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

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

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

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: Most coefficients and measures of uncertainty from models where crime rates are included in their original scale are severely biased. However, in many cases, this problem could be minimised, or altogether eliminated by log-transforming crime rates. This transforms the multiplicative measurement error observed in police recorded crime rates into a less harmful additive mechanism.

Conclusions: The validity of findings from regression models where police recorded crime rates are used in their original scale is put into question. In interpreting the large evidence base exploring the effects and consequences of crime using police statistics we urge researchers to consider the biasing effects shown here. Equally, we urge researchers to log-transform crime rates before they are introduced in statistical models.

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

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

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 still 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 theory, Cohen & Felson, 1979, Matsueda et al., 2006). Police data is also central in studies that, from a more exploratory perspective, seek to identify the 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 relationships between crime, socio-economic inequality, deprivation, and ethnic heterogeneity have been of special interest.

However, with some notable exceptions (Barnett, 1981; Brantingham, 2018; Gibson & Kim, 2008; Levitt, 1998; Mohler et al., 2019; Vollaard & Hamed, 2012), researchers have generally failed to sufficiently recognise the implications of using measurement error prone police data 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, it could be expected that the impact in regression models relying on this data will be substantial. Importantly, the magnitude and direction of those biases can be difficult to anticipate. The measurement error impact will likely propagate through the model, affecting the accuracy of not just the estimates of crime variables, but all model estimates and their respective measures of uncertainty (Nugent et al., 2000). Furthermore, given the complexity of the errors (stemming from multiple interacting mechanisms), the direction of the biases in model estimates remains unclear. It is therefore no exaggeration to suggest that police statistics represent both one of the most important data sources and biggest methodological challenges in the study of the causes and consequences of crime.

Gibson and Kim (2008) published the most complete study of the effects of measurement error on crime data to date. The authors provided a general framework to assess the potential effect of both the random and systematic errors observed in police recorded crime rates, and used it to demonstrate the existence of severe biases in studies exploring the relationship between economic inequality and crime. However, although ground-breaking in formally defining the measurement error mechanisms present in police recorded crime rates, Gibson and Kim (2008) is limited in two important ways. First, they did not consider the impact of measurement error when crime rates are used as an explanatory variable, which results in different forms of bias. Secondly, they failed to reflect on how the observed measurement error mechanisms in police recorded crime rates can be substantially simplified by applying log-transformations.

In this article, we provide a general overview of the biasing effect that measurement error in police recorded crime data exerts across the regression models typically used to explore the causes and consequences of crime. Besides raising awareness about the problem, we also offer a set of general principles that can be used to interpret the validity of previous studies relying on police recorded crime rates and minimise the impact of measurement error in future studies. For simplicity, we focus on the most widely used form of police data in exiting literature, crime rates recorded across areas at a given point in time (where crime rate has a lower bound at zero and is often right-skewed). We therefore set aside other uses of police data which may be equally prone to measurement error, albeit taking a different form. This includes problems of misclassification that affect police data when measured as a binary outcome (Caplan et al., 2011; Vandeviver et al., 2015), or the effect of measurement error in time-series analysis of police statistics (Cantor & Land, 1985; Greenberg, 2001).

The paper proceeds as follows. Section 2 describes the form and prevalence of the different measurement error mechanisms affecting police crime rates. This is explored through a brief review of the literature, followed by a comparison of annual property crime rates at the police force area level using UK police recorded crime data (recorded by the Home Office) and the Crime Survey for England and Wales (CSEW). The impact exerted by the measurement error mechanisms observed in police recorded crime rates when such data is used in a regression model is set out algebraically in Section 3, and explored empirically through simulations in Section 4. The former helps us define the specific impact that could be attributed to different measurement error mechanisms, while the latter helps us visualise their combined impact across a wide range of scenarios. Section 5 provides a discussion and summary of the implications of our findings, distilled into a few simple rules of thumb.

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

The accuracy and precision of official crime statistics has long been the subject of criticism, with de Candolle (1987 [1830]) highlighting the various forms of error present in crime data as the first maps of persons sentenced in France were being produced almost two hundred years ago. Some of the errors observed by the author related to: victims being unaware they have suffered a crime, deciding not to report it to the police, authorities failing to find out who is responsible, to arrest the known offender, or legal procedures failing to convict arrested offenders due to lack of evidence. Police recorded crime data is closer in distance to crime events than judiciary statistics, yet we can still identify several of those same sources of measurement error affecting the accuracy of police recorded crime rates (Skogan, 1977).

