

Recent Advances in Estimating The Criminal Population

Dr Razik Ridzuan Mohd Tajuddin and Dr Noriszura Ismail

Research Article

ABSTRACT

Studies estimating the criminal population have recently attracted researchers. A summary of recent advances in estimating the criminal population is investigated via bibliometric analysis. The bibliometric analysis aims to identify the trends of publications and the top relevant sources and key authors involved in estimating the size of the criminal population. A total of 55 documents published between 2001 and 2022 were retrieved from the Scopus database and considered in the analysis. Most of the documents were articles written in English by first-time authors. In 22 years, only two articles received more than 100 citations from the Scopus database. The results from the bibliometric analysis revealed that the idea of estimating the size of the criminal population is still vibrant and eye-catching. The findings can be used as a reference for future article submissions.

Keywords: *bibliometrix; capture-recapture; criminal population size; Horvitz-Thompson; offenders; prevalence*

INTRODUCTION

The size of the criminal population consists of observed and unobserved criminals, or in other words, captured and uncaptured criminals. Uncaptured criminals comprise a hidden population of criminals who fail to be captured by the authorities. The total population of active criminals for a specific crime can be determined by identifying and estimating the number of uncaptured criminals. Rossmo and Routledge (1990) emphasised the importance of knowing the population size of criminals as it helps to formulate new and up-to-date laws on specific crimes. Moreover, estimating the size of the criminal population is a crucial issue that has attracted researchers, especially in the twenty-first century.

Estimating the size of the criminal population interrelates with the capture-recapture framework adopted from biostatistics and ecology. In criminology, offenders or criminals are first captured (arrested) and, after a successful prosecution, face punishment for their crimes. They

will be released into the community and may be recaptured (rearrested) if they commit crimes again. Generally, recent research documents estimate the population of criminals in the context of prevalence (De Angelis et al., 2004; Degenhardt et al., 2016; Friedman et al., 2004) or by using an estimator proposed by Horvitz-Thompson (Böhning et al., 2004; Rossmo & Routledge, 1990; Tajuddin et al., 2022; Van Der Heijden et al., 2003). Following the work of Horvitz and Thompson (1952), several authors have studied the importance of incorporating the effect of one-inflation (a large number of cases coming from a one-time arrest) in criminological settings. There are several justifications where the rise in one-inflation is standard in criminological data – the ability and desire to avoid getting recaptured (Godwin & Böhning, 2017), which may be due to community programs (Levine & Dandamudi, 2016) or previous information from other convicts (Rossmo & Routledge, 1990). Notably, this phenomenon is closely related to the prominent deterrence theory. Furthermore, data pollution (Nolan et al., 2011) due to the miscalculation or misclassification of an offence may also contribute to one-inflation in criminology datasets.

This study offers a meta-analysis review via a bibliometric approach focusing on recent advances in estimating the criminal population. The bibliometric approach is important in a scientific study because it assesses research trends, evaluates research impact, identifies key contributors, maps knowledge networks, and evaluates the productivity of the contributors. Specifically, the following research questions (RQ) are considered in this study and can be answered using a bibliometric approach.

RQ 1: What is the current publication trend for the published documents related to the estimated criminal population?

RQ 2: How are the published documents related to the estimated criminal population distributed in terms of the types of documents, affiliations by country, and language of the documents?

RQ 3: Which journals serve as a key platform for publishing documents related to the estimated criminal population?

RQ 4: Who are the most prolific authors publishing documents related to the estimated criminal population?

RQ 5: What are the most popular documents related to estimating the criminal population based on citations?

RQ 6: What keywords are commonly used in research when estimating the criminal population?

This article is structured as follows. Section 2 explains the data acquisition process and the corresponding conditions for retrieving data. Section 3 answers the research questions in detail through the bibliometric analysis. Finally, section 4 provides concluding statements regarding this study.

DATA AND METHODS

Data from the Scopus database were retrieved to achieve the stated objectives. The Scopus database was selected due to its extensive collection of quality journals and wide selection of research articles, review articles, and conference papers. The data were retrieved from the Scopus database on 25 July 2023 and converted into CSV and BibTeX formats. Table 1 shows the data retrieval process using the Scopus database, which consists of three phases – initial retrieval, Screening I, and Screening II. In the initial retrieval phase, the TITLE field code only allows documents focusing on estimating the criminal population. Nevertheless, criminals may be referred to with different labels, such as prisoners, offenders or felons, or be defined differently according to their crimes, such as rapists or addicts. Therefore, these synonyms were also included in the query (see Table 1). This query resulted in 164 documents from as early as 1967 to the latest from 2023.

