Australian Government

Australian Institute of Criminology

AIC reports
Research Report 30

Predicting high-harm offending using national police information systems: An application to outlaw motorcycle gangs

Timothy Cubitt and Anthony Morgan

Celebrating 50 years

ISSN (Online) 2206-7280
ISBN 978 1 922877 32 1 (Online)
https://doi.org/10.52922/rr77321

Published by the Australian Institute of Criminology
GPO Box 1936 Canberra ACT 2601
Tel: (02) 6268 7166
Email: front.desk@aic.gov.au
Website: www.aic.gov.au

Please note: Minor revisions are occasionally made to publications after release. The online versions available on the AIC website will always include any revisions.

All publications in the Research Report series are subject to peer review–either through a double-blind peer review process, or through stakeholder peer review. This report was subject to double-blind peer review.

General editor: Dr Rick Brown, Deputy Director, Australian Institute of Criminology

Edited and typeset by the Australian Institute of Criminology

A full list of publications in the AIC Reports series can be found on the Australian Institute of Criminology website at **aic.gov.au**

Contents

Figures

Tables

Abstract

Risk assessment is a growing feature of law enforcement and an important strategy for identifying high-risk individuals, places and problems. Prediction models must be developed in a transparent way, using robust methods and the best available data. But attention must also be given to implementation. In practice, the data available to law enforcement from police information systems can be limited in their completeness, quality and accessibility. Prediction models need to be tested in as close to real-world settings as possible, including using less than optimal data, before they can be implemented and used. In this paper we replicate a prediction model that was developed in New South Wales to predict high-harm offending among outlaw motorcycle gangs nationally and in other states. We find that, even with a limited pool of data from a national police information system, high-harm offending can be predicted with a relatively high degree of accuracy. However, it was not possible to reproduce the same prediction accuracy achieved in the original model. Model accuracy varied between jurisdictions, as did the power of different predictive factors, highlighting the importance of considering context. There are trade-offs in real-world applications of prediction models and consideration needs to be given to what data can be readily accessed by law enforcement agencies to identify targets for prioritisation.

Executive summary

There have been significant advancements in the data available to law enforcement, including the development of national information systems that can connect databases across jurisdictional borders. This is especially important in the area of serious and organised crime, which is increasingly borderless in nature, requiring agencies to work together to share information about potential targets.

In this report, we examine the replicability of a risk assessment model developed to identify high-harm outlaw motorcycle gang (OMCG) targets in New South Wales (Cubitt & Morgan 2022) using a national police information system. In addition to assessing the predictive accuracy of the model in settings other than the jurisdiction in which the original model was developed, we assess whether it is possible to develop a predictive model with acceptable accuracy using data readily available and accessible to law enforcement agencies.

Method

The sample for this study was drawn from the National Gangs List (NGL), maintained by the Australian Criminal Intelligence Commission (ACIC), which brings together information from state and territory databases into a nationally agreed, secure and validated list of OMCG members. The recorded criminal history for these individuals was extracted from the National Police Reference System, which records the offence history of individuals who have been arrested and subject to some form of legal action by police. Data were available for 5,534 affiliates of OMCGs nationally.

Importantly, while the sample used in this study was much larger than in the earlier study, covering all Australian states and territories rather than focusing on New South Wales, the level of detail about individuals and their offences was less comprehensive.

We used the random forest algorithm to predict high-harm offending, comparing the results of each model with those produced by more traditional logistic regression methods. We performed this analysis first at the national level before repeating the analysis for the largest Australian jurisdictions (in terms of OMCG membership). In addition to the overall predictive accuracy of each model, we report the classification error, along with other relevant metrics.

Results

Overall, 19.1 percent of all OMCG members in the sample had a recorded high-harm offence in the five-year reference period.

The area under the receiver operating characteristic (AUROC) for the national model—an important measure of its predictive accuracy—was 0.801, while the overall classification error was 19.1 percent. An AUROC of 0.8 or higher is generally regarded as excellent. The false positive rate—the proportion of individuals who did not go on to commit a high-harm offence but who were incorrectly identified by the model as high-harm offenders—was just 8.3 percent. The national model was less accurate at identifying affiliates who did go on to commit a high-harm offence, with a false negative rate of 63.4 percent. A false negative occurs when an individual who was predicted to not be a high-harm offender goes on to commit a high-harm offence in the reference period. This means that 36.6 percent of OMCG members who did go on to commit a high-harm offence were predicted by the model to be a high-harm offender. The national model did not perform as well as the models for some larger states, with results from New South Wales—the largest jurisdiction in terms of OMCG membership and the poorest performing model in this instance—reducing the accuracy of the national model.

Model accuracy varied between jurisdictions. The AUROC ranged from 0.774 to 0.862. False positive rates were consistently low, not exceeding 10 percent in any of the states examined. The rate of false negatives was higher, ranging from 46.8 to 69.4 percent. Even in the best performing model, around half the OMCG members who committed a high-harm offence were not predicted to be high-harm offenders. Other intelligence sources are critical to ensure a complete picture of the risk posed by OMCG members.

There were also differences in the relative importance of different predictive factors. This suggests that context needs to be considered in developing predictive models, particularly where operational activity and information recording practices may vary and thus may shape the results of a predictive model.

Discussion

Results from this study show that, even with a limited pool of data from a national police information system, high-harm offending by OMCG members can be predicted with a relatively high degree of accuracy. However, we were unable to reproduce the same prediction accuracy as the original model (Cubitt & Morgan 2022). While false positives were relatively rare, there were higher rates of false negatives than in the original New South Wales model. In other words, the model was better at predicting who would not go on to commit a high-harm offence than who would commit a high-harm offence.

These findings demonstrate that there are trade-offs in real-world applications of prediction models and consideration needs to be given to what data can be readily accessed by law enforcement agencies to identify targets for prioritisation. The models presented in this report may perform an important role in helping to guide other intelligence activity by law enforcement agencies in their efforts to reduce gang-related crime.

Introduction

Risk assessment models are used at different stages of the criminal justice system as a mechanism for identifying high-risk individuals, places or problems. They enable decisions to be made regarding who to prioritise and what action should be taken to mitigate the risk of (especially) recidivist offending. Recently, there has been a shift towards machine learning approaches. These approaches capitalise on the high volume of data captured by criminal justice agencies. At the same time, criminal justice data can be highly variable from jurisdiction to jurisdiction in terms of completeness, quality and accessibility, which poses challenges in terms of developing, replicating and implementing validated risk assessment models, irrespective of the statistical method used. In this report, we examine the replicability of a risk assessment model developed to identify high-harm OMCG targets in New South Wales, Australia (Cubitt & Morgan 2022). We use a random forest algorithm to predict high-harm offending with a larger (in terms of number of observations and jurisdictions covered) but less comprehensive (in terms of number of variables) database on gang members. In addition to assessing the predictive accuracy of the model in settings other than the jurisdiction in which the original model was developed, we assess whether it is possible to develop a predictive model with acceptable accuracy using data from national police information systems readily available and accessible to law enforcement agencies.

Outlaw motorcycle gangs

While their origins in North America can be traced to the 1940s (Barker 2015), OMCGs first appeared in Australia in the 1960s, growing in numbers due to the membership of returned servicemen from the Vietnam War and the expansion of American OMCGs into Australia (Bain & Lauchs 2017; Lauchs 2017). Because of their militaristic origins they follow a strict hierarchical structure and enforce ideals of loyalty, secrecy and brotherhood, typically through the use of violence (Lauchs, Bain & Bell 2015). However, the preponderance of disillusioned servicemen, many of whom struggled to readjust to life at home, meant they also quickly developed a culture of extreme machoism, rebellion and 'barbarianism' (Quinn & Forsyth 2009). There are currently 38 clubs in Australia, including clubs with local (eg Rebels, Comancheros) and international (eg Hells Angels, Bandidos, Gypsy Jokers, Outlaws, Satudarah, Mongols) origins. Irrespective of their origins, these groups share similar characteristics. They are exclusively male, adhere to club rules and are obedient to senior members who hold office and direct the activities of clubs through a hierarchical structure. They expand via regional, self-managed chapters, almost like franchises, and maintain a highly secretive culture and a strong emphasis on loyalty. Members wear recognisable patches to indicate both membership and status within the club and follow a relatively rigid process whereby they enter the club as a probationary member for an extended period before they are accepted as a patched member with full voting rights (Barker 2015; Lauchs, Bain & Bell 2015).

OMCGs are considered a national organised crime threat by Australian law enforcement agencies (ACIC 2017). Public occurrences of violence between clubs, both historical and contemporary, have been the catalyst for legislative reform and the establishment of dedicated police taskforces (Ayling & Broadhurst 2014). OMCG involvement in illicit commodity markets, primarily illicit drug manufacturing, trafficking and distribution, has also attracted significant attention, particularly as it relates to Australia's burgeoning methamphetamine market (ACIC 2017). Australian clubs have also expanded internationally and established new chapters, including in South-East Asia, with the goal of capitalising on lucrative methamphetamine distribution networks (United Nations Office on Drugs and Crime 2019). This has been enabled by an apparent shift in the membership and culture of OMCGs in the Australian context towards younger members who are less motivated by the camaraderie and brotherhood traditionally offered by OMCGs and instead join in pursuit of power, prestige, profit and women, and who have a greater propensity for violent and organised crime (Dowling et al. 2021; Lauchs 2017; Voce, Morgan & Dowling 2021).

While they are regarded as an organised crime threat, OMCGs are distinct from other organised crime groups. Although they are classified as gangs, they also differ from more traditional street gangs in a number of important ways. Von Lampe and Blokland (2020) conceptualise OMCGs, street gangs and organised crime groups along two continuums: the motivation for their crime, ranging from entrepreneurial (or profit-motivated) to achieving symbolic goals; and their organisational structure, ranging from informal-diffused (less organised) to instrumental-rational (highly organised). OMCGs are involved in both entrepreneurial and symbolic crimes—like street gangs, they are focused on maintaining the collective identity of their groups through gang insignia and they engage in violence to protect territory and reputations (von Lampe & Blokland 2020). At the same time OMCGs, like organised crime groups, are involved in entrepreneurial crimes, particularly the supply and distribution of illicit commodities, for economic gain (Lauchs, Bain & Bell 2015). In terms of organisational structure, OMCGs are distinct from most organised crime groups and street gangs, which tend to have informal-diffused structures, and are more like mafias, with their rigid hierarchical structure and clearly defined membership boundaries (von Lampe & Blokland 2020). Though studies have shown the relationship between the organisational structure of OMCGs and criminal offending is more complex than this suggests (Bright & Deegan 2021; van Deuren, Kleemans & Blokland 2020), it nonetheless serves to demonstrate how OMCGs differ from street gangs and organised crime groups.