Most notably, police recorded crime rates are affected by the 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). In particular, crime reporting rates differ systematically based on victims' sex (females report more often than males) and 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 differences in reporting rates by crime types, with theft of motor vehicle and burglary typically being those with the largest 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). Much of these differences may be related to the financial, physical and emotional harms of these crimes, but they are also related to the need to report to the police before filing an insurance claim in the case of property crimes that result in significant financial losses. Reporting rates may also differ across geographic areas (Buil-Gil et al., 2021; 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). Counts of other 'victim-less' crimes may be less susceptible to these differences individual reporting tendencies, but will instead be influenced by how pro-active the police may be in detecting offences, which can again vary between police forces (and over time within the same force).

Having been brought to the attention of the police, the decision to record an incident as a crime is the result of a complex interaction of police discretion and the various counting rules and protocols followed by the police to classify and record crimes (Burrows et al., 2000). An officer must first determine whether an incident meets the legal thresholds to be considered a crime, before making an individual judgement on whether or not to proceed (Klinger & Bridges, 1997). Having decided to proceed, the crime is then classified according to pre-defined criteria, giving precedence to offence severity and counting offences based on the number of victims (as opposed to the number of criminal acts). 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).

These different sources 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-recording 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 across victims, or in the recording process across police forces and areas, could be considered random errors, as they introduce undue variability (i.e. noise) in police crime rates. Put differently, the former group of errors impact the validity of police crime rates, while the latter affects their reliability (Lohr, 2019).

Another important feature defining the type of measurement error affecting police recorded crime rates stems from its predominantly multiplicative form. Gibson and Kim (2008) argue that for the case of police crime rates the relationship between the errors (U) present in the observed and imperfectly measured crime rate variable (X^*) is related to the true but unobserved variable (X) multiplicatively, $X^* = XU$, with $E(U) \neq 1$ if the errors are systematic (i.e. non-random). This is a common representation of measurement error used in applications exploring the presence of errors in count and duration data, or similarly distributed variables (Glewwe, 2007; Pickles et al., 1996; Skinner & Humphreys, 1999). This representation implies that the magnitude of the error term is proportional to the true crime counts. We can see how that might be the case by considering crime rates as the aggregated counts of individual crime events, each one of them potentially missed by police records. From that premise follows that, on average, we should expect larger errors in areas where crime is more prevalent.

However, in some instances it may make sense to consider an additive measurement error structure, represented as, $X^* = X + U$, where $E(U) \neq 0$ if the errors are systematic. This is the case when crime rates are log-transformed; a common strategy used to normalise their often right-skewed distribution (Sutherland et al., 2013; Whitworth, 2011), to interpret effects in relative terms (Goulas & Zervoyianni, 2013; Witt & Witte, 2000), or as a result of employing generalised linear models where logs are used as the link function, such as Poisson models (Quick, 2019; Sampson et al., 1997). Crucially, log-transforming crime rates also has the effect of transforming the observed multiplicative measurement error into an additive mechanism, since: $\log(X^*) = \log(XU) = \log(X) + \log(U)$. This distinction between additive and multiplicative errors has important implications when it comes to anticipate the impact of measurement error in regression models and the strategies that could be adopted to adjust for it.

2.1. An Empirical Study Using British Data

To confirm that this is an accurate conceptualisation of the forms of measurement error affecting police recorded crime rates, we compare the rate of crimes recorded by the different police forces in England and Wales for the year ending March 2012 against estimates from the CSEW (2011/12).[1] This involves assuming that the CSEW represents a 'gold standard' that is free of measurement error; a common approach employed across studies that have explored the presence of measurement error in police data (Gibson and Kim, 2008; Vollaard, 2012). The implications of this assumption are discussed in Section 5.1.

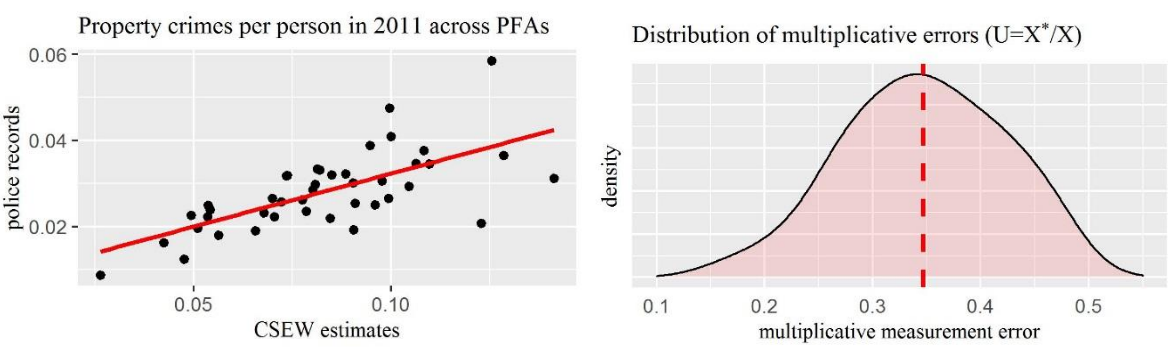
We take discrepancies between the two data sources as evidence of measurement error in police records. Specifically, we compare the average number of property crimes recorded per household in each police force area in England and Wales in 2011/12, against similar types of crimes estimated from the CSEW in that same period and areas. To obtain roughly comparable groups we aggregate the following offence categories from the police records and the CSEW: vehicle theft, bicycle theft, and residential burglary. We include all incidents measured in the CSEW irrespective of whether victims reported them to the police, allowing us to understand the full extent of the discrepancies between incidents experienced by crime victims and those that the police records. Police recorded crime data is accessed from the Home Office open data tables (https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables). 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).