If we had used the TITLE-ABS field code, which searches documents based on the query string discovered in the title and abstract, more documents could have been obtained. However, the documents may not be focused on estimating the size of the criminal population. A manuscript by Godwin and Böhning (2017) perfectly explains why using TITLE-ABS may not be the best option. Godwin and Böhning (2017) studied the development of the Horvitz-Thompson estimator based on the capture-recapture framework and excess ones in the captured subjects, which is also known as one-inflation. Generally, excess ones refer to the presence of one-valued observations which is failed to be adequately described by an ordinary Poisson distribution. In criminology, excess ones translate to a considerable increase in first-time offenders beyond what a Poisson distribution can accurately estimate.

However, estimating the population of criminals is one of many examples of datasets being used to convince readers of the capability of the estimator proposed in different settings. In addition, Tajuddin et al.'s (2022) work may have been partly based on Godwin and Böhning's (2017) research. However, the authors were fully focused on issues in criminology. Hence, Godwin and Böhning (2017) are not included in the 164 documents, yet Tajuddin et al. (2022) are included.

Phase	Field code	Query or prompt	Number of documents
Initial Retrieval	TITLE	("Estimat*") AND ("Population" OR "Size*" OR "Number*" AND "Crim*" OR "Abuse*" OR "User*" OR "Felon*" OR "Prison*" OR "Addict*" OR "Rapist*" OR "Arrest*" OR "Offender*" OR "Delinquen*")	164
Screening I	TITLE	("Estimat*") AND ("Population" OR "Size*" OR "Number*" AND "Crim*" OR "Abuse*" OR "User*" OR "Felon*" OR "Prison*" OR "Addict*" OR "Rapist*" OR "Arrest*" OR "Offender*" OR	120

	"Delinquen*") AND PUBYEAR > 2000 AND PUBYEAR < 2023
Screening II –	Manual filter from reading the title and abstract of the documents from the 55 Screening I phase

Table 1: Field code and query for data retrieval.

To align with the objectives, which focus on recent advances, the data underwent several screening phases before being used in the analysis. In the Screening I phase, articles published from 2001 until 2022 were filtered (see Table 1). Any documents published in 2023 were not included, since 2023 had not ended as of the date of retrieval, 25 July 2023. Screening I phase resulted in 120 documents. The initial query used, such as "User*," may indicate internet users or traffic users and "Crime*" may include crime mapping. Besides that, some documents discuss crime proceeds and costs and medical-related studies. Therefore, these 120 documents were further screened manually in Screening II phase to remove any irrelevant articles based on the titles and abstracts of the documents. Screening II phase resulted in 55 documents, which were used for the bibliometric analysis in Section 3. To represent global participation in advancing the field of study, the variability in documents based on different languages and article types, was not restricted.

RESULTS AND FINDINGS

An overview of the findings based on the 55 documents is discussed in this section. For the bibliometric analysis, the bibliometrics package developed by Aria and Cuccurullo (2017) using R Software (R Core Team, 2022) will be used with the help of the Biblioshiny app.

RQ 1 – Publication trend

This research question explores the publication trend from 2001 to 2022. It is important to identify the trend so that the relevance and importance of the research niche can be assessed. Figure 1 displays a fluctuating pattern in annual scientific publications from 2001 to 2022. There is a slight drop in 2020 and 2021, probably attributed to the research community being more focused on COVID-19 studies. The highest frequency of documents was published in 2016. Based on the lack of a significant decrease in publications, it can be inferred that the field of estimating the criminal population remains a viable and worthwhile area of study.

