UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools

van der Put, C.E.; Gubbels, J.; Assink, M.

DOI
10.1016/j.avb.2019.03.008

Publication date
2019

Document Version
Final published version

Published in
Aggression and Violent Behavior

Link to publication

Citation for published version (APA):
van der Put, C. E., Gubbels, J., & Assink, M. (2019). Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools. *Aggression and Violent Behavior*, *47*, 100-116. https://doi.org/10.1016/j.avb.2019.03.008

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:06 Dec 2021

Aggression and Violent Behavior 47 (2019) 100–116

Contents lists available at ScienceDirect

Aggression and Violent Behavior

journal homepage: www.elsevier.com/locate/aggviobeh

Predicting domestic violence: A meta-analysis on the predictive validity of risk assessment tools

Claudia E. van der Put*, Jeanne Gubbels, Mark Assink

Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Risk assessment tools
Domestic violence
Predictive validity
Meta-analysis
Intimate partner violence

ABSTRACT

Risk assessment tools are increasingly being used to guide decisions about supervision and treatment of domestic violence perpetrators. However, earlier review studies showed that the predictive validity of most of these tools is limited, and is reflected in small average effect sizes. The present study aimed to meta-analytically examine the predictive validity of domestic violence risk assessment tools, and to identify tool characteristics that positively moderate the predictive validity. A literature search yielded 50 independent studies (N = 68,855) examining the predictive validity of 39 different tools, of which 205 effect sizes could be extracted. Overall, a significant discriminative accuracy was found (AUC = 0.647), indicating a moderate predictive accuracy. Tools specifically developed for assessing the risk of domestic violence performed as well as risk predictions based on victim ratings and tools designed for predicting general/violent criminal recidivism. Actuarial instruments (AUC = 0.657) outperformed Structured Clinical Judgment (SCJ) tools (AUC = 0.580) in predicting domestic violence. The onset of domestic violence (AUC = 0.744) could be better predicted than recurrence of domestic violence (AUC = 0.643), which is a promising finding for early detection and prevention of domestic violence. Suggestions for the improvement of risk assessment strategies are presented.

1. Introduction

Domestic violence against women is a widespread phenomenon affecting the lives of millions of women all over the world (Alhabib, Nur, & Jones, 2010). Domestic violence has significant consequences for victims, such as serious physical and mental health problems (Campbell, 2002). To reduce domestic violence recidivism, clinicians, correctional personnel, police officers, and victim service workers continuously judge the risk and dangerousness of spousal assaulters (Dutton & Kropp, 2000). Based on these judgments, decisions are made about the appropriate (intensity of) treatment. Risk assessment tools are increasingly being used to guide decisions about supervision and treatment of perpetrators, even though the development and evaluation of these tools are still in its infancy (Dutton & Kropp, 2000; Hanson, Helmus, & Bourgon, 2007). Compared to studies on the predictive validity of risk assessment tools for sexual, general, or violent offending, only few studies have been published on the predictive validity of risk assessment tools for domestic violence (Hanson, Bourgon, & Helmus, 2007). Moreover, earlier review studies showed that the predictive validity of most risk assessment tools for domestic violence is limited, and is reflected in small average effects sizes (Hanson et al., 2007; Messing & Thaller, 2013). Therefore, it is essential to gain insight into which approaches to risk assessment perform well, and which instrument characteristics (e.g., specific instrument, length of instrument) and study characteristics (e.g., type of sample, type of assessor) influence the predictive validity positively or negatively, and thus act as moderators. Identifying moderators yields important knowledge that can be used in developing and/or improving risk assessment tools. Therefore, we conducted a three-level meta-analysis, in which we estimated the average predictive accuracy of risk assessment tools for domestic violence, and examined variables that potentially influence this accuracy.

Several terms are used interchangeably for describing domestic violence, including intimate partner violence, spousal violence, spousal assault, family violence, wife assault, wife abuse, and spouse abuse. In this study, we chose to use the term domestic violence, which we defined as physical violence, verbal violence, psychological abuse, and sexual violence against intimate partners (or ex-intimate partners), both in mild forms and more severe forms. In the risk assessment literature, there is debate surrounding the purpose of risk assessment, with some arguing that the goal is to predict recidivism, and others arguing that the goal is to prevent violence and to manage risk (Douglas & Kropp,

* Corresponding author at: Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15780, 1001 NG Amsterdam, The Netherlands.

