The bias of crime statistics: Assessing the impact of data bias on police analysis and crime mapping

\*Wait to hear back about Journal of the Royal Statistical Society Series A. If not accepted, submit to Journal of Quantitative Criminology

TASKS TO BE FINISHED BY 14th APRIL:

Angelo: 1. Email MMU’s RKE about Sam’s contract. 2. Fill first draft of Section 4

David: 1. Download original Census data and upload it onto Google Drive. 2. Fill first draft of Sections 1, 2 and 3

Sam: 1. Clean up R codes

David Buil-Gil. University of Manchester

Angelo Moretti. Manchester Metropolitan University

Samuel H. Langton. Manchester Metropolitan University

**Corresponding author**

David Buil-Gil. Humanities Bridgeford Street building, School of Social Sciences, University of Manchester, 176 Oxford Road, M13 9PL, Manchester, United Kingdom.

Email: [david.builgil@manchester.ac.uk](mailto:david.builgil@manchester.ac.uk)

**ORCID**

David Buil-Gil. 0000-0002-7549-6317

Angelo Moretti.0000-0001-6543-9418

**Short bio**

David Buil-Gil is a Research Fellow at the Department of Criminology of the University of Manchester, UK. His research interests are in small area estimation applications in criminology, environmental criminology and crime analysis, perceived disorder, fear of crime, new methods for data collection and open data.

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.

Samuel H. Langton is a Research Associate for the Crime and Well-being Big Data Centre at Manchester Metropolitan University, UK. His research focuses on examining longitudinal trends in known offender residence concentrations, the estimation of non-crime police demand and spatial data visualisation. He is also interested in promoting the use of open software in social sciences.

# Abstract

XXX

# Keywords

Unreliability; Simulation; Crime analysis; Manchester; Survey; Official statistics

# 1. Introduction

Police-recorded crimes are the main source of information used by police forces to analyse crime patterns, investigate the geographic concentration of crime, and design and evaluate spatially targeted policing strategies and crime prevention policies (Weisburd & Lum, 2005). Police statistics are also used by criminologists and crime scientists to develop theories of crime and deviance (see Bruinsma & Johnson, 2018; Wortley & Townsley, 2017). Nevertheless, crimes known to police are affected by biases and unreliability driven by unequal crime reporting rates across social groups and geographical areas and the police inability to control every area (Baumer & Lauritsen, 2010; Goudriaan et al., 2006; Hart & Rennison, 2003; Xie, 2014). The measures of error that affect the reliability of crime statistics is an issue that merits deeper scrutiny, since it affects police everyday practices, criminal policies and citizens’ everyday lives. Yet it is an understudied issue, and the implications of data biases for crime mapping are unknown. Moreover, police analyses and crime mapping are moving towards the study of smaller levels of geography than ever before, such as street segments with highly homogeneous communities (Groff et al., 2010; Weisburd et al., 2012). Maps produced from police records are used to foreground the micro places where rates of recorded offences are more prevalent in order to target more police resources in these areas. This paper presents a simulation study and an application to analyse the impact of data biases on crime maps produced from police records at the different spatial scales. It assesses whether micro-level maps are affected by a larger risk of bias than maps produced at larger scales.

A bit on origins of crime mapping

A bit on early discussions about police data bias

A bit on

DGB

# 2. The criminology of place and crime mapping

XXX

DBG

# 3. Bias in crime statistics

XXX

DBG

Susan McVie - chapter Oxford handbook of crime 2017

Leslie McAra

Jeff Brantingham

Papers about unequal crime reporting

# 4. Data and methods

Talk about research strategy

## 4.1 Generating the population and simulation steps

This simulation study consists of two steps.

**Step 1.** Aggregated data from the UK Census 2011 at the Output Area level are downloaded from the NOMIS website (<https://www.nomisweb.co.uk/census/2011>) to calculate the real parameters of each variable in each Output area. From the Census we obtain the number of citizens living in each Output area, and the mean and standard deviation of age, and the proportions of the following binary variable where in the breakets we give the reference category: proportions of sex (males), ethnicity (whites), income (not-working population), education (higher education or more). We will use this information to simulate our synthetic population, and predict the number of crimes suffered by each individual and the likelihood of each crime to be known to the police.