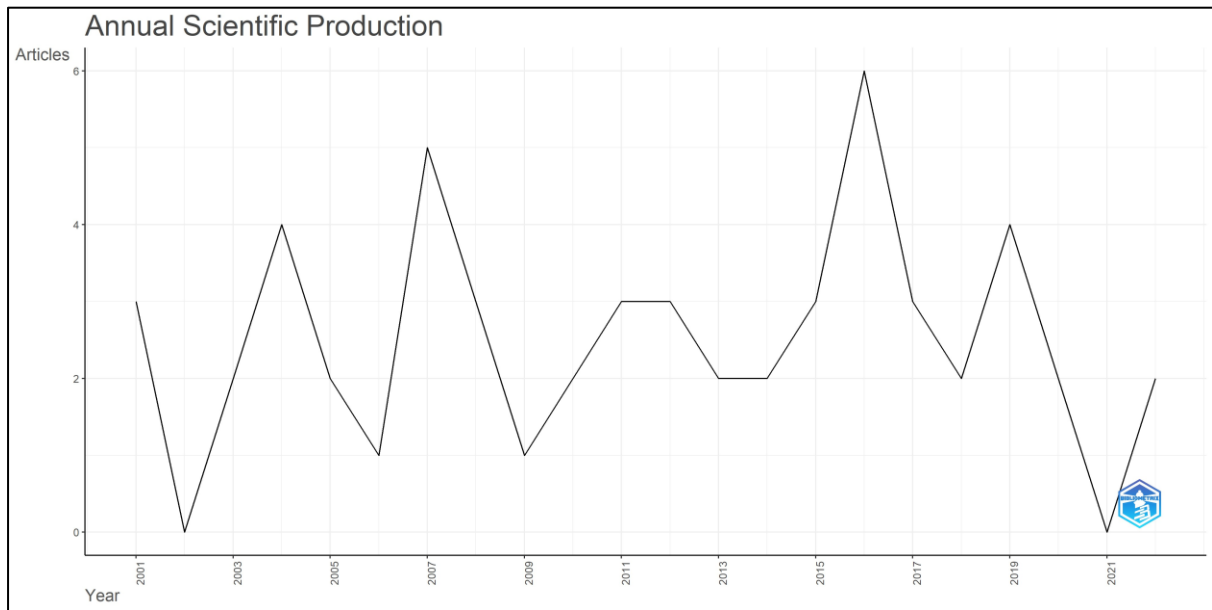


Figure 1: Annual scientific publications. Source: Scopus and Biblioshiny.

RQ 2 – Distribution by type, language, frequency of authors and affiliation

This research question examines the distributions of scientific documents based on their type, the language used, the frequency of publication by the authors, and the countries of publication in the area of estimating the criminal population. Tables 2–5 describe the number of documents by type, language, number of authors, and top affiliated countries, respectively. From Table 2, articles comprise 92.74% of the total documents in this study, followed by conference papers and letters, which are 3.63% each.

Types of documents	Number of documents	Percentage (%)
Articles	51	92.72
Conference papers	2	3.64
Letters	2	3.64
Total	55	100

Table 2: Number of documents by type. Source: Scopus and Biblioshiny.

From Table 3, most of the documents (90.90%) were written in English, followed by German (3.64%) and equally by Chinese, Japanese, and Dutch, each 1.81%.

Language of documents	Number of documents	Percentage (%)
English	50	90.90
German	2	3.64
Chinese	1	1.82
Japanese	1	1.82
Dutch	1	1.82
Total	55	100

Table 3: Number of documents by language. Source: Scopus and Biblioshiny.

From Table 4, most authors (92.27%) were involved and responsible for only one document. Only three authors were responsible for three documents – Degenhardt, Friedman, and Hickman (see Table 6).

Number of documents	Number of authors	Percentage (%)
1	215	92.27
2	15	6.44
3	3	1.29
Total	233	100

Table 4: Number of documents by number of authors. Source: Scopus and Biblioshiny.

From Table 5, most of the documents are published by researchers affiliated with the United States of America (USA), followed by Australia, the Netherlands, the United Kingdom, and Canada.

Countries	Number of documents
USA	12
Australia	5
Netherlands	5
United Kingdom	5
Canada	3

Table 5: Top 5 affiliation countries with the most documents. Source: Scopus and Biblioshiny

RQ 3 – Popular journals for publications

This research question seeks to identify the most accepted publication outlets for publishing documents on estimating the criminal population. By identifying the most popular journals for this field, researchers can target their work to these journals. The most popular publishing outlets in the field of estimating the criminal population are the *International Journal of Drug Policy* and *Journal of Urban Health*, followed by the *Addiction* and *Medical Journal of Australia*, as shown in

Figure 2. In the twenty-first century, both the *International Journal of Drug Policy* and the *Journal of Urban Health* have published three documents related to estimating the criminal population.

