The British Journal of Criminology, 2024, **64**, 1347–1361
https://doi.org/10.1093/bjc/azae018
Advance access publication 5 June 2024
Article

OXFORD

Estimating the Reliability of Crime Data in Geographic Areas

Ian Brunton-Smith, Alex Cernat, Jose Pina-Sánchez and David Buil-Gil[*]

[*]I. Brunton-Smith, Department of Sociology, University of Surrey, Guildford, UK; email: i.r.brunton-smith@surrey.ac.uk; A. Cernat, Department of Social Statistics, University of Manchester, Manchester, UK; J. Pina-Sánchez, Centre for Criminal Justice Studies, School of Law, University of Leeds, Leeds, UK; D. Buil-Gil, Department of Criminology, University of Manchester, Manchester, UK

Crime data are problematic: Crimes that are never reported undermine its validity and differences in police recording practices affect its reliability. However, the true extent of these problems is not well known, with existing studies suffering from a number of methodological limitations. We examine the quality of police recorded crime data and survey-based crime estimates recorded in England and Wales using a robust latent trait model that effectively represents the competing sources of error. We find that whilst crime rates derived from police data systematically underestimate the true extent of crime, they are substantially more reliable than estimates from survey data. Reliability is lower for violence and criminal damage and is getting worse over time.

KEY WORDS: criminal, statistics, reliability, validity, regions, UK

INTRODUCTION

Crime measurements commonly stem from one of two data sources: police records or crime victimization surveys. Police records have the benefit of being highly granular (spatially, temporally and in terms of offence types) but are generally accepted to be flawed since they fail to capture crimes that are not reported, and because of the inconsistencies stemming from different recording practices followed across police officers and forces. Victimization surveys provide less widespread coverage of the range of crimes and their precise locations and are subject to a range of methodological shortcomings that result from the use of a survey approach. But their focus on victims' experiences and the use of representative samples allow them to overcome key limitations of police recorded crimes, including recording information about crimes unknown to the police (Aebi and Linde 2014; Lohr 2019). As a result, survey data are often taken as the most valid approach to estimate national levels of crime at one point in time or across medium and long terms, and are also often used to assess the extent of measurement error in police data (Gibson and Kim 2008; Tarling and Morris 2010). However, because victimization surveys are generally carried out annually and their sample design is optimized to produce country or regional-level estimates, most research into the causes and consequences of crime is still explored using police data. This includes those studies focussing on comparatively short-time intervals (e.g. the immediate effect of a lockdown), more granular spatial resolutions (e.g. comparison of crime prevention initiatives at the city or neighbourhood level) and offence types (e.g. knife crime or homicides) (Ariel and Bland 2019; Bland and Ariel 2020).

The reliance on police data means that a good deal of crime research is suspected to be heavily affected by measurement error, and therefore bias. However, recent studies have shown that the use of police recorded crime data in criminological research does not always need to lead to substantial bias. If the *form* and *prevalence* of the measurement error mechanisms affecting police data can be estimated, we can accurately anticipate the direction and extent of the subsequent bias in multivariate analyses relying on these data (Pina-Sánchez *et al.* 2022; 2023). Yet, as much as it is widely acknowledged that police data are flawed, we still do not have accurate estimates of its measurement properties. This represents an important and surprising gap in the crime data literature that we address by directly contrasting police recorded data against crime survey records whilst correctly recognizing the errors inherent in both.

The most common approach used to explore the presence of measurement error in police data involves making direct comparisons with estimates from victimization surveys (e.g. Gibson and Kim 2008; Tarling and Morris 2010; Pina-Sánchez *et al.* 2022). Typically, this involves treating victimization surveys as a 'gold standard', which implies that any discrepancies between crime estimates derived from the survey or police sources are evidence of measurement error in the latter. Such an approach can help us assess the extent to which certain crimes are underreported (i.e. systematic errors), but it is bound to overestimate the amount of noise (i.e. random errors) in police data, and in so doing, underestimate its reliability. This is because although one could take crime estimates derived from surveys to be unbiased, they are still estimates, and, as such, uncertain. Importantly, the amount of uncertainty in survey-based crime estimates is likely to vary across areas and over time (Rosenbaum and Lavrakas 1995). Alternative approaches have relied on manual reviews of documents and recordings of incidents (Klinger and Bridges 1997; Her Majesty Inspectorate of Constabulary 2014). Reliability estimates based on this methodology are as internally valid as they can possibly be. However, this is an expensive approach, and therefore estimates are limited to relatively small samples of police forces, crime types and time points, which affects their external validity.

In this study, we take a different approach. We explicitly account for the uncertainties in both police recorded crime data and survey data using a MultiTrait–MultiMethod (MTMM) model. As a latent variable estimation method, MTMM models allow us to take crime rates derived from each data source as imperfect measures of the true extent of crime, whilst, crucially, estimating how well each of the measures captures the true (but unobserved) crime rate. That is, MTMM directly provides us with estimates of the reliability ratio of both police recorded and survey-based estimated crime rates. This allows us to assess their quality profile more accurately, eschewing the need to give one of the data sources primacy as a 'gold standard'. Specifically, we use police data from the Home Office and survey data from the Crime Survey for England and Wales (CSEW) to estimate annual crime rates across Police Forces ($N = 43$) and Community Safety Partnership (CSP) areas ($N = 312$) from 2011 to 2019.

The article proceeds as follows. We first discuss the collection of official crime records as well as the use of surveys as a way to capture the same information. We then present the data used and formally introduce the MTMM model. Finally, we present the results and discuss their implications for crime research and crime prevention practice.

THE MEASUREMENT OF CRIME

Measuring the extent of crime has a long history, with the first maps of recorded crime across France being produced as early as 1829 as part of the collection of so-called 'moral statistics' (de Candolle 1830[1987]; Sellin 1931).[1] Police records of crime have since become one of the longest running sources of government statistics, documenting a dramatic rise in criminal activity from the middle of the 20th century and an equally dramatic drop since the mid-1990s across most Western countries (Farrell *et al.* 2014; Berg and Lauritsen 2016). Yet, in spite of its long-running nature and the efforts invested to bring about consistent counting rules, police recorded data continue to be criticized for its low reliability.