The variables were generated for output areas and units according to the following distributions:

* where and denote the mean and variance of the variable OA
* , where denotes the proportion of males in OA
* denotes the proportion of citizens without any income in the OA
* denotes the proportion of citizens with high education (holding a degree) in the OA
* denotes the proportion of citizens from a white ethnic group in the OA.

Thus, we generate N=503,127 individuals across D=1,530 Output Areas across Manchester Local Authority Districts. These individuals are defined by very similar characteristics to the population of Manchester.

**Step 2.** After this initial step, we access to the Crime Survey for England and Wales 2011-2012 data and estimate negative binomial regression models where the responses are the following categorical variables: vehicle, residence, theft, violence, all crime, denoting how many times a citizen was victim of the specific crime. The independent variables are the same ones as in the previous step, but related to the Crime Survey for England and Wales. In this step, we obtain the regression model coefficient estimates, and these are used to generate the crime counts per person in the generated population.

**Step 3.** Crime counts for each person are generated in the synthetic population created in Step 1 and using the regression coefficient estimates obtained in Step 2 according to a Negative Binomial regression model.

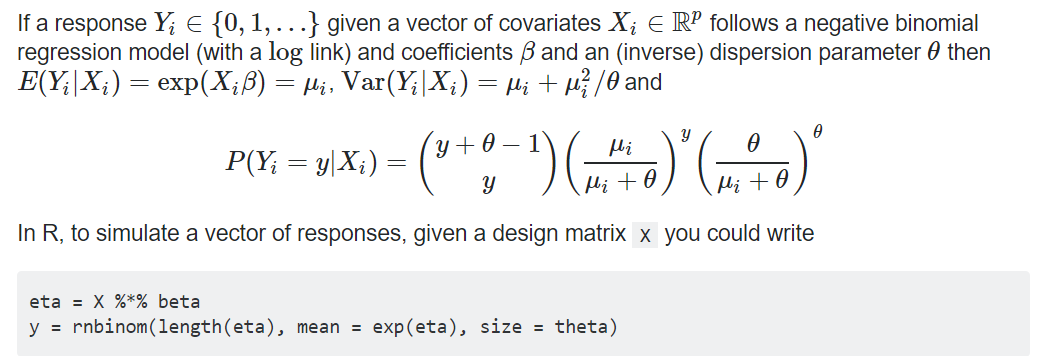


Table X. Negative binomial generalised linear models of crime victimisation fitted from CSEW 2011/12 data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vehicle crimes | Residence crimes | Theft and property crimes | Violent crimes |
| (Intercept) | -0.756\*\*\* | -0.926\*\*\* | -2.085\*\*\* | 0.439\*\* |
| Age | -0.020\*\*\* | -0.011\*\*\* | -0.023\*\*\* | -0.046\*\*\* |
| Male (0/1) | 0.108\*\* | 0.027 | 0.303\*\*\* | 0.002 |
| White (0/1) | -0.097 | -0.203\*\* | -0.092 | 0.241\* |
| Not working (0/1) | -0.359\*\*\* | 0.215\*\*\* | -0.091 | -0.301\*\*\* |
| High education (0/1) | 0.040 | -0.165\*\*\* | 0.126\* | -0.1333\* |
| AIC | 31236 | 30422 | 12513 | 22132 |

\*\*\*p-value<0.001, \*\*p-value<0.01, \*p-value<0.05

Table X. Logistic models of crimes known to police fitted from CSEW 2011/12 data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Vehicle crimes | Residence crimes | Theft and property crimes | Violent crimes |
| (Intercept) | -0.466\*\*\* | -0.808\*\*\* | -0.854\*\*\* | -0.242 |
| Age | -0.003 | 0.001 | 0.005 | 0.003 |
| Male (0/1) | -0.099 | 0.094 | -0.282\* | -0.259\*\* |
| White (0/1) | 0.106 | 0.127 | 0.064 | -0.034 |
| Not working (0/1) | -0.174\* | 0.189\*\* | -0.002 | 0.026 |
| High education (0/1) | 0.071 | 0.044 | 0.278\* | -0.189\* |
| AIC | 5868.2 | 5105.3 | 1717.1 | 3163.1 |

