

RESEARCH

Open Access



The value of criminal history and police intelligence in vetting and selection of police

Timothy I. C. Cubitt*

Abstract

Despite decades of research considering police misconduct, there is still little consensus on officer characteristics associated with misconduct, and best practice for detection and prevention. While current research focuses on the correlates of misconduct among serving police, a small but growing body of research seeks to understand factors that could improve the accuracy of vetting during the police recruitment process and thereafter. In the wake of recent high-profile misconduct by police in the UK, there is a renewed focus on whether current vetting processes are sufficient, and how they might be improved. The present research analyses data from the vetting of a sample of UK police officers to consider whether current processes accurately identified which officers were high risk for misconduct. Findings suggested that current vetting processes performed poorly at identifying officers at risk of misconduct. However, police intelligence and criminal history data from the time of recruitment performed very well at identifying which officers were high risk for serious misconduct later in their career. These findings hold important implications for how police are vetted, and subsequently selected in the UK. In particular, the importance of integrating police intelligence, and criminal records data into vetting processes.

Keywords Police misconduct, Vetting, Recruitment, Intelligence, Corruption

Introduction

Despite decades of research examining police misconduct, there is still little consensus on the officer characteristics associated with misconduct, and best practices for detection and prevention. The relatively homogeneous nature of police, and the lack of variation in characteristics and background, has made the correlates of police misconduct difficult to elicit (Grant & Grant, 1996). However, as policing agencies have made data more readily available, the characteristics of misconduct prone officers have become more clear (Cubitt, 2021; Cubitt et al., 2022; Gaub, 2020; Gaub & Holtfreter, 2021; Huff et al., 2018; Kane & White, 2009; White & Kane, 2013). Analysis of data relating to police misconduct is

central to the development of effective misconduct prevention processes within policing agencies. However, at present, the police misconduct literature focuses on information relating to serving police (eg. complaints of misconduct, service length, use of force etc.), with limited research into vetting and recruitment processes. The present research considers the process of vetting of police, and examines the types of information that may be useful in identifying applicants that are at higher risk of serious misconduct.

Literature review

Characterising misconduct

Research that has considered officers after the point of recruitment has identified several important characteristics of misconduct prone officers. Despite being overrepresented among police, male officers appear more likely to commit misconduct than women (Gaub, 2020; Greene et al., 2004; Kane & White, 2009; McElvain & Kposowa, 2008; White & Kane, 2013). Recent research has noted

*Correspondence:

Timothy I. C. Cubitt
timothy.cubitt@aic.gov.au
University of Technology Sydney, Ultimo, Australia



© The Author(s) 2023. **Open Access** 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.

that violent crime, financial, and alcohol-related crime are known to occur among police (Boateng et al., 2021). Regarding the private lives of officers, there are mixed findings on whether single or married officers are more at risk of misconduct; an important consideration given the conventional wisdom that marriage is a protective factor against misconduct (Kane & White, 2009; White & Kane, 2013) with recent research finding that divorce was protective among women (Gaub, 2020). Age of officers, more specifically being older (Greene et al., 2004; McElvain & Kposowa, 2008; Rojek & Decker, 2009; White & Kane, 2013) and obtaining some tertiary education (Kane & White, 2009; Kappeler et al., 1998) are noted as generally protective against misconduct. Although there is some evidence to suggest more serious misconduct was equally prevalent among older officers if they experienced career stagnation (Cubitt et al., 2020).

Findings relating to the tenure of an officer are particularly complicated, with some research suggesting that as tenure progresses, risk of misconduct increased (McElvain & Kposowa, 2008; Micucci & Gomme, 2005; Wolfe & Piquero, 2011), while elsewhere longer tenure was found to be either protective or feature little effect (Kane & White, 2009; Simpson & Kirk, 2022). In addition, a nonlinear effect of tenure has been suggested, with risk high during the probation year, and the period immediately following, before declining (Harris, 2014; Stinson et al., 2012). These findings are further complicated by whether misconduct is on-duty, using the powers of a police officer, or off-duty, with significant differences in the behaviour of officers (Boateng et al., 2021). The peer group of an officer may also have some sway on the likelihood of misconduct, with some suggestion of peer group effects (Quispe-Torreblanca & Stewart, 2019; Wood et al., 2019; Simpson & Kirk, 2022).

Vetting and pre-recruitment correlates of misconduct

A small but growing body of research addressing organisational characteristics and misconduct has touched on the initial vetting of police. Huff et al. (2018) found that “organizational characteristics including pre-hiring screening, accountability mechanisms and community relationships are associated with lower levels of agency misconduct” suggesting that, at least to some extent, agencies that employed a screening process were less prone to misconduct. Employment characteristics—both prior to joining the police, and while employed as a police officer—are often important predictors of misconduct. Prior to joining, those with any type of criminal history, record of discipline, or misconduct in prior careers were more likely to commit misconduct (Donner, 2019; Greene et al., 2004; Kane & White, 2009; Simpson & Kirk, 2022). Notably, where agencies employed a background

investigator to screen applicants, a negative recommendation from that investigator was a risk factor for misconduct, but only within the first 2 years, after which the effect diminished (Kane & White, 2009; White & Kane, 2013). General employment problems before becoming a police officer also suggested a higher likelihood of committing serious misconduct (Greene et al., 2004; Kane & White, 2009).

Vetting of police in the UK

In the wake of the murder of Sarah Everard by a serving member of the Metropolitan Police Service in March 2021, the Home Secretary of the UK commissioned His Majesty’s Inspectorate of Constabulary and Fire & Rescue Services (HMICFRS) to inspect the vetting and counter-corruption arrangement in policing across England and Wales (HMICFRS, 2022). As a component, HMICFRS considered the recruitment process, how recruit vetting is undertaken, and decisions in response to information received during vetting. The HMICFRS found that, of the 725 vetting files considered, 131 featured potentially flawed decisions at the recruitment stage. The report stated:

“We also found 131 cases where the decision [to employ staff] was questionable at best. In these, we found officers and staff with criminal records, or suspicions that they had committed crime (including some serious crime), substantial undischarged debt, or family members linked to organised crime. In other cases, officers and staff had given false or incomplete information to the vetting unit. We also found officers who, despite a history of attracting complaints or allegations of misconduct, successfully transferred between police forces. This is wholly unsatisfactory.

