UNIVERSITY OF LEEDS

This is a repository copy of *The Impact of Measurement Error in Regression Models Using Police Recorded Crime Rates*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/197197/

Version: Accepted Version

Article:

Pina-Sánchez, J, Buil-Gil, D, Brunton-Smith, I et al. (1 more author) (2023) The Impact of Measurement Error in Regression Models Using Police Recorded Crime Rates. Journal of Quantitative Criminology, 39. pp. 975-1002. ISSN 0748-4518

https://doi.org/10.1007/s10940-022-09557-6

© The Author(s). This is an author produced version of a article published in Journal of Quantitative Criminology. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

The Impact of Measurement Error in Regression Models Using Police Recorded Crime Rates

Jose Pina-Sánchez[1]
David Buil-Gil[2]
Ian Brunton-Smith[3]
Alexandru Cernat[4]

Suggested running head: The Impact of Measurement Error in Crime Rates

Abstract

Objectives: Assess the extent to which measurement error in police recorded crime rates impact the estimates of regression models exploring the causes and consequences of crime.

Methods: We focus on linear models where crime rates are included either as the response or as an explanatory variable, in their original scale or log-transformed. Two measurement error mechanisms are considered, systematic errors in the form of under-recorded crime, and random errors in the form of recording inconsistencies across areas. The extent to which such measurement error mechanisms impact model parameters is demonstrated algebraically using formal notation, and graphically using simulations.

Results: The impact of measurement error is highly variable across different settings. Depending on the crime type, the spatial resolution, but also where and how police recorded crime rates are introduced in the model, the measurement error induced biases could range from negligible to severe, affecting even estimates from explanatory variables free of measurement error. We also demonstrate how in models where crime rates are introduced as the response variable, the impact of measurement error could be eliminated using log-transformations.

Conclusions: The validity of a large share of the evidence base exploring the effects and consequences of crime is put into question. In interpreting findings from the literature relying on regression models and police recorded crime rates, we urge researchers to consider the biasing effects shown here. Future studies should also anticipate the impact in their findings and employ sensitivity analysis if the expected measurement error induced bias is non-negligible.

Keywords: police data; crime rates; measurement error; bias

Funding details: This work is supported by the Secondary Data Analysis Initiative of the Economic and Social Research Council (Grant Ref: ES/T015667/1).

[1] School of Law, University of Leeds, Leeds, LS2 9JT, UK, j.pinasanchez@leeds.ac.uk, +44 (0)7909414923.
[2] School of Social Sciences, University of Manchester, Manchester, UK
[3] Department of Sociology, University of Surrey, Guildford, UK
[4] School of Social Sciences, University of Manchester, Manchester, UK

1. INTRODUCTION

It is widely acknowledged that police recorded crime data is deeply flawed, subject to different forms of measurement error. This data fails to reflect incidents that are not detected by the police, leading to systematic under-estimations of the true figure of crime (Biderman & Reiss, 1967; Coleman & Moynihan, 1996; Skogan, 1977), while it is also affected by substantial recording inconsistencies between and within police forces (Boivin & Cordeau, 2011; Her Majesty Inspectorate of Constabulary, 2014).

Despite its questionable measurement properties, police data is heavily relied upon by researchers as it holds important advantages over other sources of crime data in terms of accessibility and versatility – allowing for spatiotemporal resolutions unavailable to victimisation and offenders surveys. As such, police recorded crime rates are commonly used in the process of building and testing crime theory (see for example research on social disorganisation and collective efficacy, Duncan et al., 2003, Sampson et al., 1997; or rational choice and routine activity theories, Cohen & Felson, 1979, Matsueda et al., 2006). Police data is also central in studies that, from a more exploratory perspective, seek to identify predictors of crime (Bowers and Johnson, 2005; Ellis et al., 2019). There is also a large group of studies that have relied on police recorded crime data as an explanatory variable, seeking to estimate the effect of crime on a wide range of phenomena such as fear of crime (Krahn & Kennedy, 1985; Zhao et al., 2015), or police use of force (McCarthy et al., 2019; Sobol et al., 2013). Beyond Criminology, studies making use of police data are also common in areas of Sociology (Lee & Ousey, 2005; Miethe et al., 1991), Social Policy (Machin & Meghir, 2004; Whitworth, 2012), Epidemiology (Browning et al., 2012; Messer et al., 2006), Geography (Keels et al., 2005; Morenoff & Sampson, 1997), and Economics (Han et al., 2013; Philipson & Posner, 1996), where the relationship between crime and socio-economic inequality, deprivation, or ethnic heterogeneity have been of special interest. It is therefore no exaggeration to suggest that police statistics represent the most important data source in the study of the causes and consequences of crime.

However, with some notable exceptions (see for example, Barnett, 1981; Brantingham, 2018; Fajnzylber et al., 2002; Farrell & Pease, 2003; Gibson & Kim, 2008; Levitt, 1998; Martin & Legault; 2005; Neumayer, 2005; Pepper et al., 2010; Pudney et al., 2000; Vollaard & Hamed, 2012), researchers have generally failed to sufficiently recognise the implications of using police data prone to measurement error on the validity of their results. If variables affected by measurement error are introduced in multivariate models, they will often lead to biased estimates (Fuller, 2009; Gustafson, 2003). Given the large prevalence of measurement error in police statistics, bias in regression models relying on this data may be substantial. Furthermore, the magnitude and direction of those biases can be difficult to anticipate, as the measurement error impact will likely propagate through the model, affecting the accuracy of not just crime estimates, but all model estimates and their respective measures of uncertainty (Nugent et al., 2000).

To date, we do not have a general understanding of the impact that common forms of measurement error present in police recorded crime have across typical models used in the literature. The small group of studies that have previously explored this problem have mostly focused on specific applications. That is, they have explored the impact that measurement error could exert when police data is used to investigate specific research questions. See for example discussions on how measurement error could be biasing estimates of the effect of economic inequality (Fajnzylber et al., 2002; Gibson & Kim, 2008; Neumayer, 2005), police arrests/presence (Levitt, 1998; Vollaard, 2012), or gun ownership (Maltz & Targonski, 2002; Martin & Legault, 2005) on crime. Even the most comprehensive studies in the literature, where new estimators to adjust for the impact of measurement error have been developed – either invoking a set of assumptions about the measurement error term, or auxiliary data from victimisation surveys – are limited in scope to a specific outcome model with crime data introduced in a specific form, always as the outcome variable[5].

Such focus on specific applications has only been able to provide a narrow view of what is a much larger problem. This is because the presence of measurement error varies heavily across crime types, mainly as a result of differential reporting rates (Hart & Rennison, 2003; Tarling & Morris, 2010), but also across the chosen spatial area of analysis, with crime rates measured at lower spatial units being less reliable (Buil-Gil et al., 2021a). While the impact associated to the same measurement error could vary even more intensely depending on modelling decisions such as: the type of outcome model to be specified, where in the model

[5] See for example Brantingham (2018), or Pepper et al. (2010), where modelling strategies to adjust for measurement error in police data are used for binary outcome models used in hotspot policing, or time-series analysis assessing changes in crime rates across time.

the variable affected by measurement error is introduced (i.e. as an outcome or explanatory variable), or whether the affected variable is subject to some form of transformation before or as part of the estimation process. Hence, outside the few research questions where the effect of measurement error has been actively studied, there is an overall lack of understanding about the extent to which estimates from models relying on police data are biased. This, in our view, is nothing short of the largest methodological challenge affecting the empirical literature exploring the causes and consequences of crime.

Here, we provide a more encompassing overview of the biasing effect that measurement error present in police recorded crime data exerts across standard regression models commonly used in the literature. In doing so, our aim is to raise awareness about the problem, but also to facilitate interpretations of the validity of previous studies relying on police recorded crime, and ultimately minimise its impact in future studies. For simplicity, we focus on the most widely used form of police data: crime rates recorded across geographical areas at a given point in time. We therefore set aside other uses of police data which are also prone to measurement error, albeit taking a different form, such as problems of misclassification that affect police data when measured as a binary outcome (Brantingham, 2018; Caplan et al., 2011; Vandeviver et al., 2015), or the effect of measurement error in dynamic models including autoregressive terms (Pudney et al., 2000; Cantor & Land, 1985; Greenberg, 2001).

The broader perspective sought in this study is achieved by exploring the impact associated to: i) a wider combination of the types of errors that could be expected across different crime types and spatial areas; ii) models where crime rates are introduced as the outcome variable, but also, as an explanatory variable; and iii) models where crime rates are introduced in their original form, or after they have been log-transformed. This allows us to move beyond specific applications and shed new light on the impact of measurement error in areas that have not yet been explored. For example, the consideration of interacting measurement mechanisms is key to understand what their expected impact will be since, depending on the setting, measurement error mechanisms can operate in different directions, potentially cancelling themselves out entirely, while in other instances they can operate in the same direction, reinforcing each other's biasing effects. Similarly, and as far as we are aware, no other study has assessed the impact that could be expected when police recorded crime rates are introduced as an explanatory variable. This represents a substantial gap in the literature, affecting not only previous studies exploring the consequences of crime relying on police data, but also potentially, any other study on any other subject, where police recorded crime rates are used as a control variable. Lastly, by contemplating the impact of crime rates both in their original form and log-transformed, we cover the two main forms used to introduce crime rates in regression models in the literature, while we also demonstrate how either log-transforming crime rates, or specifying them using generalised linear models with logs as the link function (Osgood, 2000), could in many cases represent a simple yet highly effective approach to minimise the impact of measurement error.

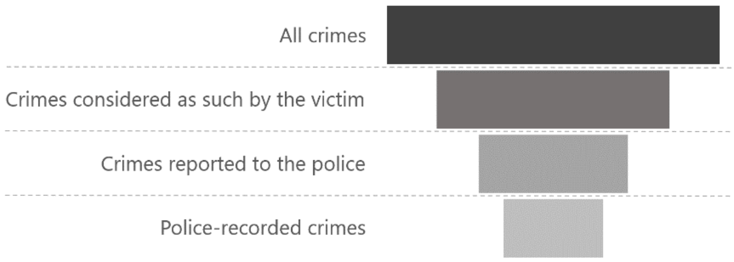
Our analytical strategy is twofold, based first on a formal approximation using algebra, further enhanced through simulations at a second stage. The former defines the specific impact that could be attributed to different measurement error mechanisms present in police recorded crime rates, while the latter facilitates visualising their combined impact across a wide range of scenarios. But first we proceed to illustrate the form and prevalence of measurement error that could be expected in police recorded crime rates. We do so both theoretically and empirically through comparisons with crime estimates derived from victimisation surveys and a register of vital statistics.

2. PREVALENCE AND NATURE OF MEASUREMENT ERROR IN POLICE RECORDED CRIME RATES

To assess the measurement properties of police recorded crime rates, we first need to define the concept that researchers are trying to capture when using such data. Generalising, researchers use police crime rates to reflect the underlying extent of crime. However, what constitutes 'crime' is not always clear-cut. Broadly speaking, we can consider four conceptualisations of crime, which can be ordered as a sequence of subsets of 'all crimes' according to their breadth, as shown in Figure 1.

