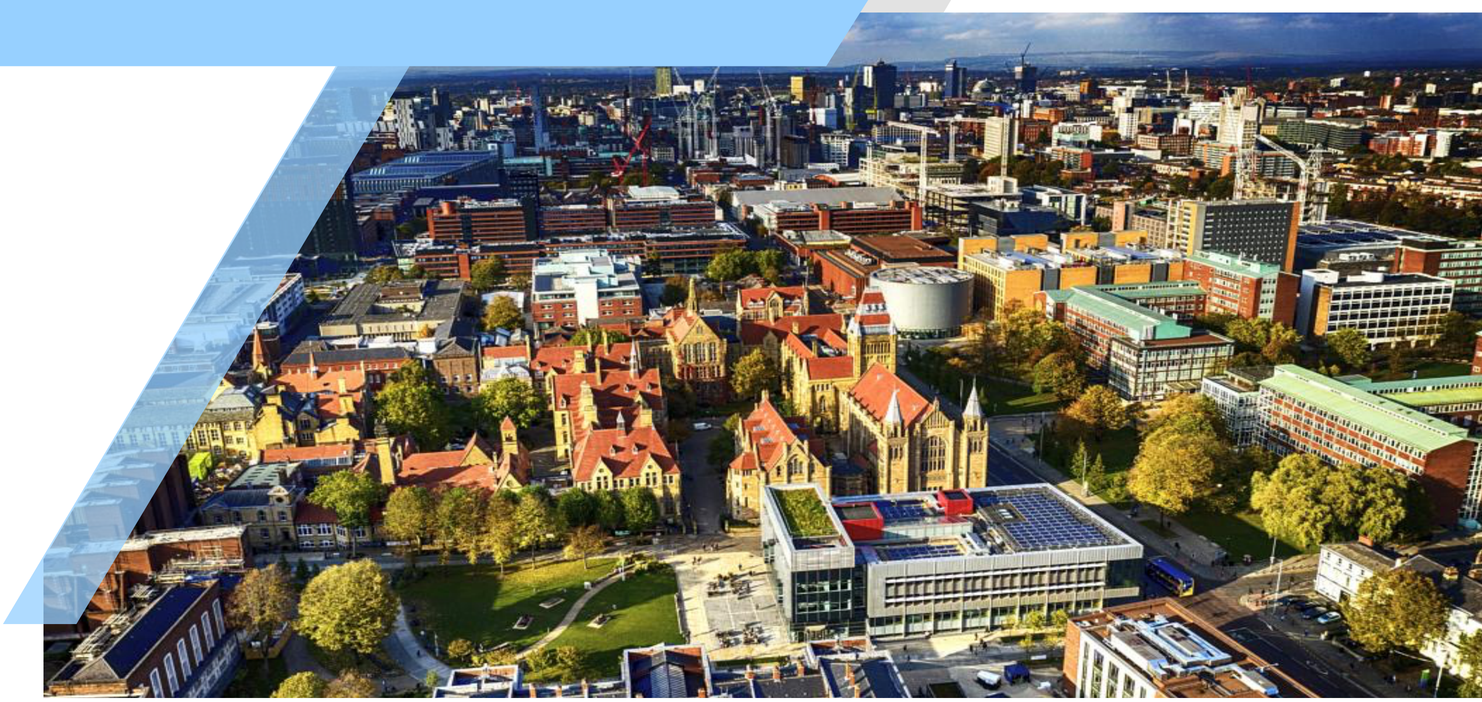
Manchester Statistical Society

Project funded by the Campion
Grant 2020/2021

Mapping the bias of police records

Project Final Report



Dr Angelo Moretti
Dr David Buil-Gil

MANCHESTER STATISTICAL SOCIETY 1833

Manchester Metropolitan University

MANCHESTER 1824 The University of Manchester

TABLE OF CONTENTS

Executive Summary 1
Purpose of the Project 2
Review of Existing Literature 2
Our Approach 3
Main Findings and Conclusions 4
Note on Authors and Contributors 7
Project Outputs 7
References 8

