Predicting Domestic Violence Revictimization at First Contact: Machine Learning for Case Prioritization in a Brazilian Second-Response Program

Matheus Nascimento[1] Ísis Lira[2] Rodrigo Serrano-Berthet[3]
Laura Schiavon[1,2] Claudio Todisco[4] Federico Veneri[3] Carolina Zancan[4]

Área de Economia Social e do Trabalho

Abstract

Protective resources for domestic violence victims are scarce relative to registered cases. We study how much ordering victims by predicted revictimization risk improves on informal screening. Using administrative records from domestic violence occurrences in Paraná, Brazil, we constructed four algorithms under two specifications: a baseline using standard police records and an augmented version adding data from the National Risk Assessment Form. Because visit capacity is fixed and below demand, we evaluate models by their yield at the top of the risk distribution, using Precision@k. Prioritizing the top 100 cases, the preferred model recovers more than twice the revictimization cases expected under random allocation, a gain above 100%. Adding the structured instrument recovers approximately 9 additional true positive cases per 100 deployments. Ordering victims by predicted risk substantially increases the share of at-risk victims reached under any fixed capacity constraint, whether for follow-up visits or other protective interventions.
Keywords: Domestic violence. Revictimization. Resource allocation. Machine learning.

Resumo

Os recursos de proteção a vítimas de violência doméstica são, no geral, escassos em relação ao número de casos registrados. Este trabalho avalia quanto a ordenação das vítimas por risco previsto de revitimização melhora em relação à triagem informal. A partir de registros administrativos de ocorrências de violência doméstica no Paraná, Brasil, construímos quatro algoritmos sob duas especificações: uma de referência, baseada nos registros policiais padrão, e uma versão ampliada, que incorpora dados do Formulário Nacional de Avaliação de Risco. Como a capacidade de atendimento é fixa e inferior à demanda, avaliamos os modelos por seu rendimento no topo da distribuição de risco, por meio da métrica Precision@k. Ao priorizar os 100 casos de maior risco, o modelo preferido recupera mais que o dobro dos casos de revitimização esperados sob alocação aleatória, um ganho superior a 100%. A inclusão do instrumento estruturado recupera aproximadamente 9 casos verdadeiros positivos adicionais a cada 100 acionamentos. A ordenação das vítimas por risco previsto aumenta substancialmente a parcela de vítimas em risco alcançadas sob qualquer restrição fixa de capacidade, seja para visitas de acompanhamento, seja para outras intervenções de proteção.
Palavras-chave: Violência Doméstica. Revitimização. Alocação de recursos. *Machine learning.*

JEL classification: K42, J12, J16, C53

[1]Federal University of Juiz de Fora (UFJF).
[2]Leme - Laboratory for Violence Reduction
[3]Inter-American Development Bank (IDB).
[4]Paraná State Secretariat of Public Security (SESP-PR).

1 Introduction

Intimate partner violence is a central public security challenge in Brazil. More than 40% of Brazilian women have experienced physical, sexual, or psychological violence by a current or former partner at some point in their lives (Fórum Brasileiro de Segurança Pública and Instituto Datafolha, 2025), a pattern consistent with global estimates that 27% of women aged 15–49 who have been in a relationship report physical or sexual abuse by an intimate partner (World Health Organization, 2021).

Violence by intimate partners differs from other interpersonal violence in one respect that shapes the institutional response: it recurs. Emotional and financial dependence, fear of retaliation, and the disruption of shared households raise the cost of leaving and keep victims exposed to the same perpetrator (Barnett, 2000, 2001), which makes repeat victimization more likely than in other offense types (Bureau of Justice Statistics, 2017). This recurrence is concentrated in time. The probability of a subsequent offense peaks in the weeks following the initial report and declines thereafter (Lloyd et al., 1994; Farrell, 1995), which defines a narrow window in which a follow-up contact can reach the victim before further escalation.

Second-response policing programs are designed to exploit this window. Specialized units conduct follow-up visits to victims already registered in administrative records, monitoring protective order compliance, providing referrals to support networks, and establishing a state presence intended to deter the perpetrator. Evaluations indicate that these programs reduce revictimization among enrolled victims (Koppensteiner et al., 2024). Their effectiveness, however, is bounded by capacity. The volume of registered occurrences exceeds the number of follow-up visits a unit can conduct, so not every eligible victim can be contacted within the elevated-risk window. In Paraná, the fifth most populous state in Brazil, registered domestic violence occurrences exceeded follow-up visit capacity by a ratio of six to one in 2023. When capacity is fixed and demand exceeds it, the order in which cases are visited determines how many at-risk victims the program reaches. Under current practice, this order is set by informal case screening rather than by a risk-based rule, and there is no guarantee that it tracks the underlying distribution of revictimization risk.

This paper asks how much a police unit gains, in true revictimization cases reached under a fixed visit capacity, by ordering cases according to predicted risk rather than at random. We estimate a predictive model for domestic violence revictimization using administrative records from Paraná and evaluate it as a case-prioritization rule under the capacity constraints that second-response units face. Revictimization is defined as a subsequent domestic violence registry involving the same victim within 180 days of an index occurrence. We estimate four algorithms, logistic regression, random forests, LightGBM, and XGBoost, under two feature specifications: a baseline drawn from the standard police record (BOU) and an augmented specification that adds 63 metrics from the National Risk Assessment Form (FONAR), a structured instrument administered by trained Civil Police personnel. Because the operational problem is yield at the top of the risk distribution rather than average discrimination across all thresholds, we evaluate models with Precision@k and its gain over random allocation, alongside AUC for comparability with the existing literature.

The results support three main conclusions. First, the logistic regression BOU+FONAR specification reaches an AUC of 0.630 on the main test set, and under a simulated weekly triage restricted to 50 visits it reaches a victim who is later revictimized in 22.7% of prioritized deployments, against 14.2% under random allocation. Second, ordering the top 100 cases by predicted risk recovers more than twice the revictimization cases expected under random allocation, a gain above 100% for the augmented specification and 43% for the baseline that relies only on standard police records. Third, adding the structured risk instrument recovers approximately 9 additional true positive cases per 100 prioritized deployments relative to the baseline, which quantifies the operational value of collecting the instrument beyond the administrative record already available at no marginal cost.

This study makes two contributions. First, it provides evidence on domestic violence revictimization prediction from a Latin American setting, using short-window, police-recorded outcomes. Validated instruments and predictive models concentrate in high-income Western jurisdictions and rely on re-arrest, conviction, or homicide outcomes measured over multi-year windows, and their performance does not transfer automatically across legal frameworks or baseline population rates (Hegel et al., 2022). Second, it evaluates predictive performance in terms tied to the allocation decision. Most validation studies report AUC, which averages discriminatory ability across all thresholds; under fixed capacity, the relevant quantity is the yield among the highest-risk cases actually visited (Shmueli, 2019; Greene et al., 2022). We report Precision@k and the gain over random allocation, which map model output directly to the deployment decision.

The remainder of the paper proceeds as follows. Section 2 reviews the empirical literature on domestic violence recurrence and risk stratification frameworks. Section 3 details the *Patrulha Maria da Penha*. Section 4 describes the administrative data sources, sample construction criteria, and predictive modeling strategy. Section 5 reports the main empirical results and validation metrics. Section 6 discusses policy implications, limitations, and directions for future research.

