Femicides, Anti-Violence Centers, and Policy Targeting

DOCUMENTO DI VALUTAZIONE N. 21

DOCUMENTO
DI VALUTAZIONE

Ufficio Valutazione Impatto
Impact Assessment Office

Senato della Repubblica

Questo *Documento di valutazione* è a cura di

AUGUSTO CERQUA, Department of Social Sciences and Economics, Sapienza University of Rome

COSTANZA GIANNANTONI, Department of Social Sciences and Economics, Sapienza University of Rome

MARCO LETTA, Department of Social Sciences and Economics, Sapienza University of Rome

GABRIELE PINTO, Department of Social Sciences and Economics, Sapienza University of Rome

I dati sono aggiornati al settembre 2025

CODICI JEL: I18, J12, J16, K42

PAROLE CHIAVE: FEMICIDES, GENDER-BASED VIOLENCE, PUBLIC POLICY, POLICY TARGETING, POLICY EVALUATION

Femicides, Anti-Violence Centers, and Policy Targeting

Ottobre, 2025

This paper investigates the socio-economic and geographic patterns associated with femicides and examines the role of local anti-violence centers (AVCs) in combating gender-based violence. First, we compile a novel granular dataset of femicide cases in Italy covering the period 2006-2022 and analyze it with machine learning techniques. This empirical analysis identifies areas at highest risk for women and pinpoints the main demographic and socio-economic predictors of the phenomenon at the territorial level. Second, we collect data on the location and timing of the opening of all local AVCs and show that our femicide risk map only partially aligns with the AVC network, suggesting that machine predictions could be employed to refine the targeting criteria of this policy. Third, using detailed information on the timing of AVC openings in each province and a staggered non-parametric difference-in-difference approach, we find that, on average, the establishment of AVCs did not significantly decrease the occurrence of femicides, while it reduced cases of sexual violence. These finding suggest ample room to improve the targeting and effectiveness of public policy interventions aimed at combating violence against women.

Questo documento analizza i modelli socioeconomici e geografici associati ai femminicidi ed esamina il ruolo dei centri antiviolenza locali (CAV) nel contrasto alla violenza di genere. Innanzitutto, abbiamo compilato un nuovo dataset granulare sui casi di femminicidio in Italia relativi al periodo 2006-2022 e lo abbiamo analizzato con tecniche di apprendimento automatico. Questa analisi empirica identifica le aree a più alto rischio per le donne e individua i principali fattori demografici e socioeconomici predittivi del fenomeno a livello territoriale. In secondo luogo, raccogliamo dati sulla posizione e sui tempi di apertura di tutti i CAV locali e mostriamo che la nostra mappa del rischio di femminicidio è solo parzialmente allineata con la rete degli CAV, suggerendo che le previsioni automatiche potrebbero essere utilizzate per affinare i criteri di selezione di questa politica. In terzo luogo, utilizzando informazioni dettagliate sui tempi di apertura degli AVC in ciascuna provincia e un approccio non parametrico differenziale differenziale scaglionato, abbiamo riscontrato che, in media, l'istituzione degli AVC non ha ridotto in modo significativo il verificarsi di femminicidi, mentre ha ridotto i casi di violenza sessuale. Questi risultati suggeriscono che vi è ampio margine per migliorare la miratezza e l'efficacia degli interventi di politica pubblica volti a combattere la violenza contro le donne.

Summary

Figures and Tablex Index

"*Every so often, in history, an Antigone stands beside the body she must bury and faces the chorus of the people and the tyrant, demonstrating that such sorrow can be a struggle*."

Valeria Parrella, an Italian writer, reflecting on Elena Cecchettin's commitment following the femicide of her sister Giulia.

1. Introduction

Gender-based violence is a systemic problem afflicting most countries worldwide, with femicide being its most extreme form. The ongoing persistence of this issue has recently drawn increased attention from the media and politics, with several governments accelerating the implementation of measures to contrast the phenomenon.[1] In May 2024, the European Council adopted the Directive on combating violence against women and domestic violence (2024/1385). This directive criminalizes both online and offline offenses, establishes requirements for protection and support, and mandates that Member States transpose it into their national law and policy by 2027.

At the same time, in recent years, public opinion has finally begun to mobilize to demand greater public efforts to combat gender-based violence. Yet, in many developed countries, a considerable share of the population continues to minimize the severity of the issue, often viewing it as a concern limited to less developed nations or economically disadvantaged communities (Pappa et al., 2022), or dismissing it as a case of unwarranted concern and media overstatement.[2] Moreover, there is evidence of increasingly popular backlash activism, supported by expanding men's right movements (Botto and Gottzén, 2024). As a consequence, femicides are often perceived—and thus dismissed—by substantial segments of public opinion as just another form of general homicide, with critics arguing that their unique characteristics and root causes do not warrant special attention or policy remedies. This lack of awareness translates into a legislative void: as of April 2025, only three European countries—Croatia, Cyprus and Malta—have recognized femicide as a crime *per se*.[3] In other European legislative systems, femicides remain undistinguished from homicides, and the gender motive in a murder is often not even recognized as an aggravating factor (European Institute for Gender Equality, 2023).

Meanwhile, in many countries, despite the overall decrease in total homicides, the frequency of femicides either remains stable or is increasing.[4] To effectively address this concerning phenomenon, it is essential to enhance the targeting—and, consequently, the effectiveness—of

[1] For instance, *Corriere della Sera*, the leading Italian newspaper, published no articles containing the word "femicide" up to 2004. From 2004 to 2014, 214 articles were published, while from 2014 to June 2024, over 2,200 articles on femicide cases were published. The number of articles containing the words 'violence against women' published from 1994 to 2004 was 1,495. This number remained relatively stable in the subsequent decade, with 1,909 articles, and almost doubled in the last decade, reaching 3,852 articles.

[2] For instance, in a 2018 survey conducted by the Italian Institute of Statistics (ISTAT), about one-quarter of respondents felt that women can provoke sexual violence by how they dress. See Italian National Institute of Statistics (ISTAT) (2018) for the report.

[3] On 7 March 2025, the Italian Council of Ministers approved a draft bill to introduce femicide as a specific criminal offense in the Italian legal system. Parliament must vote on the bill before it can become law

[4] Cfr. UN-WOMEN, Gender-related killings of Women and Girls (Femicide/Feminicide), 2022.

both existing and planned public initiatives. These include targeted communication and educational campaigns to raise awareness, the establishment and strengthening of local centers dedicated to combating gender-based violence, and other localized public interventions.

How to target these interventions in the territory? Are there areas at greater risk for women where public efforts are more required? When anti-violence policies are implemented, are they effective in reducing local episodes of gender-based violence? These are challenging questions to answer for several reasons. First and foremost, there are significant data gaps: systematic and accurate data on femicide cases are uncommon and difficult to obtain, especially at a high level of granular resolution. Second, while the literature on the economic analysis of crime has extensively emphasized the significance of a spatial approach to various types of violent crimes, femicide has primarily been associated with idiosyncratic risk factors. Territorial patterns and systematic determinants have received less attention in existing research, which has mostly focused on individual and family-related drivers.[5] Yet, a geographic approach to the issue is warranted due to the deep cultural roots of this phenomenon and the inherent connection between culture and territory. Therefore, an improved understanding of the risks and root causes of femicides is a necessary precondition to design better and more effective institutional responses. Third, while the phenomenon is widespread at the aggregate level, from a purely technical standpoint, femicides are statistically rare events if analyzed at high spatial resolution levels (e.g., municipalities), making accurate predictions inherently difficult. As a result, it is unsurprising that there is no research specifically aimed at revealing the geography of femicides in the service of policymaking.

In this paper, starting from sparse and unstructured information from several independent, hand-collected data sources, we manually assemble a novel territorial dataset of femicides for Italy covering the entire country for the period from 2006 to 2022. We combine this panel dataset with comprehensive territorial information on demographic, socio-economic, and geographic characteristics of Italian local economies. Unlike most extant data sources, our dataset offers several advantages: i) by focusing on deaths as the primary measure of gender-based violence, it is substantially less likely to be plagued by the known under-reporting issues compared to outcomes employed in other studies, such as helpline calls or police reports (Iyer et al., 2012)[6]; ii) it provides a more granular and accurate geographic localization of the killings; iii) it spans seventeen years, longer than most existing studies; iv) by integrating femicide data with other measures of gender- based violence and newly-collected data on Anti-Violence Centers' (AVCs) location and opening years, it enables analysis of both ex-ante policy targeting and ex-post evaluation.

[5] Outside of economics, there are a few exceptions in the medical literature (Srivastava et al., 2023; VanderEnde et al., 2012; Vyas and Heise, 2016).

[6] Moreover, as the under-reporting is likely to be geographically heterogeneous, a targeting analysis based on helpline calls or police reports would likely be biased.

Italy is a relevant case study due to its high rates of gender-based violence and because it is a country where the issue has recently come under intense media and political scrutiny.[7] Specifically, a femicide case which occurred in Northern Italy in November 2023, resulting in the tragic death of Giulia Cecchettin, a young Italian student brutally killed by her male partner, sparked widespread public outrage, leading to mass protests and calls for state action across the country.

We start by forecasting and mapping femicide risk at the local level. Specifically, on our rich data, we apply Machine Learning (ML) models to forecast the occurrence of femicides. We show that these ML models significantly outperform a benchmark predictor based on the historical records of femicides in each territory. The results also reveal that femicide occurrence is mainly associated with demographic predictors, such as population-related variables, rather than socio-economic ones, and that femicides are significantly less likely to occur in rural areas where female emancipation is generally lower. This finding is consistent with previous works advancing a 'backlash hypothesis' (Bulte and Lensink, 2019; Daniele et al., 2023), i.e., that female economic or political empowerment is positively associated with violence against women. We then use the best-performing among our ML forecasters to construct an indicator of femicide risk for each area that can inform the implementation of place-based policies aimed at combating violence against women. We compare this indicator of risk with the geography of existing AVCs and show that the femicide risk map only partially overlaps with the deployment of existing public support at the local level, especially in higher-risk areas. This finding suggests that the targeting criteria for establishing or strengthening these centers—and enhancing their operational capacity—across the country could be supported and improved through data-driven evidence on the spatial distribution and interaction of femicide risk factors.

Next, we exploit the staggered introduction of AVCs across the Italian territory and use a recently developed non-parametric difference-in-differences design with variation in treatment timing (Imai, Kim, and Wang, 2023) to estimate the impact of the establishment of a new AVC on gender-violence outcomes at the provincial level. Our analysis shows that, on average the local introduction of a new AVC does not significantly decrease the probability of femicide occurrence, possibly due to inadequate targeting of the areas most at risk, as highlighted by the ML analysis. Similarly, it does not have relevant effects on reported abuses against women or stalking, but it reduces the number of reported sexual violence cases by about 20%. We also find that the effects of AVCs on some outcomes (e.g., sexual violence) are more pronounced in high-risk areas identified through our ML forecasting analysis. This suggests that, while AVCs may not have a significant impact on average, they are particularly effective when implemented in red-flag areas more vulnerable to the phenomenon.

[7] See UN statistics here.

We make three main contributions to the existing literature on gender-based violence. First, we are not aware of any other comprehensive study on the socio-economic geography of femicides. Previous studies have focused on very specific aspects, transmission mechanisms, and relationships, such as the role of femicide news in increasing help-seeking behavior (Colagrossi et al., 2023); the effects of subsidized homeownership programs (Lagomarsino and Rossi, 2023), specialized courts (Garcìa-Hombrados et al., 2024), or compulsory schooling policy (Erten and Keskin, 2018; Akyol and Kirdar, 2022) on Intimate Partner Violence (IPV); the effect of female political participation femicides and rapes in the US (Pappa et al., 2022), and on Intimate Partner Femicides (IPF) in Italy (Denti and Faggian, 2022); the link between local support services and the propensity to report sexual violence (Denti and Iammarino, 2022); the distribution of IPF within metropolitan areas (Caporali, 2024); the role of warrantless domestic violence arrest laws (Chin and Cunningham, 2019); the relationship between women's economic empowerment and IPV (Bergvall, 2024); and the impact of COVID-19 lockdown (Arenas-Arroyo et al., 2021; Berniell and Facchini, 2021; Miller, Segal, and Spencer, 2024) and alcohol prohibition (Luca et al., 2015) on gender violence trends. In contrast, our analysis explicitly centers on the spatial dimensions of the phenomenon, utilizing a statistical framework based on a four-step "forecast-map-compare-assess" approach for both ex-ante targeting and ex-post evaluation. Thanks to this unique approach, we can provide fine-grained, non-parametric evidence regarding: i) the territorial patterns of the phenomenon; ii) areas most at risk to target future interventions; and iii) the effectiveness (or lack thereof) of existing local AVCs.

Second, from a methodological viewpoint, we are the first to map femicide risk at a granular level using ML techniques. In this regard, our study also speaks to the recent econometric literature leveraging the use of ML for AI-based policy design (Aiken et al., 2022; Andini et al., 2018; Antulov-Fantulin et al., 2021; Christensen et al., 2024; de Blasio et al., 2022; Carrieri et al., 2021; Johnson et al., 2023). ML algorithms enable us to develop forecasting models that can later be used to assess current policy targeting. In this regard, our paper demonstrates a proof of concept that femicide patterns can be effectively investigated by coupling ML with territorial data. In doing so, we also apply data-mining and rebalancing techniques to address the so-called statistical "rare-event" issue—specifically, the fact that only a very small number of places register at least one femicide in a given year. This methodological approach can be effectively leveraged to forecast other policy-relevant and rare-event problems that are challenging to predict, such as workplace fatalities.

Last, albeit from a territorial perspective, this ML targeting analysis also contributes to the recent and growing micro literature leveraging advanced predictive tools for risk assessment in domestic abuse cases (e.g., Grogger et al., 2021; Yu et al., 2023). This body of work shows that ML algorithms or linear predictive models trained on criminal histories of individuals fare substantially better than naive methods—such as simple Bayes rules or past outcome averages—in predicting the occurrence of domestic abuse. We add to this literature by demonstrating that the same advantage in using ML arises when focusing on local-level, rather than individual-level, prediction of gender violence outcomes, particularly femicides. This reinforces the notion that modern AI predictive tools can be employed at multiple levels (micro as well as meso) to more effectively anticipate and prevent gender violence phenomena.

The rest of this paper is arranged as follows. Section 2 provides a background on gender-based violence concepts and recent trends. Section 3 describes the data used for the analysis. Section 4 presents the methodology and main evidence from the ML targeting analysis. Section 5 illustrates the empirical frame- work, results, the underlying mechanisms and heterogeneity, and robustness of the main causal analysis on AVC openings. Section 6 concludes.

2. Background on Violence Against Women in Italy

The United Nations define gender-based violence as *"any act of gender-based violence that results in, or is likely to result in, physical, sexual, or mental harm or suffering to women, including threats of such acts, coercion or arbitrary deprivation of liberty, whether occurring in public or in private life."*[8] Despite being a long-standing phenomenon, it was only in the 1990s that worldwide institutions began to commit to tackling violence against women (Rights, 2014). It is a global issue: the World Health Organization (WHO) shows that about 30% of women worldwide have been subjected to either physical and/or sexual IPV or non-partner sexual violence in their lifetime (World Health Organization, 2021). Still, the WHO stresses that violence against women is preventable through a combination of support services.

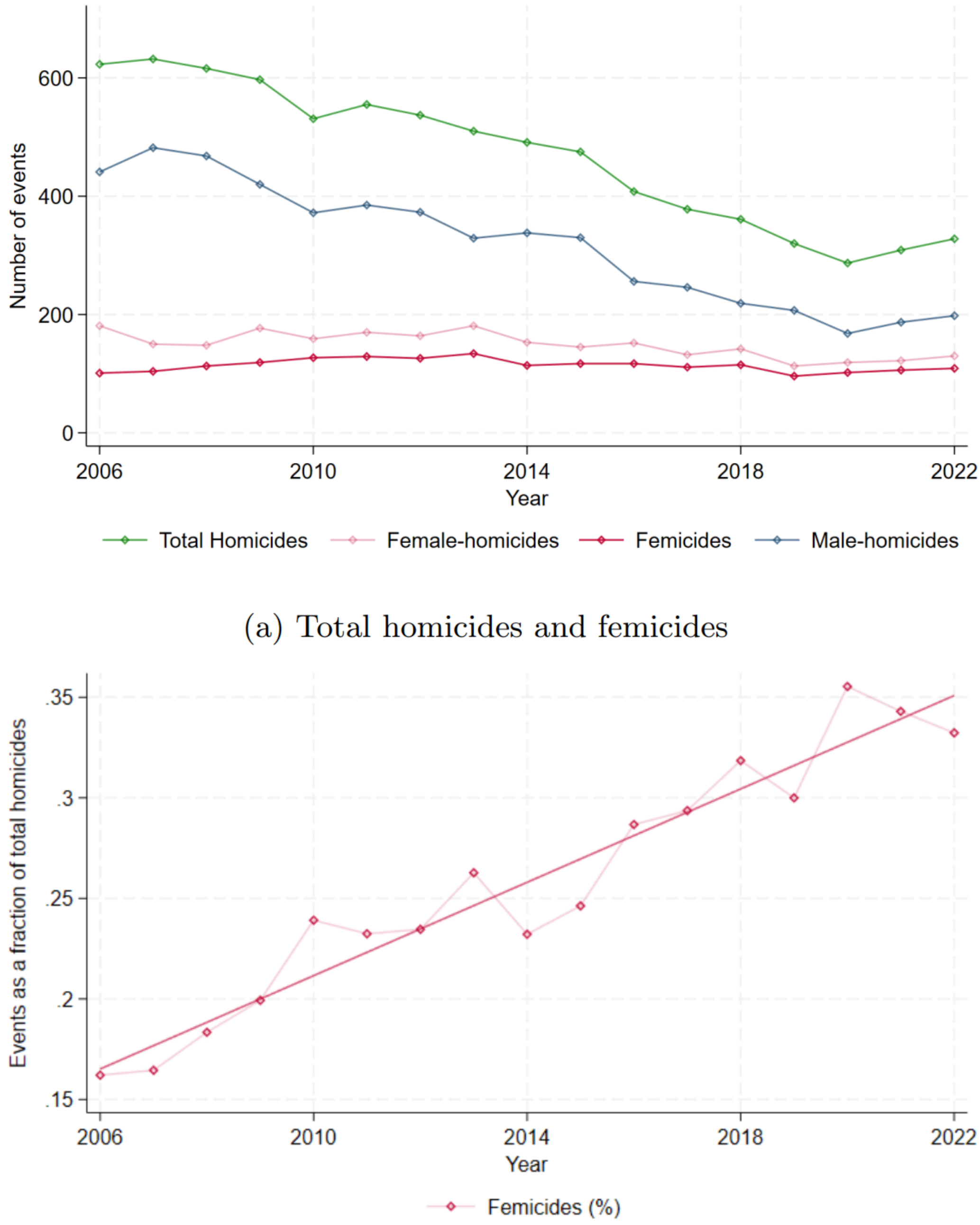
Among developed countries, Italy ranks as one of the safest in terms of general violence. According to the OECD data, Italy's homicide rate (the number of murders per 100,000 inhabitants) is 0.5, lower than the OECD average of 2.6 and comparable to that of its neighboring countries, Austria (0.5) and France (0.4) (OECD, 2020). Moreover, homicides in Italy are decreasing over time. As shown in Figure 1, the total number of homicides has halved in recent years. Nevertheless, these decreasing dynamics are highly gendered: only homicides with male victims have reduced, while homicides with female victims murdered because of their gender—i.e., femicides—have remained constant over time.[9] Indeed, from 2006 to 2022, the share of femicides rose from 0.17 to 0.36 (Figure 1b), indicating that while Italy is overall becoming a safer place, this improvement does not extend to women.

