

CRIME AND JUSTICE BULLETIN

NUMBER 278 | AUGUST 2026

Typologies of domestic and family violence perpetrators: A data-driven study

Fan Cheng

AIM

To classify Domestic and Family Violence (DFV) perpetrators into distinct groups based on their demographic characteristics and interactions with the NSW Criminal Justice, Child Protection, Housing and Homelessness systems.

METHOD

The study uses linked administrative data from seven NSW government systems, covering criminal offending and victimisation, Child Protection (CP) and Out-of-Home Care (OOHC), specialist homelessness services and housing assistance. A two-step analytical approach is used. First, Factor Analysis of Mixed Data is estimated on the full population of 196,701 individuals with at least one DFV offence finalised in court between 10 March 2008 and 1 January 2023, producing a common projection space across cohorts, genders and time periods. Second, K-prototype clustering is applied to 71,896 individuals born from 1987 onwards, for whom CP and OOHC histories are fully observed. Clustering is conducted separately for males and females, with group numbers selected using a grid search informed by separation and stability metrics. Groups are characterised by their demographic profiles, system contact patterns, and 12-month reoffending and revictimisation outcomes. Clustering for other cohorts and alternative designs is undertaken for sensitivity analysis only.

RESULTS

Six principal components (PCs) are identified, explaining one-third of variation across linked criminal justice, CP, housing and homelessness records. These PCs capture differences in lifetime intensity of crime perpetration; DFV and violent victimisation history; childhood maltreatment and youth diversions; recent DV, violent charges and AVO breaches; OOHC history, placement characteristics, and youth diversions; and housing and homelessness support service engagement. Using these components and demographic characteristics, gender-stratified clustering identifies four male and four female groups. Male groups differ mainly in the intensity and recency of criminal justice contact and multi-system exposure, while female groups are more strongly distinguished by cumulative victimisation histories, childhood system involvement and housing disadvantage. Differences in 12-month outcomes align with these profiles: groups with higher recent offending and justice contact show higher DFV reoffending, while groups, particularly among females, with extensive prior victimisation and housing engagement show higher DFV revictimisation. Results for older cohorts and alternative stratifications show similar patterns in sensitivity analyses.

CONCLUSION

DFV perpetrators in NSW can be divided into eight gender-stratified groups. The typology offers a policy-relevant framework for understanding heterogeneity in risk and system contact and may inform more targeted intervention design. Sensitivity analyses support its applicability across cohorts and time periods.

KEYWORDS

domestic violence

typology

factor analysis

cluster analysis

linked dataset

principal component

INTRODUCTION

Domestic and Family Violence (DFV) policy and practice often treat perpetrators as a homogenous group. However, there is substantial variation in perpetrators' offending patterns, backgrounds and risk profiles. Individuals differ in their demographics, relationship to victim(s) (e.g., parent, intimate partner, child, sibling), frequency and severity of offending, prior experiences as a victim of crime and interaction with government services. They also vary in less directly observable characteristics, including their motivations for violence, attitudes toward the victim and responsiveness to interventions. Perpetrators can also present with complex and intersecting needs, including histories of childhood maltreatment (Taillieu et al., 2016), substance use (Goldberg & Blaauw, 2019), and mental health issues (Fowler & Westen, 2011). Together, these differences underscore the limitations of a one-size-fits-all approach to intervention.

This study addresses the need for a more nuanced understanding of DFV by employing a typological approach (i.e., dividing perpetrators into distinct groups) using administrative data from multiple service systems. By identifying distinct (sub-)groups of perpetrators, typologies can support more targeted policy and service responses. Given the constraints of administrative data, particularly the limited capture of relationship dynamics, coercive control and motivations, the resulting typology should be interpreted as a set of data-driven administrative profiles summarising patterns of system contact and disadvantage. These profiles are intended to inform service planning and intervention design, rather than to provide a complete account of perpetrator behaviour. We focus on a younger cohort of perpetrators for whom early-life system contact can be observed in its entirety, allowing childhood maltreatment and OOHC experiences to be incorporated into the typology. The methodology can, however, be applied more broadly as data availability improves.

Existing typologies and their limitations

A substantial body of research has sought to classify DFV perpetrators into distinct types. Early typologies were largely theoretical or clinically-driven (see Boxall et al., 2015b for a summary). Johnson (2008), for example, distinguished between intimate terrorism, violent resistance, and situational couple violence, emphasising the role of coercive control. Holtzworth-Munroe and Stuart (1994) identified family-only, dysphoric/borderline, and generally violent/antisocial perpetrators, based on severity, psychopathology, and aggression. Other work differentiated emotionally dependent ("pit bulls") perpetrators from those who are calm and calculating ("cobras") (Jacobson & Gottman, 1998). While informative, these frameworks were typically based on small clinical samples and may not be generalisable to broader populations.

More recent work has identified distinct groups using empirical methods applied to large administrative datasets. For example, Wells et al. (2016) applied latent class analysis to a Canadian sample of over 1,600 male perpetrators, identifying four subgroups based on trauma, substance use, and criminal history. In the UK, Robinson and Clancy (2021) developed the Priority Perpetrator Identification Tool using police data to identify high-harm DFV offenders. In Australia, Boxall et al. (2015a) examined prior offending among DFV perpetrators in Tasmania, finding that many had extensive histories of both family and non-family violence. The Crime Statistics Agency Victoria (2020) used latent class modelling to identify seven perpetrator groups based on their demographics and offending history. Dowling et al. (2021) extended this evidence base by examining criminal career trajectories of domestic violence offenders in Australia, highlighting patterns of persistence and escalation. More recently, Howard and Kebbell (2026) examined DFV typologies using Queensland police data for 10,956 male offenders and identified a four-cluster solution comprising homicidal, generally violent/antisocial, low-level antisocial, and family-only offenders.

However, most existing typologies derived from administrative data rely on a relatively narrow set of data sources (typically police or court records), limiting their ability to capture the full spectrum of perpetrator experiences and service interactions. Another limitation is the predominant focus on Intimate Partner Violence (IPV), despite DFV encompassing violence within a wider range of relationships, including adolescent violence towards parents and extended family violence. Typologies for these forms of DFV are less well developed, as are typologies for female perpetrators. Existing exceptions (Babcock et al.,

2003; Boxall & Sabol, 2021; Howard & Kebbell, 2026; Moulds et al., 2019; Swan & Snow, 2002) suggest important difference in patterns of violence, underscoring the need for inclusive frameworks that account for gender and relationship diversity.

Choice of factors for the typology

Typologies have been promoted as a means of tailoring interventions to perpetrator characteristics. Evidence suggests they may improve treatment design and delivery by identifying those who are most likely to benefit from specific treatment models (Gondolf, 1997; Saunders, 1996), by predicting who will engage and complete programmes (Saunders, 1996), and by informing risk assessments (Hilton & Harris, 2005; Loinaz, 2014). In this way, typologies offer a structured framework for differentiating intervention strategies and allocating resources more effectively (Holtzworth-Munroe & Meehan, 2004). However, despite their promise, the application of typologies in practice remains limited. Critics of the approach argue that typologies often oversimplify complex behaviours and lack consistency across studies (Capaldi & Kim, 2007). Others caution that typologies may fail to capture dynamic risk factors and contextual variability (Johnson, 2008). Boxall et al. (2015b) also question the generalisability of existing DFV typologies to Australian contexts, suggesting that typologies be validated with local data and modified where appropriate. These limitations underscore the need for robust, empirically grounded approaches that generate actionable insights while acknowledging uncertainty.

Developing a meaningful typology therefore requires careful selection of factors consistently linked to DFV perpetration. Key literature reviews highlight several important dimensions related to violence severity, generality of offending (family-only vs. generally violent), psychopathology (e.g., dysphoric/ borderline or antisocial), and underlying motivations (e.g., coercive control vs. situational reactivity) (Alexander & Johnson, 2023; Holtzworth-Munroe & Stuart, 1994). Individual characteristics, including demographics (e.g., age, gender, and socioeconomic disadvantage), adverse childhood experiences (e.g., childhood maltreatment or exposure to violence), and behavioural factors (e.g., substance use, mental health conditions, and antisocial behaviour) also emerge as key distinguishing features (Boxall et al., 2015b; Giesbrecht et al., 2025). Some frameworks emphasise internalising features (e.g., emotional dysregulation), while others focus on externalising behaviour (e.g., aggression, antisociality), with family-only subtypes showing lower psychopathology and violence levels compared to generally violent/antisocial groups (Giesbrecht et al., 2025). Relational and situational factors are also thought to play an important role. These include the presence of coercive control, patterns of conflict escalation, victim-offender relationship type, and the stability of housing and family environments (Beckwith et al., 2023).

In designing the current study, we considered variables that: (1) are theoretically linked to DFV perpetration; (2) capture meaningful differences between groups; and (3) can be reliably measured using administrative data in NSW. This led to the inclusion of variables relating to prior DFV and general offending, crime victimisation histories, interactions with child protection and out-of-home care systems, housing and homelessness service use, and broader sociodemographic markers. These dimensions align with practical decision-making needs in policy and service delivery.

The current study

The aim of this study is to classify DFV perpetrators into distinct groups and then describe the characteristics of each group. To achieve this, the study draws on linked administrative data from seven NSW government systems, incorporating information on criminal offending, victimisation, child protection involvement, homelessness service use and public housing contact. Using this multidimensional dataset, we apply a two-step analytical process. First, Factor Analysis of Mixed Data (FAMD) is used to summarise major patterns of system contact, offending and victimisation across the full DFV sample, providing a common projection space that supports consistent interpretation across cohorts, gender and time periods. This is followed by K-prototype clustering to classify individuals into groups based on these components and demographic characteristics. The main typology is constructed within the young cohort (who have complete information) using gender-stratified clustering (four groups for males and four

for females). Sensitivity analyses test whether the group structure is robust to pooled (non-stratified) clustering, alternative cohort observability strata, and different time windows.

This study is guided by four research questions:

1. Can DFV perpetrators in NSW be classified into distinct groups using linked administrative data across criminal justice, child protection, housing and homelessness systems?
2. Which principal components best summarise major dimensions of perpetrators' offending, victimisation, and service-system contact?
3. What demographic, behavioural, and system-contact characteristics distinguish the resulting perpetrator groups, and how might these differences inform the design of DFV interventions in NSW?
4. To what extent is the resulting group structure robust to alternative modelling choices, cohort definitions and time periods?

Together, these analyses provide the first NSW-based, data-driven typology of DFV perpetrators constructed using wide-ranging linked administrative data. The resulting typology is intended to function as a descriptive framework for understanding heterogeneity among DFV perpetrators, rather than as a predictive or intervention-matching tool. The groups represent administrative profiles derived from observed system interactions and therefore reflect constellations of disadvantage, exposure and risk that may be relevant for service coordination, resource allocation and strategic planning. Sensitivity analyses are used to assess whether the broad structure of the typology generalises across cohorts and time windows, supporting cautious application beyond the primary analytical cohort where appropriate. The next section of this report describes the data sources, research sample, variables and empirical strategy utilised in this study. The section proceeding that outlines our main results and sensitivity analysis. We then conclude with a discussion of the main findings of this paper.

METHOD

Data sources

For the purposes of this study, we utilise an extract from the Department of Community and Justice's (DCJ) Linked Data Asset (LinDA). LinDA contains a wealth of information drawn from seven separate DCJ databases, each of which is described below.

1. **BOCSAR's Reoffending Database (ROD)** contains information on all individuals charged with an offence by NSW Police after 1 January 1994, that went on to be finalised¹ on or before 31 March 2024. For the purposes of this study, it is instructive to think about the information contained within ROD as consisting of three (sub-)sets (of data): court records; custodial records; and victimisation records. Court and custodial records provide information centred around the person accused of an offence. Victimisation records provide information centred around the person that was alleged to have been the victim of an offence. Further details regarding each of these three sets are provided below.

- 1.1. **Court records:** For each individual accused of an offence, ROD enables us to observe their date of birth, gender, Aboriginality, SEIFA percentile rank (Socio-Economic Indexes for Areas,

¹ Where finalisation refers to the court appearance where all legal proceedings for a given matter come to an end. For individuals that entered into a plea of guilty, the finalisation date is the date that the penalty is handed down. For individuals that entered into a plea of not guilty and thus proceeded to trial or defended hearing, the finalisation date depends on the outcome of the trial or defended hearing. If the defendant was found to be guilty, the finalisation date is the date that the penalty is handed down. If instead, the defendant was found to be not guilty, the finalisation date is the date that such a determination was made. Among matters where the prosecution withdraws all charges, the finalisation date is the date that all charges were withdrawn.

2021; Australian Bureau of Statistics, 2023)²; as well as the nature (e.g., Australian and New Zealand Standard Offence Classification code [ANZSOC; Australian Bureau of Statistics, 2011]) of each offence they have been charged with; the outcome of each charge (e.g., guilty, not guilty, withdrawn); and the resulting sanction if they are found guilty (e.g., imprisonment, fine, community service).

1.2. Custodial records: Also contained within ROD is a complete custodial history of all individuals charged with an offence by NSW Police from 2000 to 2022. For such individuals, we are able to observe the date that they entered and exited an NSW correctional facility, as well as the legal status³ at the time of reception—whether each episode was the result of a bail decision (i.e., remand) or penalty imposed following a proven⁴ offence (i.e., sentenced episodes).

1.3. Victimisation records: The victimisation information contained within ROD consists of information about individuals that NSW Police believe to have been the victim of a violent or property crime. For the purposes of this study, such information includes: a complete list of criminal incidents reported to and/or detected by NSW Police (including DFV and sexual offence related incidents); the date of the incident; where the incident took place; information regarding the nature of the incident (e.g., the type of crime that may have occurred); the relationship between the victim and Person of Interest (POI)⁵; and whether the POI was charged by police.

2. The Child Protection (CP) dataset includes information on Children and Young People (CYP) aged under 18 who are, or may be, at risk of significant harm. The CP extract contained in LinDA consists of the details of CYP identified in a concern report and/or a Risk of Serious Harm (ROSH) report between 1987 and 2022. Key variables include: the number of ROSH reports; the final risk level recorded in a concern report (e.g., very high, high, medium, low); and safety and risk evaluations conducted during helpline assessments (e.g., assessed issue types, primary CP harm types).

3. The Out-of-Home Care (OOHC) dataset captures service response details for children who could not safely live with their own families and are placed in an OOHC setting. The data ranges from 1987 to 2022, where each record provides information for an actual OOHC placement event. Records reflect the last care event for a CYP occurring on or before the end of the financial year. Such information includes, for example: total number of OOHC placements; total days spent in OOHC; types of caregivers (e.g., kinship, foster); and placement type (e.g., residential care, relative care).

4. The Specialist Homelessness Services (SHS) dataset contains information for individuals who requested assistance from a SHS provider between July 2015 and December 2022. The extract includes information about an individual's situation (e.g., with or without children); the reason(s) for requesting assistance (e.g., domestic violence, housing crisis); and the type of support provided (e.g., crisis accommodation, referral to DFV support services).

5. The Social Housing Applications dataset contains information on all social housing applications between 1970 and 2022. For each application, we are able to observe the date and nature of the application (i.e., public, community or Aboriginal housing).

6. The Public Housing Tenancy dataset contains information on all public and Aboriginal housing tenancies between 1999 and 2022. For each tenancy, we are able to observe the tenancy start and end dates, whether the stay was in public housing or Aboriginal housing, and the tenants' primary source of income (e.g., age pension, carer pension, Jobseeker, parenting allowance).

² Recorded at the time of finalisation of the defendant's court matter; for each individual, values are averaged across all matters finalised prior to 1 January 2023.

³ Recorded at the time of reception into custody and may change during the custodial episode (e.g., from remand to sentenced following court proceedings).

⁴ A 'proven' offence refers to an offence where guilt has been established (e.g., through a guilty plea or at trial). In NSW, a proven offence differs from a conviction in that an offender can have a proven offence with or without conviction. Proven offences without conviction are not identified on standard criminal record checks. In this sense, a criminal conviction is a form of penalty imposed by the courts following the establishment of guilt.

⁵ A POI is the term NSW Police use to refer to suspected offenders in connection with a criminal incident.