A growing body of research is interested in exploring the precursors and management strategies of police demand (Ashby, 2020; Laufs et al., 2020). In such cases, the specific phenomenon that researchers seek to capture is crimes that are reported (or known) to the police, represented by the second level from the bottom in Figure 1. In some other instances researchers seek to capture the broadest conceptualisation of crime, represented by the top level, including so-called 'victimless' crimes, or those where the victim is not aware of its condition, as it is often the case in fraud or cybercrime (Van de Weijer et al., 2019). However – even though this is rarely stated explicitly - most studies relying on police recorded crime rates use them as a proxy for the extent of crime that is considered as such by the victim, represented by the second level from the top in Figure 1. Hence, in this study we take the number of crimes of which victims are aware (expressed in rates) as the true value of crime that most researchers aim to capture, and consequently define measurement error as the discrepancy between that and the crime rates recorded by the police.

Fig. 1. Different conceptualisation of crime



As illustrated schematically in Figure 1, police recorded crime rates under-estimate the true extent of crime, as they are affected by victims' willingness to report an incident to the police; an effect which varies by demographic groups and crime types (Hart & Rennison, 2003; Tarling & Morris, 2010). Crime reporting rates differ systematically according to the victims' sex (females report more often than males), their relationship to the offender (reporting rates are smaller when the offender is a stranger), but also based on victims' age, ethnicity and income (Baumer, 2002; Hart & Rennison, 2003). There are also stark differences in reporting rates by crime types, 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 & Rennison, 2003; Tarling & Morris, 2010).

Table 1 presents the estimated reporting rates for crime types commonly considered in the literature. These reporting rates are derived from the Crime Survey for England and Wales (CSEW) and the National Crime Victimization Survey (NCVS), which include questions on whether crimes came to be known to the police. The two surveys are not perfectly comparable because of differences in the specific offence types included (see Appendix), however they show some important similarities. Both show how reporting rates vary markedly across crime types, with motor vehicle theft reaching close to perfect reporting rates, while fewer than half of property crimes are reported to the police.

Table I. Reporting rates for different crime types

	CSEW 2018 to 2019		NCVS 2017 to 2020	
	Cases reported in the survey	% known to police (weighted)	Cases reported in the survey	% known to police (weighted)
Violent crime	1979	38.8%	516	46.6%
Property crime	2035	36.7%	995	41.8%
Burglary	719	59.5%	248	45.4%
Motor vehicle theft	130	89.7%	33	73.5%
All crimes	7,840	37.3%	3,209	42.04%

Reporting rates may also differ across geographic areas (Buil-Gil et al., 2021b; Xie & Baumer, 2019), reflecting variations in citizens' perceptions of the police and their willingness to cooperate with police services (Jackson et al., 2016; McCandless et al., 2016). However, the extent and variability with which crimes are reported is not the only problem affecting police statistics.

Once brought to the attention of the police, the decision to record an incident as a crime is the result of a complex interaction of various counting rules and protocols, not always standardised across police forces, where personal discretion plays a large role (Burrows et al., 2000). An officer must first determine whether an incident meets the legal threshold to be considered a crime, before making an individual judgement on whether to proceed with the registration process (Klinger & Bridges, 1997). Having decided to do so, the crime is then classified according to pre-defined criteria. Here again there is evidence of considerable variability across police forces (Burrows et al., 2000; Her Majesty Inspectorate of Constabulary, 2014; von Hofer, 2000), with researchers also pointing to systematic under-counting in certain types of areas, including high-rise housing areas (Bottoms et al., 1987) and rural areas (Berg & Lauritsen, 2016), and even more flagrant data manipulation practices as a result of managerial and political pressures (Eterno et al., 2014).

These different forms of measurement error affecting police recorded crime rates can be grouped in two main categories: systematic and random errors. Problems of under-reporting and under-counting represent systematic errors, since they lead to a downward bias in the proportion of crimes recorded across all areas. By contrast, inconsistencies in crime reporting and recording processes across victims, areas and police forces could be considered random errors, as they introduce undue variability (i.e. noise) in police crime rates. That is, the former group of errors impact the validity of police crime rates, while the latter affects their reliability (Lohr, 2019).

How these types of errors relate to the unobserved true crime rates is less clear. Only a subset of the small group of studies exploring the presence of measurement error in police data have sought to define this question formally, and amongst those few not all follow the same approach. Broadly speaking we can distinguish two main groups, based on whether the measurement error is thought to be additive or multiplicative. The additive model represents the standard functional form used to conceptualise measurement error problems (Novick, 1966; Stefanski & Carroll, 1995), which has been adopted for the exploration of measurement error in crime rates in important studies such as Fajnzylber et al. (2002) and Pepper et al. (2010). Under such models, the measurement error (U) present in the observed and imperfectly measured crime rate variable (X^*) is thought to be related to the true but unobserved variable (X) additively: $X^* = X + U$, with $E(U) \neq 0$ if the errors are systematic as opposed to entirely random. Alternatively, Gibson and Kim (2008) and Pudney et al. (2000) have viewed the relationship between the errors and the true crime rate as multiplicative: $X^* = XU$, with $E(U) \neq 1$ if the errors are systematic. Such multiplicative representation implies that the magnitude of the error term is proportional to the true prevalence of crime, which has been commonly employed in applications exploring the presence of measurement error in count and duration data (Glewwe, 2007; Pickles et al., 1996; Skinner & Humphreys, 1999), which just like crime rates are left-censored and typically right-skewed.

If, as shown in Figure 1, we consider police recorded crime as a subset of the true extent of crime, the proportional relationship between true crime rates and errors posited by the multiplicative model (i.e. higher crime areas will lead to larger errors) seems appropriate. However, it is important not to dismiss the additive model entirely. Often, researchers introduce crime rates in their models after they have been log-transformed. This is done either to interpret effects in relative terms (Goulas & Zervoyianni, 2013; Witt & Witte, 2000), or to normalise right-skewed crime rates (Sutherland et al., 2013; Whitworth, 2012). The latter is also achieved through generalised linear models where logs are used as the link function such as Poisson or negative binomial models (Osborn & Tseloni, 1998; Sampson et al., 1997). The use of such log-based generalised linear models was advocated by Osgood (2000) as a strategy to improve the specification of crime rates. Incidentally, when such models are employed, besides potentially enhancing the specification of crime rates any multiplicative errors affecting crime rates will be transformed into additive errors, since: $\log(X^*) = \log(XU) = \log(X) + \log(U)$. Hence, even if the measurement error present in police recorded crime rates is assumed to be multiplicative, when considering the potential impact that such errors could have, we should refer to the additive model. This apparently minor detail has been so far overlooked by crime researchers, however, as we will see, differentiating between an additive and a multiplicative measurement error could lead to vastly different measurement error induced bias in regression models.

2.1. Empirical Assessment of the Extent and Nature of Measurement Error in Police Recorded Crime Rates

To test our conceptualisation of measurement error affecting police data we undertake two comparisons. First, we compare the rate of property crimes recorded by the different police forces in England and Wales for the year ending March 2012 against estimates from the CSEW (2011/12).[6] Specifically, we compare the

[6] The CSEW sampling approach is designed to enable the calculation of reliable victimisation estimates at the PFA level, with an average sample of 1,096 respondents in each area ($\min = 917, \max = 4{,}023$). PFA is an UK spatial

average number of property crimes recorded per household in each police force area (PFA) in England and Wales in 2011/12, against matching types of crimes estimated from the CSEW in that same period and areas. To obtain comparable groups, we aggregate the following offence categories recorded by the police and the CSEW: vehicle theft, bicycle theft, and residential burglary. Crime rates are estimated from the CSEW for a comparable subset of measured offences using the crime mappings outlined in the Office for National Statistics crime statistics user guide (see ONS, 2015: 36), including all incidents irrespective of whether or not victims reported them to the police. Police recorded crime data is accessed from the Home Office open data tables.[7] Comparing these two estimates allows us to understand the full extent of the discrepancies between incidents experienced by crime victims and those recorded by the police. Our sample consists of 42 police forces operating in England and Wales after excluding the City of London, which in 2011 recorded a property crime rate 7.9 times larger than the average police force.[8]

In comparing police recorded crime rates with similar estimates derived from the CSEW it is important to keep in mind that the latter is not a 'gold standard', i.e. free of measurement error. Whilst this is a convenient assumption commonly employed in the literature (Gibson and Kim, 2008; Vollaard, 2012), victimisation surveys are themselves subject to multiple limitations, e.g. sampling error, recall errors, interviewer effects, and more (Lohr, 2019; Schneider, 1981). As such, discrepancies between the two crime rates should not be interpreted as perfect evidence of measurement error affecting police records. Still, since it is widely accepted that the CSEW provides a more accurate reflection of the underlying true extent of crime (ONS, 2022), we will use this as the benchmark measure against which police records are compared. Acknowledging both the superior measurement properties of the CSEW, without characterising it as a gold standard involves taking discrepancies between CSEW and police recorded crime rates as predominantly - but not entirely - evidence of measurement error in the latter. Put differently, we should take discrepancies between CSEW and police recorded crime rates as the upper bound estimate of the extent of the measurement error present in the latter.

To enhance the external validity of our findings, we undertake a second comparison, where we focus on American homicide rates (per 100,000 people) across states in 2019. Police recorded homicides are taken from the Uniform Crime Reporting (UCR) and compared to data from the National Center for Health Statistics (NCHS).[9] The two measures are largely in agreement since Coroners and medical examiners will normally collaborate with law enforcement in homicide cases (Regoeczi et al., 2014). However, as before, NCHS data should not be taken as a gold standard. Discrepancies can arise for multiple reasons, such as differences in the definitions of homicides subcategories, problems of misclassification potentially affecting both measures, or as a result of the NCHS recording the homicide in the county of residence of the victim rather than the location where the incident took place. Still, if not a gold standard, NCHS data is often considered a more accurate measure of homicide rates when used to reflect aggregate rates at higher spatial levels (Cantor & Cohen, 1980; Regoeczi et al., 2014). This is so mainly as a result of the different reporting practices underlying the two measures; whereas reports from the NCHS are compulsory those from the UCR only follow voluntary practices, which are associated with inconsistencies, delays, and an overall lower case prevalence at the National level (Regoeczi et al., 2014). To limit some of the effects associated with the voluntary nature of UCR, Florida and Alabama, the two states that in 2019 did not meet the UCR guidelines, were excluded from our analysis. This restricted our sample to 48 states.

Figure 2 presents scatterplots depicting the relationship between police recorded crime rates and those derived from our two benchmarks (the CSEW and the NCHS), for property crime and homicide rates. These are complemented with histograms showing the distribution of the discrepancies between those data sources when the errors are taken to be multiplicative. Inspecting these graphs, we can identify three key

---

unit commonly used in the literature (Abramovaite et al., 2019; Han et al., 2013; Machin & Meghir, 2004), encompassing 1.3 million people on average, which makes them similar to states and large counties in the US (Barnett, 1981; Philipson & Posner, 1996).

