

RESEARCH

Open Access



A machine learning analysis of serious misconduct among Australian police

Timothy I. C. Cubitt^{*} , Ken R. Wooden and Karl A. Roberts

Abstract

Fairness in policing, driven by the effective and transparent investigation and remediation of police misconduct, is vital to maintaining the legitimacy of policing agencies, and the capacity for police to function within society. Research into police misconduct in Australia has traditionally been performed on an ad-hoc basis, with limited access to law enforcement data. This research seeks to identify the antecedents of serious police misconduct, resulting in the dismissal or criminal charge of officers, among a large police misconduct dataset. Demographic and misconduct data were sourced for a sample of 600 officers who have committed instances of serious misconduct, and a matched sample of 600 comparison officers across a 13-year period. A machine learning analysis, random forest, was utilised to produce a robust predictive model, with Partial Dependence Plots employed to demonstrate within variable interaction with serious misconduct. Prior instances of serious misconduct were particularly predictive of further serious misconduct, while misconduct was most prominent around mid-career. Secondary employment, and performance issues were important predictors, while demographic variables typically outperformed complaint variables. This research suggests that serious misconduct is similarly prevalent among experienced officers, as it is junior officers, while secondary employment is an important indicator of misconduct risk. Findings provide guidance for misconduct prevention policy among policing agencies.

Introduction

Police accountability is a notion, typically predicated on the investigation of officers for instances of misconduct. This process, by which complaints are made against officers, and subsequently investigated, allows policing agencies to remediate poor behaviour among employees and demonstrate fairness and accountability to the public (Walker and Archbold 2005). In Australia, this process frequently comprises of the receipt of a complaint against an officer, the investigation of this complaint, a decision on whether the complaint is substantiated and if so, the imposition of disciplinary or remedial action against the officer (NSW Police Force 2012). Fairness in policing, imbued by the effective and transparent investigation and remediation of police misconduct is vital to maintaining the legitimacy of police forces, and public consent

to the duties of police. One of the most effective ways to improve police misconduct, and subsequent fairness, is to develop a strong understanding of the antecedents to misconduct as a means of developing evidence-based detection, prevention and if necessary intervention policy for at-risk officers (Quispe-Torrealblanca and Stewart 2019). The current study seeks to identify antecedents of serious misconduct among police, defined as misconduct deemed to require consideration of dismissal or the criminal charge of a police officer. The relative power of these antecedents in predicting serious misconduct among officers will be considered, as a means of identifying the more important factors in disrupting police misconduct. The role of police in adequately managing misconduct of their workforce is pivotal to their function, as recent events demonstrate, where agencies do not effectively manage the misconduct of their officers, there may be devastating outcomes (Weitzer 2002).

*Correspondence: 17639226@student.westernsydney.edu.au
Western Sydney University, Parramatta, NSW 2150, Australia



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

There is a noteworthy body of research into police misconduct, often with a focus on the procedural and demographic aspects of officers, this research has yielded important insight. Quispe-Torreblanca and Stewart (2019) considered the network of workplace peer groups, particularly moving between peer groups, on incidence of misconduct. This network structure suggested that agencies may be susceptible to networks of misconduct, however misconduct is often associated with individual variables such as race, age and tenure of officers (Wood et al. 2019). Historically, the assessment of police misconduct has centred on investigation of individual areas of misconduct, or associated variables. Prior research has included disciplinary records of prior employment (Cohen and Chaiken 1973; Kane and White 2009; Greene et al. 2004), personality traits (Cutler and Muchinsky 2006), and criminal records (Greene et al. 2004; Kane and White 2009).

Correlates of police misconduct

As a byproduct of the policing environment, large quantities of data are typically produced, although access for research purposes is often limited. Despite limited access, there is a notable body of work seeking to identify important correlates of misconduct among police. Extant literature typically categorises misconduct prone officers as inexperienced (Greene et al. 2004; Harris 2008; Harris and Worden 2012) and male (Greene et al. 2004; Walker and Archbold 2005). Misconduct risk has been noted as particularly acute early in age and career (Harris 2008; Kane and White 2009), an important finding allowing for training and misconduct prevention to be a directed activity among agencies. Further, the length of time that an officer remains in the same workplace, termed police tenure, has been related to increased risk of misconduct (Alston 2010). However, the drivers of this risk are less clear, with some suggestion that extended tenure in duties likely to accrue trauma exposure, may result in health, emotional and performance impairment (Thorn-ton and Herndon, 2015). Among external factors to the workplace, secondary employment has been broached as a misconduct risk, focusing on the type of secondary employment undertaken (Kirsch 2014). This leads to the notion of fatigue, particularly as a performance related misconduct risk (Pearsall 2012; Senjo 2011).

The complaint history of officers, resulting from both on and off duty behaviour, is a rich data source regarding the antecedents to serious misconduct. Kane and White (2009) provide an extensive overview of career ending misconduct among police officers in New York. An important finding from this work were factors external to the workplace resulting in dismissal of police officers, particularly drug offences and profit related crime.

However, it was similarly important that administrative misconduct and failure to perform duties appropriately were prevalent among misconduct prone officers (Kane and White 2009).

An important piece of work in the identification of risk factors for misconduct was produced by Worden, Harris and McLean (2014), in which improper use of force, and poor performance were identified as predictors. These variables have been explored throughout literature, much of which has identified elevated use of force among officers as a risk factor (Lersch et al. 2006). Improper use of force has been embraced as a strong indicator of misconduct risk, however it is important to note that it is difficult to isolate individually as a predictor (Bazley et al. 2009). This is noteworthy in the context of the present study, there have been several instances of recent criminal charges among officers relating to improper use of force in New South Wales (NSW), the jurisdiction in which this research is situated (O'Mallon 2017; Owen 2017; Parish 2018; Le Lievre 2017; Mackenzie and White 2018).

Machine learning to analyse misconduct

A novel approach for misconduct risk identification was suggested by Helsby et al. (2018), utilising a range of data sources and introducing machine learning methods to the area of study. This approach resulted in a noteworthy success rate in predicting adverse interactions between police and members of the public (Helsby et al. 2018). It appears, however that current iterations of misconduct intervention models fail to comprehensively identify factors associated with misconduct, and subsequently have some difficulty making predictions (James, James and Dotson 2020; Bazley et al. 2009; Porter et al. 2012). The present study seeks to address this gap through analysis of comprehensive, novel data generated by the largest policing agency in Australia.

Machine learning analyses show substantial promise in the prediction of serious misconduct (Helsby et al. 2018). However, machine learning, and particularly random forest models have been used for measurement and prediction in criminology for some time. Classification trees, specifically random forest models, are suggested to determine predictive power with greater accuracy than linear modelling (Berk 2013; Breiman 2001). To achieve this goal, data sources are arranged into an ensemble of classification trees to inductively discern nonlinear functions and interactions among variables (Berk 2013). Similar models have been used in literature to predict recidivism among homicide offenders (Neuilly et al. 2011), recidivism among offenders while on parole (Berk et al. 2009), and domestic violence offending (Berk et al. 2016). Machine learning analytics have been projected to play a

considerable role in the future of criminological analysis (Chan and Bennett Moses 2015). Given the complexity of interactions present in datasets sourced from policing agencies, particularly regarding police misconduct across an extended time-period, machine learning analytics offer a strong alternative to traditional analyses.

New South Wales policing jurisdiction

While there is noteworthy research among Australian policing jurisdictions, data is typically sourced externally to policing agencies (Goodman-Delahunty 2014; Crehan 2019), utilising qualitative methodologies (Hine et al. 2018), or small samples (People 2008; People et al. 2010; Gorta 2009). While these studies provide insight into the misconduct environment, there is a paucity of empirical analysis of substantial police misconduct datasets.