7. **The Private Rental Assistance (PRA)** dataset contains information on short-term and emergency housing support provided to individuals struggling to access the private rental market between 1999 to 2022. Key information includes total payment amounts; and type of rental assistance products (e.g., bond loans, temporary accommodation, rent choice subsidy).

The linked data allow individuals’ interactions with multiple service systems to be observed over extended periods, enabling the construction of a multidimensional profile of system contact across justice, welfare and housing domains. Differences in the time coverage of the constituent datasets, particularly for CP and OOHC records, are addressed explicitly through the definition of the estimation sample and the use of sensitivity analyses described later in the paper.

Defining DFV perpetrators: estimation sample

To construct the estimation sample, we use the ROD extract to identify individuals with at least one charge (proven or not proven)⁶ for a DFV offence that was finalised in an NSW criminal court between 10 March 2008 and 1 January 2023. This period is selected for two reasons: first, the DFV flag in ROD, which enables us to differentiate between DFV and non-DFV related offences, came into effect on 10 March 2008;⁷ and second, given that our ROD extract contains information for charges finalised up to 31 March 2024, we impose the 1 January 2023 cutoff to allow for sufficient time for most subsequent (re)offences to progress through the courts so that we can investigate 12-month reoffending outcomes for all individuals in the estimation sample. In total, our estimation sample consists of information for 196,701 DV perpetrators with at least one DFV charge finalised between 10 March 2008 and 1 January 2023. These individuals constitute what we refer to as “the spine” of our estimation sample.

Before moving on, it is worth making clear that the datasets held in LinDA do not cover the same time periods. This is illustrated by Table 1 which shows the first year that information is available in each dataset. The Social Housing Applications and PRA datasets, for example, go back as far as 1970, while the SHS dataset only has information available from 2015. From Table 1, we can also see that all individuals in the spine have at least one finalised charge in ROD. This is because, by construction, to be included in the estimation sample, an individual needs to have had at least one DFV charge finalised between 10 March 2008 and 1 January 2023. We can also see that just under one-half of these individuals have spent time in custody, applied for social housing, or were the victim of a crime reported to police; and between one-third to one-quarter of individuals in the estimation sample had contact with the public housing, child protection and SHS systems.

Table 1. Databases utilised in this study by year of availability

Database	Year data available from	Number of individuals with at least one record	Proportion of individuals with at least one record (%)
ROD – Court Records	1994	196,701	100.0
ROD – Custody Records	2000	88,561	45.0
Social Housing Applications	1970	85,760	43.6
ROD – Victimisation Records	1994	84,071	42.7
Private Rental Assistance	1970	64,299	32.7
Public Housing Tenancy	1999	60,043	30.5
Child Protection	1987	53,814	27.4
Specialist Homelessness Services	2015	43,767	22.3
Out-of-home Care	1987	12,843	6.5

6 To be clear, our estimation sample consists of people charged with a DFV offence, regardless of how that charge was finalised. For ease of explanation, we refer to these individuals as ‘DFV perpetrators’.
 7 Which classifies offences as being DFV related in accordance with the *Crimes (Domestic and Personal violence) Act 2008 (NSW)*.

One notable feature of Table 1 is that less than seven per cent of individuals in the spine have a recorded contact with the OOHC system. This number should be interpreted with caution. It does not necessarily mean that only seven per cent of DFV perpetrators have had contact with OOHC system. This is because the OOHC data only extends back to 1987, and since individuals enter the OOHC system before age 18, those born before 1969 (that is, aged 54 or older at the end of our observation window in 2023) do not have any information recorded in the OOHC data (even if they did have contact with the OOHC system). As a result, the true proportion of DFV perpetrators in contact with the OOHC system is likely underestimated in Table 1. Similar reasoning applies to the CP data, which also begin in 1987 and therefore systematically under-record CP involvement among older individuals in the sample.

Because the CP and OOHC datasets only begin in 1987 and individuals can only appear in these systems before age 18, the extent to which a person's childhood history is observable varies by birth year. In our sample, approximately 36.6 per cent (71,896 individuals) have childhood years that fall entirely within the post-1987 period and are therefore fully observable; 45.4 per cent (89,344 individuals) have childhood years that are only partly observable; and 18.0 per cent (35,461 individuals) have no observable CP/OOHC history because all of their childhood years occurred before 1987. These differences reflect structural features of the data rather than true variation in system contact.

Note that the child protection data refer to instances where individuals in our sample were themselves recorded as children requiring protection, rather than as parents or guardians. Together, these data limitations mean that CP and OOHC contact is not observed consistently across all cohorts. For this reason, the main clustering analysis is restricted to the subgroup of 71,896 individuals born from 1987 onwards, for whom childhood CP and OOHC histories are fully observed. This restriction ensures that early-life system contact is measured consistently and avoids treating the absence of records arising from data limitations as evidence of no childhood system involvement. Individuals with partially or non-observable childhood histories are retained in the full dataset and are used in sensitivity analyses examining the robustness of the typological structure across cohorts and data-coverage regimes. Throughout the paper, differences in CP and OOHC prevalence are interpreted with reference to these observability constraints.

Constructed variables

Recall from the previous section that the purpose of this study is to classify individuals in the spine into distinct groups. We perform this classification using 277 variables drawn from LinDA, including four demographic variables, (i.e., gender, Aboriginality, age at first DV finalisation date, and SEIFA percentile rank). The 277 variables are aggregated at the individual level to conduct our analysis. These variables are aggregated in accordance with the nature of the information they contain. When counting events (e.g., ROSH reports, proven offences), we take the summation to 1 January 2023;⁸ when examining the presence of a condition (e.g., being male), we utilise binary variables equal to one if the condition is met, zero otherwise;⁹ and finally, we transform categorical variables (e.g., risk level identified in a child protection concern report) into a set of binary variables.

⁸ Note that the time period over which each variable is summed differs between datasets. For variables contained within the three ROD extracts, we sum variables between 10 March 2008 and 1 January 2023. For variables contained within the other datasets, we take the sum from the earliest available date (e.g., 2015 for the SHS data, 1987 for the OOHC and CP data) to 1 January 2023.

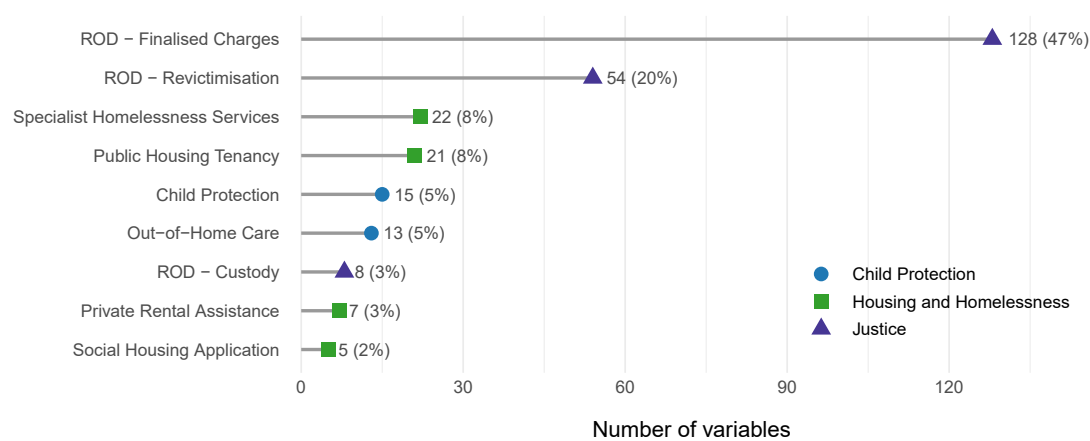
⁹ A single individual can have multiple (distinct) values for their date-of-birth, Aboriginality and gender in our sample. This occurs because, for a single individual, the value associated with these variables differs between databases. To address this issue, we use the most commonly recorded value for date-of-birth and gender in our analysis. Aboriginality is extracted from ROD using a derived person-level indicator. This indicator is constructed by linking police records that may refer to the same person and applying a 20% rule, whereby a person is classified as Aboriginal if at least 20% of linked police records with known Aboriginality are recorded as Aboriginal. Where more than one derived person record is linked to a single individual in our analysis spine, the individual is classified as Aboriginal if any linked ROD-derived indicator is classified as Aboriginal. In this report, we use the term 'Aboriginal' to refer to Aboriginal people, people that identify as Torres Strait Islander and people who identify both as Aboriginal and Torres Strait Islander.

Given the differing timespans of the constituent databases contained within LinDA, some of the aggregated variables have a large number of missing values for a given individual. We deal with this issue by either recoding the variable's value to zero, or imputing the variable's value. Variables we recode as zero refer to counts of events. For example, if an individual has zero requests for assistance from an SHS provider, all variables in SHS data will be missing. In this case, we recode their count of requests for assistance from an SHS provider to zero. We impute the value of only one variable in our analysis: SEIFA percentile rank. For each individual's SEIFA percentile rank, we impute missing values using the Multivariate Imputation by Chained Equations (MICE; van Buuren & Groothuis-Oudshoorn, 2011) method.¹⁰ Because the empirical strategy involves dimension reduction prior to clustering, most constructed variables are used to derive latent dimensions of system contact rather than being interpreted individually. Key demographic variables are retained outside the dimension-reduction process to preserve their interpretability and to support profiling of the resulting groups.

Figure 1 provides a graphical summary of “where” (i.e., in terms of which constituent database) each of the 273 variables (excluding the four demographic variables) used in this study are drawn from. From Figure 1, we can see that the majority of variables (70%) are drawn from Justice data (i.e., ROD). The number of variables from other datasets is much smaller, with 55 (20%) variables from Housing and Homelessness services and 28 (10%) from Child Protection services.

Given the number of variables utilised in this paper, we do not provide a detailed description of each variable in this section of the report.¹¹ Instead, we focus on describing what we believe to be the most policy relevant subset of these variables. That description constitutes the remainder of this subsection.

Figure 1. Number of variables from each database



Demographics

Table 2 reports the number and percentage of people sharing various demographic characteristics. In our sample, 78% of DFV perpetrators are male, and 27% identify as Aboriginal. There are 1,851 people under the age of 18, which is about 1% of the sample, and most people (64%) were aged between 30 and 54 on 1 January 2023.¹² The average SEIFA percentile rank of individuals in our sample is 41, which indicates that DFV perpetrators reside in areas more disadvantaged than about 59% of the country.

10 This technique models the missing SEIFA percentiles as a function of other observed variables in the cleaned analysis dataset, including demographic characteristics and variables capturing individuals' contact with the criminal justice, CP and housing systems, to predict plausible values for the missing SEIFA percentile.
 11 Summary statistics and detailed descriptions of all 273 variables used in this study are available from the author upon request; a subset is presented in Table 4.
 12 When referring to age in this study, unless explicitly stated otherwise, we calculate an individual's age as of 1 January 2023.

Table 2. Summary statistics for DFV perpetrators included in the spine of the estimation sample (n=196,701)

Demographics		Number of individuals	Proportion of individuals (%)
Male		152,944	77.8
Aboriginal		53,555	27.2
Age group	10-17	1,851	0.9
	18-29	33,881	17.2
	30-39	52,091	26.5
	40-54	73,417	37.3
	55+	35,461	18.0
SEIFA percentile rank (mean/SD)*		41.2	24.7

Note. * Continuous variables are reported as the average and standard deviation (SD). SEIFA percentile rank ranges from 0 to 100, where 50 represents the national median. Values below 50 indicate areas that are more disadvantaged than the national average, while values above 50 indicate more advantaged areas.

Recent reoffending and revictimisation outcomes

We construct two descriptive outcome measures for perpetrators in our sample: recent reoffending and recent revictimisation. These measures are used to describe the sample overall and across groups in the typology, but are not used in the analysis to generate the groups.

Recent reoffending was measured by identifying individuals in the estimation sample whose most recent proven offence was finalised in 2022 (and designating this as their index offence). Individuals were classified as recent reoffenders if they had at least one subsequent proven offence which occurred within 12 months and was finalised within 15 months of the index offence finalisation date.¹³ We calculated separate measures for general reoffending and DFV reoffending. For DFV reoffending, both the index offence and the subsequent offence were required to be DFV-related. The general reoffending measure included both DFV and non-DFV related offences.

Recent revictimisation was measured in a similar way. We identified individuals whose most recent police-recorded victimisation incident¹⁴ occurred in 2022 and treated this as the index victimisation event. Individuals were classified as recently revictimised if they experienced at least one subsequent police-recorded victimisation incident within 12 months of the index event. Separate measures were constructed for general revictimisation and DFV revictimisation. For DFV revictimisation, both the index incident and the subsequent incident were required to be DFV-related. The general revictimisation measure included both DFV and non-DFV related incidents.

In Table 3, we report the recent reoffending and revictimisation rates for general offences and DFV offences, defined as the proportion of individuals with a proven offence or police-recorded victimisation within 12 months of their most recent observed event. The rates are reported as percentages and are averaged across the full sample, and by gender, Aboriginality and age group. Overall, the general reoffending rate is 21.7 per cent, and the DFV reoffending rate is just under half of that at 9.6 per cent. From Table 3 we can also see that both the general and DFV reoffending rate is higher for males and Aboriginal people. Table 3 also reports the general and DFV revictimisation rates for individuals in the spine of our estimation sample. For these individuals, we can see that the general revictimisation rate is about 18 per cent and the DFV revictimisation rate is around 7 per cent. Females and Aboriginal people are about twice as likely to be a repeat victim compared with males and non-Aboriginal people, respectively.

¹³ We refer to this outcome as “recent” because it is based on the most recent period for which a complete follow-up window was available at the time of data extraction. Rates are reported using a 12-month follow-up window because this is a standard short-term measure in NSW reoffending research and it captures the period in which DFV recidivism risk is highest (BOCSAR, 2026; Trevena & Poynton, 2016). Restricting the sample to a long follow-up window (e.g., over 10 years via birth cohort) would exclude the majority of perpetrators (sample average age of 41 years) and eliminate older groups, undermining typology representativeness.

¹⁴ Which we define as NSW Police recording an individual as the victim of a violent or property crime, regardless of whether that crime went on to be proven in court.

Table 3. The 12-month reoffending and revictimisation outcomes (%) for DFV perpetrators, by gender, Aboriginality and age group

	General reoffending	DFV reoffending	General revictimisation	DFV revictimisation
Full sample	21.7%	9.6%	17.8%	7.1%
Male	23.0%	10.1%	14.7%	4.2%
Female	17.2%	8.0%	28.6%	17.0%
Aboriginal	34.2%	13.3%	27.4%	12.5%
Non-Aboriginal	17.0%	8.2%	14.2%	5.0%
Child (<18yr)	18.1%	10.7%	17.0%	8.8%
Adult	21.8%	9.6%	17.8%	7.1%

Note. This table reports the proportion of individuals in the estimation sample who: (i) commit at least one subsequent offence within 12 months of their most recent court finalisation (reoffending), or (ii) is recorded as a victim in at least one subsequent police-recorded incident within 12 months of their most recent victimisation (revictimisation). Results are shown for both general and DFV-related outcomes, where DFV measures require both the index event and the subsequent event to be DFV-related. In total, there are 196,701 individuals in our estimation sample.

Other perpetrator features

In addition to the four demographic variables, there are 273 other person-level features constructed from the linked database and used in the analysis. Table 4 reports summary statistics associated with a selection of the variables used in the analysis. From Table 4, we can see that between 2008 and 2022, on average, a DFV perpetrator in our sample was the defendant in three ADVOs; charged with seven offences, two of which were for assault; and spent over six months in custody. From Table 4, we can also see that about 10 per cent of individuals in the spine sought assistance from an SHS provider between 2015 and 2022, and about one-quarter resided in public housing between 1999 and 2022.