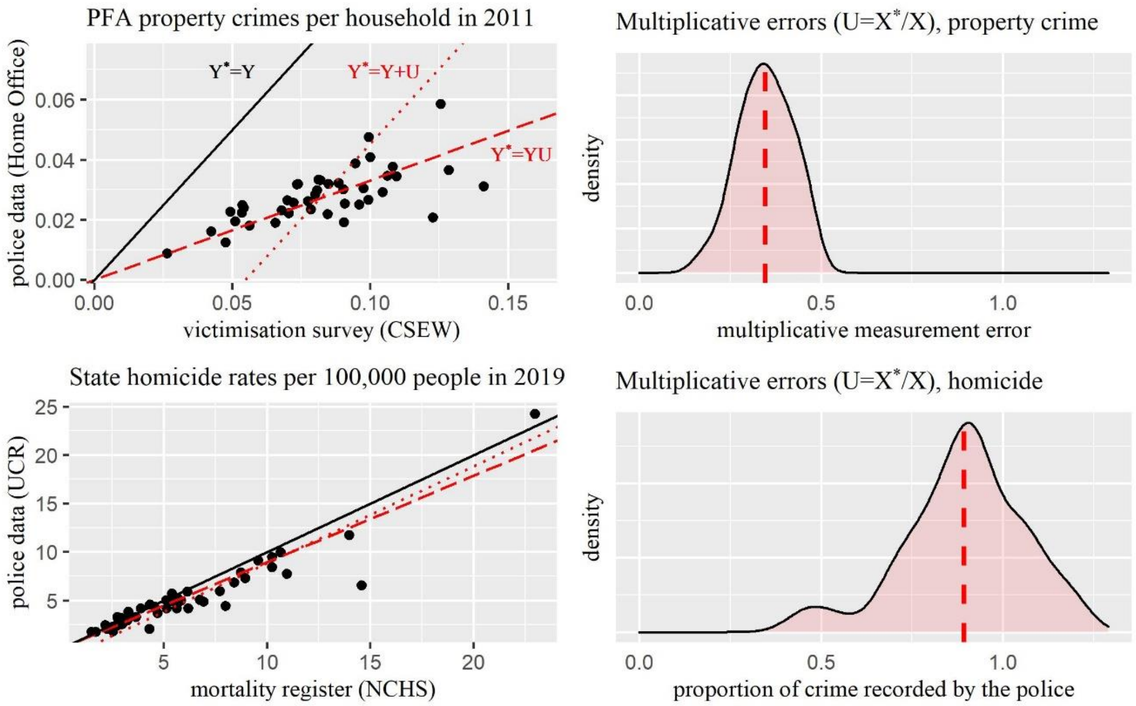
[7] Home Office data is available here: https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables.

[8] The City of London is primarily a business and financial centre with a small resident population of approximately 10,000 but a large day-time population leading to artificially high crime rates.

[9] UCR data is available here: https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/tables/table-20
NCHS data is available here: https://wonder.cdc.gov/controller/saved/D76/D99F056

properties defining the type of measurement error that could be expected to be present in police recorded crime rates.

Fig. 2. Comparison of property crime and homicide rates using data from the police, a victimisation survey (CSEW) and vital statistics (NCHS)



Systematic negative. The average crime rates recorded by the police is lower than those derived from the CSEW and the NCHS. Specifically, the estimated average recording rate is 34.7% for property crime, and 89.3% for homicides (as shown by the dashed vertical lines in the histograms).

Multiplicative. The size of the errors is proportional to the crime rates from the benchmark measures. This is shown in the scatterplots by the increasing divergence along the x-axis between the red dashed line of best fit summarising the relationship between $X^*$and X under a multiplicative model, and the continuous black line where X^* is assumed to be a perfect measure of X. Furthermore, to test whether the multiplicative model provides a better fit than the more commonly employed additive model, we have also included a visual representation of the latter, shown by the red dotted lines, and estimated the Akaike Information Criteria (AIC) for the two competing functional forms. We find that for both property crime and homicide the multiplicative models show a better fit (multiplicative AIC is -294.4 for property crimes, compared to -206.4 for the additive model, and 166.0 and 175.5 respectively for homicide).

Unreliable. As shown in the two histograms, discrepancies are not uniform but normally distributed. The standard deviation of the errors in property crime is 0.074, pointing at unequal reporting rates and/or recording practices across PFAs. The standard deviation of the errors in homicide rates is 0.17, which suggests even more substantial inconsistencies in recording practices across states in the US. However, as previously noted, those two standard deviations should be interpreted as the upper bound of the estimated variability attributed to measurement error in police data. Since neither the CSEW nor the NCHS represent gold standards, a significant share of that variability is likely stemming from measurement error affecting our benchmark measures, not police data.

3. ILLUSTRATING THE IMPACT OF MEASUREMENT ERROR IN POLICE RECORDED CRIME RATES FORMALLY

Various factors determine the extent of the impact on estimates from a regression model where one of the variables included is affected by measurement error: the specific form of the measurement error, its prevalence, the type of regression model employed, where and how the affected variable is introduced in that model, and the association between other variables included in the model with both the true value of the affected variable and its measurement error term. For clarity, and to constrain the number of scenarios to be considered, here we will invoke a few simplifying assumptions, and focus on the most common uses of police recorded crime rates.

We start by considering the impact of measurement error on a simple linear regression model, where the affected variable, X^*, is introduced as the only explanatory variable. We then move to consider the case of a multiple linear regression where a second explanatory variable is included, Z, which we take to be perfectly measured. We assume that the measurement error term is: homoscedastic, $var(U) = var(u_i)$, where u_i represents any particular value of U; independently distributed, $cov(u_i, u_j) = 0$; and non-differential, by which we mean unrelated to the response variable, $E(Y|X, X^*) = E(Y|X)$, and to any other variables included in the model, which in our case it is just Z, so $cov(U, Z) = 0$.

To ensure that the scenarios explored encompass most types of studies where police recorded crime rates are used in regression models, we also consider the impact of measurement error when crime rates are introduced as the response variable, Y^*. Scenarios presenting the systematic and random mechanisms identified in Section 2 are shown separately to distinguish their specific impact. We also consider additive and multiplicative errors separately to reflect the fact that recorded crime rates are not always introduced in their original scale but may first be log-transformed. Recall that under a multiplicative measurement error model: $\log(X^*) = \log(XU) = \log(X) + \log(U)$.

3.1. Crime Rate as an Explanatory Variable

Let us start with the case of a simple linear model where both response and explanatory variables are continuous, and the latter is affected by measurement error: $Y = \alpha + \beta X^* + \varepsilon$. Using ordinary least squares (OLS), the constant and slope of this model can be estimated by solving the following system of equations:

$$\begin{cases} \hat{\alpha}^* = \bar{Y} - \hat{\beta}\bar{X}^* \\ \hat{\beta}^* = \dfrac{S_{X^*,Y}}{S^2_{X^*}} \end{cases} \tag{1}$$

where $\bar{X}^* = E(X^*)$, $S^2_{X^*} = var(X^*)$, and $S_{X^*,Y} = cov(X^*, Y)$.

Consider first the impact of a purely systematic measurement error. Under the common assumption of additive measurement error, $X^* = X + u$, where under-recording takes the form of a scalar $u < 0$, then substituting X^* into the first line of Eq. (1) yields $\hat{\alpha}^* = \bar{Y} - (\hat{\beta}\bar{X} + \hat{\beta}u) = \hat{\alpha} + \hat{\beta}u$. The model's intercept will be biased downwards by $\hat{\beta}u$. The slope, however, will remain unbiased, as neither covariance nor variance are affected by a change of origin.[10] Since the substantive interest of regression models normally stems from the association between explanatory and response variables, we might conclude that the consequences of this type of systematic error are minimal.

However, in the presence of multiplicative systematic error, $X^* = Xu$, the picture is more problematic. In this case, the constant will continue to be biased by $\hat{\beta}u$ (although this time it will be an upward bias since $0 < u < 1$). More importantly, the slope will now be biased because both the variance and covariance are affected by a change of scale.[11] Substituting from the second line of Eq. (1) we have:

---

[10] Proof of the variance being unaffected by a change of origin:
$S^2_{X^*} = \frac{\Sigma(X^*-\bar{X}^*)^2}{n-1} = \frac{\Sigma(X+u-(\bar{X}+u))^2}{n-1} = \frac{\Sigma(X-\bar{X})^2}{n-1} = S^2_X$
Proof of the covariance being unaffected by a change of origin:
$S_{X^*,Y} = \frac{\Sigma(X^*-\bar{X}^*)\,(Y^*-\bar{Y}^*)}{n-1} = \frac{\Sigma(X+u-(\bar{X}+u))\,(Y^*-\bar{Y}^*)}{n-1} = \frac{\Sigma(X-\bar{X})\,(Y^*-\bar{Y}^*)}{n-1} = S_{X,Y}$

[11] Proof of the variance being affected by a change in scale:
$S^2_{X^*} = \frac{\Sigma(X^*-\bar{X}^*)^2}{n-1} = \frac{\Sigma(Xu-\bar{X}u)^2}{n-1} = \frac{u^2\,\Sigma(X-\bar{X})^2}{n-1} = u^2 S^2_X$
Proof of the covariance being affected by a change in scale:
$S_{X^*,Y} = \frac{\Sigma(X^*-\bar{X}^*)\,(Y^*-\bar{Y}^*)}{n-1} = \frac{\Sigma(Xu-(\bar{X}u))\,(Y^*-\bar{Y}^*)}{n-1} = \frac{\mathrm{u}\Sigma(X-\bar{X})\,(Y^*-\bar{Y}^*)}{n-1} = uS_{X,Y}$

$$\hat{\beta}^* = \frac{S_{X^*,Y}}{S^2_{X^*}} = \frac{S_{Xu,Y}}{S^2_{Xu}} = \frac{S_{X,Y}u}{S^2_X u^2} = \frac{\hat{\beta}}{u} \tag{2}$$

Thus, the slope is augmented by a factor proportional to the rate of under-recording.

Anticipating the specific impact on the slope becomes more complicated if we also consider that the observed measurement error affecting police recorded crime rates is not uniform but can vary randomly across areas. In that setting, U is a normally distributed variable, $U \sim N(\overline{U}, S^2_U)$. In the presence of additive errors, we will now observe an attenuation bias in the slope as a result of the random noise present in X^*. Specifically, under the assumption that U is non-differential (i.e. unrelated to X or Y), the covariance $S_{X^*,Y}$ will be equal to $S_{X,Y}$, but the variance $S^2_{X^*}$ will be the sum of the variance of X and the variance of U, $S^2_X + S^2_U$. Substituting the estimator of the slope in Eq. (1) we now have:

$$\hat{\beta}^* = \frac{S_{X^*,Y}}{S^2_{X^*}} = \frac{S_{X,Y}}{S^2_X + S^2_U} = \frac{S_{X,Y}}{S^2_X}\frac{S^2_X}{S^2_{X^*}} = \hat{\beta}\left(\frac{S^2_X}{S^2_X + S^2_U}\right) \tag{3}$$

The slope is attenuated by a factor equal to the proportion of signal to noise in X^* (i.e. its reliability ratio). The specific effect of the bias becomes harder to anticipate if the errors are multiplicative. In this case, under the assumption that X and Y are independent, the denominator of the bias term shown in Eq. (3) is now defined as $S^2_{X^*} = S^2_X S^2_U + S^2_X \overline{U}^2 + S^2_U \overline{X}^2$.

