Validation and Calibration of the Spanish Police Intimate Partner Violence Risk Assessment System (VioGén)

Juan José López-Ossorio[1] · José Luis González-Álvarez[1,2] · José Manuel Muñoz Vicente[3] · Carlota Urruela Cortés[1] · Antonio Andrés-Pueyo[4]

Abstract

This study describes the rationale, development, and validation of the intimate partner violence (IPV) police risk assessment forms of the VioGén System of the Spanish Ministry of Interior ($VPR_{4.0}$ and $VPER_{4.0}$), which promote greater predictive effectiveness and an improvement in the IPV law enforcement prevention. A validation study of the mentioned protocols is presented, including inter-observer reliability, estimated by the equivalence or inter-judge reliability method, while the convergent validity of these protocols was calculated with the RVD-BCN protocol. The sample consisted of 6613 new cases of IPV included in the VioGén System over a period of 2 months and which were longitudinally followed up for 6 months. The discrimination indexes are not only the summarized odds ratio (OR), area under the ROC curve (AUC), sensitivity, and specificity, but also the calibration indexes positive predictive value (PPV) and negative predictive value (NPV). The results show the suitability of using procedures which, in a coordinated manner, incorporate two risk assessment instruments, one for a first screening assessment and a second one to re-assess IPV danger situations on a regular basis. The values obtained are within the margins reported by different meta-analyses regarding this type of instruments, which supports their use for professional practice.

Keywords Intimate partner violence · Police risk assessment · Prediction tools · Validation process · VioGén system

Introduction

The law enforcement context is the victim's main gateway into the justice system. Thus, it is a privileged place for adopting the first protection measures. Moreover, on many occasions—as in the Spanish case—law enforcement forces receive institutional orders to perform this task (López-Ossorio et al. 2016). In this sense, demands coming from the law enforcement area have contributed to the exponential evolution of the development of tools for the assessment and handling of violent recidivism risk (Belfrage et al. 2012; Buchanan et al. 2012). Due to the relevant

increase of social awareness and to the magnitude of this problem, intimate partner violence (IPV) has been one of the fields where this methodological development can be noted (Women Against Violence Europe—WAVE 2011, 2012). Currently, we can count with more than 150 tools for this purpose (Arbach-Lucioni et al. 2015).

In the international context, we can highlight the law enforcement use of the following IPV risk assessment guides: Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) Second Edition (Kropp et al. 2010); Threat Assessment Systems (DV-MOSAIC) (De Becker 1997; De Becker et al. 2000); Ontario Domestic Assault Risk Assessment (ODARA) (Hilton et al. 2008); Spousal Abuse Risk Assessmental (SARA) (Kropp et al. 1995); the Spouse Violence Risk Assessment Inventory (SVRA-I) (Dayan et al. 2013); and recently, the Violence Risk Screening—Police Version (V-RISK-POL) (Roaldset et al. 2017). The Spanish context counts with the adaptation of international tools—such as SARA (Andrés Pueyo et al. 2008)—and also its own tools, such as the Prediction Scale of Serious Spouse Violence Risk—Reviewed (EPV-R) (Echeburúa et al. 2010) and the Risk Assessment Protocol for Violence against

✉ Juan José López-Ossorio
jjlossorio@gmail.com

1   Instituto de Ciencias Forenses y de la Seguridad (Institute for Forensic and Security Sciences), Universidad Autónoma de Madrid, Madrid, Spain

2   Coordination and Studies Office of the Secretary of State for Security, Spanish Ministry of Interior, Madrid, Spain

3   Supreme Court of Madrid, Madrid, Spain

4   Faculty of Psychology, Universidad de Barcelona, Barcelona, Spain

women by spouse or ex-spouse (RVD-BCN) (Álvarez et al. 2011). These tools usually contain specific elements of this type of violence and others shared with more unspecific risk of violence assessment tools (Rossegger et al. 2013).

The development of this area of research has allowed for a reduction of uncertainty and favored making more accurate decisions aimed at the prediction and handling of recidivism risk (Otto and Douglas 2010). The assessment protocols are used as a guide for professionals when making decisions that have significant implications regarding civil liberties, as well as public security, by means of a paradigm shift that goes from danger assessment to risk assessment, promoting a remarkable change of interest for evidence-based techniques when predicting the probability of recidivism (Arbach-Lucioni and Andres-Pueyo 2016). This has resulted in an improvement of predictive efficacy, an important increase of objectivity and transparency of the assessment process and mainly in organizing the resources available to handle risk in a more efficient and effective way paying particular attention to the empiric evidence of its elements and its risk profiles, preventing and neutralizing possible future aggressions (Canales et al. 2013; Folino 2015; Llor-Esteban et al. 2016; Loinaz 2017).

To achieve the necessary improvements in professional practice, it is essential to submit these tools to a validation process, different from that of psychometric instruments, in order to acknowledge their characteristics (Muñoz and López-Ossorio 2016). The validation of a tool is a continuous and dynamic process that gains consistency when adapting the tool to different cultures, with different populations and subjects (Carvajal et al. 2011). The need of using evidence-based violence risk assessment methodologies has led to elaborating international criteria and methodology rules to be met by instruments of this nature for its use by psychologists, psychiatrists, and other professionals—the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement—(Singh et al. 2015). Once these criterions are met, the professional could opt for actuarial risk assessment tools, which are characterized by its aim to minimize the assessor's subjectivity on the risk estimation, or for structured clinical judgment guidelines that provide greater discretion to the assessor in the predictive task. Different meta-analysis indicates that there are no significant differences between both methodologies regarding their predictive capacity (Singh et al. 2011, 2014).

In Spain, and for IPV, there is also an institutional mandate, which harbors law enforcement tasks. Since 2007, the Spanish law enforcement services have the Comprehensive Monitoring System for Cases of Gender Violence (VioGén) which, among other functionalities, contains a dual actuarial protocol for IPV risk assessment, with two computerized forms: Police Risk Assessment (VPR—*Valoración Policial del Riesgo*) and Police Risk Evolution Assessment (VPER—*Valoración Policial de la Evolución del Riesgo*) (López-Ossorio et al. 2016). The VPR form performs a first function of screening cases with actuarial risk indicators, and the VPER form monitors the evolution of events with risk and protection indicators.

This system, which already accumulates more than 3,000,000 risk assessments, provides access to a large sample of cases that allows a rigorous validation process with a high generalization capacity.