E-mail address: C.E.vanderPut@UvA.nl (C.E. van der Put).

https://doi.org/10.1016/j.avb.2019.03.008
Received 12 July 2018; Received in revised form 28 January 2019
Available online 01 April 2019
1359-1789/ © 2019 Elsevier Ltd. All rights reserved.

2002; Kropp, 2008). Important elements of risk management are the monitoring of changes in risk, treatment, supervision, and victim safety planning. In the present study, we focus on the use of risk assessment instruments for predicting domestic violence. Some specialized domestic violence risk assessment tools have been developed to assess the risk that a *perpetrator* will re-offend, whereas other tools assess the risk that a *victim* will be re-victimized. Further, some risk assessment tools also provide information on “the nature, form, and degree of the danger” of violence (Kropp, 2008). Depending on the (clinical) purpose of the risk assessment, different specialized instruments exist to help assessors achieve their goals.

A number of previous review studies has been published on the predictive accuracy of domestic violence risk assessment tools, including several narrative reviews (e.g., Dutton & Kropp, 2000; Hilton & Harris, 2005; Kropp, 2008; Nicholls, Pritchard, Reeves, & Hilterman, 2013) and two quantitative reviews (Hanson et al., 2007; Messing & Thaller, 2013). Messing and Thaller (2013) examined the average predictive validity of five intimate partner violence risk assessment instruments, and found a small average effect size for four of these instruments. Hanson et al. (2007) found a small to moderate predictive accuracy for the approaches most commonly applied in predicting domestic violence. These approaches are: (a) using tools that are specifically designed for assessing the risk of domestic violence recidivism, (b) using risk assessment tools designed for general or violent recidivism, and (c) using victim (partner) ratings. Hanson and colleagues found a similar predictive accuracy for these three approaches, as no differences in accuracy were found between risk assessment tools for domestic violence, risk assessment tools for general or violent recidivism, and risk assessment based on victim ratings. However, the authors concluded that the lack of evidence for superiority of any one of these approaches is likely due to limited research (Hanson et al., 2007). In the last decade, a number of new studies have been conducted on the predictive validity of risk assessment tools for predicting domestic violence. Therefore, we meta-analytically re-examined the predictive validity of risk assessment tools for domestic violence.

In the present meta-analysis, we tested which risk assessment approach (specialized domestic violence risk assessment tools, general/violence risk assessment tools, or victim-partner ratings) is to be preferred in terms of predictive accuracy. In addition, we examined risk assessment method as potential moderator of the predictive accuracy. Regarding assessment method, risk assessment tools (both tools specifically developed for domestic violence recidivism and tools for general/violent criminal recidivism) can be divided into actuarial tools and structured clinical judgment (SCJ) tools. In actuarial tools, conclusions are based solely on empirically established relationships between risk factors and the outcome of interest, whereas in SCJ tools, conclusions are based on the judgment of a professional who combines and weighs information in a subjective manner (Dawes, Faust, & Meehl, 1989). Differences in predictive accuracy between actuarial and SCJ tools have not yet been meta-analytically examined for tools specifically designed for predicting domestic violence. We expected actuarial tools to outperform SJC tools, because several review studies showed that actuarial tools outperform SCJ tools in predicting general and violent recidivism (e.g., Aegisdóttir et al., 2006; Dawes et al., 1989; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Hanson & Morton-Bourgon, 2009; Hilton, Harris, & Rice, 2006). In some actuarial tools, the professional is allowed to “override” the outcome of the assessment by adjusting the score either upwards or downwards (further referred to as actuarial tools with override). Based on previous research, we expected actuarial tools without the override option to outperform actuarial tools with the override option. Earlier studies showed that when professionals were allowed to adjust the assessment outcome, the actuarial conclusions that were correctly modified (by overriding the outcome) were outnumbered by those incorrectly modified (Dawes et al., 1989).