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

Our sample consists of 42 police forces operating in England and Wales after excluding the City of London Police force, which in 2011 recorded a crime rate 7.9 times larger than the average police force.[2] Figure 1 shows a scatterplot depicting the relationship between crime rates estimated by the CSEW and recorded by the police, and a histogram of the distribution of the discrepancies between the two measures under the assumption of multiplicative errors. Inspecting these two graphs we can identify three key properties defining the type of measurement error present in police recorded crime counts:

i. Under-estimated; the number of crimes recorded by the police is lower than those estimated by the CSEW across all areas, as can be inferred from the different ranges covered by the Y and X axes of the scatterplot; on average, the police only picks up 34.7% of the crime counts estimated by the CSEW.
ii. Multiplicative; the size of the errors is proportional to the magnitude of the observed crimes, as shown by the increasing divergence from the line summarising the relationship between $X^*$and X in the scatterplot.
iii. Inconsistent; as shown in the histogram discrepancies are not uniform but normally distributed (with a standard deviation of 0.07), pointing at unequal reporting or recording practices across areas.

Fig. 1. Comparison of property crimes derived from police data and the CSEW, and distribution of their discrepancies under the assumption of multiplicative measurement errors



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

Various factors determine the extent that a model's estimates will be impacted by measurement error: the specific form and prevalence of the measurement error, the type of regression model, how the variable is used in that model, and the way any other variables included in the model are correlated. 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 crime rates data.

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 but with just a second explanatory variable, Z, which we assume is 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 $cov(U, Z) = 0$. Lastly, to further 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^* (Tarling & Dennis, 2016; Whitworth, 2012).

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

[2] 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.

the fact that recorded crimes are not always introduced in their original scales (either as counts, the absolute number of crimes recorded in an area, or as a rate, relative to the number of residents in that area) but may first be log-transformed. Recall that: $\log(X^*) = \log(XU) = \log(X) + \log(U)$.

3.1. Crime Rate as an Explanatory Variable

Let us start with the case of a simple linear model where both the response and the explanatory variables are continuous, and the latter is affected by measurement error, which can be defined as, $Y = \alpha + \beta X^* + \varepsilon$. Using ordinary least squares, 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, $\hat{\alpha}^* = \bar{Y} - \left(\hat{\beta}\bar{X} + \hat{\beta}u\right) = \hat{\alpha} + \hat{\beta}u$. 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.[3] Since the substantive interest of regression models stems from the association between explanatory and response variables, we might conclude that the consequences of this type of systematic error are minimal.

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 also be biased because both the variance and covariance are affected by a change of scale.[4] Substituting from the second line of Eq. (1) we have,

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

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 crime records includes a random component. Here, U is a normally distributed variable, $U \sim N(\bar{U}, S^2_U)$. 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,

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

---

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

[4] 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:
$S_{X^*,Y} = \frac{\Sigma(X^* - \bar{X}^*)\,(Y^* - \bar{Y}^*)}{n-1} = \frac{\Sigma(Xu - (\bar{X}u))\,(Y^* - \bar{Y}^*)}{n-1} = \frac{\mathrm{u}\Sigma(X - \bar{X})\,(Y^* - \bar{Y}^*)}{n-1} = uS_{X,Y}$

The slope is attenuated by a factor equal to the reliability ratio of X^*. 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, $S_{X^*}^2 = S_X^2 S_U^2 + S_X^2 \bar{U}^2 + S_U^2 \bar{X}^2$.

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. The ordinary least squares will not estimate $\hat{\beta}_1$ but rather,

$$\hat{\beta}_1^* = \hat{\beta}_1 \frac{S_{X|Z}^2}{S_{X|Z}^2 + S_U^2} \tag{4}$$

which differs from the bias observed in the slope of the simple linear model, Eq. (3), since $S_{X|Z}^2$ 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,

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

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

It is clear that the impact of using variables affected by measurement error in multivariate models is not negligible, and in many cases 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 into 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 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 the regression coefficients will not be affected. However, in the presence of systematic errors all model estimates will be biased. The extent of the bias will be proportional to $\bar{U}$, which for the case of under-recorded crime rates will lead to attenuation bias.

