## Unique and new, multi-source big-data perspective

Our proposal brings a data science approach to the study of domestic abuse, by using multiple large-scale datasets including crime data from the West Midlands Police (serving a population of 2.9 million people), financial transactions data from Lloyds Banking Group (the UK’s largest retail bank with 22 million customers), and a national-level victimisation survey (Crime survey for England and Wales; CSEW). Our approach is new and ambitious, and has significant potential to derive novel, policy-relevant insights about domestic abuse in the UK and complement the existing body of knowledge on the topic, which has been predominantly generated via qualitative research methods. There is no other research team in the world with access to comparable data, skills and resources. Without funding, this unique opportunity to link datasets will be lost. With domestic abuse at this scale, any gains will help huge numbers of victims.

We were dismayed by the reviews. Our approach contrasts with the qualitative approach, and it is from this background that all but one of the reviewers appears to have come. We can certainly sympathize that the ESRC has struggled to find reviewers expert in both quantitative research and DA. This in itself emphasizes the novelty and importance of this project. Reviewer 134177025 was much more engaged with the proposed quantitative methods but “found it hard to believe” that there was not a substantial existing quantitative literature. This contrasts with the comments of the qualitative reviewers who emphasized the lack of such work and acknowledged that there is a need for quantitative research in a field dominated by qualitative work. However, much of these reviews focused upon qualitative models and theory. Reviewer 199347314 even suggests that a data driven approach is a negative aspect of the proposal, which we found particularly incongruous, given that this is a proposal for the *Secondary Data Analysis Initiative*. Furthermore, recent results from a related ESRC grant (ES/M01178X/1) have shown that existing models such as DASH have little to no explanatory power when tested empirically, thus demonstrating the desperate need for analyses that are quantitative and data driven.

Reviewer 134177025 also raised issues around the use of random forest analyses. We share their concerns around the potential for data mining with these techniques. Our reason for selecting them is to a) build upon the more robust regression analyses that they supplement, and b) to provide a descriptive summary that can identify potential complex interactions between predictors, as such patterns are particularly difficult to identify using qualitative methods. As the first to apply such a comprehensive data driven approach it is our intention that any descriptions can be robustly tested by subsequent projects.

## Research team experience

All reviewers made varying claims about the relevance of our expertise and research experience, and one states that our experience comes only from Psychology. This is incorrect. We are data scientists working at the intersection of behavioural science and economics and two of us have been members of the Network for Integrated Behavioural Science for 6 years. We have recently published in top economics journals including *American Economic Review* and *Management Science*, and Trendl holds two degrees in Economics. We would like to clarify that our proposal was written in full knowledge of the relevant literature cited by reviewers, and the lack of references in the proposal is merely a result of space limitations. Indeed, it is not possible to provide a complete review of the literature in a six page proposal whilst also clearly outlining our analysis strategy, and given the complexity of the data and proposed analyses, we focused upon the proposed methods. We are fully aware of the studies by Walby and Towers, and we agree they are important seminal works highlighting the benefits and limitations of the CSEW. However, our proposal builds upon the qualitative findings and prior analysis of the CSEW, by integrating far more data, from more sources, and using more advanced data analytic techniques. We are also aware of the ESRC grant ES/M01178X/1 aiming to improve DASH, but our proposal substantially differs from it in terms of research questions, analysis techniques, and data sources.

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. 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, 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 research, as well as a member of the charity sector and a survivor of domestic abuse (nominated by the charity).

We bring this approach to DA having already applied it successfully to other policy relevant topics. These include [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 (under review at Nature), we use Lloyds Banking Group data to show gambling is associated with increased financial distress, reduced lifestyle and health outcomes, higher rates of future unemployment and disability, and higher mortality rates, as well as to identify risk and protective factors.

## Feasibility

Reviewer 158241287 raises 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. We have director level support from Lloyds Banking Group for the project, where named postdoc Trendl is an employee in their behavioural and data science team, as well as being at Warwick. Applying data at this scale to the study of domestic abuse has never been done before but we, uniquely, have the agreements and infrastructure in place to begin now.

Having Lloyds Banking Group as a partner in the research is a huge asset. Reviewer 130837269 was concerned that we are only using data from one retail bank (the largest in the UK). It is unnecessary to have data from a second retail bank and logistically would be virtually impossible for any researcher given its secure and proprietary nature. By using weighted sampling of the 22 million customers, we can construct any representative sample and still achieve a sample size far in excess of any ever used to study domestic abuse. We fear the criticisms levelled at the specific data we hold are a case of the perfect, yet impossible, becoming the enemy of the excellent, yet eminently feasible.

## The sensitive nature of the research topic

We acknowledge that domestic abuse is an extremely sensitive topic, and any research should be done with the lived experiences of victims in mind. We take the semantic point raised by Reviewer 185645305 about describing children who witness domestic abuse as victims of domestic abuse, noting that the act of witnessing such events is what makes them victims. The reviewer expressed concern about our proposed analysis investigating criminal behaviour in later life. 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 (nor we would allow our findings to be used for such purposes), 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. We also note that our data-driven approach will identify potential protective factors (as suggested by the reviewer).

Reviewer 130837269 is 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, and again allows us to identify protective changes that could be made, such as changes to benefit payment schedules.