The assessment protocol has recently been reviewed and improved, and this paper presents the latest step in the process of the updating of the protocol forms of the Spanish Ministry of Interior, which are mainly actuarial adjusted tools. Specifically, the validation process is described as a first phase to guarantee the quality of the predictive task. The process of selection of the elements (risk/protection factors) that shape the tool was already presented in a former study (López-Ossorio et al. 2017).

Regarding validity, there are several ways of estimating it. *Predictive validity* informs through the area under the ROC curve (AUC) of the diagnostic effectiveness of the tool and its capacity to detect the presence or absence of a certain phenomenon, in this case recidivism in IPV behaviors. However, it is an incomplete performance parameter, and the accuracy of the tool must also be reported in order to predict the proportion of recidivist offenders in the different groups (calibration) and the precision with which the tool identifies recidivist offenders (discrimination). Discrimination is determined by sensitivity, specificity, the AUC value, and the odds ratio (OR), while calibration indexes are the positive predictive value (PPV) and negative predictive value (NPV) along with the analysis between the predicted cases and those observed in the different risk levels. All this is studied in order to achieve the best stratification of reported cases by risk levels.

As it has been mentioned, this study presents the work carried out in the development and validation of the law enforcement assessment and reassessment forms for recidivism risk and IPV risk (VPR$_{4.0}$ and VPER$_{4.0}$) integrated in the VioGén System and which replace former questionnaires. Thus, the aim is to report the performance parameters obtained for both forms, trusting that this information can serve as a suggestion for other countries and institutions, which carry out IPV risk assessments.

Method

Participants

The sample consisted of 6613 new cases of IPV included in the VioGén System over a period of 2 months and which were longitudinally followed up for 6 months. These female victims had an average age of 34.5 (SD = 13.03) and an age range of 13 to 68 years. As for the aggressors reported, the average age

was of 36.51 years old ($SD = 14.57$) and an age range of 14 to 68 years. The 65.7% of this sample consisted of Spanish women; the rest came from a total of 86 different countries, standing out Romania for its high frequency (5.4%), Morocco (4.2%), Ecuador (3.3%), and Colombia (2.3%). Regarding the men reported, 67.9% were Spanish, and the rest was distributed between 89 different countries, with a similar percentage distribution to the women mentioned above.

Thanks to the geographical stratification of the sample, it was possible to gather the total amount of complaints processed in police units of the national territory, urban (65%), or rural (35%), corresponding to 46 of the 52 Spanish provinces. The higher percentages were found in Madrid (18.1), Valencia (7.4%), Alicante (5.9%), Sevilla (5.5%), Málaga (5.2%), Las Palmas (4.9%), and Murcia (4.7%). The complaints were mainly collected by the National Police (62.5%) and the Spanish Guardia Civil (35%), with the remaining percentage divided among eight local police forces and the Navarre police. Table 1 shows the main sample descriptions.

Instruments

The new police risk assessment protocols ($VPR_{4.0}$ and $VPER_{4.0}$) from the VioGén System were used in this validation study. To maintain the consistency with former studies, the first form is also called *Police Risk Assessment*/*Valoración Policial del Riesgo* ($VPR_{4.0}$) and the second form is called *Police Risk Evolution Assessment*/*Valoración Policial de la Evolución del Riesgo* ($VPER_{4.0}$).

For the construction of the two updated forms, an expert group from Universidad Autónoma de Madrid (UAM) carried out a thorough bibliographic review of risk indicators of IPV in databases and documentary sources. Following, other experts from UAM and the University of Barcelona (UB) were consulted in order for them to analyze the suitability of the indicators obtained. Finally, a group of 10 police officers, who had experience in IPV and risk assessment, was established so as to know their opinion about the adequacy of the indicators for a future police assessment. All this previous exploratory work supports the validity of content through the thorough

review of studies (rational validity), queries to professional users (validity of response), and expert opinion (validation by judges).

The indicators compiled were organized in two support forms, called VPR_{2015} and $VPER_{2015}$, with which an empirical study about recidivism in the law enforcement context was conducted in order to know the OR of each indicator in relation to the violent behaviors and their parameters (López-Ossorio et al. 2017). According to those OR values, the 65 indicators that finally integrated the VPR_{2015} and $VPER_{2015}$ forms were selected, and a weighted additive algorithm was generated for each form. Some cutoff points were also established to determine five risk levels: unappreciated, low, medium, high, and extreme.

Finally, for the $VPR_{4.0}$, 39 risk indicators, which must be fulfilled in a dichotomous way to increase the objectivity, grouped into four thematic dimensions, were included. The first dimension is the severity of the reported episode. It has indicators that collect information about the IPV history. The aggressor's related factors are the second dimension divided into three categories: partner relationship, jealousy, monitoring, and harassing behaviors; indicators that point out antisocial characteristics, especially defiance and rule breaking and violent characteristics; and psychopathological and psychosocial adjustment indicators. The third dimension group is a compilation of indicators with those victims' characteristics, which put them in a vulnerable situation. The form ends with aggravating circumstances, which include situational indicators that may precipitate violent behaviors, including as well subjective aspects of the victims related to their perception of their own risk, as well as minors and other dependent relatives. We understand that this grouping of indicators by thematic dimensions has a better construct validity support than the statistic clusters resulting from more inductive factorial procedures.

The second form, the $VPER_{4.0}$, is complementary to the previous one and is constituted by 43 indicators, which are also fulfilled in a dichotomous way. From these 43 indicators, 34 are risk factors, and 9 are protective factors; they are grouped into five criminological dimensions: the four from

Table 1 Basic descriptive data of the sample ($n = 6613$ cases)

Description	Women	Men				
Age	$M = 34.5$ $SD = 13.03$	$M = 36.51$				
	Range = 13–68	$SD = 14.57$				
		Range = 14–68				
Description	Spain	Romania	Morocco	Ecuador	Colombia	Others
Nationality %	65.7	5.4	4.2	3.3	2.3	19.1
Description	Urban	Rural				
Territory %	65	35				
Description	National police	Guardia civil	Others			
Police force %	62.5	35	2.5			

the VPR$_{4.0}$ plus a new dynamic-relational indicator dimension used for monitoring risk and for updating the protection measures applied in the first instance. This form, VPER$_{4.0}$, presents two different ways, which are called with incident (VPER-C$_{4.0}$) and without incident (VPER-S$_{4.0}$), to be used by the law enforcement officers, depending on whether there have been recidivism or any similar episodes registered since the first VPR$_{4.0}$ application, or if is just a periodic assessment because time has passed without any violent incidents. The VPER-C$_{4.0}$ version includes indicators, which are specifically designed to explore the new violent episode's characteristics (first dimension), which are not contemplated in the VPER-S$_{4.0}$ version. Some examples of indicators included on both forms can be seen in Table 2. Risk and protection factors and its OR can be found at López-Ossorio et al. (2017).