\*\*\*p-value<0.001, \*\*p-value<0.01, \*p-value<0.05

## 4.2 Empirical evaluation of simulated dataset of crimes

EXPLAIN SOURCE OF POLICE DATA (POLICE.CO. UK)

ADD MAPS OF CRIMES KNOWN TO GMP COMPARED TO OUR SIMULATED CRIMES (KNOWN TO POLICE) + SPEARMAN CORRELATION

## 4.3 Assessing the results

How and why we calculate relative different

# 5. Mapping the bias of police-recorded crimes

PRESENT RESULTS OF THE SIMULATION STUDY

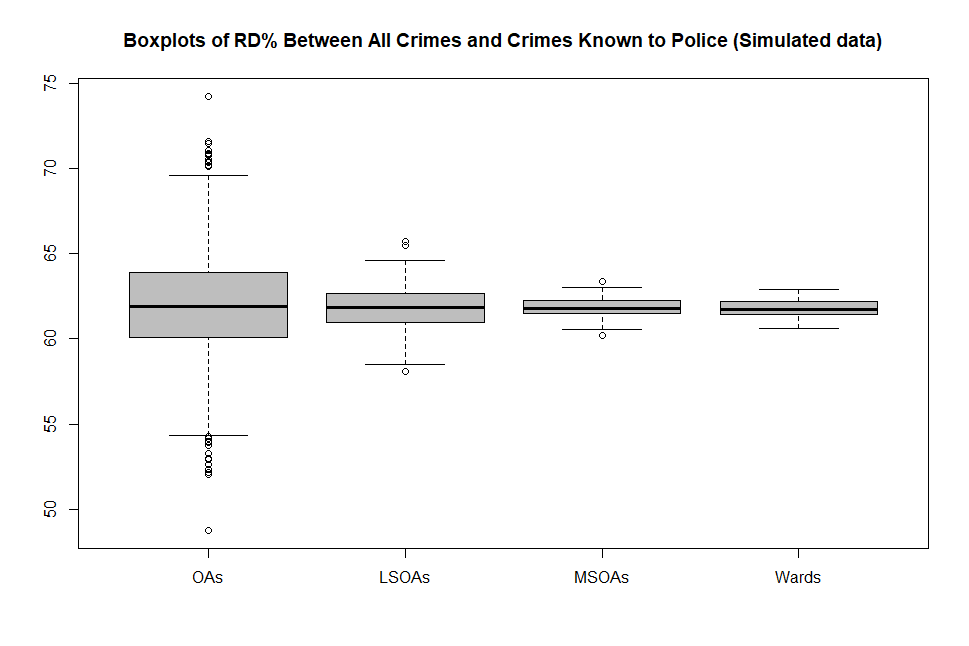
This simulation study is designed to study the bias of crime estimates at small geographical level obtained by using police records only. In particular, the estimates are related to the Output Area (OA) levels in Greater Manchester.

Two main biases are investigated here: the bias across social groups in the population and the bias across geographical areas. A synthetic population

Bias across social groups

and bias across geographical scales

XXX



# 6. Discussion

XXX

# 7. Conclusions

XXX

# References

Baumer, E. P., & Lauritsen, J. L. (2010). Reporting crime to the police, 1973-2005: A multivariate analysis of long-term trends in the National Crime Survey (NCS) and National Crime Victimization Survey (NCVS). *Criminology, 48*(1), 131-185.

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

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.

Groff, E. R., Weisburd, D., & Yang, S. M. (2010). Is it important to examine crime trends at a local “micro” level?: A longitudinal analysis of street to street variability in crime trajectories. *Journal of Quantitative Criminology, 26*, 7-32.

Hart, T. C., & Rennison, C. (2003). *Reporting crime to the police, 1992-2000*. Special Report, Bureau of Justice Statistics.

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 Place.

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.

Wortley, R., & Townsley, M. (Eds.) (2017). *Environmental criminology and crime analysis*. Second edition. Abingdon: Willan Publishing.

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