

DAVID BUIL-GIL, JOSE PINA-SÁNCHEZ, IAN BRUNTON-SMITH &amp; ALEXANDRU CERNAT

# BAD DATA, WORSE PREDICTIONS

## How measurement error in crime data affects crime prevention.

Increasingly sophisticated methods are applied to predict crime patterns from police records with little regard given to the quality of the data. We explore how this may affect crime prevention.

### MEASUREMENT ERROR IN CRIME DATA

Police-recorded crime statistics are widely used for different purposes. Police forces use crime data to analyse geographic variations in crime and assist operational decisions. Policy makers draw on police records to justify policy changes. Neighbourhood watch groups use crime data to lobby for security. Crime researchers examine recorded crime trends to build theories of crime. But police-recorded crime data are severely affected by measurement error.

Crime records are incomplete. Not all victims report crimes to the police. And the police do not record all crimes reported. Many incidents, such as drug offences, do not even have direct victims who can inform the police. All this results in what is known as the 'dark figure of crime' - i.e., crimes not recorded in statistics. The percentage of crimes unknown to the police can be as large as 67% for damage, 63% for personal property offences and 61% for threats. Crime reporting also varies across population groups and recording practices differ between police forces, with a 2014 inspection concluding that between 63 and 71% of violent incidents and between 71 and 77% of sexual crimes reported to the police were not correctly recorded in crime registers.

Due to the ever-growing evidence that crime records are inaccurate, such records had the official designation of 'UK National Statistics' removed in 2014. Yet, crime aggregates are still used for crime prevention. Here we explore how crime prevention may be flawed by an uncritical acceptance of police-recorded data.

### IMPACTS ON CRIME PREVENTION

We describe a set of ways in which police records are used in crime prevention efforts, and explore how these may be affected by underlying measurement error present in crime data.

#### Geographic crime analysis

Crime statistics are aggregated in geographic areas and visualised in maps to study the spatial concentration of crime and

identify areas where crime is most prevalent. Geographic crime analysis assists the design of targeted crime control initiatives. For geographic crime analysis to accurately highlight high-crime areas, the proportion of crimes unknown to police forces should be spatially uniform. This is not the case. The 'dark figure of crime' varies across cities and neighbourhoods. Consequently, the police may underestimate crime in places with low reporting rates and overestimate it in places with higher reporting rates.

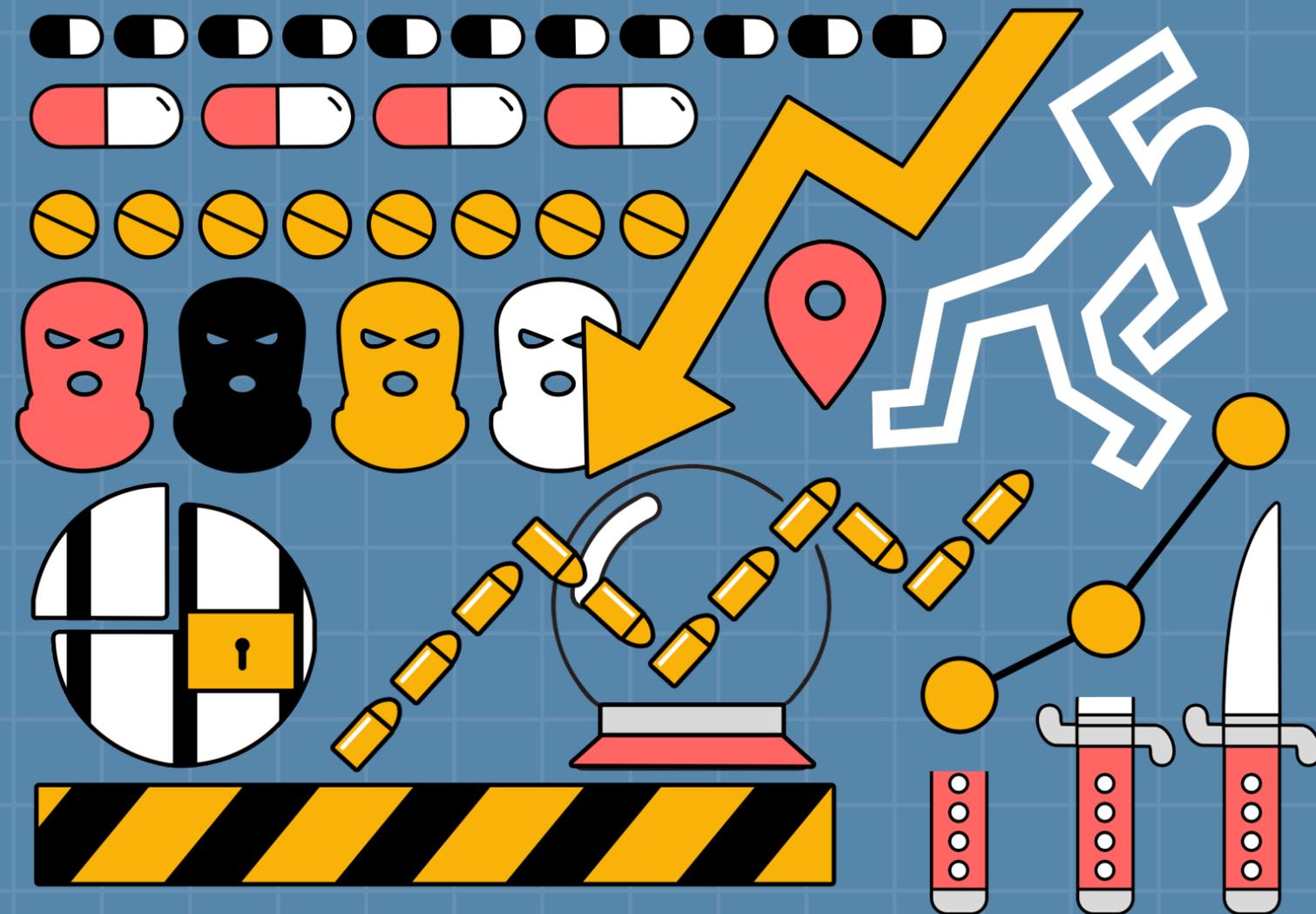
“...crime forecasts suffer from high risk of inaccuracy and may result in disproportionate police control on historically over-policed communities.”