2 Literature Review

This study relates to two strands of the literature. First, it contributes to the evidence on how institutional frameworks and public policies affect domestic violence. Second, it connects to the literature on domestic violence risk assessment, focusing on harm concentration, stratification instruments, and the predictive performance of algorithmic tools under capacity constraints.

Regarding the first strand, cross-country evidence links domestic violence prevalence to institutional and economic determinants. Legal system design and institutional capacity shape enforcement and reporting incentives (Iyengar, 2009; Chin and Cunningham, 2019). Similarly, female economic empowerment and household bargaining power affect exposure to violence (Aizer, 2010; Bhalotra et al., 2021), though gains in women's relative economic status can trigger coercive control or violent backlash (Bobonis et al., 2013; Erten and Keskin, 2018). Underlying gender norms condition how these economic and legal channels operate (Alesina et al., 2021), within a global landscape where intimate partners or family members commit approximately 55% of female homicides (United Nations Office on Drugs and Crime, 2023).

Institutional responses alter these incentives, enforcement, and reporting behaviors through targeted interventions. Specialized women's police stations decrease domestic violence and increase reporting rates (Perova and Reynolds, 2017; Amaral et al., 2021). In Brazil, where reported cases of violence against women more than doubled between 2014 and 2023 (Lima et al., 2025) despite severe under-reporting (Observatório da Mulher contra a Violência and Instituto DataSenado, 2024; Vasconcelos et al., 2024), the Maria da Penha Law reduced female homicides from household aggression by 9% (Ferraz and Schiavon, 2022). While these legal frameworks establish broad protective structures, second-response policing programs operate at a later stage of this institutional response, deploying follow-up contacts to victims after an initial report to reinforce compliance with protective orders and connect victims to support networks (Lloyd et al., 1994; Farrell, 1995; Koppensteiner et al., 2024).

Regarding the second strand, the literature on domestic violence risk assessment emphasizes that harm concentrates within a small set of victim-perpetrator dyads. Fewer than 2% of dyads generate 80% of recorded harm, as measured by the Cambridge Crime Harm Index (Bland and Ariel, 2015), and 3% of perpetrators account for 90% of total intimate partner violence harm (Barnham et al., 2017). Risk peaks four weeks after the index event (Morgan et al., 2018; López-Ossorio et al., 2019), making the operational success of follow-up programs dependent on intake screening. However, frontline professionals operate under time and capacity constraints that lead them to prioritize immediate incident characteristics over less visible but more predictive risk factors (Robinson et al., 2016; Turner et al., 2019).

To formalize risk stratification and overcome these frontline cognitive biases, efforts span unstructured professional judgment, structured professional judgment (SPJ), and actuarial instruments. Actuarial tools assign empirical weights to administrative items to produce a score that serves as the prediction (Hilton et al., 2004). Evidence on their relative performance is mixed: broader reviews find comparable validity across instrument types (Messing and Thaller, 2013; Svalin and Levander, 2020), while the largest meta-analysis to date reports an actuarial advantage (van der Put et al., 2019). Instrument type, however, is not sufficient to ensure predictive performance under field conditions. The DASH instrument displays low discriminative validity when completed by frontline officers due to measurement error, as officers tend to record immediate incident characteristics rather than stronger predictive signals (Turner et al., 2019; Myhill et al., 2023). Whether this reflects a structural limitation of checklist-based instruments or a symptom of field completion conditions remains contested.

Machine learning applied to administrative records offers a scalable alternative that avoids these field-completion constraints (Grogger et al., 2021; Quijano-Sánchez et al., 2021; Turner et al., 2022). Models trained on administrative records outperform frontline frameworks, though results remain context-dependent based on data quality, outcome definitions, and follow-up windows (Hui et al., 2023; Verrey et al., 2023; Le et al., 2026). Furthermore, these predictive models concentrate in high-income Western jurisdictions (Hegel et al., 2022). This study evaluates a predictive model for domestic violence revictimization using administrative data from Paraná, Brazil, testing the performance of structured metrics in a Latin American setting where short-window evidence remains scarce. Finally, while most validation studies report aggregate AUC metrics, we evaluate performance using Precision@k, which links predictive accuracy to deployment decisions under specific capacity constraints (Shmueli, 2019; Greene et al., 2022).

3 Institutional Context: Second-Response Policing in Paraná

This study is set in Paraná, the fifth most populous state in Brazil, with 11.9 million inhabitants (Instituto Brasileiro de Geografia e Estatística, 2025). In 2023, the state registered 23,886 cases of bodily harm in domestic violence contexts, a rate of 397.6 per 100,000 women and the sixth highest among Brazilian states (Fórum Brasileiro de Segurança Pública, 2025).

Public security in Brazil is organized at the state level. In Paraná, operational execution of domestic violence protocols is divided between the Civil Police, responsible for criminal investigation, and the Military Police (PMPR), responsible for patrol, emergency dispatch, and preventive follow-up. The PMPR organizes its response around a three-tier framework mandated across all 399 municipalities.[1] The first tier handles emergency dispatch through the 190 line. The second and third tiers, sequential follow-up and structured prevention, consist of visits to victims already registered in the administrative system and are executed by a specialized unit, the *Patrulha Maria da Penha* (PMP). These follow-up visits are the object of this study.

The PMP identifies target cases by cross-referencing occurrence records from both police forces, denunciation hotline reports, and active protective orders. Deployments take two forms: preventive visits, initiated by operational command from the analysis of recent occurrence records, and protective order compliance visits, triggered by judicial notification. During a visit, officers monitor victim safety, provide referrals to support networks, and collect updated information. Each deployment is linked to the victim's unique identifier and updates a longitudinal monitoring file.

The constraint that motivates this study is the gap between the volume of eligible cases and the capacity to visit them. In 2023, registered domestic violence offenses against women in Paraná exceeded 68,000, while the PMP conducted approximately 12,000 follow-up visits, roughly one visit

[1] Established under State Law No. 19.788/2018, later incorporated into the *Código Estadual da Mulher Paranaense* (State Law No. 21.926/2024); operational protocols follow PMPR Directive No. 003/2025.

for every six registered cases. Not all occurrences are immediately eligible, and the temporal selection window narrows the active pool at any given moment, but the disparity between demand and capacity remains wide. Because capacity is fixed and well below demand, the order in which cases are visited determines how many at-risk victims the program reaches within the elevated-risk window.

Under current practice, this order is set informally. Officers review recent police records and assign priority based on severity and recurrence indicators identifiable during screening. No quantitative risk stratification occurs at this stage, so the allocation of visits need not track the underlying distribution of revictimization risk. This is the allocation problem the paper addresses.

4 Data and Methods

4.1 Data Sources

This study integrates two primary administrative sources: police records and structured risk assessment forms. Additionally, we utilize two auxiliary administrative records exclusively to handle temporal censoring. All data tracks individual-level observations within the state of Paraná. While the complete occurrence registry spans from January 2010 through December 2023 to enable the longitudinal linkage, our primary analytical sample is restricted to the window between January 2022 and December 2023. This temporal restriction isolates the exact period where structured risk assessment data achieve sufficient operational scale for model training and evaluation.