While the New South Wales Police Force (NSWPF) report complaint rates on an annual basis (NSW Police Force 2017), experiential and qualitative data are frequently relied upon for insight (Goodman-Delahunty et al. 2014; Zulfacar et al. 2012). The present study is the first time that unique data sources such as these, including the complete complaint histories of officers, have been utilised in Australia. The intention was to assess these data for variables which interact with serious police misconduct. Serious misconduct in this jurisdiction is defined by Section 10 of the Law Enforcement Conduct Commission Act 2016 as ‘misconduct... that could result in prosecution of the officer... for a serious offence or serious disciplinary action against the officer... for a disciplinary infringement’ (Law Enforcement Conduct Commission Act 2016 (NSW)). Serious misconduct is therefore identified as the most serious end of the scale, requiring consideration of dismissal by the NSW Police Commissioner, or criminal charges brought against an officer. All instances of serious misconduct are considered by an internal police panel of senior officers, outcomes may consist of remedial action, punitive action or dismissal (NSW Police Force 2012). The intention of this study was to consider features among the demographics and misconduct history of officers, that interact with serious misconduct. Specifically, the following research questions are to be answered:

1. Are prior instances of misconduct useful in predicting serious misconduct?
2. What career stage are officers most at risk of serious misconduct?
3. Are there demographic features which are notably predictive of serious misconduct?

Data description

Data collection was facilitated by the Professional Standards Command of the NSWPF. Officer demographic and complaint histories were sourced from the misconduct database held securely on-site. Substantiated instances of misconduct were sourced by type of misconduct. A count of unsubstantiated complaints was also included as an indicative variable for the volume of complaints an officer received. Additional file 1 provides a list of variables available for analysis, and their definitions.

Data limitations

Data entry error is a substantial limitation among policing agencies around the world. Data entry occurs manually by typically time-poor staff, resulting in large scale inaccuracy (Helsby et al. 2018). Data cleaning was undertaken by cross referencing the dataset with original complaint documentation for each complaint included, to ensure each complaint was correctly categorised. It is a limitation that each complaint is assumed to be investigated on an equitable basis, however, it was not methodologically sound to independently evaluate the investigative rigour of a policing agency, as such substantiated complaints were the unit of measure available. There may have been utility in including data relating to instances of workplace injuries, however this data was not available for analysis.

Methodology

This study is best described as a secondary data analysis, applying machine learning analytics to a naturally occurring data (Lester et al. 2017). A maximum sample size of 1200 officers was available, which was determined sufficient to satisfy the methodology (Morgan and Wilson Van Voorhis 2007; Green 1991). Officers who have been subject to substantiated findings of serious misconduct ($n=600$) were randomly sampled from between January 2003 and October 2016, this constituted a 30% sample of officers who have committed serious misconduct across this time period. Data access was negotiated with the host organisation, however access was contingent on a maximum sample size of 30% across the sample period. To provide an appropriate comparison, this sample was matched to a comparison sample of 600 officers. While a stratified matching process was available, the sample and analysis of Kane and White (2009) among the New York Police Department was notably similar to the present study, as such this matching method was replicated. Each officer was deidentified and matched to a randomly selected officer from their academy class, through

randomisation of the unique NSWPF officer identification number.

Machine learning techniques have been used in interrogating data in policing and criminology for some time now (Berk et al. 2009, 2016; Helsby et al. 2018; Neuilly et al. 2011). This is largely due to the ability to identify underlying interactions and functions among data, and in the case of the random forest, ensemble classification trees are particularly good at identifying nonlinear interactions (Berk 2013) among data generated by humans. For example, the analysis produced by Helsby et al. (2018), with a similar intention to the present study, concluded that the random forest outperformed other machine learning methods in prediction of adverse police interactions. While the random forest algorithm has been identified as performing significantly better in classification exercises than logistic regression (Couronne et al. 2018), it is still an emerging methodology in this area of work. As a means of improving accessibility of results, findings of the random forest analysis here are presented alongside a logistic regression model, also serving to provide a comparator for model performance.

In accurately replicating the analytical methodology suggested by Helsby et al. (2018) among the current dataset, a ROC curve is provided as Fig. 1 with the Area Under the Curve (AUROC) to demonstrate robustness of the random forest model. The ROC curve identifies the true positive rate of classification (y-axis), compared with the false positive rate (x-axis). The closer the curve to 45 degrees, the less accurate the model. To take this analysis one step further, Partial Dependence Plots (PDP) are provided for important career variables, to show at which point in their career, officers were most associated with serious misconduct (Zhao and Hastie 2019). PDPs provide the logit contribution of the variable under consideration to the probability of classification among the

dependent variable, relative to the gini coefficient from the random forest model. In simple terms, PDPs allow for a more granular understanding of the effect within the ranges of important variables, where otherwise we may only know that the variable itself is important. For example, they may be used to show the specific age most associated with serious misconduct, and the age that is least. Where plotlines show decline, the interaction with serious misconduct is reduced, where plotlines show an incline, the interaction is increased by the logit value of the y-axis.

Modelling was performed through application of the framework for preprocess design matrices, the dataset was partitioned into 70% training set and 30% test set. Analysis was performed using statistical analysis software, R, using the 'randomForest', 'dplyr', 'pRoc', 'pdp' and 'ggplot2' packages. The random forest model was trained on findings of serious misconduct prior to exposing the model to the test set (Hyndman and Anthanasopoulos 2014). Variable importance is interpreted as Mean Decrease Gini (MDG) (Hong et al. 2016). The Gini coefficient is a measure of statistical dispersion in which coefficients resulting from random forest analysis are interpreted as a proportion of the overall model, higher coefficients indicate greater predictive power. Finally, a confusion matrix was produced to identify the prediction accuracy of the random forest model for serious misconduct in the test set. The confusion matrix compares the predicted outcomes from the random forest model, to the observed outcomes (Barnes and Hyatt 2012), providing a measure of whether the model made a correct selection, or whether it failed to predict serious misconduct among officers, and the rate at which this occurred.

Results

It was immediately clear that this was a particularly robust model, with an AUROC of 0.97, however it was also clear that one variable outperformed all others. This variable related to a prior instance of serious misconduct (MDG=0.37). The reduced burden of proof among departmental investigations, when compared with criminal charges, means officers may be found not-guilty at court while still having an instance of serious misconduct substantiated against them. Alternatively, after considering an officer for dismissal, the NSW Police Commissioner may retain them as officers, while substantiating an instance of serious misconduct against them. Subsequently, an officer may have a prior instance of serious misconduct, while retaining their position. The finding that a prior instance of serious misconduct was predictive of further serious misconduct is important, however there was further insight to be gained from this dataset. While this finding will be considered in the discussion, to

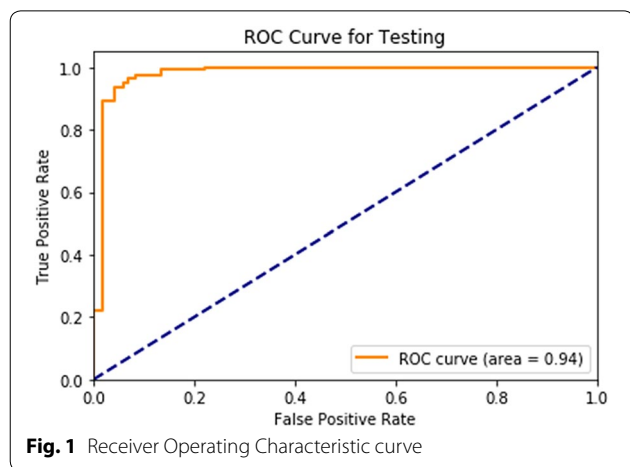


Fig. 1 Receiver Operating Characteristic curve

facilitate further insight from the available dataset, this variable was removed from the random forest model. The random forest was then performed again utilising the same method and the retained variables, which fell into three distinct categories, demographic variables, misconduct process variables, and prior minor misconduct. The subsequent analysis resulted in a marginally less robust model, as may be expected. The model was still robust with an AUROC of 0.94, as shown in Fig. 1.