Critics have argued that police records are affected by several external factors which could themselves vary in space and time. For instance, victims may be unaware of crime or choose not to report it, and the police may fail to identify or arrest the offender or choose not to record incidents for a variety of reasons. The very definition of criminal activities is also historically contingent and subject to change, whilst the policing priorities of the day and the size of the police force can present the illusion of a spike in crime whilst the underlying rate of crime remains constant. And despite the existence of a comprehensive set of counting rules and protocols for the police to follow (Home Office 2011; 2013), once brought to their attention the decision to record an incident as a crime is down to the personal discretion of the officer(s) involved (Burrows *et al.* 2000; Boivin and Cordeau 2011) as well as the prevailing culture within the specific force (Warner 1997). As such, police recoded crime data must be viewed as a record of the extent of crime which is necessarily affected by policing activities and their interactions with the public.

In part reflecting these problems with police recorded crime data, the 1960s saw the advent of a new approach to measuring crime using survey data. Pioneered in the United States as part of the President's Commission on Law Enforcement and the Administration of Justice (Ennis 1967), sample surveys measure crime by asking victims directly to report on their experiences. By adopting a probability-based sampling strategy coupled with a consistent measurement tool, they are unaffected by the sorts of variations in reporting and recording practice that have been so prevalent in recorded crime data. However, survey-based crime data are not error-free, with measurement errors arising from poor question wording, limitations with the questionnaire design, as well as inadequate training of survey personnel, sampling bias and non-response bias. Crime estimates also suffer from victims' memory failures, social-desirability bias and underestimation or exaggeration of incidents (Schneider 1981; Schneider and Sumi 1981). The sample design is also almost inevitably flawed, with the most vulnerable (and crime prone) members of society including the homeless and residents of institutions typically absent from the population registers used to derive the initial sample. Nevertheless, crime surveys are now a mainstay of counts of crime across a wide range of countries because of the opportunities they afford for capturing details of the so-called 'dark figure of crime', i.e. the share of offences that go missing from police statistics because they were never reported or recorded (Biderman and Reiss 1967; Skogan 1977).

Whilst police recorded crime statistics and victim surveys remain the principal sources of crime data, we are witnessing an increasing number of data sources about offending and victimization. These include calls for police services, probation statistics, incidents recorded by health emergency services, social media data, gunshot detection technology and self-report crime surveys (Williams *et al.* 2017; Hibdon *et al.* 2021; Sutherland *et al.* 2021; Koziarski *et al.* 2022; Piza *et al.* 2023). However, these continue to be deployed rather sporadically and there is no centralized system for collating this information at a national scale.

Researchers and practitioners have subsequently used crime surveys to demonstrate that victims' willingness to report an incident to the police is contingent on the characteristics of the victim and offence (Hart and Rennison 2003; Tarling and Morris 2010). Reporting rates

1 In the United Kingdom, the collection and analysis of crime records dates back to the 1840s (Fletcher 1849).

differing systematically according to the victims' sex, age, ethnicity and income, as well as their relationship to the offender (Baumer 2002; Hart and Rennison 2003; Xie and Baumer 2019). Reporting rates also vary by crime type, with theft of motor vehicle and burglary typically being those with the highest reporting rates, and petty crimes such as theft and shoplifting being less likely to be reported to the police (Hart and Rennison 2003; Tarling and Morris 2010).

However, these comparisons are generally simplistic, with victim survey data frequently treated as an error-free 'gold standard' that police records are compared against. This is despite the fact that survey data are themselves subject to a range of widely known measurement errors and tend to have limited sample sizes at the level of small geographic areas (Rosenbaum and Lavrakas 1995). Studies have also tended to assume that 'more is better', with the higher crime counts typically observed in victim surveys taken as evidence of systematic bias in police recorded crime data (Gibson and Kim 2008; Pina-Sánchez *et al.* 2022). And when discussing discontinuities between data sources over time, studies have (implicitly) assumed a consistent survey design in addition to the lack of confounding between changes in reporting practices and changes in levels of crime over time (Office for National Statistics 2023).

More robust approaches to comparison that combine multiple data sources to quantify measurement error are now becoming available to researchers. In particular, MTMM models have enabled researchers to go beyond the assumption of a 'gold standard' and consider measurement errors in multiple data sources concurrently. For example, Oberski *et al.* (2017) showed how the MTMM approach can be used to estimate measurement error in administrative records and survey data in the context of employment statistics, and in criminology Yang *et al.* (2018) used an MTMM to assess the measurement of convergent and discriminant validity of measures of social disorder. This approach was recently used by Cernat *et al.* (2022) in the context of crime data, although their focus was on comparisons between different survey-based estimates. They found that the distribution of survey-based offence location estimates, as opposed to victim residence estimates, is highly similar to police recorded crime statistics, and there is little trade-off in terms of the reliability and validity of offence location and victim residence measures. Here, we use the same approach to provide a more nuanced picture of the quality of crime data at the subnational level than has been possible in existing studies.

DATA

In this study, we combined police recorded crime data with the CSEW for the period 2011–19, comparing the number of crimes estimated by each data source separately for each Police Force Area (PFA) and CSP. There are a total of 43 PFAs in England and Wales, differing substantially in size and internal composition (with the smallest, the City of London, covering just 2.6 km^2 and the largest, Dyfed-Powys, spanning nearly 11,000 km^2). Each PFA contains an average of seven CSPs, with a total of 312 in England and Wales. Both geographies structure police activity in some meaningful way. PFA represents the main structure for policing, with each PFA operating as a distinct policing unit with responsibility for its own budget and for setting its own policing priorities. Each PFA is also responsible for recording and 'cleaning' their own crime data, though all police forces are required to follow common counting rules established by the Home Office (2011; 2013). CSP, by contrast, is a police administrative boundary that has primarily been established to facilitate intelligence sharing across a range of local services that feed into policing activity within each PFA, as well as contributing to local anti-social behaviour strategies. In most regions, CSPs are aligned with local authorities geographically, with only a few isolated cases exhibiting minor variations.

Data from police recorded crimes and CSEW were harmonized prior to analysis, with four distinct crime categories that could be consistently measured by both data sources: violent crime (common assault and wounding), burglary, vehicle crime (including theft and damage to vehicles) and criminal damage (to households). Counts of crime within each category were then converted into rates per 1,000 residents to ensure comparability across areas. Data from the City of London PFA were omitted from the analysis because of the very low resident population and outlier crime distribution.

Police recorded crime data

Crime counts were derived directly from police recorded crime data for each PFA/CSP from Home Office published data tables (https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables). Annual counts of crime in each crime category were first combined into the four main crime types (see Supplementary Materials for the precise offence mapping), with the annual count of crime for each crime type and the area converted into an equivalent crime rate using PFA/CSP population and household totals from the 2011 Census (including the appropriate upward adjustment for years after 2011).