Table 4. Summary statistics of a subset of variables for the sample

Variable name	Variable type	Mean* (or n)	SD (or %)
ROD – Judicial Records (2008-2022)			
Number of finalised court appearances (with proven offence/s) as a juvenile or adult in the 1 year prior to the latest index contact	Numerical	0.4	0.8
Number of finalised court appearances prior to the latest index contact with the justice system	Numerical	4.4	6.3
Number of charges for Acts Intended to Cause Injury	Numerical	1.8	1.9
Number of times an individual was granted bail	Numerical	2.3	2.8
Number of times an individual was charged with an offence	Numerical	7.2	8.3
Number of times an individual was charged with a DFV offence	Numerical	1.7	1.5
Number of times the person was found guilty in court	Numerical	5.3	6.6
Number of times an individual received a sentence of imprisonment	Numerical	0.7	2.2
Number of times an individual pleaded guilty for a DFV offence	Numerical	1.9	2.8
Number of ADVOs where an individual was the defendant in the five years prior to each index date	Numerical	3.0	5.0
Length of all prison sentences handed down to an individual (in months)	Numerical	8.3	26.7
Number of cautions issued under <i>Young Offenders Act 1997</i> (NSW)	Numerical	0.3	0.7
ROD – Custody Records (2008-2022)			
Days in custody for each individual	Numerical	208.3	660.3
Number of custodial episodes for each individual	Numerical	2.3	4.9
Whether an individual had at least one custodial episode as an adult	Binary	84515	43.0%
Whether an individual had at least one custodial episode as a juvenile	Binary	13001	6.6%

Table 4. Summary statistics of a subset of variables for the sample (*continued*)

Variable name	Variable type	Mean* (or n)	SD (or %)
<i>Social Housing Applications (1970-2022)</i>			
At least one housing application with a priority level of general	Binary	71477	36.3%
At least one housing application with a priority level of priority	Binary	29824	15.2%
<i>ROD – Victimisation Records (2008-2022)</i>			
Number of events where an individual was the victim of at least one DFV related incident	Numerical	1.6	4.1
Number of events involving this victim within one year prior to the latest index event date	Numerical	0.3	0.8
Number of incidents where the POI was an intimate partner of the victim	Numerical	0.5	1.4
Number of times the individual was a victim at a residential premises	Numerical	1.5	3.6
Number of events an individual was the victim of a violent crime	Numerical	1.0	2.2
Number of events an individual was the victim of a property crime	Numerical	0.0	0.2
<i>Private Rental Assistance (1970-2022)</i>			
Total monetary amount received in the form of private rental assistance	Numerical	1191.3	4443.3
Whether an individual was placed in temporary accommodation at least once	Binary	52564	26.7%
<i>Public Housing Tenancy (1999-2022)</i>			
Days an individual spent in public housing	Numerical	868.1	1992.7
Whether an individual lived in public housing at least once	Binary	48952	24.9%
Whether the individual ever had Newstart or Jobseeker as their primary income in public housing	Binary	18357	9.3%
<i>Child Protection (1987-2022)</i>			
Number of child protection reports where the helpline assessed issue included carer issues	Numerical	0.8	3.0
Number of child protection reports where the helpline assessed issue included child and young person issues	Numerical	1.0	7.6
Number of child protection reports where primary harm was physical	Numerical	0.0	0.9
Total number of ROSH events	Numerical	2.2	7.1
Whether an individual's average Child Protection primary risk level is high	Categorical	2596	1.3%
Whether an individual's average Child Protection primary risk level is very high	Categorical	881	0.4%
Whether an individual had contact with Child Protection system	Binary	53814	27.4%
<i>Specialist Homelessness Services (2015-2022)</i>			
Days an individual was in receipt of support from a Specialist Homelessness Service provider	Numerical	55.5	198.4
Whether an individual ever received assistance from a SHS provider due to a housing crisis	Binary	20153	10.2%
Whether an individual ever received assistance from a SHS provider due to an inadequate or inappropriate dwelling	Binary	9407	4.8%
<i>Out-of-Home Care (1987-2022)</i>			
Whether an individual had contact with Out-of-Home Care system	Binary	12733	6.5%
Days an individual spent in Out-of-Home Care	Numerical	106.8	631.4
Number of Out-of-Home Care placements for an individual	Numerical	0.2	1.0
Whether an individual in contact with OOHHC ever had an Aboriginal caregiver	Binary	4466	2.3%
Whether the individual was ever actually placed with a relative of Aboriginal Kinship during OOHHC	Binary	6304	3.2%

Note. *For summary statistics: numerical variables are presented as mean and standard deviation; binary variables as counts (n) and proportions (%); and categorical variables as counts (n) and proportions (%) for each category.

Empirical strategy

Recall that our central aim is to divide the 196,701 DFV perpetrators in our sample into a relatively small number of meaningful groups. Selecting the appropriate number of groups involves balancing analytical precision with practical relevance. That is, too many groups limit the usefulness of the findings for policymakers because it is not practical to develop and implement a large number of interventions tailored to many groups. On the other hand, creating too few groups risks not capturing the granular differences between perpetrators that might be key to the tailoring of successful interventions.

To address these competing concerns while maintaining comparability across cohorts, genders and time periods, we adopt a two-step approach. In the first step, we apply Factor Analysis of Mixed Data to reduce the number of variables. This step is essential given the large number of variables that we use to classify individuals into groups. Too many variables is an issue because many of them are correlated with one another, making them partially redundant. Of the total 277 variables in our estimation dataset, we exclude four demographic variables from the FAMD process to preserve their interpretive value. These variables include: Aboriginality, gender, age at first DV finalisation and SEIFA percentile rank. FAMD on the full population ($n = 196,701$) allows us to summarise the information across the remaining 273 variables into a small set of uncorrelated components. The result of this process is a set of PCs which retain the most useful information in the data. In the second step, we use K-prototype clustering analysis to define the main typology within the young cohort with complete childhood observability (born from 1987 onwards; $n = 71,896$), stratified by gender. This method is well-suited to datasets containing both numerical and categorical variables. The result yields a set of perpetrator groups. Further detail regarding both steps constitutes the remainder of this section.

Step 1: Factor Analysis of Mixed Data to summarise the data

In order to classify DFV perpetrators into meaningful groups, we need to reduce the number of variables from 273 to a much smaller, more manageable set. This step is important for both computational efficiency and analytical clarity. A dataset with 273 variables can be difficult to work with because it increases computational complexity, makes it harder to detect meaningful patterns, and can introduce noise or redundancy that weakens the performance of the clustering algorithm used in Step 2. To reduce the number of variables, we employ FAMD,¹⁵ a technique designed to integrate different types of variables (e.g., continuous variables like number of days in custody, and binary variables like whether an individual has spent time in public housing) into a common framework. This integration is achieved through the use of Principal Component Analysis for continuous variables and Multiple Correspondence Analysis for categorical variables, producing PCs that summarise major dimensions of offending, victimisation and multi-system service contact.

FAMD involves the construction of several PCs for all variables in the data. Each PC is a weighted (linear) combination of the original variables in the underlying data such that all the resulting PCs are uncorrelated with one another. Said differently, each PC contains information from all 273 variables, but the importance assigned to each piece of information (or variable) differs between PCs such that the PCs are uncorrelated with one another. FAMD is estimated once on the full DFV population ($n = 196,701$), producing a single fixed PC space that can be used consistently when comparing patterns across genders, cohorts and time windows. To preserve the interpretability of key demographic variables (i.e., gender, Aboriginality, age at first DV finalisation, SEIFA percentile rank), we exclude them from the FAMD process. These variables are omitted because they are of separate interest to policymakers.¹⁶ The remaining 273 variables are included in the FAMD to summarise the information in the data. Importantly, the set of variables omitted from the FAMD step guides both: the sorting of individuals into groups (described in the next step) and interpretation of the resulting groups. The FAMD solution retained six PCs, which captured information on (1) lifetime intensity of crime perpetration; (2) DFV and violent victimisation history; (3) childhood maltreatment and youth justice diversions; (4) recent DV, violent charges and AVO breaches;

¹⁵ We implement FAMD using the **FactoMineR** (Lê et al., 2008) and **factoextra** (Kassambara & Mundt, 2020) packages in R.

¹⁶ Exclusion of different variables (e.g., based on insights from policymakers and operational staff) would likely result in a different set of groups.

(5) Out-of-home care history, placement characteristics, and youth diversions; and (6) housing and homelessness support service engagement.

Step 2: K-prototype clustering to classify perpetrators into groups

Following the FAMd, we apply K-prototype clustering to classify DFV perpetrators into distinct groups within the young cohort (born from 1987 onwards; $n = 71,896$) for whom childhood CP and OOHC histories are fully observed.¹⁷ This approach extends the classical K-means algorithm to integrate mixed data types (e.g., continuous and categorical variables) by minimising squared Euclidean distances (for continuous variables) and simple matching dissimilarities (for categorical variables) (Huang, 1998). Clustering is conducted separately for males and females to allow gender-specific grouping structure. Within each gender, the clustering model uses the PCs derived from the FAMd step, along with the four excluded demographic variables, to classify individuals into groups.

The optimal number of groups, usually denoted as “ k ”, was selected separately for males and females using a pre-specified grid search over values of k ranging from 2 to 12 on a stratified subsample of 10,000 individuals. Model selection was guided by three complementary criteria: the average silhouette coefficient, the Calinski-Harabasz Index (CH Index; Caliński & Harabasz, 1974), and the bootstrap stability using the Adjusted Rand Index (ARI; 100 bootstrap resamples per k). Across both genders, these metrics consistently indicated that $k = 4$ provided the best balance between within-cluster cohesion, between-cluster separation, stability under resampling, and substantive interpretability. On this basis, four groups were retained for males and four for females in the main typology (M1-M4; F1-F4).

Once the categorisation is complete, we describe each group in terms of its size and key characteristics. Descriptive statistics are presented for both the four excluded variables (i.e., the demographics) and the PCs, as well as the DFV reoffending and DFV revictimisation outcomes. Based on the dominant traits within each group, we assign intuitive names to facilitate the interpretation of each group.

Cluster metrics and stability

To select a robust number of perpetrator groups, we evaluated each candidate K-prototype solution using three complementary cluster stability metrics calculated on a stratified subsample of 10,000 defendants from the young cohort of 71,896. The subsample was stratified by age band, Aboriginality and SEIFA percentile of socioeconomic disadvantage to preserve the joint distribution of key demographic covariates. For each gender, clustering was performed in the six-dimensional space of FAMd component scores, supplemented with the three demographic variables used in the main typology.

First, we computed the average silhouette coefficient, which compares, for each defendant, the average dissimilarity to cases in the same cluster with the average dissimilarity to cases in the nearest alternative cluster (Rousseeuw, 1987). Values range from -1 to 1 , with higher values indicating that observations are, on average, closer to their assigned cluster than to neighbouring clusters. We report the overall mean across defendants for each value of k .

Second, we calculated the CH Index, defined as the ratio of between-cluster dispersion to within-cluster dispersion, scaled by the number of clusters and observations. CH index values are positive values, with larger values indicating more compact and well-separated clusters in the FAMd space, and are commonly used to guide the choice of k in clustering algorithms.

Third, we assessed bootstrap cluster stability using the Adjusted Rand Index, which compares agreement between two clustering while correcting for chance (Hubert & Arabie, 1985). For each candidate k , we treated the K-prototype solution on the full subsample as a reference and drew bootstrap resamples with replacement. Within each resample, we re-estimated the K-prototype solution, aligned the resulting clusters to the reference using the Hungarian algorithm on centroid distances in the FAMd space, and computed the ARI between the aligned bootstrap labels and the reference labels. We summarise stability by the mean ARI and its range across resamples. Values range from -1 to 1 , with higher mean ARI and

¹⁷ We implement the K-prototype clustering procedure using the `c1ustMixType` package (Szepannek, 2018) in R.

a narrow range indicating that the partition is reproducible under small perturbations of the data and zero meaning random labelling. We also include the minimum and maximum ARI from the bootstrapping results.

Sensitivity analysis

Following selection of $k = 4$ for each gender in the main typology, a series of sensitivity analyses are conducted to assess whether the main typology is robust to alternative modelling and sample-definition choices while holding the FAMD projection space fixed.

Sensitivity 1 (gender-stratification logic check): We compare the baseline gender-stratified solution with a pooled clustering model estimated on the young cohort that includes gender as a clustering variable. For comparability, k is fixed at 4 in the pooled model. Agreement between pooled and gender-stratified solutions is assessed within each gender using centroid alignment and the ARI, supplemented by group-level diagnostics (e.g., purity).

Sensitivity 2 (cohort-gender stratification within the older cohorts): We repeat gender-stratified clustering within alternative childhood observability strata within the entire sample (full/partial/no coverage), holding k fixed at 4, and quantify agreement with the baseline solution using ARI after aligning clusters by centroid similarity.

Sensitivity 3 (time-window stability within the young cohort): We re-estimate gender-stratified clustering within three index-date windows defined by latest DFV finalisation year (2017-2018; 2019-2020; 2021-2022), holding k fixed at 4, then assess stability using ARI within gender after alignment to the baseline solution.

Together, these sensitivity analyses assess whether the baseline gender-stratified four-group structure is preserved under alternative grouping choices and whether any observed differences reflect substantive changes or shifts in cohort composition and data coverage.

RESULTS

In this section, we first summarise information from DCJ's linked administrative databases using principal components that capture major dimensions of offending, victimisation and service-system contact for the estimation sample. We then classify perpetrators into distinct groups based on these components and describe the demographic, system-contact and outcome profiles of each group. The results focus on the main gender-stratified typology for the young cohort, with additional robustness tests reported separately as sensitivity analyses.

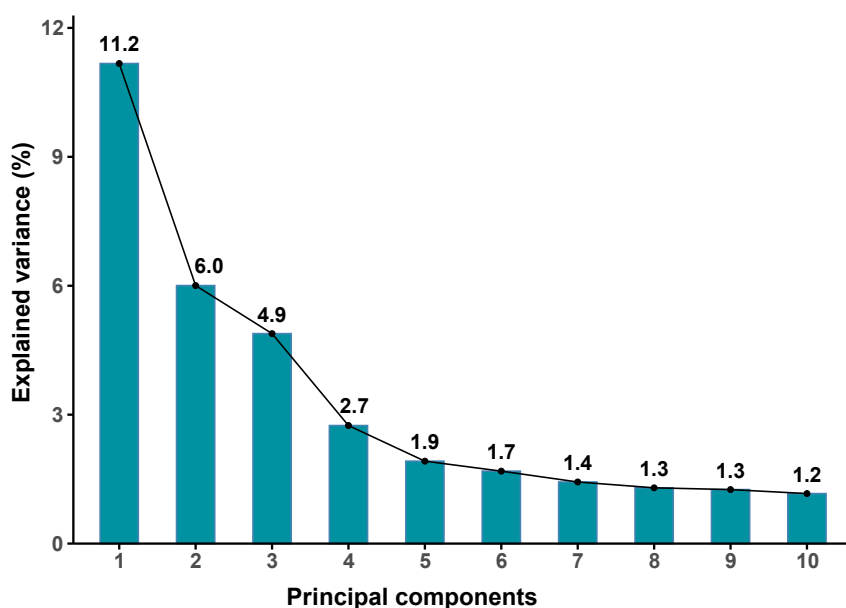
Principal Components summarising information from DCJ's linked administrative dataset

Before classifying individuals into groups using the K-prototype clustering approach described earlier, we first reduce the number of variables using FAMD. This step allows us to summarise the information contained within the 273 variables in our estimation data into a smaller number of uncorrelated components. To determine how many PCs to retain in our analysis, it is instructive to consider the fraction of the variance across the entire estimation sample that each PC can explain. In other words, the usefulness of each PC in summarising the information contained within the entire dataset.

Figure 3 presents a "scree plot" showing the proportion of the estimation dataset's variation that can be explained by each of the first ten PCs. From Figure 3, we can see that the first PC accounts for approximately 11.2 per cent of the entire dataset's variation, followed by 6.0 per cent for the second PC and 4.9 per cent for the third. The fourth PC explains 2.7 per cent of the variation. After the fourth PC, the explained variance drops sharply, with the fifth and sixth PCs contributing 1.9 per cent and 1.7

per cent, respectively, and all subsequent components contributing to less than 1.4 per cent each. We therefore acknowledge an elbow at the fourth PC, but retain six PCs to balance dimension reduction with information preservation: including PCs 5–6 captures additional, interpretable structure (>1.5 per cent each) that improves downstream clustering separation and stability without materially increasing model complexity. Taken together, the first six PCs capture about one-third of the information contained across all the major DCJ databases. For these reasons, we retain six PCs in the final clustering model, as this choice balances the trade-off between dimension reduction (i.e., the preference for fewer PCs) and variance retention (i.e., the desire to capture as much information as possible with the PCs).