Police Records

The primary data source is the *Boletim de Ocorrência Unificado* (BOU), a standardized administrative registry generated across all police forces in Paraná. Established by state legislation to unify reporting systems across the Military Police, Civil Police, and Municipal Guard, the BOU enables systematic cross-force record linkage and longitudinal individual tracking. Attending officers complete the registry at the time of the incident, capturing structured data on the offense's legal classification, the physical setting, the means employed, and the immediate police response.

Every individual registered in a BOU is identified through their full name, maternal name, date of birth, and national identification number. The protocol classifies individuals by their operational role: victim, perpetrator, witness, or involved party. This role classification forms the empirical basis for isolating victims and perpetrators in our analytical sample. Recorded demographic covariates include estimated age, biological sex, and municipality of residence. Additionally, each BOU links to administrative annexes tracking weapon or substance seizures. Ultimately, the BOU provides the core longitudinal record of an individual's contacts with the state's criminal justice system.

The National Risk Assessment Form

The National Risk Assessment Form—*Formulário Nacional de Avaliação de Risco* (FONAR)—is a structured risk instrument established by federal authority to standardize domestic violence tracking.[2] Designed for administration during the victim's first formal contact with law enforcement, the instrument features two core components. First, an objective section collects standardized data regarding the victim, the perpetrator, and the history of domestic abuse. Second, a subjective section captures qualitative risk assessments and institutional referral recommendations completed exclusively by specialized personnel.

In Paraná, the FONAR is administered by trained Civil Police personnel and linked directly to the corresponding BOU registry. The instrument can be completed either digitally through an integrated

[2]The FONAR was introduced by Joint Resolution CNJ/CNMP No. 5 of March 3, 2020, and given statutory force by federal Law No. 14,149 of May 5, 2021.

information system or via paper forms. This study utilizes digitally registered records exclusively, as paper-based forms were not available for this research. Consequently, the 34.1% of BOU records linked to a FONAR in our analytical sample does not represent the true completion rate. This figure captures only digitally registered forms and cannot distinguish cases where a paper-based FONAR was completed but unobserved from cases where no form was completed. Among the latter, our data identify victim refusal as one source of non-completion. Another source is the unavailability of trained personnel at the time of the incident, which our data cannot isolate from other reasons for non-completion.

Auxiliary Administrative Records

Two supplementary data sources are linked to the BOU to construct temporal censoring variables. First, arrest warrant records register the execution date of any judicial mandate against perpetrators associated with each occurrence. Second, mortality registries record the exact date of death of any individual involved in an incident. We link both auxiliary sources to the BOU using the unique individual identifier constructed in the linkage process. Because arrest warrant and mortality records may contain multiple entries linked to the same occurrence—such as multiple individuals or warrant records associated with a single incident—we collapse these sources into binary occurrence-level indicators. These indicators flag whether any associated perpetrator was arrested and whether any associated individual died before merging into the final dataset. Section 4.2 details the role of these indicators in the final analytical sample construction.

4.2 Sample Construction

The analytical sample is victim-centered, defining the unit of observation as a female victim's first registered domestic violence occurrence within the study period. This construction filters the raw administrative data through several sequential steps.

We restrict the baseline population to domestic violence BOU records registered between January 2022 and December 2023. As established, this window captures the initial period of systematic digital FONAR implementation in Paraná. While earlier registries dating back to 2010 inform the longitudinal individual identifier, they are excluded from this analytical sample due to sparse and unsystematic risk form coverage.

The sample is restricted to criminal incidents involving 1 to 5 victims and a maximum of two perpetrators. These thresholds exclude a small share of atypical events that likely reflect administrative data entry anomalies. Within each remaining occurrence, the female victim constitutes the primary unit of observation. We exclude registries where the perpetrator is mistakenly recorded in the victim role, as well as occurrences classified as homicide, since the revictimization outcome is undefined for these cases.

To enforce our first-event design, we utilize the record linkage identifier to isolate each victim's earliest domestic violence registry within the study period. When two separate BOU records occur for the same individual within a 24-hour window, we treat them as a single event. In these deduplication cases, we retain the earliest chronological record. This pipeline ensures that the final analytical dataset contains exactly one unique observation per victim.

The outcome variable is a binary indicator equal to one if a new domestic violence BOU involving the same victim is registered within 180 days of the index occurrence. To handle temporal censoring, we apply two primary exclusion rules. First, we exclude observations where the perpetrator was arrested and the victim experienced no subsequent incident within the 180-day window. In these cases, the absence of revictimization is attributable to perpetrator incapacitation rather than latent baseline risk. Second, we omit observations where the victim died within the 180-day follow-up window without a prior revictimization event.

To prevent right-censoring, we exclude victims whose 180-day exposure window extends beyond December 2023. For the same reason, we remove the entire month of July 2023 from the sample. Because 180 days does not map exactly to six calendar months, a share of early July cases technically satisfy the exposure criterion but lack a fully observed outcome window. Removing this cohort eliminates this residual censoring.

The final analytical sample restricts to victims with a completed and digitally registered FONAR. The training period (January 2022 to April 2023) comprises 25,452 observations with a 13.4% revictimization rate. The subsequent test period (May to June 2023) contains 2,872 observations with a 14.2% revictimization rate.

4.3 Feature Engineering

We estimate two model specifications using the analytical sample, varying exclusively in their predictor sets.

The BOU specification relies solely on covariates extracted from the primary police occurrence report (21 covariates). These predictors include the victim's estimated age and a binary indicator for residential settings. Additionally, the model incorporates thirteen binary indicators for legal classifications. These mutually non-exclusive categories aggregate raw typification strings into physical violence (high, mid, and low severity), psychological violence, sexual violence, property offenses, firearms-related offenses, substance-related offenses, kidnapping, protective order violations, legal non-compliance, and offenses involving elderly, incapacitated, or child victims. Finally, we include six binary indicators capturing the means employed during the incident: verbal or psychological coercion, physical force, firearms or explosives, blunt objects, bladed weapons, and drugs or chemical substances.

The BOU+FONAR specification supplements the baseline model with 63 additional predictors extracted from the risk assessment form. In this configuration, we replace the BOU-estimated age with the victim's self-reported age from the FONAR. These features encompass five distinct risk domains:

- Violence History: Prior severe and non-severe physical assaults, sexual assault, escalation of violence severity, and previous police contact or protective orders.
- Perpetrator Behavior: Systematic coercive control, substance abuse, mental health history, suicidal ideation, economic distress, weapons access, and threats targeting dependents, family, colleagues, or animals.
- Victim Vulnerability: Relationship dissolution, new partnerships, pregnancy, children witnessing the abuse, children fathered by the perpetrator, custody conflicts, financial dependence, and physical disabilities or housing instability.
- Relational Status: Structural type of intimate relationship, distinguishing current from former romantic partners.
- Socioeconomic Characteristics: Educational attainment for both parties, with missing observations explicitly encoded as a separate indicator category.