That said, there is compelling evidence that Australian and international OMCG members engage in both organised crime and gang-related violence. Australian and international studies show a high prevalence of criminal involvement among OMCG members (Blokland et al 2019; Klement 2016; Rostami & Mondani 2017; Tremblay et al 1989). This includes involvement in organised crime-type offending, although there is considerable heterogeneity between clubs (Barker 2015; Blokland, Soudijn & van der Leest 2017; Bright et al. 2022; Bright et al. 2023; Morgan, Dowling & Voce 2020) and evidence that certain individual and club-level factors are associated with a greater involvement in organised crime-type offences (Morgan, Dowling & Voce in press). Recorded violence is also common (Morgan, Cubitt & Dowling 2023), whether it is simply a function of the outlaw lifestyle or results from intra- and inter-club disputes relating to territorial expansions (Cubitt, Dowling & Morgan 2023), building and maintaining gang reputation, and protection and extortion activities (Lauchs, Bain & Bell 2015). Importantly, similar to other criminal groups (Ratcliffe & Kikuchi 2019), there is evidence that serious offending by OMCGs, including organised crime-type offending and serious violence, is highly concentrated among a relatively small proportion of members and chapters (Morgan, Dowling & Voce 2020). This has led to calls for a more nuanced policy response to OMCGs that reflects differences in the degree of member and club involvement in serious and organised criminal activity (Lauchs, Bain & Bell 2015; von Lampe and Blokland 2020).

While there has been a proliferation of new legislation targeting OMCGs that aims to restrict members' ability to interact with one another, prevent the exploitation of key industries and remove club infrastructure such as clubhouses and club insignia (Bartels, Henshaw & Taylor 2021), the targeting of high-risk OMCGs based on criminal intelligence is a key feature of the law enforcement response to OMCGs. There are dedicated police operations and taskforces targeting OMCGs both nationally and in each state and territory, supported by analyst capability to produce strategic and tactical intelligence to inform operational responses.

Gang databases

Police agencies in each Australian state and territory maintain a database or a list of known OMCG affiliates, along with information about these individuals. There is also a National Gangs List (NGL), maintained by the ACIC, which brings together information from these state and territory databases into a nationally agreed, secure and validated list of OMCG members. This is consistent with a growing shift internationally towards maintaining national information systems, which are better equipped to manage information on individuals and offences that move across jurisdictional boundaries (Phythian & Kirby 2022). This is especially important with OMCGs, with many clubs spanning multiple states (Morgan, Dowling & Voce in press) and with criminally mobile members (Dowling & Morgan 2021).

The use of gang databases has become the subject of considerable debate, including in the United States (Huff & Barrows 2015; Kennedy 2009) and the United Kingdom (Densley & Pyrooz 2020). Criticisms of the approach primarily focus on issues related to defining gangs and inclusion criteria for individual members, and the resulting potential for the under-identification and over-identification of members (Huff & Barrows 2015). The latter issue poses a particular threat to civil liberties, especially given the over-representation of racial and ethnic minorities (Densley & Pyrooz 2020). While the legislative response to OMCGs has certainly attracted criticism in Australia (Bartels, Henshaw & Taylor 2021), gang databases have not attracted the same scrutiny. There are important differences between OMCGs and other gangs in terms of the ability to define what constitutes a gang, how easy it may be to validate group membership, how enduring membership periods are, the racial and ethnic background of members, and groups' historical origins that may mitigate some of the risks of using databases in this context. Importantly, the NGL is governed by a clear definition of OMCGs and strict inclusion criteria. OMCG members are known to pass through a rigid, protracted recruitment process, wear visible signs of membership and remain members of clubs for an extended period (Boland et al. 2021; Lauchs, Bain & Bell 2015). These factors make it easier to identify members, although recent changes in recruitment practices (Dowling et al. 2021) and legislation prohibiting the wearing of club insignia has undoubtedly made it more difficult to track membership. That is not to suggest that the management of databases on OMCGs has not attracted controversy, but this has primarily focused on issues of accuracy and management (Goldsworthy & McGillivray 2017). Likewise, Australian police have been criticised for their management of group-involved violent crime involving young people, including their characterisation as gangs, management of offender lists and proactive enforcement targeting young people (Yeong 2021).

Risk assessment

Though not infallible, one mechanism through which concerns regarding these databases may be addressed is by incorporating structured, validated risk assessment, developed using robust and transparent methods. While violence risk assessment is common (Singh, Grann & Fazel 2011), including in relation to gang violence (Guay 2012; Valasik 2018), risk and threat assessments for organised crime have been more reliant on the subjective assessment of law enforcement officers and analysts (Albanese 2008; Ratcliffe, Strang & Taylor. 2014). The application of machine learning analytics to risk assessment is a relatively recent development in criminology, but there are a growing number of applications to law enforcement (Berk 2021; Cubitt, Wooden & Roberts 2020). There is evidence that quantitative risk assessment outperforms professional judgement in terms of accurately forecasting reoffending, that algorithms are at least as accurate as model-based alternatives, and that the process of applying machine learning can be more transparent than the cognitive processes of individual decision-makers (Berk 2021).

While there has been significant advancement in terms of the development of validated risk assessment models for use in criminal justice applications, insufficient attention has been paid to their effective implementation. Despite the effort that has gone into developing information systems within which a risk assessment process may be integrated (Phythian & Kirby 2022), we know there are technological, organisational, leadership and cultural factors that prevent those systems from being used (Koper, Lum & Willis 2014; Lum, Koper & Willis 2017), impeding efforts to implement intelligence-led policing (Darroch & Mazerolle 2013; Sanders, Weston & Schott 2015). Crime analysts frequently identify problems accessing data with which to prepare intelligence products (O'Connor et al. 2021), which may prevent them from undertaking more sophisticated analyses (O'Connor et al. 2022). Issues relating to the quality of information create inefficiencies for crime analysts, who must invest a lot of time in trying to clean and link data for the purpose of preparing intelligence products (Huey, Ferguson & Koziarski 2022). Gang databases are not immune to these problems (Densley & Pyrooz 2020). It is imperative that these data quality and access issues are considered during the development of any new risk assessment models. This is the focus of the current study.

Predicting high-harm offending among outlaw motorcycle gang members

Cubitt and Morgan (2022) developed a risk assessment model using machine learning to identify high-harm OMCG targets in New South Wales, Australia. In addition to developing a predictive model using machine learning methods, the study also examined whether focusing on recorded high-harm offending, rather than a more indiscriminate focus on any recorded offending, would produce a more accurate model that was less prone to potential criticisms. For example, such a model could address concerns that the results would be biased by proactive policing of OMCGs for low-level public and regulatory offences, or that examining any type of offending, irrespective of the harm it caused, could lead to an over-representation of certain clubs or gang members. The initial study employed an array of law enforcement and custodial data to produce an impressive prediction rate, with an area under the receiver operating characteristic (AUROC) curve of 0.914. The AUROC is an important measure of the predictive accuracy of a risk assessment—the probability that a randomly selected case with a recorded high-harm offence will receive a higher risk rating than a randomly selected individual who did not commit a high-harm offence, Typically, an AUROC greater than 0.6 is regarded as indicative of moderate accuracy, over 0.7 is good, and over 0.8 is excellent (Hosmer & Lemeshow 2004). As well as exhibiting a high AUROC score, the model demonstrated a low rate of both false positives and false negatives. In addition, the rate of false positives was significantly diminished by focusing on high-harm offending, directly addressing possible concerns of a net-widening effect.

While the results from this original study were impressive in terms of the predictive accuracy of the model, there are questions regarding the extent to which it could be operationalised with full fidelity. The original model was developed by linking data from the NSW Police Force database on gang membership with offence data from the Computerised Operational Policing System and custodial data from the NSW Bureau of Crime Statistics and Research Reoffending Database. The dataset included 130 variables. It took several months to negotiate access to these data, to link the data across agencies, and to clean and code the data for analysis. The final dataset used for analysis was unlikely to be available for routine analysis and prediction.

Current study

In this paper, we attempt to reproduce the study by Cubitt and Morgan (2022). We use the random forest algorithm to predict high-harm offending with a larger (in terms of number of observations and jurisdictions covered) but less comprehensive (in terms of breadth of information and number of variables) national database on OMCG affiliates. In addition to assessing the predictive accuracy of the model using a national sample of OMCG affiliates, and then in different jurisdictions, we assess whether it is possible to develop a predictive model with acceptable accuracy using data from national police information systems that are readily available and accessible to law enforcement agencies.

Method

High-harm offending

High-harm offending was operationalised as offences that feature in the top 10 percent of harm as defined by a modified version of the Western Australian Crime Harm Index (WACHI) (Cubitt & Morgan 2022). The WACHI was developed by House and Neyroud (2018) to assign each offence type in the Australian and New Zealand Standard Offence Classification a harm value based on equivalent prison sentences for first-time offenders. This is part of a growing trend towards the measurement of crime harm, rather than relying on offence frequency, as a way of providing a quantifiable proxy for harm to the community (Andersen & Mueller-Johnson 2018; Curtis-Ham & Walton 2018; Kärrholm, Neyroud & Smaaland 2020; Sherman, Neyroud & Neyroud 2016). This is especially useful in distinguishing between prolific and harmful offenders and is an effective way to represent the concentration of crime-related harm among offenders, victims or places (Mitchell 2019; Ratcliffe & Kikuchi 2019). WACHI values were adapted to analyse the offending of OMCG members using police data (Morgan, Dowling & Voce 2020). Similar to Cubitt and Morgan (2022), high-harm offences in this study included murder, attempted murder, manslaughter, aggravated sexual assault, importation of illicit substances, aggravated robbery, non-aggravated robbery, property damage by fire or explosion, dealing in commercial quantities of illicit substances, and serious assault causing injury. As well as representing high-harm offences generally, these offence types are also reflective of the serious offences that characterise OMCG involvement in organised criminal activity and which typify the violence that occurs during inter- and intra-club disputes (Lauchs, Bain & Bell 2015). We used the same selection of offences to reproduce this analysis nationally and in the three Australian states that feature the largest proportion of criminally active OMCG members (New South Wales, Victoria and Queensland).

Data

Cubitt and Morgan (2022) used a dataset of 130 explanatory variables emerging from police offence, intelligence and custodial data, relating to 2,246 active OMCG affiliates. These individuals were patched members or nominees of a club who were currently involved with an OMCG (ie they were not former members or deceased). The outcome of interest—high-harm offending—was primarily operationalised as a five-year reference period, while findings were also compared to a two-year reference period. The random forest algorithm was applied using the demographics, custodial history and criminal history of OMCG affiliates to predict whether they would commit a high-harm offence across a subsequent five-year reference period ending in December 2019.

In the current study, we used data on the recorded offence histories of individuals identified by law enforcement as being affiliated with an OMCG in Australia as of May 2019. Data were obtained from two intelligence databases managed by the ACIC: the NGL and the National Police Reference System (NPRS). The ACIC is a national criminal intelligence agency with responsibility for developing and maintaining national information-sharing systems, which connect state and territory law enforcement data and facilitate data sharing between agencies. This is particularly relevant to serious and organised crime, given it is not constrained by domestic or international borders, and for criminal groups that are highly mobile, including OMCGs (Dowling & Morgan 2021). The NGL is a validated list of OMCG affiliates, including patched members, prospects and nominees. (Prospects and nominees are both probationary members, but the terminology varies between clubs.) Motorcycle clubs in Australia are classified as OMCGs, and included on the NGL, if they meet a set of criteria agreed upon by law enforcement and intelligence agencies. Information about OMCG associates or supporters may also be included on the NGL; however, this information was excluded from analysis, largely because of inconsistencies between jurisdictions. The NPRS holds the offence history of individuals who have been arrested and subject to some form of legal action by police in any jurisdiction in Australia. Data for affiliates of OMCGs on the NGL were matched with the criminal history of individuals on the NPRS using their name, date of birth and address information. All matching was undertaken by the ACIC prior to transmission to the research team. This procedure resulted in a dataset of 5,669 members from 39 clubs across Australia. Four affiliates were removed due to low confidence matches, with missing data resulting in a further 20 affiliates being removed from the sample. An additional 111 affiliates were removed because they were not assigned to a chapter, resulting in a final sample of 5,534 OMCG affiliates from across Australia.