Executive Summary

Police-recorded crimes are the main source of data used by police forces and researchers to analyse the distribution of crime in geographic areas. These data, however, are affected by measurement error arising from underreporting and recording inconsistencies across police jurisdictions. Not all persons are equally willing to report crimes to the police and cooperate with police services, and practices followed by the police to record crime vary across police jurisdictions. Both these issues affect the 'dark figure of crime' (i.e., all crimes unknown to the police), which vary across areas. The open question that this project addresses is whether micro-level maps of police-recorded crimes (i.e., maps produced from aggregating crimes at very detailed spatial scales) suffer from a higher risk of bias than maps of crime produced at larger scales, such as neighbourhoods and ward. While communities unwilling to cooperate with the police may concentrate in some micro-places more than others, and thus the 'dark figure' may vary across small areas, larger geographies aggregate more heterogeneous social and demographic groups, and thus the proportion of crimes unknown to the police may be more similar across areas. We utilise data from the UK Census and the Crime Survey for England and Wales to generate synthetic crime data in Manchester, UK, and analyse if micro-level crime maps are affected by a larger risk of bias than maps of crime aggregated in neighbourhoods. The main findings of this project are:

- The proportion of crime unknown to the police varies substantially across micro-places.
- The proportion of crimes unknown to the police is similar across neighbourhoods.
- Micro-level maps of crime are affected by a larger risk of bias than maps of crimes aggregated at larger spatial scales.
- The risk of bias in micro-level crime mapping is attributed to the fact that social groups unwilling to cooperate with the police concentrate in some areas more than others. This is less of a problem when aggregating crimes in neighbourhoods.
- Future work is needed to address measurement error in crime data.

Cite as:

Moretti, A. & Buil-Gil, D. (2021). *Mapping the bias of police records: Project final report.* Manchester: Manchester Statistical Society.

Purpose of the Project

This project investigates the impact of measurement error in police-recorded crime data (e.g., varying rates of crime reporting to the police across social groups and geographic areas, recording inconsistencies across police forces) on the accuracy of maps of crime produced at different spatial scales. We use a novel approach to generate synthetic data of crimes known and unknown to the police in Manchester, UK, from parameters observed in the real world, thus allowing us to assess whether maps of crime aggregated at small spatial scales are affected by a larger risk of bias than maps produced at larger scales, such as neighbourhoods or wards.

The hypothesis underlying our project is that the population groups unwilling to report crimes to the police concentrate in some micro-places but not others, and thus the proportion of crimes unknown to the police vary substantially across small areas, while larger units of geography aggregate population groups who report and groups who do not, and thus the proportion of crimes unknown to the police is similar across all areas. In other words, we expect that while the 'dark figure of crime' (i.e., proportion of crimes unknown to the police) will vary very much across micro-places, the 'dark figure' will be similar across neighbourhoods. Hence, the risk that police records underestimate or overestimate crime rates in some micro-places more than others would be large, whereas this risk would be smaller when studying crime at the level of neighbourhoods.

Review of Existing Literature

Police-recorded crimes are the main data source used by the police and government agencies to analyse the geographic concentration of crime, study crime patterns and design and evaluate spatially targeted policing strategies and crime prevention policies (Bowers and Johnson 2014; Weisburd and Lum 2005). Police statistics are also used by researchers to develop and evaluate crime and deviance theories (Bruinsma and Johnson 2018). However, crimes known to police are affected by measurement error driven by unequal crime reporting rates across both social groups and geographical areas (Buil-Gil et al. 2021a; Goudriaan et al. 2006; Xie 2014).

It is well known in the literature that crime reporting to police forces is more common among some population groups than others. For instance, female victims report more often than male victims, and young citizens report crimes less often than adults (Hart and Rennison 2003; Tarling and Morris 2010). Some contextual factors affect crime reporting rates across geographic areas, such as economic deprivation, the degree of urbanisation, minorities concentration, and social cohesion (Goudriaan et al. 2006; Slocum et al. 2010; Xie and Baumer 2019; Xie and Lauritsen 2012). The demographic and social characteristics of residents in micro-places are generally more homogeneous compared with larger scales (Brattbakk 2014; Weisburd et al. 2012). Therefore, crime aggregates for small geographies are more likely to be affected by unequal crime reporting rates compared with aggregates produced at larger, more heterogeneous spatial scales. For example, Buil-Gil et al. (2021a) show that the variation in the 'dark figure of crime' between neighborhoods is larger than the variation between cities.

Our Approach

In this section, we briefly describe the methodological approach used to answer our research question i.e., 'Are crime maps produced at smaller, more socially homogeneous spatial scales, at a larger risk of bias compared to maps produced at larger, more socially heterogeneous scales?'. Formal details on this approach can be found in Buil-Gil et al. (2021b).