Design and Procedure

Intra-observer Reliability and Convergent Validity

In order to calculate the reliability and convergent validity, 300 reports of the cases used in this validation were downloaded from the VioGén System; 150 were recurrent cases and 150 were not, counting with different risk levels (assessed by the previous VPR$_{3.1}$ form, which was in force during the development and validation of the new versions): 18.7% in unappreciated risk level; 35% in low-risk level; 28.3% in medium-risk level; 12.7% in high-risk level; and 5.3% in extreme-risk level. Each case's information was organized by a group of 12 psychologists trained to confirm the presence of the indicators of the new forms and to calculate

the correlation with the ones indicated by the law enforcement officers (blind to this information).

Convergent validity between the risk assessment of the form and the gold standard selected, the Risk Assessment of Violence Against Women by their partners or ex-partners Protocol (RVD-BCN), a tool that includes 16 factors that allow the risk assessment by means of risk dimensions, vulnerability, and risk perception of women, was also estimated. The protocol assesses the near-term risk for committing serious violent acts either by the partner or the ex-partner, with three risk levels. The choice of the RVD-BCN protocol as a comparative tool was motivated by its approach and objectives, which have been validated with a Spanish sample following the same essential parameters and for being a recent implementation. Four of the police reports did not include the necessary information for the analyses; for that reason, 296 of them were finally used.

Validation of the VPR$_{4.0}$ Form

To validate the VPR$_{4.0}$ form, an analytical-relational level prospective longitudinal design—which subsequently allows the construction of a predictive level model based on bivariate analysis—was used on the cohort of cases of those women who reported having been victims of IPV by their partners or ex-partners from September 24 to December 29, 2015 ($n = 6613$), with a "temporal risk window" that finalized on April 29, 2016. Those cases in which a new complaint was submitted against the same aggressor are considered as recidivists in the law enforcement field, finding 490 new complaints in 6 months. Thus, the recidivism rate of IPV after a second complaint in a 3-month interval was 4.6% and 7.4% for a 6-

Table 2 Examples of indicators included in the police risk assessment forms (VPR$_{4.0}$-VPER$_{4.0}$)

Risk factors	Homicide critical indicators
Violent episode's severity	
	1. Serious or very serious physical violence
	2. Serious or very serious sexual violence
	3. Use of bladed weapon
	4. Death threats by the aggressor
	5. Aggressions and threats escalation over the last 6 months
Aggressor's factors	
	6. Over the last 6 months the aggressor shows extreme jealousy
	7. Over the last 6 months the aggressor shows harassing behaviors
	8. Over the last year the aggressor has physically assaulted third parties or animals
	9. The aggressor suffers from a mental disorder
	10. The aggressor has suicidal thoughts or suicide attempts
	11. The aggressor has any type of addiction or abusive behavioral of alcohol or drugs
Aggravating circumstances	
	12. The victim has expressed her intention to break the relationship less than 6 months ago
	13. The victim believes that the aggressor is capable of severely assaulting or even killing her

month interval. With the aim of assuring the measures' stability, a cross-validation procedure (two halves) was chosen, so that for the $VPR_{4.0}$ form development, 60% of the sample ($n = 3907$) was used, while the 40% left ($n = 2706$) was used for obtaining the validity parameters. Six-month estimates are considered suitable when short-term protection measures and continued risk reevaluations are established.

Validation of the VPER4.0 Form

For the $VPER_{4.0}$ form construction and validation, a retrospective design of cases and controls was used because it was not possible to draw on previous assessments as for the $VPR_{4.0}$. The cases were the assessments made in the first recidivism instance ($n = 246$; as a study category), and controls were the first periodic assessments ($n = 3727$; as the reference category). Therefore, the sample used for the validation of this tool consisted of 3973 cases. The sample corresponds to the first evolutional assessments conducted and is a subsample of the entire array that emanates from the referred criterion.

The data related to the severity of the registered violence served as the second dependent variable. This variable was built upon 3725 cases that were included on the periodic assessments. Those cases were both, with and without recidivism, reported by the victim—hence the difference with the 490 recidivist cases. The periodic assessments were collected through three indicators that allowed knowing the physical, psychological, and sexual violence severity. This way, those cases that specialized police officers considered as serious during the subsequent assessments with the risk indicators of psychological, physical, and sexual violence included on the support-form show 13.1% (487 registries) of serious violence and 86.9% (3238 registries) of mild or no violent episodes, once they were dichotomized.

The four cutoff point determination for the five risk levels was done assessing three important criterions: the first cutoff point should assume a low percentage of false-negatives (even if the false-positives rate increased); the risk categories, from medium risk, should collect the larger quantity of cases considered as serious; and to a smaller degree, the resulting risk cluster should maintain similar percentages to the ones provided by other equivalent tools (Álvarez et al. 2011).

Statistical Analyses

The independent variables correspond to the $VPR_{4.0}$ and $VPER_{4.0}$ risk assessment forms. The dependent variables were the recidivism in IPV cases in a period of time from 3 to 6 months for the validation sample of the $VPR_{4.0}$ ($n = 6613$; recidivism of 490 cases) and the violence severity, which is considered the most important variable of analysis because of its relevance ($n = 3725$; serious violence in 487 cases). For the

$VPER_{4.0}$, the recidivism-dependent variable comes from its own sample ($n = 3973$; recidivism of 246 cases). For the calculations, the 20th version of the IBM SPSS (Statistical Package for Social Sciences) statistical program was used.

Results

Intra-observer Reliability and Convergent Validity

The tool reliability analysis, as measuring accuracy, showed $R_{xy} = 0.734$, $p = 0.000$; CI = 95% values. The convergent validity, which is referred to an external criterion (RVD-BCN) for the ordinals risk measures, showed a Kendall's Tau-b t (296) = 0.635, $p = 0.000$, CI = 95% values.