The outcome variable was operationalised for two reference periods: the most recent two-year period and the most recent five-year period prior to May 2019. The five-year reference period was consistent with the original study and allowed for direct comparisons to be made. Explanatory variables included the age, membership status and location of each affiliate, along with information about their prior criminal history. Criminal histories of OMCG affiliates were based on the 20-year period prior to the reference period. Prior offences were grouped into nine major offence categories for analysis. These included the following offence categories: violent, property, drugs, weapons, traffic, public order, breach of orders, fraud, and other offences. As well as information about the prior offending of each affiliate, the average number of offences by other members of their chapter was calculated. Whether an affiliate had previously committed a high-harm offence and the total harm associated with their prior offending was also determined. A variable for individual mobility was included for analysis, to identify affiliates who had committed offences in a jurisdiction other than where they currently resided (see Dowling & Morgan 2021), as well as a variable relating to the mobility of other chapter members. We also included a variable measuring the versatility of offending among each affiliate using the same diversity index implemented in prior studies focusing on organised criminal offending (see Francis et al. 2013; Fuller, Morgan & Brown 2019), incorporating a bias-correction method to account for cases with a small number of offences (Francis & Humphreys 2016). This measure produces a value between 0 and 1, with a score closer to zero indicating greater specialisation, and a score closer to one indicating more diverse offending.

Analytic approach

Consistent with the prior study we implemented the random forest algorithm in a classification task to predict whether OMCG affiliates committed a high-harm offence in the reference period. Cubitt and Morgan (2022) argued that, given the substantial amount of data and the number of variables included in the analysis, the random forest algorithm was preferred over more conventional analytic methods, such as logistic regression (Berk 2013). Further, the random forest has been consistently found to outperform generalised linear modelling (Couronné, Probst & Boulesteix 2018). However, given the smaller number of variables available for the current study, and the fact that random forest is still an emerging analytical methodology in criminology, we benchmarked the analysis against the more commonly used logistic regression method.

National-level analysis

The first step in this process was to reproduce the analysis performed by Cubitt and Morgan (2022) at the national level, employing all available data, for both the five-year and two-year reference periods. To compute the random forest, data were randomised and partitioned into a 70 percent training set and a 30 percent test set. The random forest algorithm was trained on the larger set and the model was tested using the partitioned test set (Hyndman & Anthanasopoulos 2014).

While considerably fewer variables were available than in the prior study, there were still a large number of variables attributable to the 5,534 OMCG affiliates. For each random forest a logistic regression was also estimated, providing a benchmark for the predictive accuracy of the model and assessment as to the relative benefits of a machine learning approach. For each logistic regression we computed both supervised and unsupervised modelling approaches, primarily to identify the approach with strongest modelling performance, but also to account for any potential overfitting. In each circumstance the modelling performed equally well. We therefore report the supervised approach to closely adhere to the modelling of the random forest.

A receiver operating characteristic (ROC) curve, which plots the true positive rate of classification, referred to as sensitivity (y-axis) compared with the false positive rate, equal to 100 minus the specificity (x-axis) at any threshold value, was produced for each model. The area under the receiver operating characteristic (AUROC) curve was then determined to provide an overall measure of performance. We compared the AUROC for the random forest to the AUROC for the logistic regression to determine whether the random forest outperformed the logistic regression. To do this we implemented the bootstrap test for statistical significance between ROC curves to determine whether the differences between the predictive accuracy of the models were statistically significant.

To find the most robust modelling approach, we tuned the hyperparameters of the random forest model to optimise the number of iterations and variables randomly considered at each split. When optimising the random forest, model performance will typically plateau as the ntree parameter (number of trees) reaches several hundred iterations (Couronné, Probst & Boulesteix 2018). This was the case in the present research; for each model, performance was optimal at ntree=500, with 20 features randomly selected at each split. We then reported the out-of-bag error estimate, which describes the aggregate error of the random forest on the training set (Schonlau & Zou 2020). Given that the national-level sample was not equally drawn from each jurisdiction, it was important to ensure that the training and test sets used for analysis were balanced. This balancing was undertaken during the randomisation of the data into training and test sets, and was validated by comparing the out-of-bag error estimate with the aggregate classification error produced by the confusion matrix. The random forest performs a type of cross-validation, using out-of-bag samples, as a component of the training step of the modelling process.

A confusion matrix was also produced for the test set of each modelling process (see Table 1). The confusion matrix measures the performance of the trained model on the test set, providing a measure of how often the model successfully or unsuccessfully made predictions (Barnes & Hyatt 2012). For simplicity, we focus on the overall classification errors and the false positive and false negative rate, although the full range of parameters is noted in Table 1.

Table 1: Error and accuracy calculations of the confusion matrix

		Actual high-harm offender?		Classification errors	Classification accuracy
		No	Yes		
Predicted high-harm offender?	No	True negative (tn)	False negative (fn)	False omission rate fn/(fn+tn)	Negative predictive value tn/(fn+tn)
	Yes	False positive (fp)	True positive (tp)	False discovery rate fp/(tp+fp)	Positive predictive value tp/(tp+fp)
Classification errors		False positive rate fp/(fp+tn)	False negative rate fn/(tp+fn)	Aggregate classification error (fp+fn)/(tp+fp+tp+fn)	
Classification accuracy		Specificity or true negative rate tn/(fp+tn)	Sensitivity or true positive rate tp/(tp+fn)	Accuracy (tp+tn)/(tp+fp+tp+fn)	

In the *Results* section, we report the out-of-bag error estimate, the confusion matrix for prediction error on the test set and the AUROC to describe accuracy of the modelling processes (Couronné, Probst & Boulesteix 2018; Schonlau & Zou 2020; Svetnik et al 2003). Each analytical process reported in this study was undertaken using the statistical analysis software R Studio, and the ‘randomForest’, ‘dplyr’, ‘pROC’ and ‘ggplot2’ packages.

State-level analysis

It was also important to understand whether there was variation in prediction accuracy between states. The four jurisdictions with the largest proportion of OMCG affiliates—New South Wales (*n*=2,362, 42.7% of all OMCG affiliates nationally), Victoria (*n*=1,322, 23.9%), Queensland (*n*=750, 13.6%) and Western Australia (*n*=588, 10.6%)—were selected for comparison. The random forest and logistic regression modelling processes were replicated for each jurisdiction independently. The evaluation metrics were then repeated, including the out-of-bag error estimates on the training set and the AUROC. A confusion matrix was also produced for each test set.

Limitations

There are several limitations specific to this study that need to be acknowledged. There are obvious limitations with relying on recorded offence data. These are especially relevant to OMCGs, given the clandestine nature of organised criminal activity and a culture that emphasises the importance of secrecy, including around serious acts of intra- and inter-club violence. Relatedly, there is information that is readily available to law enforcement, including from the NPRS used in this research, that was not included in the data used for the current study. This includes information on warnings, warrants, firearm involvements and protection orders—information that may contribute to the predictive accuracy of a risk assessment model, especially in relation to short-term violent offending (although this is dependent on how readily accessible these data are and whether they are in a format suitable for analysis). There are also limitations that are specific to the random forest algorithm. Because we were focused on replicating the methodology in Cubitt and Morgan (2022), which was limited to individual-level data, the random forest used here prevented us from including group-level variables that may have been of interest, and which have been shown to be correlated with serious criminal activity (Morgan, Dowling & Voce, in press). There is growing evidence of co-offending between OMCG members (see Bright et al. 2022) that may have important implications for understanding patterns of offending, including high-harm offending.

Results

Table 2 provides descriptive statistics for the sample. The mean age of affiliates was 44.0 years. Three-quarters of affiliates (73.5%) were patched members in non-office bearer roles, 16.9 percent were patched members in office bearer roles (ie chapter president, sergeant of arms, etc) and 9.6 percent were prospects or nominees. Prior criminal history characteristics are based on a 20-year period prior to the five-year reference period. Three-quarters of the sample (73.2%) had a recorded history of offending. The most common prior offence types were serious traffic offences (44.3%), followed by violent offences (43.5%) and disorder offences (42.0%). One in three (33.3%) OMCG members had a prior drug offence, while nearly one in ten (9.3%) had been proceeded against for a fraud offence prior to the reference period. Affiliates had an average of 12.2 prior offences. Nearly one-third (31.0%) had at least one offence recorded in a jurisdiction other than the one in which they currently resided, indicating some level of mobility. One-third of the sample had committed a prior high-harm offence (1,915 affiliates, 34.6% of affiliates and 47.3% of criminally active affiliates).

During the five-year reference period 2,809 affiliates (50.8%) had committed at least one offence of any type, while 1,054 affiliates (19.1%) had committed a high-harm offence. During the two-year reference period 1,746 affiliates (31.6%) had committed at least one offence and 511 (9.2%) had committed a high-harm offence.