Demographics

Officers from the serious misconduct group were on average 37.2 years old. Males were overrepresented in the misconduct group, as were general duties officers, while Constables and Senior Constables dominated the cohort. Misconduct prone officers had typically spent 5.9 years at their rank, with an average length of service of 12.6 years prior to serious misconduct. Detectives accounted for 13% of the serious misconduct sample, while 50.2% had been pursued for dismissal, and 72.1% had been subject to remedial action for prior misconduct. After consideration for serious misconduct, 46% of officers were still employed. The majority of officers were employed in metropolitan areas, while secondary employment

featured among 44.3%. A comparison of demographic data is provided in Table 1.

As detailed by Table 2, improper use of force, performance issues, issues with investigations, and customer service were comparatively elevated among the serious misconduct group. It also appears that there is potential for officers to be found to commit serious misconduct more than one time.

Random forest

The random forest model yielded a ranking of variables that were most associated with serious misconduct (Table 3), while the associated MDG coefficient demonstrated relative power of each variable. These findings are presented alongside a logistic regression (Table 4), to provide model comparison. While the logistic regression was marginally less robust than the random forest model (AUROC = 0.85), it still correctly classified 85% of cases. This additional analysis allows for consideration of directionality and significance among variables, features that are not typically available among random forest models. Table 3 shows demographic variables with strong interactions with serious misconduct, among these the strongest was secondary employment, this variable bears considerable note from this analysis. Secondary employment here did not relate to an instance of misconduct,

Table 1 Demographic data

	Serious misconduct group (n = 600)	Comparison group (n = 600)
Age	37.2 years	38.1 years
Gender	21.7% female 78.3% male	23.5% female 76.5% male
Duty type	General Duties: 65.8% Specialist Command: 34.2%	General Duties: 55% Specialist Command: 45%
Final rank at time of analysis	Constable: 27% Senior Constable: 57% Sergeant: 13% Commissioned Officer: 3%	Constable: 23% Senior Constable: 60% Sergeant: 15% Commissioned Officer: 3%
Years at rank	5.9 years	6.3 years
Length of service	12.6 years	12.87 years
Socio-economic status per geographic area	5.76 (out of 10)	6 (out of 10)
Income	\$88,391.70 per year	\$89,575.00 per year
Designated Detective	13%	10%
Awards or compliments	7.23 per officer	6.68 per officer
Dismissal proceedings	50.2%	0%
Prior remedial action for misconduct	72.1%	6%
Currently active employee	46.3%	64%
Secondary employment	44.3%	9.5%
Work in metropolitan areas	60%	62%

Table 2 Average instances of prior misconduct per officer

Complaints	Serious misconduct (Range)	Comparison (Range)
Improper use of force	0.29 (0–4)	0.12 (0–3)
Firearm related complaint	0.11 (0–2)	0.02 (0–2)
Adverse discharge of a Taser	0.005 (0–1)	0
Vehicle pursuit	0.04 (0–2)	0
Drink driving	0.17 (0–3)	0.03 (0–2)
Management action	3.75 (0–26)	1.75 (0–14)
Performance issues	0.40 (0–6)	0.09 (0–2)
Improper association	0.07 (0–2)	0.01 (0–1)
Complaint related to court	0.45 (0–8)	0.22 (0–4)
Incident during arrest or custody	0.21 (0–5)	0.09 (0–1)
Alleged corruption	0.11 (0–2)	0.01 (0–1)
Local Area Commander Warning Notice	0.26 (0–2)	0.03 (0–1)
Region Commander Warning Notice	0.23 (0–3)	0.03 (0–1)
Positive drug test	0.03 (0–1)	0
Issues with an investigation	0.55 (0–7)	0.3 (0–4)
Customer service	0.69 (0–12)	0.36 (0–4)
Harassment, intimidation or bullying	0.31 (0–5)	0.12 (0–4)
Untruthfulness/dishonesty	0.32 (0–4)	0.11 (0–3)
Misuse of powers	0.11 (0–2)	0.08 (0–2)
Disobey reasonable direction	0.14 (0–7)	0.08 (0–2)
Unauthorised absence	0.12 (0–3)	0.06 (0–1)
Property or exhibits complaint	0.245 (0–3)	0.13 (0–2)
Access and disclosure of confidential information	0.325 (0–4)	0.12 (0–2)
Breach code of conduct	0.335 (0–8)	0.05 (0–1)
Social media related	0.026 (0–2)	0
Workplace transfers	3.94 (0–16)	3.63 (0–15)
Unsubstantiated complaints	5.19 (0–55)	3.58 (0–25)
Total instances of serious misconduct	1.79 (1–13)	0

rather the presence of approved secondary employment on the employment record of an officer. Awards issued to officers, age and length of service were also important. General duties featured highest among duty types, while Senior Constables were the only rank to feature.

The complaint type with the strongest interaction with serious misconduct was issues with an investigation, followed by improper use of force. Drink driving was a notable variable, while harassment, bullying and intimidation was also important. Table 3 details the 28 variables that had the strongest interactions with serious misconduct. These variables constitute 96.7% of the random forest model, the remaining independent variables constituted 3.3% of the model and therefore did not return a noteworthy interaction with serious misconduct.

To fully explore the prediction accuracy of the random forest model, a confusion matrix was produced for the test split of the dataset (Table 5). Here, two green cells are provided, identifying where the random forest correctly predicted serious misconduct, and two red

cells identifying where the model failed. In attempting to distinguish between officers who would commit serious misconduct, and those who would not, this model did remarkably well, identifying 94% of officers correctly.

Partial dependence plots

Partial dependence plots were generated for four noteworthy demographic variables, age, length of service, time spent at rank and unsubstantiated complaints. It is important to note that the changes in each PDP are relatively minor, suggesting that risk of misconduct is present at most points within these variables, however there are clear peaks and troughs. Figure 2 provides the PDP for the interaction between the age of an officer, and serious misconduct. From this plot it is clear that among officers in their late 20's, into early 30's feature lower misconduct risk. However, around the mid 30's the likelihood of serious misconduct increases, peaking in the early 40's. While there was a slight decline from this age onward, the effect remained stable, suggesting

Table 3 Variable importance produced by Random Forest model trained on findings of serious misconduct

Variable	Mean Decrease Gini (MDG)
Demographic variables	
Secondary employment	0.13949791
Awards or complimentary remarks	0.12657524
Age	0.06176403
Length of service	0.05572832
Socio-economic status	0.03953699
Years at rank	0.03666515
Transfers	0.03130221
Duty type: General duties	0.00876206
Rank: Senior Constable	0.00698986
Misconduct process variables	
Unsubstantiated complaints	0.06648072
Local Area Commander Warning Notice	0.05990792
Region Commander Warning Notice	0.05457879
Performance or conduct management plan	0.0249561
Management action	0.02337803
Prior minor misconduct	
Issues with an investigation	0.03136233
Improper use of force	0.02252512
Drink driving	0.02020685
Complaint related to court	0.01807019
Breach code of conduct	0.01761054
Harassment, intimidation or bullying	0.01534989
Allegation of corruption	0.01431663
Positive drug test	0.0108887
Disobey reasonable direction	0.01070835
Firearm related complaint	0.00983039
Untruthfulness/Dishonesty	0.00948597
Rudeness to members of the public	0.00902758
Breach privacy, unauthorised disclosure of data	0.00840011
Property or exhibits complaint	0.00767745

that officers in their late 30's to early 40's were at higher risk of serious misconduct. However, it is important to note that these changes are relatively small. While some age groups are at greater risk of serious misconduct than others, there is not an age at which there is little or no risk of misconduct.

