

## RESEARCH BRIEF

# Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases

Based on BFI Working Paper 2021-01, "[Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases](#)," by Jeffrey Grogger, Irving Harris Professor in Urban Policy at UChicago's Harris School of Public Policy; Sean Gupta, London School of Economics; Ria Ivandic, University of Oxford; and Tom Kirchmaier, London School of Economics.

### KEY TAKEAWAYS

- ✓ Domestic abuse affects nearly one-third of all women world-wide who are in a relationship.
- ✓ Authorities have employed risk assessment tools such as questionnaires to try to determine which domestic abusers might repeat.
- ✓ These assessment tools are inconsistent, owing in part to their dependence on subjective questioning and scoring.
- ✓ Machine-learning tools, coupled with criminal histories of abusers, are more effective methods to anticipate recidivism.

The numbers are staggering. Domestic abuse affects nearly one-third of all women world-wide who are in a relationship, including one-fourth of US women and one-third of English women who suffer physical or sexual abuse at the hands of an intimate partner.

For years, authorities have tried to reduce the incidence of domestic abuse (at least 80 percent of which is male-on-female) by trying to anticipate recidivism, or who may offend again. It is difficult, if not impossible, to anticipate the first case of domestic abuse, the reasoning goes, but with proper techniques we may predict who is most likely to reoffend.

One such technique is a risk assessment tool that involves some form of questionnaire for the victim or the perpetrator, followed either by a scoring rule based on the answers, or a subjective ranking by the person who administered the questionnaire. But how effective are such

methods? Can we rely on the subjective judgment of police and probation officers, victims' advocates, and others who conduct these interviews? Do the questionnaire's scores offer any predictive value? Are their better ways to determine who is most likely to reoffend?

These are the questions that motivate "Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases," by Jeffrey Grogger, Sean Gupta, Ria Ivandic, and Tom Kirchmaier. The authors show that using machine-learning algorithms improves the results of questionnaires but, more importantly, algorithms perform even better when they incorporate criminal histories. Their results have important implications for authorities hoping to efficiently direct limited resources toward preventable crimes.

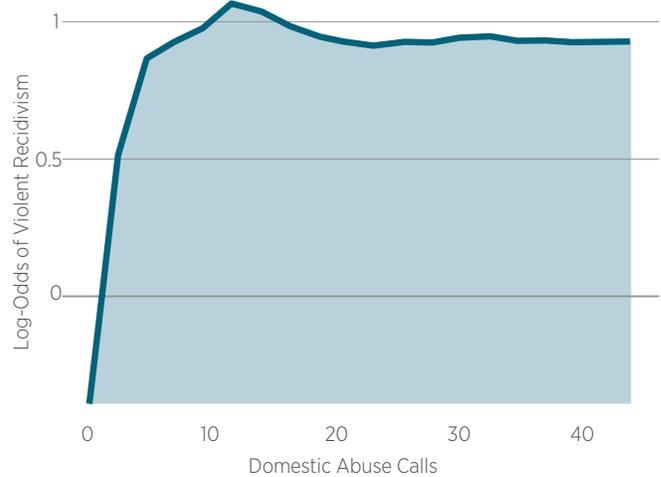
### Prediction as prevention

The assessment tools described above are used in several European countries and in many parts of Canada and the United States. The authors focus on one particular

Police officers must ask 27 questions and weigh the value of each answer to formulate a grade. Among other factors, police officers will have varying degrees of experience with domestic abuse cases and, likewise, their judgments may vary even on similar cases. Further, officers receive various amounts of training, and most officers likely never receive feedback on their previous judgments; there is no learning by doing.

**Figure 1** • Partial Dependence Plots for Selected Features from Random Forest Based on All Features

*Partial Dependence Plot for Domestic Abuse Calls, Dyad*



Note: The figure shows that the relationship between the number of prior domestic abuse calls and the likelihood of violent recidivism is highly non-linear.

assessment tool employed in England and Wales known as DASH (Domestic Abuse, Stalking, and Harassment and Honor-Based Violence). DASH consists of a questionnaire administered by a police officer, who then renders a score of likely recidivism. High-risk victims are offered protective services. But who is high-risk? Some police forces reported that fewer than 10 percent of domestic abuse cases were high risk, while others put the number at 80 percent, according to 2014 data.