VPR_{4.0} Form Performance Parameters

The performance parameters oriented to the predictive validity, which were calculated with 3 and 6 months repeated measures, were the basis of the validation. In Table 3, not only the discrimination indexes are summarized—OR, AUC, sensitivity, and specificity—but also the calibration indexes—PPV and NPV; as it can be seen, the validation parameter for the sensitivity measurement ranges between 82 and 79%, and for the specificity measurement, it ranges between 32 and 33%. Both of them report on the proportion of recurrent cases classified as risk cases and the proportion of no recurrent cases classified as unappreciated risk cases, respectively. The PPV indicates the proportion or percentage of individuals classified as risk and recidivist people, ranging between 5 and 8%—with 4.3 and 6.8% prevalence. The NPV of the proportion of individuals classified as unappreciated risk and not recidivist people ranged between 98 and 95%.

The area under the curve (AUC) is an important index to know the instrument's likelihood of classifying two people correctly as whether or not recidivist, adding value to the predictions' magnitude through the optimal threshold that balances sensitivity and specificity. The most serious violence estimation reaches a better efficiency and shows the AUC value = .658 (Fig. 1), which is accompanied by other $VPR_{4.0}$ performance parameters submitted.

The $VPR_{4.0}$ case classification by risk levels—when serious violence is estimated—provides information about the instrument calibration, and it proves necessary to efficiently organize the victims' protection. This way, despite most of the cases are at medium risk (182 of the 487 most serious cases) within the high-risk category, 25.2% corresponds to serious episodes and 35.2% corresponds to the extreme risk category, which is above any other risk levels. The II type error or false negative has a 6.1% percentage at the unappreciated risk level category.
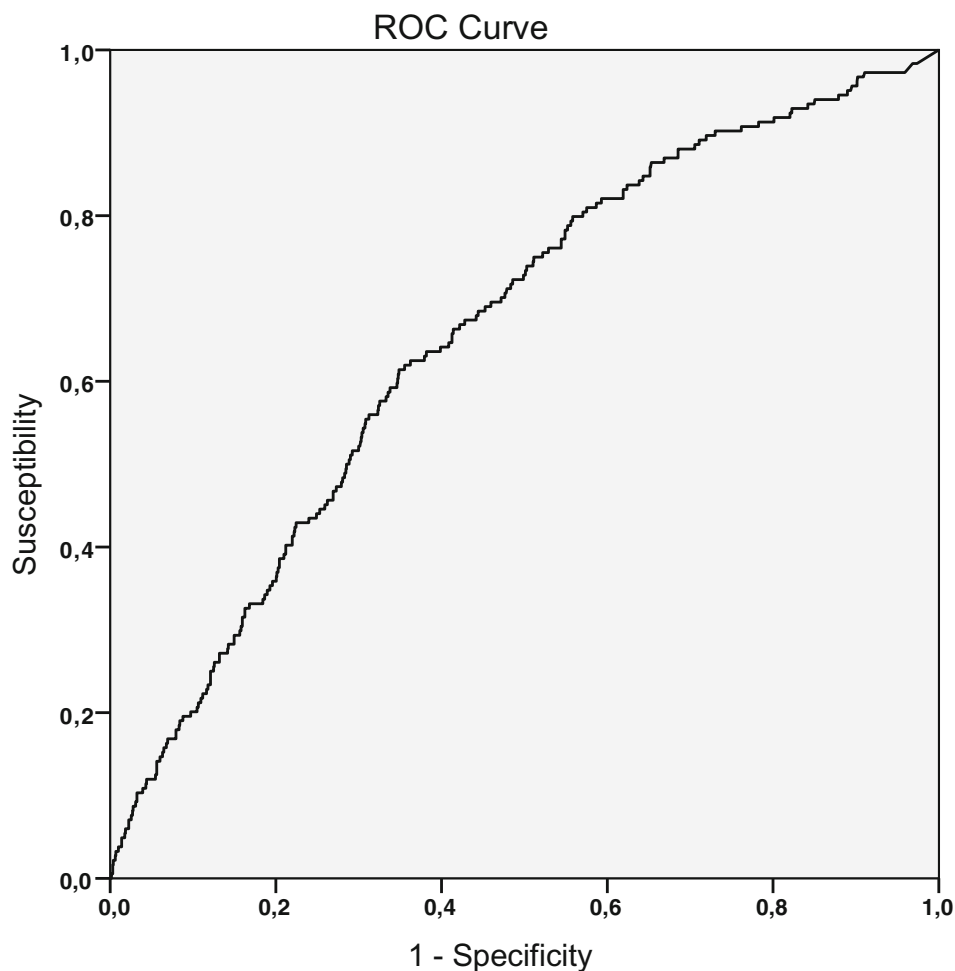
Table 3 Data of the 40% ($n = 2706$) of the sample used for the cross-validation in the different performance parameters for the VPR4.0 form and recidivism from 3 to 6 months, with prevalence of 4.6% and 7.4%, respectively, as well as its combination with serious violence

Parameters	VPR$_{4.0}$ and recidivism to 3 months	VPR$_{4.0}$ and recidivism to 6 months	[b]VPR$_{4.0}$ and serious violence	[c]VPER-S 4.0	VPER-C 4.0
Odds ratio (OR)	2.19 [1.36–3.54]	1.84 [1.28–2.66]	3.10 [1.90–5.06]	10.04 [6.92–14.58]	11.65 [7.66–17.70]
Sensitivity	.82	.79	.90	.86	.90
Specificity	.32	.33	.26	.60	.57
Area under the curve (AUC)	.63 [.58–.67]	.60 [.57–.64]	.66 [.62–.69]	.82 [.79–.85]	.82 [.79–.84]
Positive predictive value (PPV)	.05	.08	.15	.12	.12
Negative predictive value (NPV)	.98	.95	.95	.98	.99

CI = 95%

[b] Data obtained from the sample used for the violence severity estimation ($N = 3725$)

[c] Data from the sample used for the two VPER$_{4.0}$ versions validation ($n = 3973$) with 246 recidivist cases (CI = 95%)