In fact, it is not just the slope of the variable prone to measurement error that will be affected, the bias will spread through the model impacting all the regression coefficients of any additional explanatory variables introduced in the model, even if these additional variables are measured perfectly. Carroll et al. (2006) show how for the simplest case of a multiple linear regression model, $Y = \alpha + \beta_1 X^* + \beta_2 Z + \varepsilon$, where X^* is subject to random additive errors, but Z is perfectly measured, regression coefficients for both variables are biased. OLS will not estimate $\hat{\beta}_1$ but rather,

$$\hat{\beta}^*_1 = \hat{\beta}_1 \frac{S^2_{X|Z}}{S^2_{X|Z} + S^2_U} \tag{4}$$

which differs from the bias observed in the slope of the simple linear model, Eq. (3), since $S^2_{X|Z}$ represents the residual variance of the regression of X on Z. Hence, the attenuation bias is now stronger than the case of simple linear regression, and the higher the correlation between the explanatory variables the stronger the bias. Importantly, we will also find that instead of $\hat{\beta}_2$ we obtain,

$$\hat{\beta}^*_2 = \hat{\beta}_2 + \hat{\beta}_1\left(1 - \frac{S^2_{X|Z}}{S^2_{X|Z} + S^2_U}\right)\gamma \tag{5}$$

where γ is the coefficient of Z in the regression of X on Z.

It is therefore clear that the impact of using variables affected by measurement error in multivariate models is not negligible, and in most cases it is hard to anticipate, becoming harder in line with the complexity of the measurement error mechanisms and the outcome model considered.

3.2. Crime Rate as a Response Variable

We proceed to consider the case where the variable prone to measurement error is the response variable, Y^*. As before, we assume that the measurement error term, U, is homoscedastic, independently distributed, and independent from the true value, Y, and any other variables included in the model. Let us consider a linear model with two perfectly measured explanatory variables, which takes the following form,

$$Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{6}$$

In this case, if the measurement error is additive, $Y^* = Y + U$, then substituting in Eq. (6) we have, $Y + U = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, which can be further rearranged as,

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + (\varepsilon - U) \tag{7}$$

Hence, random additive measurement errors affecting the response variable will be absorbed by the model's residuals, only affecting the precision of the model's estimates. If the errors are systematic then the intercept will be biased, but all other regression coefficients will remain unbiased. This changes when the errors are multiplicative, $Y^* = YU$. Substituting in Eq. (6) we have, $YU = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$, which can be rearranged as,

$$Y = \frac{\alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon}{U} \tag{8}$$

In this case, if the errors are completely random only the precision of the regression coefficients will be affected. However, in the presence of systematic errors all model estimates will also be biased. The extent of the bias will be proportional to $\bar{U}$, which for the case of under-recorded crime rates will represent a form of attenuation bias.

In sum, even when simple scenarios are considered, we see how the type of measurement errors observed in police recorded crime rates can impact the validity of estimates from regression models. Depending on the form and prevalence of the measurement error, the type of outcome model, where in the model the error-prone variable is introduced, and the ways in which the variables included in the model are correlated, we can see radically different effects. These effects range from relatively negligible (e.g. purely systematic measurement error additively associated to an explanatory variable will only bias models' intercepts) to potentially substantial (e.g. as shown in Eq. 8, systematic multiplicative measurement error affecting the response variable will bias all regression coefficients).

4. ILLUSTRATING THE IMPACT OF MEASUREMENT ERROR IN POLICE RECORDED CRIME RATES THROUGH SIMULATIONS

Importantly, it is not just the magnitude of the impact of measurement error that matters, but the extent to which that impact can be predicted. If the biasing effect of measurement error can be anticipated simply enough – as is the case, for example, in the scenario shown in Eq. (8), where all we need is an estimate of the under-recording rate in police data – then, findings based on police data can be adjusted. However, we have seen how the specific impact of measurement error is often hard to predict, with systematic and random errors leading to different types of biases, which can operate in different directions. The combination of these types of errors in varying degrees of intensity, as seen across the different crime types or area levels considered in the literature, makes it particularly difficult to anticipate their joint effect. To better understand the impact of the types of measurement error seen in police data across a wide range of settings we use computer simulations.

We simulate the varying forms of measurement error that could be expected to affect police recorded crime rates for different crime types across different area levels. However, to frame our analysis on real data, the impact of those errors is assessed on different models investigating the relationship between property crime rates and worry about crime, with perceptions of disorder included as control, and all the variables measured at the PFA level. Worry about crime and perceptions of disorder are area-level direct estimates (i.e., weighted means) of Confirmatory Factor Analysis (CFA) factor scores derived from the CSEW. The worry about crime measure combines items tapping into worry about burglary, robbery, rape, assault and receiving insults in public places (CFI = 0.97, TLI = 0.95, SRMR = 0.03). Perceived disorder covers perceptions of noisy neighbours and loud parties, teenagers hanging around on the streets, rubbish and litter lying around, vandalism and graffiti, people using or dealing drugs, people being drunk or rowdy in public places, and people being harassed or intimidated (CFI = 0.97, TLI = 0.96, SRMR = 0.03). In both cases, factor scores were linearly transformed to [0,1] range to enable an easier interpretation of results: $\frac{F_i - \min(F)}{\max(F) - \min(F)}$, where F_i is the factor score in respondent i. For property crime we take the CSEW estimates reported in Section 2.1. Table 2 shows descriptive statistics of the three variables. Their pairwise Pearson's correlation coefficients are as follows: $\rho_{Crime,Worry} = 0.69$; $\rho_{Crime,Disorder} = 0.62$, $\rho_{Worry,Disorder} = 0.68$).

Table II. Descriptive statistics of data from the CSEW used in our empirical illustration

	Mean	Median	Min.	Max.
Property crime rate	0.08	0.08	0.03	0.14
Worry about crime	0.51	0.41	0.41	0.61
Perceptions of disorder	0.27	0.27	0.20	0.37

We consider four linear models where we cross the position of the crime rates (response vs. explanatory variable) and their distribution (original scale vs. log-transformed). Table 3 presents the estimates for these models, which we refer to as the 'benchmark' models. As could be expected from their pairwise correlations, both worry about crime and perceptions of disorder are positively associated with property crime, and for the most part these associations are statistically significant.

Table III. Regression coefficients from the 'benchmark' models (based on CSEW estimated property crime rates at the PFA level)

	Coef.	SE	p-value
Response variable:			
Worry about crime			
Intercept	0.36	0.03	<0.001
Property crime	0.67	0.20	0.002
Perception of disorder	0.36	0.12	0.004
Response variable:			
Worry about crime			
Intercept	0.54	0.07	<0.001
Log-property crime	0.05	0.02	0.002
Perception of disorder	0.35	0.12	0.006
Response variable:			
Property crime			
Intercept	-0.13	0.04	0.001
Worry about crime	0.34	0.10	0.002
Perception of disorder	0.16	0.09	0.079
Response variable:			
Log-property crime			
Intercept	-5.44	0.52	<0.001
Worry about crime	4.35	1.33	0.002
Perception of disorder	2.46	1.19	0.045

To assess the impact of measurement error we compare estimates from the benchmark models presented in Table 3 against those obtained for the same models after property crime rates derived from the CSEW are subjected to different forms of simulated errors reflecting the types of errors affecting police recorded crime rates. For the models where crime is log-transformed the simulated errors are introduced in the crime rates before they are log-transformed. We focus on the impact on the regression coefficients of the two explanatory variables included in each model, and on their standard errors. This impact is quantified using the relative bias; the proportional difference between the observed estimate using crime rates where simulated error has been introduced, and the benchmark estimate, which for the case of regression coefficients can be expressed formally as, $RBIAS_{\hat{\beta}} = (\hat{\beta}^{*} - \hat{\beta})/\hat{\beta}$, or as, $RBIAS_{\widehat{SE}_{\hat{\beta}}} = (\widehat{SE}_{\hat{\beta}^{*}} - \widehat{SE}_{\hat{\beta}})/\widehat{SE}_{\hat{\beta}}$, for their standard errors.

We simulate a range of different multiplicative measurement error scenarios.[12] To reflect the varying levels of under-recording seen across different crime types (as shown in Table 1 and Figure 2), we consider the impact of under-recording rates ranging from no systematic errors in recording rates to up to 80% of crimes being missed. There are some specific crime types for which rates of under-recording may be expected to be even higher than 80%, such as anti-social behaviour or attempted theft (Appendix). However, the range considered here is likely to reflect most settings explored in the literature, including studies that focus on all recorded crimes (Cho & Park, 2017; Matsueda et al., 2006), broad categories of property or violent crime

[12] The R code used can be found here, https://osf.io/kv3sc/

(Abramovaite et al., 2019; McCarthy et al., 2019), or more specific crime types such as homicide, burglary, or motor vehicle theft (Philipson & Posner, 1996; Reisig & Parks, 2000).

We also explore the impact associated with varying levels of random errors. This helps us assess the extent to which the two measurement error mechanisms interact, while expanding the scope of our analysis. Random errors might be a result of variations in recording practices across police forces, as well as potential differences in average reporting propensities between areas. The magnitude of this random error could be expected to be proportional to the heterogeneity of the areas under consideration, which, as shown by Buil Gil et al. (2021a), is proportionally related to the spatial resolution considered (i.e. higher heterogeneity across smaller area units, such as output areas in the UK or census blocks in the US). To capture this potential heterogeneity we explore three scenarios: one where recording rates are assumed to be uniform; a second where we simulate *half* of the variability in the multiplicative errors detected in Section 2.1 for the case of property crime across PFAs in England and Wales ($sd = 0.037$); and a third scenario with *half* of the variability seen in homicides across states in the United States ($sd = 0.083$). We halve the estimated standard deviation to reflect the fact that we could not rely on a gold standard to accurately estimate the extent of measurement error in police records. Instead, we have opted to treat them as lower bound estimates derived from assuming both police records and our two benchmark measures (the CSEW and NCHS) are equally prone to random errors. Furthermore, these estimates are derived from comparisons of crime rates measured at the PFA and state levels, which represent some of the largest spatial units used in the literature. As such, when interpreting these scenarios, it is important to note that they are based on conservative estimates.

To ensure that no negative errors were simulated in the second and third scenario, the following additional constraint was imposed: $U > 0.001$. Moreover, to minimise the presence of simulation errors, the relative bias was estimated and averaged over 10,000 iterations.