The PDP for length of service demonstrated a steady increase in the likelihood of serious misconduct across time, noticeably increasing around 15 years of service. After 17 years there is a marginal decrease, before finally around 26 years of service, risk of serious misconduct increases again (Fig. 3).

The years spent at current rank provided a measure of career stagnation among officers, by showing how

Table 4 Logistic regression estimates and odds ratios predicting serious misconduct among police

Variable	β	SE	OR
Demographic variables			
Secondary employment	1.56	0.09	4.38**
Awards or complimentary remarks	0.19	0.19	1.22**
Age	0.44	0.07	1.53*
Length of service	0.31	0.16	1.37*
Socio-economic status	- 0.72	0.27	0.48*
Years at rank	0.29	0.09	1.39
Transfers	- 0.34	0.26	0.73
Duty type: General duties	0.15	0.05	1.16
Rank: Senior Constable	0.10	0.29	1.10
Misconduct process variables			
Unsubstantiated complaints	1.14	0.18	3.15**
Local Area Commander Warning Notice	0.44	0.15	1.55
Region Commander Warning Notice	0.49	0.72	1.64
Performance or conduct management plan	0.36	0.08	1.44
Management action	0.23	0.07	1.26
Misconduct types			
Issues with an investigation	0.88	0.22	2.42
Improper use of force	0.64	0.11	2.00**
Drink driving	0.54	0.28	1.72*
Complaint related to court	0.59	0.47	1.79**
Breach code of conduct	0.59	0.72	1.80
Harassment, intimidation or bullying	0.49	0.08	1.64*
Allegation of corruption	0.22	0.20	1.25
Positive drug test	0.27	0.13	1.31
Disobey reasonable direction	0.37	0.33	1.44
Firearm related complaint	0.17	0.17	1.18
Untruthfulness/Dishonesty	0.16	0.29	1.17
Rudeness to members of the public	- 0.09	0.02	0.91
Breach privacy, unauthorised disclosure of data	- 0.45	0.21	0.64
Property or exhibits complaint	- 0.59	0.55	0.55

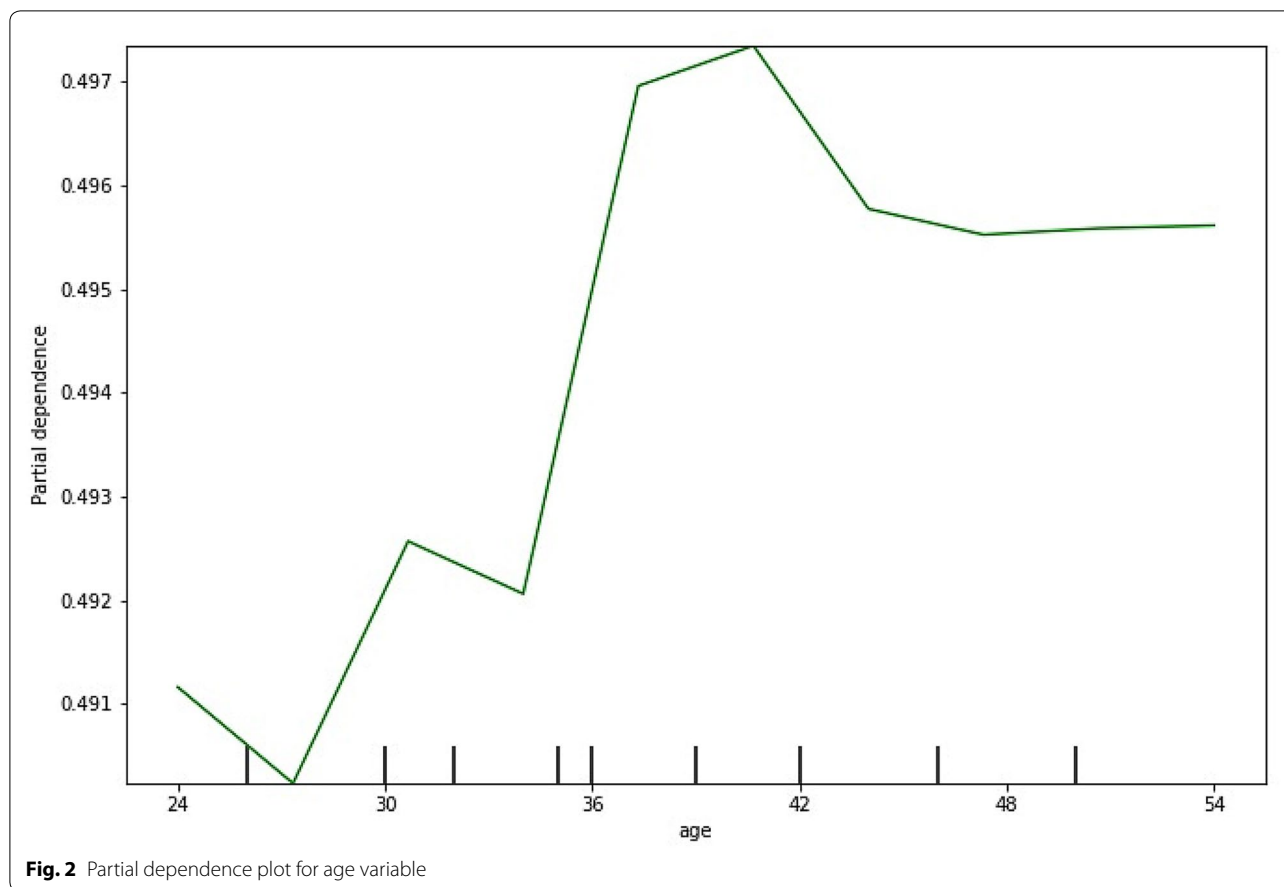
*P < 0.05

**P < 0.01

Table 5 Confusion matrix for serious misconduct among officers in test set

	Predicted positive	Predicted negative	Totals	Percent
True positive	205	17	222	61.6%
True negative	5	133	138	38.4%
Totals	210	150	360	
Percent	58.4%	41.6%		

long they had spent at their rank prior to an instance of serious misconduct. The PDP for this variable showed a relatively continuous effect, by which the longer an



officer spent at their rank, the higher the likelihood of serious misconduct (Fig. 4).

Finally, a PDP was produced for the unsubstantiated complaints variable, included as a count of the number of complaints that each officer had received, but were not substantiated. This PDP was particularly important as it demonstrated that, while substantiated complaints were important in predicting likelihood of serious misconduct, unsubstantiated complaints were important as well. Figure 5 suggests that a small number of unsubstantiated complaints does not notably increase the likelihood of serious misconduct, however, around 4 instances, this effect changes. As officers received complaints, even though these were investigated and found to be unsubstantiated, the likelihood of serious misconduct among these officers increased quickly, up to 12 unsubstantiated complaints, after which the effect declined.

