



# The Predictive Validity of Intimate Partner Violence Risk Assessments Conducted by Practitioners in Different Settings—a Review of the Literature

Klara Svalin<sup>1</sup> · Sten Levander<sup>1</sup>

Published online: 10 August 2019  
© The Author(s) 2019

## Abstract

Intimate partner violence (IPV) is a global health problem with severe consequences. One way to prevent repeat IPV is to identify the offender's risk of recidivism by conducting a risk assessment and then implement interventions to reduce the risk. In order to be effective, accurate risk assessments and effective interventions are required. Practitioners in different settings are conducting IPV risk assessments, but the predictive validity of practitioners' IPV assessments has not been studied via a comprehensive literature search. This is the overall aim of the present study. The literature search was conducted in five different databases and at three different publisher sites. The selection of studies was based on nine different inclusion and exclusion criteria. The number of studies that fulfilled the criteria was unexpectedly small ( $N = 11$ ). One of the studies was conducted in a treatment setting, the others in criminal justice settings. The predictive accuracy for the global risk assessments ranged from low to medium. The role of treatment or other interventions to prevent repeat IPV had been analyzed in one way or another in eight of the studies. There is a knowledge gap, the reasons of which are discussed.

**Keywords** Intimate partner violence · Criminal justice settings · Violence risk assessment · Predictive validity

## Introduction

Intimate partner violence (IPV) refers to “any behavior within an intimate relationship that causes physical, psychological, or sexual harm to those in the relationship” (World Health Organization [WHO] 2012, pp. 1). IPV is a global health problem with severe consequences. Victims of IPV are traumatized physically and mentally. The economic burden for society is high. With respect to serious violence, the common finding is that women are victimized by men (World Health Organization/London School of Hygiene and Tropical Medicine [WHO/LSHTM] 2010). According to García-Moreno et al. (2013), 38% of all murdered women globally are killed by an intimate partner. Motivated by the prevalence and severity of IPV against women, noting that the problem still has low priority, researchers have called for action (García-Moreno et al. 2015), for instance by educating health

workers in how to identify and support victims. The need for further research on how to manage IPV is also emphasized.

IPV offenders often relapse into new IPV offenses. For instance, in a recent study based on a population with 336 male and female victims of IPV, Rahman (2018) reported that 43% were victimized repeatedly within 12 months. Similar numbers were presented in Lin et al. (2009); 48% of the offenders relapsed into overall violence within 3 months. Petersson and Strand (2017) compared rates of IPV recidivism among antisocial and family-only offenders. After 50 months, 27% of the antisocial offenders and 13% of the family-only offenders had recidivated into new IPV offenses.

One way to prevent repeat IPV is to identify the offender's risk of recidivism by conducting a risk assessment and then implement interventions to reduce the risk. Risk assessments have been identified as a cornerstone in IPV prevention (Kropp 2004). There are different approaches to violence risk assessment—which can be classified by four components in the risk assessment procedure: identifying risk factors, measuring risk factors, combining risk factors, and produce a final risk assessment (Skeem and Monahan 2011). The least structured approach is the clinical judgment; the rater assesses the risk based on his/her professional experiences, and the most structured approach is the actuarial approach, in which all four

✉ Klara Svalin  
klara.svalin@mau.se

<sup>1</sup> Department of Criminology, Malmö University, Malmö, Sweden

components are structured (ibid). Thus, the actuarial assessment is based on risk factors identified in empirical studies, which are measured and combined due to specific guidelines (Hart 1998). The final risk assessment is usually based on statistical measures (Hart et al. 2007). There are some different kinds of semi-structured tools and approaches. For instance, the structured professional judgment (SPJ) approach which is a combination of the clinical and actuarial approach (Nicholls et al. 2013). The rater often uses a tool with risk factors,<sup>1</sup> but in addition to those, case-specific factors, if any, should be added (see for instance B-SAFER; Kropp et al. 2010). The final risk assessment is based on the risk factors and the rater's judgment (Nicholls et al. 2013). The overall aim of SPJ assessments is to prevent repeat violence, and a part of the procedure is therefore the risk management, which should be based on the risk assessment (Douglas and Kropp 2002).

IPV is a type of violence that is often faced by practitioners in criminal justice and health contexts, and many practitioners meet victims of IPV soon after the violence has occurred. Since repeat IPV victimization usually occur close to previous IPV exposure (Mele 2009), they have the potential to play a significant role in victim protection. It may even be a matter of saving lives. For instance, practitioners working in emergency departments meet victims of IPV, and they not only play an important role in the immediate situation but also in identifying the risk for repeat IPV. In some emergency departments, screening tools are used to make decisions about interventions in IPV cases (e.g., Koziol-McLain et al. 2010). Furthermore, social workers meet victims of IPV in many different situations, which Danis (2003) has argued gives them an opportunity to identify those at high risk of IPV and to intervene in those cases. This presupposes that the violence risk assessments are accurate, and there is no point and perhaps a risk in itself to administer interventions to false positives. Consequently, a key aspect concerns the accuracy of violence risk assessments (predictive validity). According to Messing and Thaller (2013), this is the most important aspect of the efficacy measures.

Knowledge on the predictive validity of IPV risk assessments has been examined and summarized in a number of recent review studies, e.g., Graham et al. (2019), Helmus and Bourgon (2011), Messing and Thaller (2013), and Nicholls et al. (2013), all of which are described below. However, the predictive validity of IPV assessments conducted by practitioners in different settings has not been the focus in any of these reviews. Citing Graham et al. (2019, pp. 18): "It is imperative that future research investigate the psychometric properties of IPV/IPH<sup>2</sup> risk assessment tools administered by service providers in real-world settings and

the feasibility of typical providers' appropriate and successful use of these tools". This will be the focus of the present study.

The following questions will be examined: How accurate are practitioners' intimate partner violence risk assessments with regard to repeat IPV? Which practitioner groups had conducted the assessments in the studies under review, and what were their characteristics in terms of violence risk assessment education/training? An important part of the violence risk assessment procedure usually involves the implementation of interventions to protect victims and prevent offenders from engaging in repeat violence. Since such interventions are intended to prevent re-victimization, they should be considered in the evaluation of predictive validity in relation to repeat violence (see Belfrage 2008). This is the next question examined in the study: the role of protective measures in the examination of predictive validity. Finally, a number of previous studies have highlighted the fact that tools are not always used as recommended in the guidelines (e.g., Wong and Hisashima 2008). These findings will be described in more detail below, and the question of whether the tools evaluated in the studies were used as recommended will also be studied. By reviewing the knowledge regarding practitioners' IPV risk assessments, we will hopefully learn more about the usefulness of such assessments and find guidance regarding the work that remains to be done in the fields of both practice and research.

### Previous Studies of the Predictive Validity of IPV Violence Risk Assessments

A recent study examined the average predictive validity of five different IPV risk assessment tools (ODARA, SARA, DA, DVSI, K-SID) and victim assessments (Messing and Thaller 2013). The data were based on results obtained in ten previous studies, all of which examined the accuracy of the tools by measuring the area under the curve (AUC) of the receiver operating characteristic (ROC). The AUC statistic is the ratio of correct predictions. If AUC is 0.75, three of four predictions are correct. The distribution of prediction errors (false positives and negatives) is analyzed separately and not considered in this context. Results from analyses in which protective actions were controlled for were not included, nor were studies of risk assessment tools that had only been evaluated once (Messing and Thaller 2013). The ODARA produced the highest average AUC score (=0.67) which corresponds to a moderate effect size. The average AUCs of the other tools and victim assessments varied between 0.54 and 0.63, and the effect sizes were small.

The predictive/postdictive validity of IPV risk assessment tools was also examined in another review study, which was based on 39 publications identified by means of a systematic literature review (Nicholls et al. 2013). These studies represented all English-language publications on the subject from

<sup>1</sup> Sometimes, victim vulnerability factors are also included in SPJ tools, for instance, SARA (Kropp and Hart 2015) and B-SAFER (Kropp et al. 2010).

<sup>2</sup> IPH = intimate partner homicide

western nations written between 1990 and 2011. For most of the tools, ROC analyses had been conducted. The AUC values varied substantially (0.48–0.92). A closer look at the studies with the highest AUC values showed that the tools used were actuarial tools based on victim/offender questionnaires or interviews. In addition, some of the risk measures were based on victim appraisals.