Figure 2. Proportion of variance in the estimation dataset (273 constructed variables) explained by each of the top ten principal component



To facilitate our understanding of the six PCs, we need to learn more about the variables contributing most to each PC. Figure 3 reports the top ten variables with the highest contribution for each PC.¹⁸ A brief description of these variables is provided below. Note that all PCs are constructed as a linear combination of the same 273 variables, but the importance or “weight” of each variable differs among PCs. As a result, individual variables may contribute to more than one PC, albeit with different weights and alongside different sets of correlated variables. For example, charges with diversionary penalty contributes to both PC3 and PC5.¹⁹ Accordingly, the presence of a variable in multiple PCs reflects its role in capturing distinct underlying dimensions of variation, rather than duplication of measures.

- PC1 captures lifetime intensity of crime perpetration. As shown in the top-left panel of Figure 3, this component is driven by criminal justice indicators, including the number of times charged with any offence, findings of guilt in court, prior finalised court appearances, custodial episodes, and repeated bail interactions. Higher values on PC1 reflect persistent and non-specific offending trajectories.
- PC2 captures DFV and violent victimisation history. The top-right panel of Figure 3 shows that this component is defined by victimisation-related variables, including victimisation incidents at residential premises, DFV and sexual victimisation events, violent crime victimisation, breach-AVO victimisation, and repeated victimisation incidents in recent years. Higher values on PC2 indicate extensive recorded histories of DFV and violent victimisation.

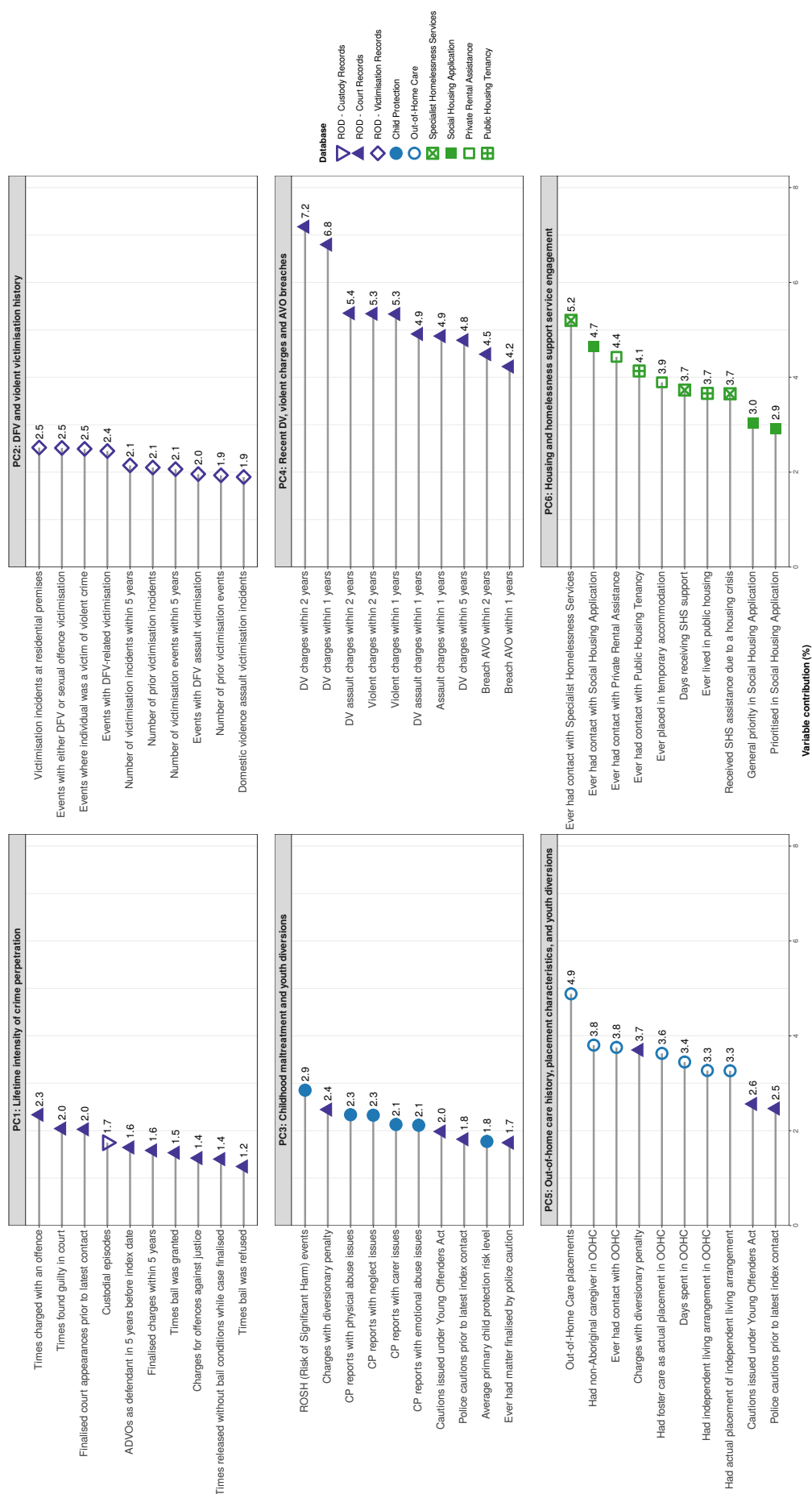
¹⁸ In the context of PC analysis, ‘contribution’ is defined as the fraction of each PC’s total variance that can be explained by a given variable.

¹⁹ Charges with diversionary penalty refers to a non-court outcome in ROD capturing diversionary justice responses. This measure is not recorded where the jurisdiction is youth conference, police caution, cannabis caution, or police warning.

- PC3 reflects childhood maltreatment and youth diversions. As illustrated in the middle-left panel, this component is driven by child protection indicators such as ROSH events, ROSH reports involving physical, emotional and neglect abuse, and average child protection risk levels. It also includes police and court-based diversionary responses under the *Young Offenders Act 1997* (NSW). PC3 therefore captures early-life exposure to maltreatment alongside early contact with the criminal justice system.
- PC4 summarises recent DFV, violent charges and AVO breaches. The middle-right panel of Figure 3 shows that this component is dominated by time-proximal offending indicators, including DFV charges and assaults, violent charges, and breaches of AVOs occurring within one to two years of an individual's most recent court finalisation date before 2023. Higher values on PC4 indicate recent DFV-related criminal justice involvement.
- PC5 represents OOHC history, placement characteristics, and youth diversions. The bottom-left panel shows that this component is driven by OOHC-related variables, including OOHC placements, placement type (such as foster care and independent living), duration spent in care, caregiver characteristics (such as whether the caregiver was non-Aboriginal), and indicators of contact with the criminal justice system at an early age. PC5 therefore reflects more contact with the OOHC and criminal justice systems as a child or young person.
- PC6 captures engagement with housing and homelessness support services. As shown in the bottom-right panel, this component is defined by contact with Specialist Homelessness Services, social housing applications, public housing tenancy, private rental assistance, temporary accommodation, and indicators of housing crisis and prioritisation. Higher values on PC6 indicate greater engagement with housing and homelessness support services and higher levels of system contact within housing assistance pathways. These measures likely reflect utilisation of, and interaction with, housing support systems rather than direct indicators of housing instability.

Overall, these six components provide a multidimensional summary of individuals' interactions with justice, child welfare and housing systems, capturing lifetime offending, victimisation histories, child protection and youth justice system involvement, recent DFV activity, and engagement with housing and homelessness support services. These components form the basis for the clustering analysis reported in the following subsection.

Figure 3. Top ten contributing variables for each of the first six principal components



Classifying perpetrators into groups

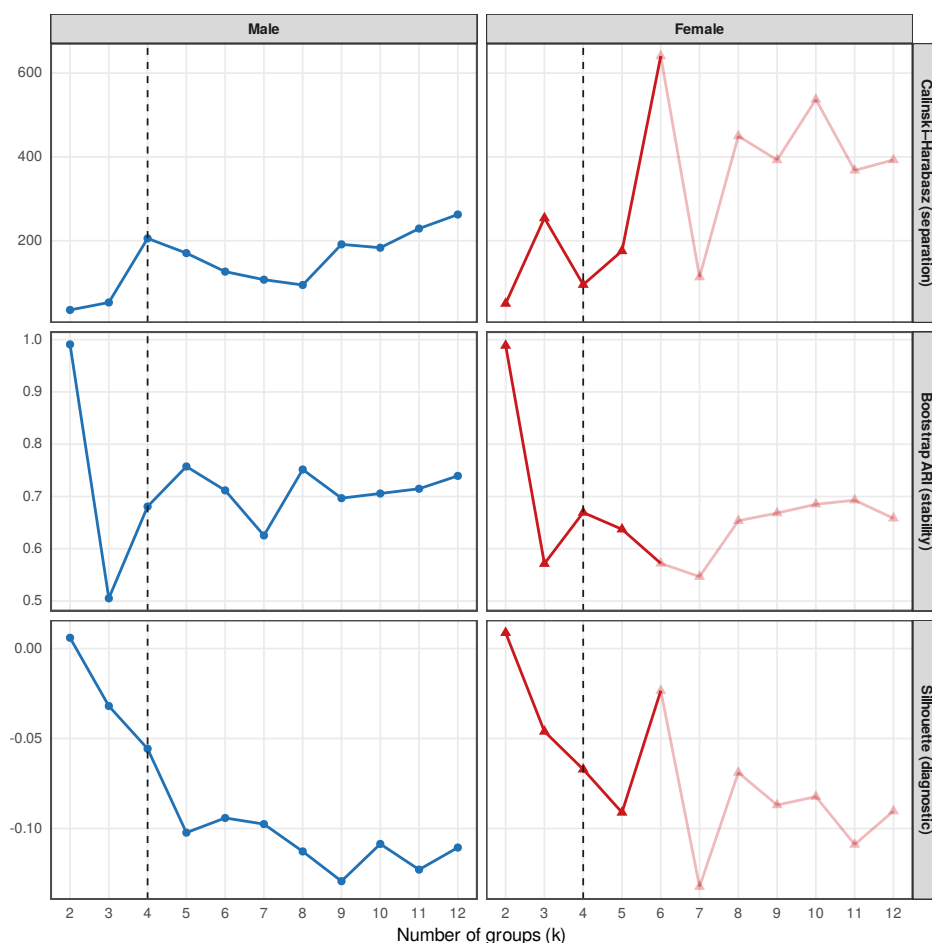
Using the six principal components described above, we classify DFV perpetrators into distinct groups based on shared patterns of offending, victimisation and service-system contact. The main results focus on the young cohort with complete childhood observability (born from 1987 onwards; $n = 71,896$) and clustering is conducted separately for males and females. This yields eight groups in total: four male groups (M1-M4) and four female groups (F1-F4).

The K-prototype clustering approach used in this paper results in many possible groups to choose from. That is, the model generates a variety of options regarding the number of groups we can classify individuals into, and the onus is then on the researcher to determine how many groups is optimal for our purposes. This is analogous to determining the number of PCs to utilise. If we select too many groups, then it may not be feasible for policymakers to develop interventions tailored to each group. If, on the other hand, we select too few groups, then we risk not capturing the very differences between groups that are important for decision-making. We take an empirical approach to guide our decision on this matter.

To determine the number of groups to select, we examine the clustering diagnostics shown in Figure 4, estimated separately for males and females in the young cohort. Figure 4 reports three complementary metrics across $k = 2-12$: the CH Index, which measures between-cluster separation; the mean bootstrap ARI, which measures the reproducibility of cluster assignments under resampling; and the average silhouette coefficient, which summarises within-cluster cohesion. In addition, Figure 4 and Appendix Table A1 report the bootstrap collapse rate at each k , defined as the proportion of bootstrap resamples in which the clustering algorithm fails to return k distinct, non-empty groups. A non-zero collapse rate indicates that the clustering solution is occasionally unstable, with one or more groups disappearing under small perturbations of the data.

For males, separation improves markedly at $k = 4$ (CH index = 205.6), while stability remains high across all values of k , with mean ARI around 0.7 and no bootstrap collapses observed. For females, higher values of k can yield stronger separation (e.g., CH index = 640.8 at $k = 6$), but bootstrap diagnostics indicate occasional instability at higher cluster counts, with a two per cent collapse rate observed for k value above 5 (presented as transparent points/lines). Across both genders, $k = 4$ provides a stable and interpretable solution, combining strong separation, consistent bootstrap agreement (mean ARI of 0.7), and zero collapse rates for both males and females. On this basis, four groups are retained for each gender in the main analysis, with higher- k solutions treated as diagnostically informative but less robust. These metric values, as well as the minimum and maximum ARI (among 100 Bootstrap repetitions) and collapse rates, are presented in the Appendix Table A1.

Figure 4. Clustering diagnostics for selecting the number of groups (k), by gender (young cohort)



A typology of DFV perpetrators in NSW

Table 6 summarises the characteristics of the eight gender-stratified groups identified in the young birth cohort (four male groups: M1-M4; four female groups: F1-F4). Results are presented across three panels covering demographic characteristics (Panel A), principal-component profiles (Panel B), and reoffending and revictimisation outcomes (Panel C). As noted at the outset, these groups should be interpreted as data-driven administrative profiles reflecting patterns of socioeconomic disadvantage, criminal justice involvement, child protection system contact and victimisation, rather than behavioural or relational “types” of DFV.

Across both genders, the proportion of individuals aged under 18 at first DFV court appearance is relatively similar among males, but shows greater variation and higher levels among females. Among males, the proportion ranges from 17.4% to 20.6%. By comparison, higher and more variable levels of early onset are observed among females, ranging from 21.0% to 28.7%, with Group F3 (28.7%) exhibiting the highest proportion. Socioeconomic context and Aboriginal representation also differ substantially across groups, both within and between genders. Males in Group M1 reside in more socioeconomically advantaged areas (mean SEIFA percentile = 79.6), while males in Groups M2 and M3 are concentrated in more disadvantaged areas (11.2 and 32.4). Aboriginal representation is lowest in Group M1 (5.3%) and highest in Group M3 (27.2%). A similar pattern is observed among females. Females in Group F1 live in more advantaged areas (SEIFA = 80.2), while females in Group F2 and F3 are located in more disadvantaged areas (11.4 and 32.2). Aboriginal representation is highest in Group F3 (32.3%) and lowest in F1 (6.9%).

Table 5. Comparison of summary characteristics for gender-stratified groups of DFV perpetrators born after 1987

Group index	M1	M2	M3	M4	F1	F2	F3	F4
Panel A. Demographics								
Male (%)	100	100	100	100	0	0	0	0
Female (%)	0	0	0	0	100	100	100	100
Aboriginal (%)	5.3	21.4	27.2	17.1	6.9	26.7	32.3	20.1
<18 years at first DFV court appearance (%)	19.6	17.4	20.6	19.3	27.2	21.0	28.7	25.7
SEIFA percentile (mean)	79.6	11.2	32.4	52.1	80.2	11.4	32.2	51.9
Panel B. Principal components score (relative scale 0-100 within each PC)								
PC1: Lifetime intensity of crime perpetration	0.0	31.2	63.4	36.0	13.8	47.0	100.0	60.8
PC2: DFV and violent victimisation history	4.0	0.0	0.3	3.3	47.4	71.3	100.0	77.3
PC3: Childhood maltreatment and youth diversions	59.2	65.6	100.0	79.4	49.5	0.0	50.9	34.4
PC4: Recent DV, violent charges and AVO breaches	68.8	100.0	81.3	78.1	37.5	12.5	3.1	0.0
PC5: Out-of-home care history, placement characteristics, and youth diversions	100.0	66.9	55.1	52.0	90.6	37.8	0.0	29.1
PC6: Housing and homelessness support service engagement	100.0	62.8	51.0	61.8	87.3	0.0	34.3	41.2
Panel C. 12-month reoffending and revictimisation								
DFV reoffending (%)	8.1	11.2	12.7	11.3	6.7	8.2	11.6	9.5
General reoffending (%)	20.4	29.3	35.2	29.0	13.6	17.8	24.8	20.6
DFV revictimisation (%)	3.0	4.4	5.1	4.6	10.8	17.2	20.3	16.6
General revictimisation (%)	12.0	14.7	18.0	16.0	20.3	29.1	35.3	30.6

Note. This table reports the characteristics of the eight groups of DFV perpetrators. Panel A reports demographic characteristics of these perpetrators. Panel B reports each group's principal component scores rescaled to range from 0 to 100 within each component, where 0 represents the lowest observed group mean and 100 the highest for that PC. Panel C reports the share of individuals within each group that have at least one DFV related reoffence and at least one DFV related revictimisation within 12 months. Continuous variables are reported as an average; categorical variables as proportion of the group.