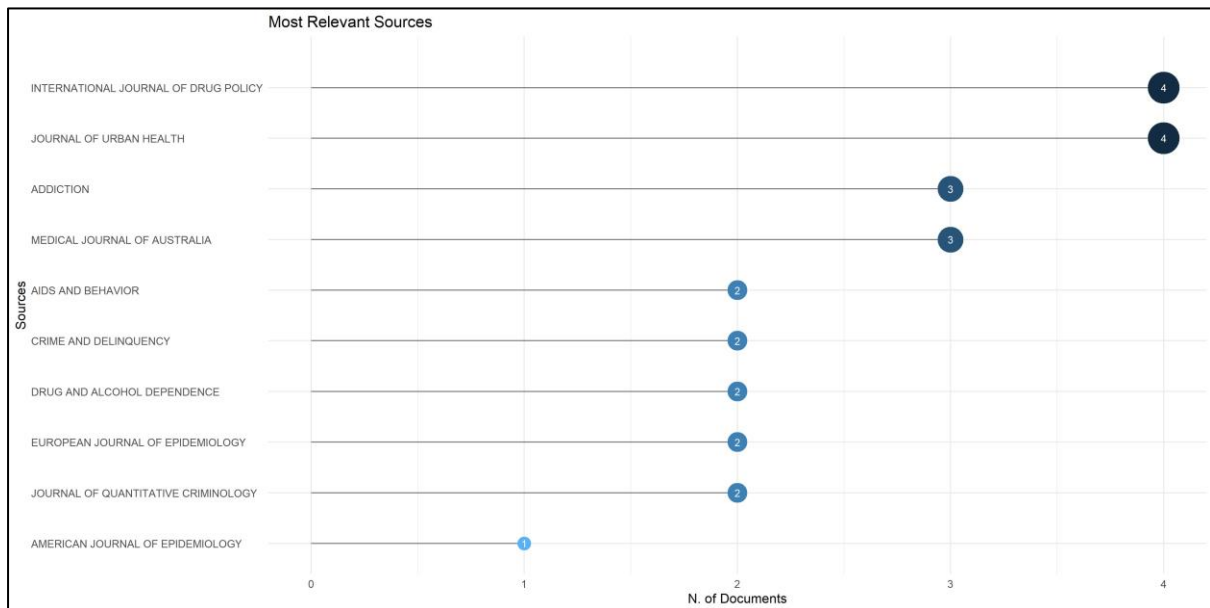


Figure 2: Top 10 most relevant sources and their associated number of documents. Source: Scopus and Biblioshiny.

RQ 4 – Prolific authors

This research question seeks to pinpoint the prolific authors who have published impactful research documents from 2001 until 2022. Budding researchers may approach these authors for research collaborations. The top five most relevant authors based on the number of documents, both total and fractionalised, are the same, yet they are ranked differently, as indicated in Table 6. Degenhardt, Friedman, and Hickman each produced three documents from 2001 until 2022. However, Bouchard has the highest fractionalised number of documents. This indicates that Bouchard produced impactful research with the least number of co-authors. This is supported by the number of citations gained from an article by Bouchard in 2007 (see Section 3.5).

Name of authors	Number of documents (total)	Number of documents (fractionalised)
Degenhardt	3	0.66
Friedman	3	0.45
Hickman	3	0.73
Bouchard	2	1.17
Buster	2	0.83

* Highest number of documents (total and fractionalised) are in bold

Table 6: Top 5 most relevant authors. Source: Scopus and Biblioshiny

RQ 5 – Popular documents

This research question determines the top-cited scientific documents and the corresponding publishing outlet. It is also another way of identifying potential collaborators and publication outlets. The top 10 most cited documents are tabulated in Table 7. Only two documents considered in this study received more than 100 citations in the Scopus database. However, it is essential to note that although Bouchard's work was published in 2007, three years after Friedman et al.'s (2004) work was published, Bouchard (2007) has only four fewer citations than Friedman et al. (2004). Furthermore, two relatively recent research documents by Degenhardt et al. (2016) and Nikfarjam et al. (2016) captured a relatively high number of citations in a short period.