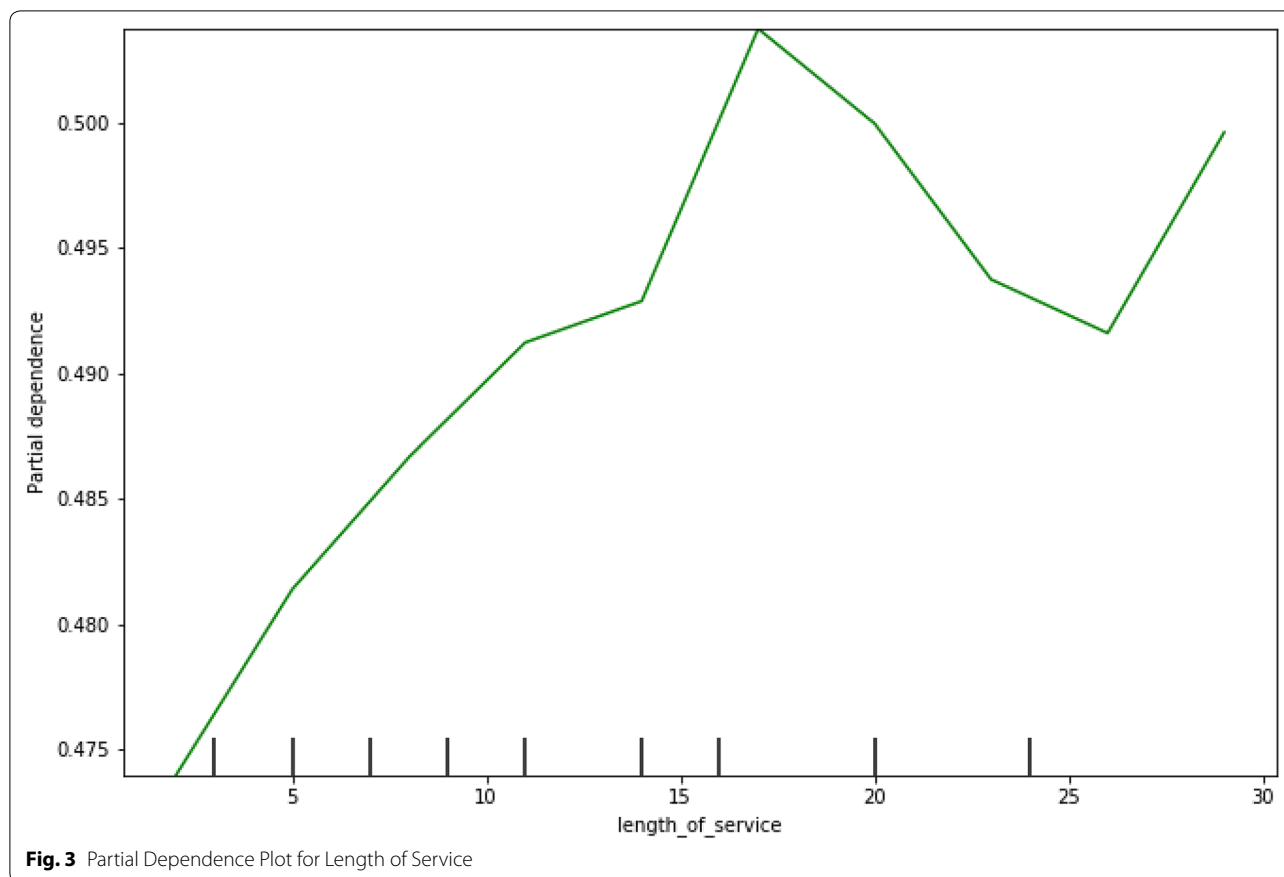
Discussion

An analysis of police misconduct by Greene et al. (2004) found that younger, male officers were more often misconduct prone, prompting suggestion that these officers may require an elevated rate of scrutiny. Although the

present study considered serious misconduct that may result in criminal charge or dismissal, risk was only elevated into the late 30 years of age. It was also noteworthy that comparison officers were typically marginally older at time of recruitment. Green et al. (2004) suggested that officers assigned to general patrol duties, referred to in this study as general duties, be placed at a higher rate of scrutiny. This finding is supported here, as general duties were found to be predictive of serious misconduct. While general duties officers were associated with serious misconduct, it was when these officers were mid-career and had been at their rank for several years, risk appears to increase. It is particularly noteworthy that general duties officers are shift-workers (NSW Police Force 2018), suggesting the elements of fatigue and harms associated with shift work (Rajaratnam et al. 2011; Senjo 2011; Pearsall 2012), including diminished decision making.

Risk factors for misconduct

On initial modelling, the random forest analysis determined that a prior instance of serious misconduct was particularly predictive of further serious misconduct, this was an important, but not unheralded finding. Kane

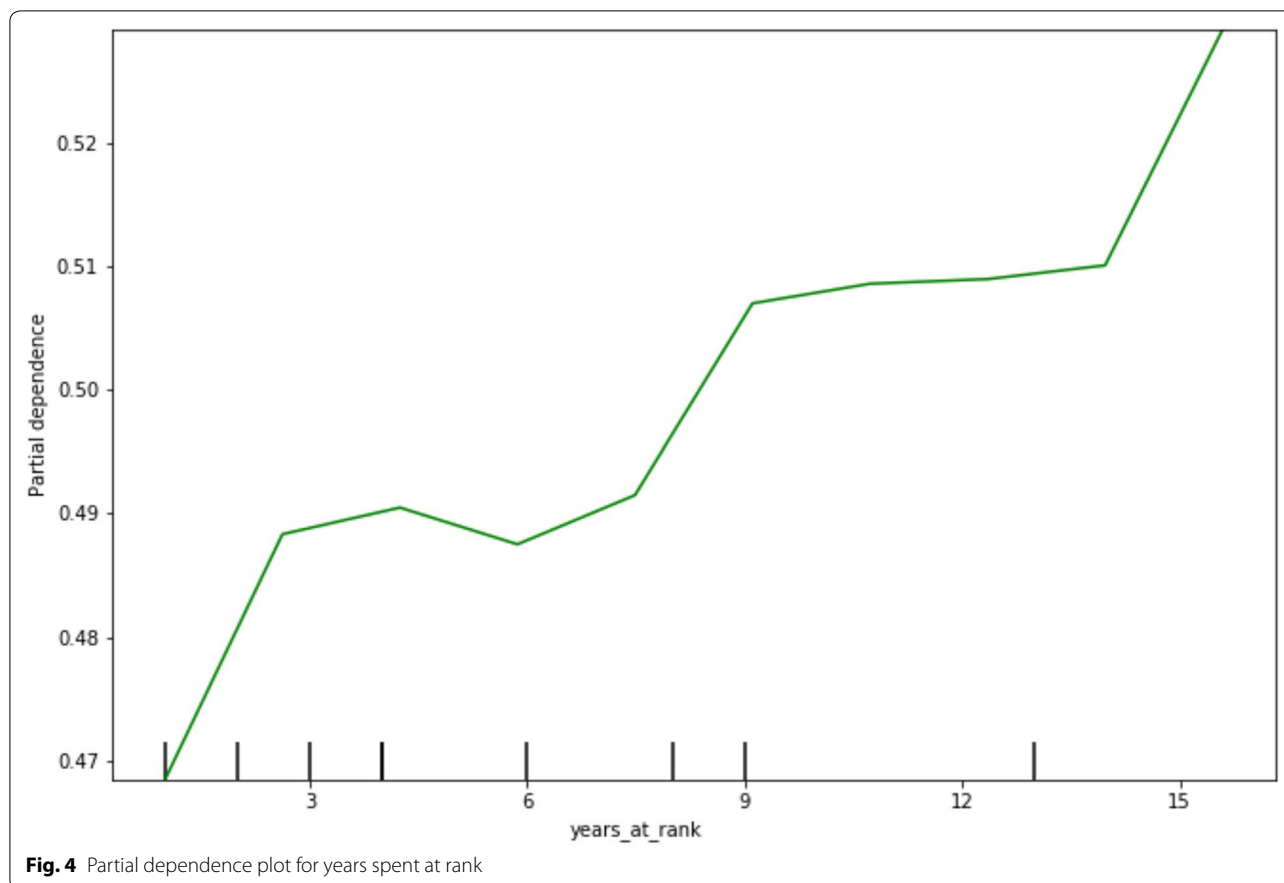


and White (2009) found that officers who had criminal records prior to joining the police, were substantially more likely to commit an instance of career ending misconduct. While serious misconduct in the present study did not necessarily meet the threshold of prior criminal charges against an officer, it did provide a measure of prior deviance. Subsequently, the notion that prior deviance was predictive of further deviance was supported here, furthermore it was the strongest predictor of serious misconduct. This suggested that either an officer who commits an instance of serious misconduct is sufficiently misconduct prone that remediation is no longer effective, or alternatively that the remediation processes for misconduct in this jurisdiction are not sufficient. However, in interpreting this finding, it is important to note that officers who have previously committed an instance of serious misconduct, may attract an elevated level of scrutiny compared to officers that had not.