In all these cases, forces had overlooked or downplayed the matter and cleared the applicants, often without any rational explanation for doing so. There were occasions when sound vetting rejections had been overruled, with dubious justification. We have concluded that many aspects of police vetting need to be clarified and strengthened.” (HMICFRS, 2022: p.1).

While vetting procedure is a considerable, and ongoing issue within policing agencies, the weak evidence for vetting practice leaves agencies with an unclear path to remedy their processes. However, the HMICFRS report identified several important components of the vetting process which require greater focus, including whether criminal histories are considered with appropriate weight, and whether police intelligence sources could be used more effectively in vetting police. Building on this report, the present research undertakes an exploratory analysis to ask (1) using criminal history, and intelligence information available to police, is it possible to predict

which officers will commit an act of corruption? We then consider the effectiveness of the vetting process by examining whether (2) the issues identified during vetting hold a relationship with serious misconduct, and if they do, whether the number of flags raised is an appropriate metric for screening out applicants. Finally, using a sample of officers who had committed serious misconduct, this research considered (3) whether there were any typologies, or whether they were largely a homogeneous group.

Data for this research emerges from the vetting and misconduct processes a policing agency in the UK, for the purposes of this research the agency has been anonymised. The threshold for serious misconduct considered here was the same as other literature in this area, that an officer is found to have committed misconduct that was sufficient to be considered for dismissal or charged with a criminal offence (Kane & White, 2009; White & Kane, 2013; Cubitt et al., 2020; Cubitt et al., 2022). If an officer commits an act of gross misconduct, a panel is convened to consider the facts of the misconduct, and any mitigating submissions by the officer, and then decides on the sanction to be implemented. For all officers considered in the present research that committed an act of serious misconduct, the panel consideration and sanction process had been completed.

Data description

This study employed a de-identified secondary dataset featuring information on sworn police available at the vetting phase of the recruitment process and thereafter. Variables in these data included whether an officer had received a substantiated complaint of serious misconduct (binary variable), the age of an officer at the time of recruitment (continuous variable), whether an individual had transferred from a different police agency (binary variable), the number of issues or flags identified during the vetting process (continuous variable), and any conditions placed on the employment of that officer after recruitment (binary variable). At the point of vetting, questionnaires are issued to the applicant to obtain information that may inform a vetting flag, for example a flag may be raised if an applicant discloses an association with an individual that has a history of serious offending or are the subject of an active investigation. A condition placed on an officer's employment may include a non-association clause, for example if an applicant is associated with a known criminal, it may be that they are offered employment with police on the condition that they no longer associate with that individual.

Two further variables were included that described information that was available to police at the time of vetting, but was not considered during the vetting process for these officers. These included a variable identifying

the number of intelligence reports on policing databases that related to each respective officer (continuous variable). An intelligence report may be recorded when an incident does not meet the threshold for a criminal charge, but may be useful or important information for officers; for example, that an individual may be known to carry a weapon, or that they are a known associate of organised crime. The final variable included in these data was whether an individual had a footprint on the Police National Computer (PNC) (binary variable), an operational policing database that shares data across all law enforcement agencies in the UK (Francis & Crossland, 2002). This variable described whether there were any records on the PNC that related to each respective officer at the time of recruitment. An incident that results in a record on the PNC may be as minor as a road traffic offence, through to a more serious criminal offence. This was coded as a simple binary variable to describe whether there was, or was not a record on the PNC relating to this individual.

Data were made available for the 7 years between January 2014 and December 2021. Across this time, 131 officers from this agency received allegations of misconduct that were (1) substantiated, and (2) of the severity that they resulted in a misconduct hearing. Of these, data were available the 99 officers whose cases had been finalised. The remaining officers were in some way still involved in the misconduct process, via appeal, criminal charge, or incomplete hearing. To provide an appropriate comparison, the matching methodology used by Kane and White (2009) and Cubitt et al. (2020) to study police misconduct was replicated. Each officer that committed an instance of serious misconduct in these data were matched to a randomly selected officer from their academy class that had not. This resulted in a matched dataset of 198 de-identified officers, half of whom had committed a substantiated instance of corruption, in the 7 years from January 2014.

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; Cubitt et al., 2020). Data cleaning, and cross referencing was undertaken prior to provision of these data to ensure that each feature referenced against each officer was validated. For the purposes of this research, it is assumed that there is equal confidence in each vetting record, intelligence report, and PNC report. Ultimately, each of these data emerge from different sources, and there is no way for researchers to

evaluate whether there are different degrees of confidence in each.

Methodology

Given the amount of data available, and limited prior research in this area, this research is best considered to be a sequence of exploratory analyses. It was first important to compare the available data. Summary statistics were provided for each variable and a simple correlation analysis was then undertaken. This correlation analysis had two purposes, first to rule out collinearity for subsequent modelling, and then to consider the basic relationship between variables. Given the differing structure of variables employed here, a heterogeneous correlation matrix was produced (Babchishin & Helmus, 2016; Brown & Benedetti, 1977; McGrath & Meyer, 2006).

Logistic regression

While this was a smaller sample of officers, it constituted a greater proportion of those that had committed serious misconduct in the agency considered than had been available in prior literature (Cubitt et al., 2020). In recent years, computational analytical methods have become more common in examining deviance among police, these have included network analytics (Wood et al., 2019; Ouellet et al., 2019; Cubitt, 2021), and machine learning analytics (Helsby et al., 2018; Cubitt et al., 2020; Jain et al., 2022; Cubitt et al., 2022) to evaluate and predict misconduct. However, neither the data structure or the intention of the study were suitable for network analyses, while the volume of the data did not lend itself to machine learning. In classification tasks machine learning modelling often outperforms more traditional statistical techniques when using high dimensional data; where the number of covariates is large, relative to the sample size (Couronné et al., 2018), that was not the case in the present research.

The logistic regression is a commonly employed statistical method using police data (Gaub, 2020; Kane & White, 2009). Given the sample size available, and the research questions, this research estimated a logistic regression using the binary outcome variable of substantiated serious misconduct among officers, and the previously described explanatory variables. The intention of this analysis was to consider their utility in predicting whether an officer posed an unacceptable risk of serious misconduct.