Crime survey data

We derived crime counts from CSEW data (Office for National Statistics 2021*a*) following the same approach used to produce national statistics (Office for National Statistics 2021*b*). Respondents to the survey are first asked whether they had experienced a range of different types of incidents in the last 12 months, with these initial screening questions used as the basis for a series of more detailed follow-up survey questions. From these 'victim forms' we identified the same four crime types and counted the total number of incidents experienced by each individual, as well as the number of these incidents subsequently reported to the police. These crime totals were then capped at a maximum of five to minimize year-to-year and area-to-area fluctuations[2] and used to estimate the weighted incidence rate per area (annually by PFA and biannually by CSP). This was calculated as the (weighted) mean number of incidents experienced by each sampled individual in each area of residence, with the incidence rate then re-scaled to rates per 1,000 people/households using the 2011 Census counts. Incidence rates were calculated annually for each PFA and biannually for each CSP to increase the area sample sizes at the smaller spatial scale.

METHOD

To assess the consistency of crime survey and police recorded crime data, we use an MTMM model to identify the unique sources of variation that contribute to differences in estimates of crime in different locations. This is a structural equation modelling approach that assumes an underlying 'true' level of crime in each PFA (or CSP), with the crime counts generated by the police and CSEW then treated as imperfect measurements of this true score. Each potential source of variation is represented by an unobserved (latent) variable, with the observed crime rates used to statistically identify each of these latent variables (Figure 1). Crime estimates derived from the same *source* (e.g. estimates of violent crimes, household burglaries, vehicle crimes and household criminal damage incidents that are recorded by the police) are used to identify three distinct latent 'Method' variables: police recorded crimes, experienced CSEW crimes and reported CSEW crimes. At the same time, rates of the same crime *type* that are derived from different sources (e.g. estimates of violent crimes that are recorded by the police, experienced CSEW crimes and reported CSEW crimes) are used to identify four distinct latent 'Trait' variables: violent crime, household burglary, vehicle crime and household criminal

2 This approach is commonly used by the Office for National Statistics to stabilize crime trend estimates using CSEW data. This approach has been criticized by Walby *et al.* (2016) for failing to properly capture the experiences of individuals (most often female victims of domestic violence) that may be subject to high levels of repeat victimization.

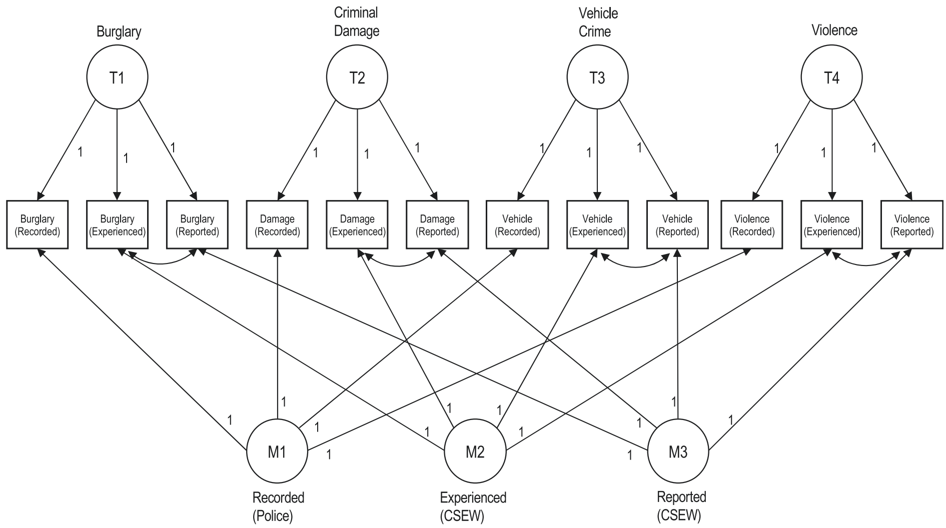

Fig. 1 Visual representation of our MTMM model. Squares represent observed variables and circles represent latent variables. Factor loadings are fixed to one for identification and correlated errors are included between observed variables measuring the same topic from the survey to account for the shared variance. Residual variances are not presented to facilitate reading.

damage. This means that each count of crime is used in the model twice, once to identify a latent method variable and once to identify a latent trait variable.

As a result, we have three latent method variables (same method/different crime type) and four latent trait variables (same crime type/different method), with each latent variable itself measured by multiple indicators. In each area, the four latent trait variables provide an estimate of the true rate of each crime type from the information that is consistent across the different sources (police recorded crimes and the two survey estimates), and the three latent method variables tell us how much this estimate differs depending on how it was measured. Estimating the MTMM on data covering all PFA (and CSP) therefore allows us to identify the unique variance associated with each method, and simultaneously identify the unique variance associated with each crime type. Any residual variation (i.e. differences in the counts of crime from the same PFA/CSP that cannot be explained by the measurement approach or type of crime) is then treated as a random measurement error.

The MTMM model can be formally written as:

$$y_{mt} = T_t + M_m + \varepsilon_{mt}$$

where y is the observed variable measured for each method m and trait t. The observed variables are explained by the latent trait variables (T_t), latent method variables (M_m) and indicator-specific residual (ε_{mt}). The traits capture the variance of interest that is consistently measured across the different methods. We refer to this as the corrected crime estimate. The method latent variable captures the variance that is specific to each measurement approach. This is the variance that is common across traits within each method. This is the complement of validity as it does not capture the concept of interest but is consistent across topics. Finally, the item-specific residual represents the random error of the measures and is the complement of reliability. In the context of this model, we define 'validity' as the proportion of the total observed variation that is due to the latent trait variable (T_t). We define lack of validity, or method effects, as the proportion of the total observation that is due to the method latent variable (M_m). This is defined as a lack of validity as it is a consistent source of variation that is present over multiple measures that do not capture the substantive topic of interest. Finally, we define lack of 'reliability' as the proportion of variation that is due to the residual (ε_{mt}). We can estimate these statistics for each question indicator (y_{mt}) and also aggregate them by trait (t), method (m) or both.