Helmus and Bourgon (2011) reviewed 15 years of the use of the SARA tool. At the time of the study, 11 studies on the predictive accuracy of the SARA tool had been published and were included in the review. The AUCs (for total score or global risk assessment) varied between 0.59 and 0.87. The highest AUC value was found in a study conducted in Spain, based on 102 provincial court cases (Andrés-Pueyo et al. 2008). The assessments were produced in retrospect and the follow-up period was 12 months (outcome: IPV recidivism). The AUC of the SARA total score/global risk assessment was 0.77/0.87. The study with the second highest AUC values (a conference presentation, Gibas et al. 2008) was conducted in Canada, based on a federal treatment sample ( $N=108$ ) with correctional staff conducting the assessments. The predictive accuracy (AUC) of the SARA total score/global risk assessment (with IPV recidivism as the outcome measure) was 0.70/0.76.

In the most recent of the review studies, Graham et al. (2019) examined the reliability and validity of IPV/IPH risk assessment tools. They also studied the feasibility of the use of such tools among practitioners. The results are based on 42 articles, including 43 studies examining 18 tools in total. In almost half of the studies ( $n=21$ ), the assessments were conducted by researchers and in the other half ( $n=22$ ) by practitioners. For 12 out of 18 tools, AUC values were provided. In line with most of the previously presented review studies, the AUCs varied substantially. The lowest AUC was 0.51 and the highest 0.86. However, since different outcome measures and follow-up times were employed, direct comparisons were not meaningful. Information on the feasibility of the use of the tools in practitioner settings was scarce; only 1 out of 43 studies specifically discussed this question. The formulation of the questions in the tool and the routines of the assessment procedure are examples that were discussed.

In summary, all four review studies included a mix of risk raters; for example, in some of the studies, risk was assessed by practitioners, in others by researchers. There was an overlap of two studies that were included in all four reviews. The AUCs in both Graham et al. (2019), Nicholls et al. (2013), and Helmus and Bourgon (2011) varied greatly, whereas the range of the AUC values (average AUC values) was smaller in Messing and Thaller (2013). The tool associated with the highest AUC value in Graham et al. (2019) and Nicholls et al. (2013) was the Danger Assessment scale (DA, AUC = 0.86 and 0.92, respectively) and in Messing and Thaller (2013), it was the ODARA (0.67). The highest AUC value

in the SARA studies reported by Helmus and Bourgon (2011) was reported as 0.87. There are many factors that influence predictive validity, e.g., information sources (for the assessment and for recidivism), definitions of IPV, length of follow-up times, and outcome measures (Nicholls et al. 2013), which means that the variety of such factors complicates comparisons between different studies. Nicholls et al. (2013) suggest the examination of more than one tool in the same study as a means of, at least in part, overcoming this problem.

Previous studies have highlighted a number of problems related to practitioners' use of violence risk assessment tools. One such issue is related to the fact that the tools are not administered in the recommended ways (Her Majesty's Inspectorate of Constabulary [HMIC] 2014; Wong and Hisashima 2008). The first of these studies (HMIC, 2014) examined the use of the DASH tool (domestic abuse, stalking and harassment, and honor based violence; Richards 2009) in a number of police areas in England and Wales. One finding was that the mandatory form, which is a part of the DASH assessment, was not completely followed in a large number of cases in one of the police areas. The second study evaluated probation officers' use of the SARA guide. Information for the assessments was drawn from a database which contained little information regarding victims (Wong and Hisashima 2008). Consequently, the risk management plans for the victims were not as meaningful as they could have been if such information had been available. The authors also concluded that a SARA assessment was completed in less than half of the cases (38%) that should have been assessed (according to specified criteria).

Further, Cattaneo and Chapman (2011) interviewed 13 practitioners working with victims of IPV in different settings, e.g., shelters, courts, and a hospital. A majority of the participants did not use any tools to assess and manage violence risk. Instead, they used their own professional experiences, their colleagues' professional experiences, and their "gut feeling." These means of determining risk were often combined. Such types of assessments are similar to the unstructured clinical approach in the first generation of violence risk assessments. Similar results were found in an inter-rater reliability study that compared police employees' violence risk assessments (Svalin et al. 2017a). Two different tools were evaluated separately. However, the main results were similar for both tools. The global risk assessments were rather consistent across different raters whereas the assessments of the factors included in the tools differed. Thus, it seemed as that the assessment of global risk was based on something other than the factors included in the tool. One suggestion was that the police employees based these assessments on tacit knowledge (gut feeling). Lack of education and training in risk assessment was discussed as a central explanation for the use of tacit knowledge in the police employees' risk assessments. Finally, Cattaneo and Chapman (2011) also studied different

practitioners' use of assessments in management decision-making and found that some of them allowed the assessment to fully guide their work, while others used it only as one part of this process.

## Method

A systematic literature search was conducted in order to find material for the review. Five different databases were chosen based on the aim of the study: to study the IPV risk assessments of practitioners working in different settings. These were Sociological Abstracts, Psychinfo, Cinahl, Pubmed, and Medline. Thus, several different topics, such as psychology, sociology, social work, medicine, psychiatry, and criminology, were covered in the searches. In addition, searches were conducted at three different publisher sites, namely, Taylor & Francis, SAGE Journals and Science Direct (Elsevier).<sup>3</sup> The searches only included studies written in English.

## Procedure

The database searches were conducted on October 24, 2017. No cut-off was specified for the earliest date on which studies were published. As has previously been noted, Nicholls et al. (2013) have recently conducted a comprehensive review on the predictive validity of IPV violence risk assessments. Their literature search was based on four clusters with related terms. The clusters were *intimate partner violence*, *measurement*, *risk assessment*, and *risk*. Since the aim of this study is similar to one of the aims in Nicholls et al. (2013, see aim d, and predictive validity, p. 85), the choice of search clusters and related terms for the present study was inspired by their choices. However, some of their search terms were excluded and some were added, in order to narrow the search further and thus make it more appropriate to the more specific aim of the present study (see Table 1). For instance, our risk cluster only included search terms related to recidivism whereas their risk cluster was broader and included outcomes as “risk” and “dangerousness.”

The final search string (below) was the same in all database searches:

“partner violence” OR “partner abuse” OR “domestic violence” OR “intimate partner violence” OR “wife abuse” OR “wife assault” OR “family violence” OR femicide OR “intimate partner homicide” OR “spouse abuse” OR “spouse assault” OR “spouse violence”) AND (“test validity” OR “statistical validity” OR accuracy OR predict\* OR sensitivity OR

specificity) AND (actuarial OR “risk assessment” OR “structured professional judgment” OR “dangerousness assessment” OR “rating scale” OR “assessment tool” OR instrument OR “Domestic violence risk appraisal guide” OR “Danger assessment” OR “Kingston screening instrument for domestic violence” OR “Ontario domestic assault risk assessment” OR “Spousal assault risk assessment guide” OR “Brief spousal assault form for the evaluation of risk” OR “domestic violence screening instrument” OR “Violence risk appraisal guide” OR “Level of service inventory” OR HCR-20) AND (relapse OR repeat OR re-victimization OR re-abuse OR recidivism).

The searches resulted in the identification of a total of 932 studies (sociological abstracts: 787 studies, Psychinfo: 70 studies, Cinahl: 13 studies, Pubmed: 34 studies and Medline: 28 studies). Once duplicates from the database searches had been excluded (manually), the total number of studies was reduced to 846.

The publisher site searches were conducted on October 25, 2017 (Taylor & Francis) and January 26, 2018 (Sage Journals and Science Direct (Elsevier)). Since the search string used in the database searches was too complex for use at the publishers' sites, the following combination of terms was used: “violence risk assessment” AND “intimate partner violence”. In total, these searches identified 71 studies (Taylor & Francis: 28 studies, SAGE Journals: 34 studies and Science Direct: 9 studies). Articles that had already been identified in the previous searches and duplicates from the different site searches were excluded, resulting in a total of 63 studies.

The next step was the sorting procedure, which was mainly carried out by reading all the abstracts of the identified studies. However, when the information in the abstract was not sufficient to determine whether or not a study would be included, parts of the full text were read. For instance, if information about whether practitioners had carried out the risk assessments in the study was missing in the abstract, the methodology part in the article was read. As soon as a criterion was found not to be fulfilled, the article was excluded. That is, not all inclusion and exclusion criterion were checked in all studies.

## Inclusion and Exclusion Criteria

A number of inclusion and exclusion criteria were used to determine whether or not a study was eligible for inclusion. The first criterion relates to the type of study. Only original articles and dissertations were included, and thus, reviews/research summaries, book chapters, conference contributions, and editorials were excluded. The studies' abstracts were also of significance in the sorting process. The abstract had to state that the predictive validity of IPV risk assessments was going to be evaluated. Thus, if there was no mention of this in the abstract, the article was not included. The type of risk assessment tool was not restricted to IPV risk assessment tools

<sup>3</sup> We also tried to conduct searches at the Springer and Wiley sites. However, their search function only allowed searches among their journals and not among specific articles.