4.4 Modeling Strategy

The prediction goal is a binary classification of revictimization within 180 days of the index occurrence. The validation design uses a temporal split: the model is trained on occurrences from January 2022 through April 2023 and evaluated on data from May and June 2023. The training set precedes the test set in calendar time. This design replicates the operational setting where a deployed model scores new cases as they arrive, following standard recommendations in the predictive policing and domestic violence prediction literatures (Tollenaar and Van Der Heijden, 2019; Turner et al., 2022).

We estimate four algorithms for each specification: logistic regression, random forests, LightGBM, and XGBoost (Breiman, 2001; Chen and Guestrin, 2016; Ke et al., 2017; Dressel and Farid, 2018). These models are standard benchmarks in the revictimization and domestic violence prediction literatures (Tollenaar and van der Heijden, 2013; Berk et al., 2016; Wijenayake et al., 2018; Turner et al., 2022). While logistic regression serves as an interpretable linear baseline, tree-based ensemble methods capture non-linearities and interactions without manual specification of functional forms.

Model performance is evaluated using three metrics. First, AUC-ROC measures discriminatory ability across all classification thresholds (Hanley and McNeil, 1982). Across domestic and intimate partner violence applications, the historical mean AUC is approximately 0.643 (van der Put et al., 2019). Second, the Brier score evaluates calibration by calculating the mean squared error between predicted probabilities and observed outcomes (Brier, 1950). This metric penalizes models with accurate case rankings but poorly calibrated probability scales, an important property for operational risk communication. Third, Precision at k (P@k) measures the proportion of true revictimization cases recovered within the top k highest-scored observations, reported at thresholds of $k \in \{100, 500, 1000\}$ (Manning et al., 2008).

Threshold-dependent metrics such as sensitivity, specificity, and F1 score are not reported. These metrics require a fixed classification threshold. However, the operational threshold depends on the patrol's capacity constraints, which vary across deployment contexts and are not determined by the model. Reporting these metrics at a single threshold would misrepresent the model's utility across different deployment scenarios. P@k addresses this constraint by measuring yield at specific operational capacities, defined by the number of visits a patrol can conduct in a given period. The ratio of P@k to its random baseline—equivalent to lift at rank k—quantifies the operational gain of algorithmic prioritization relative to random case allocation (Shmueli, 2019; Greene et al., 2022).

5 Results

5.1 Descriptive Analysis

Table 1 describes the analytical samples used in the evaluation. The training set includes 25,452 domestic violence occurrences registered between January 2022 and April 2023, with a 13.4% revictimization rate within 180 days. The test set covers May and June 2023, comprising 2,872 observations with a 14.2% revictimization rate. This sample evaluates out-of-sample predictive performance under the main validation scenario.

Table 1: Sample coverage and revictimization rates by period

Period	**N**	**Revictimized**	**Rate (%)**
Jan/2022–Apr/2023 (training)	25,452	3,416	13.4
May–Jun/2023 (test, main)	2,872	408	14.2

Notes: The analytical sample is restricted to cases with a completed and digitally registered FONAR. This represents 34.1% of all registered first domestic violence occurrences during the training period (25,452 with FONAR; 49,263 without).

Before presenting model performance, we characterize the training sample by documenting the bivariate relationship between each predictor variable and the 180-day revictimization outcome. The objective is to provide a descriptive analysis of potential risk factors by examining whether observed covariates differ systematically between revictimized and non-revictimized victims. Tables 2 and 3 report the prevalence of each variable among revictimized ($Y = 1$) and non-revictimized ($Y = 0$) victims, alongside p-values testing the null hypothesis of equal prevalence across groups.

Among occurrence record variables (Table 2), the strongest univariate associations with revictimization occur in psychological violence classifications (73.1% vs. 70.1%, $p < 0.001$), protective order violations (0.5% vs. 0.2%, $p < 0.001$), and property offenses (7.3% vs. 6.0%, $p < 0.01$). The magnitude of these differences is modest. This small baseline variation anticipates the limited predictive power of the BOU-only specification relative to the model with risk assessment data.

Table 2: Descriptive statistics by revictimization status: BOU variables

Variable	**Total (%)**	**Revictimized (%)**	**Not Revictimized (%)**	***p*-value**
Continuous				
Age (years), mean (SD)	36.9 (13.2)	36.8 (12.7)	36.9 (13.3)	0.861
Legal classification				
Physical violence — high severity	0.4	0.3	0.4	0.294
Physical violence — mid severity	34.4	33.2	34.6	0.127
Physical violence — low severity	9.2	9.8	9.1	0.200
Psychological violence	70.5	73.1	70.1	<0.001
Sexual violence	1.6	1.3	1.7	0.104
Protective order violation	0.2	0.5	0.2	<0.001
Firearms-related offense	0.5	0.4	0.5	0.507
Substance-related offense	0.2	0.3	0.2	0.061
Property offense	6.2	7.3	6.0	<0.01
Kidnapping	0.3	0.4	0.3	0.873
Offense involving elderly/child/incapacitated	0.2	0.2	0.2	0.792
Legal non-compliance	1.2	1.3	1.2	0.953
Means employed				
Verbal/psychological	35.7	36.7	35.6	0.245
Physical force	16.6	16.5	16.7	0.799
Firearm or explosive	0.8	0.9	0.8	0.560
Blunt object	0.6	0.4	0.7	0.080
Bladed weapon	3.3	3.8	3.2	0.084
Drug or substance	0.1	0.2	0.1	<0.05
Incident setting				
Residential setting	82.8	84.2	82.5	<0.05

Notes: Training set, N = 25,452. Y = 1 denotes victims who experienced a new domestic violence occurrence within 180 days of the index event. All statistics for categorical variables represent percentages within each sub-sample. p-values are derived from chi-squared tests for binary variables and a two-sample t-test for age.

The FONAR variables exhibit larger differences between groups (Table 3). Among victims who experienced revictimization, prior police reports or protective orders were recorded for 38.9%, compared to 27.3% among those who did not. This represents a gap of 11.6 percentage points. Similarly, escalation of violence severity was reported for 80.0% of revictimized victims versus 73.5% of non-revictimized individuals. Threats involving a knife occurred in 29.3% of revictimization cases versus 20.8% of non-revictimization cases. Perpetrator alcohol abuse stands at 40.8% versus 30.9%, and perpetrator suicidal ideation tracks at 42.8% versus 36.2%.