Femicides represent only the most extreme form of gender-based violence. In Italy, almost 80% of women who have suffered from gender-based violence did not report their abuser to the police (European Union Agency for Fundamental Rights, 2014). This means that the majority of cases of violence experienced by women, particularly abuses by partners, go undetected.

[8] Declaration on the Elimination of Violence against Women, Proclaimed by General Assembly resolution 48/104 of 20 December 1993.

[9] Although data from the Ministry of Interior represent the most accurate source in terms of total counts, as they capture the definitive number of women killed, they lack information on perpetrators' motives, making it impossible to distinguish femicides from other female homicides. In our dataset, we specifically identify femicides by extracting information from detailed reports that document the motive behind each killing.

Figure 1: Homicides and femicides over time in Italy



(a) Total homicides and femicides

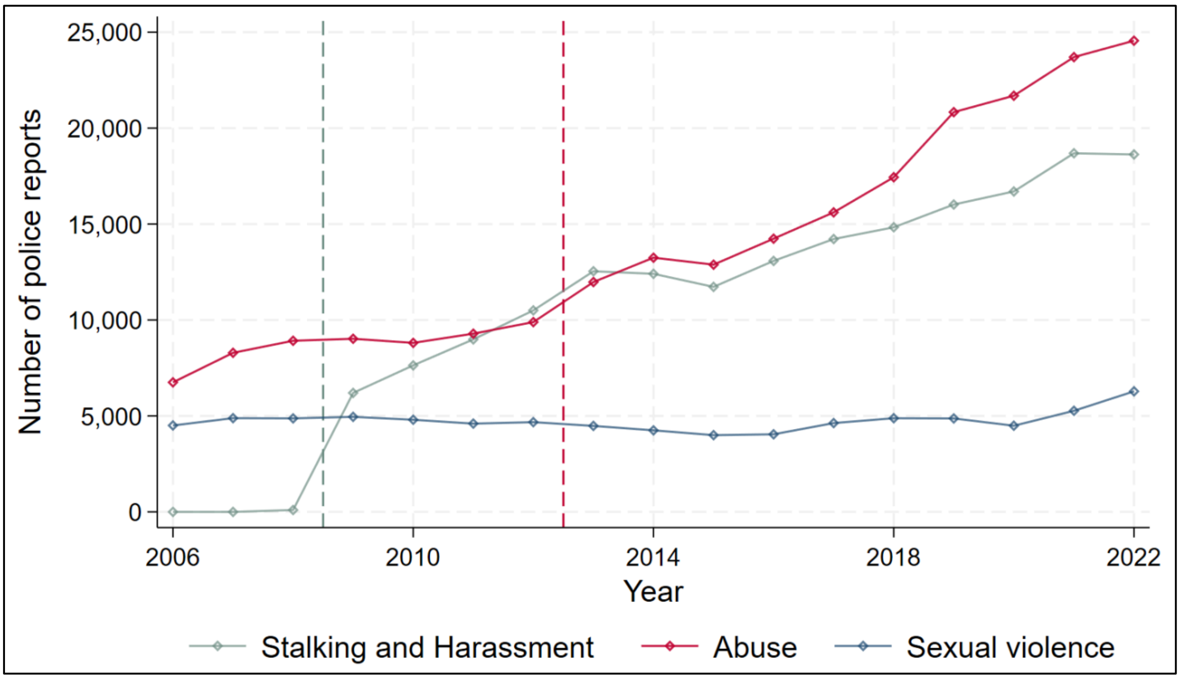
(b) Fraction of femicides

Source: Authors' elaboration of own data (femicides) and data from Italian Ministry of the Interior (other murders).

Reasons for this under-reporting may include victims' fear of retaliation from perpetrators, lack of trust in the judicial system, shame or fear of societal judgment, and other psychological, social, and economic barriers. This phenomenon of under-reporting has severe consequences, as it makes it difficult to obtain an accurate estimate of the severity of gender-based violence and complicates the development of adequate and appropriate responses to prevent it.

This evidence implies that femicide counts can be considered the only indicator for which measurement error in violence against women is minimized. Nevertheless, from a statistical standpoint, femicides are rare events. For this reason, it is still important to complement the analysis of femicides by also looking at other types of crimes against women. Figure 2 shows a significant increase in reports of gender-related crimes in recent years, indicating that awareness campaigns and efforts against gender-based violence have contributed to creating an environment that increasingly encourages women to report without fear. Specifically, the number of stalking reports has been steadily increasing since 2009, the year when stalking was introduced as a crime in Italian law.[10]

Figure 2: Composition of crimes against women



Notes: Vertical bars indicate the introduction of the crime of stalking, and the strengthening of the protection system for victims of violence. Authors' elaboration of data from Italian Ministry of the Interior.

A further increase in the number of reports of stalking and abuse is observed after 2013, with the introduction of D.L. 93/2013, which strengthened the protection system for victims of violence. In fact, in addition to increasing penalties, the law introduced immediately

[10] The crime of stalking was introduced in Italy with Decree-Law 11/2009, which, by inserting Article 612-bis into the Penal Code, provided a more concrete response in the fight against violence against women.

enforceable precautionary measures such as the immediate removal of the perpetrator from the family home, the prohibition to approach places frequented by the victim, and the possibility of monitoring the perpetrators with an electronic tag. Moreover, the D.L. 93/2013 mandated that law enforcement officers immediately report cases of gender-based violence to the public prosecutor and provided for the priority treatment of reports to ensure an immediate response. All these innovations contributed to encouraging women to report their abusers by reducing the risk of retaliation.

Apart from legislative tools, the Council of Europe Convention on Preventing and Combating Violence Against Women and Domestic Violence[11] emphasized the importance of a widespread presence of highly specialized support services such as AVCs, which are crucial to assist women suffering from all forms of violence, both in the short and long term, and to support victims in their journey of recognizing and escaping violence, focusing on raising awareness.[12] In this context, AVCs play a key role. They offer a safe space to talk anonymously and without judgment, provide legal advice and support for police reporting, and offer psychological counselling and job training to encourage labour market participation and economic independence. In the most urgent scenarios, AVCs provide safe and confidential accommodation in shelters for women and their children, protecting them from the abuser, and collaborate with the authorities to ensure adequate security measures, such as issuing protection orders.[13] Additionally, AVCs serve as crucial training hubs for the community, with volunteers and staff often leading training programs at schools and other support facilities. This helps raise awareness about gender-based violence and contributes to fostering community-wide change. Hence, the presence of an adequate network of AVCs across the territories is fundamental, particularly to reach marginalized women, helping them seek help, and raising their awareness about the risks of their situations (CETS No. 210 of the Council of Europe).

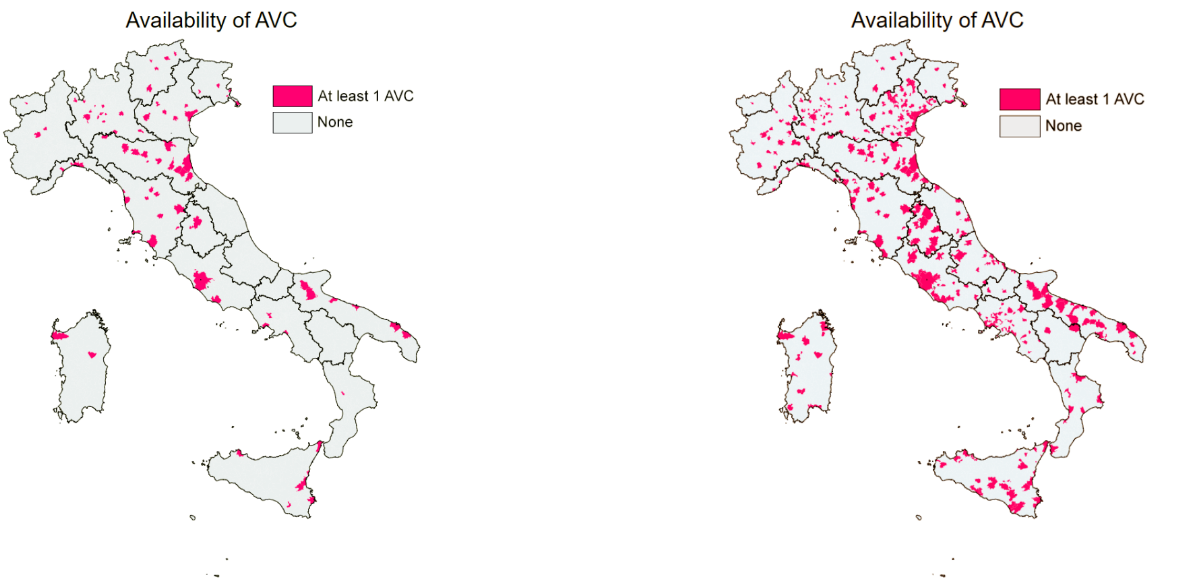
Figure 3 compares the diffusion of AVCs in Italy over the last twenty years. In 2004, there were 104 AVCs, primarily located in central regions (mainly Tuscany and Emilia Romagna), with several regions entirely lacking services against gender-based violence. By contrast, as of June 2024, coverage has significantly increased, particularly in Southern regions, with 457 centers in place and at least one AVC in every Italian province.

[11] This treaty was signed in Istanbul on 11 May 2011. As of April 2025, it has been ratified by 38 countries and the European Union.

[12] Often victims are so threatened by their abusers that they are too scared to seek help (see, for instance, Women's aid).

[13] See, for instance, the definition provided by Refuge, the largest domestic abuse organization in the UK, or "D.i.Re - Donne in Rete contro la violenza", a large net of Italian AVCs.

Figure 3: Presence of Anti-Violence Centers across Italian municipalities



(a) Anti-Violence Centers in 2004 (b) Anti-Violence Centers in 2024

Notes: The data is updated as of June 2024. Authors' elaboration of own data.

3. Data

The empirical analyses are based on the combination of several data sources on: i) femicides; ii) AVCs' locations and opening years; iii) reports to the police on crimes against women; iv) calls to the helpline against gender-based violence; v) socio-economic, territorial, and demographic characteristics.

We collected yearly data on femicides at the municipal level digitizing information from reports provided by the non-governmental organization "*Casa delle Donne per Non Subire Violenza*". These reports, issued annually since 2006, provide detailed information about each femicide incident in Italy, including the name and age of the woman killed, the municipality where the incident occurred, the weapon used, and a brief description of the context in which the femicide took place. A nontrivial challenge in assembling these data was that—especially in earlier report editions—the recorded municipality was often misspelled, identified solely by the frazione (borough) name within the municipality, or replaced by the nearest medium- or large- sized city rather than the actual site of the killing.[14] For this reason, for all femicide cases in our dataset, we complement these data with contextual information drawn from local newspaper articles to retrieve the actual murder municipality and refine the final dataset. Table 1 highlights the importance of this additional cross-check across different sources: it reports that, for two example years for which we conducted this inspection, without this further step, the share of femicides incorrectly assigned to wrong locations would have been extremely high (around 27% on average). This refinement, therefore, allows us to achieve greater precision in the geographic attribution of femicide cases compared to previous literature.

Table 1: Mismatch in the locations of femicides without case-by-case verifications (example years)

	Year	
	2008	2016
Share of mismatched locations	37.86%	16.53%

Notes: This table presents the share of femicides that, in selected sample years, would have been incorrectly assigned to the wrong locations without further refinement based on checks from local newspapers. The data for the initial assignment of locations were extracted from reports provided by the non-governmental organization "Casa delle Donne per Non Subire Violenza."

By crossing these sources, we gathered data on 1,942 femicides that occurred in Italian municipalities over the period 2006-2022. Besides the longer time-span and the aforementioned

[14] For example, a femicide that happened in Terracina is reported in Latina.

geographic refinement, our data differ from previous literature (e.g., Colagrossi et al. (2023)), as we do not restrict our analysis to IPF and consider *all* femicide cases, not just those perpetrated by intimate partners, i.e., IPF. Indeed, although a large part of murderers are current or former partners, we deem it important to also take into account femicides committed by relatives (mostly brothers or fathers), colleagues, or unknown individuals. [15] To give an idea, on average, since 2006, episodes of IPF have accounted for 60% of total femicides, while the other 40% were committed by men with whom the victims did not have, or had previously had, a sentimental relationship.[16] That being said, we also perform our analyses using only IPF data and find consistent results.

We then retrieved information about the location of all AVCs in the Italian territory from the website of the Department for Equal Opportunities. For each of the 457 AVCs, we supplemented the location information with timing data by manually collecting the exact year of opening. This information was derived from multiple sources: regional spokespersons, official AVC websites, local newspapers, and direct contact with AVC directors. In doing so, we took into account that for most AVCs, the *official* information on the opening year was not reliable, as the regional registers consider the *opening date* to be the date of the AVC's official recognition in the national network. It is often the case that an AVC operates effectively for many years in a territory before being included in the network.[17] By digitizing all this information, we created a unique, geolocalized dataset on femicides and AVC locations and openings across Italian municipalities for the period 2006-2022.[18] These variables will be used for both the targeting and causal impact exercises. In the following, we report separately the data sources and features of the other variables used in each analysis.

3.1 Data for the targeting analysis

For the ML-based targeting analysis (detailed in subsection 4.1), we chose the Local Labor Market (LLM) level as the most appropriate unit of spatial analysis. [19] Each LLM is an

[15] Moreover, by looking at all femicides, we can also account for sex-workers murdered by their clients, whose deaths are often overlooked in the news.

[16] Out of 1,942 femicides collected, 61 involved multiple femicides within the same day and the same family unit (typically, partner and children, partner and her sisters, partner and mother). We registered these episodes as single events to avoid over-reporting in a single municipality.

[17] In 2014, Italian Government imposed a minimum standard of activity as a requisite to be eligible for public funding (see Senate Commission of Inquiry into Femicide, 2017).

[18] An other important measure to tackle violence against women is the establishment of women's shelters, which provide safe temporary accommodation for women and children escaping domestic violence. In our analysis, we do not consider the presence of women's shelters because, for safety reasons, it is not possible to retrieve their exact locations or other relevant information

[19] The criteria used to determine Italian LLMs are similar to those used to define Metropolitan Statistical Areas in the US or Travel to Work Areas in the UK.

aggregation of two or more neighboring municipalities, defined by the Italian National Institute of Statistics (ISTAT) based on daily commuting flows from place of residence to place of work. Choosing LLMs instead of municipalities as the unit of spatial analysis is a good compromise between the granularity of the data and the relative rarity of the femicide event. We employ yearly LLM data covering the period from 2006 to 2022.[20] We cover all 610 Italian LLMs, and the outcome is a dummy variable, which is equal to 1 if, in a given year, there is at least one femicide occurrence in the LLM, and 0 otherwise.

The initial pre-treatment information set consists of over 100 variables. They include the industrial structure, macro-regional dummies, labour market characteristics, socio-economic and demographic data (including information about weddings), features of the housing market, information on the electoral turnout and the related gender gap, local politics, and crime rankings and scores. The latter two variables are constructed from a composite of crime indicators compiled by the newspaper Il Sole 24 Ore. All these data are included at the LLM level.[21] See Table A.1 in Appendix A for a detailed description of the variables and their sources.[22] In this set of covariates, we included the first two lags of all the predictors, considering also as predictors the lags of the outcome variable. This implies that we collapse the original 2006–2022 dataset into a dataset covering the period 2008–2022.

3.2 Data for the AVC impact evaluation

The first causal analysis—on the effect of AVC openings—is conducted at the provincial-level due to the sensitive nature of the data, which made it impossible to get some of the key variables at a more disaggregated geographical level.[23] Despite having lower granularity compared to municipal and LLM data, the provincial level still offers relatively high spatial resolution, as Italian provinces (NUTS-3 units) can be considered as roughly comparable to United States counties (Barone et al., 2022). In particular, we use administrative data on reports to the police regarding gender violence. Data are provided by the Department of Public Safety of the Italian Ministry of the Interior. We collect information related to crimes of stalking, abuse

[20] We have collected data for all variables beginning in 2006, except for calls to the 1522 helpline, which are available only from 2013 onward. A separate targeting analysis incorporating 1522 calls as predictors is presented in Appendix C, covering the period 2013–2022.

[21] Crime rankings, and crime scores are collected at the provincial level. For these variables, we assign the crime ranking and score of the corresponding province to each LLM. The number of 1522 helpline calls, also collected at the provincial level, has been distributed based on the population of each LLM within a province.