Table 2: Descriptive statistics (*n*=5,534)

Variables	*n*	%	Mean	SD
Age			44.0	12.5
Rank and membership status				
Patched member (non-office bearer)	4,069	73.5		
Patched member (office bearer)	933	16.9		
Prospect or nominee	532	9.6		
State or territory				
New South Wales	2,362	42.7		
Victoria	1,322	23.9		
Queensland	750	13.6		
Western Australia	588	10.6		
Tasmania	279	5.0		
South Australia	176	3.2		

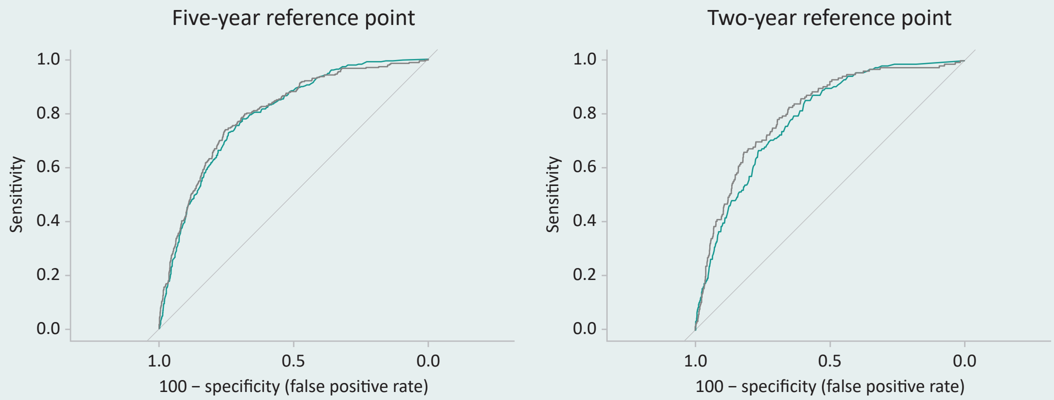
Table 2: Descriptive statistics (*n*=5,534) (cont.)				
Variables	***n***	**%**	**Mean**	**SD**
Australian Capital Territory	32	0.6		
Northern Territory	24	0.4		
Prior offences				
Violence	2,406	43.5	2.4	5.0
Property (excluding fraud)	1,694	30.6	1.8	6.5
Fraud	516	9.3	0.3	3.2
Drug	1,844	33.3	1.6	4.1
Weapons	1,463	26.4	0.9	2.5
Serious traffic	2,449	44.3	1.8	3.9
Disorder	2,324	42.0	1.6	3.4
Breach	1,279	23.1	0.9	2.9
Other	1,670	30.2	0.8	2.0
Total	4,050	73.2	12.2	21.6
Prior offences among fellow chapter members				
Violence			2.4	1.8
Property (excluding fraud)			1.8	2.1
Fraud			0.3	0.7
Drug			1.6	1.5
Weapons			0.9	0.9
Serious traffic			1.9	1.6
Disorder			1.6	1.3
Breach			0.9	1.2
Other			0.8	0.8
Total			12.2	8.8
Prior harm				
Total			568.9	1,749.3
High-harm offence	1,915	34.6	1.2	2.7
Prior mobility				
Prior offending mobility	1,713	31.0		
Prior offending mobility among fellow chapter members	4,862	87.9		
Offence specialisation				
Specialist offender (prior offenders only)	903	16.3 (22.3)		
Diversity index			0.6	0.3
Members per chapter			24.0	19.5

Note: Prior criminal history is based on the 20-year period prior to the five-year reference period
Source: OMCG criminal history database [computer file]

National-level analysis

In first reproducing the analysis at the national level, the random forest model was trained using the partitioned training sample to predict whether affiliates would commit a high-harm offence within the five-year reference period. The training and the test samples were randomly selected and reflected the distribution of affiliates from each state. This modelling resulted in an out-of-bag error of 17.8 percent. For the five-year reference period, there was little difference between the random forest (AUROC=0.801) and the logistic regression (AUROC=0.795) in terms of their predictive accuracy (p=0.587), as reflected in the ROC curves in Figure 1. We then reproduced the analysis using a two-year reference period. The random forest model featured an out-of-bag error of 9.5 percent. There was a small difference between the random forest (AUROC=0.810) and the logistic regression (AUROC=0.786) results, but the difference did not achieve statistical significance (p=0.12; Figure 1).

Figure 1: Receiver operating characteristic (ROC) curves for random forest (grey) and logistic regression (green) predicting high-harm offending among OMCG affiliates at the national level



Source: OMCG criminal history database [computer file]

We then produced the confusion matrix for the random forest, given this was the preferred approach in prior research, to more closely consider classification accuracy (Table 3). For the five-year reference period, the classification error on the test set was 19.1 percent; the model was more successful at predicting which affiliates would not commit a high-harm offence than those who would. The false positive rate was 8.3 percent (or a specificity of 91.7%), while the model had a false negative rate of 63.4 percent. For the model using a two-year reference period, the aggregate classification error on the test set was 9.0 percent. However, this model was much more successful in predicting affiliates who would not offend than those who would. In fact, while the model correctly classified more than 99.3 percent of affiliates who did not commit a high-harm offence, it failed to accurately classify almost all high-harm offenders in the test set, with a false negative rate of 97.2 percent (a sensitivity of just 2.8%). On this basis, the two-year model was deemed not viable, and the remainder of the analysis proceeded with a five-year reference period.

Table 3: Confusion matrix for random forest model trained on high-harm offending at the national level

		Actual high-harm offender?		Classification errors
		No	Yes	
Five-year reference period				
Predicted high-harm offender	No	1,225	206	14.4%
	Yes	111	119	48.3%
Classification errors		8.3%	63.4%	19.1%
Two-year reference period				
Predicted high-harm offender	No	1,507	140	8.5%
	Yes	10	4	71.4%
Classification errors		0.7%	97.2%	9.0%

Source: OMCG criminal history database [computer file]

In the *Appendix*, we present the results of a restricted model limited to only those individuals who had a prior recorded history of offending based on the five-year reference period. This is the same approach that was used by Cubitt and Morgan (2022). Results indicate that there was little advantage in limiting the sample to OMCG members with a prior recorded criminal history, and that models generally performed slightly worse when the sample was restricted. Specifically, the AUROC was lower and the rate of classification errors was higher. For this reason, the analysis in the main section of this report includes the full sample of individuals in the NGL, irrespective of whether they had a recorded history of offending.

State-level analysis

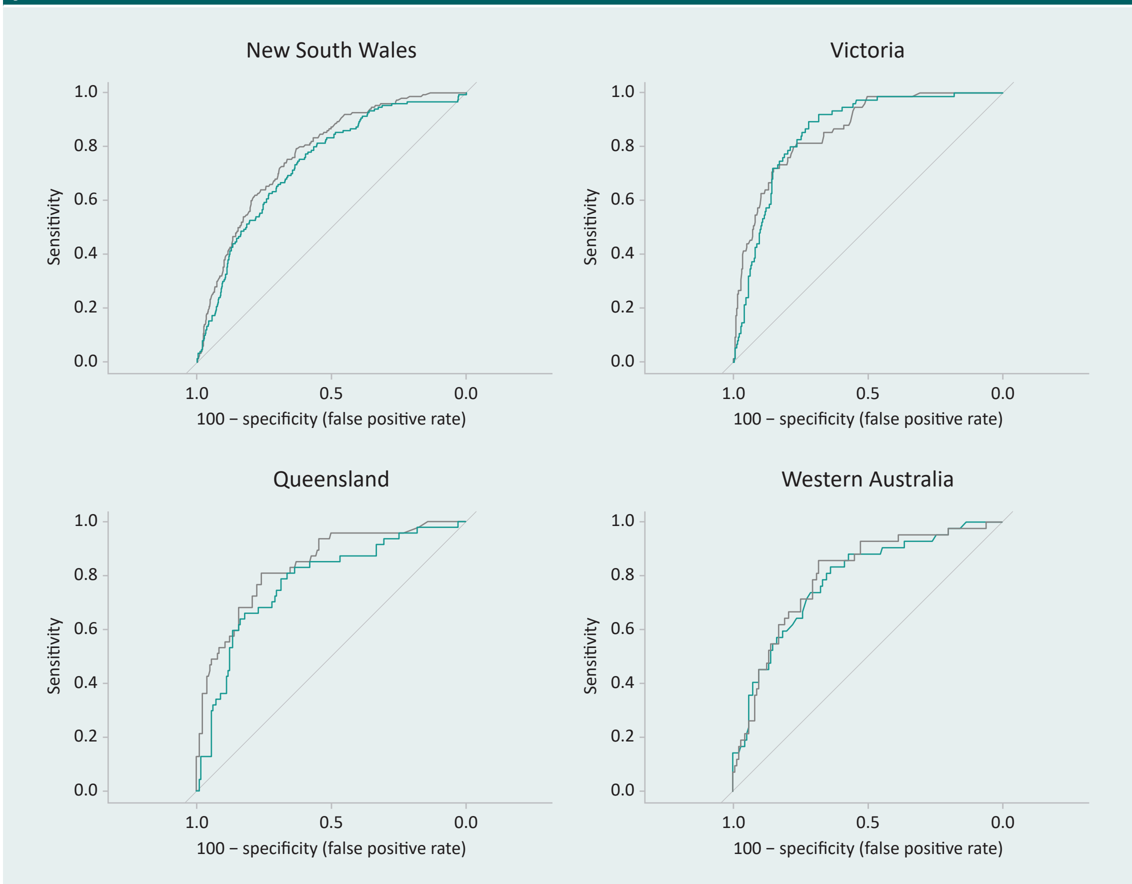
Random forest modelling

The next stage of the analysis involved developing prediction models for the four largest Australian jurisdictions, in terms of both population and OMCG membership. This was based on the location of the chapter in which each affiliate was a member. Separate models were developed using a training set restricted to OMCG members from each jurisdiction. The first jurisdiction, New South Wales, included data for 2,362 affiliates. This was the same jurisdiction analysed by Cubitt and Morgan (2022), which allows for more direct comparison (although the observation period was different). The same analytical process was implemented as with the national model but only using the five-year reference period for high-harm offending. The random forest model produced an out-of-bag error of 20.7 percent and outperformed the logistic regression (AUROC=0.774 vs AUROC=0.738, $p<0.05$), reflected in the ROC curves in Figure 2.

The second jurisdiction examined was Victoria, where there were 1,322 recorded OMCG affiliates. The out-of-bag error for the random forest was 15.8 percent. Although model performance was similar across all states, the highest rate of classification success was achieved using data from Victoria. There was little difference between the random forest (AUROC=0.862) and the logistic regression (AUROC=0.857, *p*=0.802), with both methods achieving an excellent level of predictive accuracy, as illustrated by the ROC curves in Figure 2.

The random forest for Queensland (*n*=750) produced an out-of-bag error of 15.6 percent, and an AUROC of 0.838, outperforming the logistic regression (AUROC=0.775, *p*<0.05). The final jurisdiction considered in this study was Western Australia (*n*=588). In computing this model, the random forest produced an out-of-bag error of 19.0 percent, and an AUROC of 0.798. The difference between the AUROC for the random forest and the logistic regression (AUROC=0.784, *p*=0.70) was not statistically significant.

Figure 2: Receiver operating characteristic (ROC) curve for random forest (grey) and logistic regression (green) predicting high-harm offending among OMCG affiliates of each jurisdiction



Source: OMCG criminal history database [computer file]

Confusion matrices were then produced for each of these jurisdictions (Table 4). In New South Wales, the aggregate classification error on the test set was 14.2 percent. The rate of false positives (7.9%) and false negatives (69.4%) was similar to the national model. However, while the rate of false positives was marginally higher than in the original study using data from New South Wales (1.1%), the rate of false negatives was substantially higher (11.8%). For the Victorian model the aggregate classification error on the test set was 9.3 percent. The random forest model for high-harm offending among affiliates from Victoria performed better than the national model, while still favouring the prediction of who would not commit a high-harm offence. In addition to a high negative predictive value (94.6%) and low false positive rate (4.3%), the rate of false negatives was lower than the national and NSW models (48.9%)

For the model using data from Queensland, the aggregate classification error on the test set was 15.5 percent. Classification errors were similar to Victoria—the model performed well in predicting affiliates who would not commit a high-harm offence, with a false positive rate of 7.3 percent; however, it was less successful with those who would (false negative rate of 46.8%). Finally, in Western Australia, the aggregate classification error on the test set was 21.0 percent. The model again performed well in predicting affiliates who would not commit a high-harm offence (with a false positive rate of 9.6%). The false negative rate fell close to the range for other states at 54.8 percent.