In Panel B, we present information on the PC values for each of the eight groups. For ease of interpretation, we have rescaled the mean value of each PC to a relative rank ranging from 0 to 100 within that PC, where 0 denotes the lowest observed group mean and 100 the highest. These values indicate the extent to which each group is represented by the corresponding PC relative to the other seven groups. Thus, higher values indicate stronger representation on that PC dimension, and lower values indicate weaker representation. That is, lower values (i.e., values closer to zero) indicate that, on average, relative to individuals in another group, those in a given group are less represented by the corresponding PC, and higher values (i.e., values closer to 100) indicate that individuals within that group are more represented by the corresponding PC. For example, from Table 5 we can see that individuals in Group M1 have a PC1 rank of 0, while individuals in Groups M3 and F3 have PC1 ranks of 63.4 and 100.0, respectively. This indicates that individuals in Group M1 have comparatively low levels of lifetime involvement in the criminal justice system as a defendant, while Groups M3 (63.4) and F3 (100.0) have substantially higher levels of exposure to the criminal justice system.

The scaled values in Panel B are derived from the underlying raw principal component scores, which are centred around zero. A value of zero represents the sample average on that PC dimension, with positive values indicate above-average representation on a given dimension, while negative values indicate below-average representation. The raw PC scores are available in Table A2 of Appendix A. For the same PC1 example as above, in Table A2, Group M1 has an average PC1 score of -1.53, indicating comparatively lower than average offending intensity, whereas Groups M3 and F3 have scores of 2.01 and 4.05, indicating substantially higher levels of offending intensity.

Panel B of Table 5 shows that group differentiation among males is driven primarily by lifetime criminal justice involvement and recent DFV-related offending. Lifetime offending intensity (PC1) is lowest in Group M1 (0.0) and highest in M3 (63.4). Recent DFV, violent charges and AVO breaches (PC4) are most pronounced in Groups M2 (100.0), M3 (81.3) and M4 (78.1), indicating elevated recent justice contact for these groups. Child protection and youth justice diversions (PC3 and PC5) is substantial across all male groups but is highest in Groups M1 and M2, while housing and homelessness service engagement (PC6) is particularly pronounced in Group M1 (100.0).

Among females, group differentiation is dominated by lifetime criminal justice involvement, DFV and violent victimisation history, child protection and youth justice diversions, and housing and homelessness service engagement. DFV and violent victimisation history (PC2) is highest for Group F3 (100.0) but is also elevated for Groups F2 (71.3) and F4 (77.3), while housing assistance and support engagement (PC6) and OOHC history is most pronounced for Group F1. In contrast to males, recent DFV-related offending (PC4) is comparatively low across all female groups, particularly for Groups F3 (3.1) and F4 (0.0). However, Groups F3 and F4 are distinguished by their greater lifetime involvement with the criminal justice system as indicated by their higher scores on PC1.

Outcome patterns reported in Panel C align closely with the underlying PC profiles. Among males, DFV reoffending rates are higher for groups with increased lifetime and recent offending intensity (i.e., Groups M2, M3 and M4), with Group M3 showing the highest rate (12.7%) and Groups M2 and M4 showing similarly elevated rates (11.2% and 11.3%, respectively). General reoffending rates are higher than DFV reoffending rates across all groups, but again follow a similar pattern, being the highest for males in Group M3 (35.2%). The rate of revictimisation is much more pronounced for females than it is for males. The DFV revictimisation rate is particularly high for Group F3 (20.3%). This finding is consistent with their higher value on PC2 which reflects prior DFV victimisation recorded in police data. It is worth noting that general revictimisation rates are also high in Group F3 (35.3%). Female DFV reoffending rates are lower overall (relative to males) but remain elevated in Groups F3 (11.6%) and F4 (9.5%).

Based on these observations, Table 6 summarises the key characteristics of individuals within the eight groups, including a brief descriptive profile, a short label, group size, and the percentage share of the total estimation sample (n = 71,896). These group summaries reflect relative patterns across the PCs, which capture weighted combinations of 273 variables, rather than absolute differences on any single underlying measure.

Table 6. Descriptive summary of the eight DFV perpetrator groups, including short labels, group descriptions, group size and percentage of the total sample

	Group short label	Group summary description	Group size	% of total sample
M1	Low level offenders from more advantaged areas	Males residing in socioeconomically advantaged areas who have very low lifetime offending, a distinct pattern of childhood OOHC and youth diversions, high housing and homelessness service engagement and lower rates of reoffending	8,236	11.5
M2	Moderate offenders with recent high DFV charge volume from highly disadvantaged areas	Males residing in highly disadvantaged areas who have moderate lifetime offending, the highest recent DFV-related offending and elevated housing and homelessness service engagement	14,513	20.2
M3	Chronic high-rate general and DFV offenders	Males residing in disadvantaged areas who have the highest lifetime offending, elevated child protection reports and youth diversions, moderate housing and homelessness service engagement and the highest risk of DFV and general reoffending	14,744	20.5
M4	Moderate offenders with recent DFV charges	Males residing in mid-socioeconomic areas who have moderate lifetime offending, recent DFV-related offending and elevated child protection reports and youth diversions	15,669	21.8
F1	Low level DFV victim-offenders from more advantaged areas	Females residing in socioeconomically advantaged areas who have low lifetime offending, recent DFV charges, moderate DFV and violent victimisation histories, elevated OOHC involvement and high housing and homelessness service engagement	2,898	4.0
F2	Low-moderate offenders with high DFV victimisation risk from highly disadvantaged areas	Females residing in highly disadvantaged areas who have moderate lifetime offending, high DFV and violent victimisation history, lower child protection reports and youth diversions, low housing and homelessness service engagement and elevated revictimisation risk	5,040	7.0
F3	Early-onset high-risk victim-offenders	Females residing in highly disadvantaged areas who have high lifetime offending, early DFV court contact, the most extensive DFV and violent victimisation histories, moderate child protection reports and youth diversions, and the highest revictimisation risk	5,212	7.2
F4	Moderate general offenders with high victimisation risk	Females residing in mid-socioeconomic areas who have elevated lifetime offending, early DFV court contact, high DFV and violent victimisation history, moderate housing and homelessness service engagement and elevated revictimisation risk	5,584	7.8

Although the clustering is conducted separately for males and females, several structural similarities are evident across the resulting groups. For both genders, higher levels of socioeconomic disadvantage and Aboriginal representation are associated with greater multi-system contact. Criminal justice and child protection involvement and housing and homelessness service engagement also feature prominently across multiple groups. At the same time, the dominant dimensions of differentiation vary by gender: male groups are primarily distinguished by recent criminal justice involvement for DFV, whereas female groups are more strongly differentiated by victimisation histories, child protection system involvement and housing and homelessness service engagement. The typology also highlights meaningful variation in policy-relevant risk profiles. Groups characterised by elevated lifetime and recent DFV-related offending exhibit higher rates of DFV reoffending (e.g., Group M3), while groups characterised by extensive victimisation histories and moderate housing assistance engagement, particularly among females, exhibit higher rates of DFV revictimisation (e.g., Group F4).

Among males, Group M3 appears to represent the highest-risk profile, given its substantially higher lifetime offending intensity (PC1), elevated recent DFV-related offending (PC4), and the higher DFV reoffending rates (12.7%). Groups M2 and M4 can be considered a second tier of risk, with similar elevated DFV reoffending rates (11.2% and 11.3%, respectively), but different underlying profiles. Group M4 combines moderate lifetime offending with elevated recent DFV-related activity and elevated child protection system involvement, suggesting a more generalised pattern of justice and system contact. By contrast, Group M2 is characterised by lower lifetime offending intensity but the highest recent DFV-related offending and marked socioeconomic disadvantage, suggesting a more concentrated recent DFV risk. By comparison, Group M1 has comparatively low lifetime offending intensity, despite elevated OOHC, youth diversion and housing system contact, indicating a distinct profile of childhood and service involvement rather than high offending risk.

Among females, the most policy-relevant distinctions are driven by differences in DFV and violent victimisation history and lifetime criminal justice involvement, rather than by recent DFV-related offending. Groups F3 and F4 show the highest levels of DFV and violent victimisation (PC2) and the highest DFV revictimisation rates (20.3% and 16.6%), alongside substantial socioeconomic disadvantage, early DFV court contact and, in the case of F4, moderate housing and homelessness service engagement. These groups therefore represent cohorts with heightened revictimisation risk and repeated system contact as victim-survivors. At the same time, both groups also show elevated lifetime criminal justice involvement, with PC1 ranks of 100.0 for F3 and 60.8 for F4. Group F3 is particularly notable, having the highest lifetime offending intensity, the highest victimisation history, and the highest DFV and general revictimisation rates. In contrast, Individuals in Group F1 are from socioeconomically advantaged areas but have elevated OOHC involvement and housing and homelessness service engagement, while F2 combines high disadvantage and victimisation with comparatively lower housing assistance engagement and moderate reoffending.

Sensitivity analysis

A series of sensitivity analyses were conducted to assess the robustness of the typology to alternative modelling and sample definition choices. Robustness is evaluated using two complementary measures of agreement between clustering solutions: the Adjusted Rand index (ARI) and cluster purity. ARI provides a summary measure of overall agreement between two clustering solutions, while cluster purity offers a group-level diagnostic indicating how consistently individuals within a baseline group are assigned to the same cluster under an alternative specification. Together, these measures allow assessment of both global stability and group-specific sensitivity. This framework is applied consistently across three sensitivity analyses: (1) gender-stratified versus pooled clustering, (2) childhood observability and cohort coverage, and (3) temporal stability across DFV index windows. Detailed purity measures are reported in Appendix B.

Gender-stratified versus pooled clustering

The main typology is estimated separately by gender to account for differences in offending and victimisation patterns. A pooled clustering specification was estimated to assess the impact of this modelling choice. The results indicate that pooling reduces cluster separation and alters group composition, particularly among female groups. Agreement between pooled and gender-stratified solutions is moderate for both males and females (ARI = 0.51), indicating that gender-stratified clustering better captures meaningful variation in administrative profiles. Additional diagnostics and comparisons are reported in Appendix B1.

Childhood observability and cohort coverage

Three birth cohorts are defined based on the availability of Child Protection and Out-of-Home Care data: Cohort A (pre-1969; no coverage), Cohort B (1969-1987; partial coverage), and Cohort C (post-1987; complete coverage, baseline sample). To assess generalisability beyond the baseline cohort, individuals from Cohorts A and B are assigned to the baseline clustering solution estimated on Cohort C. Across cohorts, group positions in demographic space remain broadly consistent with those observed in the baseline cohort, indicating that the baseline groups map to similar demographic profiles when applied to older cohorts. Differences across cohorts are primarily reflected in the distribution of individuals across groups rather than changes in group location, consistent with differences in age composition and the completeness of childhood data coverage.

Consistency measures indicate moderate to high agreement across cohorts. Agreement is strongest for Cohort C, with perfect correspondence for both males and females (ARI = 1.00). For Cohort A (pre-1969), agreement remains high for both males (ARI = 0.83) and females (ARI = 0.78). For Cohort B (1969-1987), agreement is more variable, with moderate agreement for males (ARI = 0.50) and high agreement for females (ARI = 0.87). Detailed cohort-level diagnostics, including purity measures, are provided in Appendix B2.

Temporal stability across DFV index windows

Temporal stability is assessed by re-estimating clustering across DFV index windows and comparing group assignments with the baseline typology. ARI values indicate that the core structure of the typology is preserved across time periods, although stability varies by gender and time window. For males, ARI values range from 0.52 in 2017-2018 (W1) to 0.98 in 2019-2020 (W2) and 0.95 in 2021-2022 (W3). For females, ARI values are 0.95 in W1 and 0.53 and 0.55 in W2 and W3, respectively.

Across both genders, these results indicate that group profiles remain recognisable over time, although variation occurs in the distribution of individuals across groups and in boundary sharpness. These differences are consistent with changes in cohort composition and system contact over time rather than structural changes in the underlying typology. Full results, including purity measures, are reported in Appendix B3.

Taken together, these sensitivity analyses indicate that the eight-group, gender-stratified typology is robust to alternative clustering specifications, differences in cohort composition, and variation across time periods. While pooled models and refits within older cohorts and shorter index windows recover broadly similar structural patterns, agreement is consistently strongest when clustering is conducted separately by gender within cohorts with complete childhood observability. Gender stratification therefore provides clearer differentiation of offending and victimisation profiles and supports more interpretable administrative grouping, and is retained as the primary specification for the main analysis.

DISCUSSION

This study developed a data-driven typology of DFV perpetrators in NSW using linked administrative data spanning criminal justice, child protection, OOHC, housing and homelessness systems. By applying a two-step approach, dimension reduction using a single global factor structure followed by gender-stratified clustering, the analysis identified eight distinct profiles for a cohort of DFV perpetrators born after 1987 (four male groups: Groups M1-M4; four female groups: Groups F1-F4).

The eight groups identified in this study differ meaningfully in terms of their socioeconomic characteristics, Aboriginal representation, criminal justice, child protection and OOHC system involvement, housing and homelessness service engagement, and DFV-related outcomes. Individuals in Groups M2 and F2 reside in areas of high socio-economic disadvantage, while individuals in Groups M1 and F1 (the groups which are the smallest in size) reside in less socio-economically disadvantaged areas. Aboriginal representation is highest in Groups F3 (32.3%) and M3 (27.2%) and lowest in Groups M1 (5.3%) and F1 (6.9%). Groups are also differentiated based on their lifetime criminal justice involvement, with Groups M3 and F3 having substantially higher levels of prior exposure to the criminal justice system as a defendant compared with other groups. DFV victimisation history is much more pronounced for female groups than it is for male groups, while recency of DFV charges and AVO breaches differentiate the male groups. Early onset offending is more prominent for female groups (particularly Groups F1, F3 and F4) compared with male groups. These patterns indicate that DFV perpetrators in contact with the NSW criminal justice system are a diverse population, suggesting the need for tailored approaches to service delivery.

Differences in reoffending and revictimisation rates reinforce the practical value of the typology for system-level planning. Among males, DFV reoffending is highest in Groups M3 (12.7%), with Groups M2 (11.2%) and M4 (11.3%) showing similarly elevated rates. These groups also have relatively high general reoffending rates (e.g., 35.2% and 29.0% respectively). DFV revictimisation rates are much higher across all female groups compared with males, with the highest DFV revictimisation rates observed for Groups F3 (20.3%) and F2 (17.2%). Notably, Group F3 also has a comparatively high rate of general revictimisation (35.3%).

Sensitivity analyses indicate that the typology is reasonably robust and interpretable for planning purposes, but some caution is warranted. Pooling genders produces only moderate agreement with the gender-stratified solution (ARI = 0.51 for both males and females), supporting the use of gender stratification to preserve within-gender heterogeneity. Cohort-stratified refits show high agreement in the oldest cohort (pre-1969: ARI = 0.83 males; 0.78 females), but lower agreement for males in the partially observable cohort (1969-1987: ARI = 0.50). Temporal refits across DFV index windows also show that agreement can vary by gender and period (e.g., males: ARI = 0.52 in W1 and 0.95-0.98 in W2-W3; females: ARI = 0.95 in W1 and around 0.53-0.55 in W2-W3). Together, these results suggest that the core administrative profiles persist, but that group boundaries and relative distribution across groups can shift across cohorts and time windows. This is consistent with using the typology as a descriptive planning tool rather than as a forecasting or individual-level intervention-matching instrument.