ROC Curve

	Area	SE	Sig	Low.	Upp.	CI 95%
VPR$_{4.0}$	.658	.017	.000	.625	.692	

Fig. 1 ROC curve values for VPR$_{4.0}$ and serious violence using 40% of the validation sample
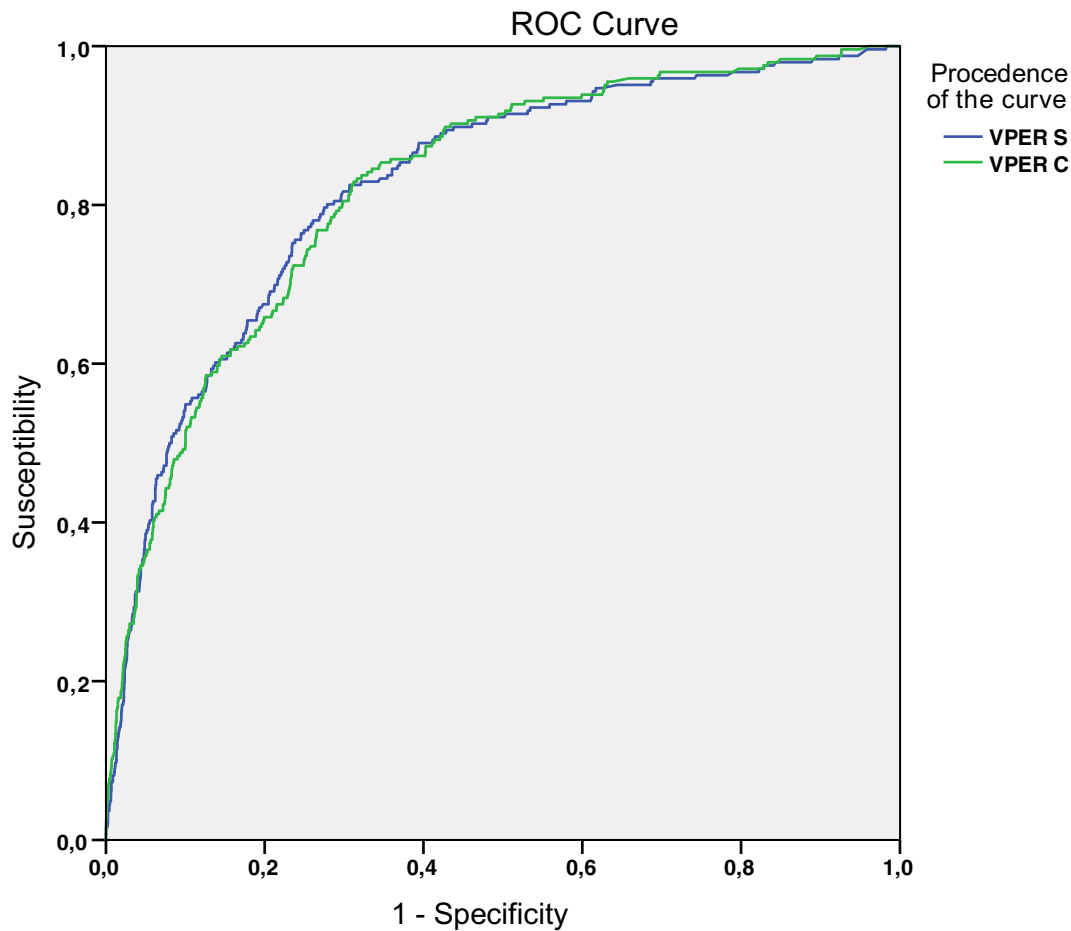
VPER$_{4.0}$ Form Performance Parameters

The VPER-C$_{4.0}$ version (with incident) showed an OR = 11.65 (CI = 7.66–17.70), and the VPER-S$_{4.0}$ version (without incident) showed an OR = 10.04 (CI = 6.92–14.58)—with a low-risk cut-off point. The sensitivity reached by the C and S versions ranges between 90%–95%, with 28%–38% specificity, respectively. Table 3 shows the parameters obtained with both versions to estimate the recidivism from cases used in both instrument versions construction with its validation sample (n = 3973) and 246 recidivist cases, as well as the VPR$_{4.0}$ form parameters.

The NPV shows how the form rules out risk with a high degree of accuracy, exceeding both versions a 95% (.99–.98). This way, the proportion of individuals that were classified as low risk and not recidivist redounds in a very high NPV for this risk level, which is intended to rule out any risk, an objective of vital interest in police matters. The AUC shows predictive values above .80 for both VPER$_{4.0}$ versions (Fig. 2).

The VPER-C$_{4.0}$ form classifies a 6.5% of recidivist cases at unappreciated risk level, and the VPER-S$_{4.0}$ version classifies a 7.3% (see Table 4). The high-risk level classifies the highest recidivism percentage, reaching a 35.4%. The recidivist case classification by risk levels increases progressively until reaching the extreme level, and 69 cases of the 202 in that category were recidivist cases (34.2%). Both versions keep very similar classification parameters in relation to their predictive capacity, showing differences at the highest-risk levels, especially at the extreme risk level, where the VPER-C$_{4.0}$ classifies 69 cases against the 40 cases classified by the VPER-S$_{4.0}$. That way, the VPER-C$_{4.0}$ form classifies 4.5% of the serious cases with unappreciated risk with a 1.4% error within its category. The sum of the medium, high, and extreme categories collects 85.8% of the cases considered as serious cases, and the extreme-risk category classifies

Fig. 2 ROC curve recidivism values for VPER$_{4.0}$ with and without incident

VPER 4.0	Area	SE	Sig	Low.	Upp.	CI 95%
VPER-C $_{4.0}$	.823	.014	.000	.796	.849	
VPER-S $_{4.0}$	.824	.014	.000	.797	.851	

Table 4 Contingency table for the VPER-C4.0 form according to whether risk exists or not and the violence severity

VPER-C4.0 assessment	Serious violence	Mild/absent violence	Total
Risk	445	1250	1695
% high risk level	26.3	73.7	100
% violence severity	91.4	38.6	45.5
[b]Low risk	42	1988	2030
% low-risk level	2.1	97.9	100
% violence severity	8.6	61.4	54.5
Total	487	3238	3725

Note: Data from the total sample ($n = 3725$) with 487 serious cases

[a] The odds ratio for risk is 16.851 [12.185–23.303]

[b] The low-risk category includes not estimated and low levels

56.1% of those cases as extreme-risk cases. The VPER-C$_{4.0}$ classifies correctly 91.4% of the most serious cases, considered as medium, high, and extreme risk (Table 4).

Discussion

The objective of the forms shown is to identify the subjects with higher risk to assign protection resources in the most efficient way. The assessment protocol and the validated forms in this study offer an answer to this need, aiming by combining the VioGén System dual protocol (VPR$_{4.0}$ and VPER$_{4.0}$) to cut down recidivism—especially the most serious cases—due to the identification of high-risk subjects. The risk management approach is oriented to the inactivation of dynamic-relational risk indicators starting from the law enforcement intervention adjusted to the permanent classification of risk levels.