Table 3: Descriptive statistics by revictimization status: FONAR variables

Variable	Total (%)	Revictimized (%)	Not Revictimized (%)	*p*-value
Victim–perpetrator relationship				
Current partner or spouse	9.3	9.5	9.3	0.627
Former partner or spouse	14.3	13.8	14.3	0.413
Relationship unknown†	68.9	69.6	68.8	0.331
Violence history				
Escalation of violence	74.4	80.0	73.5	<0.001
Prior report or protective order	28.8	38.9	27.3	<0.001
Prior strangulation	11.4	13.3	11.1	<0.001
Prior hanging	19.4	23.2	18.9	<0.001
Prior suffocation	12.3	14.3	12.0	<0.001
Prior burning	1.0	1.6	0.9	<0.001
Prior stabbing	2.7	3.8	2.5	<0.001
Prior punching	42.6	48.1	41.8	<0.001
Prior kicking	35.9	41.2	35.1	<0.001
Prior slapping	46.7	53.0	45.8	<0.001
Prior pushing	58.0	62.3	57.3	<0.001
Prior sexual assault	20.1	23.5	19.6	<0.001
Prior medical care	15.7	17.9	15.4	<0.001
Threats				
Threat with firearm	6.0	7.1	5.8	<0.01
Threat with knife	22.0	29.3	20.8	<0.001
Threat with other weapon	45.1	47.3	44.8	<0.01
Coercive control				
Financial or object control	41.6	50.1	40.3	<0.001
Phone harassment	45.8	52.9	44.7	<0.001
Digital surveillance	38.9	42.0	38.5	<0.001
Isolation from family	29.6	34.7	28.8	<0.001
Interference with work	20.1	23.4	19.6	<0.001
Financial control	12.3	13.6	12.1	<0.05
Perpetrator profile				
Alcohol use	32.3	40.8	30.9	<0.001
Drug use	60.8	64.7	60.2	<0.001
Mental health condition	6.0	6.9	5.8	<0.05
Suicidal ideation	37.1	42.8	36.2	<0.001
Economic distress	34.5	39.6	33.7	<0.001
Firearms used	3.3	3.7	3.3	0.189
Firearms threatened	13.3	16.7	12.7	<0.001
Firearms access	18.9	21.1	18.6	<0.001
Threat to dependents	21.9	25.6	21.4	<0.001
Threat to family members	22.8	26.3	22.2	<0.001
Threat to animals	3.5	4.3	3.4	<0.01
Victim vulnerability				
Recent or ongoing separation	66.5	70.6	65.8	<0.001
Victim in new relationship	12.9	11.4	13.1	<0.01
Financial dependence on perpetrator	16.4	16.8	16.3	0.511
Victim disability	5.3	5.6	5.2	0.355
Violence during pregnancy	22.5	26.0	22.0	<0.001
Children fathered by perpetrator	54.3	54.9	54.2	0.469
Children witnessed violence	56.9	61.4	56.2	<0.001
Custody or alimony conflict	15.1	14.0	15.2	0.060

Notes: Training set, $N = 25{,}452$. Y = 1 denotes victims who experienced a new domestic violence occurrence within 180 days. p-values are derived from chi-squared tests. †The high prevalence of "relationship unknown" (68.9%) reflects missing values in the relationship field of the FONAR, which occurs when the field was not completed by the attending specialist. This category is encoded as a separate binary indicator in the model.

Most FONAR items are statistically significant at conventional levels. However, several variables associated with risk in the literature, such as victim financial dependence, disability, and children fathered by the perpetrator, exhibit no significant differences across groups. These bivariate comparisons offer a preliminary characterization of individual risk factors, but they do not establish which variables drive predictive performance when evaluated jointly with correlated covariates. Section 5 addresses this question directly through multivariate predictive models.

5.2 Out-of-sample Performance

This section evaluates model performance on the held-out test set covering May and June 2023. These observations track incidents registered one to two months after the training period. We compare two model specifications: the BOU-only model, which utilizes variables from the police occurrence report, and the BOU+FONAR model, which augments the baseline with 63 variables from the risk assessment form. Each specification is estimated using four algorithms: logistic regression, random forests, LightGBM, and XGBoost. This setup yields eight distinct models for comparison.

We evaluate predictive performance across three dimensions. First, the AUC-ROC measures overall discriminatory ability. Second, the Brier score measures probability calibration by calculating the mean squared error between predicted probabilities and observed outcomes. Third, Precision at k (P@k) quantifies operational yield at specific capacity constraints, as detailed below.

AUC-ROC and Brier Score

Table 4 reports the AUC-ROC and Brier scores for all eight models. Two patterns emerge from these estimates. First, the BOU+FONAR specification outperforms the BOU baseline across all four algorithms, with AUC-ROC gains ranging from 0.03 to 0.07 points. This consistent gap exceeds the performance variation across algorithms within each specification. This pattern indicates that the structured risk assessment data, rather than algorithmic complexity, drives the predictive advantage. Second, logistic regression achieves the highest AUC-ROC among the BOU+FONAR models (0.630), performing competitively against the tree-based ensemble methods. Finally, Brier scores remain concentrated near baseline levels across both specifications, indicating calibrated probability estimates.

Table 4: AUC-ROC and Brier score: main validation scenario

Specification	**Algorithm**	**AUC-ROC**	**Brier score**
BOU+FONAR	Logit	0.630	0.119
BOU+FONAR	LGBM	0.615	0.120
BOU+FONAR	XGB	0.609	0.120
BOU+FONAR	RF	0.601	0.120
BOU	Logit	0.561	0.121
BOU	LGBM	0.557	0.122
BOU	XGB	0.565	0.121
BOU	RF	0.565	0.121

Notes: Estimation uses training data from January 2022 to April 2023 and test data from May and June 2023 (N = 2,872; revictimization rate is 14.2%).

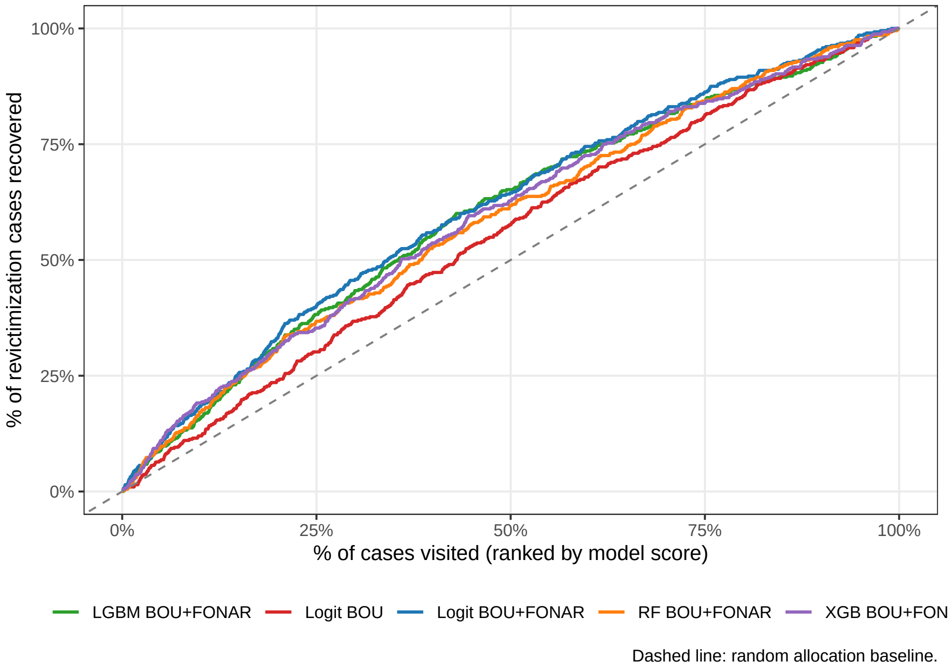
Cumulative Gains Curves

Evaluating model performance requires grounding the analysis in the operational context of the PMP. The patrol conducts preventive follow-up visits to domestic violence victims, but personnel and time limitations restrict overall capacity. Because eligible cases exceed daily visit capacity, the order of case prioritization determines the efficiency of resource allocation. Under a random allocation strategy, the proportion of revictimized victims reached scales linearly with the volume of visits conducted. An algorithmic model aims to concentrate visits among high-risk cases, capturing a higher share of revictimization events within the same capacity constraint.