**Table 1** Inspired by Nicholls et al. (2013), p. 86

Cluster	Search terms
Intimate partner violence	Partner violence, partner abuse, domestic violence, intimate partner violence, wife abuse, wife assault, family violence, femicide, intimate partner homicide, spouse abuse, spouse assault, spouse violence
Measurement	Test validity, statistical validity, accuracy, predict*, sensitivity, specificity
Risk assessment	Actuarial, risk assessment, structured professional judgment, dangerousness assessment, rating scale, assessment tool, instrument, Domestic violence risk appraisal guide, Danger assessment, Kingston screening instrument for domestic violence, Ontario domestic assault risk assessment, Spousal assault risk assessment guide, Brief spousal assault form for the evaluation of risk, domestic violence screening instrument, Violence risk appraisal guide, Level of service inventory, HCR-20
Risk	Relapse, repeat, re-victimization, re-abuse, recidivism

however. For example, evaluations of general risk assessment tools in samples consisting of IPV offenders were included, in line with Nicholls et al. (2013). We also, like Nicholls et al. (2013), included new tools, which means that there were no requirements regarding previous evaluations. Further, since the aim of the study was to examine practitioners' violence risk assessments, studies based on assessments conducted by other actors, e.g., researchers, were excluded. It was also important that the practitioners had conducted the assessments in the specific setting with which they were affiliated. For instance, in a new study by Gerth et al. (2017), psychologists conducted risk assessments based on police data. This study was excluded, since the raters did not work in this setting normally, but only conducted assessments on behalf of the specific study. Studies of risk assessments based only on victims' self-reports/perceptions were also excluded. Finally, the dependent/outcome variable was restricted to IPV recidivism (any definition was acceptable, e.g., police-reported IPV, self-reported re-victimization etc.). However, while the victim did not have to be the same as in the index crime in example, the crime that had resulted in the risk assessment, it had to be a current or former intimate partner. Thus, studies of violence in other family relations (labeled IPV) were excluded (the same applies to the index crime). Finally, studies in which IPV in the current and/or past situation was predicted were excluded.

## Results

Eleven studies were included in the review (Belfrage and Strand 2012; Belfrage et al. 2012; Hendricks et al. 2006; Hilton et al. 2010; Lauria et al. 2017; Rettenberger and Eher

2013; Shepard et al. 2002; Storey et al. 2014; Svalin et al. 2017b, 2018; Williams and Houghton 2004). In total, nine different tools/versions of tools<sup>4</sup> had been used in the studies (for a complete list of the tools see Tables 2 and 3). In two of the studies, two tools had been employed (Rettenberger and Eher 2013; Williams and Houghton 2004) and in the rest of the studies, one tool had been used.<sup>5</sup> The SARA or the B-SAFER had been evaluated in 6 of the 11 studies, the ODARA had been evaluated twice, and the rest of the tools were found only once. In Williams and Houghton's (2004) study, SARA assessments were used to evaluate the concurrent and discriminant validity of the DVSI tool. Thus, the focus of the study was on the latter tool, and not the SARA. Further, Shepard et al. (2002) examined a batterer categorization rather than a risk assessment tool. However, the study was included nonetheless because the categorization included risk levels (ranging from (1) little risk–(4) serious risk). Some of the tools (3) were general tools not specialized on a certain type of violence, but most of them (5) had been developed for the evaluation of IPV/domestic violence. One study had employed the psychopathy checklist—revised (PCL-R 2nd ed., Hare 2003). In the majority of the settings, IPV risk assessments had been conducted to guide the implementation of interventions. These were either interventions primarily intended to protect the victim from repeat crimes or interventions intended to affect the offender and thereby prevent further offenses. The study samples ranged between 65 and 1465 participants. In nine studies, the suspects/offenders were men; in one study, the sample was a mix of both male and female suspects (Svalin et al. 2018); and one study lacked information regarding the sex of the offenders (Hendricks et al. 2006).

## Setting and Raters

All the studies had been conducted in criminal justice settings, with the exception of one that had been

<sup>4</sup> In two studies, it was uncertain which version of the B-SAFER had been used (Storey et al. 2014; Svalin et al. 2018) and in one study it was unclear which version of the SARA had been used (Williams and Houghton 2004).

<sup>5</sup> In addition to the LSI-OR, Hilton et al. (2010) also evaluated the ODARA. However, since the ODARA assessments were conducted by researchers, the results of the ODARA analysis were not included in the present study.

**Table 2** Results

Reference	Sample	Rater/setting	Assessment tool/s	Assessment tool/s used as recommended	Outcome	Analytical strategy (main analyses)
Befrage and Strand (2012)	216 male IPV offenders (suspected) in Sweden	82 trained police officers	B-SAFER version 2 (Kropp et al. 2010)	The global risk assessment was changed: imminent spousal violence risk and very serious/fatal violence risk were assessed on a 3-grade scale: low, medium, or high	Police-reported IPV recidivism (42%). The same victim as in the assessment. Follow-up, 28–48 months	Chi-squared test
Befrage et al. (2012)	429 male IPV offenders (suspected) in Sweden	Trained police officers. The assessments were conducted as a part of the police officers normal duties, and they were blind to the outcome (prospective design)	SARA version 2 (Kropp et al. 1995)	After completing the risk assessment officer went through the documentation together with his/her supervisor. This procedure is not mandatory according to the SARA guide	Police-reported IPV recidivism (21%). The same victim as in the assessment. Follow-up, 18 months	Receiver operating characteristic (ROC) analysis, logistic regression analysis, Sobel test
Hendricks et al. (2006)	200 IPV offenders (charged) in Marathon County, USA	Master's level clinician/s Setting: Children's Service Society of Wisconsin (which is related to the Children's Hospital of Wisconsin)	LSI-R (Andrews and Bonta 1995)	Not specified	Court recorded IPV recidivism (17.5%). Any victim. Follow-up, 0–6 months (8.5%), 6–12 months (4%), and 12–18 months (3%) after completed treatment (or withdrawal from treatment). 2% recidivated before treatment had started	Logistic regression analysis, sensitivity, specificity
Hilton et al. (2010)	140 <sup>a</sup> male IPV offenders (incarcerated <sup>b</sup> ) in Ontario, Canada	Institutional staff (correctional treatment institution)	LSI-OR (Andrews et al. 1995) <sup>c</sup>	Not specified	IPV recidivism (criminal charge) (27%). Any victim. Mean follow-up 7.98 years. Mean time-at-risk 5.10 years	Receiver operating characteristic (ROC) analysis
Lauria et al. (2017)	200 male IPV offenders (suspected) in Australia	Police officers	ODARA (Hilton et al. 2004)	Not specified	IPV recidivism (police-reported or criminal charge) (18.5%). The same victim as in the assessment. Follow-up, 16–216 days	Chi-squared test, receiver operating characteristic (ROC) analysis
Rettenberger and Eher (2013)	66 male IPV <sup>d</sup> offenders (convicted) in Austria	The violence risk assessment was a part of a forensic evaluation, conducted by at least three forensic psychologists/psychiatrists.	ODARA (Hilton et al. 2004), PCL-R (Hare 2003) <sup>e</sup>	Yes (PCL-R). ODARA assessments were rated retrospectively	IPV recidivism (official criminal record and info from the criminal courts) (21.2%). Any victim. Follow-up 5.43–96.03 months (after imprisonment)	Receiver operating characteristic (ROC) analysis
Shepard et al. (2002)	798 male IPV offenders (court-ordered or voluntarily participated in a treatment program) in the USA	Probation officers	EDAPI, a working process/method: Enhanced Domestic Abuse Intervention Project. The assessment tool used in the project did not have a specific name)	Not specified, since the study describes the development of a working method <sup>f</sup>	IPV recidivism (investigated, charged, convicted, criminal justice databases). Any victim. Follow-up, 6, 12, and 18 months (year 94–97) 6 and 12 (year 98) (after the treatment started). 1994 36%, 46%, 51% 1996 31%, 41%, 46% 1997 28%, 39%, 44% 1998 20%, 33%, –	Spearman correlation
Storey et al. (2014)	249 male IPV offenders (suspected) in Sweden	Trained police officers. The assessments were conducted as a part of the police officers day-to-day	B-SAFER (Kropp et al. 2005, 2010)	After completing the risk assessment and management procedure, the officer went through the documentation together with his/her	Police-reported IPV recidivism (24%). The same victim as in the assessment. Follow-up, 11 months	Receiver operating characteristic (ROC) analysis, logistic regression analysis,