In sum, the manner in which the type of measurement errors observed in police recorded crime rates can impact the validity of estimates from regression models is not uniform. Depending on the form and prevalence of the measurement error, the type of outcome model, where in the model the errors are introduced, and the way 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 the model's intercept) to potentially substantial (e.g. as shown in Eq. 8, systematic multiplicative measurement error affecting the response variable will lead to bias across all regression coefficients).

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

Importantly, it is not just the magnitude of the impact of measurement error that matters, but the extent to which the impact can be predicted. If the specific biasing effect can be anticipated simply enough – as is the case for the scenario shown in Eq. (8), where all we need would be the level of under-recording in police data – then the extent to which the measurement error can lead to incorrect inferences is reduced. However, the specific biasing effect of measurement error is often harder to predict, with systematic and random forms of errors leading to different types of biases. The combination of these types of errors in varying degrees of intensity across different crime types makes it particularly hard to anticipate their joint effect. To better understand the impact that can arise from the interaction of these two distinct measurement error mechanisms we use computer simulations.

For our empirical illustration, we use linear models investigating the relationship between worry about crime and police recorded crime rates, while taking perceptions of disorder as a control variable. All variables are measured at the police force area 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. Table 1 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 I. Descriptive statistics of data used in empirical illustration

	Mean	Median	Min.	Max.
Property crime	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 2 presents the estimates for these models, which are considered the 'true' models, unaffected by measurement error. 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 II. Regression coefficients from the 'true' models

	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 the estimates from the true models presented in Table 2 against the estimates obtained for the same models after the crime variable is subject to different forms of simulated errors. For the models where crime is log-transformed this is done *after* the simulated errors have been introduced in the crime variable. We focus on the impact on the coefficients for worry about crime and perceptions of disorder, and their measures of uncertainty. This is quantified using the relative bias; the difference between the observed ($\hat{\beta}^*$) and true estimate ($\hat{\beta}$) in relation to the strength of the true estimate, which for the case of a regression coefficient can be expressed as, $R.BIAS = (\hat{\beta}^* - \hat{\beta})/\hat{\beta}$.

Table 3 presents the estimated recording rates for the most commonly used crime categories in existing literature, derived from victimisation survey data on whether crimes came to be known to the police in the CSEW and the National Crime and Victimization Survey (NCVS).[5] The two sources are not directly comparable because of differences in the specific offence types included in each (see Appendix 1), however the levels of under-recording for these crime categories are not too dissimilar. More substantial differences can be observed across crime types, with motor vehicle theft showing close to perfect recording rates, while fewer than half of the property crimes is recorded.

Table III. Recording rates for crime types commonly used in the literature relying on police statistics

	CSEW 2018/2019		NCVS 2018 and 2019	
	Cases reported in the survey	% known to police	Cases reported in the survey	% known to police
Violent crime	1979	41.5%	152	67.8%
Property crime	2035	37.3%	166	42.6%
Burglary	719	60.8%	69	66.7%
Motor vehicle theft	130	92.3%	17	82.4%
All crimes	7,840	38.6%	1,549	55.1%

[5] The recording rates in Table 3 refer to a specific one-year period (covering parts of 2018 and 2019) in the CSEW and two full years in the NCVS. Research on recording rates point at a relative stability in the ranking of offences (Tarling & Morris, 2010), yet the specific under-recording rate can vary substantially. As such, we should only consider them as approximations.

We simulate a wide range of different multiplicative measurement error scenarios.[6] To reflect the different levels of under-recording observed in Table 3, we consider the impact of under-recording rates ranging from no under-recording 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 1). However, the range considered here is likely to reflect the vast majority of situations, including studies that focus on all recorded crimes (Cho & Park, 2017; Matsueda et al., 2006), broad categories of property or violent crime (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, enhancing the external validity of our findings. Random errors might be a result of variations in recording practices across police forces, as well as potential differences in reporting propensities between areas. We might anticipate that the magnitude of this random error could be expected to be proportional to the heterogeneity of the areas under consideration, which, as recently shown by Buil-Gil et al. (2021), is inversely related to the spatial resolution used; 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 consider three scenarios. One where recording rates are assumed uniform, another where we simulate the same variability in the multiplicative errors detected in Section 2 for the case of property crime ($sd = 0.07$), and a third scenario where that observed variability is doubled ($sd = 0.14$). In interpreting these scenarios, it is important to consider that the estimate of random error used in our second scenario reflects the observed variability across police force areas, which represent one of the largest spatial units used in the literature, and should be expected to reflect only moderate inconsistencies across areas (Witt & Witte, 2000).