Figure 1 plots the cumulative gains curves for each model. These curves display the proportion of revictimization cases recovered on the y-axis against the proportion of the population visited on the x-axis, with cases ranked by descending risk score. The dashed diagonal represents the random allocation baseline. Under this baseline, visiting 25% of cases recovers 25% of subsequent revictimizations. Curves positioned above this diagonal indicate predictive performance superior to chance.

The BOU+FONAR specifications position above the diagonal across the entire distribution. At a 25% visit threshold, the Logit BOU+FONAR model recovers approximately 40% of revictimization cases within the May and June 2023 test set, compared to the 25% random baseline. At a 50% visit threshold, this model captures approximately 65% of subsequent revictimizations. While the BOU-only model also outperforms the random baseline, it remains below the BOU+FONAR curves. This performance gap is widest at lower visit thresholds, which represent the resource-constrained operating region. This distribution indicates that the risk assessment data provide the largest marginal predictive gain at the top of the risk distribution.

Figure 1: Cumulative gains curves: main validation scenario (May–June 2023)



Notes: Curves plot the proportion of revictimization cases recovered on the y-axis against the proportion of the population visited on the x-axis. The dashed diagonal represents the random allocation baseline. The BOU+FONAR models include Logit (blue), LGBM (orange), XGB (purple), and RF (green). The BOU-only baseline is represented by the Logit model (red).

Precision@k

While the cumulative gains curve provides a global overview of model performance, operational planning requires metrics tied to the specific capacity constraints of the patrol. Precision at k (P@k) measures the proportion of true revictimization cases among the k highest-scored observations. For example, if a patrol has the resources to conduct 100 visits within a given period and prioritizes according to model scores, P@100 reports how many of those 100 deployments will target a victim who experiences a subsequent occurrence. This metric reflects the operational yield of the triage protocol, mapping directly to specific capacity thresholds rather than averaging performance across all possible deployment scenarios like the AUC-ROC.

Table 5 reports the P@k results for $k \in \{100, 500, 1000\}$ across all eight models. The baseline expected value under random allocation is defined as $k \times \hat{p}$, where $\hat{p} = 14.2\%$ represents the observed revictimization rate in the test set. At the $k = 100$ threshold, the Logit BOU+FONAR model identifies 29 true revictimization cases. This volume doubles the 14 cases expected under random selection. At the $k = 1000$ threshold, this same specification recovers 207 cases against a random baseline expectation of 142 cases, representing a 46% marginal gain.

The BOU-only specifications also outperform the random baseline but achieve lower absolute yields. For example, the top-performing BOU-only specification (RF) recovers 20 true cases at the $k = 100$ threshold, compared to the 14 expected cases. This output represents a 43% gain over the baseline, compared to the 107% gain generated by the Logit BOU+FONAR model at the same evaluation threshold.

Table 5: Precision@k: main validation scenario (May–June 2023)

k	Exp.	BOU+FONAR				BOU			
		Logit	LGBM	XGB	RF	Logit	LGBM	XGB	RF
100	14	29	28	30	30	20	19	16	20
500	71	116	114	111	110	88	85	82	81
1,000	142	207	201	193	184	166	175	173	170

Notes: Cells report the number of true revictimization cases recovered within the top k highest-scored observations. Exp. represents the expected baseline count under random case allocation ($k \times \hat{p}$, where $\hat{p} = 14.2\%$).

5.3 Operational Performance

The P@k results summarize model performance at static thresholds. However, they do not reflect the dynamic operational framework of the PMP, where cases arrive continuously and triage occurs over weekly cohorts of newly registered victims. This section simulates this operational workflow. In each week of the test period, the model selects the 50 highest-scoring cases registered within that specific week for a follow-up visit. This threshold of 50 weekly visits serves as an illustrative benchmark representing the capacity of a mid-sized PMP unit; it is not intended as an optimal policy prescription.

Two statistics are evaluated for each model. *Precision* represents the proportion of visited cases that match victims who experience revictimization within 180 days. This metric measures the efficiency of the patrol's visit capacity. Under random allocation, precision equals the sample prevalence of 14.2%. A higher precision indicates a lower volume of uninformative visits. *Coverage* represents the proportion of all revictimization cases in the test period reached by a prioritized visit, measuring the overall scale of contact across the at-risk population. These two statistics capture distinct operational dimensions: precision measures efficiency per unit of deployment, while coverage measures aggregate population reach.

Table 6 reports these results for the Logit BOU+FONAR and Logit BOU specifications. The nine-week test period from May to June 2023 includes 2,872 registered first domestic violence occurrences, with 408 resulting in revictimization within 180 days. Under the 50-visit protocol, the patrol conducts a total of 450 visits across this nine-week window.

Table 6: Weekly triage simulation: 50 visits per week, May–June 2023

Model	**Cases**	**Revict.**	**Visited**	**TP**	**Coverage**	**Precision**
Logit BOU+FONAR	2,872	408	450	102	25.0%	22.7%
Logit BOU	2,872	408	450	77	18.9%	17.1%

Notes: In each of the nine weeks of the test period, visits are allocated to the 50 highest-scored cases registered that week. TP (true positives) represents the number of visited victims who experienced revictimization within 180 days. Coverage equals TP divided by total revictimizations (408). Precision equals TP divided by total visited cases (450). The random allocation baseline yields an expected value of 64 true positives (450 × 14.2%).

The Logit BOU+FONAR model identifies 102 true revictimization cases among the 450 conducted visits, achieving a precision of 22.7%. This output stands 8.5 percentage points above the 14.2% random allocation baseline. This performance implies that approximately one in four model-prioritized visits reaches a victim who experiences subsequent revictimization, compared to one in seven under random selection. The coverage rate of 25.0% indicates that prioritized deployments reached 102 of the 408 victims who experienced revictimization. Under this capacity constraint, the remaining 306 cases were not contacted within the nine-week test window.

This 25.0% coverage rate reflects the capacity constraints of the simulation window. During the nine-week test period, 408 revictimization events occurred. The patrol conducted 450 visits in total, averaging approximately one deployment per newly registered case per week. Because visits are restricted to the week of initial registration and high-risk cases accumulate faster than weekly capacity allows, this coverage represents the operational limit under a 50-visit constraint. Under a random allocation strategy, the patrol would reach 15.7% of these cases. The Logit BOU-only specification reaches 18.9% of at-risk victims using the same 450 visits, recovering 77 true cases. This baseline represents 25 fewer true-positive contacts than the BOU+FONAR specification, indicating a marginal gain of approximately 3 additional true-positive identifications per week from incorporating the risk assessment instrument.

6 Discussion

The Logit BOU+FONAR specification achieves an AUC of 0.630 on the main test set and, under a simulated weekly triage restricted to 50 visits, identifies a revictimized victim in 22.7% of prioritized deployments against 14.2% under random allocation.