**Table 2** (continued)

Reference	Sample	Rater/setting	Assessment tool/s	Assessment tool/s used as recommended	Outcome	Analytical strategy (main analyses)
Svalin et al. (2017b)	65 male IPV offenders (suspected) in Sweden	work and they were blind to the outcome (prospective design). 8 police employees (some of them were police officers, some of them were not). The assessments were conducted as a part of the police employees' day-to-day work, and they were blind to the outcome (prospective design).	PST-VC	supervisor. This procedure is not mandatory according to the B-SAFER guide Yes <sup>g</sup>	Police-reported IPV recidivism (48%). The same victim as in the assessment. Follow-up, max 16–28 months	mediating analysis (binary_mediation for Stata 11.2) Receiver operating characteristic (ROC) analysis, logistic regression analysis
Svalin et al. (2018)	301 IPV offenders (suspected) (287 males, 11 females, 3 missing) in Sweden	3 police employees <sup>h</sup> (some of them were police officers, some of them were not)	B-SAFER (Kropp et al. 2005, 2010)	Information from victims was not included in the assessment. The global risk scales were also changed: likelihood and severity of repeat IPV were assessed between 1 = low risk/no threat to 5 = high risk/significant threat	Police-reported IPV recidivism (31%). The same victim as in the assessment. Follow-up, max 10–32 months	Receiver operating characteristic (ROC) analysis, logistic regression analysis
Williams and Houghton (2004)	1465 arrested male IPV perpetrators in Colorado A DVSI was conducted in all cases (N = 1465) and in 434 of those cases, a SARA assessment was conducted	Trained probation officers (POs). The assessments were conducted as a part of a pilot study to evaluate the DVSI tool (prospective design)	DVSI and the SARA (Kropp et al. 1994, 1995, 1999)	Yes	Total sample: IPV recidivism (violations of restraining orders and arrests for IPV) (29%). Any victim or the same victim as in the assessment is unspecified/unclear. Follow-up, 18 months. A small sample of victims (N = 125) were interviewed regarding repeat victimization after 6 months. Over 30% were repeatedly exposed to violence	Receiver operating characteristic (ROC) analysis

*B-SAFER* Brief Spousal Assault Form for the Evaluation of Risk (Kropp et al. 2005, 2010), *DVSI* Domestic Violence Screening Instrument, *EDAIIP* Enhanced Domestic Abuse Intervention Project, *LSI-OR* LSI-Ontario Revision (Andrews et al. 1995), *LSI-R* Level of Service Inventory-Revised (Andrews and Bonta 1995), *ODARA* The Ontario Domestic Assault Risk Assessment (Hilton et al. 2004), *PCL-R* psychopathy checklist—revised (2nd ed.) (Hare 2003), *PST-VC* Police Screening Tool for Violent Crimes, *SARA* the Spousal Assault Risk Assessment Guide (Kropp et al. 1994, 1995, 1999)

<sup>a</sup> The total sample was 150, but LSI-OR assessments were only available in 140 of those cases

<sup>b</sup> The offenders had not to be incarcerated because of an IPV offense. However, they were all eligible for a domestic violence program, because they had either a police record or self-reported violence towards a current or former intimate partner

<sup>c</sup> The ODARA was also used in the study, but since researchers examined the risk, the ODARA results were not included in the study

<sup>d</sup> The offenders had committed sexually motivated IPV towards their current or former intimate partners

<sup>e</sup> According to the corresponding author (Eher, R., 21–22/11–17), the ODARA assessments were conducted by the forensic psychologists/psychiatrists and the DVDRAG assessments were conducted by the researchers. Hence, the ODARA results were included in this study, but not the DVDRAG results. The victim in the follow-up could be the same victim as previously or another victim/s

<sup>f</sup> A risk tool was developed to be used as a part of this working method, which was not evaluated in the study. The risk assessment evaluation described in the present study concern the probation officers batterer categorization

<sup>g</sup> Information regarding whether the PST-VC was used as recommended or not was missing in the study and hence added for this study specifically

<sup>h</sup> Information regarding the police employees' education/training in violence risk assessment was missing in the study and hence added for this study specifically

conducted in a treatment setting.<sup>6</sup> In six studies, the assessments had been conducted by police employees. Five of these studies focused on Swedish police settings (Belfrage and Strand 2012; Belfrage et al. 2012; Storey et al. 2014; Svalin et al. 2017b, 2018) and one on an Australian police setting (Lauria et al. 2017). In three of the studies, the risk assessments had been conducted by probation officers or correction institutional staff (in Canada and the USA) (Hilton et al. 2010; Shepard et al. 2002; Williams and Houghton 2004), and in the treatment study, the IPV risk assessments had been carried out by master's level clinician/s (in the USA) (Hendricks et al. 2006). Finally, in one study, the assessments had been conducted by forensic psychologists/psychiatrists at a federal evaluation center for violent and sexual offenders in Austria (Rettenberger and Eher 2013). The offender sample in this study differed from the other offender samples, since the offenders in this case had been convicted of sexually motivated violent offenses towards (current or former) intimate partners. The suspects in the other studies had committed a wider range of IPV crimes.

Overall, the studies included a rather limited amount of information regarding the level of training/experience that the raters had in assessing violence risk or regarding their professional experience. Seven studies included a brief description and in three studies, this issue was not mentioned at all.<sup>7</sup> The descriptions included, for instance, who had been responsible for the training (e.g., one of the authors of a tool, see Storey et al. 2014), the overall content (e.g., theory and practice), and the length of the training (e.g., 2 days) (see Belfrage and Strand 2012). Overall, the amount of training appeared to be rather limited. For instance, the police officers in the study by Lauria et al. (2017) had not been given any training in the use of the ODARA. Further, the probation officers in the Shepard et al. (2002) study had recommended sentences based on their offender risk categorization, and in a survey presented in the study, they expressed their satisfaction regarding the training they had received in sentencing recommendations. However, the interventions had nonetheless been implemented inconsistently.

Previous studies have highlighted the fact that violence risk assessment tools are not always used in accordance with the guidelines for a given tool (e.g., HMIC, 2014). Overall, the reviewed studies provided little information regarding the administration of the assessments. Five studies lacked information regarding whether or not the

tools had been utilized as recommended (Hendricks et al. 2006; Hilton et al. 2010; Lauria et al. 2017; Shepard et al. 2002<sup>8</sup>; Svalin et al. 2017b<sup>9</sup>). One study stated that the tools' guidelines had been followed (Williams and Houghton 2004). In yet another study, two tools had been used, one of them according to the recommendations (PCL-R) while the other (ODARA) had been used to assess cases retrospectively (Rettenberger and Eher 2013). In two studies, the global risk rating had been changed (Belfrage and Strand 2012; Svalin et al. 2018), and in addition, in the latter of these two studies, the information base had been less extensive than recommended (no information had been collected from victims). Finally, in two studies, the police officers who conducted the assessments had carried out their assessments together with their supervisor. Since this is not a mandatory procedure in the SARA or the B-SAFER, it might be viewed as an additional quality check (Belfrage et al. 2012; Storey et al. 2014).

### Predictive Validity

The predictive validity of the IPV risk assessments in relation to IPV recidivism was measured in a number of ways. However, the most common main analysis employed was the ROC analysis (conducted in eight studies). The AUC values for global risk assessments/numerical total scores, with the outcome IPV recidivism varied between 0.49 and 0.72 in the studies. The highest AUC was presented by Lauria et al. (2017) and related to the ODARA total scores with the outcome non-physical assault against the same victim as in the risk assessment. In total, 22 AUC values (using global risk assessments/numerical total score as test variables) were presented, with the results being evenly distributed between the highest and lowest values. Ten AUC values were lower than .60 and twelve higher, hence the median AUC is around .61. Overall, the predictive validity ranged from low (not predictive at all) to moderate, the median AUC effect size is small.

Predictive validity was also measured in other ways than by means of ROC analysis. For instance, Belfrage and Strand (2012) and Shepard et al. (2002) compared the recidivism rates for different risk groups/categories. The first study compared the recidivism rates between the low-, medium-, and high-risk groups for imminent/acute risk of IPV and severe/fatal IPV risk. No statistical differences were found for any of the categories (Belfrage and Strand 2012). The second study compared the rate of recidivism in four different battering

<sup>6</sup> The offenders had been referred to the Children's Service Society of Wisconsin, which is associated with the Children's Hospital of Wisconsin.

<sup>7</sup> One of those articles which lacked information regarding training/experiences of violence risk assessment was Svalin et al. (2018) study. Such information was therefore added for the purposes of this review.

<sup>8</sup> The reason to the lack of information regarding the administration of a tool was probably because a batterer categorization rather than a tool was evaluated.

<sup>9</sup> Information regarding whether the PST-VC was used as recommended or not was missing in the study, and hence added for this study specifically (in Table 2).