VPR$_{4.0}$ and VPER$_{4.0}$ forms (in both versions) were created and validated following the technical criteria and using the parameters that specialized literature advises for violence risk assessment tools (Rice et al. 2013; Rossegger et al. 2014; Singh 2013). After a prior study, risk indicators with empirical evidence have been considered, generating a calculation algorithm through bivariate statistical analysis when realizing that the interactions between some indicators were normal for these tools and not a problem, as it happens in psychometric instruments.

The calculations made with logistic regression tests show that some indicators could be mathematically disregarded. However, they were included because in doing so, the predictive capacity was not altered; in addition, the indicators serve the law enforcement agents in managing the victim's risk in an individualized manner. Appropriate adjustments should be made in subsequent robustness studies of the forms, with a different sample and with another temporal period.

The obtained results, especially regarding predictive validity, are comparable with the ones reported in the national and international literature for similar tools, especially the most famous ones (Muñoz and López-Ossorio 2016), since meta-analysis studies show that the found AUC values are in the ranges obtained in other research (Fazel et al. 2012). Here, the VPR$_{4.0}$ got an AUC value of .658, and both versions of the VPER$_{4.0}$ form obtained a value slightly over .80. So, we consider that the prediction parameters of this kind of behaviors cannot be greatly improved and that, as it happens in other contexts, the progress will come, as integrated concepts, from the right classification of cases according to their risk, optimizing the risk management.

The new VPR$_{4.0}$ form includes risk of violence indicators, specific of IPV, as well as indicators of recidivism in violent acts, integrated in different thematic nucleus noted in the specialized literature. The scale distribution of the cutoff points for the VPR$_{4.0}$ according to the recidivism probability, severity, and law enforcement response capability shows good results for a screening tool used by the agents during the first approach to the case through limited information. The data shows that it has a high sensibility, with a good margin of error to discard risk from the 20% of recidivism and a measurement error of 5.1% for its level of risk (false negatives). The adjustment of the severe cases stabilizes at a rate of false negatives of 6.1%. In this way, it intends to group the severe cases in the higher risk categories. As the studies that analyze this kind of tools conclude (Fazel et al. 2012), the VPN is significant and suggests that this tool can efficiently detect individuals with low recidivism risk, sacrificing the VPP due to the context nature. So, actuarial tools seem to fit better in the risk management plans (especially in law enforcement context) in selected risk groups, comparing individuals with the reference group.

The AUC value obtained in VPR$_{4.0}$ in severity of violence is higher than in estimating the recidivism (.66. vs. .63); they are both in the average reported by other similar tools, where the AUC values of the review studies range between .54 and .67 (Messing and Thaller 2013). The aim of stabilizing the predictions, key element for making a more appropriate risk management, has been reached. The OR also shows that the tool's risk assessment increases more than twice the risk of recidivism in probabilistic terms and more than three times the risk of it being severe—the OR average values published in international review studies ranges between 1.2 and 7.9 (Singh et al. 2011). In future maintenance of the VPR$_{4.0}$, this should be an issue for further research with the aim of improving the form. The most plausible explanation of obtaining moderate predictive results of VPR$_{4.0}$ comes from the difficulties law enforcement officers have when gathering information in early stages, inherent limitation to the initial exploration context.

The great challenge of the developed research work, aimed at improving protocol forms, was to create and validate a new

VPER form containing recidivism risk indicators as well as protective factors related to recidivism. The new validated VPER$_{4.0}$ form is built as a tool of semi-actuarial nature because it uses dynamic and static indicators (of risk and protection) through a combination that is not performed when following professional structured judgment procedures. This form improves the performance parameters of VPR$_{4.0}$ overcoming its access and quality limitations of the available information during the first assessment, positioning itself as the key element of the general protocol. When the protective factors included in the VPER$_{4.0}$ are present, they provide a good prediction of the cases. This new form version has a completely autonomous functioning with its own algorithm, and it is capable of making predictions through a mixed set (risk-protection) of indicators that provides the possibility of assessing the probability of recidivism and at the same time, monitoring the changes that occur over time.

The VPER$_{4.0}$ form assessments are sensitive to the different risk scenarios based on the factors and indicators integrated in thematic dimensions of great range, accuracy, and suitability. Thus, the building of two versions of the VPER$_{4.0}$ form allows addressing the two situations given in the cases: positive evolution of cases and negative evolution with new violent incidents. These scenarios are so different that they have justified the need for building two different versions of the VPER$_{4.0}$ form.

The VPER-S$_{4.0}$ version is capable of monitoring the progress of the events and to reduce the risk level, if possible, to an unappreciated level. At the same time, it is sensitive to incidences and characteristics of the aggressor that may increase the probabilities of a new violent event. The VPER-C$_{4.0}$ not only includes the shorter version's indicators, but it also integrates a set of indicators, which are focused on investigating the violent event or the characteristics of a possible violation of the judicial measures with the additional goal of reducing the multi-recidivism.

With the same goal, the form ends with an indicator oriented to know the victim's perception of its own risk, as it also happens in the VPR$_{4.0}$, even though after recording the victim's response, the professional performing the assessment must answer a control question to establish to what extent the victim may be underestimating her own risk. This estimation allows the police officers, specialized in victim's protection, to contribute through an interview to the better adjustment of the victim's perception of its own risk and self-protection.

For its correct use, professionals must be adequately trained on the principles of the risk-based approach and the use of risk assessment tools, with specific IPV training and assessment guides' support (Kroop 2008). The need for incorporating interpretative manuals to guide the professionals in this labor is mentioned by many risk assessment tools (Douglas et al. 2014) and is part of the assessment protocol.

The IPV risk assessment performed by police officers is an important part of criminal policy aimed at reducing recidivism

levels from a quantitative and qualitative point of view (Capdevila 2015). However, the law enforcement tasks are always limited, and it is essential that other agents help in the victim's protection, integrating other agencies in an integrated and multidisciplinary way and articulating agile and appropriate mechanisms. Accordingly, there is a pending challenge in the women's homicide risk assessment in the intimate partner context, showing that in the homicide phenomenon and the recidivism of violent behavior, even though it is possible that they share some risk indicators—still to be determined by empirical research that is taking place in Spain—they present important differences that restrict the use of violence risk assessment tools. Therefore, we should create and validate specific instruments for each kind of event.