This result can be situated against similar systems developed elsewhere, with the Spanish VioGén system offering the closest point of reference. For a six-month recidivism horizon, which is comparable to the 180-day window used in this study, its police-completed initial screener (VPR4.0) reports an AUC of 0.60 [95% CI: 0.57–0.64] (López-Ossorio et al., 2019), overlapping with the AUC achieved by the Logit BOU+FONAR specification.

This comparison is illustrative rather than probative. The baseline revictimization rate in the VioGén validation sample is lower than in the present study (7.4% at six months, against 14.2% in the main test set), and the two systems further differ in population, data type, and validation design. Overlapping AUC ranges do not imply formal equivalence; they only show that the model's discriminatory performance falls within the range reported for another operational second-response risk assessment system evaluated over a similar follow-up window.

This study contributes to the literature in two aspects. First, it provides evidence on the predictive value of a structured, specialist-administered risk assessment instrument collected as part of routine administrative practice rather than a purpose-built research protocol. The FONAR variables most strongly associated with revictimization in the descriptive analysis—escalation of violence severity, prior police contact, perpetrator alcohol abuse, perpetrator suicidal ideation, and coercive control—correspond to risk domains identified in the broader intimate partner violence literature (Campbell et al., 2009), indicating that the instrument captures conceptually relevant information within an operational data collection environment.

Second, the study quantifies the operational gains a predictive model can generate under the resource constraints that characterize second-response policing in Brazil. In Paraná, registered domestic violence occurrences exceed follow-up visit capacity by a ratio of six to one. Under a fixed-capacity triage protocol prioritizing the top 100 scored cases, the BOU+FONAR specification recovers more than twice the number of revictimization cases expected under random allocation, a gain exceeding 100%. The BOU-only specification, which does not depend on FONAR coverage, still yields a 43% gain over the same baseline. This magnitude of gain is relevant to jurisdictions where second-response units routinely operate below the volume of eligible cases, a condition common across Brazilian states and other settings with comparable institutional capacity constraints.

Three limitations qualify these findings. First, the analytical sample is restricted to occurrences with a FONAR digitally linked to the BOU, representing 34.1% of first-time domestic violence occurrences registered during the training period. As discussed in Section 4, this proportion conflates genuine non-completion with paper-based forms not captured in the digital extract used here. If either source of missingness correlates with incident severity, the BOU+FONAR performance estimates may not generalize to the full caseload. The BOU-only results, unaffected by this restriction, indicate that prioritization leverage remains available for cases without a digitally linked FONAR. Second, the outcome variable captures officially reported revictimization, which undercounts true incidence given the documented gap between the prevalence of intimate partner violence and police reporting rates. Third, this evaluation is conducted within a single subnational jurisdiction, meaning the generalizability of these findings to other Brazilian states or institutional contexts requires empirical verification.

Taken together, these results indicate that administrative data already recorded by the Military Police of Paraná contain a predictive signal capable of improving the allocation of scarce protective resources relative to the informal prioritization protocol currently in use. Because the incremental gains of the augmented specification apply only to the 34.1% of occurrences with a digitally linked FONAR, expanding this coverage is a natural policy direction for extending these gains to the larger share of the caseload currently served by the BOU-only baseline.

Referências

Aizer, Anna, "The Gender Wage Gap and Domestic Violence," *American Economic Review*, 2010, *100* (4), 1847–1859.

Alesina, Alberto, Benedetta Brioschi, and Eliana La Ferrara, "Violence Against Women: A Cross-Cultural Analysis for Africa," *Economica*, 2021, *88* (349), 70–104.

Amaral, Sofia, Sonia R Bhalotra, and Nishith Prakash, "Gender, crime and punishment: Evidence from women police stations in India," Technical Report, IZA Discussion Papers 2021.

Barnett, Ola W, "Why battered women do not leave, part 1: External inhibiting factors within society," *Trauma, Violence, & Abuse*, 2000, *1* (4), 343–372.

_ , "Why battered women do not leave, part 2: External inhibiting factors—social support and internal inhibiting factors," *Trauma, Violence, & Abuse*, 2001, *2* (1), 3–35.

Barnham, Laura, Genevieve C. Barnes, and Lawrence W. Sherman, "Targeting Escalation of Intimate Partner Violence: Evidence from 52,000 Offenders," *Cambridge Journal of Evidence-Based Policing*, 2017, *1* (2), 116–142.

Berk, Richard A., Susan B. Sorenson, and Geoffrey Barnes, "Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions," *Journal of Empirical Legal Studies*, 2016, *13* (1), 94–115.

Bhalotra, Sonia et al., "Intimate Partner Violence: The Influence of Job Opportunities for Men and Women," *The World Bank Economic Review*, 2021, *35* (2), 461–479.

Bland, Matthew and Barak Ariel, "Targeting Escalation in Reported Domestic Abuse: Evidence From 36,000 Callouts," *International Criminal Justice Review*, 2015, *25* (1), 30–53.

Bobonis, Gustavo J., Melissa González-Brenes, and Roberto Castro, "Public Transfers and Domestic Violence: The Roles of Private Information and Spousal Control," *American Economic Journal: Economic Policy*, 2013, *5* (1), 179–205.

Breiman, Leo, "Random forests," *Machine learning*, 2001, *45* (1), 5–32.

Brier, Glenn W., "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, 1950, *78* (1), 1–3.

Bureau of Justice Statistics, "Repeat Violent Victimization, 2005–14," 2017.

Campbell, Jacquelyn C., Daniel W. Webster, and Nancy Glass, "The Danger Assessment: Validation of a Lethality Risk Assessment Instrument for Intimate Partner Femicide," *Journal of Interpersonal Violence*, 2009, *24* (4), 653–674.

Chen, Tianqi and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in "Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining" 2016, pp. 785–794.

Chin, Yoo-Mi and Scott Cunningham, "Revisiting the Effect of Warrantless Domestic Violence Arrest Laws on Intimate Partner Homicides," *Journal of Public Economics*, 2019, *179*, 104072.

de Vasconcelos, Nádia Machado et al., "Subnotificação de violência contra as mulheres: uma análise de duas fontes de dados," *Ciência & Saúde Coletiva*, 2024, *29*, e07732023.

Dressel, Julia and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Science Advances*, 2018, *4* (1), eaao5580.

Erten, Bilge and Pinar Keskin, "For Better or for Worse? Education and the Prevalence of Domestic Violence in Turkey," *American Economic Journal: Applied Economics*, 2018, *10* (1), 64–105.

Farrell, Graham, "Preventing Repeat Victimization," in Michael Tonry and David P. Farrington, eds., *Building a Safer Society: Strategic Approaches to Crime Prevention*, Vol. 19 of *Crime and Justice: A Review of Research*, Chicago: University of Chicago Press, 1995, pp. 469–534.

Ferraz, Claudio and Laura Schiavon, "Crime, punishment, and prevention: The effect of a legal reform on violence against women," *Available at SSRN 4354206*, 2022.

Fórum Brasileiro de Segurança Pública, "Anuário Brasileiro de Segurança Pública, 18ª ed.," 2025.

_ **and Instituto Datafolha**, "Visível e Invisível: a Vitimização de Mulheres no Brasil, 5ª edição," março 2025.