Further, Kane and White (2009) found that a large proportion of officers had prior history of 'Administrative/failure to perform' incidents, in fact this was the most frequent factor attributed to officers who had committed career ending misconduct (Kane and White 2009: P.751). A similar finding was produced in the present study, in

which 'issues with an investigation' was the strongest predictor of serious misconduct among complaint variables. Additionally, being placed on a performance or conduct management plan, was associated with serious misconduct. The findings of this study largely support that of Kane and White (2009) in finding that there are antecedents among policing data that are predictive of career ending misconduct. When considering the propensity of an officer to commit misconduct, accessible evidence of antecedent deviance, such as substantiated complaints, is a valuable measure. However, not all possible data types are equal in this endeavour, as evidenced by the results of the random forest analysis in Table 3.

Secondary employment has been acknowledged within Australia as being a misconduct risk (People 2010). Prior research has taken differing approaches to financial debt among officers. Greene et al. (2004) suggested that officers with a mortgage, and with children were less associated with misconduct, while Cubitt and Judges (2018) found that financial debt, particularly driven by mortgage stress and supporting dependents, was contributory toward serious misconduct. Evidently secondary employment is not itself misconduct, rather it is suggested that this variable indicates extraneous factors increasing



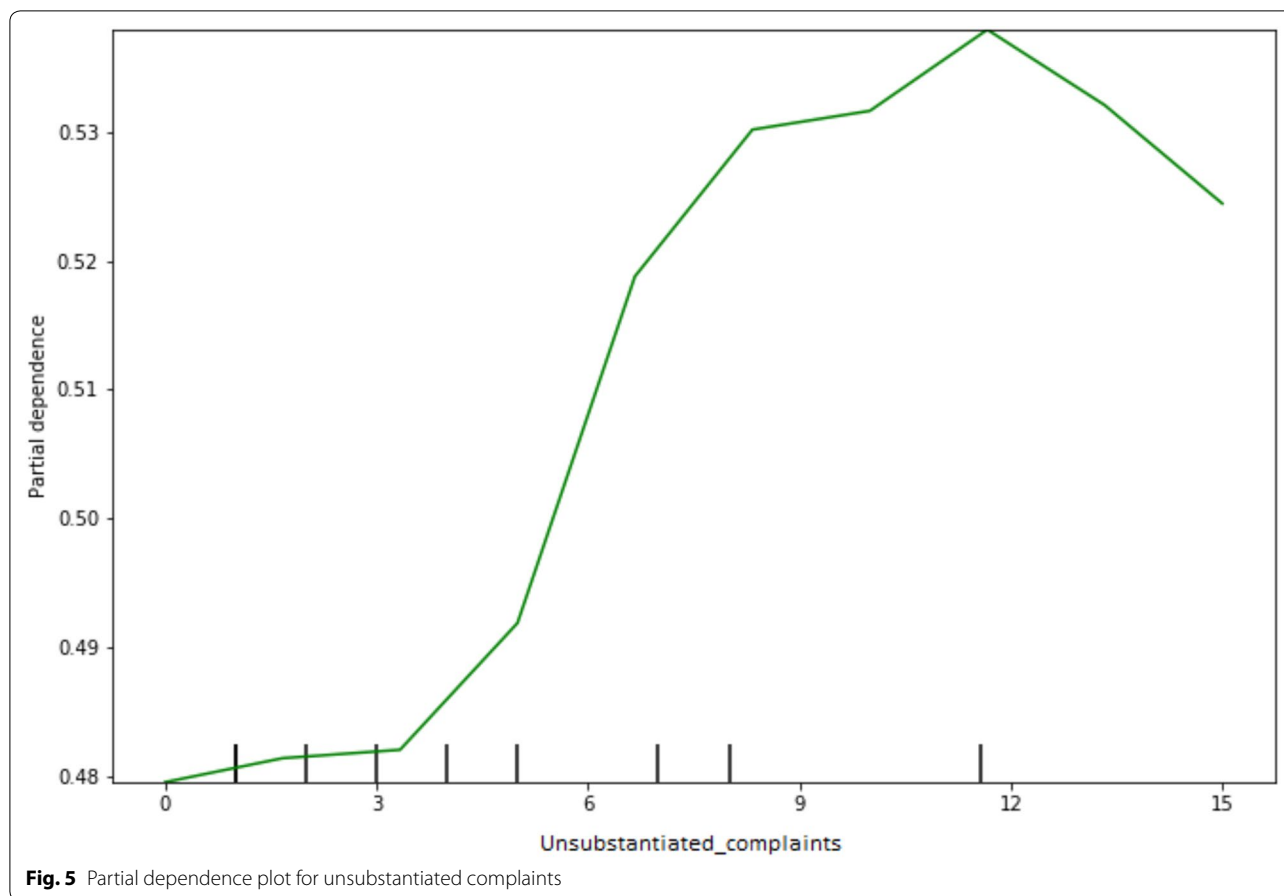
misconduct propensity, such as stress and burnout, or an inability to service loans (Lyle 2015; Cubitt and Judges 2018). As previously mentioned, fatigue is a substantial risk in a policing environment (Senjo 2011; Pearsall 2012), a factor to which secondary employment is likely to contribute. Given that secondary employment featured among 44.3% of the serious misconduct group, and comparatively only 9.5% of the comparison group, this result requires substantial consideration by policing agencies.

Given the prevalence of improper use of force in the literature, some consideration was warranted. While there is predictive utility in the improper use of force variable, the finding that it features behind a range of demographic variables in predictive power for serious misconduct is important. Improper use of force has traditionally been considered an important indicator of misconduct risk (Worden et al. 2014), frequently featuring in policing analyses (Borilla 2015; Gottschalk 2011). While there are legitimate instances of use of force in the policing environment, this is a highly visible misconduct type which has implications for community safety, police legitimacy and community consent (Borilla 2015). In an Australian context, there is noteworthy research into improper use of force (Hine et al. 2018; McCarthy et al. 2018; Baker

2009), however this research has not identified improper use of force as a predictor of serious misconduct, and has not been able to locate it among other features of misconduct. Subsequently, this finding is important as it identifies use of force as a noteworthy predictor. However, from an analytical standpoint it is neither the most important variable in this model, nor is the effect substantially greater than other complaint types.

Implications for policy

The model produced here was robust, as a sum of their parts, these variables demonstrated efficacy in predicting serious misconduct among police. Prior instances of serious misconduct in isolation were particularly good at predicting further serious misconduct. Practically, this finding suggests that where an officer commits an instance of serious misconduct, if they are to retain their employment, substantial effort should be placed into remediation. Alternatively, this may suggest that remediation procedures across the time period were insufficient. Further research is required to fully understand the attribution of this finding.