Conclusions

The protocol's instruments identify the recidivism and severe violence risk through the empiric and additive weighing of the associated weight to its factors, with the best available and optimized professional knowledge for the law enforcement context. The validated VioGén System actuarial tools, VPR$_{4.0}$ and VPER$_{4.0}$, support the strong evidence of the need to plan the victim's protection according to the risk assessment, being preferable to the JPE methodologies, because of their transparency and reliability not being designed for the assessment of psychological aspects or constructs but to make predictions (Hart 2008). It also has been deemed as relevant to report performance parameters following the recommended indications for the actuarial tools, easing the transparency and comprehension of the data provided during the assessments (Neller and Frederick 2013). On the other hand, the limitation for its moderated PPVs is presented, circumstance that would restrict its utility in other contexts, but in the law enforcement context are advisable because of its low impact or interference in the aggressors' life (Martínez 2016).

The VPR$_{4.0}$ form works as an initial screening in the assessment process that articulates the global protocol, where the VPER$_{4.0}$ is notable for carrying the weight of the assessment during the risk management performed by the specialized officers. Its two versions—with and without incident—show high quality technical properties and appropriate discrimination and calibration parameters. In conclusion, these risk assessment forms are one of the greatest assessment and risk management tools for IPV close to the most renowned tools in the international context. These forms have a high impact in the law enforcement context, since risk assessment and risk management, as connected pieces, constitute the cornerstone of modern crime prevention strategies, especially in the case of IPV.

This study faces some important limitations. Some of them are related to the nature of the transnational and ecological

perspective of this type of research, which is conducted in real life and applied settings, as opposed to a laboratory or a controlled setting. Taking this into regard, the practitioners who completed the protocols were experts and senior law enforcement officers, and this could have affected the performance of the risk assessment tool. Although this research approach has many advantages, it may also reduce inherent guarantees of the designed experiment.

Here, it is worth mentioning that another consequence of studies that focuses on the predictive validity of the tools, which estimate the probability of future violence with a subsequent process of risk management, is that it is not possible to know for certain if the police intervention had any effects on the final recidivism outcome. This may affect the higher-risk cases during the first months.

Another methodological limitation is related to the law enforcement approach of prioritizing the security and the safety of the victim within the scope of a police intervention. It is important to bear in mind that the actuarial instruments seek to compare a subject with a reference group, demonstrating again the dilemma between the ideographic and the nomothetic perspective where the risk variables of a specific case may not be well represented in the actuarial assessments (Nicholls et al. 2013). In order to avoid the limitations of the procedures based on fixed rules, it is recommended to shift to "adjusted" procedures where the evaluator can modify the automatically provided level of risk.

Yet, another relevant limitation is related to the research design. The main methodological strengths of the case-control designs turn out to also be a relative limitation. The case-control design used for the validation of the VPR$_{4.0}$ is in our opinion appropriate for this second tool, but this type of design does not offer the same guarantees of prospective longitudinal studies. The threats inherent to these research designs of an epidemiological nature were considered with the intention of reducing their impact as much as possible. We focused on the selection and information biases taking great care when accessing the primary data, and we ensured that the risk factor was estimated before the outcome (recidivism) was recorded.

Compliance with Ethical Standards

Conflict of Interests The authors declare that there is no conflict of interest.

Ethical Statement This research was not funded. This article does not contain any studies with human participants performed by any of the authors. This research did not require any informed consent.

References

Álvarez M, Andrés Pueyo A, Augé M, Choy A, Fernández C, Foulon H et al (2011) Protocolo de Valoración del Riesgo de Violencia contra la Mujer por parte de su pareja o ex pareja (RVD-BCN). Circuito Barcelona Contra la Violencia hacia las Mujeres, Barcelona

Andrés Pueyo A, López S, Álvarez E (2008) Valoración del riesgo de violencia contra la pareja por medio de la SARA. Papeles del Psicólogo 29(1):107–122

Arbach-Lucioni K, Andres-Pueyo A (2016) Violence risk assessment practices in Spain. In: Signh JP, Bjorkly S, Fazel S (eds) International perspectives in violence risk assessment. Chapter 19. Oxford University Press, Oxford

Arbach-Lucioni K, Desmarais S, Hurducas C, Condemarin C, Kimberlie D, Doyle et al (2015) La práctica de la evaluación del riesgo de violencia en España. Revista de la Facultad de Medicina 63(3): 357–366. https://doi.org/10.15446/revfacmed.v63n3.48225

Belfrage H, Strand S, Storey J, Gibas A, Kropp P, Hart S (2012) Assessment and management of risk for intimate partner violence by police officers using the Spousal Assault Risk Assessment Guide. Law Hum Behav 36(1):60–67. https://doi.org/10.1037/h0093948

Buchanan A, Binder R, Norko M, Swartz M (2012) Resource document on psychiatric violence risk assessment. Am J Psychiatry 169:1–10. https://doi.org/10.1176/appi.ajp.2012.169.3.340

Canales D, Macaulay A, McDougall A, Wei R, Campbell J (2013) A brief synopsis of risk assessment screening tools for frontline professionals responding to intimate partner violence. Centre for Criminal Justice Studies, New Brunswick

Capdevila M (ed) (2015) Tasa de reincidencia penitenciaria 2014. CEJFE, Barcelona. Retrieved February 20, 2019 from: http://cejfe.gencat.cat/web/.content/home/recerca/cataleg/crono/2015/taxa_reincidencia_2014/tasa_reincidencia_2014_cast.pdf

Carvajal A, Centeno C, Watson R, Martínez M, Sanz A (2011) ¿Cómo validar un instrumento de medida de salud? An Sist Sanit Navar 34(1):63–71

Dayan K, Fox S, Morag M (2013) Validation of spouse violence risk assessment inventory for police purposes. J Fam Violence 28(8): 811–821. https://doi.org/10.1007/s10896-013-9547-4

De Becker G (1997) The gift of fear. Little, Brown y Co., Boston

De Becker G et al (2000) Domestic violence method (DV MOSAIC). Retrieved February 20, 2019 from: https://www.mosaicmethod.com

Douglas KS, Shaffer C, Blanchard AJE, Guy LS, Reeves K, Weir J (2014) HCR-20 violence risk assessment scheme: overview and annotated bibliography, HCR-20 violence risk assessment white paper series, 1. Mental Health, Law, and Policy Institute. Simon Fraser University, Burnaby, Canada