4.1. Simulation-Derived Impact from Using Police Crime Rates as an Explanatory Variable

Figure 3 shows the impact of the simulated measurement error mechanisms seen in police crime rates used as an explanatory variable. If crime rates are introduced in their original scale, we observe a clear augmentation bias in the crime coefficient, β_1, that grows as the percentage of under-recorded cases increases. Yet the magnitude of this bias is substantially attenuated as random errors become more prevalent. This reflects the opposing effect that multiplicative systematic negative and random errors have when they are present in one of the explanatory variables (see Eq. 2 and 4). The impact on the standard errors generally mirrors that observed in the regression coefficients, although some discrepancies can be observed in instances of extreme under-recording and random error. Here, the bias in the standard error becomes larger than that of the regression coefficient.

The coefficient for perceptions of disorder, β_2, is also affected by measurement error in the recording of crime, although this impact is less severe. Furthermore, contrary to what we observed for β_1, we can now see how the two measurement error mechanisms operate in the same direction, rendering those settings characterised by both substantial average under-recording and considerable random error most problematic. However, the standard error remains largely unaffected.

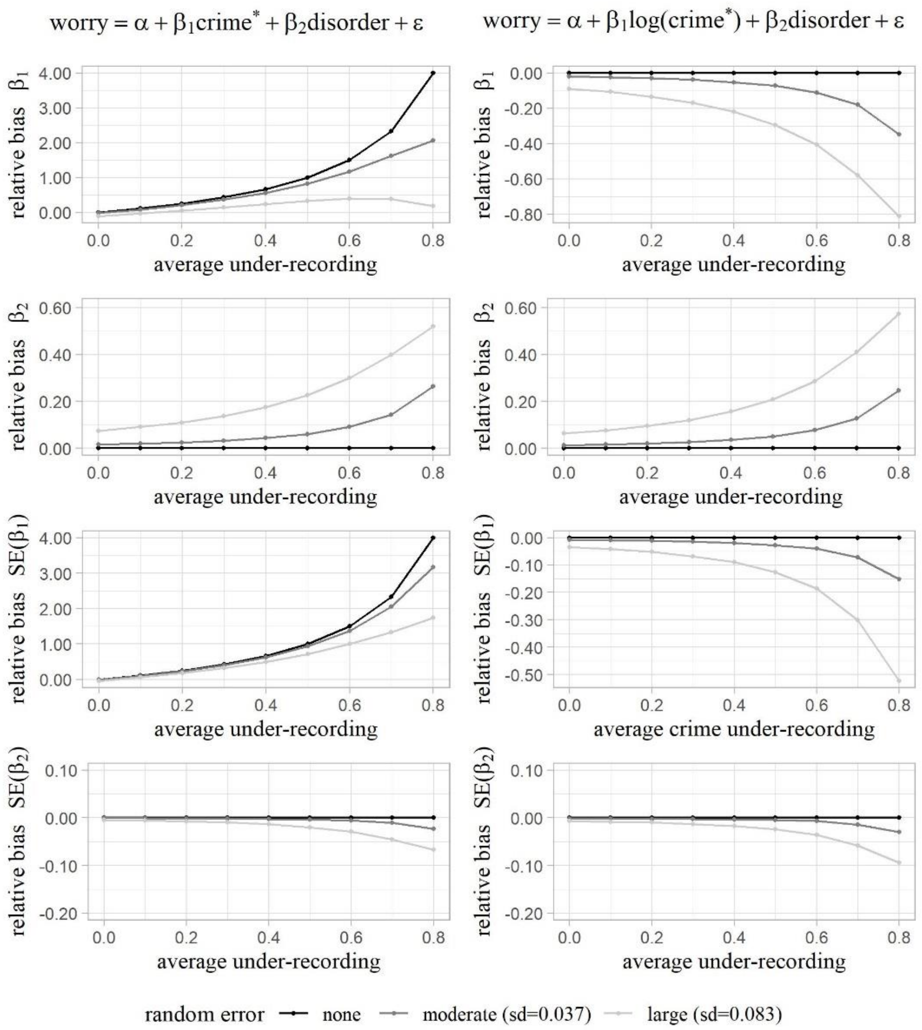
We can also observe some important differences when crime rates are introduced after being log-transformed and, thus, making the measurement error additive. The impact on the regression coefficient for crime and its standard error now takes the form of an attenuation bias, although it is somewhat less severe than was observed when crime rates were introduced on their original scale. This bias can be seen to result as a combination of the systematic and random mechanisms, operating in the same direction. By contrast, the impact on the regression coefficient for perceptions of disorder remains similar to the untransformed case, with its standard error also relatively unaffected.

In sum, we can anticipate that estimates of the effect of crime from studies where crime rates are introduced as an explanatory variable on their original scale may be severely inflated. The extent of the bias is directly proportional to the level of under-recording, but inversely related to the magnitude of the random error, to the point that in scenarios of large random error the bias can be practically cancelled out. Effects for crime types commonly used in the literature, such as violent or property crime, with recording rates near 40% (Table 1), measured at a spatial unit such as PFAs, may be expected to be overestimated by as much as

100%. Perhaps reassuringly, despite the large impact detected in the effect size of crime, the bias observed in its standard error seems similar enough to rule out a widespread problem of false positives.

When crime rates are log-transformed, both random and systematic errors operate in the same direction, leading to an attenuation bias in the effect of crime. In this case, the extent of the bias is less severe. Considering again crime types with recording rates around 40%, we can see the magnitude of the bias being relatively negligible (attenuating the true effect size by around 20%) unless recording variability across areas is as large as what could be attributed for the case of state homicide rates recorded by the UCR. The impact on the standard error follows a similar pattern, rendering the presence of false negatives relatively marginal.

Fig. 3. Impact of different measurement error mechanisms affecting police recorded crime rates when used as an explanatory variable (note change of y scale across graphs)



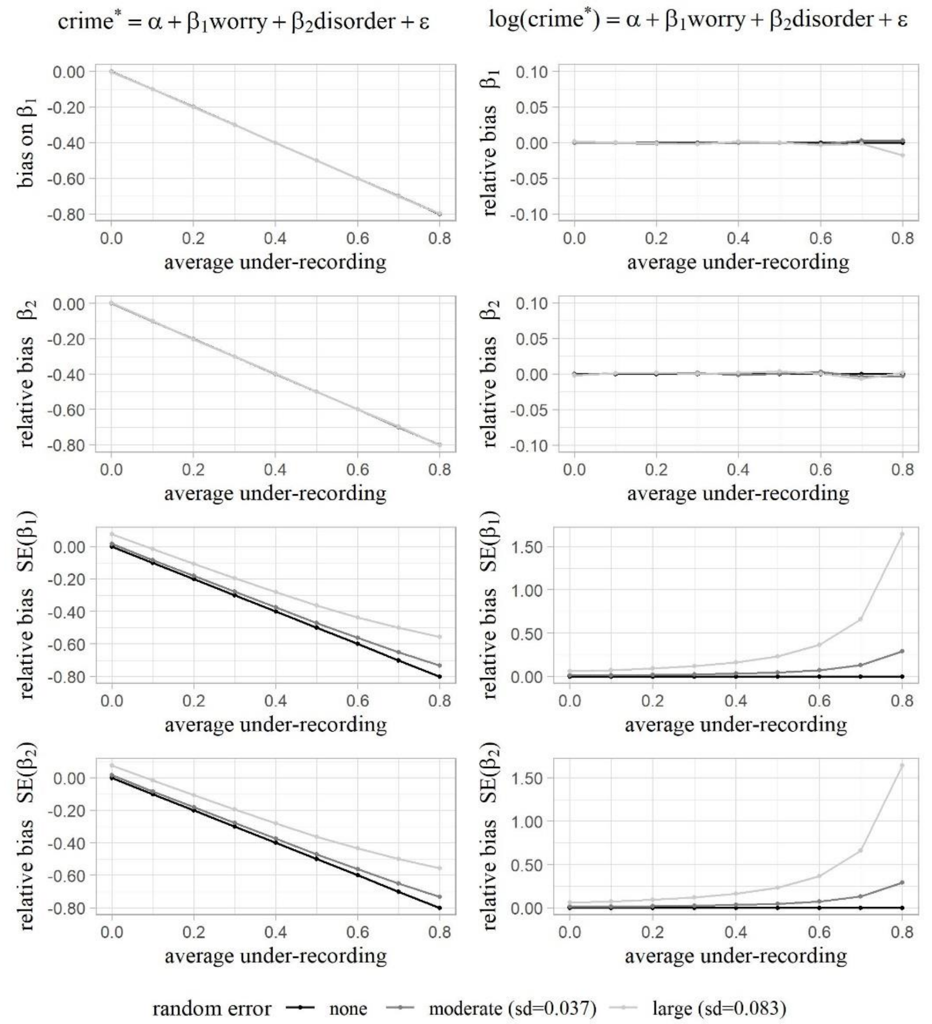
In addition, regardless of whether crime rates are log-transformed or not, we have observed that the bias does not just affect the regression coefficient for crime, but also spreads to the regression coefficients of other variables included in the model. We have only explored this effect for one variable, perceptions of disorder, which is positively correlated with property crime rates derived from the CSEW ($\rho = 0.62$). For the specific regression coefficient of perceptions of disorder, and considering again the type of measurement error seen in police recorded property crime (40% recording rate moderately varying across areas), we could expect an augmentation bias of at least 10%, regardless of whether crime is introduced in its original scale, or log-transformed. Importantly, in this case we could expect a similar impact even if we consider crime types such as homicides where average under-recording is negligible but highly variable across areas.

We have considered linear models with only two explanatory variables. However, similar effects should be expected on regression coefficients of any additional explanatory variables included in the outcome model: they will be biased as a result of the measurement error affecting police recorded crime rates, even if all of the additional variables are perfectly measured, and unrelated to the measurement error term. The specific impact on the regression coefficient of any other explanatory variable could be derived from Eq. (5). The magnitude of the bias will depend on the prevalence and variability of the under-recording, and on the associations between the true crime rate and the response variable and the explanatory variables considered, conditional on all other explanatory variables included in the model.

4.2. Simulation-Derived Impact from Using Police Crime Rates as the Response Variable

When considering models using crime rates as the response, we can observe a radically different impact depending on whether crime rates are log-transformed or not (Figure 4). If they are log-transformed, then, all regression coefficients remain unbiased, while their standard errors will only be affected by an augmentation bias when large systematic and random errors are present simultaneously. Such setting contrasts strikingly with the substantial attenuation biases observed across all regression coefficients and their standard errors when crime rates are specified in their original scale. As anticipated in Eq. (8), this type of bias is proportional to the rate of under-recording affecting the crime type considered.

Fig. 4. Impact of different measurement error mechanisms affecting police recorded crime rates when used as the response variable (note change of y scale across graphs)



The impact of measurement error is therefore much easier to anticipate when crime rates are used as the response variable. If log-transformed, the impact will be either null, or negligible in cases where the presence of measurement error is extreme, i.e. crime rates simultaneously affected by a large under-recording rate (over 80%) that is highly variable across areas. In similar cases where the two forms of measurement error are large, we should however anticipate an important loss of statistical power leading to the widespread presence of false negatives, a problem affecting all regression coefficients included in the model.