A Receiver Operating Characteristic (ROC) curve was used to identify the accuracy of the logistic regression, through the Area Under the Receiver Operating Characteristic (AUROC) curve. The ROC curve identifies the true positive rate of classification (y-axis), compared with the false positive rate (x-axis) at any threshold value. The AUROC, which we refer to in simple terms as the

accuracy of the model, represents the probability that a randomly selected case will be accurately classified.

At this point several validation exercises were undertaken. Both supervised, and unsupervised logistic regressions were computed using the same data structure to identify the approach with strongest modelling performance, but also to account for any potential overfitting. Each model performed equally well in this task, we therefore report the unsupervised approach, to align with prior research in this field. To validate the assumption that the logistic regression was the preferred approach due to the dimensionality of the data, a random forest was then computed using the same data structure, with a 70% training and 30% testing set (Cubitt et al., 2020). Model accuracy was compared with the logistic regression by repeating the ROC curve procedure and comparing the AUROC. A bootstrap test for statistical significance between ROC curves was then implemented to identify whether the difference between accuracy of the logistic regression and the random forest were statistically significant. These analyses were undertaken using the statistical analysis software, R version 4.2.0, and the 'randomForest', 'dplyr', 'pROC', 'PerformanceAnalytics', and 'ggplot2' packages.

Developing typologies – K-medoids

Several methodologies were considered for the development of typologies. Given the number and structure of groupings was unknown prior to analysis, neither a discriminant function analysis, or supervised clustering methodology were useful. A latent class analysis was considered, however, prior research has suggested that machine learning clustering algorithms may outperform latent class analyses with similar data (Brusco, Shireman & Steinley, 2017). As a result, the k-medoids clustering algorithm was preferred for analysis. K-medoids is particularly useful in uncovering the existence of latent groups in complex data (Brennan & Oliver, 2013). To satisfy the final research question, and uncover any sub-groups among misconduct prone police, this analysis only considered officers who had committed serious misconduct, with the remaining officers set aside. The k-medoids analysis was then implemented as an unsupervised analysis. This analytical process required two elements prior to analysis, a test of whether these data were indeed clusterable, and knowledge of the optimal number of clusters. To do this, the Hopkins statistic was first computed. A Hopkins statistic of less than 0.5 indicates that the data are more clusterable than not, however for these data to be meaningfully clusterable a Hopkins statistic of less than 0.25, indicating a clustering tendency at the 90% confidence interval, was preferred.

To identify the optimal number of clusters (k) in these data, prior to implementation of the k -medoids algorithm, the elbow method was used. Put simply, the elbow method groups data into a specified number of clusters, from $k=1$ to $k=10$, and then computes and presents the average quality of each cluster, allowing decision on the most accurate model.

Once the number of clusters is specified, k -medoids randomly selects and plots a centroid from data points provided, it then iteratively develops clusters (Amer, 2020). The algorithm then recomputes the centroids to best fit the clusters and repeats the process, thereby refining the accuracy of the model (Amer, 2020). This process is repeated until all data points remain in the same cluster (Amer, 2020; Jain & Dubes, 1988).

To provide a visual representation of the clusters, a Principal Component (PC) Analysis (PCA) was computed separately. PCA was used to condense several variables into two vectors, that best described the extent to which individuals are similar or different. In the PCA cluster plot provided here, the first and second PCs are selected, and plotted on the x and y axes, titled Dim1 and Dim2 respectively. In brackets, on each axis, the proportion of variance accounted for by each PC is provided.

To assess the performance of k -medoids, the Silhouette coefficient was used to evaluate the quality of clusters (Batool & Hennig, 2021). A silhouette coefficient is between -1 and 1 (Lleti et al., 2004), a negative score indicates low confidence in the clustering, while a positive score indicates that we can be confident in the accuracy of data attributed to that cluster. This analysis was performed using statistical analysis software R, version 4.2.0, and the 'dplyr', 'cluster', and 'factoextra' packages.

Results

Summary statistics and correlation

The mean age of recruitment marginally differed between groups, with officers who committed serious misconduct marginally younger at time of recruitment than those who did not. Table 1 suggests that the serious misconduct group exclusively did not transfer in from other policing agencies, while there was a noteworthy rate of prior policing experience among the comparison group. Similarly, where conditions were placed on the employment of officers, these were exclusively in the comparison group, with none occurring in the serious misconduct group. However, the serious misconduct group featured a higher rate of issues or flags identified during vetting, intelligence reports, and a higher proportion were identified on the PNC at time of recruitment. To test whether the serious misconduct group and the comparison group were significantly different, Wilcoxon rank sum test, and Chi2 test of independence were implemented respectively depending on variable structure. Statistically significant differences were found between groups in the age at recruitment ($p < 0.05$), number of police intelligence reports ($p < 0.01$), and presence on the PNC database ($p < 0.01$) variables.

The correlation analysis suggested that the relationship between variables was limited. The most notable correlation (Table 2) coefficient was between the number of issues or flags identified during vetting, and a presence on the PNC, and ($r = 0.46$, $p < 0.01$). Of the 198 officers in these data, 116 were flagged as a potential risk during their vetting, and of those that were flagged, 58 were found to commit serious misconduct. While 8.08 percent of the total sample featured a presence on the PNC, of the 116 that were flagged during vetting, 13.79 percent featured a presence on the PNC. Although these variables were not correlated to the extent that their relationship may influence modelling, it is important to note that vetting processes that may raise flags appear entirely independent of PNC and intelligence checks. To put it

Table 1 Summary statistics for matched sample

Variables	Serious misconduct group (n = 99)	Comparison group (n = 99)	Full sample (n = 198)
Serious misconduct (response variable)	100%	0%	50%
Age at recruitment (yrs.)	25.07	27.01	26.04
Sex			
Male	84.8	85.9	85.9
Female	15.2	14.1	14.1
Transferred from another police agency (percent)	0%	18.18%	9.09%
Number of issues or flags identified during vetting (per officer)	1.19	0.86	1.03 (0–4)
Conditions placed on employment at time of recruitment (%)	0%	1.01%	0.51%
Number of police intelligence reports (per officer)	0.56 (0–6)	0.09 (0–3)	0.33 (0–6)
Presence on the Police National Computer database (%)	12.12%	4.04%	8.08%