Loadings are fixed to one for identification purposes (Cernat and Oberski 2019; 2021). This is used as a way to define the latent variables as the common variation for all items measuring either the same trait (t) or method (m). This does not imply that the validity or reliability is the same across the indicators as these depend on the standardized coefficients which vary due to different amounts of item variation. We also include correlated errors between observed variables measuring the same topic and coming from the CSEW to account for the survey-specific shared variance. We estimated the model separately across all PFA and CSP at each time point.

Final models use the logged crime rate to address the skewed distribution of crimes across areas, and a full information maximum likelihood approach to adjust for missingness (areas with a CSEW estimated crime rate of zero due to sparse data are assumed to be Missing At Random given our measurement model). Models were estimated using Mplus Version 8.

RESULTS

First, we visually display recorded trends in crime across all PFAs based on our three different ways of capturing crime data. This is done to inspect the data and observe any obvious differences between data sources. Second, we explore bivariate correlations for the logged crime rates across geographic areas. And third, we present the full results from our MTMM models, presenting the estimated reliability ratios for overall crime rates, then considering the four different crime types separately, and finally exploring differences in data quality between police recorded crime data and survey estimates of crime.

Figure 2 plots annual crime rates across all PFAs in England and Wales for each crime category using the three measurement approaches. A similar trend is evident across all four crime types, with the survey data tending to identify a higher number of crimes than recorded crime data. This is confirmed when considering the intercepts from our MTMM model. There is also a general reduction in experienced crimes (the dotted line) and, to a lesser extent, reported crimes (the dashed line) over time. In contrast, there has been a general increase in police recording (the black line) that results in the gap between the three estimated crime lines generally closing.

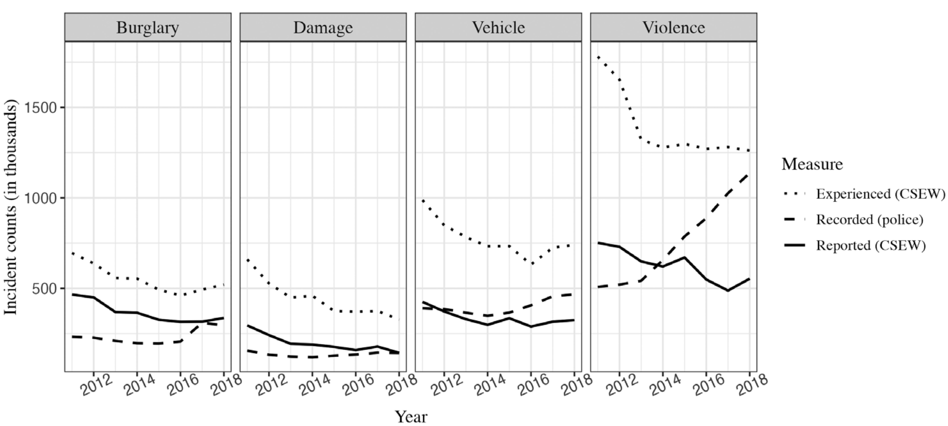

Fig. 2 Crime trends using police recorded crime and CSEW data (2011/12–2018/19). Annual crime counts are calculated by summing the total number of crimes per PFA (n = 43).

So much so that recording rates actually mirror (or even exceed) reported rates at the latter part of the data. This is most notable for violent and vehicle crimes.

Examination of the bivariate correlations for the (logged) crime rates presents a consistent picture, with strong positive correlations and modest evidence of stronger correlations between the estimates from the same methods than between the survey and police (Supplementary Figure S1). There is, however, some indication that the correlations have got weaker over time. A similar picture is evident, albeit with more modest correlations, when focus is moved to biannual estimates at the CSP level (Supplementary Figure S2).

The full results from our MTMM models are reported in Supplementary Tables S1 and S2, where the generally higher intercepts from the two survey estimated crime rates confirm the systematic undercount commonly associated with police recorded crime data. We also see the modest convergence between the datasets over time, with the logged (conditional) recording rates of vehicle and violent crime becoming higher than the equivalent CSEW estimates. However, of most interest are the estimated reliability ratios (Figures 3–5), which describe the quality

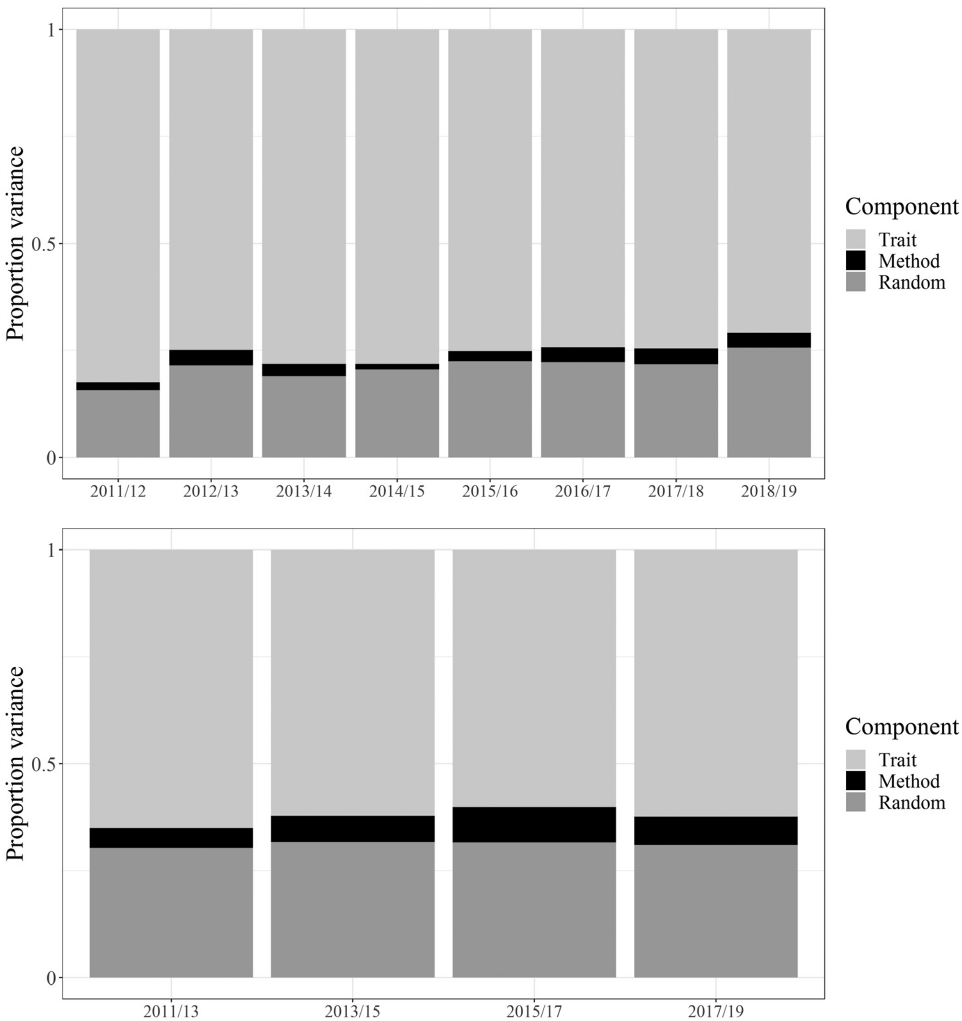

Fig. 3 Measurement error over time (all crimes) across PFAs (n = 42) and CSPs (n = 312).