#### Crime trends analysis

Crime data are used to analyse changes in crime over time to identify whether crime is increasing and why. For crime trend analysis to accurately reflect changes in crime, the proportion of crimes missing from police records should remain stable across time. This is not always the case. Crime records are affected by changes in the way data are recorded, and crime reporting rates vary across years. In turn, trends recorded in police records and crime surveys vary significantly.

#### Predictive policing

Law enforcement makes use of predictive analytics to identify potential criminal activity before it takes place, to determine targets of police interventions. Historical crime data is used to train machine learning algorithms to forecast incidents. For predictive policing to accurately forecast crime, data used to train algorithms should not be affected by measurement error (or algorithms should account for it). This is rarely the case. Crime forecasts suffer from a high risk of inaccuracy and may result in disproportionate police control on historically over-policed communities.



#### Exploring the causes of crime

Crime records are used by researchers to explore the causes of crime. Statistical modelling is used to estimate the effect of a range of social, legal and environmental constructs on crime to assess if the presence of certain social conditions, policies or urban features is causing crime to increase. For statistical models to accurately estimate the effect of a variable on crime, it is necessary that crime data are not affected by measurement error. Crime researchers rarely account for this. Error induced by underreporting, underrecording and random errors (i.e., the 'dark figure of crime') may bias model estimates of the impact of security measures, economic conditions, disorder and other variables on crime.

#### A WAY FORWARD

The biasing effect of measurement error on crime research and crime prevention is widely recognised - but there are ways forward. Victim surveys record periodical data from randomly selected samples of respondents and provide relevant information about crimes known and unknown to police. Matching survey data with police records allows us to identify the prevalence of measurement error in crime records, and researchers are using this information to identify its potential effects on statistical outputs and to generate adjusted

crime estimates. For instance, it enables identification of the geographic and temporal variation of the 'dark figure of crime', and accounts for the proportion of crimes missing from police records in statistical analyses, crime forecasts and crime prevention. We are also using it to test potential methodological solutions to mitigate the biasing effect of measurement error on model results. Simulation studies show that, in many cases, this problem can be minimised, or altogether eliminated by log-transforming crime rates. Of course, victim surveys are not error-free. This is why we compare and combine multiple crime data sources to enable crime estimates of improved precision for crime prevention.

*Dr David Buil-Gil is a lecturer in Quantitative Criminology at the Department of Criminology of the University of Manchester, and a member of the Manchester Centre for Digital Trust and Society.*

*Dr Jose Pina-Sánchez is an associate professor in Quantitative Criminology at the School of Law of the University of Leeds.*

*Prof Ian Brunton-Smith is a professor of Criminology and Research Methods at the Department of Sociology of the University of Surrey.*

*Dr Alexandru Cernat is an associate professor in Social Statistics at the School of Social Sciences of the University of Manchester.*