Inconsistency seems the norm for DASH, and clearly victims of domestic abuse are not well served by such variation. The authors challenged themselves to develop a method for predicting recidivism that outperforms DASH, and their work offers three key findings:

- DASH predictions are only trivially more accurate than those based on a simple classification technique that makes no use of DASH assessments. Further, a machine-learning approach can provide better forecasts of violent recidivism than a

structured-judgement approach based on data from the same assessment protocol.

- Criminal history matters. Machine learning methods applied to two-year criminal history data provide better forecasts of violent recidivism than the same methods applied to the data from the assessment protocol.
- Finally, adding data from the assessment protocol does little to improve the performance of machine-learning forecasts based on criminal history data.

Why does DASH perform so poorly? The authors review relevant literature and surmise that DASH is hobbled, in part, by too many questions. Police officers must ask 27 questions and weigh the value of each answer to formulate a grade. Among other factors, police officers will have varying degrees of experience with domestic abuse cases and, likewise, their judgments may vary even on similar cases. Further, officers receive various amounts of training, and most officers likely never receive feedback on their previous judgments; there is no learning by doing.

Complicating matters even more is the environment in which the DASH questionnaires are administered, which involves an interaction between the victim and a police officer, typically at a moment of considerable stress. It is not hard to imagine that the willingness of the victim to provide information may depend on a number of factors, including the officer's sex and ethnicity, as well as other circumstances, such as the mind-set of the victim at the time.

Machine-learning methods, on the other hand, are true to their name and benefit from learning from tens of thousands of incidents, for which they see the outcome in each

case. On the basis of that information, these algorithms learn weights that are optimized to forecast failure, taking into account the relative cost of false negatives to false positives.

All told, the authors speculate that these and other factors explain why the machine learning predictions based on criminal history perform better than models based on DASH features. However, the authors also stress that machine learning forecasts have limitations of their own. One is that they are only able to predict violent recidivism that is reported to police. Although more serious incidents of domestic abuse are more likely to be reported, the authors are unable to observe, and therefore to predict, violent domestic incidents that are not reported. Another limitation is that machine-learning methods cannot predict other types of serious harm that may result from domestic abuse.

## Conclusion

Machine learning, combined with criminal history, is a better predictor of domestic abuse recidivism than subjective questionnaires and scoring techniques. The authors' model, which is a good predictor of who is unlikely to violently recidivate, has important real-world implications. Call handlers, for example, could have a computer dashboard with access to a criminal history database, a coded-up version of the forecasting model, and an interface to pass data from the database to the model. When a call comes in, the call handler would take enough information to identify the victim and perpetrator in the database. With their identities established, criminal histories would be digitally retrieved and passed to the model, which would output a prediction. If the prediction were for no violent recidivism, the case would be given a low priority score. If the prediction were for violent recidivism, the

call handler would gather other information to determine whether to assign the call a high or intermediate priority.

Of course, this method is more complicated than described here, and the authors provide great detail in their paper about the underpinnings and processes surrounding their approach. But the bottom line is clear: some domestic abuse victims are in danger of experiencing further violence, and we have the tools to help them.

#### CLOSING TAKEAWAY

Machine-learning methods, on the other hand, are true to their name and benefit from learning from tens of thousands of incidents, for which they see the outcome in each case. On the basis of that information, these algorithms learn weights that are optimized to forecast failure, taking into account the relative cost of false negatives to false positives. In other words, conventional accuracy may not be the correct objective for the prediction exercise.

#### READ THE WORKING PAPER

NO. 2021-1 · JANUARY 2021

### Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases

[bfi.uchicago.edu/working-paper/20211](https://bfi.uchicago.edu/working-paper/20211)

#### ABOUT OUR SCHOLAR



#### Jeffrey Grogger

*Irving Harris Professor in Urban Policy,  
Harris School of Public Policy*

[harris.uchicago.edu/directory/jeffrey-grogger](https://harris.uchicago.edu/directory/jeffrey-grogger)



*Written by David Fettig, BFI Senior Writer and Editor*