Literature typically frames serious misconduct as a problem among inexperienced and young officers, while older officers with a stable home life, mortgage and children, were suggested to be less associated with misconduct. Findings here suggest this may be a nuanced notion. Experienced officers were at no less risk for serious misconduct than younger, junior officers. Similarly, the presence of secondary employment as a strong predictor of serious misconduct suggested that financial issues, and fatigue may have a role to play in this paradigm. Practically, findings here suggest that asset distribution for misconduct prevention among law enforcement agencies, while typically targeted toward young and junior officers, should equally be targeted toward mid-career officers. Further, risk assessments performed when an officer obtains secondary employment must be validated, and continuously improved. It is evident that the role of financial issues and fatigue among officers is not fully understood. While further research into this relationship would be beneficial, the process of obtaining secondary employment by officers is an opportune misconduct prevention avenue for agencies.

Limitations

There are several important limitations to this study. Sampling, in particular identifying an effective comparison group for officers who commit serious misconduct, is a notoriously difficult task. Kane and White (2009) offered a robust sampling method, and this has been replicated here. This research assumes that all serious misconduct is considered to be equal, however it is clear that some misconduct is more severe than others. It was not within scope of this paper, and was not appropriate for authors to delineate between misconduct severity beyond the data provided by the NSW Police Force, however it is a limitation. There is some suggestion that random forest models may bias toward categorical variables with multiple levels, given the data utilised here, this is a limitation that must be acknowledged. Further, there are some ethical considerations to the notion of predicting human behaviour, particularly when considering deviance. While this paper identifies features that show noteworthy association with misconduct, it does not suggest that they should be operationalised for prediction and intervention among at-risk officers. Rather, it is suggested that these mechanisms be leveraged for evidence-based

misconduct prevention, support mechanisms and more accurate asset distribution among agencies for the reduction of misconduct. Finally, multicollinearity bears some consideration, particularly given the dataset used here. Among naturally occurring data, collinear variables are an important risk to address. While bootstrap aggregation, or bagging, in the random forest means prediction accuracy of the model is typically unaffected, individual feature importance may be impacted by collinear variables. Steps were taken early in this analysis to identify and address potential collinearity; however, it bears note and future research employing similar methods among naturally occurring data must account for collinearity prior to analysis.

Conclusion

The findings of this research support the use of data driven analytics in the analysis of police misconduct, however many of these results adhere to conventional wisdom. The finding that prior behaviour is predictive of future behaviour, particularly regarding deviance, was not novel. In fact, this finding strongly supported findings of prior research. Conversely, there were findings, particularly regarding age, experience level and the role of secondary employment which were novel for this field of research.

The policing environment is unique in the powers afforded to officers, frequent adverse interactions with the public, and the opportunities for misconduct. While prior research has suggested the importance of pre-employment screening and education, this is not the finding of this research. There is an ongoing question of why officers who are at relatively low risk early in their career go on to become misconduct risks. Whether this is a unique function of policing agencies, to engender criminogenic features among their workforce, or perhaps a function of the burnout and trauma exposure in this environment, is not well understood. Regardless, policy among policing agencies should reflect ongoing vigilance to the likelihood of serious misconduct among officers throughout their career, not only in the early years, and particularly in the presence of secondary employment. Policing agencies consistently change across time, with policy and procedure developments. Continued research into these environments is pivotal to identifying the changing nature of misconduct among law enforcement agencies.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40163-020-00133-6>.

Additional file 1: Variables included for analysis and definitions.

Acknowledgements

The authors thank the Professional Standards Command of the NSW Police Force for access to the data used in this research. We also want to thank Dr Michael Kennedy and Dr Erin Kruger for their insight regarding this research.

Authors' contributions

All authors approved of this manuscript.

Funding

No outside funding was used to support this work.

Availability of data and materials

Data used in this research was strictly confidential, and is therefore unavailable.

Competing interests

The authors declare that they have no competing interests.

Received: 17 June 2020 Accepted: 20 October 2020

Published online: 31 October 2020

References

- Alston, R. E. (2010). *The relationship between police occupational deviance and length of service in a large police agency* (Unpublished doctoral thesis). MN: Walden University.
- Australian Bureau of Statistics. (2013). Socio-Economic indexes for Areas (SEIFA). In Statistics (ed). Canberra: Australian Bureau of Statistics
- Baker, D. (2009). Police confirmation of use of force in Australia: "To be or not to be?" *Crime, Law and Social Change*, 52(2), 139–158.
- Barnes, G. C., & Hyatt, J. M. (2012). *Classifying adult probationers by forecasting future offending*. Washington DC: US Department of Justice.
- Bazley, T. D., Mieczkowski, T., & Lersch, K. M. (2009). Early intervention program criteria: evaluating officer use of force. *Justice Quarterly*, 26(1), 107–124.
- Berk, R. (2013). Algorithmic criminology. *Security Informatics*, 2(5), 1–14.
- Berk, R., Sherman, L., Barnes, G., Kurtz, E., and Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. *Statistics in Society*, 172(1), 191–211.
- Berk, R. A., Sorenson, S. B., and Barnes, G. (2016). Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. *Journal of Empirical Legal Studies*, 13(1), 94–115.
- Bertoia, T. (2008). *Developing an early intervention system for police misconduct in a law enforcement agency research and issue papers* (Vol. 1). Sydney, NSW: NSW Police Integrity Commission.
- Borilla, M. M. (2015). *The relationship between proactive and reactive policing in allegations of misconduct*. (Ph.D.), Alliant International University, Ann Arbor.
- Breiman, L. (2001). Statistical modeling: the two cultures. *Statistical Science*, 16(3), 199–231.
- Chan, J., and Bennett Moses, L. (2015). Is Big Data challenging criminology? *Theoretical Criminology*, 20(1), 21–39.
- Cohen, B., and Chaiken, J. M. (1973). *Police background characteristics and performance*. Lexington, Massachusetts: Lexington Books.
- Couronne, R., Probst, P. and Boulesteix, A.L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*, 19(270).
- Crehan, A. C., and Goodman-Delahunty, J. (2019). Procedural justice and complaints about police. *Salus Journal*, 7(1), 58–87.
- Cubitt, T. I. C., and Judges, S. J. H. (2018). Show cause analysis: A qualitative assessment of the factors influencing police misconduct from the perspective of the officers who commit it. *Policing: An International Journal*, 41(6), 782–797.
- Cuttler, M. J., and Muchinsky, P. M. (2006). Prediction of law enforcement training performance and dysfunctional job performance with general mental ability, personality, and life history variables. *Criminal Justice and Behavior*, 33(1), 3–25.
- Goodman-Delahunty, J., Beckley, A., and Martin, M. (2014). Resolving or escalating disputes? Experiences of the NSW Police Force complaints process. *Australasian Dispute Resolution Journal*, 25(2), 79–90.