Besides risk assessment approach and method, we examined specific (/individual) instruments and length of instrument (number of items) as potential moderators of the predictive accuracy of risk assessment tools. In addition, the following study design characteristics were examined: focus of the assessment (predicting the *recurrence* of domestic violence versus predicting the *onset* of domestic violence in high-risk or general samples), study design (retrospective versus prospective design), sample that was used for validating the instrument (validation versus construction sample), length of follow-up, type of follow-up, and source of outcome measure. Below, we elaborate on the rationale for testing these characteristics as potential moderators.

1.1. Specific instrument

We examined the moderating effect of specific or individual instruments to gain knowledge about which instruments perform relatively well. This information is relevant for professionals in selecting an appropriate instrument for use in practice.

1.2. Length of instrument

We examined the number of items a tool is comprised of, because the predictive validity may vary with the length of the tool. For example, Schwalbe (2007) conducted a meta-analysis on juvenile justice risk assessment tools and found that brief tools yielded smaller effect sizes than tools of longer length. However, Van der Put, Assink, and Boekhout van Solinge (2017) did not find a significant moderating effect for length of risk assessment instruments for predicting child maltreatment.

1.3. Focus of the assessment

Most tools for predicting domestic violence are aimed at predicting the recurrence of domestic violence. However, tools are sometimes used to assess the onset of domestic violence in high-risk groups or even in the general population, which is important for early detection and prevention of domestic violence. We examined a potential difference in the predictive validity of tools assessing the recurrence versus tools assessing the onset of domestic violence.

1.4. Study design

We examined the effect of study design on predictive validity, because prospective studies may produce different predictive accuracies than retrospective studies. Some researchers have argued that risk assessment tools can be examined retrospectively, using file information from sources such as institutional files, psychological reports, and/or court reports (e.g., De Vogel, De Ruiter, Hildebrand, Bos, & Van de Ven, 2004). In contrast, other researchers have argued that prospective research is required to adequately examine the predictive validity of a risk assessment tool (Caldwell, Bogat, & Davidson II, 1988).

1.5. Sample used for validation

In some studies, the predictive validity of an instrument is examined in the same sample that was used to construct the instrument, whereas in other studies, the predictive validity is examined in a sample independent of the construction sample. We expected the predictive validity to be lower in validation samples than in construction samples, because random sampling error arising from testing an instrument in a sample that differs from a construction sample, results in reduced predictive validity estimates. In fact, models built in a construction (or training) sample tend to “overfit” the data (i.e., capitalizing on random variation). Thus, predictive validity estimates reported for construction samples are commonly inflated. In addition, we made a distinction between dependent and independent validation samples. We coded the validation sample as dependent when the total sample of a study was randomly divided into a sample to construct the tool and a different sample to validate the tool (in other words, the construction and validation samples are randomly selected from the same sample). The validation sample was coded as independent, in case a new sample was used to validate the tool. We expected the predictive validity to be lower in independent samples than in dependent samples.

1.6. Follow-up length, type, and source of outcome measure

The potential moderating effect of follow-up length of studies was examined, because the predictive validity may vary over time and differences in follow-up length are frequently observed between studies. Further, studies on the predictive validity of risk assessment tools vary in the outcome that is predicted. We examined whether the predictive validity is influenced by type of domestic violence that is predicted (general domestic violence, physical and sexual domestic violence, non-physical domestic violence, severe/near fatal domestic violence, or violation of restraining order), and source of outcome (official records, partner reports, and self-reports).

Finally, a number of additional variables was exploratively tested as potential moderating variables. These variables were: publication year, percentage of males in a sample, percentage of cultural minorities in a sample, and the mean age of a sample.

In summary, domestic violence risk assessment tools are increasingly being used, despite the fact that earlier review studies showed limited predictive accuracy of most of these tools. Therefore, the aim of the present meta-analysis was to examine whether and how the predictive validity is influenced by study and instrument characteristics. This knowledge is not only scientifically important, but also clinically relevant, as it provides guidance on improving risk assessment tools and/or implementing the most effective tools. Consequently, this review may contribute to better decision making in domestic violence cases. A three-level random-effects meta-analysis was performed to estimate the overall predictive validity of risk assessment tools for domestic violence, and to identify variables that moderate this predictive validity.