By generating a typology of DFV perpetrators across NSW, this paper provides information that can be used to tailor DFV interventions toward specific groups of people. In the gender-stratified typology, the most policy-relevant contrasts differ by gender. Among males, groups with elevated lifetime and recent DFV-related offending show higher DFV and general reoffending rates (for example, M3 and M4), pointing to the importance of court-based responses, supervision and compliance-focused interventions. Among females, differences are driven more by victimisation histories and housing and homelessness service engagement than by recent DFV-related offending (for example, F3 and F4 exhibit the highest DFV revictimisation rates), underscoring the need for integrated responses spanning justice, housing and victim-support services. Appendix Table A3 on geographic distribution (e.g., major city versus regional residence) indicate that some groups are more concentrated in metropolitan areas while others are more regionally dispersed. These geographic differences are important from a program-design perspective as social and cultural norms are known to vary across communities (Campo & Tayton, 2015), as does the capacity for program managers to deliver certain interventions in a cost-effective manner.

At the same time, it is important to recognise the limits of what typologies can tell us. Cluster membership cannot reveal the motivations underlying DFV behaviour, the presence or absence of coercive control, situational triggers, or the dynamics of intimate or family relationships. These dimensions require complementary assessment methods and cannot be inferred from administrative data alone. A further consideration when interpreting these groups, is the concern that women might be misidentified as the primary aggressor in DFV events (Barlow & Walklate, 2025; Boxall et al., 2020; Reeves, 2021). While our method reduces the influence of single-incident misclassification by drawing on patterns of repeated system contact across criminal justice, victimisation and child protection records, it cannot conclusively exclude misidentification. This limitation is therefore important to acknowledge when interpreting group composition and profiles.

While our results are largely consistent with prior research into the demographic and risk profile of DFV perpetrators, to date there has been no effort in NSW to classify individuals into distinct groups using integrated multi-system administrative datasets. Descriptions of DFV perpetrators have been developed in Tasmania (Boxall et al., 2015b), Victoria (Crime Statistics Agency Victoria, 2020), and more recently in Queensland using police data (Howard & Kebbell, 2026). These studies demonstrate that administrative justice data can be used to derive meaningful offender groupings, including groups with distinct patterns of offending and harm. However, they rely primarily on police or court records and focus on offence histories. In contrast, our typologies are based on criminal justice information as well as information relating to housing, homelessness and child protection. The inclusion of six components that explicitly separate lifetime justice exposure, recent DFV activity, victimisation histories, child protection involvement, youth diversions, and housing/homelessness contact demonstrates the value of multi-system linkage for understanding population-level heterogeneity.

In situating our work within existing typological literature, it is important to note that traditional frameworks, such as Johnson's (2001) distinction between "intimate terrorism" and "situational couple violence", and Holtzworth-Munroe and Stuart's (1994) classification of family-only, dysphoric/borderline, and generally violent/antisocial perpetrators, are grounded in behavioural patterns or psychological traits.

Our typology, by contrast, emerges from administrative system contacts and demographic patterns, and does not seek to infer underlying motivation or psychological profiles. Similarly, while prior research has identified specific typologies for adolescent-to-parent and intimate partner violence, our data-driven approach identifies groups that differ primarily in risk, socioeconomic status and patterns of service/system contact, rather than relationship dynamics. Our clusters therefore complement, rather than replicate, existing behavioural and psychological frameworks.

The present study is, of course, not without its caveats. The first is the incompleteness and uneven coverage of the administrative data used to classify individuals into groups. For example, child protection and OOHC records only date back to 1987, and the seven datasets begin in different years. As a result, we cannot observe all relevant system contacts for older cohorts, nor can we construct comparable longitudinal histories for the full cohort of 196,701 DFV perpetrators. This cross-sectional representation is well suited to describing patterns of system contact and relative risk, but it cannot capture developmental or etiological pathways or establish whether contacts with systems such as OOHC or housing typically precede or follow DFV-related offending. Questions of sequencing and progression are better addressed by a prospective longitudinal cohort study with consistent follow-up across systems.

A second caveat is that, although we used a data-driven approach to classify individuals into groups, the analysis necessarily involved choices that may have affected the resulting typology. One important set of choices concerns what information is excluded from the FAMD step. In this study, we excluded demographic information and DFV reoffending/revictimisation measures from the construction of the principal components so that these variables could be used to describe and compare groups after classification. Different inclusion/exclusion decisions would likely yield different principal components and therefore different clustering results. For example, including reoffending and revictimisation in the FAMD step would mechanically embed outcomes into the component space and could change group separation. Similarly, excluding child protection measures would likely shift the typology towards a structure that places less emphasis on early-life system involvement.

Relatedly, the choice of the number of principal components and the number of groups can materially affect the typology. We retained six principal components after inspecting the variance-explained scree pattern (with an elbow around PC4) and to improve clustering separation and stability. We then selected four groups within each gender using a grid search over candidate values of k , jointly considering internal separation and stability metrics (average silhouette coefficient, Calinski-Harabasz Index, and bootstrap stability using the Adjusted Rand Index), alongside a diagnostic check on bootstrap collapse rates. The grid-search results illustrate why these choices matter: alternative values of k can yield higher separation in some cases but can also introduce instability (for example, a non-zero bootstrap collapse rate for higher female k values), and different choices of k would change group boundaries and potentially split or merge profiles. For these reasons, the typology should be interpreted as one robust and practically useful representation of heterogeneity rather than the unique or only valid partition of the data. Policymakers should therefore use the typology to guide system planning, resource allocation and service pathway design, and avoid treating group membership as deterministic or as a substitute for individual-level assessment.

In summary, the findings from our study highlight the heterogeneity of the DFV perpetrator population in NSW and demonstrate the value of typology in supporting system-level planning. The eight gender-stratified groups capture differences in lifetime intensity of crime perpetration; DFV and violent victimisation history; childhood maltreatment and youth justice diversion; recent DV, violent charges and AVO breaches; OOHC history and placement characteristics; and housing and homelessness service engagement. This information can assist policymakers in identifying where particular cohorts are concentrated, where service demand may be highest, and which parts of the service system may require additional capacity or tailored delivery models. Continued collaboration between policymakers and researchers is essential, not to generate discrete treatment programs, but to ensure that improvements in data quality and system integration can enhance the sector's ability to plan, allocate resources, and coordinate responses effectively across diverse DFV perpetrator populations.

ACKNOWLEDGEMENTS

The author thanks the Women, Family and Community Safety (WFCS) team for their support and funding for this project. Special thanks are extended to Francy Bulic and Elaine Ortega from the LinDA data team for assistance with data extraction and preparation. The author is grateful to Brigitte Gilbert for early-stage insightful discussions. The author further acknowledges Steve Yeong and Suzanne Poynton for their supervision and guidance provided throughout the research process. The author is also grateful to the two anonymous peer reviewers for their insightful feedback, to Dominic Furneaux-Bate and Suzanne Poynton for proofreading, and to Florence Sin for desktop publishing.

REFERENCES

- Alexander, E. F., & Johnson, M. D. (2023). On categorizing intimate partner violence: A systematic review of exploratory clustering and classification studies. *Journal of Family Psychology*, 37(5), 743–752. <https://doi.org/10.1037/fam0001075>
- Australian Bureau of Statistics. (2011). Australian and New Zealand Standard Offence Classification (ANZSOC) (3rd ed., Cat. No. 1234.0). Retrieved 20 July 2025 from <https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-offence-classification-anzsoc/latest-release>
- Australian Bureau of Statistics. (2023). Socio-Economic Indexes for Areas (SEIFA), *Australia, 2021* (Cat. No. 2033.0.55.001). Retrieved 20 July 2025 from <https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/latest-release>
- Babcock, J. C., Miller, S. A., & Siard, C. (2003). Toward a typology of abusive women: Differences between partner-only and generally violent women in the use of violence. *Psychology of Women Quarterly*, 27(2), 153–161. <https://doi.org/10.1111/1471-6402.00095>
- Barlow, C., & Walklate, S. (2025). Who is the victim? Exploring the complexities of misidentification. *Criminology & Criminal Justice*, 17488958251355378. <https://doi.org/10.1177/17488958251355378>
- Beckwith, S., Lowe, L., Wall, L., Stevens, E., Carson, R., Kaspiw, R., MacDonald, J. B., McEwen, J., Willoughby, M., & Gahan, L. (2023). *Coercive control literature review*. Australian Institute of Family Studies. Retrieved 20 July 2025 from <https://aifs.gov.au/all-research/research-reports/coercive-control-literature-review>
- Boxall, H., Dowling, C., & Morgan, A. (2020) Female perpetrated domestic violence: Prevalence of self-defensive and retaliatory violence. *Trends & issues in crime and criminal justice* no. 584. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti04176>
- Boxall, H., Payne, J., & Rosevear, L. (2015a). Prior offending among family violence perpetrators: A Tasmanian sample (Trends & Issues in Crime and Criminal Justice No. 493). *Australian Institute of Criminology*. <https://doi.org/10.52922/ti168454>
- Boxall, H., Rosevear, L., & Payne, J. (2015b). Domestic violence typologies: What value to practice? *Trends & Issues in Crime and Criminal Justice*, (494), 1–9. <https://doi.org/10.52922/ti168448>
- Boxall, H., & Sabol, B. (2021). Adolescent family violence: Findings from a group-based analysis. *Journal of Family Violence*, 36, 787–797. <https://doi.org/10.1007/s10896-020-00191-1>
- Bureau of Crime Statistics and Research. (2026). *Reoffending*. Retrieved 19 February 2026 from <https://bocsar.nsw.gov.au/topic-areas/re-offending.html>
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>

- Campo, M., & Tayton, S. (2015). Domestic and family violence in regional, rural and remote communities. *Australian Institute of Family Studies*. Retrieved 20 July 2025 from <https://aifs.gov.au/resources/policy-and-practice-papers/domestic-and-family-violence-regional-rural-and-remote>
- Capaldi, D. M., & Kim, H. K. (2007). Typological approaches to violence in couples: A critique and alternative conceptual approach. *Clinical Psychology Review*, 27(3), 253–265. <https://doi.org/10.1016/j.cpr.2006.09.001>
- Crime Statistics Agency Victoria. (2020). Finding typologies of family violence perpetrators using police records. *Crime Statistics Agency*. Retrieved 20 July 2025 from <https://www.crimestatistics.vic.gov.au/research-and-evaluation/publications/finding-typologies-of-family-violence-perpetrators-using>
- Dowling C, Boxall H & Morgan A (2021). The criminal career trajectories of domestic violence offenders. *Trends & issues in crime and criminal justice* no. 624. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78016>
- Fowler, K. A., & Westen, D. (2011). Subtypes of partner violence perpetrators among male juvenile offenders. *Journal of Child Psychology and Psychiatry*, 52(3), 273–273. <https://doi.org/10.1111/j.1469-7610.2010.02300.x>
- Giesbrecht, C. J., Keown, L. A., & Bruer, K. C. (2025). A new typology of men who perpetrate intimate partner violence: Differentiating perpetrators based on criminal history and antisocial attitudes. *Criminal Justice and Behavior*. Advance online publication. <https://doi.org/10.1177/00938548251362812>
- Goldberg, A. E., & Blaauw, E. (2019). Parental substance use disorder and child abuse: Risk factors for child maltreatment? *Psychiatry, Psychology and Law*, 26(6), 911–924. <https://doi.org/10.1080/13218719.2019.1664276>
- Gondolf, E. W. (1997). Batterer programs: What we know and need to know. *Journal of Interpersonal Violence*, 12(1), 83–98. <https://doi.org/10.1177/088626097012001006>
- Hilton, N. Z., & Harris, G. T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence, & Abuse*, 6(1), 3–23. <https://doi.org/10.1177/1524838004272463>
- Holtzworth-Munroe, A., & Meehan, J. C. (2004). Typologies of men who are maritally violent: Scientific and clinical implications. *Journal of Interpersonal Violence*, 19(12), 1369–1389. <https://doi.org/10.1177/0886260504269693>
- Holtzworth-Munroe, A., & Stuart, G. L. (1994). Typologies of male batterers: Three subtypes and the differences among them. *Psychological Bulletin*, 116(3), 476–493. <https://doi.org/10.1037/0033-2909.116.3.476>
- Howard, R. A., & Kebbell, M. R. (2026). Examining Domestic and Family Violence Typologies Using Queensland Police Data: Differences in Risk of Harm. *Journal of Interpersonal Violence*, 0(0). <https://doi.org/10.1177/08862605261447033>
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3), 283–304. <https://doi.org/10.1023/A:1009769707641>
- Jacobson, N. S., & Gottman, J. M. (1998). When men batter women: New insights into ending abusive relationships. *Simon and Schuster*.
- Johnson, M. P. (2001). Conflict and control: Gender symmetry and asymmetry in domestic violence. *Violence Against Women*, 7(9), 1003–1018. <https://doi.org/10.1177/10778010122182839>
- Johnson, M. P. (2008). A typology of domestic violence: Intimate terrorism, violent resistance, and situational couple violence. *Northeastern University Press*.

- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and visualize the results of multivariate data analyses* (R package version 1.0.7). Retrieved 20 July 2025 from <https://CRAN.R-project.org/package=factoextra>
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1–18. <https://doi.org/10.18637/jss.v025.i01>
- Loiniz, I. (2014). Typologies, risk and recidivism in partner-violent men with the B-SAFER: A pilot study. *Psychology, Crime & Law*, 20(2), 183–198. <https://doi.org/10.1080/1068316X.2013.770854>
- Moulds, L. G., Mayshak, R., Mildred, H., Day, A., & Miller, P. (2019). Adolescent violence towards parents: A case of specialisation? *Youth Justice*, 19(3), 206–221. <https://doi.org/10.1177/1473225419880004>
- Reeves, E. (2021). 'I'm Not at All Protected and I Think Other Women Should Know That, That They're Not Protected Either': Victim–Survivors' Experiences of 'Misidentification' in Victoria's Family Violence System. *International Journal for Crime, Justice and Social Democracy*, 10(4), 39–51. <https://doi.org/10.5204/ijcjsd.1992>
- Robinson, A. L., & Clancy, A. (2021). Systematically identifying and prioritising domestic abuse perpetrators for targeted intervention. *Criminology & Criminal Justice*, 21(5), 687–704. <https://doi.org/10.1177/1748895820914380>
- 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. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saunders, D. G. (1996). Feminist-cognitive-behavioral and process-psychodynamic treatments for men who batter: Interaction of abuser traits and treatment models. *Violence and Victims*, 11(4), 393–414. <https://pubmed.ncbi.nlm.nih.gov/9210279/>
- Swan, S. C., & Snow, D. L. (2002). A typology of women's use of violence in intimate relationships. *Violence Against Women*, 8(3), 286–319. <https://doi.org/10.1177/10778010222183071>
- Szepannek, G. (2018). clustMixType: User-friendly clustering of mixed-type data in R. *Journal of Statistical Software*, 87(3), 1–30. <https://doi.org/10.32614/RJ-2018-048>
- Taillieu, T. L., Brownridge, D. A., Sareen, J., & Afifi, T. O. (2016). Childhood emotional maltreatment and mental disorders: Results from a nationally representative adult sample from the United States. *Child Abuse & Neglect*, 59, 1–12. <https://doi.org/10.1016/j.chiabu.2016.07.007>
- Trevena, J., & Poynton, S. (2016). *Does a prison sentence affect future domestic violence reoffending?* (Crime and Justice Bulletin No. 190). Bureau of Crime Statistics and Research. Retrieved 20 July 2025 from <https://bocsar.nsw.gov.au/documents/publications/cjb/cjb151-200/cjb190-report-does-a-prison-sentence-affect-future-domestic-violence-reoffending-2016.pdf>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Wells, L., Ellis, D., & Jaffe, P. (2016). Exploring typologies of domestic violence perpetrators: A latent class analysis of perpetrator risk factors. *Children and Youth Services Review*, 65, 70–77. <https://doi.org/10.1016/j.childyouth.2016.06.004>

APPENDIX

APPENDIX A. CLUSTERING METRICS AND SUMMARY STATISTICS

Table A1. Clustering metrics for grid search of k from 2 to 12, with the optimal k = 4 highlighted

Gender	k	Silhouette Coefficient	Calinski-Harabasz Index	Adjusted Rand Index (mean)	Adjusted Rand Index (min)	Adjusted Rand Index (max)	Bootstrap collapse rate ¹
Male	2	0.0	34.8	1.0	1.0	1.0	0%
	3	0.0	52.6	0.5	0.2	1.0	0%
	4	-0.1	205.6	0.7	0.4	1.0	0%
	5	-0.1	170.5	0.8	0.4	1.0	0%
	6	-0.1	126.3	0.7	0.4	1.0	0%
	7	-0.1	107.1	0.6	0.5	1.0	0%
	8	-0.1	94.3	0.8	0.4	1.0	0%
	9	-0.1	191.4	0.7	0.4	0.8	0%
	10	-0.1	183.3	0.7	0.5	1.0	0%
	11	-0.1	229.1	0.7	0.5	0.9	0%
	12	-0.1	262.6	0.7	0.6	0.9	0%
Female	2	0.0	50.0	1.0	1.0	1.0	0%
	3	0.0	254.2	0.6	0.2	1.0	0%
	4	-0.1	95.1	0.7	0.4	1.0	0%
	5	-0.1	175.6	0.6	0.4	1.0	0%
	6	0.0	640.8	0.6	0.4	1.0	2%
	7	-0.1	113.1	0.5	0.4	1.0	2%
	8	-0.1	449.8	0.7	0.5	1.0	2%
	9	-0.1	392.6	0.7	0.5	1.0	2%
	10	-0.1	537.3	0.7	0.4	0.9	2%
	11	-0.1	367.9	0.7	0.5	1.0	2%
	12	-0.1	392.9	0.7	0.5	0.9	2%

Note. ¹ Bootstrap collapse rates are based on 100 repeated bootstrap resamples. No collapses were observed for any male models (0% across all values of k). For females, a small number of collapses were observed when k > 5 (2% of bootstrap resamples), indicating occasional instability at higher cluster counts.