[22] It would also have been informative to include additional predictors—such as the rate of conscientious objection among gynecologists (Muratori, 2022) and the gender composition of the police force (Miller and Segal, 2019)—but these data were not available at a disaggregated level for the period under analysis.

[23] Over the period considered, the number of provinces has changed slightly, especially in the region of Sardinia. We use the 107 provinces as defined in 2023, with the only exception being the province of Southern Sardinia, which has been aggregated with the province of Cagliari due to the unavailability of separate data for both provinces for all years.

against family members and cohabitants, and sexual violence. For all these data, which cover the period 2006-2022, we have information on whether the victim is a woman. We complement this information with our original data on femicides, distinguishing between general femicides and IPF.

We also include provincial-level data on calls to the national helpline against gender-based violence (1522). Since these data are available only from 2013 onward, we will use them as an alternative outcome variable when analyzing provinces that opened an AVC from 2014 onward. Finally, we have collected demographic (population) and economic (per capita income) variables, which will be used to normalize the outcome measures per 1,000 inhabitants and as additional controls in a robustness check.

4. Machine Learning, Femicide Risk, and Policy Targeting

4.1 Methodology

We build an ML pipeline to forecast the likelihood of femicide occurrence in a given year and LLM. This probability can be interpreted as the risk that at least one femicide will occur in a certain territory in the near future. The primary aim of this exercise is to use machine predictions to assess whether territories at higher risk are sufficiently addressed by existing preventive policies, and to determine if ML-based targeting rules could more effectively guide interventions across territories.

The phenomenon we aim to forecast has several characteristics that are important to bear in mind, making our exercise inherently challenging. First, we are faced with a phenomenon which, although widespread on a national scale, is, from a statistical point of view, a relatively rare event. In fact, among our territorial units (LLMs), the likelihood that at least one femicide occurs in a given unit and year is only 12.44%. Second, while the phenomena are undoubtedly linked to the socioeconomic and cultural characteristics of the territories and communities, it is also characterized by an inherent idiosyncratic component related to the peculiar characteristics of the perpetrators who commit femicides. These characteristics might not necessarily correlate with those of the community to which they belong (and, in some cases, not even with the characteristics of their families). Third, while undoubtedly many femicides are the result of a history of violence and oppression, there remain some unpredictable components that would be impossible to forecast, even if one were to study the phenomenon at the individual level. In this respect, we acknowledge that our aim is not to provide forecasts meant to inform action on an individual case basis. Instead, the information we provide should be used at the same level as the information we use in our analysis, which is on a territorial basis.

First, since we aim to equip the policymaker with an ML tool that forecasts future outcomes based on present data, we include only lagged (rather than contemporaneous) values of the predictors in the ML pipeline. Second, a key departure from the standard ML routine is that we do not use a random split to divide the sample into two disjoint subsamples, namely the training set and the testing set. Instead, we adopt a criterion of non-random split on time to carry out our forecasting task (Cerqua et al., 2024). The rationale for this choice lies in the peculiar nature of panel data: we need to preserve the temporal structure of the data and avoid time-dependent data leakage, i.e., the fact that using a random split would result in having future years in the training set to retrospectively predict the past in the testing set. Therefore, from a forecasting perspective, we use an expanding-window approach for model assessment, selection, and validation, focusing on one-step-ahead forecasts (Hyndman and Athanasopoulos, 2018). This means that in each iteration, for all units in our sample, we use all the information from the present and previous years (with data starting from 2006), to predict the likelihood of the occurrence of at least one femicide in the next year. For example, to predict what happens in year 2019, we train the model to forecast training instances of the outcome based on the predictor information from the period 2006-2017. After training the model, for each area, we plug the predictor data related to year 2018 to forecast femicide occurrence (or lack thereof) in 2019.

To alleviate the rare-event issue (i.e., the strong imbalance of our binary outcome variable), after splitting the dataset into a training and testing set based on the non-random split on time, we artificially rebalance the training set observations using a Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002). SMOTE over-samples the underrepresented class of the outcome variable (femicide occurrence, in our case) to create a larger (1.75 ratio) synthetic but almost perfectly balanced training dataset. In this synthetic training set, the frequency of the variable measuring the likelihood of femicide is 50%. Since we also use categorical variables (such as those related to the province to which the LLMs belong), and the classic version of SMOTE can be problematic when dealing with categorical predictors (Elreedy and Atiya, 2019), we apply a more recent version of SMOTE named SMOTENC, which is specifically designed for this purpose.[24] Importantly, we exclusively apply the outcome rebalancing routine to the training data, allowing our ML models to efficiently learn how to forecast the outcome. However, we test their out-of-sample forecasting ability exclusively on the original testing set, which features the real-world skewed distribution of the outcome variable.

As a benchmark to assess the quality of the performance of the ML models, we create a naive forecaster that is the average historical outcome of femicides in the five years preceding the forecast year for each LLM. This provides a credible and intuitive measure of risk often used in policy reports, serving as our reference to evaluate the usefulness of our empirical exercise.

The rest of the pipeline follows a standard approach where we fine-tune a set of popular ML models, namely random forest, neural networks[25], extreme gradient boosting (XGBoost), bagging, and the more traditional logistic regression (logit). We compare the performance of all these techniques in forecasting femicide occurrence on testing set years using different—and complementary—metrics such as the Area Under the Curve (AUC) of the ROC (Receiving Operator Characteristics), Sensitivity (True Positive Rate), Specificity (True Negative Rate), and Balanced Accuracy (the arithmetic mean of Sensitivity and Specificity). We will use Balanced Accuracy to select the best-performing model, as it provides a more reliable accuracy assessment for imbalanced data.

Finally, to ensure the robustness and reliability of our forecasts and provide statistical significance estimates for the performance metrics, we run 100 iterations using different random seed numbers. We then report average measures across all iterations along with their

[24] See this link for more information about the SMOTENC routine.

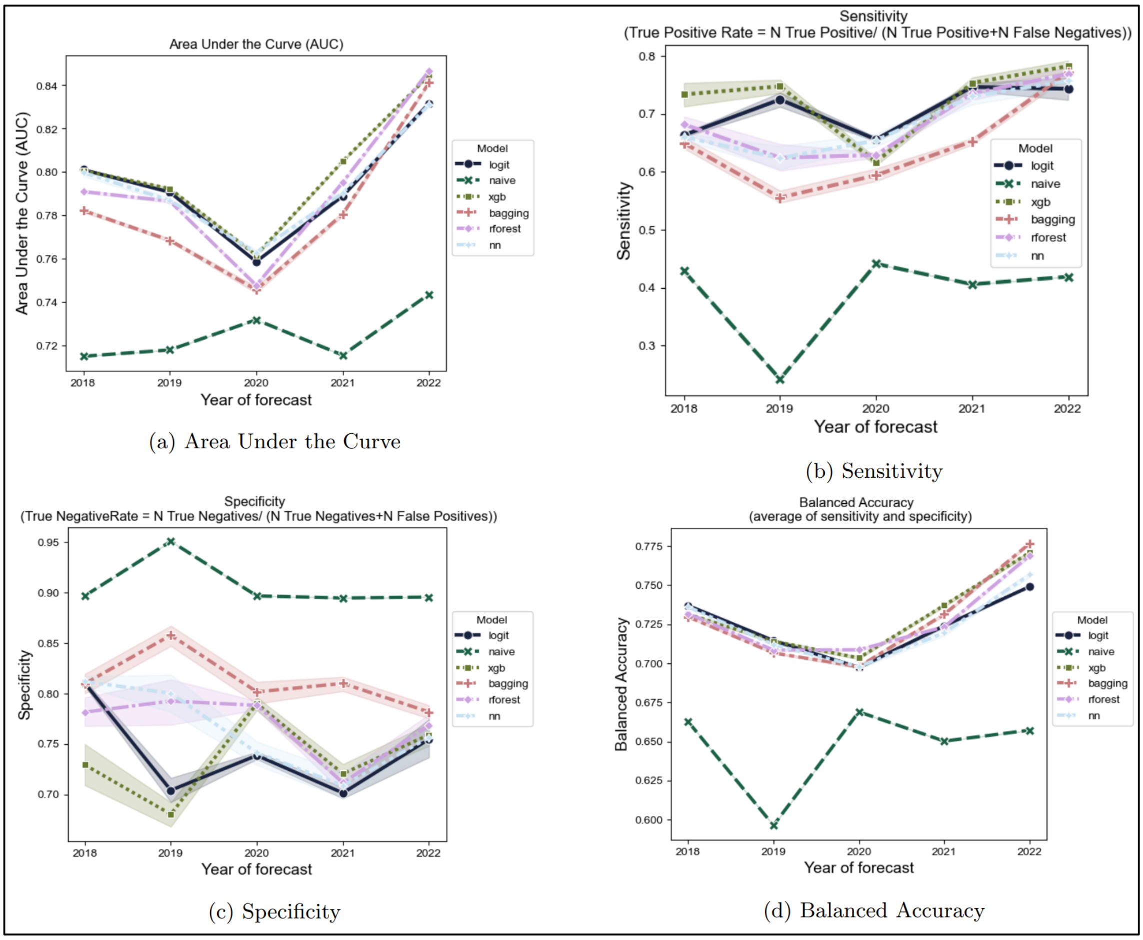
[25] We employ a basic fully connected network made of three layers (64, 32, 1).

estimated standard deviation. For the reader unfamiliar with ML, Appendix B provides more details on the ML algorithms and the performance metrics.

4.2 Main results

The results of our forecasting exercise on the testing set years 2018-2022 are shown in Figure 4.[26]

Figure 4: Performance Metrics of ML Models



(a) Area Under the Curve

(b) Sensitivity

(c) Specificity

(d) Balanced Accuracy

Notes: Performance metrics over 100 iterations with highlighted 95% confidence intervals.

It is evident that all models perform better than the naive model, with the exception of the Specificity metric. Compared to this naive forecaster, a lower value of Specificity of our models implies that: i) we are less likely to forecast correctly the absence of femicides in a given area (fewer true negatives); ii) we are more likely to forecast a femicide in an LLM where, in fact, a femicide did not happen (more false positives). Conversely, a higher value of Sensitivity implies

[26] We report the ROC curves in Figure C.1 in the Appendix.

that we are more likely to correctly forecast a femicide in an LLM where a femicide did happen (i.e., we identify many more true positive instances). In our context, we argue that ceteris paribus, false negatives are substantially costlier than false positives. Thus, Specificity should be less valued than Sensitivity, or, from another perspective, it is relatively better to have a conservative indicator of risk. Furthermore, when we examine the Balanced Accuracy (see Table C.1 in Appendix C), we find that all models perform significantly better than the naive historical predictor.

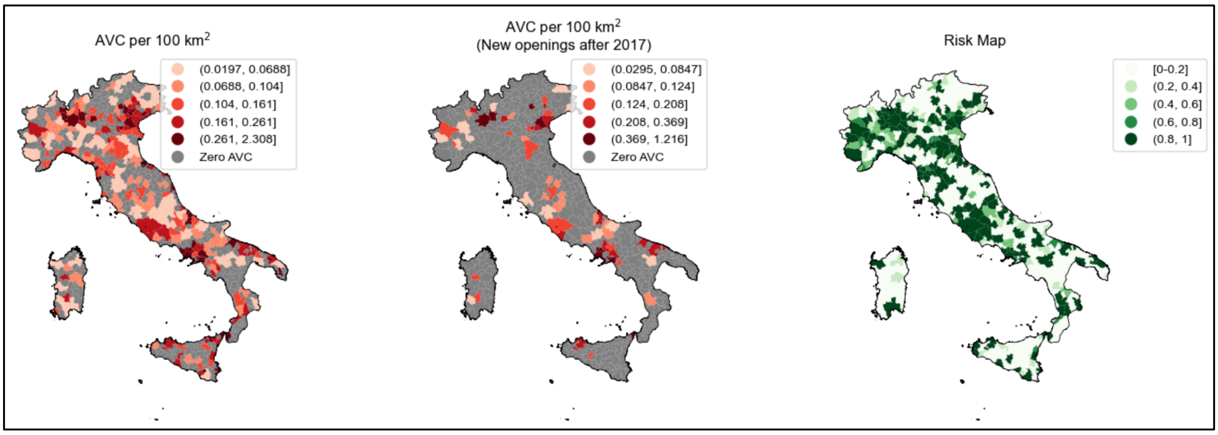
Among our proposed models, XGBoost slightly outperforms the others, although the statistical differences in terms of performance (with the exception of the naive model) are very small. Overall, the good performance of the models documents that, despite our caveats on the inherently complex nature of this social phenomenon and the prominent role of idiosyncratic factors and individual components, predictive models trained on aggregate territorial data can capture a substantial share of the out-of-sample variation and evolution of the phenomenon at the local level. This, in turn, fully supports the adoption of a data-driven territorial approach to analyze femicide risk.

For interpretability, Figure C.2 in the Appendix reports the average SHAP values (Lundberg and Lee, 2017) of the most important predictors for the best-performing XGBoost model (for comparison, we also report those of the logit model in Figure C.3), while Figure C.4 reports the mean absolute SHAP values of the top 10 predictors across all models. Overall, the most important predictors associated with femicides are related to the demographic characteristics of local economies. While this was not unexpected, partly due to the mechanical correlation of femicide occurrence in more populated areas, the prominent role played by predictors such as a dummy for rural areas—associated with a *lower* probability of femicide occurrence despite the inclusion of several population-related predictors—suggests that the traditional view of gender- based violence being more prevalent in close-knit, isolated, and less modern communities is outdated and that, if anything, femicide risk is higher in areas where gender emancipation is generally greater, consistent with the "backlash hypothesis" linking women's emancipation with increased gender-based violence (Bulte and Lensink, 2019; Daniele et al., 2023).

We also note that the XGBoost model and other non-linear models do not consistently outperform the simpler logit model. This is an important finding from the perspective of the real-world usability and applicability of ML models for targeting public policies. In particular, we acknowledge the nontrivial complexity vs. interpretability trade-off in AI-powered policy design and recognize that transparency is a key issue when ML is intended for use in public policy (Athey, 2017). For this reason, the fact that easily interpretable models like logit can perform almost as well as complex and harder-to-interpret learners such as random forest and XGBoost, and even black-box models like neural networks, is significant. It highlights that, should policy- makers be unwilling to leverage the potential of ML forecasting for femicide risk targeting due to concerns about transparency, explainability, and accountability, they might be willing to trade off a bit of forecasting accuracy in exchange for more interpretability and communicability of the targeting policy rule, and choose simpler and more explainable models over more complex ones. At the same time, this does not mean abandoning the ML framework for AI-powered policy design. When used within a predictive framework, rather than causal inference, and relying on the conventional ML pipeline based on training, testing, and out-of-sample prediction aimed at minimizing variance in outcome prediction on unseen data—which is not the case in traditional econometrics, where logit is typically used for in-sample estimation of parameters of interest, with an exclusive focus on reducing bias, even at the cost of overfitting and extreme out-of-sample variance—logit also fully belongs to the family of ML models. Just like Ordinary Least Squares or the LASSO family, it is a basic and parametric ML model based on linear and functional form assumptions (Hastie et al., 2009).

We then construct an ML-based territorial risk measure for femicides based on the aggregation of the binary predictions of all the models, with the exception of the naive model. Specifically, the risk indicator is constructed by averaging the binary predictions of each model over the 5-year out-of-sample forecasting period (2018-2022) and across each ML model.[27]

Figure 5: AVC coverage vs. ML-based femicide risk indicator

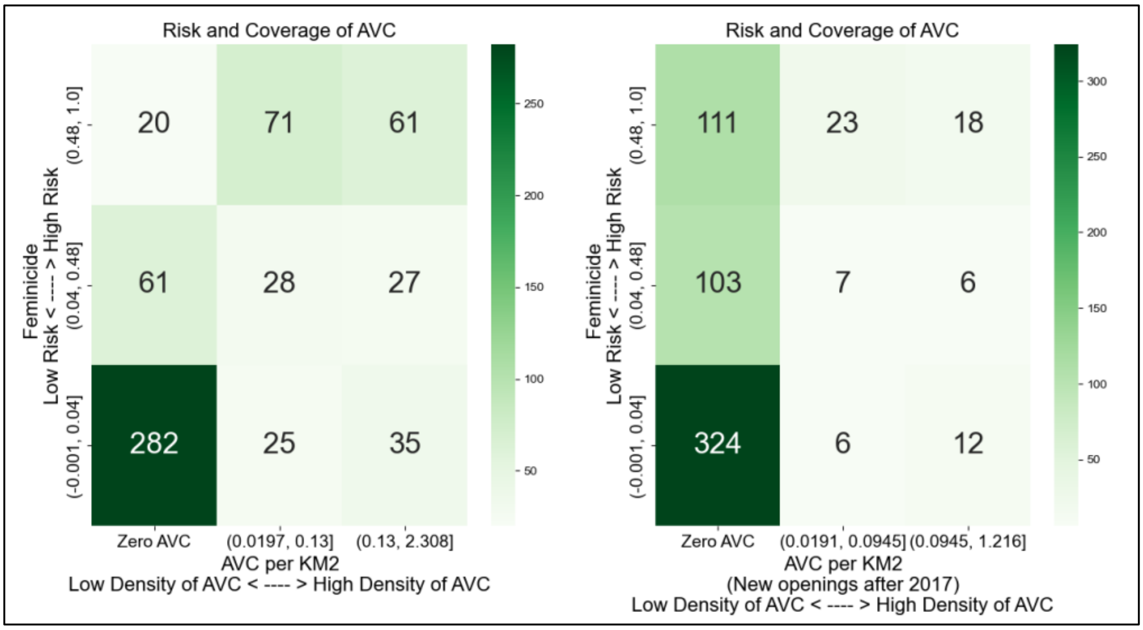


From a targeting perspective, we then compare machine predictions with actual policy implementation at the local level. In Figure 5, we display the coverage of AVCs (in the left panel we consider all AVCs as in 2022, while in the center panel we only consider AVC that opened after 2017) and compare it with the femicide risk indicator described above. While there is a good correspondence between areas at the lowest risk and a scarce or null coverage of AVC, we observe that high-risk areas are much less covered by an adequate presence of AVCs. This result suggests that the coverage of AVCs falls short where they are needed most. This finding

[27] Therefore, within a given LLM, the indicator ranges from 0 to 1, where a value of 1 means that, across all 5 years, every ML model predicted at least one femicide occurrence, while a value of 0 indicates that no ML model predicted any femicide throughout the entire period. We opted to average across all ML models due to the absence of statistically significant differences in performance, as depicted in Table C.1.

is also confirmed by Figure 6, which shows the degree of overlap between the ML-based femicide risk indicator and AVC coverage by tertiles. In the figure, the left panel shows all AVC openings, while the right panel shows only those that occurred between 2018 and 2022. Based on this ML analysis, we conclude that there is significant room to refine policy targeting, also from a cost-effectiveness perspective, by concentrating resources and policy efforts in the areas at greatest risk that are currently not adequately covered.[28]

Figure 6: Openings of AVC vs. ML-based femicide risk indicator (2018-2022)



To further validate this targeting analysis, we conducted several sensitivity analyses. First, we repeated the analysis using a narrower definition of femicides, restricting it to IPF. Second, to include helpline calls to 1522 among the predictors, we re-ran the baseline analysis over a shorter time-span beginning in 2013. Results are strongly consistent with the main findings (Appendix C). This reinforces the notion that current targeting rules for public measures against gender-based violence remain suboptimal, and that local policy interventions are not more prevalent where they are most needed.