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

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

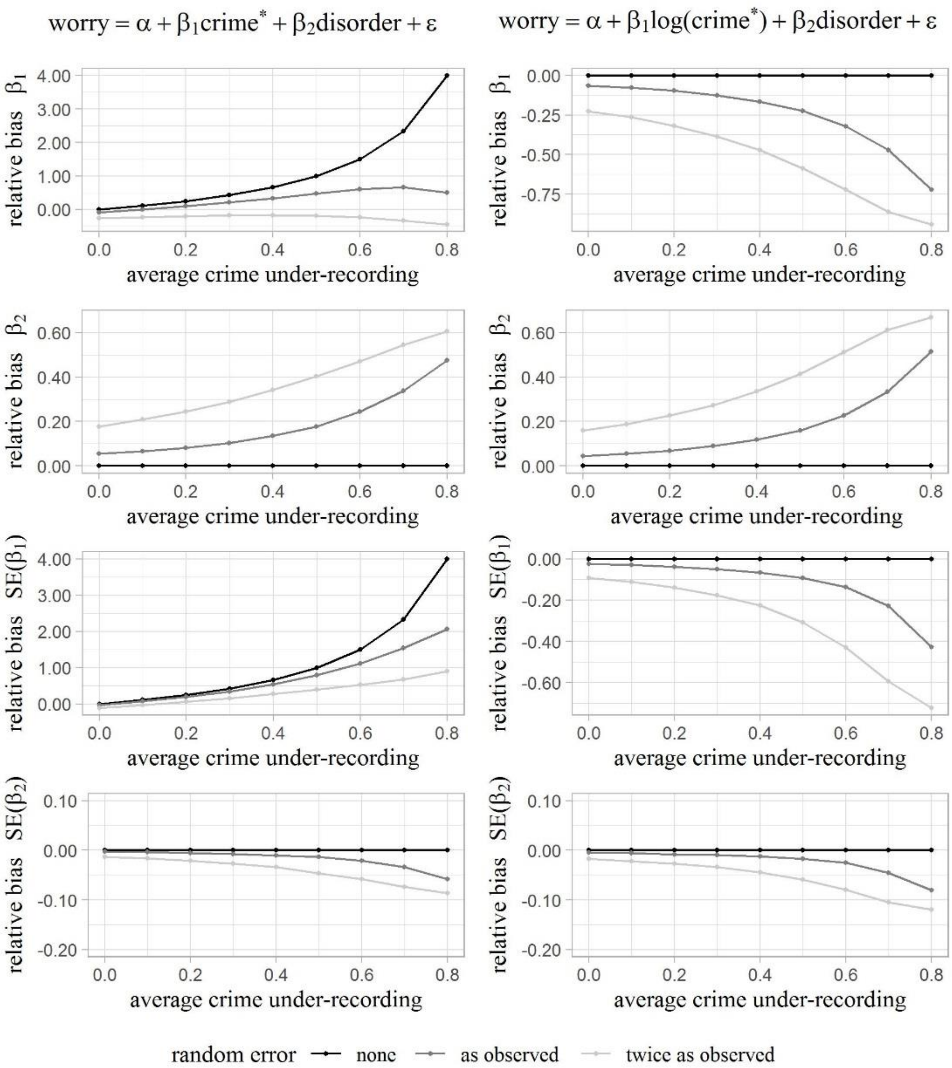
The impact of the different measurement error mechanisms when police crime rates are used as an explanatory variable is shown in Figure 2. When crime rates are introduced using their original scale, we observe a clear augmentation bias in the crime coefficient, β_1, that grows as the percentage of under-recorded cases increases. Yet the magnitude of this bias may be 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 crime recording, although this impact is less severe. Importantly, now the two measurement error mechanisms appear to 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 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 used on their original scale. This bias can be seen to result as a combination of the systematic and random mechanisms, which now operate in the same direction. By contrast, the impact on the regression coefficient for perceptions of disorder remains similar to the non-transformed case, with its standard error also relatively unaffected.

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

Fig. 2. Impact of the measurement error (note change of y scale across graphs) in police recorded crime rates used as an explanatory variable



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 extreme random error (twice as large as the recording variability observed for property crime at the police force area level) the bias can be practically eliminated. Effects for crime types commonly used in the literature, such as violent or property crime, with recording at rates near 40% (Table 3), measured at a spatial unit such as police force areas may be expected to be overestimated by as much as 50%. Perhaps reassuringly, in spite of 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.