Table A2. Comparison of average and median raw PC scores for gender-stratified groups of DFV perpetrators born after 1987

Group index	M1	M2	M3	M4	F1	F2	F3	F4
<i>Panel D. Raw principal components scores (mean)</i>								
PC1: Lifetime intensity of crime perpetration	-1.53	0.21	2.01	0.48	-0.76	1.09	4.05	1.86
PC2: DFV and violent victimisation history	-0.01	-0.26	-0.24	-0.05	2.73	4.24	6.05	4.62
PC3: Childhood maltreatment and youth diversions	1.92	2.06	2.81	2.36	1.71	0.63	1.74	1.38
PC4: Recent DV, violent charges and AVO breaches	0.25	0.35	0.29	0.28	0.15	0.07	0.04	0.03
PC5: Out-of-home care history, placement characteristics, and youth diversions	0.97	0.55	0.4	0.36	0.85	0.18	-0.3	0.07
PC6: Housing and homelessness support service engagement	0.12	-0.26	-0.38	-0.27	-0.01	-0.9	-0.55	-0.48
<i>Panel E. Raw principal components scores (median)</i>								
PC1: Lifetime intensity of crime perpetration	-3.05	-1.24	0.46	-1.15	-2.69	-0.29	2.48	0.15
PC2: DFV and violent victimisation history	0.01	-0.09	-0.15	-0.05	1.49	3.06	4.47	3.23
PC3: Childhood maltreatment and youth diversions	0.68	0.81	1.29	0.88	0.48	0.27	0.47	0.4
PC4: Recent DV, violent charges and AVO breaches	-0.35	-0.35	-0.36	-0.33	-0.37	-0.46	-0.43	-0.41
PC5: Out-of-home care history, placement characteristics, and youth diversions	0.15	0.09	0.13	0.08	0.16	0.09	0.1	0.13
PC6: Housing and homelessness support service engagement	0.53	0.16	0	0.23	0.53	-0.49	-0.55	-0.18

Figure A1. Boxplots for the distributions of eight numerical characteristics across eight groups of DFV perpetrators

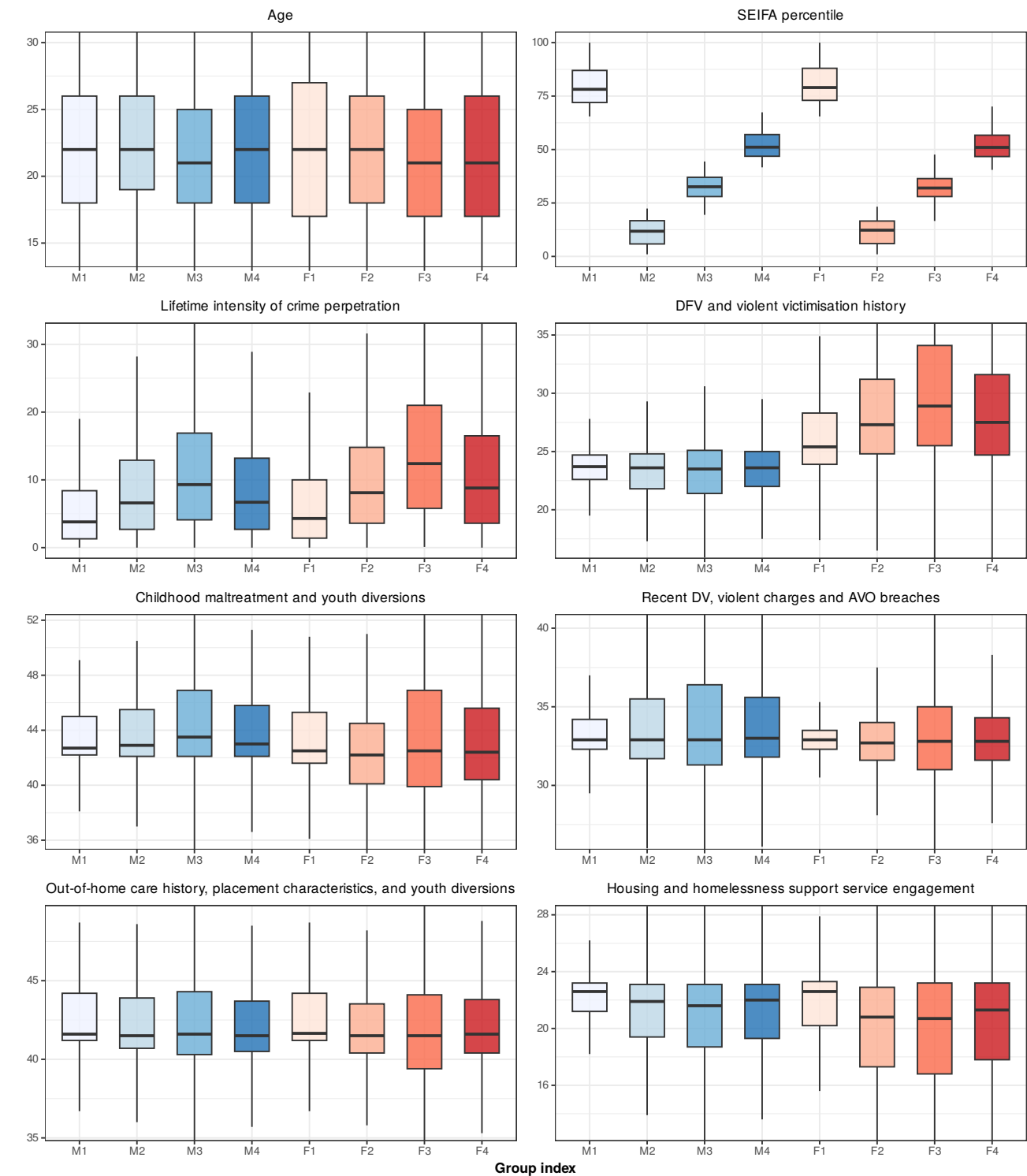


Table A3. Group-level summary statistics for a selected subset of variables, reported in their original measurement scale

Database	Variable name	M1	M2	M3	M4	F1	F2	F3	F4
ROD - Court Records	Age 10-13 (%)	2.2	1.9	2.4	2.4	2.7	2.3	4.6	4
ROD - Court Records	Age 14-17 (%)	17.4	15.5	18.2	16.9	24.5	18.7	24.1	21.7
ROD - Court Records	Age 18-24 (%)	44.8	49.6	49.5	47.6	37.7	45.9	43.6	42.9
ROD - Court Records	Age 25-29 (%)	24.1	23.8	21.7	22.9	20.9	22.8	19.2	21.1
ROD - Court Records	Age 30-35 (%)	11.5	9.1	8.2	10.1	14.1	10.3	8.4	10.3
ROD - Court Records	Number of times an individual was charged with an offence	5.5	5.6	5.5	5.4	11.1	11.2	11.2	11.4
ROD - Court Records	Number of times the person was found guilty in court	3.4	3.4	3.4	3.3	7.4	7.6	7.5	7.6
ROD - Court Records	Number of finalised court appearances prior to the latest index contact with the justice system	2.6	2.6	2.6	2.5	6.4	6.6	6.5	6.7
ROD - Court Records	Number of ADVOs where an individual was the defendant in the five years prior to each index date	1.8	1.8	1.7	1.7	4.7	4.8	4.9	5
ROD - Court Records	Number of times an individual was granted bail	1.5	1.5	1.5	1.5	2.8	2.8	2.8	2.9
ROD - Court Records	Number of times an individual was released without bail conditions while their case was finalised	1.8	1.9	1.8	1.8	3.6	3.7	3.7	3.7
ROD - Court Records	Number of times an individual was refused bail	0.3	0.3	0.3	0.3	1.1	1.2	1.1	1.2
ROD - Court Records	Number of charges for Offences Against Justice	0.6	0.6	0.6	0.6	1.6	1.6	1.6	1.7
ROD - Court Records	Number of times an individual received youth diversions or cannabis caution penalty	1	1	1	1	1.6	1.5	1.6	1.6
ROD - Court Records	Number of cautions issued under Young Offenders Act 1997 (NSW)	0.6	0.6	0.6	0.6	0.9	0.9	1	1
ROD - Court Records	Number of police cautions prior to the latest index contact with the justice system	0.5	0.5	0.5	0.5	0.9	0.9	0.9	0.9
ROD - Court Records	Number of times an individual received a juvenile control order	0.1	0.1	0.1	0.1	0.2	0.2	0.2	0.2
ROD - Court Records	Number of events where the individual was the victim of an incident and NSW Police charged the suspected offender	0.6	0.5	0.6	0.5	2.2	2	2.1	2
ROD - Court Records	Whether the individual has ever resided in an inner regional area at time of charge (%)	28.4	28.8	28.6	29.3	46.3	47.8	47.2	48.2
ROD - Court Records	Whether the individual has ever resided in a major city area at time of charge (%)	71.1	70.7	70.5	70.2	69.2	68.7	69.2	68.4
ROD - Court Records	Whether the individual has ever resided in an outer regional area at time of charge (%)	9.6	9.5	9.2	9.6	16.5	15.8	17.3	16
ROD - Court Records	Whether the individual has ever resided in a remote area at time of charge (%)	1.5	1.5	1.4	1.5	3.6	2.8	2.9	3.1
ROD - Court Records	Whether the individual has ever resided in a very remote area at time of charge (%)	0.9	1	0.9	0.9	1.6	1.9	1.8	1.7
ROD - Custody Records	Number of custodial episodes for each individual	1.5	1.5	1.4	1.4	4.6	4.7	4.7	4.9
ROD - Custody Records	Days in custody for each individual	113.5	113.6	110	108.4	337.2	360.5	334	355.3

Table A3. Group-level summary statistics for a selected subset of variables, reported in their original measurement scale (*continued*)

Database	Variable name	M1	M2	M3	M4	F1	F2	F3	F4
ROD - Victimisation Records	Number of events where an individual was the victim of at least one DFV related incident	0.9	0.9	1	0.9	3.9	3.6	3.6	3.5
ROD - Victimisation Records	Number of times the individual was a victim at a residential premises	0.9	0.9	0.9	0.8	3.5	3.3	3.3	3.2
ROD - Victimisation Records	Number of events an individual was the victim of a violent crime	0.7	0.7	0.7	0.7	2.4	2.3	2.3	2.2
ROD - Victimisation Records	Number of events where the individual was a victim in at least one DFV Assault incident	0.5	0.5	0.5	0.5	1.8	1.7	1.7	1.7
ROD - Victimisation Records	Number of events where the individual was a victim in at least one breach AVO incident	0.2	0.1	0.2	0.1	0.9	0.8	0.8	0.8
ROD - Victimisation Records	Number of incidents where the POI was an intimate partner of the victim	0.2	0.2	0.3	0.2	1.2	1.1	1.1	1.1
Child Protection	Total number of ROSH events	4.3	4.3	4.4	4.3	9.6	9.9	10	9.6
Child Protection	Number of child protection reports where the helpline assessed issue included neglect	1.8	1.9	1.9	1.9	4.6	4.5	4.7	4.5
Child Protection	Number of child protection reports where the helpline assessed issue included physical abuse	1.7	1.7	1.7	1.8	3.8	4.2	4.1	3.9
Child Protection	Number of child protection reports where the helpline assessed issue included carer issues	1.5	1.6	1.6	1.6	3.5	3.5	3.6	3.5
Child Protection	Number of child protection reports where the helpline assessed issue included sexual abuse	0.8	0.8	0.8	0.8	1.8	1.9	1.8	1.7
Child Protection	Number of child protection reports where the helpline assessed issue included child and young person issues	2	2	1.9	1.9	4.9	5.2	5	4.4
Child Protection	CP final risk level (very high) (%)	1	1.2	1.1	1.2	1.6	1.6	1.3	1.6
Child Protection	CP final risk level (high) (%)	3.4	3.3	3.6	3.5	3.7	3.9	4.2	3.9
Out-of-Home Care	Days an individual spent in Out-of-Home Care	167.3	177.3	160.1	164.6	459.0	462.7	465.1	490.4
Out-of-Home Care	Number of Out-of-Home Care placements for an individual	0.2	0.2	0.2	0.3	0.8	0.9	0.9	0.9
Out-of-Home Care	Whether an individual in contact with OOHC ever had an non-Aboriginal caregiver	8.3	8.1	7.7	8.2	24.6	24.2	24.5	24
Out-of-Home Care	Whether an individual had contact with Out-of-Home Care system	9	9.1	8.9	9.2	27.3	27.2	27.3	26.6
Out-of-Home Care	Whether the individual was ever actually placed in foster care during OOHC	5.7	5.9	5.5	5.8	18	18.5	18.5	18.1
Private Rental Assistance	Whether an individual was placed in temporary accommodation at least once	6.6	6.6	6.9	6.7	89.3	89	88.1	88.5
Public Housing Tenancy	Days an individual spent in public housing	673.1	708.4	692.7	686.7	1574	1625.3	1624.8	1579.5
Specialist Homelessness Services	Whether an individual ever received assistance from a SHS provider due to a housing crisis	5.4	5.4	5.6	5.5	35.8	35.8	37.6	37.1
Specialist Homelessness Services	Days an individual was in receipt of support from a Specialist Homelessness Service provider	31.2	33.3	31.7	32.1	206.4	203.2	211.1	215
Specialist Homelessness Services	Whether an individual had contact with Specialist Homelessness Services	16.7	17	16.6	17.1	67.1	67.5	67	67.4

Note. This table reports the characteristics of the five groups of DFV perpetrators. Group 1 is labelled as G1, Group 2 is labelled as G2 and so on. Continuous variables are reported as an average; categorical variables as proportion of the group.

APPENDIX B. SENSITIVITY ANALYSIS

Appendix B1. Gender stratification versus pooled clustering

This section reports detailed results comparing the gender-stratified clustering solution used in the main analysis with an alternative pooled clustering specification.

The baseline gender stratified solution (Groups M1-M4 for males and F1-F4 for females) is compared with a pooled solution estimated jointly for males and females, yielding four pooled groups (P1-P4). In the pooled model, gender is included as a clustering variable, and the number of clusters is fixed at four to maintain comparability with the baseline solution. Pooled groups are aligned to gender specific groups using centroid matching.

Table B1 reports agreement between pooled and gender stratified solutions. Agreement is moderate and consistent across genders (ARI = 0.51 for both males and females). Additional group level diagnostics are reported in Table B2, including cluster purity and the dominant pooled group for each baseline group. These diagnostics illustrate how baseline groups distribute across pooled clusters and provide a group level complement to the ARI results.