Our apporach involves four steps:

1. <u>Simulating a synthetic population of Manchester residents from Census 2011</u>

The first step is to generate a synthetic population consistent with the social, demographic and spatial characteristics of Manchester. First, we download 2011 Census data aggregated in Output Areas in Manchester. Second, we obtain empirical parameters of the following demographic variables in areas: age, sex, income, education, ethnicity. Third, we generate a synthetic population following the empirical parameters in each area.

2. <u>Simulating crime victimisation from the Crime Survey for England and Wales (CSEW) 2011/12</u>

First, we obtain access to the CSEW and estimate Negative Binomial regression models at the individual level for the following crime types: (i) violent crime, (ii) residence crime, (iii) theft and property crime, (iv) vehicle crime. The same independent variables as in Step 1 are used. The Negative Binomial Regression model is a widely used model in criminology. Second, we obtain regression parameter estimates and simulate crime victimisation in the synthetic population following Negative Binomial regression models.

3. <u>Simulating whether each crime is known to the police</u>

The third step consists of estimating whether each simulated crime is known to the police or not. Thus, we can analyse the difference between all crimes (generated in step 2) and those known to the police (to be estimated in step 3) in each geographic area. First, we estimate Logistic Regression models of crimes being known to police in the CSEW. We use the same independent variables as in Step 1. Second, we obtain the regression parameter estimates and simulate if each simulated crime is known to police following a Bernoulli probability distribution.

Moreover, we also estimate if each crime took place in the residents' local area or elsewhere, and remove from all simulated crimes that do not take place within 15-min walking distance from the persons' household. This is done to effectively estimates crimes where they happen instead of crimes where victims live (for technical details, see Appendix 2 in Buil-Gil et al., 2021b). After these steps, we obtain a final synthetic dataset of 359,248 crimes across 1,530 Output Areas in Manchester. We aggregate these in LSOAs, MSOAs and Wards. Output Areas are the smallest geographic units we analyse. The average population size is 125 households. LSOAs generally contain between four and six OAs, with an average population size of 1500, and MSOAs have an average population size of 7200. The largest scale used are wards.

4. Empirical Evaluations

There are two sets of evaluations we carry out. Once the synthetic data is generated, we use victimisation data recorded in the CSEW and data on crimes known to Greater Manchester Police (GMP) to evaluate whether our simulated dataset of crimes matches the empirical values of crime. This is used to evaluate the quality of our synthetic data of crimes. The empirical evaluation of our synthetic dataset is satisfactory.

Main Findings and Conclusions

Table 1 shows the measures of relative difference (RD) and relative bias (RB) between crimes known to the police and all crimes (simulated dataset) across four spatial scales. Figure 1 shows the same results but graphically. The RD is close to 62% for all the spatial scales (i.e., on average, 62% of crimes are unknown to police at each spatial scale), however, the measures of dispersion vary considerably depending on the spatial level under study. The standard deviation of the RD between all crimes and police-recorded offences is the largest at the level of OAs, whereas it is much smaller when crimes are aggregated at the LSOA level. It becomes almost zero at the level of MSOAs and wards: RD has a large variability across small areas, but it is minimal when using larger geographies.

Table 1 *Relative difference (RD%) and relative bias (RB%) between crimes known to police and all crimes (simulated data)*

		OA	LSOA	MSOA	WARD
RD%	Mean	62.2	62.2	62.2	62.2
	SD	3.3	1.4	0.8	0.6
	Min	48.7	57.2	60.7	61.3
	Max	74.1	65.9	64.6	63.5
RB%	Mean	-62.2	-62.2	-62.2	-62.2
	SD	3.3	1.4	0.8	0.6
	Min	-74.1	-65.9	-64.6	-63.5
	Max	-48.7	-57.2	-60.7	-61.3