- Gorta, A. (2009). Illegal drug use by police officers: using research and investigations to inform prevention strategies. *International Journal of Police Science and Management*, 11(1), 85–96.
- Gottschalk, P. (2011). Management challenges in law enforcement: the case of police misconduct and crime. *International Journal of Law and Management*, 53(3), 169–181.
- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behav Res*, 26(3), 499–510.
- Greene, J. R., Piquero, A.R., Hickman, M.J., Lawton, B.A. (2004). Police Integrity and Accountability in Philadelphia: Predicting and Assessing Police Misconduct (pp. 123): U.S. Department of Justice.
- Harris, C. J. (2008). Exploring the relationship between experience and problem behaviors: a longitudinal analysis of officers from a large cohort. *Police Quarterly*, 12(2), 192–213.
- Harris, C. J. (2014). The onset of police misconduct. *Policing: An International Journal of Police Strategies and Management*, 37(2), 285–304.
- Harris, C. J., and Worden, R. E. (2012). The Effect of sanctions on police misconduct. *Crime and Delinquency*, 60(8), 1258–1288.
- Helsby, J., Carton, S., Joseph, K., Mahmud, A., Park, Y., Navarrete, A., et al. (2018). Early intervention systems: predicting adverse interactions between police and the public. *Criminal Justice Policy Review*, 29(2), 190–209.
- Hickman, M. J. (2006). *Citizen Complaints About Police Use of Force*. Washington DC, USA: US Department of Justice.
- Hine, K. A., Porter, L. E., Westera, N. J., Alpert, G. P., and Allen, A. (2018). Exploring police use of force decision-making processes and impairments using a naturalistic decision-making approach. *Criminal Justice and Behavior*, 45(11), 1782–1801.
- Hong, H., Xiaoling, G., and Hua, Y. (2016, 26–28 Aug. 2016). Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. Paper presented at the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS).
- Hyndman, R. J., and Athanasopoulos, G. (2014). *Forecasting: Principles and practice*. Melbourne, Victoria: Monash University.
- James, S., James, L., and Dotson, L. (2020). Evaluating the effectiveness of a police department's early intervention system. *J Exp Criminol*. <https://doi.org/10.1007/s11292-019-09397-8>
- Kane, R. J., and White, M. D. (2009). Bad cops: a study of career-ending misconduct among New York City police officers. *Criminology and Public Policy*, 8(4), 737–769.
- Kirsch, N. (2014). *Misconduct risks associated with authorised secondary employment: a review of compliance with NSW police force policies and guidelines*. Sydney, NSW: Police Integrity Commission.
- Law Enforcement Conduct Commission Act (2016) (NSW). Retrieved from <https://www.legislation.nsw.gov.au/view/html/inforce/current/act-2016-061>.
- Le Lievre, K. (2017, 01/08/2017). Police officer charged with assault during Queanbeyan arrest. Sydney Morning Herald. Retrieved from <https://www.smh.com.au/national/nsw/queanbeyan-police-officer-pleads-not-guilty-to-assault-charges-20170801-qxn3mg.html>
- Lersch, K. M., Bazley, T., and Mieczkowski, T. (2006). Early intervention programs: an effective police accountability tool, or punishment of the productive? *Policing: An International Journal of Police Strategies and Management*, 29(1), 58–76.
- Lester, J. N., Muskett, T., O'Reilly, M. (2017). Naturally Occurring Data Versus Researcher-Generated Data. In M. O'Reilly, Lester, J., Muskett, T. (Ed.), *A Practical Guide to Social Interaction Research in Autism Spectrum Disorders*. (pp. 87–116). London, UK: Palgrave Macmillan.
- Lyle, P. L. (2015). Moonlighting police: Policies that regulate secondary employment – Possible stress and job burnout issues. (Ph.D.), Capella University, Ann Arbor.
- Mackenzie, B., White, L. (2018). NSW Police officer who hit 16yo boy 'as hard as he could' during Byron Bay arrest may face criminal charges. ABC News. Retrieved from <https://www.abc.net.au/news/2018-07-23/nsw-police-officer-could-face-legal-action-over-byron-bay-arrest/10024932>
- McCarthy, M., Porter, L., Townsley, M., and Alpert, G. (2018). Influence of community characteristics on serious police use of force events in an Australian policing jurisdiction: a test of minority threat, social disorganisation, and ecological contamination theories. *Policing and Society*, 1–18.
- Morgan, B. L., and Wilson Van Voorhis, C. R. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, 3(2), 43–50.
- Neuilly, M.-A., Zgoba, K. M., Tita, G. E., and Lee, S. S. (2011). Predicting recidivism in homicide offenders using classification tree analysis. *Homicide Studies*, 15(2), 154–176.
- NSW Police Force. (2012). NSW Police Force Complaint Handling Guidelines. Retrieved from https://www.police.nsw.gov.au/_data/assets/pdf_file/0011/226379/NSWPF_Complaint_Handling_Guidelines.pdf
- NSW Police Force. (2017). *Annual Report 2016–2017*. Sydney, NSW: New South Wales Police Force.
- NSW Police Force. (2018). General Duties. Retrieved 19/07/2018, 2018, from https://www.police.nsw.gov.au/recruitment/the_career/general_duties
- O'Mallon, F. (2017, 16/12/2017). Ex-NSW police sergeant Nigel Alfred Davey guilty of perjury, assault. Canberra Times. Retrieved from <https://www.canberratimes.com.au/national/act/exnsw-police-sergeant-nigel-alfred-davey-guilty-of-perjury-assault-20171213-h042ha.html>
- Owen, B. (2017, 9/11/2017). Police officer charged over alleged assault in Bon-nells Bay. Newcastle Herald. Retrieved from <https://www.theherald.com.au/story/5046334/police-officer-charged-with-assault/>
- Parish, R. (2018, 15/01/2018). Sydney cop to face alleged assault charges of woman, 22, in Mt Druitt after internal investigation. Blacktown Advocate. Retrieved from <https://www.dailytelegraph.com.au/newslocal/the-stand-ard/sydney-cop-to-face-alleged-assault-charges-of-woman-22-in-mt-druitt-after-internal-investigation/news-story/7f3796f1715c96efee33473c38faf1bf>
- Pearsall, B. (2012). Sleep Disorders, Work Shifts and Officer Wellness. *NIJ Journal*, 270, 36–39.
- People, J. (2008). *Unauthorised disclosure of confidential information by NSW police officers*. Sydney, NSW: Police Integrity Commission.
- People, J. (2010). *Characteristics of complaints about misconduct by Off Duty NSW Police Officers*. Sydney, NSW: Police Integrity Commission.
- People, J., Kirsch, N., Barnett, P. (2010). *Improper Associations in the NSW Police Force: A Review of Compliance with Policies and Guidelines* Sydney, NSW: Police Integrity Commission.
- Police Act 1990 (NSW). Retrieved from <https://www.legislation.nsw.gov.au/view/html/inforce/current/act-1990-047>.
- Porter, L. E., Prenzler, T., and Fleming, J. (2012). Complaint reduction in the Tasmania Police. *Policing and Society*, 22(4), 426–447.
- Quispe-Torreblanca, E.G., and Stewart, N. (2019). Causal peer effects in police misconduct. *Nature: Human Behaviour*, 3: 797–807.
- Rajaratnam, S. M. W., Barger, L. K., and Lockley, S. W. (2011). Sleep disorders, health, and safety in police officers. *Journal of the American Medical Association (JAMA)*, 306(23), 2567–2578.
- Reiss, A. J. (1988). *Private Employment of Public Police*. Washington, DC: National Institute of Justice.
- Senjo, S. R. (2011). Dangerous fatigue conditions: a study of police work and law enforcement administration. *Police Practice and Research*, 12(3), 235–252.
- Walker, S. E., and Archbold, C. A. (2005). *The New World of Police Accountability*. Thousand Oaks, California: SAGE Publications.
- Weitzer, R. (2002). Incidents of police misconduct and public opinion. *Journal of Criminal Justice*, 28(4), 313–346.
- Western Australia Police. (2008). *Submission to the Parliament of NSW: Inquiry into Early Intervention Systems to Identify Officers at Risk of Corruption*. Sydney, NSW: Parliament of NSW.
- Wood, G., Roithmayr, D. and Papachristos, A.V. (2019). The Network Structure of Police Misconduct. *Socius: Sociological Research for a Dynamic World*, 5: 1–18.
- Worden, R. E., Harris, C., and McLean, S. J. (2014). Risk assessment and risk management in policing. *Policing*, 37(2), 239–258.
- Zhao, Q., and Hastie, T. (2019). Causal Interpretations of Black-Box Models. *Journal of Business and Economic Statistics*, 1–19.
- Zulfacar, R., Nair, G., and Goodman-Delahunty, J. (2012). *The NSW police Force Complaints Process: Experiences of Client Advocates and Legal Practitioners* Police Accountability Project. Sydney, NSW: Community Legal Centres NSW.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.