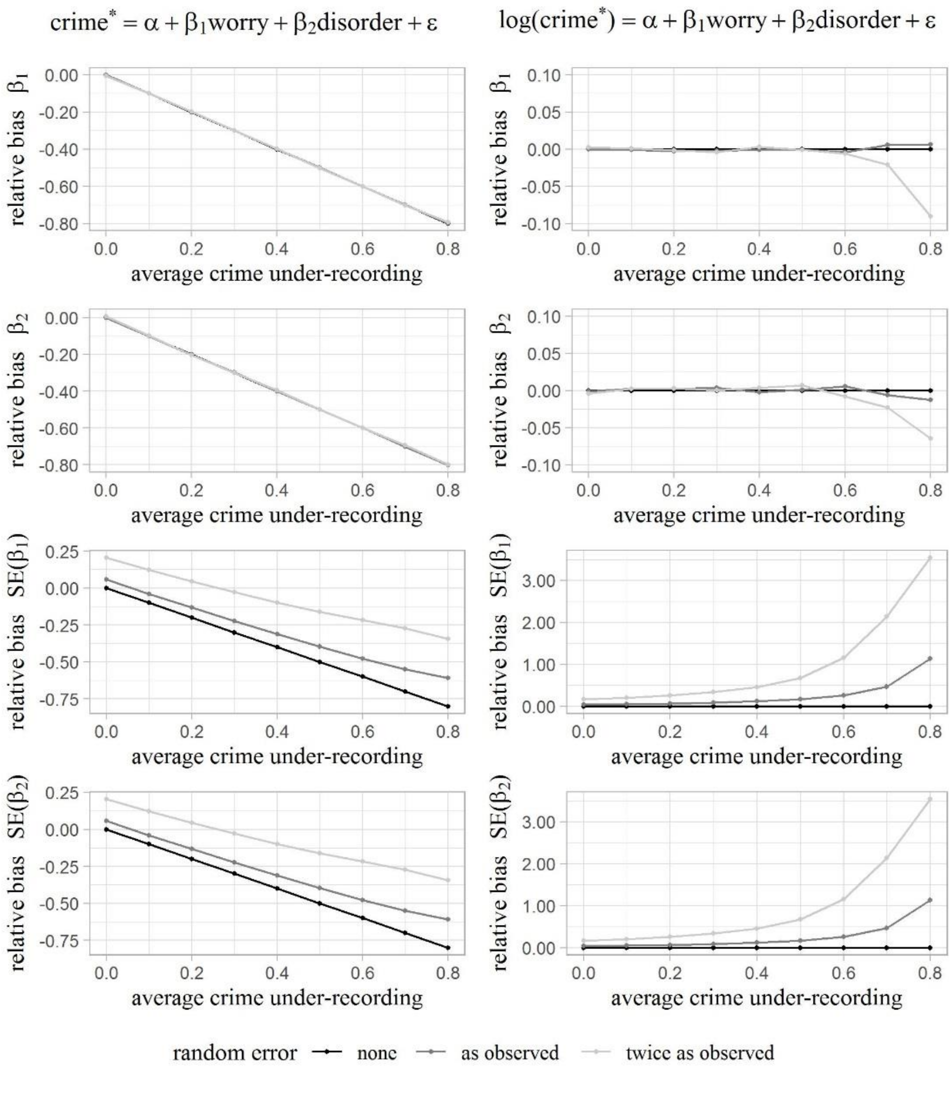
When crime rates are log-transformed, both random and systematic errors operate in the same direction, leading to an attenuation bias in the effect of crime. In this case, the extent of the bias is less severe. Taking again crime types with recording rates around 40%, we can see the magnitude of the bias being relatively negligible (attenuating the true effect size by around 25%) unless recording variability across areas is extreme. The impact on the standard error follows a similar pattern, rendering the presence of false negatives relatively marginal.

In addition, regardless of the scale used for crime rates, we have observed that the bias does not just affect the regression coefficient for crime, but also spreads to regression coefficients from other variables included in the model. We have only explored this for one variable, perceptions of disorder, which is positively correlated with crime ($\rho = 0.62$). For the specific regression coefficient of perceptions of disorder, and under the same reference crime described above (40% recording rates measured at policed force area level or equivalent), we could expect an augmentation bias of roughly 25% regardless of whether crime is introduced in its original scale, or log-transformed. However, the specific impact to be observed in any other explanatory variables included in the model is harder to anticipate. Depending on the relationship between crime and the response variable, and crime and any other explanatory variable, the magnitude and direction of the bias will vary.

4.2. 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. If they are log-transformed, then all regression coefficients remain unbiased, while their standard errors will only be affected by an augmentation bias when extreme systematic and random errors are present simultaneously. Compare this to the substantial attenuation biases observed across all regression coefficients and their standard errors when crime rates are introduced in their original scale. This is a type of bias that - as anticipated in Eq. (8) - is proportional to the rate of under-recording.