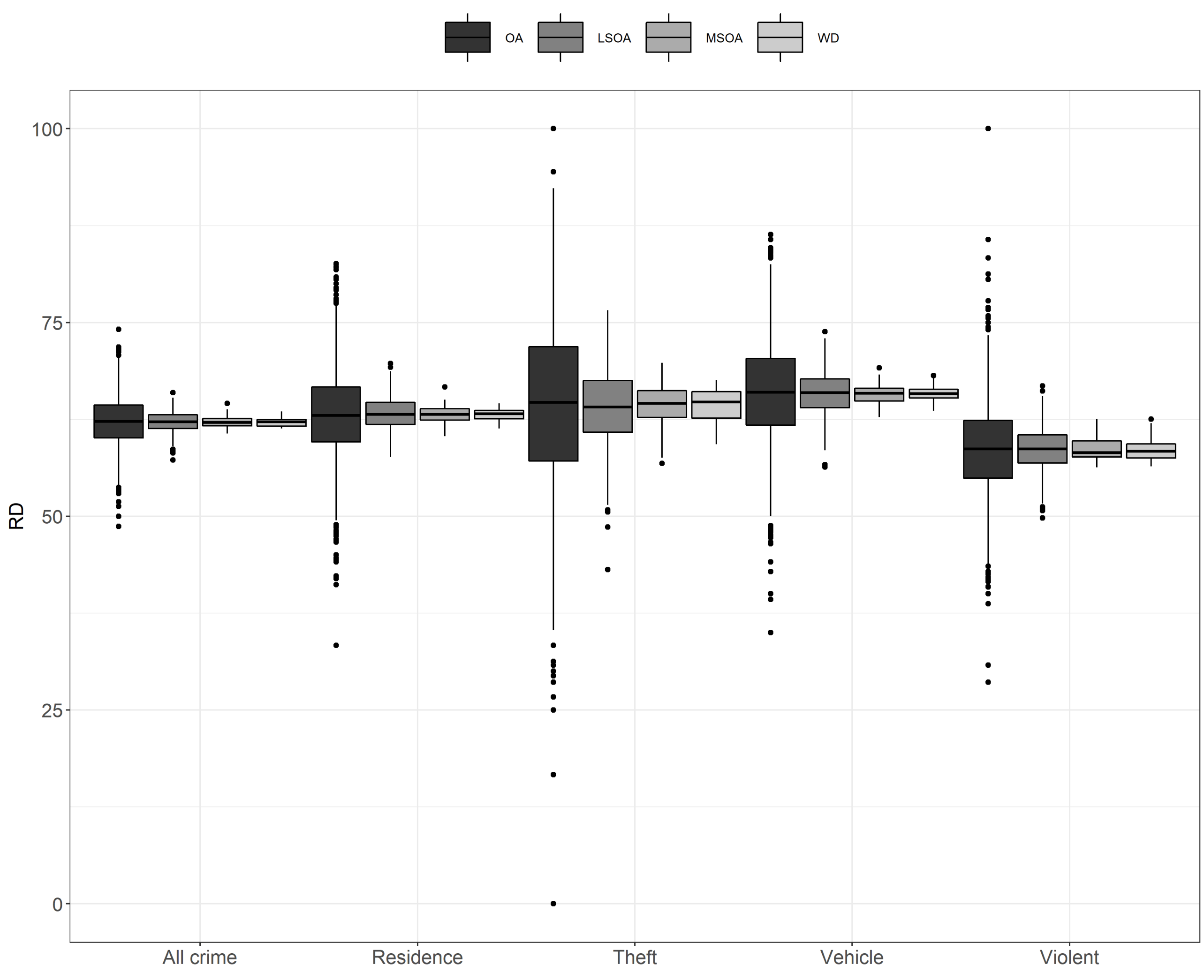


Figure 1 *Boxplots of RD% between all crimes and crimes known to police at the different spatial scales (simulated dataset)*

In other words, this reflects that the police might be aware of the vast majority of crimes in one small area, and in another one, very few. Therefore, geographic crime analysis produced from police records at highly localised spatial scales, such as OAs, may show high concentrations of crime in some areas but simply as an artefact of the variability in the crimes known to police. This varies, however, across crime types, with the 'dark figure' of theft and property crime varying very much across OAs. Figure 2 depicts the RD between all property crimes and property crimes known to the police at the level of OAs, LSOAs, MSOAs and wards in Manchester. This is crucial to better show the impact of measurement error on maps of crime produced at the different spatial scales. Figure 2 shows that the RD varies considerably across OAs, whereas the RD between all crimes and police-recorded crimes becomes homogeneous when crimes are aggregated at the scales of MSOAs and wards.