[28] We acknowledge that there might be other criteria, beyond femicide risk, that led the policymaker to establish or not establish AVCs in a given area. This is the issue of so-called 'omitted payoffs' (Kleinberg et al., 2018), which is common in the literature on ML for policy targeting. However, we note that while there might be other payoffs for the policymaker that we do not observe, reducing the risk of the most extreme form of gender-based violence is certainly a priority among them.

5. The Causal Effect of AVC Openings

In this section, we explore the effect of local public policies on reported violence against women by exploiting the staggered introduction of AVCs across the Italian territory.

5.1 Methodology

The primary econometric challenge in analyzing the effects of an AVC opening is that provinces with AVC openings may systematically differ from other provinces. To identify suitable counterfactuals for the provinces treated, we employ a recent evaluation technique proposed by Imai, Kim, and Wang (2023), which is a non- parametric generalization of the Difference-in-Differences (DiD) estimator designed specifically for time-series cross-sectional (TSCS) data (e.g., regions, provinces, LLMs).

With this approach, for each province treated *i*, we first select a set of control provinces that did not open a new AVC within a certain number of years (as described below) before and after the year of the AVC opening in province *i*. We then reweight this matched set, denoted as M_i, using the covariate balancing propensity score (CBPS) weighting (Imai and Ratkovic, 2014). This method assigns a higher weight to control provinces within M_i that exhibit greater similarity to the province treated *i* in terms of pre-treatment trends of the outcome and control covariates. Specifically, we use the lagged values of all dependent variables— stalking, abuse against family members and cohabitants, sexual violence, femicides and IPF—as pre-treatment covariates. To account for the varying province sizes, all variables are measured per 1,000 inhabitants. For each province treated, we estimate the counterfactual outcome by computing the weighted average of the control provinces. Finally, we calculate the DiD estimate for each treated province and average these to obtain the average treatment effect on the treated (ATT).

An important step in this procedure involves selecting a non-negative integer F as the number of leads. This value represents the outcome of interest measured at F time periods after the treatment is administered, where F=0 represents the contemporaneous effect. By choosing F = 5 we analyze the treatment effect on the outcome up to five years after the AVC opening. Additionally, we select another non-negative integer L as the number of lags to adjust for. This choice must consider the bias-variance trade-off: while a larger value enhances the credibility of the parallel-trend assumption (PTA)[29], it also reduces the number of untreated provinces, making it more challenging to find matches with similar pre-treatment trends. In our case, we opted for L=5.

[29] Equivalence of pre-trends becomes more demanding with an increase in the number of pre-treatment periods. This makes intuitive sense in the DiD setup, where equivalence of pre-trends in a larger number of periods is regarded as stronger evidence for the plausibility of the PTA (Dette and Schumann, 2024).

We define the ATT as:

$$\delta(F,L) = \mathbf{E}\Big\{ Y_{(i,T+F)}\left(X_{i,t}=1, X_{(i,T-1)}=0, (X_{i,T-l})_{(l=2)}^{L}\right) + \\ -Y_{(i,T+F)}\left(X_{i,T}=0, X_{(i,T-1)}=0, (X_{(i,T-l)})_{(l=2)}^{L}\right) | X_{i,T}=1, X_{i,T-1}=0 \Big\} \tag{1}$$

where the provinces treated are those that opened the AVC, i.e., $X_{(i,T-1)} = 0$ and $X_{(i,T)} = 1$.

In this definition $Y_{(i,T+F)}\left(X_{i,t}=1, X_{(i,T-1)}=0, (X_{i,T-l})_{(l=2)}^{L}\right)$ is the (observed) potential outcome under a treatment change, whereas $Y_{(i,T+F)}(X_{i,T}=0, X_{i,T-1}=0, (X_{i,T-l})_{(l=2)}^{L})$ represents the potential outcome without the AVC opening. In both cases, the resto g the treatment history, i.e., $\left(X_{(i,T-l)l=2}^{L}\right) = \left\{X_{i,T}, \ldots, X_{i,T-L}\right\}$, is set to the realized history. In our case, $\delta(5, 5)$ represents the average causal effect of an AVC opening on the outcome five years after the treatment, assuming that the potential outcome depends on the treatment history up to five years earlier. Given that the data are available from 2006 to 2022, we only consider as treated those provinces where the first AVC opening occurred between 2011 and 2017, or where an additional (in most cases, the second) AVC was opened during the same period, provided there were no AVC openings in the previous five years. See Table A.2 in the Appendix for more details on the 22 provinces treated. Simultaneously, for each province treated, only provinces with no AVC openings in the five years before and after the year of the AVC opening of province *i* will be included in the matched set M_i.

We then compute the DiD estimate of the ATT for each treated observation and then average it across all treated observations. Formally:

$$\hat{\delta}(F,L) = \frac{1}{\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{i,t}} \sum_{i=1}^{N}\sum_{t=L+1}^{T-F} D_{i,t} \Big\{ \left(Y_{i,T+F} - Y_{i,T-1}\right) + \\ - \sum_{i' \in M_i} \omega_{it}^{i'} \left(Y_{i',T+F} - Y_{i',T-1}\right) \Big\} \tag{2}$$

Where $D_{i,t}$ is the treatment dummy. The non-parametric generalization of the DiD estimator is based on three assumptions.

Assumption 1. Absence of the carryover effect. This implies that the potential outcome for province *i* at time *T+F* does not depend on the previous treatment status of the same province after *L* time periods. In other words, we allow for the possibility that past treatments affect future outcomes up to *L* years.

Assumption 2. Absence of interference. The potential outcome for province *i* at time *T+F* does not depend on the treatment status of other provinces.

Assumption 3. PTA holds after conditioning on the treatment, outcome, and covariate histories.

$$\begin{aligned} &\mathbf{E}\big[Y_{i,T+F}\big(X_{i,T}=0, X_{i,T-1}=0, (X_{i,T-l})_{l=2}^{L}\big) \\ -Y_{i,T-1} \mid X_{i,T}=1, X_{i,T-1}=0, &(X_{i,T-l}, Y_{i,T-l})_{l=2}^{L}, (Z_{i,T-l})_{l=0}^{L}\big] = \\ &= \mathbf{E}\big[Y_{i,T+F}\big(X_{i,T}=0, X_{i,T-1}=0, (X_{i,T-l})_{l=2}^{L}\big)+ \\ -Y_{i,T-1} | X_{i,T}=0, X_{i,T-1}=0, &(X_{i,T-l}, Y_{i,T-l})_{l=2'}^{L} (Z_{i,T-l})_{l=0}^{L}\big] \end{aligned} \tag{3}$$

where $Z_{i,t}$ is a vector of observed time-varying confounders for province i at year t. Therefore, the conditioning set includes the treatment history, the lagged outcome, and covariate history, which in our empirical analysis corresponds to the other dependent variables. Choosing a relatively large value of L (in our case L = 5) increases the credibility of a limited carryover effect and of the PTA.

This set of assumptions is less stringent than those employed by the most common methodologies for analyzing TSCS data, such as the two-way fixed effects estimator (TWFE), dynamic panel models, matching methods, and the DiD estimator (Imai, Kim, and Wang, 2023).[30] Additionally, unlike other recently proposed estimators (e.g., the DiD with multiple time-period estimators by De Chaisemartin and d'Haultfoeuille (2024) and Callaway and Sant'Anna (2021)), this approach explicitly tests for pre-treatment trend differences in all covariates and does not rely on parametric assumptions. Nonetheless, in Appendix D we show that our results remain robust when using alternative matching and weighting methods to refine the matched set of control units, as well as De Chaisemartin and d'Haultfoeuille (2024) and Callaway and Sant'Anna (2021) estimators (see Table D.1).

Before presenting the estimates, it is important to assess the credibility of Assumptions 2 and 3 to guarantee the validity of the empirical analysis. Assumption 2 would be violated if women were able to move across provinces to attend an AVC. We believe this not to be concerning in our setting. Indeed, it is unlikely that a woman suffering abuse or violence would travel to another province to regularly attend an AVC and then report the crime in that province. The support provided by AVCs typically requires frequent sessions over time, making cross-province travel impractical for most victims. Therefore, the potential outcome of a province is unlikely to be affected by the availability of AVCs in neighboring provinces. As for Assumption 3, the non-parametric generalization of the DiD estimator allows examination of the covariate balancing between treated and matched control observations. This enables investigation of whether the treated and matched control observations are comparable with respect to the pre-treatment trends of the variables of interest (Imai, Kim, and Wang, 2023). The covariate balance plot is reported in Figure 7. Each line represents the balance of each main

[30] Recent literature has shown that estimates from TWFE models in which the timing of policy changes are staggered over time can be biased when the treatment effect is heterogeneous (see, among others, Callaway and Sant'Anna (2021); De Chaisemartin and d'Haultfoeuille (2024); Imai and Kim (2021)).

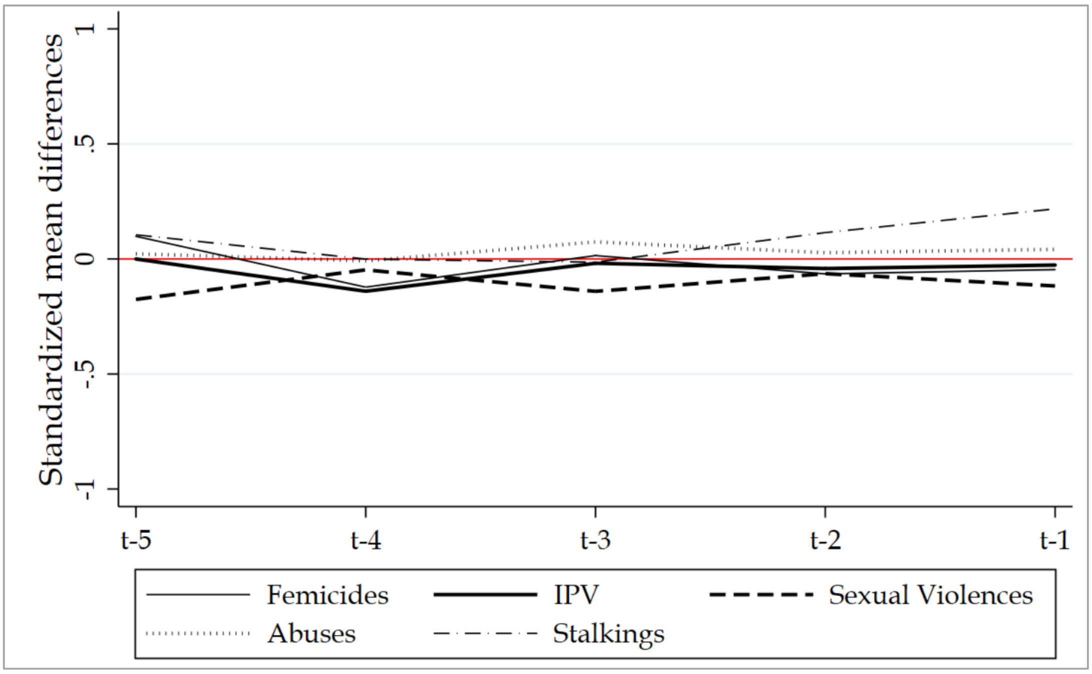
lagged outcome. It clearly emerges that the level of imbalance remains stable across the 5 pre-treatment years and fully within the (-1, 1) range of the standard deviation. It is important to note that we control for the same covariates (the five dependent variables) across all empirical analyses, meaning that the covariate balance plot presented in Figure 1 is the same for all the analyses presented below. This figure provides evidence supporting the credibility of Assumption 3 in our analysis.

5.2 Main results

Figure 8 illustrates the pre- and post-treatment evolution for both treated and control provinces (after the CBPS reweighting) for all dependent variables. The vertical black bars represent 95% confidence intervals for the estimated counterfactual outcome. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)). This graphical representation follows the approach outlined by Cattaneo et al. (2025).

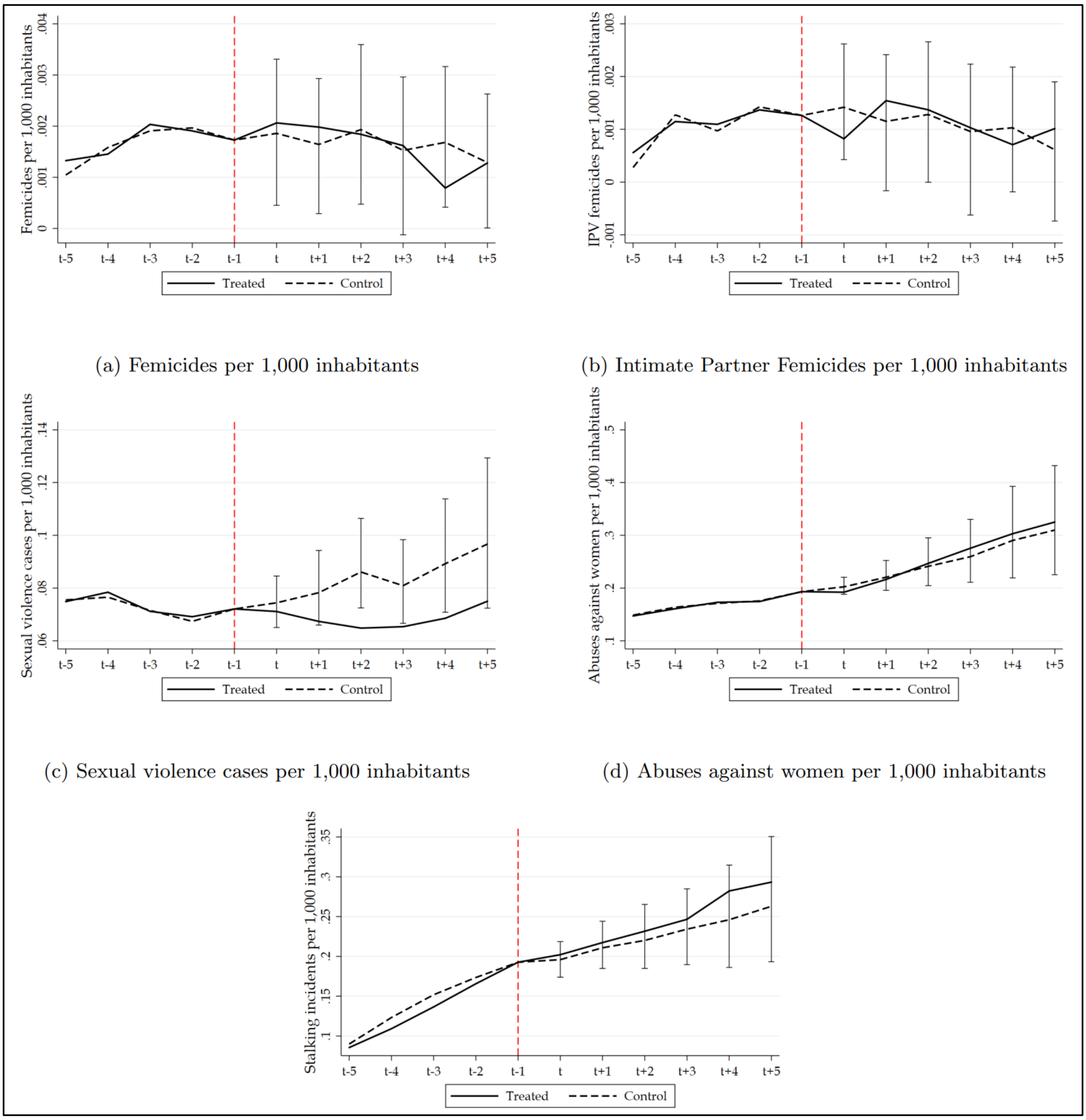
Panel (a) shows that the opening of a new AVC at the local level leads to a slight reduction in the probability of femicide occurrence; however, this impact is not statistically significant. When considering only IPF, the observed reduction disappears, suggesting that new AVC openings have no differential effect on general femicides compared to those committed in intimate partner relationships. These results are consistent with the findings of Garcìa-Hombrados et al. (2024), who find no significant effects of specialized IPV courts on IPF, suggesting that existing institutional interventions may face limitations in preventing the most severe forms of gender-based violence. In contrast, Panel (c) demonstrates a more pronounced reduction in the number of reported sexual violence incidents, with statistically significant estimates (at the 5% level) at times $t + 2$, $t + 3$, and $t + 4$. Conversely, Panels (d) and (e) show a slight increase in the number of reported abuses against women and stalking incidents in provinces with an AVC opening, but none of these estimates is statistically significant.

Figure 7: Balance of covariates



We perform several robustness checks to assess the sensitivity of our results (see Appendix D, Table D.1). First, we use propensity score matching (PSM) (Rosenbaum and Rubin, 1983), Mahalanobis distance matching, and the inverse propensity score weighting (IPW) method (Hirano et al., 2003) to provide different matching and weighting methods to refine the matched set of control units. Second, we check whether additional covariates affect our estimates by adding population and per capita income to the vector of observed time-varying confounders. Third, we use the the dynamic version of the DiD with multiple time periods estimators developed by De Chaisemartin and d'Haultfoeuille (2024) and the semi-parametric DiD estimator proposed by Callaway and Sant'Anna (2021). These results generally confirm those of the main analysis.

