for the project. Applying data at this scale to the study of domestic abuse has never been done before.

No one else has or is currently proposing to take a data science view of domestic abuse across multiple, diverse datasets. As 199347314 describes, “most research on domestic abuse in the UK context is qualitative in nature”. Quantitative work in the area has only incorporated a single data source (such as the CSEW, *or* data from a police force), and does not incorporate external data which can provide additional data on potential causes and environmental risk factors.

1. As the application of multiple, very large datasets to research in domestic abuse is novel, there is no one with the required skills in econometrics and data science who also has a substantial track record in domestic abuse research to conduct this research. Requiring such a background would make it impossible to complete the research, because no such individuals exist. We have acknowledged this from the start, hence our working closely with domestic abuse leads in the police, and why we propose including an advisory panel.

## Theory

## With Data Access Secured Already, the Project is Viable

1. We already have access to all relevant data, as we stated in the proposal. We hold the police data and CSEW data here on servers at Warwick. Postdoc Trendl is already working with Lloyds data and is an employee of Lloyds Banking Group as well as Warwick University. Approvals and data sharing agreements are already signed. We stated in the proposal that access to data is already in place and so were dismayed to see the issue raised by Reviewer 158241287.
2. Having Lloyds Banking Group as a partner in the research is a huge asset. 130837269 was concerned that we are only using data from one retail bank, but this criticism is a misunderstanding. Lloyds are the largest retail bank with 22 million customers. It is unnecessary to have data from a second retail bank and would be logistically difficult. By using weighted sampling of the 22 million, we can construct any representative sample and still achieve a sample size far in excess of any ever used to study domestic abuse.
3. We are aware of the issues when using police data and CSEW data. For example, Walby and Towers have found that the CSEW underestimates the true extent of domestic abuse as it caps the maximum number of incidents (whereas domestic abuse is often repeated). We are extremely familiar with the West Midlands Police data, and have worked with WMP on projects in stop and search, domestic abuse, risk assessment in custody, and alcohol and police resources. We are mitigating the issues with each data set by using multiple data sets .

Walby and Towers have found issues with the CSEW - in that it underestimates the true extent of domestic abuse given that it caps the maximum number of incidents (whereas domestic abuse is often repeated). We are aware of this and plan to combine CSEW data with data from other sources.

## The sensitive nature of the research topic

We take the semantic point about describing children who witness male to female intimate partner violence as victims of domestic abuse, noting that the act of witnessing such events is what makes them victims. The user reviewer has made clear their views of the language used to describe such circumstances. The reviewer is clearly reluctant for this component of the research to be completed. We believe this arises from a difference of perspectives: we aim to identify the risk factors for later criminal behaviour, not out of a desire to stereotype or assign blame, but to better identify vulnerable individuals. It is an empirical question, whether such individuals are more likely to commit crime later. If they are, we have a societal responsibility to offer support to help them avoid this, not to simply blame.

The reviewer is also reluctant about the investigation of certain socio-economic factors such as the effect of benefit payment systems, expressing concern about potential working class stereotypes. Their assertion is absolutely correct that domestic abuse exists in all sections of society. But note, the word class appears nowhere in the proposal—our proposed research is not about class. Prior research shows a significant role of financial poverty, with the CSEW showing domestic abuse rates 3.5 times higher in very low income homes compared to middle and high income. This is an uncomfortable fact, but ignoring it does not make it false.

## Specific points

# ################## Unused Text Below Here

%We have studied the currently funded ESRC project INSERT PROJECT NAME AND PI. Our proposal is quite different and compliments this work without duplication.

## Anna comments

### no previous experience in domestic abuse research/no collaboration with experts

not sure what’s the appropriate response to this

### too data-driven/no clear theoretical framework or mention or previous studies/issues about originality as some of the questions have either been examined in the context of other countries or with a qualitative approach

we are combining data from various sources to address key questions that either have never been examined or have only been examined using one approach - fresh perspectice argument, and emphasize the importance of data driven research. The mentioned {r} [ESRC grant] (<https://gtr.ukri.org/projects?ref=ES/M01178X/1>) looks at improving the DASH using “more sophisticated algorithms than a traditional logistic regression framework.” – I can’t find more details on this, but they seem to be only using police data, whereas we are taking a broader view by relying on the very detailed information in the CSEW as well. We are also aware of the US literature on domestic abuse victimisation risk factors (which informed our selection of explanatory variables in the proposed analyses), however, we believe that we still lack the proper evidence base for understanding domestic abuse victimisation in the Uk. Despite of the documented issues with the CSEW, it is an incredibly rich and useful source of information on domestic abuse victimisation, which has not been exploited. Combining insights from the CSEW and police data will complement and validate existing knowledge from qualitative invetsigations.

### issues with the CSEW

#### data quality/data access issues/timescales

#### “worrying” assumptions

### specific points:

young person living at the address – we’ll know this from police records police mis-reporting – comparison of csew and police reports reliability of data, ethnicity – children, only a year of data // different explanatory variables for question 1– random forest not explaiend– causal effects – no domestic abuse journals– no mention of cost–