Table 2 Correlation matrix for data on serious misconduct by police officers

Corrupt conduct	- 0.15*	- 0.32**	0.15*	- 0.07	0.27**	0.15*
Age at recruitment	0.13	- 0.11	- 0.06	0.00	0.00	- 0.03
Transferred from another police agency		- 0.22**	0.23**	- 0.10	- 0.10	- 0.03
Number of issues or flags identified during vetting			0.00	0.19**	0.19**	0.46**
Conditions placed on employment at time of recruitment				- 0.03	- 0.03	- 0.02
Number of police intelligence reports					0.12	0.12
Presence on the Police National Computer database						0.12

* <0.05, ** <0.01

differently, of the 82 officers who did not have flags or issues raised during their vetting process, 20.73 percent featured police intelligence reports, and exactly half went on to commit serious misconduct.

Logistic regression

The logistic regression was estimated alongside the random forest for comparison, each model attempted to predict serious misconduct among these officers. The prior assumption, that the dimensionality of the data may result in the logistic regression outperforming the random forest, held true. The logistic regression featured an AUROC of 0.8086, outperforming the random forest, AUROC=0.7273. However, the difference between the AUROC of these models did not achieve statistical significance (p=0.158). An AUROC of greater than 0.7 is considered to represent a noteworthy prediction rate (Grogger et al., 2021), while each model met this criterion, the logistic regression appeared to be a particularly strong prediction model, suggesting that the variables included in these data were quite good at discriminating which officers would and would not commit serious misconduct. The ROC curve for the logistic regression is provided as Fig. 1, given that it outperformed the random forest, this research proceeds in describing the findings of the logistic regression.

The findings of the logistic regression (presented in Table 3) suggested that the variables detailing prior experience as a police officer, and any conditions placed on the employment of a recruit were not particularly useful. These occurred so rarely among these data that they were

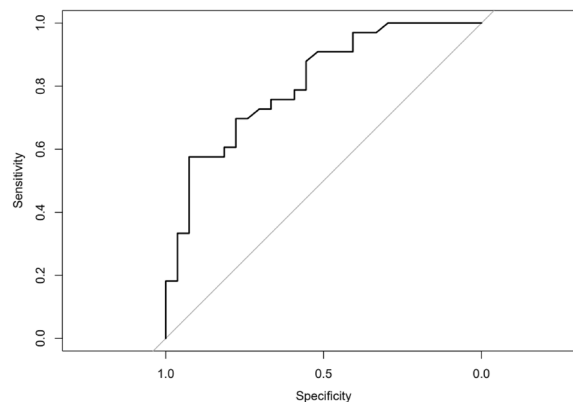


Fig. 1 Receiver Operating Characteristics Curve for logistic regression predicting serious misconduct by officers (AUROC = 0.8086)

analytically unreliable. Three variables returned a statistically significant result, those were the age of an individual at time of recruitment, the number of police intelligence reports, and whether an applicant had a presence on the PNC database. The effect of the age variable was marginal, however it suggested that as the age of an applicant increased at time of recruitment, the odds that they would commit serious misconduct marginally decreased.

The findings in relation to police intelligence holdings, and whether a recruit had a presence on the PNC were particularly important. These results suggest that if there were intelligence reports relating to an applicant, the odds that officer would commit serious misconduct were significantly greater than among applicants who were not associated with any intelligence reports, this effect increased as the number of intelligence reports relating to the applicant increased. Finally, recruits who had a presence on the PNC

Table 3 Logistic regression estimates and odds ratios

Variable	Log-odds	SE	OR
Age at recruitment	- 0.06	0.028	0.946*
Transferred from another police agency			
No (ref.)			
Yes	- 17.59	914.742	0.00
Number of issues or flags identified during vetting	- 0.12	0.165	0.888
Conditions placed on employment at time of recruitment			
No (ref.)			
Yes	- 0.04	4060.556	1.041
Number of police intelligence reports	1.33	0.407	3.773**
Presence on the Police National Computer database	1.33	0.771	3.790*

* <0.05, ** <0.01

database prior to joining were at notably higher odds of serious misconduct. Findings for both of these variables were statistically significant, suggesting that they may be particularly useful in the vetting process for recruitment and thereafter.

Typologies

The Hopkins statistic was computed to test the tendency of these data to cluster. Here, the Hopkins statistic returned 0.147, satisfying the threshold value of 0.25

meaning there was high confidence that these data were clusterable.

Latent cluster quality and content

The elbow method was implemented (Fig. 2) and suggested that clustering these data into two groups was appropriate, with a mean silhouette width of 0.51.

Silhouette coefficients for each of the clusters were computed and provided as Fig. 3. The two clusters were similar in size, with the first comprising of 52 officers, and the second comprising of 47. There was marginally greater confidence in Cluster 1, returning a Silhouette