Figure 8: Main estimates



(a) Femicides per 1,000 inhabitants

(b) Intimate Partner Femicides per 1,000 inhabitants

(c) Sexual violence cases per 1,000 inhabitants

(d) Abuses against women per 1,000 inhabitants

Notes: The vertical black bars report 95% confidence intervals for the outcome of the estimated counterfactual scenario. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)).

We interpret our results through the lens of two counterbalancing mechanisms. The establishment of AVCs likely generates both a *reporting* effect (increasing victims' propensity to report crimes) and a *violence reduction* effect (decreasing the actual incidence of violence). On one hand, a new AVC provides support to women experiencing abuse, violence, and stalking by helping them recognize the crimes committed against them and encouraging formal reporting to authorities. As a matter of fact, a distinctive characteristic of such abuses is that victims often endure such behavior for extended periods, leading to normalization that obscures their awareness of the situation's severity. In this context, the opening of a new AVC in the area enables abused and stalked women to recognize their victim status and take concrete steps toward reporting their abusers—what we conceptualize as the *reporting* effect. On the other hand, the opening of an AVC contributes to building greater awareness of the patriarchal culture through intensive training activities addressed to local institutions and society. This, in turn, should help to reduce extreme acts of violence against women, such as sexual violence. We call this a *violence reduction* effect.[31]

The presence of this effect is supported by the observed reduction in reported sexual violence and by suggestive—but not statistically significant—evidence of a decrease in femicides. This reduction could theoretically stem from two alternative mechanisms: either the opening of a new AVC inhibits women's willingness to report sexual violence, or a new opening contributes to reducing—even slowly and with a lagged effect— violence presence. While the former explanation is unlikely given the centers' mission and protocols, the latter is reasonable and consistent with the documented experiences of AVC staff and volunteers.[32]

In this context, the null effects observed for reported abuses and stalking can be understood as the consequence of these opposing mechanisms—the *reporting* effect and the *violence reduction* effect—approximately offsetting each other, resulting in no detectable net change in reported crimes. Two additional considerations enhance our interpretation of these findings. First, abuse and stalking crimes are more often associated with IPV cases, where the incentive to report is weaker. Second, the slight effect on the number of reported abuses against women and stalking incidents aligns with the Italian AVCs' reported experiences: indeed, as claimed by "D.i.Re - Donne in Rete contro la violenza", a large net of Italian AVCs, only 28% of the assisted women decide to pursue legal action. One reason behind this evidence is the so-called *institutional secondary victimization*.[33] Indeed, institutions that engage directly with women—such as social services, law enforcement, and the judiciary—should work to further strengthen trust-building processes that help reassure women who seek to pursue justice. For this kind of mechanism, Italian institutions have been condemned by the European Court of Human Rights and the European Parliament.[34]

[31] We thank an anonymous referee for their suggestions regarding the formalization of the effects.

[32] Volunteers reported this experience to us in direct conversations during data collection and in interviews conducted to gain a comprehensive understanding of the support network's functioning.

[33] For instance, after a complaint is filed, mothers frequently face charges of *parental alienation*, which threatens their ability to obtain custody or limits their parental rights.

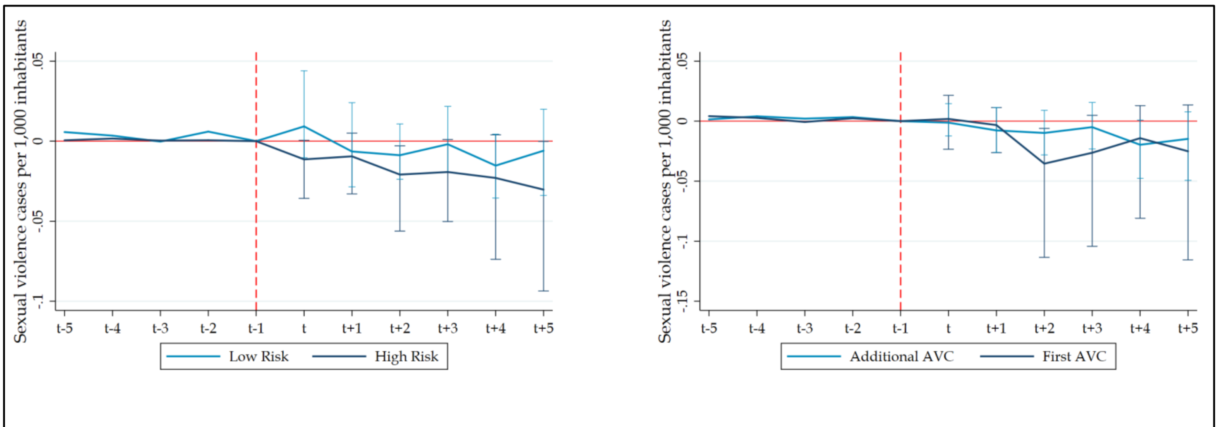
[34] For a deeper explanation of these mechanisms, see the D.i.Re Annual Report and the Report on Institutional victimization

Finally, we explore the effect of AVC openings on helpline calls. Results show that the openings have no significant effect on helpline calls (Figure D.1).[35] Still, the insignificant effect is unsurprising: the typical sequence of events operates in the opposite direction: victims generally contact the 1522 helpline for initial support, after which helpline operators recommend visiting local AVCs. Once a woman accesses an AVC and becomes integrated into a structured support pathway, there is diminished need to call the helpline, as the 1522 service typically functions as an entry point rather than a secondary resource in the support process. Furthermore, awareness of the 1522 helpline is primarily generated through television campaigns, social media, pharmacies, and supermarkets, making it unreasonable to expect increased helpline visibility as a consequence of AVC openings within a given area.

5.3 Heterogeneity analysis

In this subsection, we explore possible mechanisms behind these effects along two dimensions. First, the effectiveness of local interventions may vary according to the area-specific risk level. Second, AVC openings may have diminishing returns, meaning that initial positive effect decreases as additional AVCs are established

Figure 9: Heterogeneity - Sexual violence cases per 1,000 inhabitants



(a) High risk *vs* low risk areas

(b) First vs. already operating AVC

Notes: The vertical black bars report 95% confidence intervals for the outcome of the estimated counterfactual scenario. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)). Panel A: Treated units are those with the risk indicator above/below the median. The risk indicator is computed as the population-weighted average of the LLM risk indicators in each province. Panel B: The number of provinces with no pre-existing AVCs before receiving the treatment is 8, while those with one or more AVCs already operating is 14.

In subsection 4.2 we document the suboptimal spatial distribution of AVCs relative to areas with great risk exposure, as measured by our risk indicator. This misalignment raises the

[35] A caveat applies: data on helpline calls are only available starting from 2013, meaning that we have 12 treated provinces from 2014 onward to have at least one pre-treatment year.

question of whether AVC effectiveness varies with local risk characteristics. To investigate this potential mechanism, we split our sample by the median value of our risk indicator. Figure 9a shows that the impact on sexual violence crimes is greater in high-risk areas. Results for other outcomes are reported in Appendix D, Figure D.2. This heterogeneity in treatment effects points out the importance of accurately mapping the spatial distribution of gender-based violence risk to optimize resource allocation and intervention targeting.

To test the second mechanism, we divided our sample of treated provinces into two groups: those with no pre-existing AVCs before receiving the treatment, and those with one or more AVCs already operating. We then replicated our estimates for all five outcomes. The results (reported in Appendix D, Figure D.3) reveal no significant differences between these two groups of provinces regarding total femicides, abuses, and stalking crimes. However, we observe a slight decrease in IPF immediately after the opening in provinces with no pre-existing centers. Moreover, the significant reduction in sexual violence cases shown in Figure 8c appears to be largely driven by provinces experiencing their first AVC opening (Figure 9b). This indicates that the effectiveness of AVC openings is greater in areas previously lacking support services for violence victims.

6. Conclusions

The fight against gender violence is increasingly moving to the core of the political agenda, but implementing effective strategies requires stronger efforts. Still, eradicating femicides and, more generally, gender-based violence remains a mounting task, due to the intrinsic complexity of the problem but also the incomplete awareness and understanding regarding root causes, risk factors, and potential policy remedies specific to

this social problem.

In this paper, we have contributed to filling this knowledge gap by conducting a comprehensive study on the territorial patterns associated with femicides, gender-based violence, and local policy interventions. First, we forecasted and mapped femicide risk at the local level using ML models and found evidence consistent with the backlash hypothesis, according to which femicides are less likely to occur where female emancipation is lower. Our indicator of femicide risk indicates ample room to improve current policy efforts against gender- violence, supporting the use of data-driven targeting criteria to enhance place-based policy planning. Second, we studied the effect of AVC openings on femicides and other felonies against women with a staggered non- parametric DiD approach. We assessed that the opening of a new AVC results in a weak decrease in femicides, but a large and significant reduction in sexual violence.

Our findings have important policy implications. The establishment of AVCs follows a bottom-up approach, with centers often initiated through local volunteer efforts but subsequently funded through decisions made at higher government levels. Our targeting exercise shows that relying only on bottom-up initiatives for AVC location is suboptimal—a more effective strategy would involve opening new centers in areas identified as high-risk through data-driven methods. Moreover, the limited impact of AVCs on femicide reduction suggests that these local support structures should be complemented with innovative programs specifically targeting male behavior patterns and addressing the root causes of gender-based violence. This integrated approach would leverage the strengths of local support networks while addressing the systemic factors that perpetuate the most extreme forms of gender-based violence. For instance, this approach could complement the "National Strategic Plan to Combat Violence Against Women and Domestic Violence 2025–2027", which was approved by the Minister for family, birth rate, and equal opportunities by decree on 16 September 2025.

Overall, the results reveal that a holistic territorial approach to the problem yields valuable insights for both targeting and evaluating policy interventions aimed at combating violence against women. This, in turn, underscores that the role played by systemic, community-level, and local factors in shaping the evolution and heterogeneity of the phenomenon and the associated policy remedies is a key avenue to explore for future research.

References

Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock (2022). "Machine learning and phone data can improve targeting of humanitarian aid". In: *Nature* 603.7903, pp. 864–870.

Akyol, Pelin and Murat Guray Kırdar (2022). "Compulsory schooling reform and intimate partner violence in Turkey". In: *European Economic Review* 150, p. 104313.

Andini, Monica, Emanuele Ciani, Guido de Blasio, Alessio D'Ignazio, and Viola Salvestrini (2018). "Targeting with machine learning: An application to a tax rebate program in Italy". In: *Journal of Economic Behavior & Organization* 156, pp. 86–102.

Antulov-Fantulin, Nino, Raffaele Lagravinese, and Giuliano Resce (2021). "Predicting bankruptcy of local government: A machine learning approach". In: *Journal of Economic Behavior & Organization* 183, pp. 681–699.

Arenas-Arroyo, Esther, Daniel Fernandez-Kranz, and Natalia Nollenberger (2021). "Intimate partner violence under forced cohabitation and economic stress: Evidence from the COVID-19 pandemic". In: *Journal of Public Economics* 194, p. 104350.

Athey, Susan (2017). "Beyond prediction: Using big data for policy problems". In: *Science* 355.6324, pp. 483– 485.

Barone, Guglielmo, Fabiano Schivardi, and Enrico Sette (2022). "Interlocking directorates and competition in banking". In.

Bergvall, Sanna (2024). "Women's economic empowerment and intimate partner violence". In: *Journal of Public Economics* 239, p. 105211.

Berniell, Inés and Gabriel Facchini (2021). "COVID-19 lockdown and domestic violence: Evidence from internet-search behavior in 11 countries". In: *European Economic Review* 136, p. 103775.

Botto, Matteo and Lucas Gottzen (2024). "Swallowing and spitting out the red pill: Young men, vulnerability, and radicalization pathways in the manosphere". In: *Journal of Gender Studies* 33.5, pp. 596–608.

Bulte, Erwin and Robert Lensink (2019). "Women's empowerment and domestic abuse: Experimental evidence from Vietnam". In: *European Economic Review* 115, pp. 172–191.

Callaway, Brantly and Pedro HC Sant'Anna (2021). "Difference-in-differences with multiple time periods". In: *Journal of Econometrics* 225.2, pp. 200–230.

Caporali, Carlo (2024). "Local Determinants of Violence". PhD Thesis. Unpublished doctoral thesis.

Carrieri, Vincenzo, Raffele Lagravinese, and Giuliano Resce (2021). "Predicting vaccine hesitancy from area-level indicators: A machine learning approach". In: *Health Economics* 30.12, pp. 3248–3256.

Cattaneo, Matias D, Yingjie Feng, Filippo Palomba, and Rocio Titiunik (2025). "Uncertainty quantification in synthetic controls with staggered treatment adoption". In: *Review of Economics and Statistics*, pp. 1– 46.

Cerqua, Augusto, Marco Letta, and Fiammetta Menchetti (Oct. 2024). "Causal Inference and Policy Evaluation without a Control Group". Available at SSRN: https://ssrn.com/abstract=4315389 or http://dx.doi.org/10.2139/ssrn.4315389.

Chawla, Nitesh V, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer (2002). "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16, pp. 321–357.

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. San Francisco, California, USA: ACM, pp. 785–794. isbn: 978-1-4503-4232-2. doi: 10.1145/2939672.2939785. url: http://doi.acm.org/10.1145/2939672.2939785.

Chin, Yoo-Mi and Scott Cunningham (2019). "Revisiting the effect of warrantless domestic violence arrest laws on intimate partner homicides". In: *Journal of Public Economics* 179, p. 104072.

Christensen, Peter, Paul Francisco, Erica Myers, Hansen Shao, and Mateus Souza (2024). "Energy efficiency can deliver for climate policy: Evidence from machine learning-based targeting". In: *Journal of Public Economics* 234, p. 105098.

Colagrossi, Marco, Claudio Deiana, Davide Dragone, Andrea Geraci, Ludovica Giua, and Elisa Iori (2023). "Intimate partner violence and help-seeking: The role of femicide news". In: *Journal of Health Economics* 87, p. 102722.

Daniele, Gianmarco, Gemma Dipoppa, and Massimo Pulejo (2023). "Attacking women or their policies? Understanding violence against women in politics". In: *Understanding Violence against Women in Politics (September 29, 2023). BAFFI CAREFIN Centre Research Paper* 207.

de Blasio, Guido, Alessio D'Ignazio, and Marco Letta (2022). "Gotham city. Predicting 'corrupted' municipalities with machine learning". In: *Technological Forecasting and Social Change* 184, p. 122016.

De Chaisemartin, Clément and Xavier d'Haultfoeuille (2024). "Difference-in-differences estimators of intertemporal treatment effects". In: *Review of Economics and Statistics*, pp. 1–45.

Denti, Daria and Alessandra Faggian (2022). *The Councilwoman's Tale: Countering Intimate Partner Homicides by electing women in local councils*. Tech. rep. 31. London, UK: Department of Geography and Environment, LSE.

Denti, Daria and Simona Iammarino (2022). "Coming Out of the Woods. Do local support services influence the propensity to report sexual violence?" In: *Journal of Economic Behavior & Organization* 193, pp. 334– 352.

Dette, Holger and Martin Schumann (2024). “Testing for equivalence of pre-trends in Difference-in-Differences estimation”. In: *Journal of Business & Economic Statistics*, pp. 1–13.

Elreedy, Dina and Amir F Atiya (2019). “A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance”. In: *Information Sciences* 505, pp. 32–64.

Erten, Bilge and Pinar Keskin (2018). “For better or for worse? Education and the prevalence of domestic violence in turkey”. In: *American Economic Journal: Applied Economics* 10.1, pp. 64–105.

Garcìa-Hombrados, Jorge, Marta Martìnez-Matute, and Carmen Villa (2024). “Specialised courts and the reporting of intimate partner violence: Evidence from Spain”. In: *Journal of Public Economics* 239, p. 105243.

Grogger, Jeffrey, Sean Gupta, Ria Ivandic, and Tom Kirchmaier (2021). “Comparing conventional and machine-learning approaches to risk assessment in domestic abuse cases”. In: *Journal of Empirical Legal Studies* 18.1, pp. 90–130.

Hastie, Trevor, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer.

Hirano, Keisuke, Guido W Imbens, and Geert Ridder (2003). “Efficient estimation of average treatment effects using the estimated propensity score”. In: *Econometrica* 71.4, pp. 1161–1189.

Hyndman, Rob J and George Athanasopoulos (2018). *Forecasting: principles and practice*. OTexts.

Imai, Kosuke and In Song Kim (2021). “On the use of two-way fixed effects regression models for causal inference with panel data”. In: *Political Analysis* 29.3, pp. 405–415.

Imai, Kosuke, In Song Kim, and Erik H Wang (2023). “Matching methods for causal inference with time-series cross-sectional data”. In: *American Journal of Political Science* 67.3, pp. 587–605.

Imai, Kosuke and Marc Ratkovic (2014). “Covariate balancing propensity score”. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 76.1, pp. 243–263.

Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova (2012). “The power of political voice: women’s political representation and crime in India”. In: *American Economic Journal: Applied Economics* 4.4, pp. 165–193.

Johnson, Matthew S, David I Levine, and Michael W Toffel (2023). “Improving regulatory effectiveness through better targeting: Evidence from OSHA”. In: *American Economic Journal: Applied Economics* 15.4, pp. 30–67.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2018). “Hu- man decisions and machine predictions”. In: *The quarterly journal of economics* 133.1, pp. 237–293.

Lagomarsino, Bruno Cardinale and Martin A Rossi (2023). "JUE insight: The unintended effect of Argentina's subsidized homeownership lottery program on intimate partner violence". In: *Journal of Urban Economics* 103612.

Lemaitre, Guillaume, Fernando Nogueira, and Christos K. Aridas (2017). "Imbalanced-learn: A Python Tool- box to Tackle the Curse of Imbalanced Datasets in Machine Learning". In: *Journal of Machine Learning Research* 18.17, pp. 1–5. url: http://jmlr.org/papers/v18/16-365.html.