**Table 3** Results

Reference	Education/training/experiences	Predictive validity of global risk measures <sup>a</sup>	The role of the interventions	Additional information
Belfrage and Strand (2012)	2 days training (theory and practice) by one of the B-SAFER authors	No statistical differences between the recidivism rates in the different risk groups (low/medium/high) of severe/fatal or imminent/acute violence	Kendall's $\tau_b$ : summary risk assessment-imminent violence and protective actions, $\tau_b = 0.30$ ( $p = 0.000$ ) and severe/fatal violence and protective actions, $\tau_b = 0.25$ ( $p = 0.000$ ). Main interpretations of the weak predictive validity of the global risk assessment on IPV recidivism—(1) no implemented interventions in most of the cases and (2) effective interventions implemented in high-risk cases (significant differences ( $\chi^2$ ) in IPV recidivism in the different risk groups (fatal violence), among those who recidivated)	
Belfrage et al. (2012)	Two of the B-SAFER authors trained the police officers	ROC-analyses: the global risk assessment, outcome repeat IPV, $AUC = 0.57$ . ( $SE = 0.033$ ). Numerical total score, outcome repeat IPV, $AUC = 0.63$ . ( $SE = 0.032$ )	Logistic regression analyses: numerical total score, outcome repeat IPV, $OR = 1.17$ ( $p = .001$ ). Number of recommended management strategies, outcome repeat IPV, $OR = 1.38$ ( $p = .032$ ). Numerical total score + number of recommended management strategies, outcome repeat IPV, $OR = .98$ ( $p = .030$ ). The rate of IPV recidivism in low- and high-risk cases differed depending on the number of recommended management strategies, in concordance with the RNR principles. Sobel test = 2.99 ( $SE = 0.009$ , $p = .003$ ) The effectiveness of treatment programs on IPV recidivism was analyzed. However, the effect of the variables, treatment, and LSI-R scores on the outcome and IPV recidivism was not disentangled	Final conclusion: the SARA tool is effective in preventing violence
Hendricks et al. (2006)	Not specified	Logistic regression: overall classification accuracy. LSI-R risk and need scales 64%. LSI-R total score 66%. Total score, cutoff 11.5, specificity 67% and sensitivity 60%		
Hilton et al. (2010)	Not specified	ROC analysis: LSI-OR total score, with outcome repeat IPV, $AUC = 0.502$ . ( $SE = 0.05$ , 95% $CI = 0.397–0.607$ )	Negative significant relationship between the number of treatment modules that the offenders began and IPV recidivism ( $r = -.16$ , $p < .05$ ). No significant relationship between the number of completed treatment modules and IPV recidivism ( $r = .04$ , ns). No information regarding LSI-OR scores in relation to treatment (e.g., completed treatment)	The restricted range in the sample affected the predictive accuracy
Lauria et al. (2017)	The police officers did not have any training in the use of the ODARA	ROC analyses: the ODARA total scores, outcome repeat IPV (physical assault), $AUC = 0.68$ (95% $CI = 0.55–0.81$ ), and outcome repeat IPV (non-physical assault), $AUC = 0.72$ (95% $CI = 0.61–0.82$ )	Not specified	
Rettenberger and Eher (2013)	Theoretically/practically experienced forensic psychologists or psychiatrists. They had all	ROC analysis: the ODARA risk categories, outcome repeat IPV, $AUC = 0.71$ ( $p < .05$ ; 95% $CI = 0.58–0.84$ ).	Treatment had been suggested to all participants, during or after the imprisonment. However, the	No general conclusions regarding IPV perpetrators could be made, since the sample under study was a high-risk sample. The

**Table 3** (continued)

Reference	Education/training/experiences	Predictive validity of global risk measures <sup>a</sup>	The role of the interventions	Additional information
Shepard et al. (2002)	participated (at the minimum) at one official PCL-R training workshop  The probation officers were trained in the use of the tool. There is no information regarding the content of the training	ROC analysis: the PCL-R total score, outcome repeat IPV, AUC = 0.56. (ns; 95% CI = 0.39–0.74)  IPV recidivism at 18 months follow-up: cat 1 36% (10/28), cat 2 45% (23/51), cat 3 64% (28/44), cat 4 50% (2/4)	authors had no information regarding who had participated in treatment or not  Completed treatment was related to lower rates of IPV recidivism. There was no information regarding how the batterer categories + treatment were related to recidivism	sample size was small ( $N = 66$ ), and the number of IPV recidivism cases was 14  Further evaluations of the validity and reliability of the tool is required
Storey et al. (2014)	One of the B-SAFER authors trained the police officers	ROC analysis: the global risk assessment, outcome repeat IPV, AUC = 0.65. ( $SE = 0.04$ ). Numerical total score, outcome repeat IPV, AUC = 0.70 ( $SE = 0.04$ )	Logistic regression analysis: numerical total score, outcome repeat IPV, OR = 3.78 ( $p < .001$ ). Number of recommended management strategies, outcome repeat IPV, OR = 1.73 ( $p = .207$ ). Numerical total score + recommended management strategies, outcome repeat IPV, OR = 0.47 ( $p = .245$ ). Logistic regression analysis: global risk assessment, outcome repeat IPV, OR = 7.59 ( $p = .001$ ). Number of recommended management strategies, outcome repeat IPV, OR = 1.83 ( $p = .089$ ). Numerical total score + recommended management strategies, outcome repeat IPV, OR = 0.18 ( $p = .039$ )	The small sample size possibly affected the results of the mediation analysis
Svalin et al. (2017b)	The assessors' experiences of assessing violence risk were overall rather low. Their professional knowledge varied	ROC-analyses: the global risk assessment, with outcome repeat IPV, AUC = 0.66 ( $SE = 0.068$ ) and outcome violent recidivism, AUC = 0.67 ( $SE = 0.078$ )	The rate of IPV recidivism in low- and high-risk cases differed, depending on the number of recommended management strategies, in line with the RNR principles.  No statistically significant mediation effect of the recommended management strategies on IPV recidivism	Due to the small sample size, the results are preliminary

**Table 3** (continued)

Reference	Education/training/experiences	Predictive validity of global risk measures <sup>a</sup>	The role of the interventions	Additional information
Svalin et al. (2018)	The assessors' experiences of assessing violence risk were overall rather low. Their professional knowledge varied but was more extensive <sup>b</sup>	ROC analyses: likelihood, outcome repeat IPV, AUC = 0.57, and outcome violent recidivism, AUC = 0.54. Severity, outcome repeat IPV, AUC = 0.55, and outcome violent recidivism, AUC = 0.50	Logistic regression analyses (stepwise): severity, outcome repeat IPV (step 1), OR = 1.563 (ns). The implemented interventions, outcome repeat IPV (step 2), OR = 1.481 (ns). Likelihood, outcome repeat IPV (step 1), OR = 1.803 ( $p < .05$ ). The implemented interventions, outcome repeat IPV (step 2), OR = 1.487 (ns). Severity, outcome repeat violence (step 1), OR = 1.036 (ns). The implemented interventions, outcome repeat violence (step 2), OR = 0.800 (ns). Likelihood, outcome repeat violence (step 1), OR = 1.403 (ns). The implemented interventions, outcome repeat IPV (step 2), OR = 0.768 (ns). The rate of IPV recidivism in low- and high-risk cases did not differ depending on the number of implemented management strategies, in the way suggested by the RNR principles	
Williams and Houghton (2004)	The Colorado Department of Probation Services (DPS) project manager trained the POs in the use of the DVSI and the SARA	ROC analysis: DVSI numerical total score, outcome repeat IPV, AUC = 0.61 ( $p = .000$ ). The small sample (based on victim interviews): the predictive validity of the DVSI numerical total score, outcome repeat IPV measured as (1) controlling behaviors = AUC = 0.58, ( $p = .14$ ), (2) less threatening behaviors, AUC = 0.56, ( $p = .26$ ), (3) less serious physically violent behaviors, AUC = 0.49, ( $p = .92$ ), (4) severe threatening behaviors, AUC = 0.68, ( $p = .001$ ), (5) very severe physical violence, AUC = 0.65, ( $p = .041$ ). SARA numerical total score and a modified version of the SARA, outcome repeat IPV, AUC = 0.65, ( $p < .000$ ) (for both measures)		

<sup>a</sup> Other follow-up measures than repeat IPV (e.g., recidivism in general) was not presented in this study. In addition to summary risk assessments, global risk measures include, e.g., the total score of a tool

<sup>b</sup> This information lacked in the study and was added for this study specifically

categories ((1) no battering history, (2) low-level/not escalating, (3) clear pattern/likely to escalate, (4) high risk of serious harm) (Shepard et al. 2002). The pattern was almost the same on all follow-up occasions (6, 12, and 18 months): the rate of recidivism increased by batterer category, with one exception. The rate of recidivism was higher in category 3 than in category 4, at both the 12- and 18-month follow-ups. However, category 4 included a total of only four offenders. The categories were significantly correlated with recidivism at the 6-, 12-, and 18-month follow-ups, but the relationship was weak ( $r = .20-.21$ ,  $p \leq .05-.01$ ). Further, Hendricks et al. (2006) evaluated the classification accuracy for LSI-R risk and need scales and LSI-R total scores on repeat IPV. The classification was correct in 64% and 66% of the cases, respectively, and the sensitivity and specificity (best balance) of the total scores (cut-off 11.5) were 67% and 60%, respectively. In sum, the predictive validity was rather low in all three studies.