When crime rates are used as the response variable on their original scale the impact is substantial. All model's estimates are attenuated by a factor proportional to the under-recording rate of the crime type being modelled. This means that even in the presence of the more accurately recorded crime types, such as homicide, with a recording rate of roughly 90%, we should expect estimates to be attenuated by 10%. When considering other crime types commonly used in the literature, such as property or violent crimes, characterised by recording rates close to 40%, an attenuation bias of around 60% the size of the true estimates should be expected. On the positive side, these impacts can be easily anticipated, offering researchers the opportunity to evaluate the true effect size of the model's estimates by considering the potential under-recording rate affecting the crime type under analysis. Lastly, standard errors are affected by a similar form of attenuation bias, although in this case random errors push the bias in the opposite direction, which could lead to instances of false positives when the variability in recording or reporting practices across areas is high, as we saw was the case for homicides.

5. DISCUSSION

Crime analysis is becoming an increasingly sophisticated research subject. New data collection tools based on mobile apps are generating an ever wider range of crime related measures (Hughes et al., 2021; Solymosi et al., 2020), bottom-up approaches such as agent-based modelling techniques have provided new perspectives with which to test theories and explain outcomes arising from complex systems (Birks et al., 2012; Groff et al., 2019), whereas the growing adoption of directed acyclic graphs is leading to more transparent disclosures of the causal assumptions involved in traditional studies based on observational data (Young, 2014). Yet, in the midst of these remarkable advances, a central methodological problem lying at the core of the discipline remains paradoxically unattended. Despite their known flawed measurement properties, police statistics continue to be the most used data source in the study of the causes or consequences of crime, with very little done to explore their biasing effect.

Police statistics present important advantages in terms of versatility and accessibility, but relying on such data comes with important implications that are poorly understood. Presented with descriptive statistics or graphs comparing crime rates, researchers can speculate about the impact of common measurement error mechanisms affecting police data. For example, potential changes in reporting or detection rates are often invoked to explain questionable crime trends, while inconsistencies in recording practices can explain dubious spatial distributions. However, the specific impact that could be attributed to the type of measurement error seen in police data becomes much harder to trace when that data is used in regression models. This is a major problem. Countless studies are published each year introducing police recorded crime rates in regression models either as the response or as an explanatory variable. We suspect that estimates reported in those studies could be severely biased, but we do not know in which way, nor do we know how much so.

Amongst the few studies that have shed light on this question, the largest part has focused on the identification of the effect of measurement error in specific applications. That is, considering the measurement error present in specific crime types, which are then introduced in specific outcome models, to explore specific associations concentrated on a few topics such as gun ownership, economic inequality, or police arrests. Rather than focusing on one or a few applications, we have aimed to provide a more comprehensive overview of the impact that could be expected in the types of regression models where police recorded crime rates are commonly used to explore the causes and consequences of crime. We have considered: i) combinations of the under-recording (systematic errors) and recording inconsistencies (random errors) observed across different crime types; ii) using crime rates as the model's response variable, but also as an explanatory variable; and iii) the introduction of crime rates in their original scale and log-transformed.

Our findings demonstrate the pertinence of the more comprehensive overview of the impact of measurement error adopted here, as we show how the overall impact in model estimates is widely variable across different settings. Hence, results from previous explorations of the impact of measurement error seen in police data may not be generalisable beyond the specific settings considered in those studies. The size and direction of the biasing effects attributable to using error-prone police data in regression models is strongly dependent on the outcome model, crime type and spatial unit considered. With some studies being potentially unaffected, whereas in many others the biases could be expected to be larger than the true causal effects of interest, operate in opposing directions to those effects, or lead to over and under-estimated standard errors, each of those instances likely leading to wrong inferences. This new and more comprehensive overview of the impact of measurement error in models relying on police data does not just provide a wider understanding of the problem, it also offers a strategy to adjust for it. Once able to approximate the specific impact associated with the use of police data, researchers will be able to consider simple sensitivity analysis to communicate the expected impact in their estimates, which could be reported using uncertainty intervals. Similarly, being able to anticipate the impact associated to these errors means that researchers could assess the validity of findings from previous studies where no attempt was made to adjust for this problem.

As first identified by Gibson and Kim (2008), we show how studies where crime rates are introduced as a response variable in their original scale are likely to be severely biased. We expand on this and demonstrate how the impact of measurement error can also be severe in models where crime rates are used as an explanatory variable; as it is the case across studies aiming to estimate the causal effect of crime on a wide range of outcomes such as perceptions of insecurity (Cho & Park, 2017), residential segregation (Keels et al., 2005), or population change (Morenoff & Sampson, 1997). Importantly, substantial biases can also be found in other explanatory variables included in the model, even if they are perfectly measured. This finding points at a potentially more widespread - and ultimately more pervasive - effect than initially anticipated. We have focused our analysis on the validity of studies exploring the causes and consequences of crime, however, police recorded crime rates are also commonly used as controls in studies where the substantive interest lies on ascertaining different causal relationships (see, for example, Harmon, 2013; or Xie & Lauritsen, 2012). Those other causal estimates are also likely biased because of the measurement error present in police recorded crime rates.

Another important finding from our analysis challenges the view that crime types for which recording rates are high - such as homicides or vehicle theft - are 'safe' to use. When random variability in recording practices across areas is high, as it seems to be the case for state homicide rates derived from the UCR, model estimates relying on this data could also be biased, even if more mildly than for other crime types where recording rates are not as high. Once again, this finding underlines the relevance of adopting a broad perspective, considering the interacting effect of different measurement error mechanisms seen in police data.

5.1. Turning Multiplicative into Additive Errors

We have also shown that the potentially severe impact of the type of measurement error seen in police recorded crime rates can be considerably minimised - and in certain settings altogether eliminated – by applying a simple log-transformation. This is a transformation commonly undertaken in the analysis of crime rates for different reasons: i) to normalise their often right-skewed distributions; ii) as a result of the use of generalised linear models such as Poisson or negative-binomial; or iii) just to express the relationship between crime and other variables in relative terms. Here we showed how log-transformations have the added benefit of reducing the impact of measurement error by turning the more damaging multiplicative errors observed in police recorded crime rates into less harmful additive errors.

The only exceptions appear to be those instances where crime types affected by strong variability in recording rates across areas are used as an explanatory variable, in which case the impact of an additive measurement error might be more severe than that seen for multiplicative errors. We could expect that to be the case in two main settings: i) low frequency crime types, such as homicide, where small measurement inconsistencies could lead to relatively high variability across areas, and ii) when any other, more frequent crime types are considered but the spatial resolution at which they are measured is high, e.g. when considering crime rates at the street, neighbourhood or small output area level, rather than cities, states, or police force areas.

Leaving aside the case where crime rates are affected by strong variability across areas used as an explanatory variable, there are practically no disadvantages from adopting log-transformations. When crime rates are used as the outcome variable, we could also encounter potential model misspecifications if these crime rates in their original scale are already normally distributed rather than right-skewed, which could result in the model's residuals being not normally distributed. This is unlikely, but even in those instances the impact would be limited to the accuracy of the model's measures of uncertainty, whereas keeping the measurement error in its systematic multiplicative form will bias all regression coefficients included in the model proportionally to the rate of under-recorded crime.

Osgood (2000) urged researchers to abandon linear models and adopt Poisson-based models for the specification of crime rates. His advice is based on the more realistic parametric assumptions offered by Poisson or similar generalised models like negative binomial models. Here we echo Osgood's advice, not just to improve the specification of the response variable, but as a way to adjust the impact of measurement error. Using log-transformed crime rates as the response variable would almost invariably eliminate the strong attenuation bias affecting all regression coefficients as a result of the under-recording seen in police statistics. Further, we extend Osgood's advice to consider log-transforming crime rates that are used as explanatory variables, which, in those instances where recording inconsistencies across areas are not large, will contribute to mitigate the impact associated to multiplicative measurement errors.

To assess the extent to which Osgood's advice has been heeded, we undertook a rapid literature review of articles published in the Journal of Quantitative Criminology from 2001 to the 11th August 2021 containing the keyword 'UCR'. Out of a sample of exactly 100 articles that met the inclusion criteria, 40 used police recorded crime rates as the response variable in a regression model, with nine of them (22.5%) introducing crime rates in its original scale without relying on Poisson-based models. If we consider instances where crime rates are used as explanatory variables in a regression model, we counted sixteen articles, with thirteen of them (about 81%) introducing crime rates in their original scale. This demonstrates that the use of log-based generalised linear models or simple logarithmic transformations of crime rates before they are introduced in regression models is not uniformly adopted. There is therefore further scope for mitigating the impact of measurement error in crime analysis substantially by considering a simple data transformation.

5.2. Caveats and Future Avenues of Research

The precision with which we can interpret - and subsequently adjust - the impact of measurement error in police data hinges on how well we can estimate the prevalence and nature of those errors. We explored the presence of measurement error in police data by comparing police recorded crime rates with estimates from victimisation surveys. In addition, to consider other important crime types that are not captured by victimisation surveys, we have also compared state recorded homicide rates against rates derived from vital statistics.

The accuracy of such an approach hinges on how well crime rates derived from victimisation surveys and vital statistics are measured. For reasons of convenience, when similar comparisons are undertaken in the literature, they are often based on the assumption that the benchmark measures against which police recorded crime rates are compared can be considered a 'gold standard', free of measurement error. However, we know that both victimisation surveys and vital statistics are affected by different limitations, such as sampling error or the fact that both measures reflect the location of the victim's residence rather than where the incident took place (Cernat et al., 2021). As a result, we can deduce that the extent of measurement error attributed to police data in the literature has likely been exaggerated, but how much so is currently unclear. Here we opted to take a conservative estimate and only considered half of the variability in recording rates detected after comparing police crime rates with our two benchmarks as evidence of measurement error in the former, i.e. we have assumed that both police recorded crime rates and our two benchmark measures are equally affected by recording inconsistencies. This, together with our focus on crime rates measured at relatively high-level spatial units such as PFA and states, where a share of the inconsistencies prevalent at lower-level areas are cancelled out (Buil-Gil et al., 2021a), has probably led us to underestimate the impact that could be attributed to measurement error in regression models.

It is therefore essential that future studies explore the extent to which the different limitations of victimisation surveys and vital statistics, but also other proxy measures of crime such as medical emergency services data (Hibdon et al., 2017; Sutherland et al., 2021), may affect our estimates of measurement error in police statistics. Measurement error estimation methods that do not rely on a gold standard, such as multitrait-multimethod latent variable models (Oberski et al., 2017; Yang et al., 2018) or hidden Markov models (Pavlopuolos et al., 2020), offer a particularly promising avenue of research to do so. These can be used to estimate the validity and reliability of variables tapping on the same underlying concept. However, to the best of our knowledge, they have not yet been employed to study the problem of measurement error in crime data.