Table B1. Agreement between pooled and gender-stratified clustering solutions

Gender	Cohort size	k (gender-stratified)	k (pooled)	ARI (pooled vs gender-stratified)
Male	53,162	4	4	0.51
Female	18,734	4		0.51

Table B2. Purity of baseline gender-stratified groups relative to pooled clustering solution (young cohort)

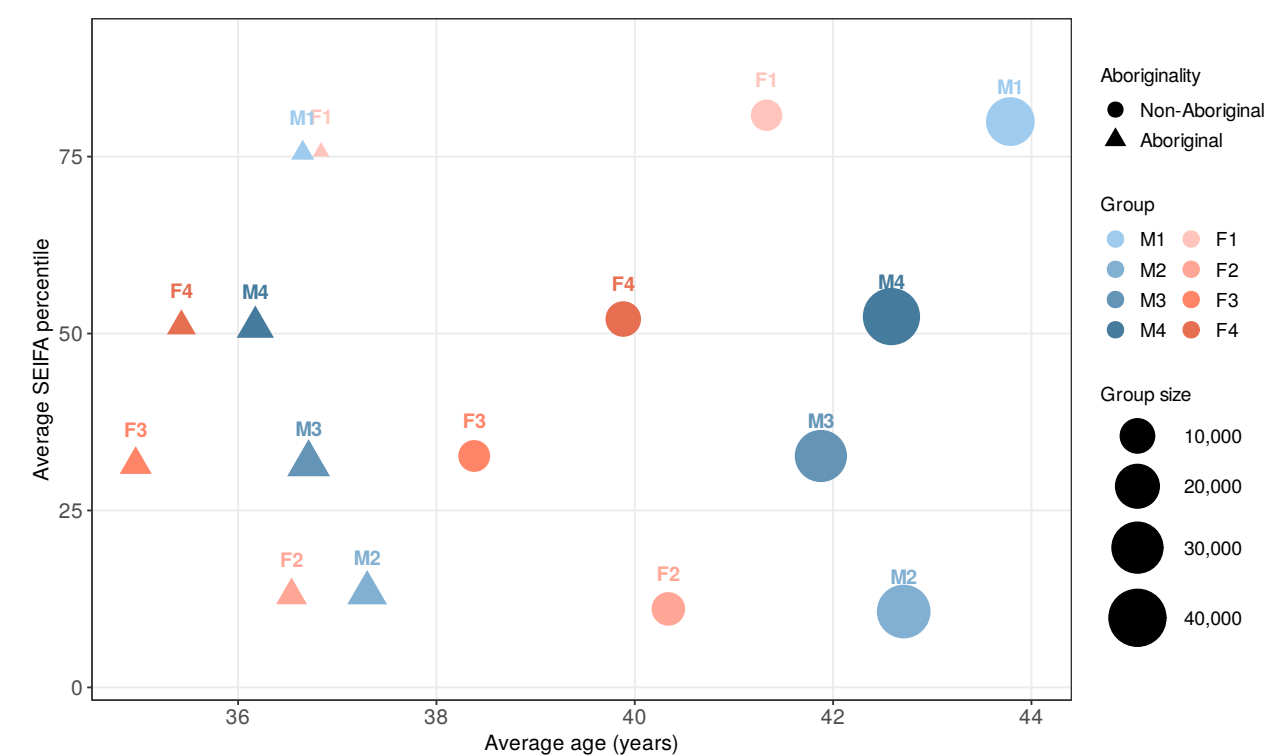
Metric	M1	M2	M3	M4	F1	F2	F3	F4
Group size (n)	8,236	14,513	14,744	15,669	2,898	5,040	5,212	5,584
Dominant pooled group index	P2	P3	P4	P4	P2	P3	P1	P4
n in dominant pooled group	8,236	13,087	6,544	13,420	2,898	4,398	2,156	4,726
Purity (%)	100	90.2	44.4	85.6	100	87.3	41.4	84.6

Appendix B2. Childhood observability and cohort coverage

This section reports detailed results for the reassignment of individuals from older cohorts to the baseline clustering solution.

Figure B1 shows the distribution of the eight baseline groups when individuals from all cohorts are assigned to the baseline clustering solution. Group positions in demographic space are broadly consistent across cohorts, while differences are reflected primarily in the distribution of individuals across groups.

Figure B1. Distribution of eight baseline groups of DFV perpetrators in terms of demographics, with all cohorts assigned



To assess structural agreement, clustering is re estimated separately within each cohort using the same principal component space and number of groups ($k = 4$ by gender). Cohort specific solutions are aligned to the baseline solution using centroid matching.

Table B3 reports agreement between cohort specific refits and the baseline solution. Agreement is highest for Cohort C (ARI = 1.00 for males and females), remains high for Cohort A (males = 0.83; females = 0.78), and is more variable for Cohort B (males = 0.50; females = 0.87). Table B4 reports group-level purity measures for cohort specific refits, illustrating variation in assignment consistency across groups and cohorts. Additional descriptive statistics for cohort specific groups are reported in Tables B5 and B6.

Table B3. Agreement between cohort-specific and baseline clustering solutions

Birth cohort	Gender	Cohort size	k (baseline)	k (cohort refit)	ARI (refit vs baseline)
A_pre1969	Male	29,034	4	4	0.83
	Female	6,427	4	4	0.78
B_1969-1987	Male	70,748	4	4	0.50
	Female	18,596	4	4	0.87
C_post1987	Male	53,162	4	4	1
	Female	18,734	4	4	1

Table B4. Purity of baseline gender-stratified groups relative to cohort-specific refits, by birth cohort

Birth cohort	Metric	M1	M2	M3	M4	F1	F2	F3	F4
A_pre1969	Group size (n)	6,223	7,966	6,431	8,414	1,464	1,841	1,225	1,897
	Dominant refit group	M1	M2	M3	M4	F1	F2	F3	F4
	n in dominant group	5,588	7,707	6,426	7,327	1,287	1,799	1,225	1,498
	Purity (%)	89.8	96.7	99.9	87.1	87.9	97.7	100	79
B_1969-1987	Group size (n)	13,002	17,929	18,540	21,277	3,589	5,012	4,428	5,567
	Dominant refit group	M1	M2	M4	M4	F1	F2	F3	F4
	n in dominant group	13,002	15,821	9,849	17,982	3,326	4,858	4,426	5,082
	Purity (%)	100	88.2	53.1	84.5	92.7	96.9	100	91.3

**Table B5. Comparison of summary characteristics for gender-stratified groups of DFV perpetrators:
birth cohort A (before 1969)**

Group index for birth cohort A_pre1969	A_M1	A_M2	A_M3	A_M4	A_F1	A_F2	A_F3	A_F4
Number of individuals	5,615	7,707	7,754	7,958	1,591	1,877	615	2,344
Panel A. Demographics								
Male (%)	100	100	100	100	0	0	0	0
Female (%)	0	0	0	0	100	100	100	100
Aboriginal (%)	1.3	9.3	12.3	5.7	3.6	0.5	97.9	4.1
Age at first DV finalisation (mean)	54.7	54.1	52.3	53.5	53.9	53.2	50.9	52.7
SEIFA percentile (mean)	81.9	10.4	33.7	54.7	79.6	13.9	25.9	46.8
Panel B. Principal components scores								
PC1: Lifetime intensity of crime perpetration	0	14	31.7	19.3	9.4	16	100	24.5
PC2: DFV and violent victimisation history	11.9	8.7	0	4.5	37.2	46.7	100	48.4
PC3: Childhood maltreatment and youth diversions	100	93.2	85.5	91.8	85.2	77.5	0	71.6
PC4: Recent DV, violent charges and AVO breaches	74.2	41.9	27.3	59.9	57.1	44.4	0	100
PC5: Out-of-home care history, placement characteristics, and youth diversions	22.3	9.1	0	7.7	38.6	39.7	100	40.6
PC6: Housing and homelessness support service engagement	100	37	18.2	37.6	57	11.6	0	17.2
Panel C. Recent reoffending and revictimisation rates								
DFV reoffending (%)	5.6	7.8	10.5	8.1	4.5	5.6	8.5	6.7
General reoffending (%)	9.6	13.2	18.4	15.4	8.4	9.4	20.7	11.8
DFV revictimisation (%)	2.6	3.7	4.7	4	7.5	10.6	28	12.8
General revictimisation (%)	9.4	12.7	15.3	14.2	14.6	19.9	39.3	22
Panel D. Raw principal components scores (mean)								
PC1_raw: Lifetime intensity of crime perpetration	-3.3	-2.2	-0.9	-1.8	-2.6	-2.1	4.2	-1.5
PC2_raw: DFV and violent victimisation history	-0.5	-0.7	-1.2	-0.9	1.1	1.6	4.8	1.7
PC3_raw: Childhood maltreatment and youth diversion justice involvement	-0.4	-0.8	-1.3	-0.9	-1.3	-1.7	-6.2	-2.1
PC4_raw: Recent DV, violent charges and AVO breaches	-0.1	-0.2	-0.3	-0.2	-0.2	-0.2	-0.3	-0.1
PC5_raw: Out-of-home care history, placement characteristics, and youth diversion	-0.3	-0.4	-0.5	-0.4	-0.1	-0.1	0.6	-0.1
PC6_raw: Housing and homelessness support service engagement	0.6	0.2	0.1	0.2	0.4	0.1	0.0	0.1

Note. This table reports summary characteristics for gender-stratified DFV perpetrator groups within birth cohort A (born before 1969), restricted to individuals with no observable Child Protection or Out-of-Home Care records due to data coverage commencing in 1987. Group labels (A_M1–A_M4; A_F1–A_F4) correspond to the nearest-centroid assignment from the baseline gender-stratified typology. Panel A reports demographic characteristics of these perpetrators. Panel B reports the average percentile rank for each group's principal component. Panel C reports the share of individuals within each group that have at least one DFV related reoffence and at least one DFV related revictimisation within 12 months. Continuous variables are reported as an average; categorical variables as proportion of the group.

**Table B6. Comparison of summary characteristics for gender-stratified groups of DFV perpetrators:
birth cohort B (between 1969 and 1987)**

Group index for birth cohort B_1969-1987	B_M1	B_M2	B_M3	B_M4	B_F1	B_F2	B_F3	B_F4
Number of individuals	12,345	17,848	20,015	20,540	3,326	4,860	5,065	5,345
Panel A. Demographics								
Male (%)	100	100	100	100	0	0	0	0
Female (%)	0	0	0	0	100	100	100	100
Aboriginal (%)	2.1	12.7	16.5	8.2	3.8	18.5	24.4	12.1
Age at first DV finalisation (mean)	37.6	36.2	35.8	36.7	38.5	37.1	36.9	37.5
SEIFA percentile (mean)	80.3	11.2	33.2	53.4	81.9	11.3	33.1	53.4
Panel B. Principal components scores								
PC1: Lifetime intensity of crime perpetration	0	31.8	60.4	34.3	2.4	44.5	100	45.1
PC2: DFV and violent victimisation history	19.5	9.1	0	8.5	56.7	78.9	100	70.5
PC3: Childhood maltreatment and youth diversions	100	91.1	85.3	91	76	40.9	0	48.6
PC4: Recent DV, violent charges and AVO breaches	71.2	0	2.8	36.5	86.9	51	100	75.1
PC5: Out-of-home care history, placement characteristics, and youth diversions	37.6	19.7	0	15.4	64.6	67.9	100	51.4
PC6: Housing and homelessness support service engagement	100	74.6	56.7	64.6	72.2	11	0	7.3
Panel C. Recent reoffending and revictimisation rates								
DFV reoffending (%)	7.1	9.9	11.7	10.2	5	7.6	8.6	7.6
General reoffending (%)	15	20.4	26.3	22.1	9.9	15.6	22	16.4
DFV revictimisation (%)	2.9	4	5.2	4.3	10.9	18.5	25.9	17.4
General revictimisation (%)	11	14.1	17.1	15.2	18.6	29.7	38.9	28.1
Panel D. Raw principal components scores (mean)								
PC1_raw: Lifetime intensity of crime perpetration	-2.3	-0.4	1.4	-0.2	-2.1	0.4	3.7	0.4
PC2_raw: DFV and violent victimisation history	-1.0	-1.7	-2.3	-1.7	1.6	3.1	4.6	2.5
PC3_raw: Childhood maltreatment and youth diversion justice involvement	-0.4	-0.8	-1.0	-0.8	-1.4	-2.9	-4.6	-2.6
PC4_raw: Recent DV, violent charges and AVO breaches	0.0	-0.2	-0.2	-0.1	0.1	-0.1	0.1	0.0
PC5_raw: Out-of-home care history, placement characteristics, and youth diversion	-0.1	-0.3	-0.4	-0.3	0.0	0.1	0.3	-0.1
PC6_raw: Housing and homelessness support service engagement	0.6	0.3	0.1	0.2	0.3	-0.4	-0.5	-0.4

Note. This table reports summary characteristics for gender-stratified DFV perpetrator groups within birth cohort B (born between 1969 and 1987), for whom Child Protection (CP) and Out-of-Home Care (OOHC) histories are only partially observable owing to administrative data coverage commencing in 1987. Group labels (B_M1–B_M4; B_F1–B_F4) correspond to the nearest-centroid assignment from the baseline gender-stratified typology. Panel A reports demographic characteristics of these perpetrators. Panel B reports the average percentile rank for each group's principal component. Panel C reports the share of individuals within each group that have at least one DFV related reoffense and at least one DFV related revictimisation within 12 months. Continuous variables are reported as an average; categorical variables as proportion of the group.

Appendix B3. Temporal stability across DV index windows

This section reports clustering stability across DFV index windows. To assess temporal stability, clustering is re estimated within three DFV index windows: W1 (2017-2018), W2 (2019-2020), and W3 (2021-2022). Within each window, gender-stratified clustering is performed using the same principal component space and number of groups, and solutions are aligned to the baseline using centroid matching.

Table B7 reports agreement between window-specific refits and the baseline solution. ARI values vary by gender and time window, with male agreement ranging from 0.52 in W1 to 0.98 in W2 and 0.95 in W3, while female agreement is highest in W1 (0.95) and lower in subsequent windows (0.53 in W2 and 0.55 in W3). Table B8 reports group level purity measures for window specific refits. Figure B2 shows the distribution of individuals assigned to baseline groups across DFV index windows. Baseline centroids are applied consistently across windows to illustrate changes in group prevalence and demographic composition over time.

Table B7. Temporal stability of gender-stratified clustering across DFV index windows (young cohort)

Time window	Gender	Cohort size	k (baseline)	k (cohort refit)	ARI (refit vs baseline)
W1_2017-2018	Male	8,054	4	4	0.52
	Female	2,778	4	4	0.95
W2_2019-2020	Male	11,274	4	4	0.98
	Female	4,023	4	4	0.53
W3_2021-2022	Male	19,038	4	4	0.95
	Female	6,649	4	4	0.55

Table B8. Purity of baseline gender-stratified groups relative to time-window-specific refits (young cohort)

Birth cohort	Metric	M1	M2	M3	M4	F1	F2	F3	F4
W1_2017-2018	Group size (n)	1,369	2,161	2,147	2,377	452	718	791	817
	Dominant refit group	M1	M2	M4	M4	F1	F2	F3	F4
	n in dominant group	1,369	2,095	1,264	1,780	441	718	784	781
	Purity (%)	100	96.9	58.9	74.9	97.6	100	99.1	95.6
W2_2019-2020	Group size (n)	1,820	3,078	3,055	3,321	675	1,062	1,074	1,212
	Dominant refit group	M1	M2	M3	M4	F1	F2	F4	F4
	n in dominant group	1,815	3,004	3,038	3,321	675	1,060	503	967
	Purity (%)	99.7	97.6	99.4	100	100	99.8	46.8	79.8
W3_2021-2022	Group size (n)	2,586	5,186	5,597	5,669	1,009	1,781	1,873	1,986
	Dominant refit group	M1	M2	M3	M4	F1	F2	F3	F4
	n in dominant group	2,586	5,161	5,416	5,481	1,009	1,650	901	1,670
	Purity (%)	100	99.5	96.8	96.7	100	92.6	48.1	84.1

Figure B2. Distribution of eight baseline groups of DFV perpetrators in terms of demographics, with all cohorts assigned