Fig. 3. Impact of the measurement error (note change of y scale across graphs) in police recorded crime rates used as the response variable



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 close to null unless the crimes analysed are affected by large under-recording rates (over 60%) and there is substantial random error. In those cases, we should also 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 estimates are attenuated by a factor proportional to the under-recording rate of the crime type being modelled. That is, even in the presence of the more accurately recorded crime types, such as motor vehicle crime, with a recording rate of roughly 90%, we should expect all model's estimates to be attenuated by 10%. When considering the crime types that are more commonly used in existing research, with recording rates closer to 40% (e.g. property or violent crimes), an attenuation bias of around 60% 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. Standard errors are affected by a similar form of attenuation bias, although in this case the random error pushes the bias in the opposite direction. This form of augmentation bias in the standard errors is however smaller compared to the attenuation produced by the systematic under-recoding mechanism. Hence, we can rule out the presence of widespread false negatives unless variability in recording across areas is extreme.

5. DISCUSSION

Most users of police statistics know the data is deeply flawed. Yet, with the exception of a few studies focused on specific applications (Barnett, 1981; Gibson and Kim, 2008; Levitt, 1998; Vollaard and Hamed, 2012), the extent to which the measurement error present in police recorded crime rates impacts statistical models is unknown. In this article, we have provided a comprehensive overview of the impact that could be expected in the types of regression models where police recorded crime rates are commonly used in the literature. We have considered: i) combinations of the under-recording (systematic errors) and recording inconsistencies across areas (random errors) observed across different crime types and spatial units; ii) the inclusion of crime rates as the model's response variable, but also as an explanatory variable; and iii) the inclusion of crime rates in their original scale and log-transformed.

The need to adopt an overarching perspective, and move beyond explorations of the impact of measurement error under specific settings, is clearly demonstrated by our findings, which show different effects depending not only on the type and prevalence of the measurement error, but on how are crime rates introduced in the regression model. 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, we have also shown that the potentially severe impact of measurement error seen in police records can be considerably minimised - and in certain settings altogether eliminated – by log-transforming crime rates. This is a transformation commonly undertaken for different reasons: i) to normalise their often right-skewed distributions (Sutherland et al., 2013; Whitworth, 2011); ii) as a result of the use of generalised linear models such as Poisson or negative-binomial (Osborn & Tseloni, 1998; Sampson et al., 1997); iii) or just to express the relationship between crime and other variables in relative terms (Han et al., 2013; Machin & Meghir, 2004). Here we have shown that, in some scenarios, log-transformations have the added benefit of reducing the impact of measurement error by turning the more damaging multiplicative error observed in police recorded crime rates into a more benign additive error. 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 based on the log-transformation of crime rates. Here we echo Osgood's advice and extend it to instances where crime rates are used as explanatory variables.

By illustrating the specific impact that can be attributed to different measurement error mechanisms present in police recorded crime rates we have also shown how the systematic and random components of the error can amplify or oppose each other depending on the setting. The combined impact of these interactions is identified for different quantities of each of the two measurement error mechanisms, providing a broad overview of their impact. This new understanding of the impact attributable to each of the different measurement error mechanisms present in police data can be used to anticipate the specific biases to be encountered across a wide range of scenarios, based on different models, crime types, and spatial units.

Being able to approximate the specific impact associated with the use of police data across different settings represents an important contribution to the empirical study of the causes and consequences of crime. Undertaking simple sensitivity analysis future researchers will be able 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 approximate the impact of measurement error in previous studies to adjust for any potential biases in the reported effects.

5.1. Caveats and Future Avenues of Research

The accuracy 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. Following the most common approach used in the literature, we have operationalised measurement error in police data as any discrepancy observed between police recorded crime and crime estimates from victimisation surveys. This approach invokes the assumption that victimisation surveys are a gold standard, free of measurement error. This is a convenient assumption, but one we know is not entirely correct.

Victimisation surveys are subject to multiple limitations that ought to be considered to assess the accuracy and precision of our estimates of measurement error. The validity of those estimates is likely affected by differences in the definition of crime types used in victimisation surveys and police statistics, and by the sampling error present in survey data. Crime rates estimated from victimisation surveys are by definition uncertain; hence, when we take them as a gold standard we are overestimating the extent of measurement error in police statistics. The reliability of our estimates of measurement error is also affected by discrepancies in the location where the crime is registered. Estimates of crime from victimisation surveys normally refer to the geographic area of residence of the interviewee, whereas the location attributed to police recorded crimes is that where the offence took place. Such discrepancy may lead to an overestimation of the variability of recording rates across areas, which will become larger the smaller the spatial unit used to express crime rates.