A comparison of total citations for the top 10 most cited documents in the Google Scholar (GS) and Web of Science (WoS) databases is presented in Table 7. The current CiteScore (CS) from the Scopus index and the Impact Factor (IF) from the WoS database are also included. From the WoS database, only one document, Friedman et al. (2004), received over 100 citations. However, from the GS database, six documents (Bouchard, 2007; Degenhardt et al., 2016; Friedman et al., 2004; Nikfarjam et al., 2016; Van Der Heijden et al., 2003; Wejnert et al., 2012) obtained over 100 citations. Generally, these articles are in the top 10 most cited documents category because they are published in high CS and IF journals.

Name of authors	Name of journal (CS, IF)	Scopus citations	WoS citations	GS citations
Friedman et al. (2004)	Journal of Urban Health (CS: 8.7, IF: 6.6)	108	106	135
Bouchard (2007)	Journal of Quantitative Criminology (CS: 7.2, IF: 3.6)	104	71	180
Wejnert et al. (2012)	AIDS and Behavior (CS: 6.3, IF: 4.4)	82	78	120
Nikfarjam et al. (2016)	International Journal of Drug Policy (CS: 7.4, IF: 4.4)	80	70	117
Degenhardt et al. (2016)	Medical Journal of Australia (CS: 8.5, IF: 11.4)	72	69	112
De Angelis et al. (2004)	American Journal of Epidemiology	71	64	97

(CS: 7.0, IF: 5.0)				
Böhning et al. (2004)	European Journal of Epidemiology (CS: 18.6, IF: 13.6)	53	51	92
Okal et al. (2013)	Sexually Transmitted Infections (CS: 5.9, IF: 3.6)	52	51	84
Van Der Heijden et al. (2003)	Statistica Neerlandica (CS: 2.1, IF: 1.5)	49	30	101
Law et al. (2001)	Addiction (CS: 10.1, IF: 6.1)	43	48	75

Table 7: Top 10 most cited documents. Source: Scopus, Biblioshiny, WoS, and GS.

RQ 6 – Popular keywords

This final research question focused on the most popular keywords used by the authors to represent the main concepts of their research. These keywords highlight areas of study and methodologies used by the researchers. Figure 3 illustrates the word cloud for the frequently used keywords by the authors. It shows that capture-recapture and population size estimation are the keywords most used in the studies. The relationship between the capture-recapture framework and population size estimation in criminology cannot be denied. This is attributed to most studies using the framework as a base to develop models and formulas for estimating the size of the criminal population (Böhning et al., 2004; Bouchard, 2007; Godwin & Böhning, 2017; Rossmo & Routledge, 1990; Tajuddin et al., 2022; Van Der Heijden et al., 2003).

Besides capture-recapture and population size estimation keywords, the keywords injection/injecting drug users are studied by the authors, as this niche population relates to many illnesses. This includes the Human Immunodeficiency Virus (HIV) and Hepatitis C Virus (HCV) (Caiaffa et al., 2003; Cathcart et al., 2007; Johnston et al., 2011; Kolarić et al., 2010; Lieb et al., 2004; McDonald et al., 2014; Okal et al., 2013; Sutton et al., 2008; Wand et al., 2009).

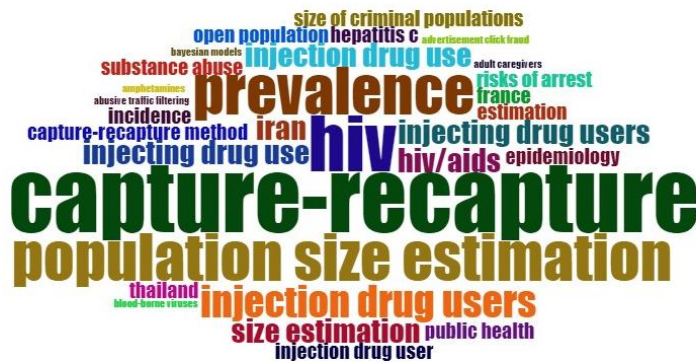


Figure 3: Word cloud for the frequently used keywords. Source: Scopus and Biblioshiny.