Greene, Travis, Galit Shmueli, Jan Fell, Ching-Fu Lin, and Han-Wei Liu, "Forks over knives: Predictive inconsistency in criminal justice algorithmic risk assessment tools," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 2022, *185* (Supplement_2), S692–S723.

Grogger, Jeffrey, Suzanne Gupta, Rožle Ivandic, and Tom Kirchmaier, "Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases," *Journal of Empirical Legal Studies*, 2021, *18* (1), 90–130.

Hanley, James A. and Barbara J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, 1982, *143* (1), 29–36.

Hegel, Jordana, Kyle D. Pelletier, and Mark E. Olver, "Predictive Properties of the Ontario Domestic Assault Risk Assessment (ODARA) in a Northern Canadian Prairie Sample," *Criminal Justice and Behavior*, 2022, *49* (3), 411–431.

Hilton, N. Zoe, Grant T. Harris, Marnie E. Rice, Carol Lang, Catherine A. Cormier, and Kimberly J. Lines, "A Brief Actuarial Assessment for the Prediction of Wife Assault Recidivism: The Ontario Domestic Assault Risk Assessment," *Psychological Assessment*, 2004, *16* (3), 267–275.

Hui, Victoria, Rose E. Constantino, and Young Ji Lee, "Harnessing Machine Learning in Tackling Domestic Violence: An Integrative Review," *International Journal of Environmental Research and Public Health*, 2023, *20* (6), 4984.

Instituto Brasileiro de Geografia e Estatística, "Estimativas da População dos Municípios e Unidades da Federação Brasileiros," 2025. Portaria IBGE nº 1.098, 27 ago. 2025.

Iyengar, Radha, "Does the Certainty of Arrest Reduce Domestic Violence? Evidence from Mandatory and Recommended Arrest Laws," *Journal of Public Economics*, 2009, *93* (1–2), 85–98.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, 2017, *30.*

Koppensteiner, Martin Foureaux, Jesse Matheson, and Réka Plugor, "The impact of improving access to support services for victims of domestic violence on demand for services and victim outcomes," *American Economic Journal: Economic Policy*, 2024, *16* (1), 292–324.

Le, Bao Linh, Laura Trafford, Selim Firtala, and Brian Williams, "Beyond physical violence: A machine learning framework for predicting IPV victimisation using multidimensional predictors," *European Journal of Criminology*, 2026. Online first.

Lima, G. C. C. et al., "Temporal trend and epidemiological profile of notifications of violence against women in Brazil: 2014–2023," *Epidemiologia e Serviços de Saúde*, 2025, *34*, e20240475.

Lloyd, Sam, Graham Farrell, and Ken Pease, "Preventing Repeated Domestic Violence: A Demonstration Project on Merseyside," Technical Report 49, Home Office Police Research Group, London 1994.

López-Ossorio, Juan José, José Luis González-Álvarez, José Manuel Muñoz Vicente, Cecilia Urruela Cortés, and Antonio Andrés-Pueyo, "Validation and Calibration of the Spanish Police Intimate Partner Violence Risk Assessment System (VioGén)," *Journal of Police and Criminal Psychology*, 2019, *34* (4), 439–449.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze, *Introduction to Information Retrieval*, Cambridge: Cambridge University Press, 2008.

Messing, Jill Theresa and Jonel Thaller, "The Average Predictive Validity of Intimate Partner Violence Risk Assessment Instruments," *Journal of Interpersonal Violence*, 2013, *28* (7), 1537–1558.

Morgan, Anthony, Hayley Boxall, and Rick Brown, "Targeting repeat domestic violence: Assessing short-term risk of reoffending," *Trends and issues in crime and criminal justice*, 2018, (552), 1–16.

Myhill, Andy, Katrin Hohl, and Kelly Johnson, "The 'Officer Effect' in Risk Assessment for Domestic Abuse: Findings from a Mixed Methods Study in England and Wales," *Feminist Criminology*, 2023.

Observatório da Mulher contra a Violência and Instituto DataSenado, "Mapa Nacional da Violência de Gênero: Pesquisa Nacional de Violência contra a Mulher, 10ª edição," Technical Report, Senado Federal 2024.

Perova, Elizaveta and Sarah Anne Reynolds, "Women's police stations and intimate partner violence: Evidence from Brazil," *Social Science & Medicine*, 2017, *174*, 188–196.

Quijano-Sánchez, Lara, Federico Liberatore, Guillermo Rodríguez-Lorenzo, Rosa E. Lillo, and José Luis González-Álvarez, "A Twist in Intimate Partner Violence Risk Assessment Tools: Gauging the Contribution of Exogenous and Historical Variables," *Knowledge-Based Systems*, 2021, *234*, 107586.

Robinson, Amanda L, Andy Myhill, Julia Wire, Jo Roberts, and Nick Tilley, "Risk-led policing of domestic abuse and the DASH risk model," *What Works: Crime Reduction Research. Cardiff & London: Cardiff University, College of Policing and UCL Department of Security and Crime Science*, 2016.

Shmueli, Galit, "Lift up and act! classifier performance in resource-constrained applications," *arXiv preprint arXiv:1906.03374*, 2019.

Svalin, Klara and Sten Levander, "The Predictive Validity of Intimate Partner Violence Risk Assessments Conducted by Practitioners in Different Settings: A Review of the Literature," *Journal of Police and Criminal Psychology*, 2020, *35* (1), 115–130.

Tollenaar, Nikolaj and Peter GM van der Heijden, "Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 2013, *176* (2), 565–584.

_ **and Peter GM Van Der Heijden**, "Optimizing predictive performance of criminal recidivism models using registration data with binary and survival outcomes," *PloS one*, 2019, *14* (3), e0213245.

Turner, Emily, Graham Brown, and Juanjo Medina-Ariza, "Predicting Domestic Abuse (Fairly) and Police Risk Assessment," *Psychosocial Intervention*, 2022, *31* (3), 145–157.

_ **, Juanjo Medina, and Graham Brown**, "Dashing Hopes? The Predictive Accuracy of Domestic Abuse Risk Assessment by Police," *British Journal of Criminology*, 2019, *59* (5), 1013–1034.

United Nations Office on Drugs and Crime, "Gender-Related Killings of Women and Girls (Femicide/Feminicide)," Technical Report, United Nations Office on Drugs and Crime, Vienna 2023.

van der Put, Claudia E., Jeanne Gubbels, and Mark Assink, "Predicting Domestic Violence: A Meta-Analysis on the Predictive Validity of Risk Assessment Tools," *Aggression and Violent Behavior*, 2019, *47*, 100–116.

Verrey, Jack, Barak Ariel, Vincent Harinam, and Louise Dillon, "Using machine learning to forecast domestic homicide via police data and super learning," *Scientific Reports*, 2023, *13*, 22932.

Wijenayake, Senuri, Timothy Graham, and Peter Christen, "A decision tree approach to predicting recidivism in domestic violence," in "Pacific-Asia Conference on Knowledge Discovery and Data Mining" Springer 2018, pp. 3–15.

World Health Organization, "Violence Against Women Prevalence Estimates, 2018," 2021.