## Unique and new big-data perspective

Our proposal aims to bring 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, such as case studies and interviews. Our research team is uniquely placed to deliver this research project, both in terms of the research skills required to carry out the proposed analyses, and access to these datasets that have never been linked before to investigate domestic abuse.

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 seem to have acknowledged that there is a need for quantitative research in a field dominated by qualitative work and that there is scope for the CSEW to be exploited to understand domestic abuse, they subsequently focused their reviews on qualitative models and theory (or lack thereof). 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 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.

## Research team experience

All reviewers made the point that our research team does not have extensive experience in domestic abuse research. While we do 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, and the lack of references in the proposal is merely a result of space limitations. Indeed, it is not possible to provide a complete (or even a short) review of the literature in a six page proposal, whilst also clearly outlining our analysis strategy, especially given the scale of the proposed 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. 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 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, 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, and a member of the charity sector (!!!!!!NAME CHARITY!!!!!!!!) and a survivor of domestic abuse (nominated by the charity).

There is a vast scope for a data-driven approach to generate novel insights to complement existing qualitative research, 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.

## 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. 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.

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 would be logistically difficult. 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.

## 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 would also like to note that our data-driven approach will also be able to 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.

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”. Previous quantitative work in the area has generally been structured around a single data source (such as the CSEW, *or* data from a police force), and does not incorporate external sources which can provide additional data on potential causes and environmental risk factors.