Table 4: Confusion matrix for random forest model trained on high-harm offending in four jurisdictions

		Actual high-harm offender?		Classification errors
		No	Yes	
New South Wales				
Predicted high-harm offender?	No	515	104	16.8%
	Yes	44	46	48.9%
Classification errors		7.9%	69.4%	14.2%
Victoria				
Predicted high-harm offender?	No	335	23	6.4%
	Yes	15	24	38.5%
Classification errors		4.3%	48.9%	9.3%
Queensland				
Predicted high-harm offender?	No	165	22	12.2%
	Yes	13	25	34.2%
Classification errors		7.3%	46.8%	15.5%
Western Australia				
Predicted high-harm offender?	No	122	23	15.9%
	Yes	13	19	40.6%
Classification errors		9.6%	54.8%	21.0%

Source: OMCG criminal history database [computer file]

The model developed for New South Wales clearly did not perform as well as those produced for other jurisdictions. Given that the representation of states and territories in the national training and test samples was aligned with the actual distribution of affiliates, and New South Wales accounted for the largest proportion of affiliates nationally, the results for New South Wales may have negatively impacted the overall accuracy of the national-level model.

We know there have been important changes to the offending behaviour of OMCG affiliates, particularly young members, associated with the changing culture of OMCGs (Dowling et al. 2021; Voce, Morgan & Dowling 2021), and that there are also important differences in offending associated with member status within clubs (which is also associated with age; Morgan, Dowling & Voce forthcoming). By definition, younger affiliates will have shorter criminal histories, which may also influence the outcomes of risk assessment models—especially with a model that relies almost exclusively on the recorded offence history of the individuals for whom predictions are being made. Importantly, while a similar proportion of affiliates in New South Wales were below the age of 40 years (47.7%) as in Western Australia (44.4%), this proportion was much higher than in Victoria (36.1%) and Queensland (33.9%)—states which produced much more accurate models.

To explore whether this might help explain the results, affiliates from New South Wales were stratified by age into one group who were over 40 years of age (*n*=1,235), and a group who were 40 years old and under (*n*=1,128). We hypothesised that the prediction accuracy would be poorer for the younger sample and better for the older sample. This hypothesis held true, with the random forest model for the older sample significantly outperforming the model for the younger sample (AUROC=0.833 vs AUROC=0.654). This suggests that the difference in predictive accuracy between states may be explained, at least in part, by the different composition of affiliates in each jurisdiction.

Variable importance for predicting high-harm offending

Another explanation for the high rate of false negatives in the national model may be that it does not take into account differences between the states in terms of which variables are important in predicting high-harm offending. The results of each random forest model were interpreted using Mean Decrease Gini (MDG; Hong, Xioling & Hua 2016). The Gini coefficient is a measure of statistical dispersion, in which results attributed to variables are interpreted as a proportion of the overall random forest model, relative to the AUROC. Each variable is assigned an MDG coefficient identifying its importance in the accuracy of the predictions. The Gini coefficients for each model are provided in Table 5. The five most important variables in each model are highlighted in bold. These demonstrate that there was some variation between jurisdictions in the relative importance of the information that was used to make the predictions.

For example, while the age of an OMCG member was an important variable in each jurisdiction, it was more important in Victoria and Queensland than it was in New South Wales and Western Australia. Total prior harm was an important variable in the national and state models. Conversely, the club with which members were associated was an important predictor in New South Wales and Victoria; however, in Queensland and Western Australia, it was considerably less important. There were more similarities in the MDG for variables in each model than there were differences. However, the degree of variation suggests that there are likely differences in the patterns of recorded offending by OMCG affiliates in each jurisdiction that may not have been captured if the analysis was only performed at the aggregate, national level. We cannot be sure as to the reasons for these differences—they may reflect actual differences in offending behaviour or differences in the composition of OMCG members in each state, or they may be a function of differences in operational activity and information recording practices between the states.

Table 5: Feature importance for each model developed to predict high-harm offending

Variable	National	NSW	Vic	Qld	WA
Age	**15.73**	**11.09**	**17.79**	**18.21**	**12.09**
Total harm produced by prior offences	**9.20**	**7.62**	**5.52**	**8.27**	**7.76**
Club	**8.67**	**7.09**	**9.58**	1.22	2.45
Total number of prior offences	**3.86**	**6.19**	**4.27**	**6.98**	3.63
Total number of prior high-harm offences	**3.53**	2.15	2.86	1.19	1.61
Diversity of prior offences	3.19	4.26	**4.59**	2.27	**4.06**
Number of jurisdictions in which the chapter is present	2.96	2.43	2.53	1.02	0.51
Average number of prior disorder offences by other members of the same chapter	2.79	2.79	2.32	2.36	3.79
Average number of prior property offences by other members of the same chapter	2.71	2.82	2.03	2.74	3.97
Average number of prior traffic offences by other members of the same chapter	2.67	3.12	2.24	**5.21**	**5.60**
Average number of prior drug offences by other members of the same chapter	2.64	2.72	2.60	2.97	2.78
Average number of prior offences by other members of the same chapter	2.63	2.96	2.49	4.26	3.57
Number of prior violent offences	2.55	**4.44**	4.09	1.75	1.53
Average number of prior breach offences by other members of the same chapter	2.39	2.29	1.91	1.98	2.37
Average number of prior violent offences by other members of the same chapter	2.38	2.50	1.97	2.40	2.68
Average number of prior weapons offences by other members	2.29	2.73	1.71	2.07	3.00

Table 5: Feature importance for each model developed to predict high-harm offending (cont.)

Variable	National	NSW	Vic	Qld	WA
Average number of prior otherwise categorised offences by other members of the same chapter	2.27	2.72	1.84	3.61	3.08
Average number of prior criminal enterprise offences by other members of the same chapter	2.20	2.18	2.43	2.49	2.46
Number of prior disorder offences	2.17	3.85	3.72	**5.52**	2.71
Number of prior traffic offences	2.04	2.33	2.15	2.36	2.60
Number of members per chapter	1.99	2.11	1.62	2.06	1.54
Average number of prior high-harm offences by other members of the same chapter	1.84	1.69	1.63	2.13	1.91
Average harm produced by prior offences by other members of the same chapter	1.84	1.73	1.48	2.06	1.89
Average number of prior fraud offences by other members of the same chapter	1.72	2.03	1.46	1.42	1.18
Number of prior property offences	1.66	2.85	1.84	3.27	2.26
Number of prior breach offences	1.65	2.1	1.79	2.33	**8.99**
Number of prior drug offences	1.39	1.25	1.78	1.77	1.56
Rank within club	1.21	1.39	2.82	0.00	1.17
Prior offences otherwise categorised	1.15	1.35	0.99	1.77	2.68
Number of prior weapons offences	1.01	1.12	1.57	0.78	1.04
Number of prior criminal enterprise offences	0.92	1.02	0.98	0.64	0.99
Number of prior fraud offences	0.67	0.77	0.72	0.39	0.79
Club has presence in multiple jurisdictions	0.55	0.58	1.19	0.78	1.08
Offended in multiple jurisdictions	0.39	0.27	1.45	0.96	0.64
Other members in the same club have offended in multiple jurisdictions	0.09	0.09	0.05	0.72	0.06

Note: bold indicates the five most important variables in each model
Source: OMCG criminal history database [computer file]

Summary of the findings

This research reproduced the analysis by Cubitt and Morgan (2022), using more limited data, to consider relative model accuracy, trade-offs of the approach, and implications. Table 6 summarises the findings of the original model for New South Wales and the national and state models presented in this report. While the aggregate prediction accuracy produced in this study was not as strong as reported in the original paper, that was expected given the comparatively limited dataset. Nevertheless, the accuracy of each of these models was acceptable according to conventional standards. The random forest was particularly effective in minimising false positives—the proportion of individuals who did not go on to commit a high-harm offence but who were predicted to be a high-harm offender never exceeded 10 percent. Conversely, none of the models were as accurate at identifying which affiliates should be targeted. The proportion of cases that did go on to commit a high-harm offence, but were not predicted to be a high-harm offender (ie false negatives), was much higher in each model than in the original model developed by Cubitt and Morgan (2022).

Table 6: Summary findings

Sample size	Sample size	Predictive accuracy	False positives (people we should not target)	False negatives (people we should target but miss)
Original model (NSW)[a]	2,246	91.4	1.1	11.8
National model[b]	5,534	80.1	8.3	63.4
New South Wales[b]	2,362	77.4	7.9	69.4
Victoria[b]	1,322	86.2	3.4	48.9
Queensland[b]	750	83.8	7.3	46.8
Western Australia[b]	588	79.8	9.6	54.8

a: Produced by Cubitt and Morgan (2022)
b: Source: OMCG criminal history database [computer file]

Discussion

The aim of this study was to reproduce a risk assessment model developed using machine learning for a specific cohort of offenders and to see how changing certain parameters—namely, the availability of data—impacted on the predictive accuracy of the model. We diverge slightly from being a true direct replication study in the purest sense (Peels 2019). Rather than being focused solely on whether the findings themselves would replicate, including in different settings, our goal was to also assess whether it would be possible to reproduce the findings in closer to a real-world context. Specifically, we assessed whether it would be possible to accurately predict high-harm offending among OMCG members without the depth of criminal history information or linked custodial data that was used in the original model.

Model accuracy: The trade-offs of using suboptimal data

Overall, the results indicate that the predictive accuracy of the random forest was—according to conventional thresholds—very good to excellent, primarily attributable to the ability to predict which affiliates would not commit a high-harm offence. That we could still predict high-harm offending with a relatively high degree of accuracy, despite a more limited set of variables, is important. Australian and international research has shown that OMCG members commit a high volume of violent and organised crime offences (von Lampe & Blokland 2020). Research has also shown that recorded offending and related harm is concentrated among a relatively small proportion of individuals and chapters (Blokland, Soudijn & van der Leest 2017; Dowling & Morgan 2021; Morgan, Dowling & Voce 2020). There is evidence that the rate of involvement in serious crime, including violence, is increasing among younger members (Voce, Morgan & Dowling 2021), most likely as a function of changing recruitment practices and a shift towards criminal enterprise and profit-motivated offending (Dowling et al. 2021; Lauchs 2017). Further, offending varies quite significantly between clubs, with some clubs having a much higher proportion of members with a recent history of serious crime, and some clubs having very little criminal justice involvement (Blokland et al. 2019; Morgan, Dowling & Voce in press; Morgan, Dowling & Voce 2020).

It is evident, however, that there are trade-offs associated with the use of less than optimal data. The ease with which information can be accessed and used to predict high-harm offending must be balanced alongside the weaker prediction rates produced in this study than by Cubitt and Morgan (2022). That earlier study demonstrated that the inclusion of custodial data to produce a weighted harm score for prior offending was important in the predictive accuracy of that model. So too was the detail about the offences, including characteristics of the incident (substance use, domestic violence involvement) and the location in which it occurred. Australia's federated system of government, where responsibility for the management of correctional systems rests with state and territory governments, means that linked custodial data is not available nationally. There are also challenges with accessing these data on a routine basis. Though it is difficult for us to estimate how much this impacted the predictive accuracy of our risk assessment models in this study, the variable importance of weighted variables in the original risk assessment model suggests this did weaken our overall model. However, the model we developed in this paper is much more likely to be implemented by a national agency.