profile for each source of crime data over time. Looking first at overall crime rates (Figure 3), we find that crime can be measured with a generally high degree of accuracy, with at least 75 per cent of the crime variation allocated to 'true' trait variance across all years. There is, however, a modest reduction in data quality over time, with both the random and method-specific variance increasing. The picture is similar when looking across PFA (top panel) and CSP level (bottom panel), albeit with somewhat reduced data quality overall when looking at the more spatially granular CSP estimates. This is not altogether surprising, with the CSP data presenting a noisier picture of the true crime landscape and method-specific variation also more evident.

A similar picture emerges when the four different crime types are considered separately (Figure 4), with a gradual worsening of data quality between 2011 and 2019 for all crime types and a reduced quality profile when considering crimes measured across CSPs (bottom panel). In addition, we note substantially more random error evident when considering criminal damage and violent crime (~30 per cent) when compared to burglary and vehicle crimes (~13 per cent). This may reflect the fact that violence and criminal damage offence categories are

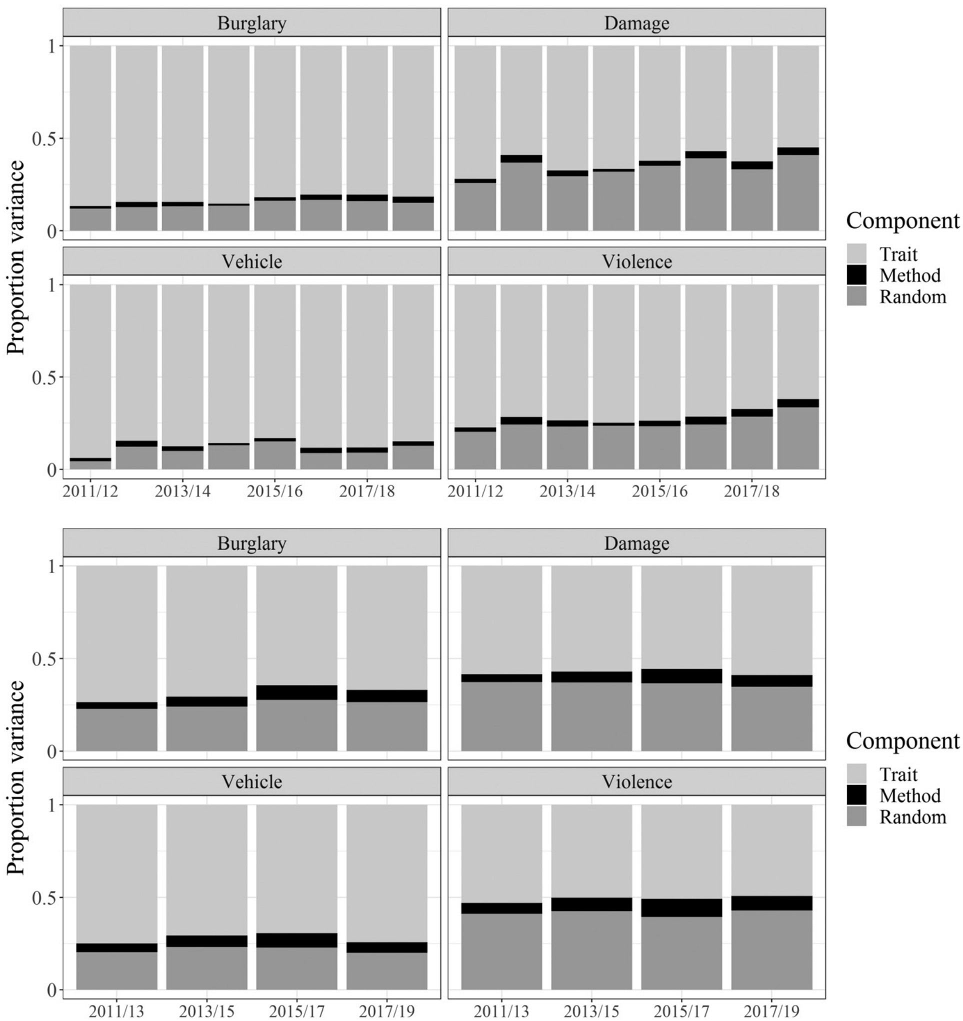

Fig. 4 Measurement error across crime types at PFA level (n = 42) and CSPs (n = 312).

somewhat broader in their scope, and suggests that there is more scope for individual differences in meaning and interpretation when considering these offences. As a result, they are generally measured with less precision. In contrast, method-specific errors remain generally similar across all crime types.

The results so far suggest that crime can be measured with a reasonable degree of accuracy at both the PFA and CSP levels, particularly for those crime categories like burglary and vehicle crime that are subject to the least amount of discretion and conceptual ambiguity. However, of most interest to us are the *relative* differences in quality between police recorded crime data and survey estimates of crime (Figure 5). Here, a more complex picture emerges. Whilst it is clear that the same pattern is evident overall—more error amongst violent crime and criminal damage, and a generally worsening picture over time and at the smaller spatial scale—this is not uniform across the data sources. In particular, there is considerably less random error when considering police recorded crime (4.5 per cent) and no clear indication that the magnitude