Luca, Dara Lee, Emily Owens, and Gunjan Sharma (2015). "Can alcohol prohibition reduce violence against women?" In: *American Economic Review* 105.5, pp. 625–629.

Lundberg, Scott M and Su-In Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Curran Associates, Inc., pp. 4765–4774. url: http: //papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Martìn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Man´e, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Vi´egas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. url: https://www.tensorflow.org/.

Miller, Amalia R and Carmit Segal (2019). "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence". In: *The Review of Economic Studies* 86.5, pp. 2220–2247.

Miller, Amalia R, Carmit Segal, and Melissa K Spencer (2024). "Effects of the COVID-19 pandemic on domestic violence in Los Angeles". In: *Economica* 91.361, pp. 163–187.

Muratori, Caterina (2022). *Is TRAP a trap? The impact of abortion access on violence against women*. Tech. rep. CHEPS Working Paper.

Pappa, Evi, Veronica Frisancho, and Chiara Santantonio (2022). *When Women Win: Can Female Representation Decrease Gender-Based Violence?* Tech. rep. CEPR Discussion Papers.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duch- esnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Rights, HTMF (2014). "Violence against women: An EU-wide survey". In: *Luxembourg: Publications Office of the European Union*.

Rosenbaum, Paul R and Donald B Rubin (1983). "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1, pp. 41–55.

Srivastava, Swati, Kaushalendra Kumar, Lotus McDougal, Ajit Kumar Kannaujiya, Ankit Sikarwar, Anita Raj, and Abhishek Singh (2023). "Spatial heterogeneity in intimate partner violence across the 640 dis- tricts of India: a secondary analysis of a cross-sectional, population-based survey by use of model-based small-area estimation". In: *The Lancet Global Health* 11.10, e1587–e1597.

VanderEnde, Kristin E, Kathryn M Yount, Michelle M Dynes, and Lynn M Sibley (2012). "Community-level correlates of intimate partner violence against women globally: A systematic review". In: *Social science & medicine* 75.7, pp. 1143–1155.

Vyas, Seema and Lori Heise (2016). "How do area-level socioeconomic status and gender norms affect partner violence against women? Evidence from Tanzania". In: *International journal of public health* 61, pp. 971– 980.

Yu, Rongqin, Yasmina Molero, Paul Lichtenstein, Henrik Larsson, Lewis Prescott-Mayling, Louise M Howard, and Seena Fazel (2023). "Development and validation of a prediction tool for reoffending risk in domestic violence". In: *JAMA network open* 6.7, e2325494–e2325494.

Appendix A – Descriptive statistics

Table A.1: List of predictors used in the targeting analysis

Predictor	Source
LLM classification (without specialization, non-manufacturing, made in Italy, and other manufacturing)	Istat
Geographical dummies (North-East, North-West, Centre and South)	Istat
Population	Istat
Employment rate, unemployment rate, activity rate	Istat
Per capita income	Ministry of Economy and Finance
Number of workers per 1,000 inhabitants	Istat
Share of workers employed in manufacturing	Istat
Share of workers employed in the construction sector	Istat
Average number of workers per local unit	Istat
Share of population located in municipalities considered as peripheral or ultra-peripheral according to the SNAI classification	Istat
LLM with at least one industrial district	Istat
Turnout at the previous European elections	Ministry of the Interior
Gender gap in turnout at the previous European elections	Ministry of the Interior
Share of foreign population	Istat
Share of old population (65+)	Istat
Share of male population	Istat
Share of graduate mayors	Ministry of the Interior
Average age of the mayors	Ministry of the Interior
Share of female mayors	Ministry of the Interior
Average price per square meter (house and villa separately)	Osservatorio del Mercato Immobiliare – Agenzia delle Entrate
Number of newborns per 1,000 inhabitants	Istat
Number of deaths per 1,000 inhabitants	Istat
Newborn-deaths ratio	Istat
Number of weddings per 1,000 inhabitants	Istat
Share of religious weddings	Istat
1522 helpline calls (provincial-level data)	Istat
Crime ranking by Il Sole 24 Ore (provincial-level data)	Il Sole 24 Ore
Crime score by Il Sole 24 Ore (provincial-level data)	Il Sole 24 Ore
Male/Female Self-Employed Ratio;	Istat - Census
Share of Female Employment in the Non-Retail Tertiary Sector;	Istat - Census
Female Unemployment Rate;	Istat - Census
Incidence of Graduates and Diploma Holders Among Population Aged 6 and Older;	Istat - Census
Average Household Size;	Istat - Census
Average stadium attendance (Serie A team)	Stadia Postcards
Total stadium attendance (Serie A team)	Stadia Postcards
Number of ordinary hospital beds per 1,000 inhabitants	Istat
Number of day-hospital beds per 1,000 inhabitants	Istat
Percentage of deaths directly attributable to alcohol abuse	Istat
Percentage of deaths directly attributable to substance abuse	Istat
Percentage of hospitalizations directly attributable to alcohol abuse	Istat
Percentage of hospitalizations directly attributable to substance abuse	Istat

Notes: This table shows the list of predictors used for the ML targeting analysis. Data on 1522 helpline calls are available only from 2013 onward and have been used exclusively in the targeting analyses reported in Figure C.6 and Figure C.7.

Table A.2: Anti-violence center openings

Province	First opening year	Number of AVCs opened prior to treatment
Foggia	2011	2
Macerata	2011	0
Varese	2011	1
Lecce	2012	1
Padova	2012	1
Siracusa	2012	3
Bologna	2013	5
Brescia	2013	1
Cosenza	2013	1
Mantova	2013	1
Perugia	2014	1
Avellino	2015	0
Benevento	2015	0
Grosseto	2015	1
Treviso	2015	1
Verbania	2015	0
Campobasso	2016	0
Bergamo	2017	5
Caltanissetta	2017	0
Monza e della Brianza	2017	1
Novara	2017	0
Potenza	2017	0

Notes: This table shows the provinces belonging to the treatment group. These provinces did not experience any AVC openings in the five years before the treatment.

Appendix B – Machine Learning Pipeline

The ML pipeline has been performed using Python 3.9 and the following packages: **Scikit-learn** (Pedregosa et al., 2011) for the pre-processing of the data, the logistic regression, the bagging and the random forest algorithms; **XGboost** (Chen and Guestrin, 2016) for the xgboost algorithm; **imblearn** (Lemaitre et al., 2017) for resampling; **tensorflow** (Martın Abadi et al., 2015) for the neural networks; and **shap** Lundberg and Lee, 2017 for the computation of SHAP values.

Here is a brief description of the models and the performance metrics used in the ML pipeline:

Random Forest is an ensemble learning technique that constructs many decision trees during training and aggregates their predictions to increase out-of-sample accuracy by reducing overfitting risk. Each tree is built using a random subset of the training data, which helps in reducing variance and making the model more robust. The final prediction is determined by averaging the outputs of all the trees. We use the following parameters: 500 trees, with a maximum depth of 5, with minimum 2 samples leaf, and 5 minimum samples split.

Neural Networks, also known as artificial neural networks (ANNs), are a class of NL models built upon a structure made of nodes and edges. An ANN consists of layers of interconnected nodes (neurons), where each node applies a mathematical function to its inputs and passes the result to the next layer. These networks can learn extremely complex patterns and representations from data, making them especially suitable for tasks like image recognition, natural language processing, and time series forecasting. The more layers and neurons a network has, the more powerful and capable it becomes at learning from data, though it may also become prone to overfitting. We use a fully connected 64*32 network, with rectified linear unit activation (*relu*), with an *adam* optimizer.

Extreme Gradient Boosting is an ensemble technique that builds models in a sequential manner, where each new model tries to correct the errors of the previous ones. **XGBoost** improves this approach by optimizing computational efficiency through parallelization, using regularization to prevent overfitting, and handling missing data. It is widely used in ML applications on structured/tabular data due to its ability to handle complex datasets with minimal tuning. We use it with the following parameters: objective = binary:logistic, learning rate= 0.01, max depth = 2, min child weight = 5, gamma= 1, subsample = 0.5, colsample by tree = 0.8.

Bagging, short for Bootstrap Aggregating, is an ensemble learning technique that improves the stability and accuracy of ML models by reducing variance. In bagging, multiple instances of a model (e.g., decision trees in our case) are trained on different subsets of the training data, which are generated by random sampling with replacement (bootstrap sampling). Each model makes a prediction, and the final result is obtained by averaging the predictions. Bagging is especially effective for high-variance models like decision trees, as it helps mitigate overfitting and provides more robust and generalizable predictions. We use it with the following parameters: the input learner is a decision tree with max depth=3, min samples split=10 and min samples leaf=5. Then, the Bagging algorithm optimizes the results on 500 of these estimators.

In the following, we describe the performance metrics used in the ML pipeline:

Sensitivity, also known as the True Positive Rate (TPR) or Recall, measures the proportion of actual positive cases that the model correctly identifies as positive. It is calculated as the ratio of true positives to the total number of actual positives (i.e., true positives + false negatives). Sensitivity is crucial in situations where the cost of missing positive cases (false negatives) is high, such as in medical diagnosis or fraud detection. In our case study, a low sensitivity implies that we are underscoring the risk of femicide, that is evidently a high cost. A sensitivity value of 1 indicates that the model correctly identifies all positive instances, whereas lower values suggest that some positive cases are being missed.

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{4}$$

Specificity, also known as the True Negative Rate (TNR), measures the proportion of actual negative cases that the model correctly identifies as negative. It is calculated as the ratio of true negatives to the total number of actual negatives (i.e., true negatives + false positives). Specificity is particularly important when false positives are costly or problematic, such as in screening tests. In our case, a high specificity implies that we are overrating the risk of a femicide, that has costs in terms of preventive policies. A specificity value of 1 means the model correctly identifies all negative instances, while lower values suggest that some negative cases are being incorrectly classified as positive.

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{5}$$

The Area Under the Curve (**AUC**) of the Receiver Operating Characteristics (**ROC**) curve is a metric used to evaluate the performance of binary classification models. The ROC curve plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity) across different classification thresholds. AUC is the likelihood that a model will rank a randomly chosen positive instance higher than a randomly chosen negative one. AUC values range from 0 to 1, where a value of 0.5 indicates no discriminative power (random guessing), and a value of 1 indicates perfect classification. A higher AUC value suggests better model performance, as it reflects a higher true positive rate and a lower false positive rate at various thresholds.

Balanced Accuracy is a metric used to evaluate the performance of a binary classifier when the class distribution is imbalanced. It is the arithmetic mean of Sensitivity (True Positive

Rate) and Specificity (True Negative Rate), providing an overall measure that takes into account both correct positive and correct negative predictions. Unlike standard accuracy, which can be misleading in the presence of class imbalance, Balanced Accuracy ensures that both types of errors (false positives and false negatives) are considered equally. A value of 1 indicates perfect classification, while values closer to 0 suggest poor performance.

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{6}$$

These metrics help assess model performance from different angles, ensuring that both positive and negative class predictions are accurately evaluated, especially in imbalanced datasets.

Appendix C – Additional results for the ML targeting analysis

This Appendix reports additional results for the ML-based targeting analysis.

Additional results for the main analysis Here we report additional results for the main targeting analysis reported in the paper. More specifically, we show below the average SHAP values for the best-performing XGBoost model and the logit Model, as well as the ROC curves for the forecasting exercise and graphs for the Balanced Accuracy.