Svalin et al. (2018) conducted ROC analysis using test variables based on predictive values from stepwise/enter regression models with risk and victim vulnerability factors as independent variables and the global risk assessment as the outcome variable. The AUCs for predictions of repeat IPV varied between 0.51 and 0.57 for predictions of repeat IPV and repeat violence (identical range for both outcomes). Lauria et al. (2017) examined the predictive validity (AUC) of each risk factor in the ODARA, with physical and non-physical assault as the outcome variables. The results ranged between 0.46–0.63 and 0.52–0.66 respectively. The ODARA items were also examined in Rettenberger and Eher (2013), using correlations between each item and IPV recidivism. Five of the 13 ODARA items correlated significantly (range 0.25–0.44,  $p < .05-.001$ ).

### The Significance of Interventions on the Predictive Accuracy

All but three of the reviewed studies (Lauria et al. 2017; Williams and Houghton 2004; Rettenberger and Eher 2013<sup>10</sup>) analyzed treatment or other interventions in one way or the other. In one of these, both risk assessment and treatment were related to the outcome,<sup>11</sup> but the relationship between the risk assessment and the treatment was not clear (Shepard et al. 2002). Further, in Hilton et al. (2010), the predictive accuracy of the LSI-OR scores on IPV recidivism was low (AUC = 0.50) and there was a negative significant

relationship between the number of initiated offender treatment modules and IPV recidivism ( $r = -.16$ ,  $p < .05$ ). However, the number of *completed* treatment modules and IPV recidivism were not correlated, and there were no information regarding LSI-OR scores for those offenders who completed the treatment.

Hendricks et al. (2006) examined the accuracy of the LSI-R and the effect of two IPV offender treatment programs on IPV recidivism. They run into problems interpreting the results. For example, offenders who completed one of the treatment programs (SAFE) had a lower likelihood of IPV recidivism compared to those who did not complete the program. However, since the offenders who completed the specific program had significantly lower LSI-R scores than those who did not, it was not clear whether the effect was due to the treatment or their lower risk.

The five studies regarding IPV risk assessment in Swedish police settings evaluated whether the recommended and/or implemented interventions mediated the relationship between the risk assessment and IPV recidivism. In one of these studies, the implemented protective actions correlated significantly with the global risk assessment, but not with IPV recidivism (Belfrage and Strand 2012). Focusing on IPV recidivism cases only in relation to implemented protective actions, the authors found a significant difference between the rates of recidivism in the low-, medium-, and high-risk groups (severe/fatal violence). The higher the assessed risk, the lower the rate of recidivism is. These results were suggested to be due to the effectiveness of protective actions implemented in the most severe cases. Belfrage et al. (2012) and Storey et al. (2014) found risk assessment to predict the number of recommended protective actions and IPV recidivism, and that risk assessment<sup>12</sup> together with the number of recommended protective actions predicted IPV recidivism. Further, in Belfrage et al. (2012), the number of recommended protective actions also predicted IPV recidivism and mediated the relationship between risk assessment and IPV recidivism. In both studies, the rate of repeat IPV was lower in high-risk cases with a high level of interventions, compared to the recidivism rate in high-risk cases with a low level of interventions (Belfrage et al. 2012; Storey et al. 2014).

In line with both Belfrage et al. (2012) and Storey et al. (2014), the Svalin et al. (2017b) study also examined the effect of recommended protective actions on predictive accuracy. The results showed that the risk assessment (low/high risk) and the recommended protective actions (low/high level of protective actions) in interaction did not significantly predict repeat IPV, with one exception. In high-risk cases with a high level of recommended interventions, the risk of repeat IPV was significantly increased compared to the reference

<sup>10</sup> In this study, treatment had been suggested in connection with the prison sentence. However, the authors had no information regarding who had then participated in treatment (Rettenberger and Eher 2013).

<sup>11</sup> The direction of the correlation differed depending on how the treatment variable was measured. For instance, completed treatment was associated with lower rates of IPV recidivism (Shepard et al. 2002). At the same time, court-mandated treatment was associated with higher rates of IPV recidivism in the same study.

<sup>12</sup> Measured as the SARA total score in Belfrage et al. (2012) and the global risk assessment in Storey et al. (2014).

category (low-risk cases with a low level of recommended protective actions). However, due to the small sample, the findings were considered preliminary. The question regarding the significance of protective interventions was also evaluated in a more recent study (Svalin et al. 2018), but this time with a larger sample and with a follow-up of the interventions.<sup>13</sup> The results showed that the risk of repeat IPV was significantly increased in high-risk cases (likelihood) with or without any implemented protective actions, compared to the reference category (that is, cases assessed as low risk in which no interventions were implemented). The low-risk cases with at least one implemented protective action did not significantly predict the outcome.

In sum, three studies included no information regarding treatment and other interventions (Lauria et al. 2017; Williams and Houghton 2004; Rettenberger and Eher 2013). In three other studies, different kinds of interventions were analyzed in one way or another, but it was difficult to draw conclusions regarding the role of the interventions in relation to the risk assessment and the outcome (Shepard et al. 2002; Hendricks et al. 2006; Hilton et al. 2010). In the remaining five studies, interventions were analyzed/discussed as possible mediating factors, with somewhat varying results. One article showed that the recommended protective actions mediated the relationship between the risk assessments and repeat IPV and concluded that the risk assessment had prevented repeat victimization (Belfrage et al. 2012). Storey et al. (2014) did not find a similar effect but an interaction between risk assessment and recommended interventions in relation to IPV recidivism. The implemented interventions in Belfrage and Strand (2012) did not correlate with repeat IPV recidivism. However, the rate of IPV recidivism was lower in the high-risk group compared with the low- and medium-risk groups (among recidivism cases only). On the other hand, the final two studies (Svalin et al. 2017b, 2018), did not find any support for the risk assessments with subsequent interventions to be violence preventive. Finally, none of these studies said anything about the importance of specific protective actions or about whether some actions are effective in some cases but not in others.

## Discussion

The main aim of the present study has been to examine the predictive validity of IPV risk assessments conducted by practitioners in different settings. In a majority of the studies, the predictive validity for the global risk assessments/numerical total scores was measured using the AUC of ROC with IPV

recidivism as the outcome. The AUC values ranged between 0.49 and 0.72; only three AUCs were 0.70 or higher. Thus overall, the predictive accuracy was small. The results were similar in the three studies that measured predictive validity in other ways than by means of ROC: There were no differences between the rates of recidivism in the different risk groups (Belfrage and Strand 2012). The sensitivity and specificity measures that represented the best balance were relatively low (Hendricks et al. 2006), and the correlation between the risk categories and IPV recidivism was non-significant (Shepard et al. 2002).

A wide range of AUC values has also been noted in previous review studies. For instance, Nicholls et al. (2013) presented AUC values ranging between 0.48 and 0.92, Graham et al. (2019) between 0.51 and 0.86, and Helmus and Bourgon (2011) between 0.59 and 0.87. Messing and Thaller (2013) presented the average AUCs of different tools, which had a smaller range (0.54–0.67). As was mentioned previously, predictive validity is influenced by many factors, which makes it difficult to compare results from different studies. However, the overall accuracy was slightly higher in both Nicholls et al. (2013), Graham et al. (2019), and Helmus and Bourgon (2011) compared to the AUCs reported in the present study (0.49–0.72). The studies with the highest AUC values in those reviews included actuarial and SPJ tools assessed by risk evaluators (in some cases practitioners, while no information was presented in other studies) or based on self-reports by victims and offenders. Thus, more research is needed to develop the knowledge on predictive validity and on what is required to produce accurate assessments in different settings. Some researchers have highlighted the need to shift the focus from those who are the subject of the risk assessment to those who examine the risk, since they argue we have reached a “predictive glass roof” (Sturup et al. 2013).