CONCLUSIONS

From the bibliometric analysis using the Biblioshiny app, it is apparent that estimating the size of the criminal population has often been studied in the twenty-first century (until 2022). Based on the 55 documents retrieved from the Scopus database, it was discovered that most of the documents were articles written in English by first-time authors in this area. A good portion of the documents were by USA-based authors.

Researchers in this area should consider the most relevant sources and the top authors' publication records to decide which journal to target. This bibliometric analysis only used the Scopus database and could be further improved by also considering the WoS, GS, and other databases. However, caution must be exercised, especially when calculating the number of citations, since Scopus is a more extensive database than WoS, yet smaller than GS (refer to Table 7). Thus, the number of citations may differ from one source to another. Hence, a systematic literature review may also be conducted for further investigation.

Also, since the capture-recapture technique has been consistently used in estimating the size of the criminal population, researchers are advised to explore this further in diverse settings. This includes different crimes, races, and age groups, and investigating whether the capture-recapture frameworks still work in targeted groups and if the previous models introduced are relevant and useful.

ACKNOWLEDGEMENTS

This study was funded by Geran Galakan Penyelidik Muda from Universiti Kebangsaan Malaysia (Grant no.: GGPM-2023-052). The authors are also thankful for the constructive comments from the editor and reviewers. The authors have no conflicts of interest.

About the authors:

Razik Ridzuan Mohd Tajuddin (PhD) is a Senior Lecturer in Universiti Kebangsaan Malaysia. His research focus is on investigating spatiotemporal effects of narcotics crime in Malaysia. He applies quantitative techniques such as statistical modelling on crime data including estimating population size of criminals. He is interested to work with law enforcement agencies such as police, prison and rehab units related to narcotics offences with an aim to create a narcotics-free society.

Noriszura Ismail (PhD) is a Professor in Universiti Kebangsaan Malaysia. Her research focus is on the applications of statistical knowledge including theoretical works as well as applications to wide area of studies such as quantitative criminology, actuarial science, and risk management. She also works on investigating spatiotemporal effects of narcotics crime in Malaysia.

REFERENCES

- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959-975.
- Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., & Viwatwongkasem, C. (2004). Estimating the number of drug users in Bangkok 2001: A capture–recapture approach using repeated entries in one list. *European Journal of Epidemiology*, *19*, 1075-1083.
- Bouchard, M. (2007). A capture–recapture model to estimate the size of criminal populations and the risks of detection in a marijuana cultivation industry. *Journal of Quantitative Criminology*, *23*, 221-241.
- Caiaffa, W. T., Mingoti, S. A., Proietti, F. A., Carneiro-Proietti, A. B., Silva, R. C., Lopes, A. C. S., & Doneda, D. (2003). Estimation of the number of injecting drug users attending an outreach syringe-exchange program and infection with human immunodeficiency virus (HIV) and hepatitis C virus: the AjUDE-Brasil project. *Journal of Urban Health*, *80*, 106-114.
- Cathcart, S., Patton, R., Hickman, M., Hendry, C., & Ashworth, M. (2007). Comparing population estimates for drug users, the prevalence of blood-borne viruses and the provision of services in one London borough: Evidence of a gap in provision? *Drugs: Education, Prevention and Policy*, *14*(3), 205-215.
- De Angelis, D., Hickman, M., & Yang, S. (2004). Estimating long-term trends in the incidence and prevalence of opiate use/injecting drug use and the number of former users: back-calculation methods and opiate overdose deaths. *American Journal of Epidemiology*, *160*(10), 994-1004.
- Degenhardt, L., Larney, S., Chan, G., Dobbins, T., Weier, M., Roxburgh, A., Hall, W. D., & McKetin, R. (2016). Estimating the number of regular and dependent methamphetamine users in Australia, 2002–2014. *Medical Journal of Australia*, *204*(4), 153-153.
- Friedman, S. R., Tempalski, B., Cooper, H., Perlis, T., Keem, M., Friedman, R., & Flom, P. L. (2004). Estimating numbers of injecting drug users in metropolitan areas for structural analyses of community vulnerability and for assessing relative degrees of service provision for injecting drug users. *Journal of Urban Health*, *81*, 377-400.
- Godwin, R. T., & Böhning, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *66*(2), 425-448.
- Horvitz, D. G., & Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663-685.
- Johnston, L., Saumtally, A., Corceal, S., Mahadoo, I., & Oodally, F. (2011). High HIV and hepatitis C prevalence amongst injecting drug users in Mauritius: findings from a population size

estimation and respondent driven sampling survey. *International Journal of Drug Policy*, 22(4), 252-258.