Data quality and accessibility is a perennial challenge for law enforcement agencies and, in many circumstances, the data available to researchers from data linkage exercises may not be readily accessible by intelligence analysts or other frontline users (O'Connor et al. 2021). However, it is important to note that there has been significant, and recent, advancement in law enforcement data holdings. Indeed, this research was undertaken within the context of the development of the ACIC's new National Criminal Intelligence System (NCIS). The NCIS, which became available for use by law enforcement and intelligence agencies in active operations in early 2021, aims to provide a secure mechanism through which policing information from multiple datasets can be shared nationally using a single interface (ACIC 2021). At the same time, building linked databases, especially those that traverse jurisdictional boundaries, is a notoriously challenging task (Hollywood & Winkelman 2015), and the existence of national information-sharing databases does not guarantee they will be used consistently (Phythian & Kirby 2022). There are numerous technological, organisational, leadership and cultural factors that can inhibit the optimal sharing and use of important information (Koper, Lum & Willis 2014; Lum, Koper & Willis 2017). These can limit the use of technology and stymie efforts to innovate, including the adoption of intelligence-led policing (Darroch & Mazerolle 2013; Sanders, Weston & Schott 2015).

The differences we found in the predictive accuracy of the random forest between the three most populous Australian states—particularly in terms of false negative rates—is noteworthy. There may be a number of explanations for these differences. The findings may be influenced by the focus of police efforts targeting OMCG members. The poorest predictive accuracy was for members in New South Wales, which was the focus of the original study by Cubitt and Morgan (2022). This state has arguably had the most intensive and prolonged response to OMCGs, since the inception of Strike Force Raptor in 2009, which is relevant given the data used in this study relate to legal action by police. New South Wales is a target for criminally mobile OMCG members, offering attractive illicit markets and acting as the central hub for some of the largest clubs (Dowling & Morgan 2021). We also found a much higher proportion of younger affiliates in New South Wales than in Victoria and Queensland. Patterns in recorded offending may therefore be quite different among OMCG members in that state, which may influence the relative predictive accuracy of the model, especially in the absence of additional detail and linked custodial histories. The results for New South Wales may have also impacted the accuracy of the national model, especially given New South Wales accounted for the largest proportion of OMCG affiliates included in the study.

We also observed some differences between states in the variables that were important in predicting high-harm offending. This too may be a consequence of actual differences in offending behaviour, differences in the composition of OMCG members in each state, or a function of differences in operational activity and information recording practices between the states. This might also explain why the national model did not perform as well as some of the state models, like Queensland and Victoria. The random forest approach aims to minimise aggregate classification errors—meaning that it tries to find the best pathway based on all of the data on hand to predict the outcome of interest. Outcomes of a random forest for different populations will therefore naturally vary when these pathways differ because of the relative importance of variables in the model. Jurisdictional comparisons like this are uncommon, and further work is needed to understand why prediction models may perform better in some jurisdictions than others.

Relatedly, the random forest did not consistently outperform logistic regression. Meta-studies have found that random forest outperforms logistic regression in the majority of cases, but that this is contingent on several factors, including the sample size and number of variables (or features) in the data (Couronné, Probst & Boulesteix 2018). There were far fewer features in the random forest used in the current study than in the original study by Cubitt and Morgan (2022), who had access to a much wider array of data and more than 100 different variables. There are other examples where random forest has not outperformed logistic regression in criminal justice settings (Etzler et al. 2022). However, we argue here that random forest remains the preferred approach, since when there were differences between models they favoured random forest.

Implications and disruption opportunities

We are not ignorant of the sensitives around the management of gang databases or lists (Densley & Pyrooz 2020; Kennedy 2009). Of course, there are important differences between OMCGs and other gangs in terms of how easy it may be to validate group membership. For example, some groups have specific membership periods, members may have particular racial and ethnic backgrounds, and clubs may have different historical origins, serving to mitigate some of the risks of using databases in this context. Certainly, the results presented here cannot be applied to individuals who might be included on gang lists as OMCG associates, or to other offender groups, including those for which intelligence-led models have attracted criticism in Australia (Sentas & Pandolfini 2017). Applying structured risk assessment methods that have been developed and validated using transparent methods, as was the case with Cubitt and Morgan (2022) and in the current study, may help to mitigate the risks of individuals being targeted solely because of their membership on a list. Further, the use of harm as the primary outcome measure, rather than any offending, may also alleviate concerns (Ratcliffe & Kikuchi 2019). Cubitt and Morgan (2022) showed that focusing on high harm was much less likely to result in the targeting of individuals who did not go on to commit high-harm offences. An important finding from Cubitt and Morgan (2022) was the low rate of false positives and false negatives. Here, the rate of false positives was still low. This is important as it limits the risk of OMCG members being unfairly targeted, and is also important in terms of the efficient use of policing resources. Nevertheless, because high-harm offenders represent the minority of individuals in the sample (19.1%), a low false positive rate still means that between 34 and 49 percent of individuals identified by the model as high-harm offenders were, in fact, not. This has important implications in terms of the implementation of the model and how this information is used.

Conversely, a low false negative rate means that eventual high-harm offenders are not being overlooked. The rate of false negatives in this study were much higher than in the previous study (11.8%; Cubitt & Morgan 2022). At a national level, the model correctly identified nearly 40 percent of those OMCG members who went on to commit a high-harm offence as a high-harm offender—a false negative rate of more than 60 percent. However, it is important to identify a baseline for acceptability. For example, in the absence of modelling of this type to inform targeting, it is possible that the error rate may be higher than is seen here. Despite the false prediction rate among these models exceeding previous research that had access to more detailed data (the trade-off), it may still be an improvement on naïve predictions, or a simple informed guess. While it is important to remain circumspect about the quality of the modelling, and the identified rate of error, there is still some utility to the information provided by these approaches in identifying law enforcement targets.

It is also important to consider the context and approach of the intervention. The gold standard of risk assessment is an absolute rate of prediction or, in other words, a model that produces no false positives or false negatives. However, this is an extremely unlikely outcome, particularly in criminal justice risk assessments. As a result, there is another trade-off to be made relating to the degree of consideration given to the outcomes of the model when balancing the selected intervention and the potential consequences of that intervention. Ultimately, the threshold for an acceptable rate of false positives should reflect the consequences of the decisions or interventions made in response to the findings (Wynants et al. 2019). Applying this logic to the present context, a direct policing intervention based solely on the predictions from this research is likely inappropriate. However, the false positive rate achieved in this research may be useful for resource allocation and as an additional tool alongside traditional intelligence. As Wynants et al. (2019) note, there is no universally optimal threshold for decision-making from prediction models; rather, the pivotal consideration is striking a balance between the rate of false positives and the intended intervention.

Irrespective of the quality of data that is available for analysis, the outcomes of a structured risk assessment are not an absolute measure of whether an individual will or will not commit a high-harm offence, nor can they be relied upon—in isolation—to identify all priority targets. Intelligence-led policing of OMCGs remains the cornerstone of the law enforcement response, and there are various techniques employed by law enforcement agencies as part of this strategy, including the use of surveillance and human sources. Though unusual, the recent example of the Federal Bureau of Investigation and Australian Federal Police-led Operation Ironside, which disrupted several planned murders, some allegedly involving OMCG members (Australian Federal Police 2021), illustrates the benefit of alternative intelligence sources. We propose this model as an additional tool to guide the efforts of law enforcement. More specifically, results from this model may be useful in helping to direct other intelligence-gathering exercises, especially where an individual who is identified as being a potential high-harm offender has not been identified as a target using more conventional intelligence-gathering methods.

A related issue is that of implementation and how to operationalise the results from this study. First, it is important to note that the data used for this study were reflective of OMCG composition and offending at the time of data extraction. Unlike other models, such as those designed to predict domestic violence repeat offending, the model was not an assessment of the likelihood of further offending by an OMCG member when they came into contact with law enforcement for an offence. This means that, as time progresses and the information available changes, these findings may date. These models are not intended to be perpetual. This approach requires continual updating. This would require the modelling to be reproduced to generate contemporary assessments with the intention of assisting decision-making for the operational tasking of police.

There are also questions about what specific action should be taken in response to risk assessment, which is a critical consideration when it comes to the effective implementation of these models. Clearly, the reference period used in this study—whether the shorter two-year period or more optimal five-year period—is not intended to be used to guide decisions about the duration of any proactive targeting of individuals or the clubs to which they belong. To do so over an extended period would be highly inefficient and unethical. Rather, results can be used to help guide the focus of other targeted intelligence-gathering strategies, which could then inform policing activity. They might also be used to help guide efforts to encourage disaffiliation from gangs, including through the relatively novel (in Australia) gang exit programs (see Boland et al. 2021). Other prevention strategies may also be informed by the outcomes of risk assessment, such as strategies to reduce recruitment into clubs with a high proportion of members at risk of committing high-harm offences. The principal aim of developing structured risk assessment is to encourage more nuanced responses that reflect the heterogeneity of clubs and club involvement in serious crime. Whatever strategies are developed and implemented, they should be rigorously evaluated, both as a way of testing the efficacy of risk assessment but also as a way of building an evidence base for reducing crime by OMCGs, which is sorely lacking (Dowling & Morgan 2022; van Ruitenburg & Blokland 2022).

Data availability and access is pivotal for a path forward

In the best performing state, the model correctly identified more than half of the OMCG members who went on to commit a high-harm offence. Identifying a significant proportion of high-harm offenders is central to this analysis. However, these results—and the comparison with results from the previous study—demonstrate the trade-offs associated with using less than optimal data, the benefit of collecting and linking data that can improve the predictive accuracy of risk assessment models, and the importance of alternative sources of intelligence.

To support the development of these types of approaches, and to limit the need for trade-offs in their implementation, more complete data is required. While this research used data that were readily available to policing agencies and which could be shared with researchers, it is also clear that there are data within the criminal justice system which, when linked, can improve the ability to anticipate high-harm offending by OMCG members. In Australia there are secure databases, such as the NCIS, that are government-regulated, contain comprehensive data relating to their subject matter, and are available to agencies that benefit from such access. National-level analyses are important, particularly for offending cohorts like OMCGs that move across jurisdictional borders (Dowling & Morgan 2021; Phythian & Kirby 2022). The additional information contained within the NCIS—not available in the data from the NGL and NPRS available for this study—may benefit this task.

However, this research also reinforces the importance of being able to link data between police and corrections for the purpose of risk assessment. Cubitt and Morgan (2022) used data on custodial episodes of OMCG members alongside offence-level data to predict high-harm offending. This cannot be done at a national level, nor is it as easy to achieve in some states as it is in places like New South Wales. Correctional data would be an invaluable addition to a national-level collection on police-recorded offending of OMCG members. It is also important that these linked data be available to state-level analyses of this kind. The findings of this research and other studies like it demonstrate how powerful prediction models can be in the analysis of crime and in supporting decisions for resource allocation or disruption approaches by police. More comprehensive data may yield stronger and more reliable risk assessments—so long, of course, as the relevant practical, legal and ethical requirements can be met.