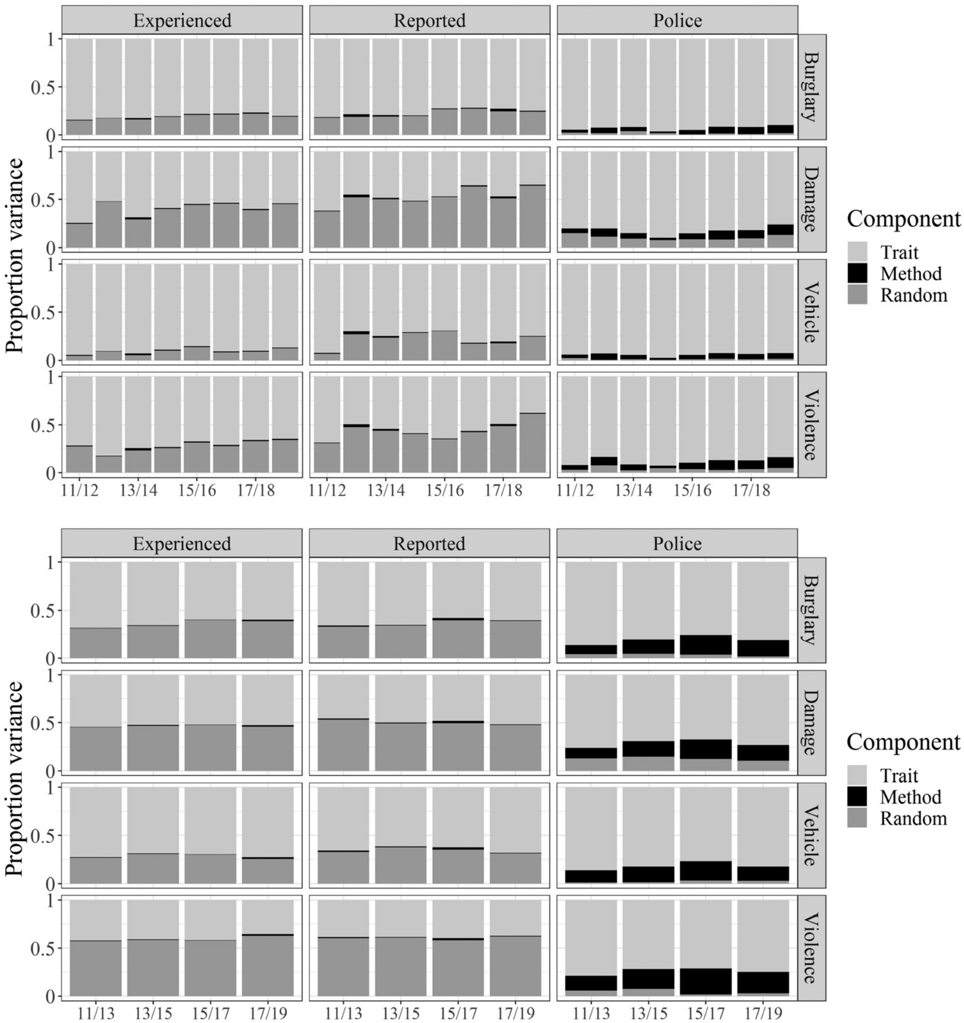

Fig. 5 Comparing the magnitude of measurement error between crime survey and police recorded crime data across PFAs (n = 42) and CSPs (n = 312).

of this error has increased over time. Interestingly, whilst police recorded crime data does not appear to be subject to much random error, we do observe a comparatively larger contribution of method effects (around 6 per cent), meaning that the validity estimates for police recorded crime data are around 90 per cent. Method effects are particularly noticeable for police recorded crime when the focus shifts to CSP.

In contrast, both survey estimates have substantial random errors (24 and 35 per cent for experienced and reported incidents at the PFA level), with reported incidents of violence and criminal damage particularly prone to random error across all years. The picture for CSEW is similar when considering CSP estimates, although here we do not see such a clear difference between experienced and reported crimes. Unlike police recorded crime data, method effects are far less prominent when survey estimates are considered.

DISCUSSION

In this article, we have used new developments in latent variable modelling to investigate data quality in crime measures. Whilst the general problems we outline are well known in the literature, our study represents an important advance over existing studies by providing an accurate quantification of the validity and reliability of existing approaches to counting crime. By treating different crime estimates (police recorded crimes, experienced CSEW crimes and reported CSEW crimes) as different methods that measure the same concepts of interest (burglary, criminal damage to a dwelling, vehicle crime and violence), we have been able to bypass many of the limitations that are intrinsic to each data source. Our results identify a complex pattern in data quality. At its core there appears to be a trade-off between the systematic bias in the means—with police records capturing on average fewer crimes—and consistent variance estimates—with surveys showing higher levels of random error.

Overall, then, when the research goal is not to retrieve true estimates of crime, but rather explore crime variability across areas, then police data may not be as bad as commonly thought. In fact, we find that the reliability of police data is high and stable, with only around 10 per cent of the variation between areas identified as error (including both random and method-specific variation). Whilst we have no way of knowing that this picture will remain ad infinitum, or that it would apply equally across different jurisdictions, there is no obvious reason to think that the crime recording landscape in the United Kingdom should behave in a markedly different fashion to other countries. This means that these data are well equipped to explore the variability of crime across time, at least when crime records are aggregated annually or biannually. This finding should not be understated. Recent work from Pina-Sánchez *et al.* (2022) demonstrated that it is straightforward to adjust for the potential biases resulting from the systematic undercounting inherent in crime data, but pointed to a more uncertain way forward when random errors are considered and a need to incorporate more complex sensitivity analyses in this context. The fact that we find these random errors are modest in nature suggests they may be reasonably ignored in most analyses unless there are reasons to expect the levels of undercounting are associated with other focal variables of interest.

Another important takeaway is that, whilst the measurement quality of these four types of crimes is overall good, it does vary significantly by topic and aggregation level. The average amount of variance that is attributed to the trait over all the questions, years and methods is 76 per cent. This implies that around a quarter of the observed variance in crime measures does not capture the concept of interest. The measurement quality is better when burglary and vehicle crimes are considered (83 and 87 per cent of the variance is from the trait, respectively) than when measuring damage and violence (63 and 71 per cent, respectively). Also, using data aggregated at the CSP level shows worse quality (62 per cent trait variance) than when using PFA, confirming that crime data quality worsens at smaller spatial levels (Buil-Gil *et al.* 2022). Trait variance is likely to be even larger in crime records aggregated at the highly detailed levels of analysis typically considered in place-based criminology and policing, such as neighbourhoods, micro places and street segments (Braga and Weisburd 2010). Further work is thus needed to study the measurement qualities of estimates of crime obtained from police and survey data for small geographic areas.