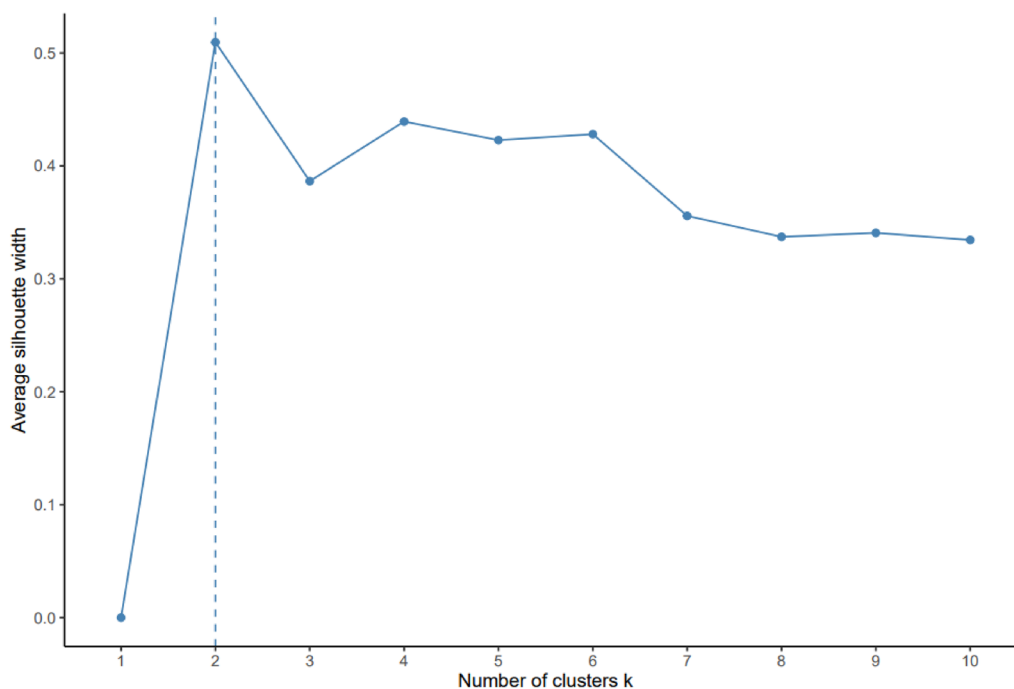


Fig. 2 The elbow method to identify the optimal number of clusters

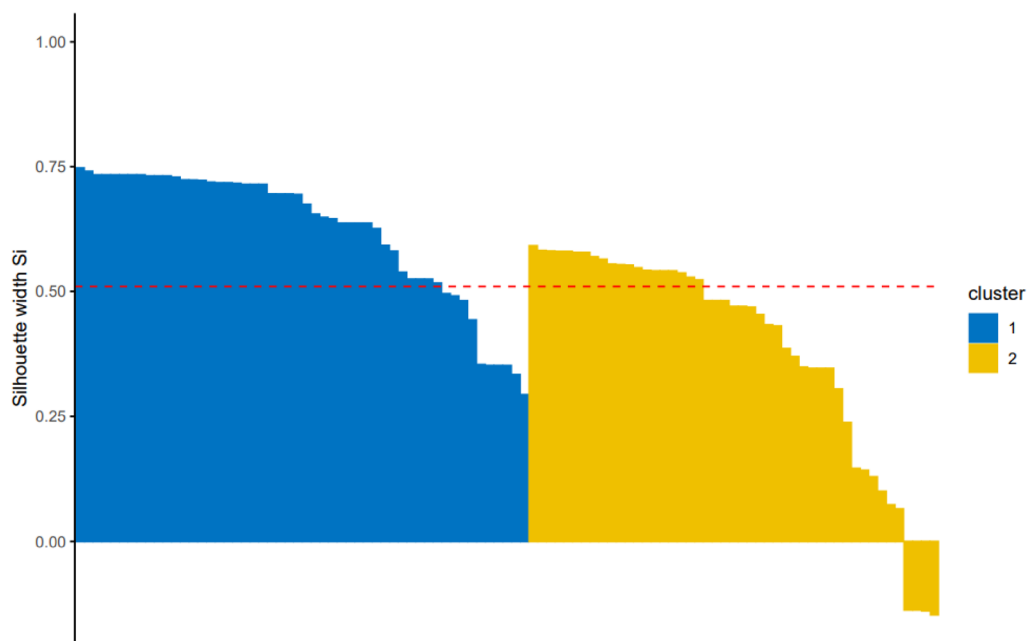


Fig. 3 Silhouettes for typologies of officers with substantiated allegations of corrupt conduct (mean silhouette width = 0.51)

coefficient of 0.62, than Cluster 2 which returned a Silhouette coefficient of 0.39.

The characteristics the developed typologies

While the Hopkins statistic and silhouette coefficients suggested that there could be confidence that these data were clusterable, Fig. 4 identified that the differences between groups were likely marginal. It is notable that the first cluster featured younger officers at the time of recruitment (see Table 4) who had a marginally greater mean number of flags identified during the vetting process. While the second cluster who were older, with a mean age of 30.04 years, featured a marginally higher mean number of intelligence reports on police databases, and were marginally more likely to have a record on the PNC database. Although these were discernable groups, and suggest that there were some differences at the time of recruitment, the differences were ultimately marginal, and adhered to the assertion that officers who committed serious misconduct were largely homogeneous.

Discussion

The evidence base for best practice in vetting police is limited, however selection of the right officers is central facet of agency legitimacy and corruption prevention. The findings of this research held notable similarities to other work considering misconduct among police. For example, it is relatively established that officers with a criminal history, record of discipline in alternate workplaces, or

prior misconduct at different agencies were more associated with misconduct (Donner, 2019; Greene et al., 2004; Kane & White, 2009; Simpson & Kirk, 2022). These were similar findings to those here, with records on police databases associated with an increased likelihood of serious misconduct. Further, Kane and White (2009) and White and Kane (2013) found that negative recommendations by background investigators prior to employment was associated with later career ending misconduct. This is similar to the police intelligence data included in the present study. Although it does not represent formal criminal charges, it is an indicator of adverse behaviour, that the present study has found to be a strong predictor of serious misconduct.

This research principally set out to consider whether criminal history and intelligence information available to police offered an opportunity to improve vetting, and whether the flags currently developed during vetting appropriately identified high risk officers. Findings suggested that flags currently used were negatively associated with serious misconduct among officers. In other words, officers who did not have issues or flags identified during vetting were more associated with serious misconduct, than those that did, meaning that these flags held little practical utility in assessing risk. These flags are only used during the vetting process, they are not used to implement ongoing monitoring, and therefore it is unlikely that the presence of flags would influence officers to be less likely to commit misconduct after they are