The use of transparent machine learning in police settings

Notably, in mid-2023, police agencies from Australia and New Zealand endorsed the Australia New Zealand Policing Advisory Agency (ANZPAA) principles for the use of artificial intelligence (AI). ANZPAA proposed nine principles for the use of AI by policing agencies in the region, including transparency, explainability, accountability, skills and knowledge, human oversight, fairness, proportionality and justifiability, reliability, and privacy and security (see ANZPAA 2023 for further detail). These principles were proposed to guide agencies in the ethical and responsible use of AI, to minimise any potential harm associated with the use of AI, and to maintain community confidence in the adoption and deployment of AI systems by police.

Machine learning is a domain of AI, meaning that the analytics used to develop the risk assessments in this research should adhere to the nine principles endorsed by police agencies in Australia and New Zealand. When AI is used for decision-making, particularly relating to matters impacting the general public, there is a reasonable concern that—even among high-performing analytics—opaque models may hide biased results from scrutiny (Lo 2022). These 'black-box' type modelling procedures, common in AI, are emblematic of the concerns leading to the principles proposed by ANZPAA.

These principles helped guide the approach to this research. In producing this research, transparency was central to the task. It was important that we could interrogate the models produced to understand why they were more or less successful, and whether they adhered to conventional logic on what we know about OMCGs. If we were unable to measure the distribution of true and false positives and negatives, or the variables that were important in making predictions, we could not be certain if these models were trustworthy. The analyses presented in this report were designed with ethical and transparent use of machine learning in mind. Adhering to these principles can help ensure public trust, accountability and effective oversight of the implementation of AI models in police settings.

Conclusion

This research highlights the importance of giving attention to the practical application of risk assessment models developed for use by law enforcement to identify high-risk individuals, places and problems. Transparency is important in model development, as is using the best data available; however, the data available to law enforcement are often limited in terms of completeness, quality and accessibility. By reproducing a prediction model that was originally developed to predict high-harm offending among OMCGs, we show that it is possible to predict high-harm offending using machine learning with a relatively high degree of accuracy, including a low rate of false positives, using the data readily available to intelligence analysts. However, there are clear trade-offs in real-world applications of prediction models using operational data available from national police information systems.

References

URLs correct as at December 2023

ACIC—see Australian Criminal Intelligence Commission

Albanese JS 2008. Risk assessment in organized crime: Developing a market and product-based model to determine threat levels. *Journal of Contemporary Criminal Justice* 24(3): 263–273. https://doi.org/10.1177/1043986208318225

Andersen HA & Mueller-Johnson K 2018. The Danish Crime Harm Index: How it works and why it matters. *Cambridge Journal of Evidence-Based Policing* 2(1): 52–69. https://doi.org/10.1007/s41887-018-0021-7

ANZPAA—see Australian and New Zealand Policing Advisory Agency

Australian Criminal Intelligence Commission (ACIC) 2021. *2020–21 Annual report*. https://www.acic.gov.au/publications/annual-reports

Australian Criminal Intelligence Commission (ACIC) 2017. *Organised crime in Australia 2017*. https://www.acic.gov.au/publications/unclassified-intelligence-reports/organised-crime-australia

Australian Federal Police 2021. *AFP-led Operation Ironside smashes organised crime*. Media release, 8 June. https://www.afp.gov.au/news-centre/media-release/afp-led-operation-ironside-smashes-organised-crime

Australian and New Zealand Policing Advisory Agency (ANZPAA) 2023. *Police and artificial intelligence principles*. https://www.anzpaa.org.au/ArticleDocuments/2350/Australia New Zealand Police Artificial Intelligence Principles 2023.PDF.aspx

Ayling J & Broadhurst R 2014. Organized crime control in Australia and New Zealand. In L Paoli (ed), *The Oxford handbook of organized crime.* Oxford: Oxford University Press: 612–633

Bain A & Lauchs M 2017. *Understanding the outlaw motorcycle gangs*. Durham: Carolina Academic Press.

Barker T 2015. *Biker gangs and transnational organised crime*, 2nd ed. New York: Anderson Publishing

Barnes GC & Hyatt JM 2012. *Classifying adult probationers by forecasting future offending.* https://nij.ojp.gov/library/publications/classifying-adult-probationers-forecasting-future-offending

Bartels L, Henshaw M & Taylor H 2021. Cross-jurisdictional review of Australian legislation governing outlaw motorcycle gangs. *Trends in Organized Crime* 24(3): 343–360. https://doi.org/10.1007/s12117-021-09407-0

Berk R 2021. Artificial intelligence, predictive policing, and risk assessment for law enforcement. *Annual Review of Criminology* 4(1): 209–237. https://doi.org/10.1146/annurev-criminol-051520-012342

Berk R 2013. Algorithmic criminology. *Security Informatics* 2(1): 1–14. https://doi.org/10.1186/2190- 8532-2-5

Blokland A, Soudijn M & van der Leest W 2017. Outlaw bikers in the Netherlands: Clubs, social criminal organizations, or gangs? In A Bain & M Lauchs (eds), *Understanding the outlaw motorcycle gangs: International perspectives* Durham NC: Carolina Academic Press: 91–114

Blokland A, van Hout L, van der Leest W & Soudijn M 2019. Not your average biker: Criminal careers of members of Dutch outlaw motorcycle gangs. *Trends in Organized Crime* 22: 10–33. https://doi.org/10.1007/s12117-017-9303-x

Boland D et al. 2021. Effects of outlaw motorcycle gang membership and the support needs of former members. *Trends & issues in crime and criminal justice* no. 614. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78078

Bright D & Deegan S 2021. The organisational structure, social networks and criminal activities of outlaw motorcycle gangs: Literature review. *Trends & issues in crime and criminal justice* no. 621. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti04961

Bright D, Sadewo G, Cubitt TIC, Dowling C & Morgan A 2022. Co-offending networks among members of outlaw motorcycle gangs across types of crime. *Trends in Organized Crime* 2022. https://doi.org/10.1007/s12117-022-09467-w

Bright D, Sadewo G, Lerner J, Cubitt TIC, Dowling C & Morgan A 2023. Investigating the dynamics of outlaw motorcycle gang co-offending networks: The utility of relational hyper event models. *Journal of Quantitative Criminology* 2023. https://doi.org/10.1007/s10940-023-09576-x

Couronné R, Probst P & Boulesteix A-L 2018. Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics* 19(1): 270. https://doi.org/10.1186/s12859-018-2264-5

Cubitt TIC, Dowling C & Morgan A 2023. Crime by outlaw motorcycle gang members during club conflicts. *Trends & issues in crime and criminal justice* no. 667. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78900

Cubitt TIC & Morgan A 2022. Predicting high-harm offending using machine learning: An application to outlaw motorcycle gangs. *Trends & issues in crime and criminal justice* no. 646. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78528

Cubitt TIC, Wooden KR & Roberts KA 2020. A machine learning analysis of serious misconduct among Australian police. *Crime Science* 9(1): 22. https://doi.org/10.1186/s40163-020-00133-6

Curtis-Ham S & Walton D 2018. The New Zealand Crime Harm Index: Quantifying harm using sentencing data. *Policing: A Journal of Policy and Practice* 12(4): 455–467. https://doi.org/10.1093/police/pax050

Darroch S & Mazerolle L 2013. Intelligence-led policing: A comparative analysis of organizational factors influencing innovation uptake. *Police Quarterly* 16(1): 3–37. https://doi.org/10.1177/1098611112467411

Densley JA & Pyrooz DC 2020. The Matrix in context: Taking stock of police gang databases in London and beyond. *Youth Justice* 20(1–2): 11–30. https://doi.org/10.1177/1473225419883706

Dowling C, Boland D, Morgan A, Webster J, Chiu Y & Lowe R 2021. The changing culture of outlaw motorcycle gangs in Australia. *Trends & issues in crime and criminal justice* no. 615. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78054

Dowling C & Morgan A 2022. Regulatory approaches to preventing organised crime among outlaw motorcycle gangs. *Trends & issues in crime and criminal justice* no. 652. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78665

Dowling C & Morgan A 2021. Criminal mobility of outlaw motorcycle gangs in Australia. *Trends & issues in crime and criminal justice* no. 619. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti04992

Etzler S, Schönbrodt F, Pargent F, Eher R & Rettenberger M 2022. Machine learning and risk assessment: Random forest does not outperform logistic regression in the prediction of sexual recidivism. *PsyArXiv*. https://doi.org/10.31234/osf.io/z6ky2

Francis B & Humphreys L 2016. Investigating the relationship between the diversity index and frequency of offending. *Journal of Developmental and Life-Course Criminology* 2(4): 397–416. https://doi.org/10.1007/s40865-016-0034-5

Francis B, Humphreys L, Kirby S & Soothill K 2013. *Understanding criminal careers in organised crime*. Home Office Research Report no. 74. London: Home Office. https://www.gov.uk/government/publications/understanding-criminal-careers-in-organised-crime

Fuller G, Morgan A & Brown R 2019. Criminal histories of Australian organised crime offenders. *Trends & issues in crime and criminal justice* no. 567. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti09623

Goldsworthy T & McGillivray L 2017. An examination of outlaw motorcycle gangs and their involvement in the illicit drug market and the effectiveness of anti-association legislative responses. *International Journal of Drug Policy* 41: 110–117. https://doi.org/10.1016/j.drugpo.2016.12.009

Guay J-P 2012. *Predicting recidivism with street gang members*. Corrections Research: User Report. Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2012-02-prsgm/index-en.aspx

Hollywood JS & Winkelman Z 2015. *Improving information-sharing across law enforcement: Why can't we know?* RAND Corporation. https://www.rand.org/pubs/research_reports/RR645.html

Hong H, Xioling G & Hua Y 2016. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on random forest. Paper to 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 26–28 August

Hosmer DW & Lemeshow S 2004. *Applied logistic regression*. New Jersey: John Wiley & Sons

House PD & Neyroud PW 2018. Developing a crime harm index for Western Australia: The WACHI. *Cambridge Journal of Evidence-Based Policing* 2(1–2): 70–94. https://doi.org/10.1007/s41887-018- 0022-6

Huey L, Ferguson L & Koziarski J 2022. The irrationalities of rationality in police data processes. *Policing and Society* 32(8): 947–962. https://doi.org/10.1080/10439463.2021.2007245

Huff CR & Barrows J 2015. Documenting gang activity: Intelligence databases. In SH Decker & DC Pyrooz (eds), *The handbook of gangs.* John Wiley & Sons: 59–77. https://doi.org/10.1002/9781118726822.ch4

Hyndman RJ & Anthanasopoulos G 2014. *Forecasting: Principles and practice*. Melbourne: Monash University

Kärrholm F, Neyroud P & Smaaland J 2020. Designing the Swedish Crime Harm Index: An evidence-based strategy. *Cambridge Journal of Evidence-Based Policing* 4(1): 15–33. https://doi.org/10.1007/s41887-020-00041-4

Kennedy DM 2009. Gangs and public policy: Constructing and deconstructing gang databases. *Criminology and Public Policy* 8(4): 711–716. https://doi.org/10.1111/j.1745-9133.2009.00587.x