Kolarić, B., Štajduhar, D., Gajnik, D., Rukavina, T., & Wiessing, L. (2010). Seroprevalence of blood-borne infections and population sizes estimates in a population of injecting drug users in Croatia. *Central European Journal of Public Health*, 18(2).

Law, M. G., Lynskey, M., Ross, J., & Hall, W. (2001). Back-projection estimates of the number of dependent heroin users in Australia. *Addiction*, 96(3), 433-443.

Levine, J. A., & Dandamudi, K. (2016). Prevention of child sexual abuse by targeting pre-offenders before first offense. *Journal of Child Sexual Abuse*, 25(7), 719-737.

Lieb, S., Friedman, S. R., Chitwood, D. D., Liberti, T. M., Gates, G. J., Metsch, L. R., Maddox, L. M., & Kuper, T. (2004). An HIV prevalence-based model for estimating urban risk populations of injection drug users and men who have sex with men. *Journal of Urban Health*, 81, 401-415.

McDonald, S. A., Hutchinson, S., Schnier, C., McLeod, A., & Goldberg, D. (2014). Estimating the number of injecting drug users in Scotland's HCV-diagnosed population using capture–recapture methods. *Epidemiology & Infection*, 142(1), 200-207.

Nikfarjam, A., Shokoohi, M., Shahesmaeili, A., Haghdoost, A. A., Baneshi, M. R., Haji-Maghsoudi, S., Rastegari, A., Nasehi, A. A., Memaryan, N., & Tarjoman, T. (2016). National population size estimation of illicit drug users through the network scale-up method in 2013 in Iran. *International Journal of Drug Policy*, 31, 147-152.

Nolan, J. J., Haas, S. M., & Napier, J. S. (2011). Estimating the impact of classification error on the “statistical accuracy” of uniform crime reports. *Journal of Quantitative Criminology*, 27, 497-519.

Okal, J., Geibel, S., Muraguri, N., Musyoki, H., Tun, W., Broz, D., Kuria, D., Kim, A., Oluoch, T., & Raymond, H. F. (2013). Estimates of the size of key populations at risk for HIV infection: men who have sex with men, female sex workers and injecting drug users in Nairobi, Kenya. *Sexually Transmitted Infections*, 89(5), 366-371.

R Core Team. (2022). *R: A language and environment for statistical computing*. In <https://www.R-project.org/>

Rossmo, D. K., & Routledge, R. (1990). Estimating the size of criminal populations. *Journal of Quantitative Criminology*, 6, 293-314.

Sutton, A., Hope, V., Matheï, C., Mravcik, V., Sebakova, H., Vallejo, F., Suligoï, B., Brugal, M., Ncube, F., & Wiessing, L. (2008). A comparison between the force of infection estimates for blood-borne viruses in injecting drug user populations across the European Union: a modelling study. *Journal of Viral Hepatitis*, 15(11), 809-816.

Tajuddin, R. R. M., Ismail, N., & Ibrahim, K. (2022). Estimating Population Size of Criminals: A New Horvitz–Thompson Estimator under One-Inflated Positive Poisson–Lindley Model. *Crime & Delinquency*, 68(6-7), 1004-1034.

Van Der Heijden, P. G., Cruyff, M., & Van Houwelingen, H. C. (2003). Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statistica Neerlandica*, 57(3), 289-304.

Wand, H., Spiegelman, D., Law, M., Jalaludin, B., Kaldor, J., & Maher, L. (2009). Estimating population attributable risk for hepatitis C seroconversion in injecting drug users in Australia: implications for prevention policy and planning. *Addiction*, 104(12), 2049-2056.

Wejnert, C., Pham, H., Krishna, N., Le, B., & DiNenno, E. (2012). Estimating design effect and calculating sample size for respondent-driven sampling studies of injection drug users in the United States. *AIDS and Behavior*, 16, 797-806.