Echeburúa E, Amor PJ, Loinaz I, De Corral P (2010) Escala de Predicción del Riesgo de Violencia Grave contra la Pareja-Revisada-(EPV-R). Psicothema 22(4):1054–1060

Fazel S, Singh J, Doll H, Grann M (2012) Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people. Systematic review and meta-analysis. Br Med J 345:4692. https://doi.org/10.1136/bmj.e4692

Folino JO (2015) Predictive efficacy of violence risk assessment instruments in Latin-America. European Journal of Psychology Applied to Legal Context 7(2):51–58. https://doi.org/10.1016/j.ejpal.2014.11.006

Hart S (2008) Preventing violence: the role of assessment and management. In: Constanza BA, Willem WF (eds) Intimate partner Violence prevention and intervention. Nova Science Publishers Inc., New York, pp 7–18

Hilton Z, Harris G, Rice M, Houghton R, Eke A (2008) An indepth actuarial assessment for wife assault recidivism: the domestic violence risk appraisal guide. Law Hum Behav 32(2):150–163. https://doi.org/10.1007/s10979-007-9088-6

Kroop P (2008) Intimate partner violence risk assessment and management. Violence Vict 2:202–220

Kropp P, Hart S, Webster CD, Eaves D (1995) Manual for the spousal assault risk assessment guide, 2nd edn. British Columbia Institute on Family Violence, Vancouver

Kropp P, Hart S, Belfrage H (2010) Brief spousal assault form for the evaluation of risk (B-SAFER) second edition. User manual. Proactive Resolutions, Canadá

Llor-Esteban B, García-Jiménez JJ, Ruiz-Hernández JA, Godoy-Fernández C (2016) Profile of partner aggressors as a function of risk of recidivism. Int J Clin Health Psychol 16:39–46. https://doi.org/10.1016/j.ijchp.2015.05.004

Loinaz I (2017) Manual de evaluación del riesgo de violencia. Metodología y ámbitos de aplicación. Madrid, Pirámide

López-Ossorio JJ, González JL, Andrés-Pueyo A (2016) Eficacia predictiva de la valoración policial del riesgo de la violencia de género. Psychosoc Interv 25:1–7. https://doi.org/10.1016/j.psi.2015.10.002

López-Ossorio JJ, González JL, Buquerín S, García L, Buela-Casal G (2017) Risk factors related to intimate partner violence police recidivism in Spain. Int J Clin Health Psychol 17:107–119. https://doi.org/10.1016/j.ijchp.2016.12.001

Martínez L (2016) Errores conceptuales en la estimación de riesgo de reincidencia. La importancia de diferenciar sensibilidad, valor predictivo y estimaciones de riesgo absolutas y relativas. Revista Española de Investigación Criminológica 14:1–31

Messing J, Thaller J (2013) The average predictive validity of intimate partner violence risk assessment instruments. J Interpers Violence 28(7):1537–1558. https://doi.org/10.1177/0886260512468250

Muñoz JM, López-Ossorio JJ (2016) Valoración psicológica del riesgo de violencia: alcance y limitaciones para su uso en el contexto forense. Anuario de Psicología Jurídica 26(1):130–140. https://doi.org/10.1016/j.apj.2016.04.005

Neller D, Frederick R (2013) Classification accuracy of actuarial risk assessment instruments. Behav Sci Law 31:141–153. https://doi.org/10.1002/bsl.2047

Nicholls TL, Pritchard MM, Reeves KA, Hilterman E (2013) Risk assessment in intimate partner violence: a systematic review of contemporary approaches. Partner Abuse 4(1):76–168. https://doi.org/10.1891/1946-6560.4.1.76

Otto R, Douglas K (eds) (2010) Manual de evaluación del riesgo de violencia. Taylor & Francis Group, Nueva York

Rice M, Harris G, Lang C (2013) Validation of and revision to the VRAG and SORAG: the Violence Risk Appraisal Guide—Revised (VRAG-R). Psychol Assess 25(3):951–965. https://doi.org/10.1037/a0032878

Roaldset JO, Hartvig P, Bjørkly S (2017) Psychometric properties and predictive validity of a police version of a violence risk screen—a pilot study. Int J Law Psychiatry. https://doi.org/10.1016/j.ijlp.2017.06.007

Rossegger A, Gerth J, Seewald M, Urbaniok M, Singh J, Endrass J (2013) Current obstacles in replicating risk assessment findings: a systematic review of commonly used actuarial instruments. Behav Sci Law 31:154–156. https://doi.org/10.1002/bsl.2044

Rossegger A, Endrass J, Gerth J, Singh JP (2014) Replicating the violence risk appraisal guide: a total forensic cohort study. PLoS One 9(3):e91845. https://doi.org/10.1371/journal.pone.0091845

Singh J (2013) Predictive validity performance indicators in violence risk assessment: a methodological primer. Behav Sci Law 31:8–22. https://doi.org/10.1002/bsl.2052

Singh J, Grann M, Fazel S (2011) A comparative study of risk assessment tools: a systematic review and metaregression analysis of 68 studies involving 25,980 participants. Clin Psychol Rev 3:499–513. https://doi.org/10.1016/j.cpr.2010.11.009

Singh J, Fazel S, Gueorguieva R, Buchaman A (2014) Rates of violence in patients classified as high risk by structured risk assessment instruments. Br J Psychiatry 204:180–187. https://doi.org/10.1192/bjp.bp.113.131938

Singh JP, Yang S, Mulvey E (2015) Reporting guidance for violence risk predictive validity studies: the RAGEE statement. Law Hum Behav 39:15–22. https://doi.org/10.1007/s10979-011-9278-0

WAVE- Women Against Violence Europe (2011) Proyect—identifying and protecting high risk victims of gender based violence—an overview, Report of the DAPHNE Protect. Vienna. Retrieved February 20, 2019 from: http://fileserver.wave-network.org/trainingmanuals/PROTECTI_Protecting_High_Risk_Victims_2011_English.pdf

WAVE- Women Against Violence Europe (2012) Proyect II—European network and European info CENTRE AGAINST VIOLENCE, Report of the DAPHNE Protect. Vienna. Retrieved February 20, 2019 from: http://fileserver.wave-network.org/trainingmanuals/PROTECTII_Risk_Assessment_and_Safety_2012_English.pdf