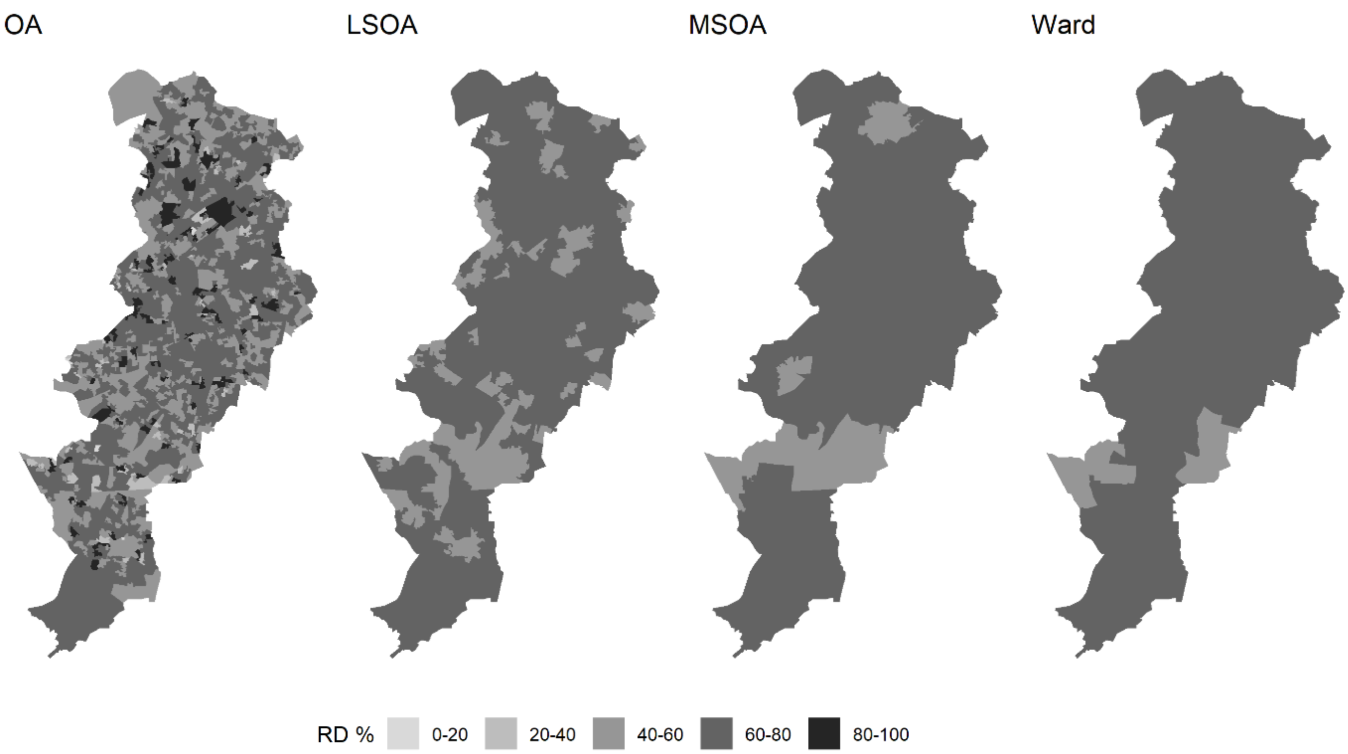


Figure 2 *Maps of RD% between all property crimes and property crimes known to police at the different spatial scales (simulated dataset). Breaks based on equal intervals*

In conclusion, we can say that aggregating crimes known to the police at very detailed levels of analysis increases the risk of inaccurate maps. Maps of police-recorded crimes produced for neighbourhoods and wards (larger spatial scales) show a more accurate image of the geography of crime. Our work has some important limitation too, which future research will address. First, our data captures area victimisation rates instead of area offence rates. Second, the CSEW does not record data about so-called victimless crimes.

Note on Authors and Contributors

Angelo Moretti is a Lecturer in the Department of Computing and Mathematics at Manchester Metropolitan University, UK. His research interests cover topics in small area estimation, survey statistics, data integration, statistical modelling and multivariate statistics with strong emphasis on crime, wellbeing and poverty indicators at small geographical level.

David Buil-Gil is a Lecturer in Quantitative Criminology at the Department of Criminology of the University of Manchester, UK. His research areas cover geographic criminology, small area estimation applications in criminology, crime mapping, measurement error in criminological research, emotions about crime, perceptions about the police, new methods for data collection and open data.

The research also included two researchers: **Samuel H. Langton** (February 2021), who is currently a Postdoc at the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), and **Yongyu Zeng** (March-April 2021), PhD Candidate at the University of Manchester. Samuel H. Langton contributed to the first phases of the project with the production of user-friendly software and outputs that were used in the article published as a result of this project. He also contributed with literature review and writing up of the article. Yongyu Zeng joined the project in the last phase of the project (March-April 2021). She contributed to the final dissemination activities and identification of gaps that will be addressed in future research projects.

Project Outputs

The main output of this project was a journal article published in the *Journal of Experimental Criminology*. This article is published in open access thanks to the support from the University of Manchester Library.