Klement C 2016. Outlaw biker affiliations and criminal involvement. *European Journal of Criminology* 13(4): 453–472. https://doi.org/10.1177/1477370815626460

Koper CS, Lum C & Willis JJ 2014. Optimizing the use of technology in policing: Results and implications from a multi-site study of the social, organizational, and behavioural aspects of implementing police technologies. *Policing* 8(2): 212–221. https://doi.org/10.1093/police/pau015

Lauchs M 2017. Nike bikies. In A Bain & M Lauchs (eds), *Understanding the outlaw motorcycle gangs: International perspectives*. Durham NC: Carolina Academic Press: 115–138

Lauchs M, Bain A & Bell P 2015. *Outlaw motorcycle gangs: A theoretical perspective*. London: Palgrave Macmillan

Lo FTH 2022. The paradoxical transparency of opaque machine learning. *AI & Society*. https://doi.org/10.1007/s00146-022-01616-7

Lum C, Koper CS & Willis J 2017. Understanding the limits of technology's impact on police effectiveness. *Police Quarterly* 20(2): 135–163. https://doi.org/10.1177/1098611116667279

Mitchell RJ 2019. The usefulness of a crime harm index: Analyzing the Sacramento Hot Spot Experiment using the California Crime Harm Index (CA-CHI). *Journal of Experimental Criminology* 15(1): 103–113. https://doi.org/10.1007/s11292-017-9318-y

Morgan A, Cubitt T & Dowling C 2023. Outlaw motorcycle gangs and domestic violence. *Trends & issues in crime and criminal justice* no. 670. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78962

Morgan A, Dowling C & Voce I in press. Outlaw motorcycle gangs in Australia: Exploring variability in gang member involvement in organised crime. In A Blokland & K von Lampe (eds), *Outlaw bikers as organized crime*. Milton Park: Routledge

Morgan A, Dowling C & Voce I 2020. Australian outlaw motorcycle gang involvement in violent and organised crime. *Trends & issues in crime and criminal justice* no. 586. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti04282

O'Connor CD, Ng J, Hill D & Frederick T 2022. Police analysts on the job in Canada: Work experiences, data work, and the move towards evidence-based policing. *Police Practice and Research* 23(4): 458–472. https://doi.org/10.1080/15614263.2021.2022483

O'Connor CD, Ng J, Hill D & Frederick T 2021. Thinking about police data: Analysts' perceptions of data quality in Canadian policing. *The Police Journal: Theory, Practice and Principles.* Advance online publication. https://doi.org/10.1177/0032258X211021461

Peels R 2019. Replicability and replication in the humanities. *Research Integrity and Peer Review* 4(1): Article 2. https://doi.org/10.1186/s41073-018-0060-4

Phythian R & Kirby S 2022. What does the UK Police National Database tell us about the future of police intelligence? *Policing: A Journal of Policy and Practice* 17: paac074. https://doi.org/10.1093/police/paac074

Quinn JF & Forsyth CJ 2009. Leather and Rolexs: The symbolism and values of the motorcycle club. *Deviant Behavior* 30(3): 235–65. https://doi.org/10.1080/01639620802168700

Ratcliffe JH & Kikuchi G 2019. Harm-focused offender triage and prioritization: A Philadelphia case study. *Policing: An International Journal* 42(1): 59–73. https://doi.org/10.1108/PIJPSM-08-2018-0118

Ratcliffe JH, Strang SJ & Taylor RB 2014. Assessing the success factors of organized crime groups: Intelligence challenges for strategic thinking. *Policing: An International Journal of Police Strategies & Management* 37(1): 206–227. https://doi.org/10.1108/PIJPSM-03-2012-0095

Rostami A & Mondani H 2019. Organizing on two wheels: Uncovering the organizational patterns of Hells Angels MC in Sweden. *Trends in Organized Crime* 22(3): 34–50. https://doi.org/10.1007/s12117-017-9310-y

Sanders CB, Weston C & Schott N 2015. Police innovations, 'secret squirrels' and accountability: Empirically studying intelligence-led policing in Canada. *The British Journal of Criminology* 55(4): 711–729. https://doi.org/10.1093/bjc/azv008

Schonlau M & Zou RY 2020. The random forest algorithm for statistical learning. *The Stata Journal: Promoting communications on statistics and Stata* 20(1): 3–29.

Sentas V & Pandolfini C 2017. *Policing young people in NSW: A study of the Suspect Targeting Management Plan*. Sydney: Youth Justice Coalition

Sherman L, Neyroud PW & Neyroud E 2016. The Cambridge Crime Harm Index: Measuring total harm from crime based on sentencing guidelines. *Policing: A Journal of Policy and Practice* 10(3): 171–183. https://doi.org/10.1093/police/paw003

Singh JP, Grann M & Fazel S 2011. A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review* 31(3): 499–513. https://doi.org/10.1016/j.cpr.2010.11.009

Svetnik V, Liaw A, Tong C, Culberson C, Sheridan RP & Feuston BP 2003. Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Science* 43(6): 1947–1958. https://doi.org/10.1021/ci034160g

Tremblay P et al. 1989. Carrières criminelles collectives: Évolution d'une population délinquante (les groupes de motards). *Criminologie* 22(2): 65–94. https://doi.org/10.7202/017282ar

United Nations Office on Drugs and Crime 2019. *Transnational organised crime in Southeast Asia: Evolution, growth and impact*. https://www.unodc.org/unodc/en/frontpage/2019/July/organised-crime-syndicates-are-targeting-southeast-asia-to-expand-operations_-unodc.html

Valasik M 2018. Gang violence predictability: Using risk terrain modeling to study gang homicides and gang assaults in East Los Angeles. *Journal of Criminal Justice* 58: 10–21. https://doi.org/10.1016/j.jcrimjus.2018.06.001

van Deuren S, Kleemans E & Blokland A 2020. Outlaw motorcycle gangs and their members' crime: Examining the social organization of crime and its relationship to formal club hierarchy. *European Journal of Criminology* 19(6): 1461–1482. https://doi.org/10.1177/1477370820980440

van Ruitenburg T & Blokland A 2022. The Dutch approach to outlaw motorcycle gangs. *Trends & issues in crime and criminal justice* no. 655. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78726

Voce I, Morgan A & Dowling C 2021. Early-career offending trajectories among outlaw motorcycle gang members. *Trends & issues in crime and criminal justice* no. 625. Canberra: Australian Institute of Criminology. https://doi.org/10.52922/ti78030

von Lampe K & Blokland A 2020. Outlaw motorcycle clubs and organized crime. *Crime and Justice* 49: 521–578. https://doi.org/10.1086/708926

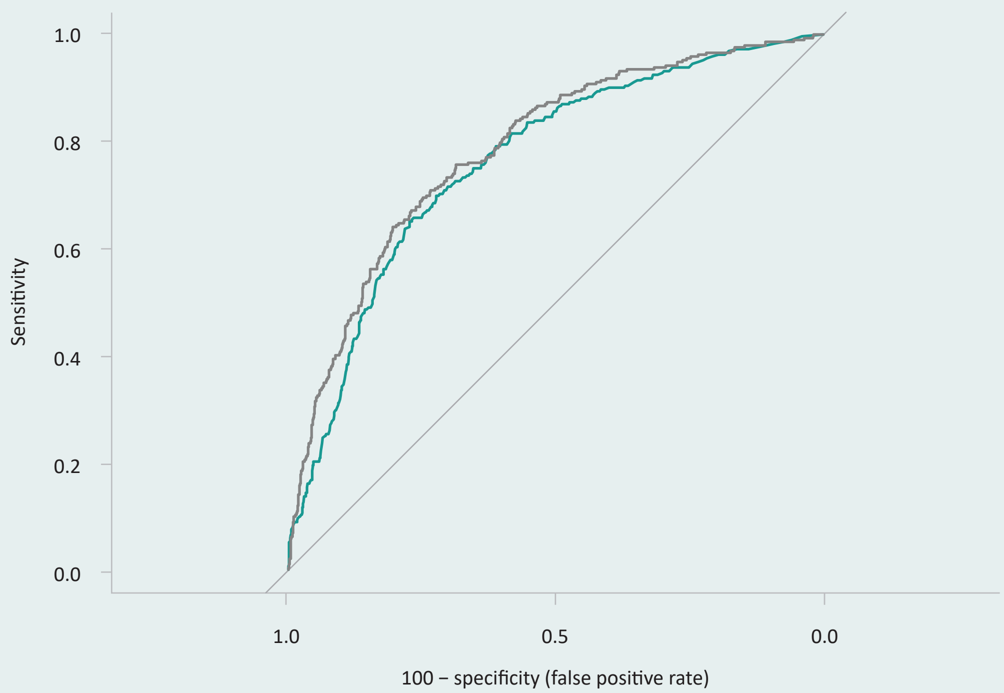
Wynants L, van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW & Van Calster B 2019. Three myths about risk thresholds for prediction models. *BMC Medicine* 17: Article 192. https://doi.org/10.1186/s12916-019-1425-3

Yeong S 2021. An evaluation of the Suspect Target Management Plan. *Crime and Justice Bulletin* no. 233. Sydney: NSW Bureau of Crime Statistics and Research

Appendix: Removing individuals with no prior recorded offending

This additional analysis limited the sample to outlaw motorcycle gang (OMCG) members with at least one recorded criminal offence (*n*=4,050). In computing this model, the random forest produced an out-of-bag error of 22.2 percent, and an area under the receiver operating characteristic (AUROC) of 0.785, marginally outperforming the logistic regression (AUROC=0.762, $p<0.05$). The receiver operating characteristic (ROC) curves are provided in Figure A1.

Figure A1: Receiver operating characteristic (ROC) curve for random forest (grey) and logistic regression (green) predicting high-harm offending among OMCG affiliates using a restricted sample



Source: OMCG criminal history database [computer file]

The aggregate classification error on the test set was 23.5 percent. Table A1 shows that the model again performed well in predicting affiliates who would not commit a high-harm offence (with a false positive rate of 9.9%); however, it was less accurate in classifying affiliates who did commit a high-harm offence, with a false negative rate of 65.8 percent. These results indicate that there was little advantage in limiting the sample to OMCG members with a prior recorded criminal history. In fact, models generally performed slightly worse when the sample was restricted. Results for the states using the restricted sample are available on request.

On the basis of this analysis, the models presented in the main report are based on the full sample of OMCG members, irrespective of whether they had a prior recorded criminal history.

Table A1: Confusion matrix for random forest model trained on high-harm offending using restricted sample

		Actual high-harm offender?		Classification errors
		No	Yes	
Predicted high-harm offender?	No	831	192	18.8%
	Yes	92	100	48.0%
Classification errors		9.9%	65.8%	23.5%

Source: OMCG criminal history database [computer file]

AIC reports
Research Report

Dr Timothy Cubitt is Acting Research Manager of the Online Sexual Exploitation of Children research program at the Australian Institute of Criminology.

Anthony Morgan is the Research Manager of the Serious and Organised Crime, Cybercrime and Radicalisation research program at the Australian Institute of Criminology.

Australia's national research and
knowledge centre on crime and justice

www.aic.gov.au