Figure C.1: ROC Curves

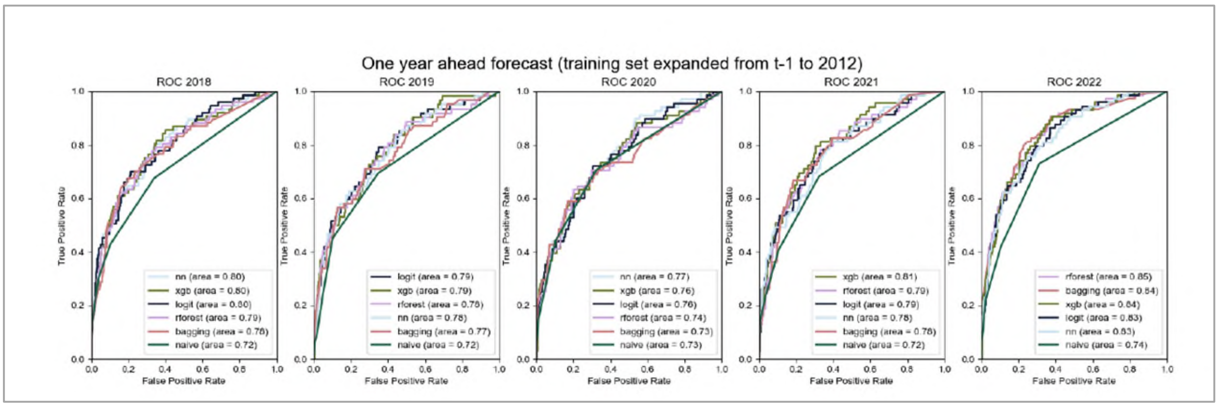


Figure C.2: SHAP Values of the XGBoost model

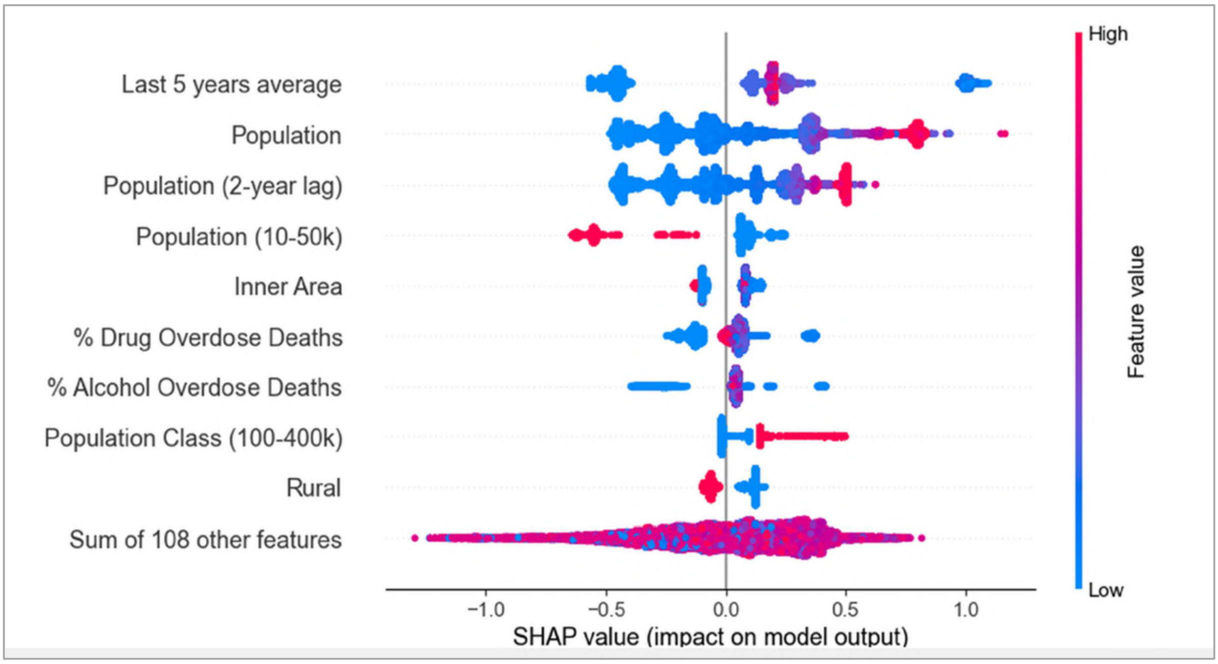


Figure C.3: SHAP Values of the logit model

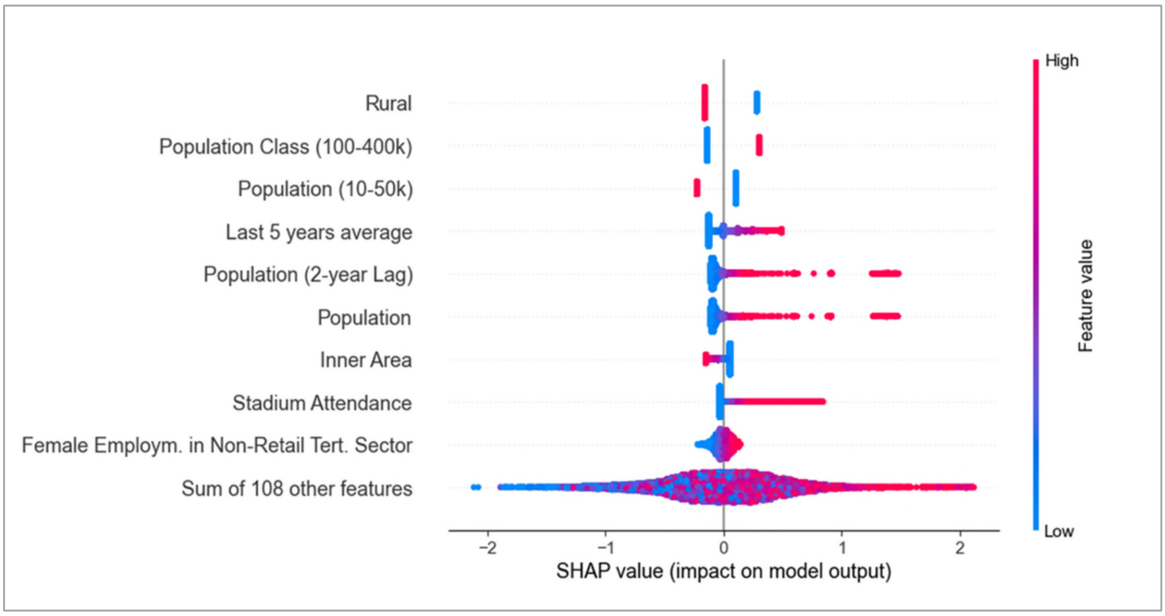


Figure C.4: Mean absolute SHAP values of all the models

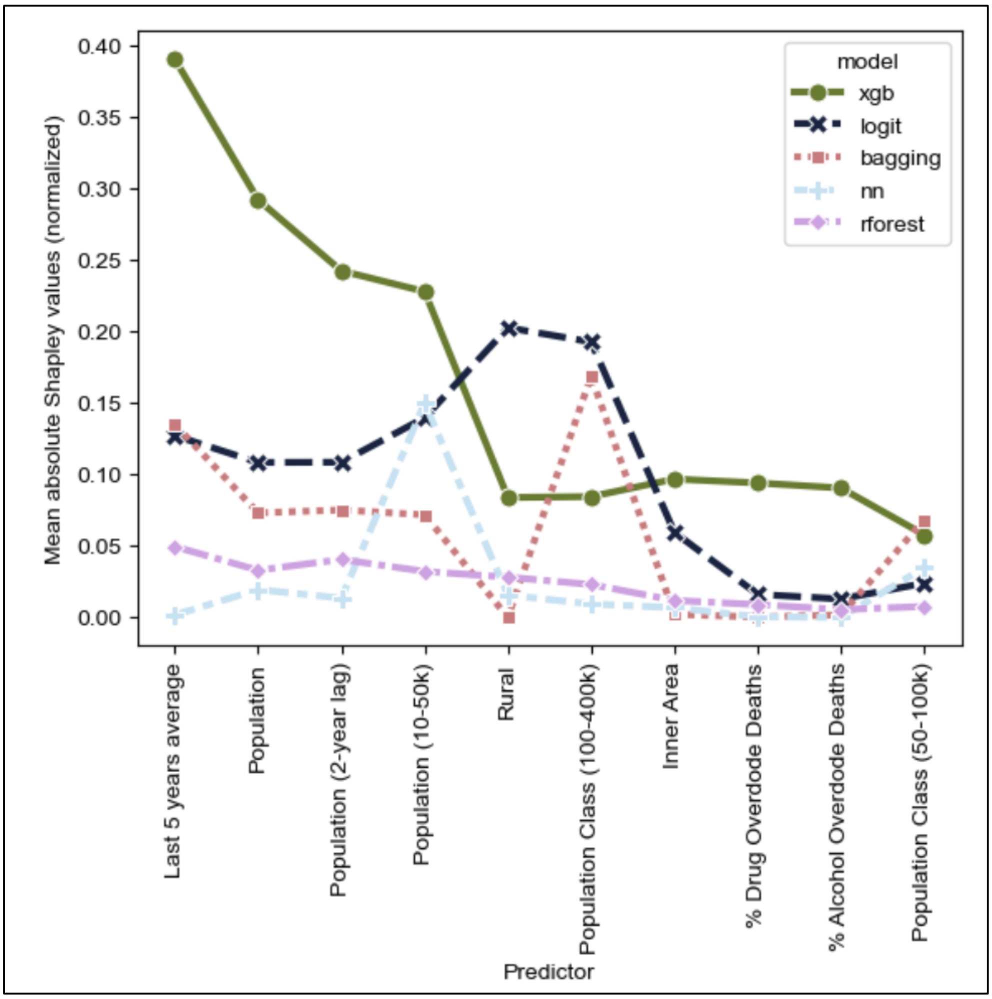
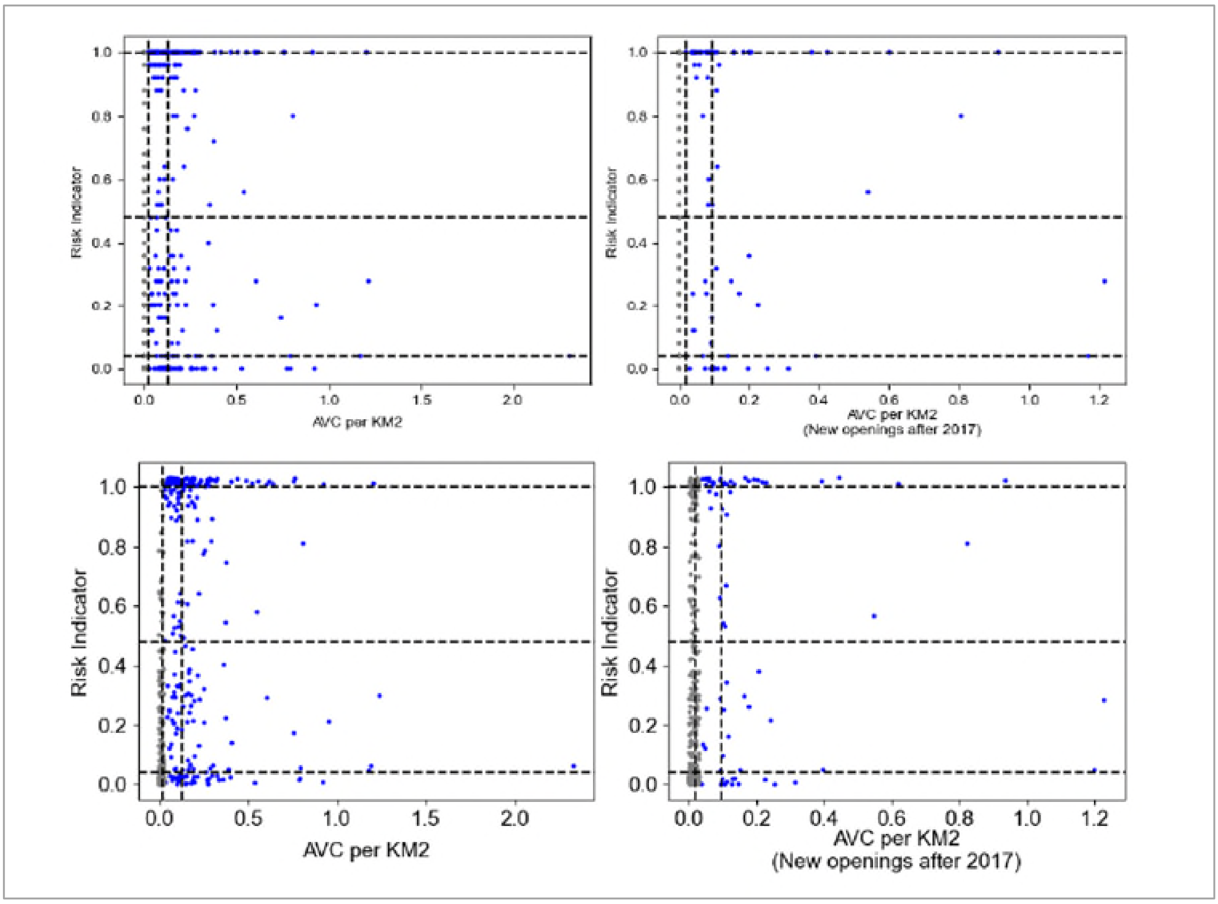


Figure C.5: Openings of AVC vs. ML-based femicide risk indicator (2018-2022)



Notes: Panel a shows the scatterplot of the risk indicator and the AVC coverage (i.e. the same data shown in Figure 6). Panel b shows a jittered version of the graph to account for overlap between points.

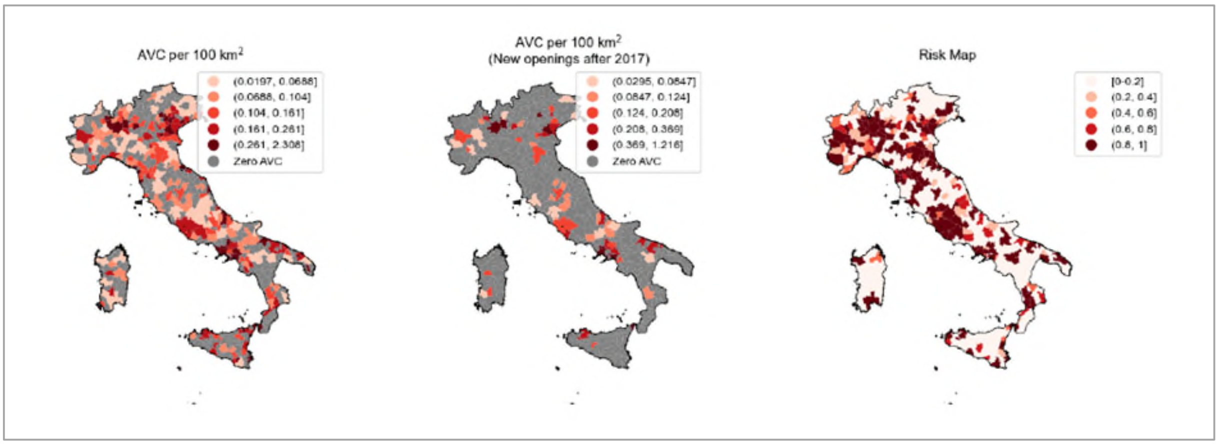
Table C.1: Balanced accuracy

Model	Mean	Std	Min	25%	50%	75%	Max
Naive	0.647078	0.026142	0.596333	0.650218	0.657221	0.662691	0.668928
Logit	0.724375	0.018375	0.687107	0.711296	0.724390	0.739236	0.759883
Neural Networks	0.724436	0.021503	0.682440	0.707632	0.720799	0.740820	0.776472
Random Forest	0.728228	0.022633	0.699859	0.709413	0.722836	0.734572	0.786028
Bagging	0.728438	0.027712	0.679862	0.705910	0.728272	0.734627	0.785549
XGBoost	0.731386	0.023521	0.692750	0.711051	0.730221	0.740657	0.784616

Alternative targeting analysis Here, we present the results of an alternative targeting analysis using also lagged 1522 emergency calls as a predictor. Figure C.6 compares the AVCs' coverage with the alternative femicide risk indicator, while Figure C.7 shows the degree of overlap by tertiles. Note that 1522 calls data are only available from 2013, so that the window on which the data are trained is shorter (from 2015, since we use two lags of the predictors).

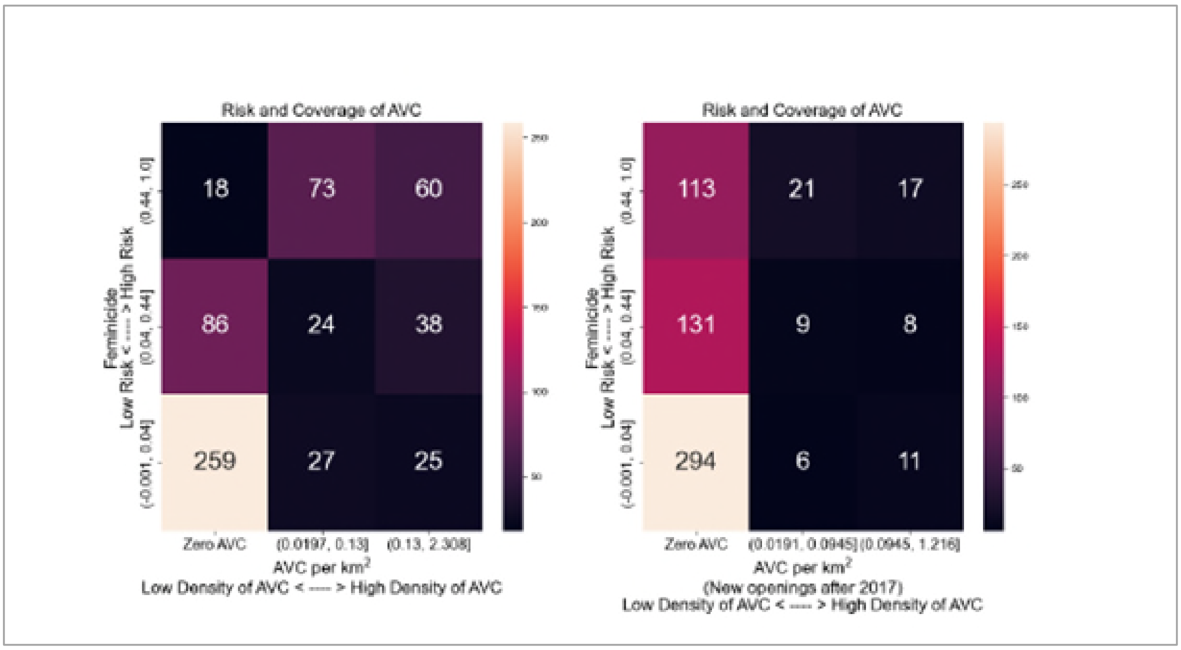
Despite these differences, the key insights—and, particular, the limited overlap of the risk indicator with the distribution of AVCs—are fully consistent with the main results.

Figure C.6: AVC coverage vs. ML-based femicide risk indicator (incl. 1522 calls)



Notes: For a comparison with the corresponding set of maps reported in the main text, see Figure 5.

Figure C.7: Openings of AVC vs. ML-based femicide risk (incl. 1522 calls) indicator (2018-2022)



Notes: This figure shows the degree of overlap between the ML-based femicide risk (computed using data from 2013 including 1522 helpline calls) and AVC coverage by tertiles. For comparison, see Figure 6.

Alternative targeting analysis using only IPF as main predictor This paragraph shows an additional targeting analysis, where the main outcome is replaced with an alternative one that captures only IPF.

Figure C.8: AVC coverage vs. ML-based femicide risk indicator (IPF only)

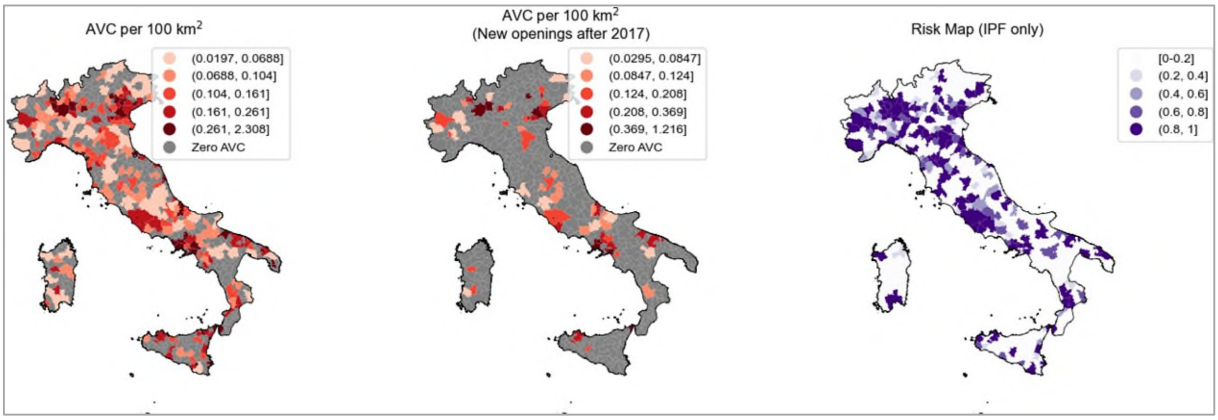
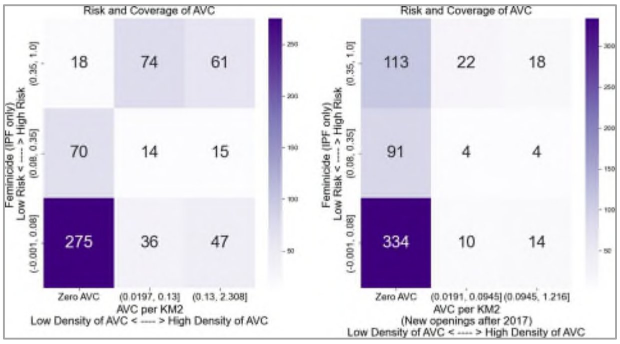
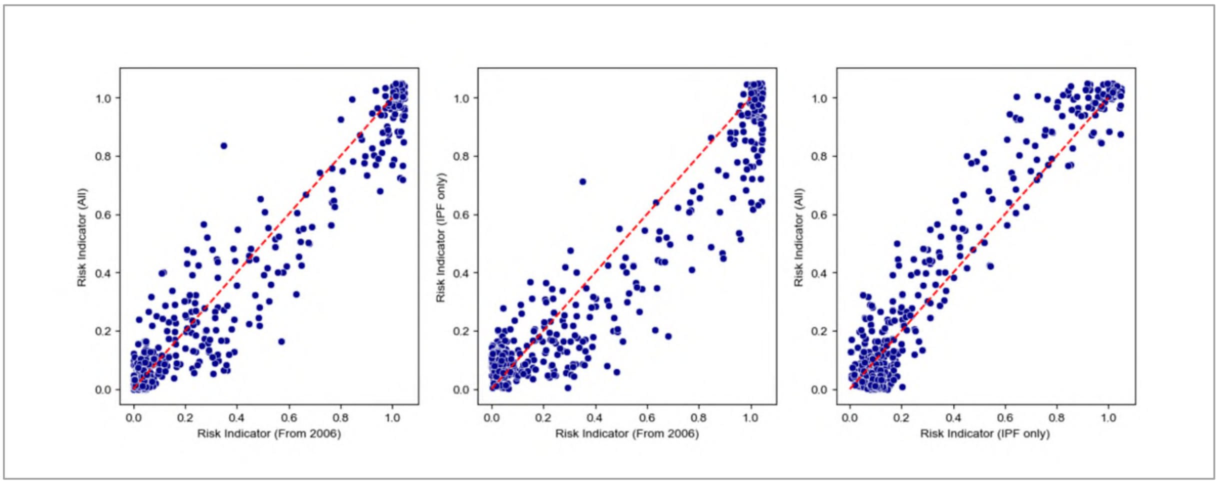


Figure C.9: Openings of AVC vs. ML-based femicide risk (IPF only) indicator (2018-2022)



Notes: This figure shows the degree of overlap between the ML-based femicide risk (IPF only) and AVC coverage by tertiles. For comparison, see Figure 6.

Figure C.10: Comparison between the three indicators



Notes: This figure shows the correlations between the three risk indicators (the baseline data that goes back to 2006, the one including 1522 calls with data from 2013, and the one only using IPF). The markers have been slightly jittered (i.e. we have added a random number between 0 and 0.05) so to make all points visible. The correlation between the three risk indicators is very high and ranges from 0.952 to 0.974.