It would also be important that future studies exploring the measurement error affecting police data consider the presence of additional error mechanisms. We have illustrated the impact of systematic under-recording and random variability in recording rates across areas, since these are two general mechanisms that apply to all settings where police recorded crime rates are used. However, we have placed strong simplifying assumptions over those two measurement error mechanisms. One of those being that the errors are independent from the response variable and other explanatory variables included in the model. The extent to which this assumption holds across some of the most important variables that are used in the study of the causes and consequences of crime should be explored. Considering how other measurement error mechanisms interact with the more general processes seen here, should help enhance the precision with which the impact of measurement error can be estimated.

Lastly, we have illustrated the impact of measurement error in relatively simple linear models. It would be useful if future studies were to expand this by considering the impact on more complex models, such as discrete data models (Machin & Meghir, 2004; Sobol et al., 2013), or when systems of equations are employed (Krahn & Kennedy, 1985; Yesberg et al., 2021). In those instances, the biasing effect of measurement error, and how it is propagated through the different parts of the model, will be harder to trace (Carroll et al., 2006). Hence, findings will likely be even less generalisable than we have seen here. In the absence of a general understanding of the potential impact exerted by measurement error, it would be key that researchers attempt to assess that impact empirically, as a form of sensitivity analysis. Regardless of the complexity of the outcome model, this could be done using simulations, just as we have done in this study (the R code employed has been included in the supplementary material). Other flexible methods that could be used as sensitivity analysis tools are simulation-extrapolation (Biewen et al., 2008; Pina-Sánchez, 2016), multiple over-imputation (Blackwell et al., 2017), or Bayesian adjustments (Gustafson, 2003; Pina-Sánchez et al., 2019).

6. CONCLUSION

We urge researchers introducing police recorded crime rates – or any other kind of police recorded crime data - in regression models, to consider how their estimates might be impacted by the presence of measurement error. We have shown how the impact can be large enough to lead to diametrically incorrect conclusions. Yet, we have also shown how, based on an understanding of the validity and reliability of police records, and on how and where they are introduced in the model, that impact can be approximated, and therefore - to some extent - adjusted. Here, we summarise the impact that should be expected across different settings in five simple general principles, which ought to be considered in revisiting findings from the literature under a more accurate and critical perspective, and to help minimise the problem of measurement error in police recorded crime rates in the future.

i. Studies using linear models with police recorded crime rates as the response variable will be biased. All regression coefficients and their standard errors are attenuated in a proportion similar to the extent of the under-recording of the crime type explored.
ii. That attenuation bias is often eliminated when crime rates are log-transformed, rendering such transformations essential in future studies, regardless of whether crime rates are normally distributed in their original scale or not.
iii. Studies including police recorded crime rates in their original scale as an explanatory variable should expect the effect of crime to be biased. The direction of the bias will depend on the dominating measurement error mechanism, an augmentation bias will arise proportionally to the under-recording affecting the crime type considered, but this will be opposed by an attenuation bias directly related to the variability in recording rates across areas.
iv. If crime rates introduced as an explanatory variable are log-transformed, we will instead observe an attenuation bias in their coefficient and standard error. This bias is proportional to both the average under-recording and the recording variability across areas. The magnitude of this bias could be expected to be, in most cases, smaller than if crime rates were introduced in their original scale.

v. Regression coefficients for other explanatory variables included in the model alongside crime rates will also be biased, even if those explanatory variables are perfectly measured, and unrelated to the measurement error term. The direction of the bias will depend on the sign of the relationship between these explanatory variables and crime, and that of crime and the response variable, conditional on all other explanatory variables included in the model, which makes it hard to anticipate. This means that measurement error in police recorded crime rates does not only affect studies exploring the causes or consequences of crime, but studies where police recorded crime rates are used as controls are also affected.

REFERENCES

Abramovaite, J., Bandyopadhyay, S., Bhattacharya, S., and Cowen, N. (2019). Alternatives to custody: Evidence from police force areas in England and Wales. The British Journal of Criminology. 59(4): 800--822.

Ashby, M. P. (2020). Changes in police calls for service during the early months of the 2020 coronavirus pandemic. Policing: A Journal of Policy and Practice. 14(4): 1054--1072.

Barnett, A. (1981). The deterrent effect of capital punishment: A test of some recent studies. Operations Research. 29(2): 346--370.

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 Jr, A. J. (1967). On exploring the 'dark figure' of crime. The Annals of the American Academy of Political & Social Science. 374(1): 1--15.

Biewen, E., Nolte, S., and Rosemann, M. (2008). Perturbation by multiplicative noise and the Simulation Extrapolation method. Advances in Statistical Analysis. 92(4): 375--389.

Birks, D., Townsley, M., and Stewart, A. (2012). Generative explanations of crime: using simulation to test criminological theory. Criminology. 50(1): 221--254.

Blackwell, M., Honaker, J., and King, G. (2017). A unified approach to measurement error and missing data: overview and applications. Sociological Methods & Research. 46(3): 303--341.

Boivin, R., and Cordeau, G. (2011). Measuring the impact of police discretion on official crime statistics: A research note. Police Quarterly. 14(2): 186--203.

Bottoms, A. E., Mawby, R. I., and Walker, M. A. (1987). A localised crime survey in contrasting areas of a city. British Journal of Criminology. 23(2): 125--154.

Bowers, K. J., and Johnson, S. D. (2005). Domestic burglary repeats and space-time clusters: The dimensions of risk. European Journal of Criminology. 2(1): 67--92.

Brantingham, P. (2018). The logic of data bias and its impact on place-based predictive policing. Ohio State Journal of Criminal Law. 15(2): 473--486.

Browning, C. R., Cagney, K. A., and Iveniuk, J. (2012). Neighborhood stressors and cardiovascular health: Crime and C-reactive protein in Dallas, USA. Social Science & Medicine. 75(7): 1271--1279.

Buil-Gil, D., Moretti, A., and Langton, S. H. (2021a). The accuracy of crime statistics: Assessing the impact of police data bias on geographic crime analysis. Journal of Experimental Criminology.

Buil-Gil, D., Medina, J., & Shlomo, N. (2021b). Measuring the Dark Figure of Crime in Geographic Areas: Small Area Estimation from the Crime Survey for England and Wales. The British Journal of Criminology. 61(2): 364--388.

Burrows, J., Tarling, R., and Mackie, A. (2000). Review of police force's crime recording practices. Home Office Research Study.

Cantor, D., and Cohen, L. E. (1980). Comparing measures of homicide trends: Methodological and substantive differences in the vital statistics and Uniform Crime Report time series (1933-1975). Social Science Research. 9: 121--145.

Cantor, D., and Land, K. C. (1985). Unemployment and crime rates in the post-World War II United States: A theoretical and empirical analysis. American Sociological Review. 317--332.

Caplan, J. M., Kennedy, L. W., and Miller, J. (2011). Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. Justice Quarterly. 28(2): 360--381.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). Measurement Error in Nonlinear Models: A Modern Perspective, CRC press, Boca Raton.

Cernat, A., Buil-Gil, D., Brunton-Smith, I., Pina-Sánchez, J., and Murrià-Sangenís, M. (2021). Estimating crime in place: Moving beyond residence location. Crime & Delinquency

Cho, J. T., and Park, J. (2017). Exploring the effects of CCTV upon fear of crime: A multi-level approach in Seoul. International Journal of Law, Crime and Justice. 49: 35--45.

Cohen, L. E., and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. American Sociological Review. 588--608.

Coleman, C., and Moynihan, J. (1996). Understanding Crime Data: Haunted by the Dark Figure, Open University Press, Buckingham.

Duncan, T. E., Duncan, S. C., Okut, H., Strycker, L. A., and Hix-Small, H. (2003). A multilevel contextual model of neighborhood collective efficacy. American Journal of Community Psychology. 32(3-4): 245--252.

Ellis, L., Farrington, D. P., and Hoskin, A. W. (2019). Handbook of Crime Correlates, Academic Press, London.

Eterno, J. A., Verma, A., & Silverman, E. B. (2016). Police manipulations of crime reporting: Insiders' revelations. Justice Quarterly, 33(5): 811--835.

Fajnzylber, P., Lederman, D., and Loayza, N. (2002). Inequality and violent crime. The Journal of Law and Economics. 45(1): 1--39.

Farrell, G., and Pease, K. (2001). Measuring and interpreting repeat victimization using police data: An analysis of burglary data and policy for Charlotte, North Carolina. Crime Prevention Studies. 16: 265--289.

Fuller, W. A. (2009). Measurement error models, John Wiley & Sons, New York.

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(2): 247--254.

Glewwe, P. (2007). Measurement Error Bias in Estimates of Income and Income Growth among the Poor: Analytical Results and a Correction Formula. Economic Development and Cultural Change. 56: 163--189.

Goulas, E., and Zervoyianni, A. (2013). Economic growth and crime: Does uncertainty matter? Applied Economics Letters. 20(5): 420--427.

Greenberg, D. F. (2001). Time series analysis of crime rates. Journal of Quantitative Criminology. 17(4): 291--327.

Groff, E. R., Johnson, S. D., and Thornton, A. (2019). State of the art in agent-based modeling of urban crime: an overview. Journal of Quantitative Criminology. 35(1), 155--193.

Gustafson, P. (2003). Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments, CRC Press, Boca Raton.

Han, L., Bandyopadhyay, S., and Bhattacharya, S. (2013). Determinants of violent and property crimes in England and Wales: A panel data analysis. Applied Economics. 45(34): 4820--4830.

Harmon, M. G. (2013). "Fixed" sentencing: The effect on imprisonment rates over time. Journal of Quantitative Criminology. 29(3): 369--397.

Hart, T. C., and Rennison, C. (2003). Reporting crime to the police, 1992-2000. Special Report. Bureau of Justice Statistics. Available at: https://static.prisonpolicy.org/scans/bjs/rcp00.pdf

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 at: https://www.justiceinspectorates.gov.uk/hmicfrs/wp-content/uploads/crime-recording-making-the-victim-count.pdf

Hibdon, J., Telep, C. W., and Groff, E. R. (2017). The concentration and stability of drug activity in Seattle, Washington using police and emergency medical services data. Journal of Quantitative Criminology. 33(3): 497--517.

Hughes, L. A., Schaible, L. M., and Kephart, T. (2021). Gang graffiti, group process, and gang violence. Journal of Quantitative Criminology.

Jackson, J., Bradford, B., Stanko, B. and Hohl, K. (2013). Just Authority? Trust in the Police in England and Wales, Routledge, Abingdon.

Keels, M., Duncan, G. J., DeLuca, S., Mendenhall, R., and Rosenbaum, J. (2005). Fifteen years later: Can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty. Demography. 42(1): 51--73.

Klinger, D. A., and Bridges, G. S. (1997). Measurement error in calls-for-services as an indicator of crime. Criminology. 35(4): 705--726.

Krahn, H., and Kennedy, L. W. (1985). Producing personal safety: The effects of crime rates, police force size, and fear of crime. Criminology. 23(4): 697--710.