Like all studies, this work is not without its limitations. In particular, whilst the MTMM model has enabled us to relax the assumption that survey data represent an error-free 'gold standard', this is replaced by other identifying assumptions. One of the most important being that the different 'methods' capture the same phenomenon. Differences observed across crime data sources will likely not be due solely to measurement error. How crime is conceptualized is also central, with different sources of crime data measuring distinct, but related, phenomena. For instance, police statistics record crimes that happen in an area, whilst estimates of crime obtained from surveys show crimes in places where victims live, thus systematically underestimating crime in places with a low residential population and a large 'ambient' population (Cernat *et al.* 2022). The crimes captured by the two sources may also be different, with many of the incidents identified in a victim survey unlikely to be reported to the police (perhaps because they are not deemed serious enough, or because of public distrust or fear of reprisals) and many of those incidents reported to the police not ultimately recorded (if they are less serious or details are ambiguous). This last point is confirmed by the presence of a substantial latent method variance in police data, which implies that whilst recording practices are internally consistent, they are systematically different from the survey estimates. Our results are also specific to the data-generating context of the two data sources when they were collected, with any substantial changes to the method of data collection potentially leading to different results. This is most relevant when CSEW estimates are considered, with the ONS having recently consulted on widespread changes to the sample design following the changes in survey delivery during covid (ONS 2022). However, whilst modest reductions in the size of the random component of error might be anticipated if sample sizes increased substantially, it is unlikely that these will ever reach the scale of police recorded crime data. They might also be plausibly offset by increases to the method effect.

CONCLUSION

We have provided compelling evidence that whilst recorded crime data may systematically undercount the true extent of crime, it exhibits a high level of precision, outperforming victimization survey data when considering the spatial patterning of crime. This is important, with existing work demonstrating that systematic errors are considerably more straightforward to anticipate and correct without the need for recourse to complex measurement error approaches. Consequently, whilst survey data remain the most important resource to provide evidence of the dark figure of crime and the nature of crime underreporting, recorded crime data can and should be given primacy in empirical studies that aim to examine the causes and consequences of crime across and within geographic areas. Of course, the pattern elicited here refers to relatively large geographic areas. When considering geographic scales like cities or neighbourhoods the difference in reliability between police data and national victimization surveys could be expected to be even starker as crime survey estimates become increasingly unstable.

Our findings therefore support the approach adopted by ONS when reporting survey data on the extent of crime (e.g. ONS 2023), with CSEW data generally unsuitable for producing subnational crime estimates in its raw form. In contrast, the approach to police recorded crime data may be too conservative, with the errors exhibiting a remarkable level of temporal stability that makes them suitable for exploring crime trends. Our work also presents a new way forward for the potential *integration* of CSEW data with police recorded crimes. The current study focussed on the correct estimation of the quality profile of each data source, but the MTMM model also contains a 'true' crime count in each area that correctly reflects the presence of measurement errors (the latent traits). The routine estimation and publication of these latent traits alongside estimates from police recorded crime data by ONS as part of their quarterly publications could go some way to providing a more robust picture of the systematic undercounting from police data across PFA (and CSP).

FUNDING

This work was supported by the Economic and Social Research Council [grant number: ES/T015667/1].

SUPPLEMENTARY MATERIAL

Supplementary material is available at *British Journal of Criminology* online.

REFERENCES

Aebi, M. F. and Linde, A. (2014), 'National Victimization Surveys', in G. Bruinsma and D. Weisburd, eds., *Encyclopedia of Criminology and Criminal Justice*, 3228–42. New York: Springer.

Ariel, B. and Bland, M. (2019), 'Is Crime Rising or Falling? A Comparison of Police-Recorded Crime and Victimization Surveys', in M. Deflem and D. M. D. Silva, eds., *Methods of Criminology and Criminal Justice Research*, 7–31. Emerald Publishing Limited.

Baumer, E. P. (2002), 'Neighborhood Disadvantage and Police Notification by Victims of Violence', *Criminology*, 40: 579–616.

Berg, M. T. and Lauritsen, J. L. (2016), 'Telling a Similar Story Twice? NCVS/UCR Convergence in Serious Violent Crime Rates in Rural, Suburban, and Urban Places (1973–2010)', *Journal of Quantitative Criminology*, 32: 61–87.

Biderman, A. D. and Reiss, A. J. (1967), 'On Exploring the "Dark Figure" of Crime', *The Annals of the American Academy of Political and Social Science*, 374: 1–15.

Bland, M. P. and Ariel, B. (2020), *Targeting Domestic Abuse with Police Data*. Springer.

Boivin, R. and Cordeau, G. (2011), 'Measuring the Impact of Police Discretion on Official Crime Statistics: A Research Note', *Police Quarterly*, 14: 186–203.

Braga, A. A. and Weisburd, D. L. (2010), 'Editors' Introduction: Empirical Evidence on the Relevance of Place in Criminology', *Journal of Quantitative Criminology*, 26: 1–6.

Buil-Gil, D., Moretti, A. and Langton, S. H. (2022), 'The Accuracy of Crime Statistics: Assessing the Impact of Police Data Bias on Geographic Crime Analysis', *Journal of Experimental Criminology*, 18: 515–41.

Burrows, J., Tarling, R., Mackie, A., Lewis, R. and Taylor, G. (2000), 'Review of Police Forces' Crime Recording Practices', *Home Office Research Study 204*. London: Home Office Research Studies.

Cernat, A., Buil-Gil, D., Brunton-Smith, I., Pina-Sánchez, J. and Murrià-Sangenís, M. (2022), 'Estimating Crime in Place: Moving Beyond Residence Location', *Crime & Delinquency*, 68: 2061–91.

Cernat, A. and Oberski, D. L. (2019), 'Extending the Within-Persons Experimental Design: The Multitrait-Multierror MTME Approach', in P. Lavrakas, M. Traugott, C. Kennedy, A. Holbrook, E. de Leeuw and B. West, eds., *Experimental Methods in Survey Research*, 1st edn, 481–500. Wiley.

—— (2021), 'Estimating Stochastic Survey Response Errors Using the Multitrait-Multierror Model', *Journal of the Royal Statistical Society Series A: Statistics in Society*, 185: 134–55.

de Candolle, A. (1830[1987]), 'Considérations sur la statistique des délits', *Déviance et Société*, 11: 352–5.

Ennis, P. H. (1967), *Criminal Victimization in the United States. A Report of a National Survey*. Washington, DC: US Government Printing Office.