Nine of the 11 studies analyzed interventions in one way or another. In four of these, it was difficult to interpret the role of the interventions in relation to the risk assessment and/or the outcome. In the study by Shepard et al. (2002), for example, both the risk assessment and treatment were related to IPV recidivism, although it was not clear how the risk assessment and treatment related to one another. The results of the five studies that examined the role of the interventions as possible mediators were inconclusive. In some of the studies, the protective actions were shown to, or suggested to, have an influence on IPV recidivism (Belfrage and Strand 2012; Belfrage et al. 2012; Storey et al. 2014), while in other studies, similar results were not found (Svalin et al. 2017b, 2018). The rates of IPV recidivism varied in those studies between 21 and 48%, which is similar to previous studies. For instance, in Petersson and Strand (2017), the prevalence of IPV recidivism was approximately 13% and 27% for family-only and antisocial offenders, respectively. Rahman (2018) reported repeat IPV for 43% of the offenders, and in Lin et al. (2009), 48% relapsed

<sup>13</sup> The interventions were measures implemented by the police or with the knowledge of the police. Thus, interventions implemented by other actors or victims themselves without the knowledge of the police were not included in the analysis.

into overall violence within 3 months. However, overall conclusions of the effectiveness of violence risk assessment and management cannot be drawn based on the rates of IPV recidivism, since there are significant methodological differences in the studies (e.g., follow-up time, weather protective actions were implemented or not etc.). More research is needed regarding the predictive validity of IPV assessments in different settings, and specifically regarding the effectiveness of crime-preventive and victim-protective actions, and whether different measures are suitable for different types of IPV offenders.

The low number of studies included in the review is itself an important result, since it indicates that there is a knowledge gap regarding the accuracy of practitioners' IPV risk assessments in different settings. There are a number of possible reasons for this finding. First, violence between intimate partners is not always separated from violence between other family members in studies of the predictive validity of IPV assessments (e.g., violence between parents and children, siblings, etc., see for example Dayan et al. 2013). Thus, by choosing to study violence between intimate partners only, studies using the broader IPV definition were excluded from the review. Looking specifically at the definition of repeat IPV used in this review, it is actually rather inclusive, even though it only refers to intimate partners. All kinds of repeat IPV conducted towards former or current intimate partners (both the same victim as in the index crime and new victims) were included, as were studies based on information from any kind of sources (e.g., self-report, police registers, etc.). As has previously been noted, however, different definitions of key terms are problematic when comparing the results from different studies and must be kept in mind when interpreting the review's findings.

Other possible reasons for the low number of studies are that practitioners conduct IPV risk assessments (1) without the use of assessment tools, (2) by means of general violence risk assessment tools together with other types of violence, or (3) that tools are used, but that their predictive accuracy has not been evaluated. These possible reasons will be discussed one by one below.

1. The unstructured clinical approach is the most commonly used approach historically (Hart, 2008), that is, assessments conducted without the use of a tool. There are indications that this is still a common way of assessing violence risk. For instance, Cattaneo and Chapman (2011) found that practitioners working with victims of IPV in different settings used their own or their colleagues' professional experiences and tacit knowledge to assess violence risk rather than a risk assessment tool. In another study, police employees were found to base their global risk assessments on information other than the factors included in the tools employed (Svalin et al. 2017a). A suggestion was that they were instead using their tacit knowledge.

2. General violence risk assessments, which include all kinds of violence, are conducted in some settings and thus not IPV assessments specifically. According to Hilton et al. (2010), at least one third of incarcerated male offenders have committed intimate partner violence and as a result of the low number of studies of IPV assessments conducted in correctional settings found in this review, one could speculate that other tools have been used in these cases. Further, Rettenberger and Eher (2013) note that different violence risk assessment tools are used in the Austrian prison system, although no IPV risk assessment tools had been used prior to the initiation of their own study, which evaluated the ODARA and DVRAG.
3. The absence of evaluations of IPV assessments may be due to a number of different reasons. Perhaps it is simply a matter of prioritization or of difficulties related to the evaluation procedure, such as difficulties obtaining access to follow-up data, which are required for this kind of evaluation. A number of studies were excluded in the sorting procedure because the assessments had been conducted by researchers and not practitioners (e.g., Buchanan 2009). It is reasonable to assume that some settings rely on the results from such evaluations. However, since conditions vary between different settings in general, and thus between different raters, for example with regard to the level of risk assessment training, access to information, the amount of time available to produce assessments, etc., it is problematic to apply the results of evaluations conducted in other settings. Finally, this study confirms that IPV risk assessment tools are sometimes used in other ways than recommended. Since this may affect the accuracy of assessments, one cannot expect the results from different settings to be applicable under such circumstances.

The overall conclusion of this review is that the research regarding the accuracy of practitioners' IPV risk assessments is limited. Only 11 studies met the inclusion criteria and all but one were conducted in criminal justice settings. Possible reasons for the low number of studies have been discussed, for example, that IPV risk assessments in practical settings are still being conducted without the use of risk tools. There was little information regarding the risk raters' training in assessing IPV risk, but based on the information that was available, this seemed to be limited for many of the raters. Information on whether or not the risk tools were used as recommended was also limited to a few studies; in three of the studies, actual changes had been introduced into the tools or the assessments had been conducted retrospectively. The level of predictive validity was rather low overall, and the role played by protective actions in relation to the risk assessment and the outcome measure was not clear. IPV risk assessment has the potential to play an important role in preventing repeat

violence and protecting victims. The studies included in this review indicate that there is more work to be done in order to achieve this. However, in order to develop a more complete picture, the IPV risk assessments conducted in different settings and their related risk management strategies must be evaluated. A possible consequence of not knowing which tools are suitable in different practical settings is that practitioners use tools that do not fit the context and, the worst case scenario, that decisions based on such assessments aggravate the situation of victims of IPV instead of preventing repeat IPV victimization.

Finally, information regarding which inclusion criteria were not met in the studies that were excluded in the sorting procedure was not collected, and this constitutes a limitation of the current study. Although the exact numbers are unknown, many studies were excluded because the predictive validity of a risk tool was not evaluated or because the type of violence evaluated was not IPV. Nicholls et al. (2013) reviewed 39 studies which included all English-language studies regarding the predictive validity of IPV risk assessments conducted in western countries between 1990 and 2011. Thus, not much has been written regarding the predictive validity of IPV risk assessments, and there is even less research available when the sample is limited to studies in which the risk assessments have been conducted by practitioners.

**Funding Information** Open access funding provided by Malmö University. This work was supported by the Department of Criminology, Malmö University.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical Approval** The article has not undergone an ethical review since it is a review study based on literature.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Andrés-Pueyo A, López S, Álvarez E (2008) Assessment of the risk of intimate partner violence and the SARA. *Papeles del Psicólogo* 29(1):107–122
- Andrews DA, Bonta JL (1995) *The Level of Service Inventory-Revised (LSI-R) manual*. Multi-Health Systems, North Tonawanda, New York
- Andrews DA, Bonta J, Wormith JS (1995) *The LSI-OR: Interview and scoring guide*. Ontario Ministry of Solicitor General and Correctional Services, Toronto, Canada
- Belfrage H (2008) Police-based structured spousal violence risk assessment: the process of developing a police version of the SARA. In: Baldry AC, Winkel FW (eds) *Intimate partner violence prevention and intervention. The risk assessment and management approach*. Nova Science Publishers, New York, pp 33–44
- Belfrage H, Strand S (2012) Measuring the outcome of structured spousal violence risk assessments using the B-SAFER: risk in relation to recidivism and intervention. *Behav Sci Law* 30(4):420–430. <https://doi.org/10.1002/bsl.2019>
- Belfrage H, Strand S, Storey JE, Gibas AL, Kropp PR, Hart SD (2012) Assessment and management of risk for intimate partner violence using the spousal assault risk assessment guide. *Law Hum Behav* 36(1):60–67. <https://doi.org/10.1007/s10979-011-9278-0>
- Buchanan K (2009) *Risk assessment and spousal violence: predictive validity and cultural applicability*. (Doctoral thesis). Regina: University of Regina
- Cattaneo LB, Chapman AR (2011) Risk assessment with victims of intimate partner violence investigating the gap between research and practice. *Violence Against Women* 17(10):1286–1298. <https://doi.org/10.1177/1077801211424477>
- Danis FS (2003) Social work response to domestic violence: encouraging news from a new look. *Affilia* 18(2):177–191. <https://doi.org/10.1177/0886109903251407>
- Dayan K, Fox S, Morag M (2013) Validation of spouse violence risk assessment inventory. *J Fam Violence* 28(8):811–821. <https://doi.org/10.1007/s10896-013-9547-4>
- Douglas KS, Kropp RP (2002) A prevention-based paradigm for violence risk assessment. *Clinical and research applications. Crim Justice Behav* 29(5):617–658. <https://doi.org/10.1177/009385402236735>
- García-Moreno C, Pallitto C, Devries K, Stockl H, Watts C, Abrahams N, Petzold M (2013) *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. World Health Organization, Geneva, Switzerland
- García-Moreno C et al (2015) Addressing violence against women: a call to action. *Lancet* 385(9978):1685–1695. [https://doi.org/10.1016/S0140-6736\(14\)61830-4](https://doi.org/10.1016/S0140-6736(14)61830-4)
- Gerth J, Rossegger A, Bauch E, Endrass J (2017) Assessing the discrimination and calibration of the Ontario domestic assault risk assessment in Switzerland. *Partn Abus* 7:168–189
- Gibas A, Kropp PR, Hart SD, Stewart L (2008) Validity of the SARA in a Canadian sample of incarcerated males. Paper presented at the 8th Annual Conference of the International Association of Forensic Mental Health Services, Vienna, Austria
- Graham LM, Sahay KM, Rizo CF, Messing JT, Macy RJ (2019) The validity and reliability of available intimate partner homicide and reassault risk assessment tools: a systematic review. *Trauma Violence Abuse*:1–23. <https://doi.org/10.1177/1524838018821952>
- Hare RD (2003) *Manual for the psychopathy checklist-revised, 2nd edn*. Multi Health Systems, Toronto
- Hart SD (1998) The role of psychopathy in assessing risk for violence: conceptual and methodological issues. *Leg Criminol Psychol* 3(1):121–137. <https://doi.org/10.1111/j.2044-8333.1998.tb00354.x>
- Hart SD, Michie C, Cooke DJ (2007) Precision of actuarial risk assessment instruments. Evaluating the ‘margins of error’ of group v. individual. *Br J Psychiatry* 190(49):60–65. <https://doi.org/10.1192/bjp.190.5s60>
- Helmus L, Bourgon G (2011) Taking stock of 15 years of research on the spousal assault risk assessment guide (SARA): a critical review. *Int J Forensic Ment Health* 10(1):64–75. <https://doi.org/10.1080/14999013.2010.551709>