It is therefore essential to explore the extent to which the different limitations of victimisation surveys may affect our estimates of measurement error in police statistics. We can deduce that the under-recording rates reported in the literature - and throughout this article - are exaggerated, but how much so is currently unclear. 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. These can be used to estimate the validity and reliability of variables measuring the same concept, and, 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 on the measurement error in police data explore and formally define the presence of any other potential 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 response variables or other key 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 enhance the precision with which the impact of measurement error can be estimated.

Lastly, we have illustrated the impact in linear models commonly used in the literature. However, these models have also been importantly simplified by including just two explanatory variables, which were moderately positively correlated with each other and with CSEW estimated crime rates. It would be useful if future studies were to expand the exploration of the impact of measurement error in police recorded crime rates by considering scenarios with different and additional explanatory variables. Furthermore, police recorded crime rates are also used in more complex non-linear models (Machin & Meghir, 2004; Sobol et al., 2013), or even as part of systems of equations (Krahn & Kennedy, 1985; Yesberg et al., preprint). In those instances, the biasing effect of measurement error, and how it is propagated through the different parts of the model, might be harder to trace out (Carroll et al., 2006). Further work is needed to explore the impact of measurement error in those instances. Even if the impact happens to be difficult to anticipate, and regardless of the complexity of the outcome model, we could still rely on flexible methods such as Bayesian measurement models (Gustafson, 2003; Pina-Sánchez et al., 2019), or simulation-extrapolation (Biewen et al., 2008; Pina-Sánchez, 2016) to adjust for the impact of measurement error.

6. CONCLUSION

We urge researchers exploring the causes and consequence of crime by employing police statistics in regression models to consider how estimates from their models might be impacted by the presence of measurement error. We have shown how the impact associated to the type of measurement error present in police recorded crime rates is potentially substantial. Yet, we have also shown how, based on 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 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 explored.
ii. That attenuation bias is almost completely eliminated when crime rates are log-transformed, rendering such transformations essential in future studies.
iii. Studies including police recorded crime rates in their original scale as an explanatory variable should expect a strong augmentation bias in the effect of crime, proportional to the level of under-recording affecting the crime type considered, but inversely 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 is, however, 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. 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, making them harder to anticipate.

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.

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.

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., Medina, J., and Shlomo, N. (2020). 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.

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

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

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.

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.

de Candolle, A. (1830 [1987a]). Considérations sur la statistique des délits. Déviance et Société. 11(4): 352--355.

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.

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.

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.

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

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.

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.

Mohler, G., Brantingham, P. J., Carter, J., and Short, M. B. (2019). Reducing bias in estimates for the law of crime concentration. Journal of Quantitative Criminology: 35: 747--765.

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.

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

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.

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.

Quick, M. (2019). Multiscale spatiotemporal patterns of crime: a Bayesian cross-classified multilevel modelling approach. Journal of Geographical Systems. 21(3): 339--365.

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.

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.

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.

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 Dennis, R. (2016). Socio‐Economic Determinants of Crime Rates: Modelling Local Area Police‐Recorded Crime. The Howard Journal of Crime and Justice. 55(1-2): 207--225.

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.

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.

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/

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 1. SPECIFIC OFFENCES USED TO DEFINE BROADER CRIME TYPES IN TABLE 3

CSEW 2018/2019			NCVS 2018 and 2019		
Crime type	Cases reported in the interview	% known to police	Crime type	Cases reported in the interview	% known to police
Violent crime	1979	41.5 %	**Violent crime**	152	67.8 %
Hit with fists or weapon	538	48.7 %	Assault	54	70.4 %
Threaten to use force or violence on you	1319	38.7 %	Attempted assault	89	65.2 %
Sexually assaulted	85	41.2 %	Rape	4	100 %
Violent from household member	37	40.5 %	Unwanted sexual contact from household member	5	60 %
Property crime	2035	37.3 %	**Property crime**	166	42.6 %
Something stolen out of hands or pockets	304	46.1 %	Larceny	190	79.5 %
Other theft	360	27.2 %			
Tried to steal	203	13.3 %	Attempt larceny	19	63.2 %
			Robbery	6	33.3 %
			Attempted robbery	1	100 %
Something stolen off car	796	40.5 %			
Bike theft	372	46.5 %			
Burglary	719	60.8 %	**Burglary**	69	66.7 %
Get in previous house to steal	38	68.4 %	Burglary	62	66.1 %
Get in previous house and cause damage	10	70 %			
Get in house since moved in to steal	8	75 %			
Get in current house to steal	250	76.4 %			
Get in current house and cause damage	37	67.6 %			
Try to get in previous house to steal/damage	21	19 %	Attempted burglary	6	83.3 %
Try to get in current house to steal/damage	355	50.1%			
Motor vehicle theft	130	92.3%	**Motor vehicle theft**	17	82.4%