Laufs, J., Bowers, K., Birks, D., and Johnson, S. D. (2020). Understanding the concept of 'demand' in policing: a scoping review and resulting implications for demand management. Policing and society.

Lee, M. R., and Ousey, G. C. (2005). Institutional access, residential segregation, and urban black homicide. Sociological Inquiry. 75(1): 31--54.

Levitt, S. D. (1998). Why do increased arrest rates appear to reduce crime: deterrence, incapacitation, or measurement error? Economic Inquiry. 36(3): 353--372.

Lohr, S. H. (2019). Measuring crime: Behind the statistics, CRC Press, Boca Raton.

Lynch, J. P., and Addington, L. A. (2006). Understanding Crime Statistics: Revisiting the Divergence of the NCVS and the UCR, Cambridge University Press, Cambridge.

Lynch, J. P., and Jarvis, J. P. (2008). Missing data and imputation in the Uniform Crime Reports and the effects on national estimates. Journal of Contemporary Criminal Justice. 24(1): 69--85.

Machin, S., and Meghir, C. (2004). Crime and economic incentives. Journal of Human Resources. 39(4): 958--979.

Maltz, M. D., and Targonski, J. (2002). A note on the use of country-level UCR data. Journal of Quantitative Criminology. 18(3): 297--318.

Matsueda, R. L., Kreager, D. A., and Huizinga, D. (2006). Deterring delinquents: A rational choice model of theft and violence. American Sociological Review. 71(1): 95--122.

McCandless, R., Feist, A., Allan, J. and Morgan, N. (2016). Do initiatives involving substantial increases in stop and search reduce crime? Assessing the impact of Operation BLUNT 2. Home Office Report. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/508661/stop-search-operation-blunt-2.pdf

McCarthy, M., Porter, L., Townsley, M., and Alpert, G. (2019). Influence of community characteristics on serious police use of force events in an Australian policing jurisdiction: a test of minority threat, social disorganisation, and ecological contamination theories. Policing & Society. 29(9): 1091--1108.

Messer, L. C., Kaufman, J. S., Dole, N., Savitz, D. A., and Laraia, B. A. (2006). Neighborhood crime, deprivation, and preterm birth. Annals of Epidemiology. 16(6): 455--462.

Miethe, T. D., Hughes, M., and McDowall, D. (1991). Social change and crime rates: An evaluation of alternative theoretical approaches. Social Forces. 70(1): 165--185.

Morenoff, J. D., and Sampson, R. J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990. Social Forces. 76(1): 31--64.

Neumayer, E. (2005). Inequality and violent crime: Evidence from data on robbery and violent theft. Journal of Peace Research. 42(1): 101--112.

Novick, M. R. (1966). The axioms and principal results of classical test theory. Journal of Mathematical Psychology. 3(1): 1--18.

Nugent, W., Graycheck, L., and Basham, R. (2000). A Devil Hidden in the Details: The Effects of Measurement Error in Regression Analysis. Journal of Social Service Research. 27: 53--75

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(520): 1477--1489.

ONS (2015). User guide to crime statistics for England and Wales. Office for National Statistics.

ONS (2022). Crime in England and Wales: year ending September 2021. Office for National Statistics.

Osborn, D. R., and Tseloni, A. (1998). The distribution of household property crimes. Journal of Quantitative Criminology. 14(3): 307--330.

Osgood, D. W. (2000). Poisson-Based Regression Analysis of Aggregate Crime Rates. Journal of Quantitative Criminology. 16: 21--43.

Pavlopoulos, D., Pankowska, P. K., Bakker, B. F., and Oberski, D. (2020). Modelling error dependence in categorical longitudinal data. In Cernat, A., Sakshaug, J. (eds.), Measurement Error in Longitudinal Data, Oxford University Press.

Pepper, J., Petrie, C., and Sullivan, S. (2010). Measurement error in criminal justice data. In Piquero, A. R., Weisburd, D. (eds.), Handbook of Quantitative Criminology, Springer.

Philipson, T. J., and Posner, R. A. (1996). The economic epidemiology of crime. The Journal of Law & Economics. 39(2): 405--433.

Pickles, A., Pickering, K., and Taylor, C. (1996). Reconciling Recalled Dates of Developmental Milestones, Events and Transitions: A Mixed Generalized Linear Model with Random Mean and Variance Functions. Journal of the Royal Statistical Society: Series A. 225--234.

Pina-Sánchez, J. (2016). Adjustment of recall errors in duration data using SIMEX. Metodološki zvezki-Advances in Methodology & Statistics. 13(1): 27--58.

Pina-Sánchez, J., Koskinen, J., and Plewis, I. (2019). Adjusting for measurement error in retrospectively reported work histories: An analysis using Swedish register data. Journal of Official Statistics. 35(1): 203--229.

Regoeczi, W. C., Banks, D., Planty, M., Langton, L., and Warner, M. (2014). The nation's two measures of homicide. Program Report, Bureau of Justice Statistics, U.S. Department of Justice.

Reisig, M. D., and Parks, R. B. (2000). Experience, quality of life, and neighborhood context: A hierarchical analysis of satisfaction with police. Justice Quarterly. 17(3): 607--630.

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 (pp. 285--314). Monsey: Criminal Justice Press.

Rotton, J., and Cohn, E. G. (2003). Global warming and US crime rates: An application of routine activity theory. Environment & Behavior. 35(6): 802--825.

Sampson, R. J., Raudenbush, S. W., and Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. Science. 277(5328): 918--924.

Schneider, A. L. (1981). Methodological problems in victim surveys and their implications for research in victimology. Journal of Criminal Law and Criminology. 72(2): 818--838.

Sellin, T. (1931). The basis of a crime index. Journal of Criminal Law & Criminology. 22(3): 335--356.

Skinner, C., and Humphreys, K. (1999). Weibull Regression for Lifetimes Measured with Error. Lifetime Data Analysis. 5: 23--37.

Skogan, W. G. (1977). Dimensions of the dark figure of unreported crime. Crime & Delinquency. 23(1): 41--50.

Sobol, J. J., Wu, Y., and Sun, I. Y. (2013). Neighborhood context and police vigor: A multilevel analysis. Crime & Delinquency. 59(3): 344--368.

Solymosi, R., Buil-Gil, D., Vozmediano, L., and Guedes, I. S. (2020). Towards a place-based measure of fear of crime: A systematic review of app-based and crowdsourcing approaches. Environment and Behavior.

Stefanski, L. A., and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. Journal of the American Statistical Association. 90(432): 1247--1256.

Sutherland, A., Strang, L., Stepanek, M., Giacomantonio, C., Boyle, A., and Strang, H. (2021). Tracking violent crime with ambulance data: How much crime goes uncounted? Cambridge Journal of Evidence-Based Policing.

Sutherland, A., Brunton-Smith, I., and Jackson, J. (2013). Collective efficacy, deprivation and violence in London. British Journal of Criminology. 53(6): 1050--1074.

Tarling, R., and Morris, K. (2010). Reporting crime to the police. The British Journal of Criminology. 50(3): 474--490.

Vandeviver, C., Neutens, T., Van Daele, S., Geurts, D., and Vander Beken, T. (2015). A discrete spatial choice model of burglary target selection at the house-level. Applied Geography. 64: 24--34.

van de Weijer, S. G. A., Leukfeldt, R., and Bernasco, W. (2019). Determinants of reporting cybercrime: A comparison between identity theft, consumer fraud, and hacking. European Journal of Criminology. 16(4): 486--508.

Vollaard, B., and Hamed, J. (2012). Why the police have an effect on violent crime after all: evidence from the British Crime Survey. The Journal of Law & Economics. 55(4): 901--924.

von Hofer, H. (2000). Crime statistics as constructs: The case of Swedish rape statistics. European Journal on Criminal Policy & Research. 8: 77--89.

Whitworth, A. (2012). Inequality and crime across England: A multilevel modelling approach. Social Policy & Society. 1(14): 48--68.

Witt, R., and Witte, A. (2000). Crime, prison, and female labor supply. Journal of Quantitative Criminology. 16(1): 69--85.

Xie, M., and Baumer, E. P. (2019). Neighborhood immigrant concentration and violent crime reporting to the police: A multilevel analysis of data from the National Crime Victimization Survey. Criminology. 57(2): 237--267.

Xie, M., & Lauritsen, J. L. (2012). Racial context and crime reporting: A test of Black's stratification hypothesis. Journal of Quantitative Criminology. 28(2): 265--293.

Yang, S. M., Hinkle, J. C., and Wyckoff, L. A. (2018). Using multitrait-multimethod (MTMM) techniques to examine the convergent and discriminant validity of social disorder. Journal of Research in Crime & Delinquency. 55(5): 571--608.

Yesberg, J., Brunton-Smith, I., and Bradford, B. (preprint). Police visibility, trust in police fairness, and collective efficacy: A multilevel Structural Equation Model. Available at: https://osf.io/preprints/socarxiv/k7ren/

Young, J. T. (2014). A sensitivity analysis of egocentric measures of peer delinquency to latent homophily: a research note. Journal of Quantitative Criminology. 30(3): 373--387.

Zhao, J. S., Lawton, B., and Longmire, D. (2015). An examination of the micro-level crime–fear of crime link. Crime & Delinquency. 61(1): 19--44.

APPENDIX. SPECIFIC OFFENCES USED TO DEFINE BROADER CRIME TYPES IN TABLE 1

CSEW 2018 to 2019			NCVS 2017 to 2020*		
Crime type	Cases reported in the interview	% known to police (weighted)	Crime type	Cases reported in the interview	% known to police (weighted)
Violent crime	1979	38.8 %	**Violent crime**	516	46.6 %
Hit with fists or weapon	538	46.6 %	Assault	200	49.5 %
Threaten to use force or violence on you	1319	36.4 %	Attempted assault	299	44.7 %
Sexually assaulted	95	28.2 %	Rape	8	**
Violent from household member	37	36.5 %	Unwanted sexual contact from household member	9	**
Property crime	2035	36.7 %	**Property crime**	995	41.8 %
Something stolen out of hands or pockets	304	46.2 %	Larceny	927	40.7 %
Other theft	360	24.8 %			
Tried to steal	203	11.7 %	Attempt larceny	52	53.5 %
			Robbery	16	59.6 %
Something stolen off car	796	40.0 %			
Bike theft	372	46.2 %			
Burglary	719	59.5 %	**Burglary**	248	45.5 %
Get in previous house to steal	38	69.0 %	Burglary	194	45.1 %
Get in previous house and cause damage	10	79.3 %			
Get in house since moved in to steal	8	**			
Get in current house to steal	250	75.7 %			
Get in current house and cause damage	37	70.3 %			
Try to get in previous house to steal/damage	21	15.4 %	Attempted burglary		
Try to get in current house to steal/damage	355	48.0 %		54	47.5 %
Motor vehicle theft	130	89.7 %	**Motor vehicle theft**	33	73.5 %

*Estimates from the NCVS are derived from a wider timeframe to obtain a larger sample size.

**Crime types with samples smaller than 10 are only used to calculate the overall proportion of cases known to the police, not to calculate their crime specific proportion.