- Buil-Gil, D., Moretti, A., & Langton, S.H. (2021). The accuracy of crime statistics: assessing the impact of police data bias on geographic crime analysis. *Journal of Experimental Criminology*. DOI:10.1007/s11292-021-09457-y

The codes and data produced for the article were also published in an open-access Github repository: https://github.com/davidbuilgil/crime_simulation2

We also presented our work in the following conferences:

- Moretti, A., Buil-Gil, D., & Langton, S.H. (2020, December). Crime mapping and the dark figure of crime: Assessing the impact of police data bias on maps of crime produced at different spatial scales. Paper presented at *UK Data Service Crime Surveys User Conference 2020*. Online due to COVID-19.
- Buil-Gil, D., Moretti, A., & Langton, S. H. (2020, November). The bias of crime statistics: Assessing the impact of data bias on police analysis and crime mapping. Paper presented at the *European Survey Research Association BigSurv20 (Big Data Meets Survey Science) Conference*, Utrecht, Netherlands. Online due to COVID-19.

- Buil-Gil, D., Moretti, A., & Langton, S. H. (2020, November). Mapping the bias of police records. A simulation study about the impact of data bias on crime mapping. Paper presented at the *Annual Meeting of the Criminology Consortium*. Online due to COVID-19.

In order to further disseminate the main findings of the project, we will publish a blog post in the Policy@Manchester blog.

As a final event of the project, we organised a webinar with invited speakers, which took place on 19th May at 2pm-4:15pm BST titled 'Mapping the bias of police records: An assessment of the impact of police data bias for crime mapping'.

References

Bowers, K., & Johnson, S. D. (2014). Crime mapping as a tool for security and crime prevention. In M. Gill (Ed.), *The Handbook of Security* (2nd ed., pp. 566–587). London: Palgrave Macmillan. DOI:10.1007/978-1-349-67284-4_25

Brattbakk, I. (2014). Block, neighbourhood or district? The importance of geographical scale for area effects on educational attainment. *Geografiska Annaler: Series B, Human Geography*, 96(2), 109–125. DOI:10.1111/geob.12040

Bruinsma, G. J. N., & Johnson, S. D. (Eds.) (2018). *The Oxford handbook of environmental criminology*. New York: Oxford University Press. DOI:10.1093/oxfordhb/9780190279707.001.0001

Buil-Gil, D., Medina, J., & Shlomo, N. (2021a). Measuring the dark figure of crime in geographic areas: Small area estimation from the Crime Survey for England and Wales. *British Journal of Criminology*, 61(2), 364–388. DOI:10.1093/bjc/azaa067

Buil-Gil, D., Moretti, A., and Langton, S.H. (2021b). The accuracy of crime statistics: Assessing the impact of police data bias on geographic crime analysis. *Journal of Experimental Criminology*. DOI:10.1007/s11292-021-09457-y

Goudriaan, H., Wittebrood, K., & Nieuwbeerta, P. (2006). Neighbourhood characteristics and reporting crime: Effects of social cohesion, confidence in police effectiveness and socio-economic disadvantage. *British Journal of Criminology*, 46(4), 719–742. DOI:10.1093/bjc/azi096

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

Slocum, L. A., Taylor, T. J., Brick, B. T., & Esbensen, F. A. (2010). Neighborhood structural characteristics, individual-level attitudes, and youths' crime reporting intentions. *Criminology*, 48(4), 1063–1100. DOI:10.1111/j.1745-9125.2010.00212.x

Tarling, R., & Morris, K. (2010). Reporting crime to the police. *British Journal of Criminology*, 50(3), 474– 490. DOI:10.1093/bjc/azq011

Weisburd, D., & Lum, C. (2005). The diffusion of computerized crime mapping in policing: Linking research and practice. *Police Practice and Research*, 6(5), 419–434. DOI:10.1080/15614260500433004

Weisburd, D., Groff, E. R., & Yang, S. M. (2012). *The criminology of place. Street segments and our understanding of the crime problem*. New York: Oxford University Press. DOI:10.1093/acprof:oso/9780195369083.001.0001

Xie, M. (2014). Area differences and time trends in crime reporting: Comparing New York with other metropolitan areas. *Justice Quarterly*, 31(1), 43–73. DOI:10.1080/07418825.2012.742126

Xie, M., & 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. DOI:10.1111/1745-9125.12204

Xie, M., & Lauritsen, J. (2012). Racial context and crime reporting: A test of Black's stratification hypothesis. *Journal of Quantitative Criminology*, 28, 265–293. DOI:10.1007/s10940-011-9140-z