- Hendricks B, Werner T, Shipway L, Turinetti GJ (2006) Recidivism among spousal abusers. Predictions and program evaluation. *J Interpers Violence* 21(6):703–716. <https://doi.org/10.1177/0886260506287310>
- Her Majesty's Inspectorate of Constabulary [HMIC] (2014) Everyone's business: improving the police response to domestic abuse. Retrieved from [www.hmic.gov.uk](http://www.hmic.gov.uk)
- Hilton NZ, Harris GT, Rice ME, Lang C, Cormier CA, Lines KJ (2004) A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment* 16, 267–275. <https://doi.org/10.1037/1040-3590.16.3.267>
- Hilton NZ, Harris GT, Popham S, Lang C (2010) Risk assessment among incarcerated male domestic violence offenders. *Crim Justice Behav* 37(8):815–832. <https://doi.org/10.1177/0093854810368937>
- Kozioł-McLain J, Garrett N, Fanslow J, Hassall I, Dobbs T, Henare-Toka TA, Lovell V (2010) A randomized controlled trial of a brief emergency department intimate partner violence screening intervention. *Ann Emerg Med* 56(4):413–423. <https://doi.org/10.1016/j.annemergmed.2010.05.001>
- Kropp PR, Hart SD, Webster CD, Eaves D (1994) Manual for the spousal assault risk assessment guide. British Columbia Institute on Family Violence, Vancouver, Canada
- Kropp PR, Hart SD, Webster CD, Eaves D (1995) Manual for the spousal assault risk assessment guide (2nd ed). British Columbia Institute on Family Violence, Vancouver, Canada
- Kropp PR, Hart SD, Webster CD, Eaves D (1999) Spousal assault risk assessment guide (SARA). Multi-Health Systems, Inc., Toronto, Canada
- Kropp PR (2004) Some questions regarding spousal assault risk assessment. *Violence Against Women* 10(6):676–697. <https://doi.org/10.1177/1077801204265019>
- Kropp PR, Hart SD, Belfrage H (2005) Brief spousal assault form for the evaluation of risk (B-SAFER): User manual. Proactive Resolutions Inc., Australia, Canada, United States
- Kropp PR, Hart SD (2015) SARA-V3: user guide for the third edition of the spousal assault risk assessment guide. Proactive Resolutions Inc., Vancouver
- Kropp PR, Hart SD, Belfrage H (2010) Brief spousal assault form for the evaluation of risk (B-SAFER), version 2: user manual. Proactive Resolutions Inc., Vancouver
- Lauria I, McEwan TE, Luebbers S, Simmons M, Ogloff JRP (2017) Evaluating the Ontario domestic assault risk assessment in an Australian frontline police setting. *Crim Justice Behav* 44(12):1545–1558. <https://doi.org/10.1177/0093854817738280>
- Lin S-C, Su CY, Chou FHC, Chen SP, Huang JJ, Wu GTE, Chen WJ, Chao SS, Chen CC (2009) Domestic violence recidivism in high-risk Taiwanese offenders after the completion of violence treatment programs. *J Forensic Psychiatry Psychol* 20(3):458–472
- Mele M (2009) The time course of repeat intimate partner violence. *J Fam Violence* 24(8):619–624. <https://doi.org/10.1007/s10896-009-9260-5>
- Messing JT, 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>
- Nicholls TL, Pritchard MM, Reeves KA, Hilterman E (2013) Risk assessment in intimate partner violence: a systematic review of contemporary approaches. *Partn Abus* 4(1):76–168. <https://doi.org/10.1891/1946-6560.4.1.76>
- Petersson J, Strand S (2017) Recidivism in intimate partner violence among antisocial and family-only perpetrators. *Crim Justice Behav* 44(11):1477–1495
- Rahman S (2018) Assessing the risk of repeat intimate partner assault. Crime and Justice Bulletin. NSW Bureau of Crime Statistics and Research, Sydney
- Rettenberger M, Eher R (2013) Actuarial risk assessment in sexually motivated intimate-partner violence. *Law Hum Behav* 37(2):75–86. <https://doi.org/10.1037/b0000001>
- Richards L (2009) DASH: domestic abuse, stalking and harassment and honour-based violence risk identification and assessment and management model, DASH risk checklist. Retrieved from <https://www.dashriskchecklist.co.uk>
- Shepard M, Falk DR, Elliott BA (2002) Enhancing coordinated community responses to reduce recidivism in cases of domestic violence. *J Interpers Violence* 17(5):551–569
- Skeem JL, Monahan J (2011) Current directions in violence risk assessment. *Curr Dir Psychol Sci* 20(1):38–42. <https://doi.org/10.1177/0963721410397271>
- Storey JE, Kropp PR, Hart SD, Belfrage H, Strand S (2014) Assessment and management of risk for intimate partner violence by police officers using the brief spousal assault form for the evaluation of risk. *Crim Justice Behav* 41(2):256–271. <https://doi.org/10.1177/0093854813503960>
- Sturup J, Forsman M, Haggård U, Karlberg D, Johansson P (2013) Riskbedömning i kriminalvård och rättspsykiatri. Sammanfattningsrapport. (Projektnummer 2013:269) [Violence risk assessment in prison service and forensic psychiatry. A summary report]. Norrköping: Kriminalvården
- Svalin K, Mellgren C, Torstensson Levander M, Levander S (2017a) The inter-rater reliability of violence risk assessment tools used by police employees in Swedish police settings. *Nordisk Politiforskning* 4(1):9–28. <https://doi.org/10.18261/ISSN.1894-8693-2017-01-03>
- Svalin K, Mellgren C, Torstensson Levander M, Levander S (2017b) Assessing and managing risk for intimate partner violence: police employees' use of the police screening tool for violent crimes in Scania. *International Journal of Law and Psychiatry* 18(1):84–92. <https://doi.org/10.1080/14043858.2016.1260333>
- Svalin K, Mellgren C, Torstensson Levander M, Levander S (2018) Police employees' violence risk assessments: the predictive validity of the B-SAFER and the significance of protective actions. *Int J Law Psychiatry* 56:71–79. <https://doi.org/10.1016/j.ijlp.2017.09.001>
- Williams KR, Houghton AB (2004) Assessing the risk of domestic violence reoffending: a validation study. *Law Hum Behav* 28(4):437–455
- Wong T, Hisashima MA (2008) Domestic violence exploratory study on the DVSI and SARA, state of Hawaii, 2003–2007. Hawaii: Interagency Council on Intermediate Sanctions. Retrieved from <http://icis.hawaii.gov/wp-content/uploads/2013/07/SARA-DVSI-Exploratory-Study-Oct-2008.pdf>
- World Health Organization [WHO] (2012) Understanding and addressing violence against women. Intimate partner violence. World Health Organization, Geneva
- World Health Organization/London School of Hygiene and Tropical Medicine [WHO/LSHTM] (2010) Preventing intimate partner and sexual violence against women: taking action and generating evidence. World Health Organization, Geneva

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.