2. Method

2.1. Inclusion criteria

For the selection of studies, several criteria were formulated. First, studies were selected when they examined the predictive validity of risk assessment tools for predicting domestic violence. Domestic violence could be physical violence, verbal violence, psychological abuse, and sexual violence against intimate partners (or ex-intimate partners), both in mild forms and more severe forms. We excluded studies examining tools that predicted general violence or general criminal recidivism. Second, studies had to examine the predictive validity of instruments for the risk of domestic violence (i.e., domestic violence in the future). When only the immediate risk for (or threat of) domestic violence was examined, the study was excluded, because this refers to the victim's safety assessment (as opposed to risk assessment). Third, both prospective and retrospective studies were included. Finally, studies had to report either an actual effect size of the predictive validity of an instrument (e.g., an Area Under the receiver operating characteristics Curve (AUC) value, a correlation (r), or Cohen's d), or sufficient statistical information for manually calculating an effect size.

2.2. Search strategy

The electronic databases PsycINFO, Web of Science, ScienceDirect, Sociological Abstracts, and Google Scholar were searched for articles, books, chapters, dissertations, and reports. Until April 2018, studies were collected using the following keywords regarding study design, instrument features, study outcomes, and participants, that were combined in different ways: 'predictive validity', 'predictive accuracy', 'AUC', 'ROC', 'sensitivity', 'specificity', 'risk assessment', 'risk measure', 'risk instrument', 'screen*', 'risk tool' 'intimate partner violence', 'domestic violence', 'family violence', 'batterer*', 'battered women', 'spousal assault', 'wife assault', 'partner assault', 'spousal abuse', 'wife abuse', 'partner abuse', 'femicide', and 'domestic homicide'. These keywords were also combined with the names of some well-known risk assessment instruments for domestic or general violence. To assess the studies against the inclusion criteria, titles, abstracts and, if necessary, full article texts were examined. In addition, the reference list of several review studies were screened in order to find relevant studies. First, we screened the reference lists of review studies on the predictive accuracy of domestic violence risk assessment tools (i.e., Dutton & Kropp, 2000; Hanson et al., 2007; Hilton & Harris, 2005; Kropp, 2008; Messing & Thaller, 2013; Nicholls et al., 2013). Second, we screened the reference lists of review studies on screening instruments, referring to the victim's safety assessment as opposed to risk assessment (i.e., Arkins, Begley, & Higgins, 2016; Canales, Macaulay, McDougall, Wei, & Campbell, 2013; Rabin, Jennings, Campbell, & Bair-Merritt, 2009;). Third, we screensed the reference lists of review studies on individual or specific types of instruments (i.e., female victim risk appraisals [Bowen, 2011]; the SARA [Helmus & Bourgon, 2011]), and a review on risk assessment for dating violence (i.e., Tapp & Moore, 2016). Finally, we screened one non-English review for additional studies (i.e., Kilvinger, Rossegger, Urbaniok, & Endrass, 2012).

In total, the search procedure yielded 280 studies. After thoroughly screening these studies, 54 studies that met the inclusion criteria were included. A flow chart of the search procedure is presented in Fig. 1.

2.3. Coding of studies

Following the guidelines of Lipsey and Wilson (2001), a coding scheme was developed to code all relevant study, design, sample, instrument, and outcome characteristics that could moderate the predictive accuracy of domestic violence risk assessment tools. First, the coded *study characteristics* included publication type (peer-reviewed, research report, or thesis/dissertation), publication year, publication status (published vs. not published), and the country in which the study was carried out (USA, Canada, European countries, Australia/New Zealand, or other countries). Second, as a *study design characteristic*, we coded whether a study was prospective or retrospective. Third, the coded *sample characteristics* were the percentage of males in the sample, mean age of the sample, the percentage of minorities in the sample, the type of the sample (independent validation sample, dependent validation sample, or construction sample), and the type of participants (domestic violent participants or general/high risk participants). Fourth, as for *instrument characteristics*, we coded the name of the instrument, instrument length (number of items), the risk assessment method (actuarial, structured clinical judgment, or actuarial with override), the assessment approach (using tools specifically designed for domestic violence, using tools for general violence or general recidivism, using victim's risk ratings, or assessing psychopathic traits), type of administrator (professional, researcher/research assistant, or any other type of administrator), type of respondent (offender/batterer, victim or both), and whether the instrument was filled out in practice or by using file information. Some studies on the predictive validity of SCJ tools for assessing domestic violence (e.g., the Spousal Assault Risk Assessment Guide (SARA)) also examined the predictive validity of a total risk score (the sum of scores on individual items). In these studies, we examined the predictive validity of both type of assessment methods. So, for the SARA, both the predictive validity of the SCJ assessment as the actuarial assessment was examined (e.g., Belfrage et al., 2011).