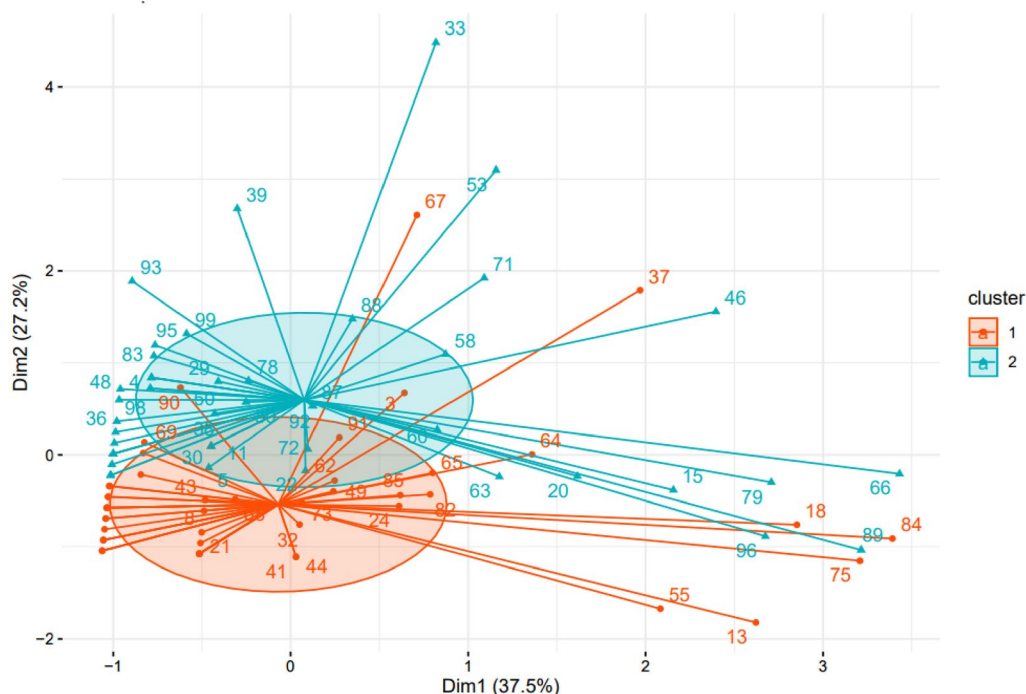


Fig. 4 Clusters of officers with substantiated allegations of serious misconduct

employed. These flags therefore appear to be poor metric for decision making during recruitment.

In contrast to current vetting processes, police intelligence databases and PNC records may be essential sources of information when vetting applicants. Those that featured in police intelligence records at the time of their application were significantly more likely to commit serious misconduct than those who did not. Similarly, officers who had any PNC record at the time of recruitment were also significantly more likely to commit serious misconduct if they were recruited into the police. These are important findings for the vetting and

recruitment of police, while current methods for flagging officers did not appear to be effective, there is a clear avenue to improving vetting and recruit selection using police intelligence, and PNC records.

Finally, the typologies of officers developed through application of k-medoids identified two clusters among officers who committed serious misconduct. The principal difference between these two groups was age, with some officers being marginally younger, and some marginally older at the time they were recruited. The lack of variation in the background, and characteristics of officers has, in the past, made characterising misconduct

Table 4 Summary statistics for developed typologies of officers who committed corrupt conduct

Variables	Cluster 1 (n = 52)	Cluster 2 (n = 47)	Overall serious misconduct group (n = 99)
Age at recruitment (yrs.)	20.58	30.04	25.07
Gender			
Male	82.69%	87.23%	85.9%
Female	17.31%	12.77%	14.1%
Transferred from another police agency (percent)	0%	0%	0%
Number of issues or flags identified during vetting (per officer)	1.23	1.15	1.19
Conditions placed on employment at time of recruitment (%)	0%	0%	0%
Number of police intelligence reports (per officer)	0.52 (0–6)	0.62 (0–6)	0.56 (0–6)
Presence on the Police National Computer database (%)	9.62%	14.89%	12.12%

prone officers difficult (Grant & Grant, 1996). The findings of this analysis supported the assertion that there is often little material difference between these officers at the time of recruitment. Here, they were found to be a relatively homogeneous group.

When considering the vetting of recruits, HMICFRS (2022) provided a relatively simple explanation for why the current system of flagging officers may be flawed, that officers were required to proactively disclose their circumstances and, when doing so, they may provide “false or incomplete information” (HMICFRS, 2022: P. 1). The present research found that officers who were flagged during vetting were less likely to commit serious misconduct. Using the logic of the HMICFRS report, it is possible that this finding is an artefact of officers being flagged during vetting simply because they were more honest than those that were not. In other words, vetting flags may, at least in part, be inadvertently measuring which officers had a higher propensity for honesty, and therefore current vetting flags may in fact be associated with more desirable recruits than undesirable. However, even if this is the case, these flags did not appear to be robust metrics for vetting decisions. Despite the poor performance of current metrics, the findings of this research support the assertion by the HMICFRS, that there are avenues to clarify and strengthen vetting at the recruitment stage. The principal avenue to improve this process is making police intelligence databases, and PNC records a core component of vetting, with considerably greater degree of hesitance in recruiting individuals with records in either of these domains.

While current definitions of predictive policing focus on the prediction of crime, often employing hot-spots analytics, intended to impact and reduce crime volume (Mohler et al, 2015), it is a similar task that is undertaken here, instead focusing on the police themselves. There are therefore similar drawbacks, and considerations to the implementation of these approaches. Meijer and Wessels (2019) noted that the implementation of these types of analytics often lead to mixed results, ultimately suggesting that analytical models are often less successful when practically implemented. Further, it is important to acknowledge concerns regarding transparency should such modelling be implemented, and legitimacy among officers either implementing or subject to the outcomes of predictions (Meijer & Wessels, 2019). This raises the importance of false predictions, both positive and negative. For example, while applicants with a PNC footprint appear at higher risk for serious misconduct, that does not mean that all of these applicants will commit an act of serious misconduct. However, the opposite is also true, the absence of a record does not ensure a sound applicant. While these analyses may serve to inform and

refine vetting processes toward greater accuracy, there is nuance to the implementation, and the agency tolerance to trade offs between false positives and negatives. Ultimately, these findings may help aid in decision making, but sole reliance on these variables is likely inadvisable.