Farrell, G., Tilley, N. and Tseloni, A. (2014), 'Why the Crime Drop?', *Crime and Justice*, 43: 421–90.

Fletcher, J. (1849), 'Moral and Educational Statistics of England and Wales', *Journal of the Statistical Society of London*, 12: 189–335.

Gibson, J. and Kim, B. (2008), 'The Effect of Reporting Errors on the Cross-country Relationship Between Inequality and Crime', *Journal of Development Economics*, 87: 247–54.

Hart, T. C. and Rennison, C. (2003), *Reporting Crime to the Police, 1992–2000*. Special Report. Bureau of Justice Statistics.

Her Majesty Inspectorate of Constabulary. (2014), *Crime-Recording: Making the Victim Count—The Final Report of an Inspection of Crime Data Integrity in Police Forces in England and Wales*. Justice Inspectorates, available online at https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/crime-recording-makingthe-victim-count.pdf

Hibdon, J., Telep, C. W. and Huff, J. (2021), 'Going Beyond the Blue: The Utility of Emergency Medical Services Data in Understanding Violent Crime', *Criminal Justice Review*, 46: 190–211.

Home Office. (2011), 'National Standard for Incident Recording Counting Rules', available online at https://www.gov.uk/government/publications/the-national-standard-for-incident-recording-nsir-counting-rules

—— (2013), 'Counting Rules for Recorded Crime', available online at https://www.gov.uk/government/publications/counting-rules-for-recorded-crime#full-publication-update-history

Klinger, D. A. and Bridges, G. S. (1997), 'Measurement Error in Calls-for-Service as an Indicator of Crime', *Criminology*, 35: 705–26.

Koziarski, J., Ferguson, L. and Huey, L. (2022), 'Shedding Light on the Dark Figure of Police Mental Health Calls for Service', *Policing: A Journal of Policy and Practice*, 16: 696–706.

Lohr, S. H. (2019), *Measuring Crime: Behind the Statistics*. Boca Raton: CRC Press.

Oberski, D. L., Kirchner, A., Eckman, S. and Kreuter, F. (2017), 'Evaluating the Quality of Survey and Administrative Data with Generalized Multitrait-Multimethod Models', *Journal of the American Statistical Association*, 112: 1477–89.

Office for National Statistics. (2021*a*), *Crime Survey for England and Wales, 2011–2020: Secure Access, Low-Level Geographic Data* [data collection], 7th edn. UK Data Service. SN: 7311.

—— (2021*b*), 'User Guide to Crime Statistics for England and Wales: March 2020', available online at https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/methodologies/userguidetocrimestatisticsforenglandandwales#comparison-of-the-csew-and-police-recorded-crime

—— (2022), *Consultation on the Redesign of the Crime Survey for England and Wales*, available online at https://consultations.ons.gov.uk/external-affairs/crime-consultation/supporting_documents/Consultation%20Document%20%20CSEW%20Consultation%20%20Final.pdf

——(2023), 'Crime in England and Wales: Year Ending December 2022', available online at https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/crimeinenglandandwales/yearendingdecember2022

Pina-Sánchez, J., Brunton-Smith, I., Buil-Gil, D. and Cernat, A. (2023), 'Exploring the Impact of Measurement Error in Police Recorded Crime Rates Through Sensitivity Analysis', *Crime Science*, 12: 14.

Pina-Sánchez, J., Buil-Gil, D., Brunton-Smith, I. and Cernat, A. (2022), 'The Impact of Measurement Error in Regression Models Using Police Recorded Crime Rates', *Journal of Quantitative Criminology*, 39: 975–1002.

Piza, E. L., Hatten, D. N., Carter, J. G., Baughman, J. H. and Mohler, G. O. (2023), 'Gunshot Detection Technology Time Savings and Spatial Precision: An Exploratory Analysis in Kansas City', *Policing*, 17: paac097.

Rosenbaum, D. P. and Lavrakas, P. J. (1995), 'Self-reports About Place: The Application of Survey and Interview Methods to the Study of Small Areas', in J. E. Eck and D. Weisburd, eds., *Crime and Place. Crime Prevention Studies*, 285–314. Criminal Justice Press.

Schneider, A. L. (1981), 'Methodological Problems in Victim Surveys and Their Implications for Research in Victimology', *The Journal of Criminal Law and Criminology*, 72: 818–38.

Schneider, A. L. and Sumi, D. (1981), 'Patterns of Forgetting and Telescoping: An Analysis of LEAA Survey Victimization Data', *Criminology*, 19: 400–10.

Sellin, T. (1931), 'The Basis of a Crime Index', *Journal of Criminal Law and Criminology (1931–1951)*, 22: 335–56.

Skogan, W. G. (1977), 'Dimensions of the Dark Figure of Unreported Crime', *Crime & Delinquency*, 23: 41–50.

Sutherland, A., Strang, L., Stepanek, M., Giacomantonio, A., Boyle, A. and Strang, H. (2021), 'Tracking Violent Crime with Ambulance Data: How Much Crime Goes Uncounted?', *Cambridge Journal of Evidence-Based Policing*, 5: 20–39.

Tarling, R. and Morris, K. (2010), 'Reporting Crime to the Police', *British Journal of Criminology*, 50: 474–90.

Walby, S., Towers, J. and Francis, B. (2016), 'Is Violent Crime Increasing or Decreasing? A New Methodology to Measure Repeat Attacks Making Visible the Significance of Gender and Domestic Relations', *British Journal of Criminology*, 56: 1203–34.

Warner, B. D. (1997), 'Community Characteristics and the Recording of Crime: Police Recording of Citizens' Complaints of Burglary and Assault', *Justice Quarterly*, 14: 631–50.

Williams, M., Burnap, P. and Sloan, L. (2017), 'Crime Sensing with Big Data: The Affordances and Limitations of Using Open-Source Communications to Estimate Crime Patterns', *The British Journal of Criminology*, 57: 320–40.

Xie, M. and Baumer, E. P. (2019), 'Crime Victims' Decisions to Call the Police: Past Research and New Directions', *Annual Review of Criminology*, 2: 217–40. https://doi.org/10.1146/annurev-criminol-011518-024748

Yang, S.-M., Hinkle, J. and Wyckoff, L. (2018), 'Using Multitrait-Multimethod (MTMM) Techniques to Examine the Convergent and Discriminant Validity of Social Disorder', *Journal of Research in Crime and Delinquency*, 55: 571–608.