Appendix D – Additional results for the AVC analysis

D.1 Robustness

We assess the sensitivity of the estimates of the AVC counterfactual analysis through several robustness checks and summarize the results in Table D.1. All these estimates align with those of the main analysis: the treatment results in a statistically significant reduction only in sexual violence cases, while the effects on the other outcomes are not statistically significant. The only exception is a decrease in the number of femicides per 1,000 inhabitants at time t+4 that is statistically significant at the 10% level when using PSM. In addition, we test the sensitivity of our method by adopting the dynamic version of the DiD with multiple time period estimators developed by De Chaisemartin and d'Haultfoeuille (2024). Small differences emerge in the point estimates reported in Table D.1. For instance, this estimator suggests a statistically significant decrease in the number of femicides per 1,000 inhabitants at time t + 4 and an even more statistically significant drop in the number of reported sexual violence cases. Finally, we adopt the semi-parametric DiD estimator proposed by Callaway and Sant'Anna (2021), which mostly delivers similar results, except for stalking cases. Indeed, the treatment generates a significant increase in reported stalking crimes at times t+3, t+4 and t+5.

Table D.1: Robustness checks

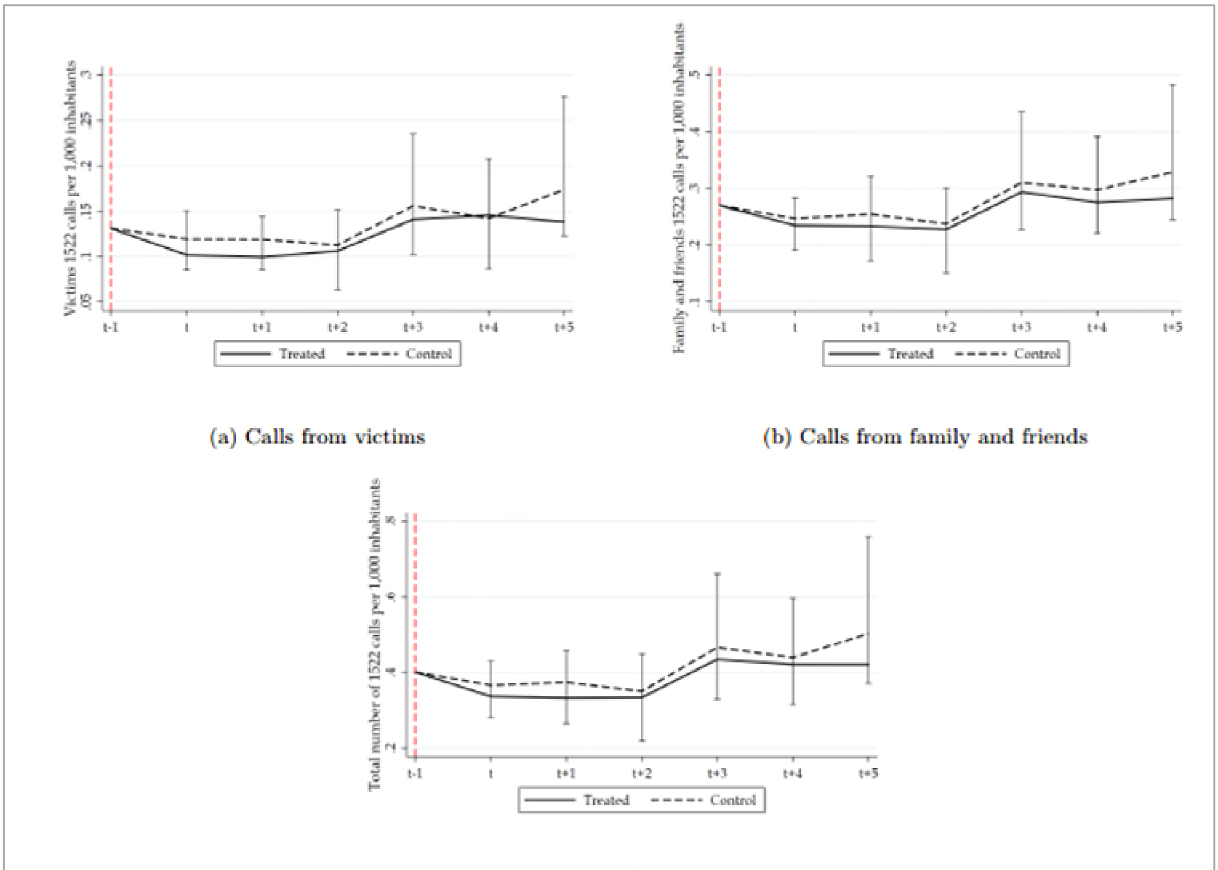
	Years after treatment					
	T	*T+1*	*T+2*	*T+3*	*T+4*	*T+5*
Panel A – Femicides per 1,000 inhabitants						
Main Estimates	0.00020	0.00034	-0.00009	0.00009	-0.00089	-0.00001
PSM with 3 neighbors	0.00042	0.00067	0.00039	0.00029	-0.00065	0.00036
Mahalanobis with 3 neighbors	-0.00050	-0.00030	-0.00016	-0.00020	-0.00093	-0.00088
IPW	0.00013	0.00026	-0.00028	0.00004	-0.00096	-0.00012
Additional covariates	-0.00103	0.00167	-0.00063	0.00085	-0.00114	0.00003
de Chaisemartin et al. (2024) estimator	0.00073	-0.00007	-0.00010	0.00014	-0.00053	-0.00029
Callaway & Sant'Anna (2021) estimator	-0.00032	-0.00003	0.00014	0.00019	-0.00078	-0.00004
Panel B – Intimate Partner Femicides per 1,000 inhabitants						
Main Estimates	-0.00060	0.00039	0.00009	0.00007	-0.00032	0.00040
PSM with 3 neighbors	-0.00022	0.00053	0.00046	0.00019	-0.00015	0.00062
Mahalanobis with 3 neighbors	-0.00103**	0.00005	0.00019	-0.00010	-0.00045	-0.00017
IPW	-0.00064	0.00028	0.00001	0.00000	-0.00038	0.00031
Additional covariates	-0.00128	0.00130	-0.00079	0.00087	-0.00080	0.00055
de Chaisemartin et al. (2024) estimator	-0.00063	0.00016	0.00015	-0.00005	0.00019	0.00024
Callaway & Sant'Anna (2021) estimator	-0.00093*	0.00048	0.00004	0.00024	-0.00028	0.00056
Panel C – Sexual violence per 1,000 inhabitants						
Main Estimates	-0.00330	-0.01089	-0.02126**	-0.01554**	-0.02068**	-0.02168*
PSM with 3 neighbors	-0.00737	-0.01085	-0.01806***	-0.01304*	-0.02245**	-0.02607**
Mahalanobis with 3 neighbors	-0.00440	-0.00937	-0.01263**	-0.01262*	-0.01510**	-0.01922*
IPW	-0.00323	-0.01139	-0.02124**	-0.01621**	-0.02138**	-0.02066*
Additional covariates	0.00433	-0.01832**	-0.01564*	-0.01380**	-0.01862**	0.01981
de Chaisemartin et al. (2024) estimator	-0.00306	-0.00717	-0.01551***	-0.01506***	-0.02296***	-0.02175***
Callaway & Sant'Anna (2021) estimator	-0.00592	-0.01468*	-0.01594**	-0.01949**	-0.02487***	-0.01853
Panel D – Abuses per 1,000 inhabitants						
Main Estimates	-0.01023	-0.00415	0.00554	0.01601	0.01279	0.01523
PSM with 3 neighbors	-0.00533	-0.00940	0.00181	0.01327	0.00602	0.01306
Mahalanobis with 3 neighbors	-0.01260	-0.00597	-0.00089	0.00165	0.01092	0.00329
IPW	-0.00976	-0.00203	0.00728	0.01667	0.01286	0.01639
Additional covariates	0.00255	0.00578	0.01897	0.02972	0.02679	0.05271
de Chaisemartin et al. (2024) estimator	-0.01403*	-0.00980	-0.00406	0.00440	-0.00265	-0.00487
Callaway & Sant'Anna (2021) estimator	-0.01291*	-0.01320	-0.01344	-0.00210	-0.00496	-0.01180
Panel E – Stalkings per 1,000 inhabitants						
Main Estimates	0.00624	0.00655	0.01154	0.01234	0.03610	0.03035
PSM with 3 neighbors	0.00345	0.01196	0.01744	0.01680	0.04261	0.03745
Mahalanobis with 3 neighbors	0.00186	0.00816	0.00776	0.01426	0.04764	0.04674
IPW	0.00698	0.00777	0.01380	0.01302	0.03633	0.02971
Additional covariates	-0.01250	0.00582	0.00885	0.01750	-0.03570	0.01234
de Chaisemartin et al. (2024) estimator	0.00463	0.00323	0.01104	0.01099	0.03822	0.03737
Callaway & Sant'Anna (2021) estimator	0.00227	-0.00305	0.00118	0.00438	0.02147	0.02220

*Notes: The estimates with the non-parametric generalization of the DiD estimator are based on the R package PanelMatch, while the Stata commands did multiplegt dyn and csdid were used to implement the De Chaisemartin and d'Haultfoeuille (2024) and Callaway and Sant'Anna (2021) estimators, respectively. Province-specific trends included for the implementation of the De Chaisemartin and d'Haultfoeuille (2024) estimator. We implement the doubly robust DiD estimator of Callaway and Sant'Anna (2021), which combines stabilized inverse-probability weighting with ordinary least squares, using as controls both never-treated provinces and those not yet treated. ***p < 0.01, **p < 0.05, *p < 0.1.*

D.2 The effect of AVC openings on 1522 calls

In this subsection, we report results on the effect of opening new AVCs on calls to 1522 helpline.

Figure D.1: 1522 calls



(a) Calls from victims

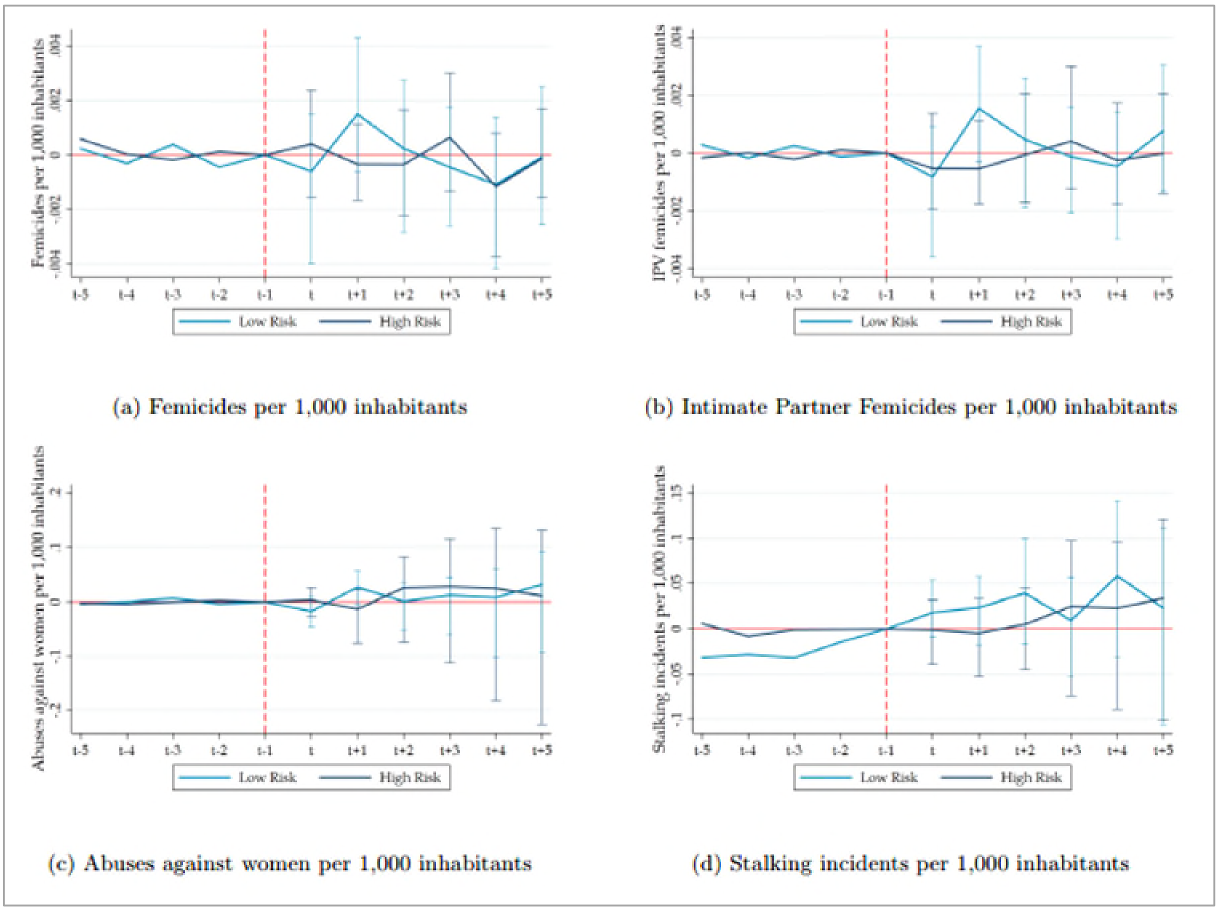
(b) Calls from family and friends

(c) Total calls

Notes: The vertical black bars report 95% confidence intervals for the outcome of the estimated counterfactual scenario. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)). The number of treated units is 12 as data on helpline calls are only available starting from 2013, meaning that we only have treated units from 2014 onwards to have at least 1 pre-treatment year.

D.3 Mechanisms - additional results

Figure D.2: High risk vs low risk areas



(a) Femicides per 1,000 inhabitants

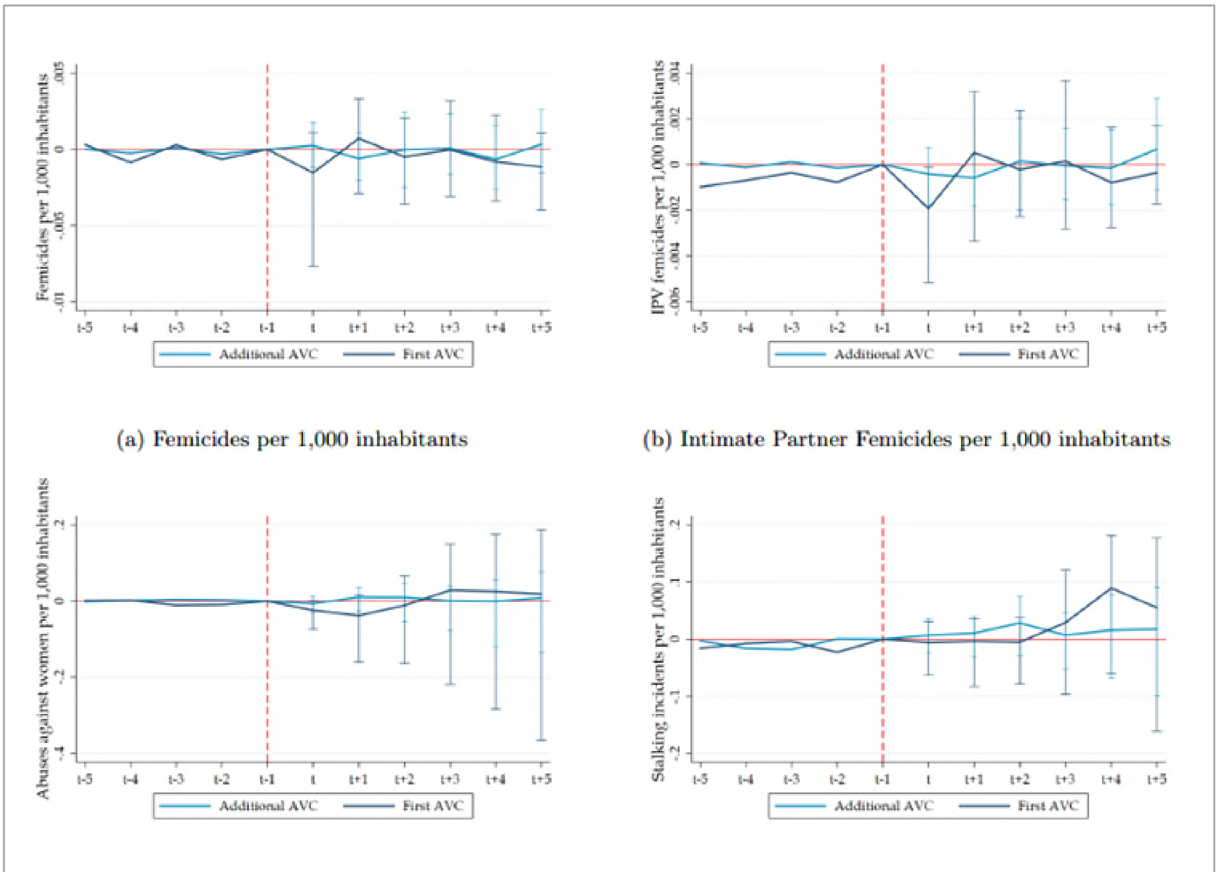
(b) Intimate Partner Femicides per 1,000 inhabitants

(c) Abuses against women per 1,000 inhabitants

(d) Stalking incidents per 1,000 inhabitants

Notes: The vertical black bars report 95% confidence intervals for the outcome of the estimated counterfactual scenario. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)). Treated units are those with the risk indicator above/below the median.

Figure D.3: First vs already operating AVC



(a) Femicides per 1,000 inhabitants

(b) Intimate Partner Femicides per 1,000 inhabitants

(c) Abuses against women per 1,000 inhabitants

(d) Stalking incidents per 1,000 inhabitants

Notes: The vertical black bars report 95% confidence intervals for the outcome of the estimated counterfactual scenario. Confidence intervals are computed using a block-bootstrap procedure (see Imai, Kim, and Wang (2023)). The number of provinces with no pre-existing AVCs before receiving the treatment is 8, while those with one or more AVCs already operating is 14.

SENATO DELLA REPUBBLICA

UFFICIO VALUTAZIONE DI IMPATTO
IMPACT ASSESSMENT OFFICE

www.senato.it/ufficiovalutazioneimpatto