Limitations

There were several limitations to this research. As previously noted, reporting was an important limitation, however it is particularly important to note here as these data represent a mix of reported police misconduct and potential criminal offences. The ability to assess unreported misconduct is limited, however it is likely that the rates of misconduct are not representative of actual rate of misconduct committed by police, here this means that we likely underestimate the actual extent of serious misconduct as well. In addition to underreporting misconduct, the underreporting of criminal activity is a known limitation in criminological study. Two important variables here related to intelligence, and criminal activity data held by police, each of which are likely subject to underreporting. Further, there were several important variables that were unavailable in the present analysis, including officer education level, prior employment type, and duty types undertaken across time. This research was also not able to consider the type of serious misconduct, for example whether it was a serious use of force or sexual misconduct, only that it was serious misconduct, and resulted in a hearing. Future research may benefit from inclusion of these further variables relating to officers, and the specific complaint type that was considered to be serious misconduct.

Early steps were taken in this analysis to identify potential collinearity using correlation analyses, a validation exercise was also undertaken to consider the possible influence of overfitting in the logistic regression. However, we cannot rule out the influence of confounding variables. As a result, we do not interpret these findings with a view toward causality, but toward the relationship they likely have as indicators of a propensity for serious conduct.

There are several limitations to the use of k-medoids in this research. The metrics employed to analyze k-medoids can be influenced by outliers (Raykov et al, 2016). To address this, the number of clusters were optimized using the elbow method to minimize the influence of within cluster outliers. Finally, k-medoids assumes that the number of clusters in the data is known prior to analysis, this metric is used to evaluate accuracy (Raykov et al, 2016). Given this research sought to identify latent sub groups through clustering, this was not the case here. This limitation was mitigated somewhat by identifying

the optimal number of clusters and then validating the cluster ensemble with the silhouette coefficients.

Implications & conclusion

The vetting of police is an exceptionally important avenue for screening out unsuitable applicants before they go on to obtain the powers and privilege afforded to sworn police officers. The findings of this research hold important implications for the way that vetting is undertaken, and the relative importance of the information obtained during this process. The current process of developing flags during vetting to identify officers who may be a risk did not appear to be effective. In fact, those that were flagged during vetting were less likely to go on to commit serious misconduct than those that were not. This finding supported the report by HMICFRS, which suggested that the current vetting processes may not be accurate or sufficient for the task undertaken. However, these findings also suggested that the use of police intelligence and PNC records may be promising. In totality, the present study finds that greater consideration should be given to whether an individual has a record on either of these databases, than the information that is proactively disclosed during vetting. If an applicant features on either of these sources, policing agencies should carefully consider whether that applicant is suitable for the powers afforded to a police officer.

Author contributions

The author read and approved the final manuscript.

Funding

TC conducted the entirety of the research. The author received no outside funding to support this work.

Availability of data and materials

The data used in this research is the property of the policing agency referred to in this work. It is therefore unavailable.

Declarations

Competing interests

The author declares that they have no competing interests.

Received: 30 December 2022 Accepted: 13 March 2023

Published online: 29 March 2023

References

- Amer, A. A. (2020). On K-means clustering-based approach for DDBSs design. *Journal of Big Data*, 7(31), 1–31.
- Anitha, P., & Patil, M. M. (2019). RFM model for customer purchase behavior using K-Means algorithm. *J King Saud University – Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2019.12.011>
- Babchishin, K. M., & Helmus, L. M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world data. *Behavior Research Methods*, 48, 1021–1031.
- Banerjee, A. & Dave, R.N. (2004). Validating clusters using the Hopkins statistic. Proceedings of the 2004 IEEE International Conference on Fuzzy Systems, December 2004. Budapest, Hungary. <https://doi.org/10.1109/FUZZY.2004.1375706>
- Batool, F., & Hennig, C. (2021). Clustering with the average Silhouette width. *Computational Statistics & Data Analysis*. <https://doi.org/10.1016/j.csda.2021.107190>
- Boateng, F. D., Pryce, D. K., & Hsieh, M. L. (2021). The criminal police officer: Understanding factors that predict police crime in the United States. *Crime and Delinquency*. <https://doi.org/10.1177/00111287211054732>
- Bora, D. J., & Gupta, A. K. (2014). Effect of different distance measures on the performance of K-means algorithm: An experimental study in Matlab. *International Journal of Computer Science and Information Technologies*, 5(2), 2501–2506.
- Brennan, T., & Oliver, W. L. (2013). The emergence of machine learning techniques in criminology. *Criminology & Public Policy*, 12(3), 551–562.
- Brown, M. B., & Benedetti, J. K. (1977). On the mean and variance of the tetrachoric correlation coefficient. *Psychometrika*, 42, 347–355. <https://doi.org/10.1007/BF02293655>
- Brusco, M. J., Shireman, E., & Steinley, D. (2017). A comparison of latent class, K-means, and K-median methods for clustering dichotomous data. *Psychological Methods*, 22(3), 563–580.
- Chappell, A. T., & Piquero, A. R. (2004). Applying social learning theory to police misconduct. *Deviant Behavior*, 25(2), 89–108. <https://doi.org/10.1080/01639620490251642>
- Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics*, 19(1), 1–14.
- Cubitt, T. I. C., Wooden, K. R., & Roberts, K. A. (2020). A machine learning analysis of serious misconduct among Australian police. *Crime Science*, 9(1), 1–14. <https://doi.org/10.1186/s40163-020-00133-6>
- Cubitt, T. I. C. (2021). Using network analytics to improve targeted disruption of police misconduct. *Police Quarterly*. <https://doi.org/10.1177/10986111211057212>
- Cubitt, T. I. C., Gaub, J. E., & Holtfreter, K. (2022). Gender differences in serious police misconduct: A machine-learning analysis of the New York Police Department (NYPD). *Journal of Criminal Justice*. <https://doi.org/10.1016/j.jcrimjus.2022.101976>
- Donner, C. M. (2019). “The best predictor of future behavior is...”: Examining the impact of past police misconduct on the likelihood of future misconduct. *Journal of Crime and Justice*, 42(3), 300–315. <https://doi.org/10.1080/0735648X.2018.1537882>
- Francis, B. & Crossland, P. (2002). The Police National Computer and the Offenders Index: Can they be combined for research purposes. London: Home Office. <https://eprints.lancs.ac.uk/id/eprint/50042/1/pncandoir170.pdf>. Accessed 28 December 2022.
- Gaub, J. E. (2020). Understanding police misconduct correlates: Does gender matter in predicting career-ending misconduct? *Women and Criminal Justice*, 30(4), 264–289. <https://doi.org/10.1080/08974454.2019.1605561>
- Gaub, J. E., & Holtfreter, K. (2021). Keeping the women out: A gendered organizational approach to understanding early career-ending police misconduct. *Crime and Delinquency*. <https://doi.org/10.1177/0011128721999332>
- Grant, J. D., & Grant, J. (1996). Officer selection and the prevention of abuse of force. In W. Geller, & H. Toch (Eds.), *Police violence: Understanding and controlling police abuse of force* (pp. 151–162). New Haven, CT: Yale University Press.
- Greene JR, Piquero AR, Hickman MJ, Lawton BA. (2004). Police integrity and accountability in Philadelphia: Predicting and assessing police misconduct. Washington, D.C. <https://www.ojp.gov/pdffiles1/nij/grants/207823.pdf>. Accessed 28 December 2022.
- Grogger, J., Gupta, S., Ivandic, R., & Kirchmaier, T. (2021). Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases. *Journal of Empirical Legal Studies*, 18(1), 90–130. <https://doi.org/10.1111/jels.12276>
- Harris, C. J. (2014). The onset of police misconduct. *Policing: An International Journal of Police Strategies & Management*, 37(2), 285–304. <https://doi.org/10.1108/PIJPSM-01-2012-0043>
- Helsby, J., Carton, S., Joseph, K., Mahmud, A., Park, Y., Navarrete, A., & Ghani, R. (2018). Early intervention systems: Predicting adverse interactions

- between police and the public. *Criminal Justice Policy Review*, 29(2), 190–209.
- His Majesty's Inspectorate of Constabulary and Fire & Rescue Services (HMIC-FRS). (2022). An inspection of vetting, misconduct and misogyny in the police service. London: HMICFRS. <https://www.justiceinspectors.gov.uk/hmicfrs/wp-content/uploads/inspection-of-vetting-misconduct-and-misogyny-in-the-police.pdf>. Accessed 28 December 2022.
- Huff, J., White, M. D., & Decker, S. H. (2018). Organizational correlates of police deviance: A statewide analysis of misconduct in Arizona, 2000–2011. *Policing An International Journal*, 41(1), 465–481.
- Jain, A., & Dubes, R. (1988). *Algorithms for clustering data*. Englewood Cliffs: Prentice Hall.
- Jain, A., Sinclair, R., & Papachristos, A. V. (2022). Identifying misconduct-committing officer crews in the Chicago police department. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0267217>
- Kane, R. J., & White, M. D. (2009). Bad cops: A study of career-ending misconduct among New York City police officers. *Criminology & Public Policy*, 8(4), 737–769. <https://doi.org/10.1111/j.1745-9133.2009.00591.x>
- Kappeler, V. E., Sluder, R. D., & Alpert, G. P. (1998). *Forces of deviance: Understanding the dark side of policing* (2nd ed.). Long Grove: Waveland Press.
- Lleti, R., Ortiz, M. C., Sarabia, L. A., & Sanchez, M. S. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimizes the silhouettes. *Analytica Chimica Acta*, 515(1), 87–100.
- Maheswari, K. (2019). Finding the best possible number of clusters using the k-means algorithm. *International Journal of Engineering and Advanced Technology*, 9, 533–538.
- McElvain, J. P., & Kposowa, A. J. (2008). Police officer characteristics and the likelihood of using deadly force. *Criminal Justice and Behavior*, 35(4), 505–521. <https://doi.org/10.1177/0093854807313995>
- McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of the *r* and *d*. *Psychological Methods*, 11, 386–401.
- Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*, 42(12), 1031–1039.
- Micucci, A. J., & Gomme, I. M. (2005). American police and subcultural support for the use of excessive force. *Journal of Criminal Justice*, 33(5), 487–500. <https://doi.org/10.1016/j.jcrimjus.2005.06.002>
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomised controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512), 1399–1411.
- Ouellet, M., Hashimi, S., Gravel, J., & Papachristos, A. V. (2019). Network exposure and excessive use of force: Investigating the social transmission of police misconduct. *Criminology & Public Policy*, 18(3), 675–704.
- Quispe-Torreblanca, E. G., & Stewart, N. (2019). Causal peer effects in police misconduct. *Nature Human Behaviour*, 3(8), 797–807. <https://doi.org/10.1038/s41562-019-0612-8>
- Raykov, Y. P., Boukouvelas, A., Baig, F., & Little, M. A. (2016). What to do when K-means clustering fails: A simple yet principled alternative algorithm. *PLoS ONE*, 11(9), 1–28.
- Rojek, J. J., & Decker, S. H. (2009). Examining racial disparity in the police discipline process. *Police Quarterly*, 12(4), 388–407. <https://doi.org/10.1177/1098611109348470>
- Rousseeuw, P. J. (1987). Silhouettes – a graphical aid to the interpretation and validation of cluster-analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Simpson, C. R., & Kirk, D. S. (2022). Is police misconduct contagious? Non-trivial null findings from Dallas. *Journal of Quantitative Criminology*. <https://doi.org/10.1007/s10940-021-09532-7>
- Stinson, P. M., Liederbach, J., & Freiburger, T. L. (2012). Off-duty and under arrest: A study of crimes perpetrated by off-duty police. *Criminal Justice Policy Review*, 23(2), 139–163. <https://doi.org/10.1177/0887403410390510>
- White, M. D., & Kane, R. J. (2013). Pathways to career-ending police misconduct: An examination of patterns, timing, and organizational responses to officer malfeasance in the NYPD. *Criminal Justice and Behavior*, 40(11), 1301–1325. <https://doi.org/10.1177/0093854813486269>
- Wolfe, S. E., & Piquero, A. R. (2011). Organizational justice and police misconduct. *Criminal Justice and Behavior*, 38(4), 332–353. <https://doi.org/10.1177/0093854810397739>
- Wood, G., Roithmayr, D., & Papachristos, A. V. (2019). The network structure of police misconduct. *Socius: Sociological Research for a Dynamic World*. <https://doi.org/10.1177/2378023119879798>

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