Finally, as *outcome characteristics*, we coded the follow-up duration in months, the starting point of the follow-up (directly after the assessment, after treatment or punishment, after the index crime, or another starting point), whether the outcome was corrected for cases in

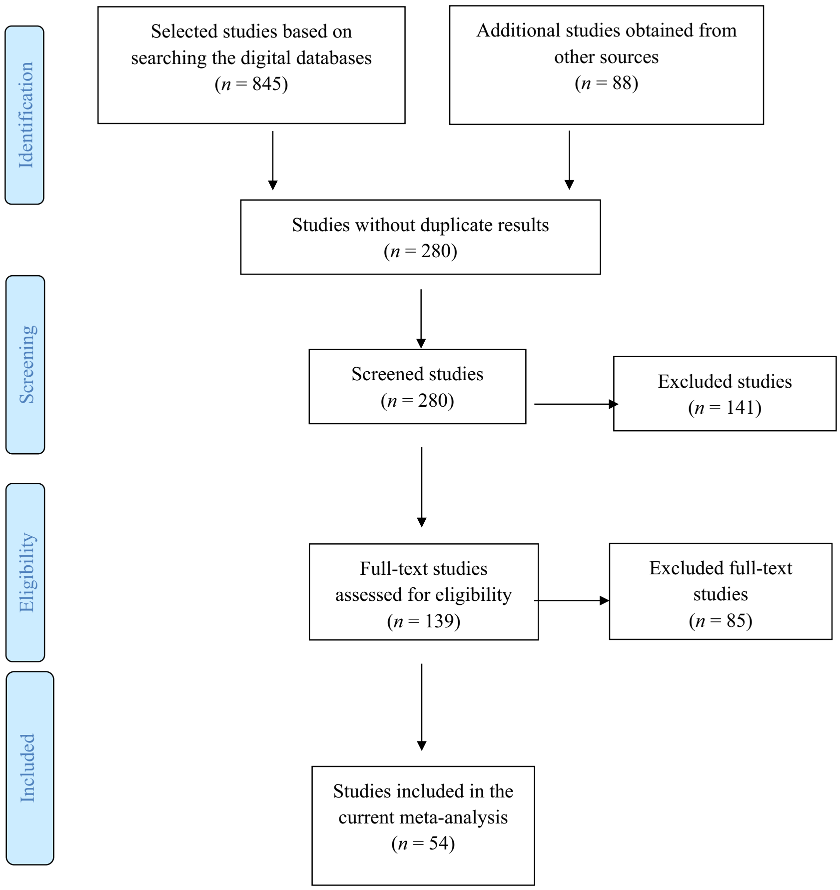

Fig. 1. Flowchart of the search procedure.

Test Result	Outcome: Positive	Outcome: Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Fig. 2. 2 × 2 contingency table comparing risk assessment tool predictions and outcomes.

which an intervention was started, and the source of the domestic violence outcome (official reports, partner reports, or self-reports/other).

2.4. Calculation of effect sizes

In the current meta-analysis, all outcomes of the primary studies were transformed into AUC values (Area Under the Curve from Receiver Operator Characteristic Curve (ROC) analysis), as an AUC value is the most common performance indicator for the predictive validity of risk assessment tools. The AUC value is a preferred measure of predictive accuracy (e.g., Swets, Dawes, & Monahan, 2000), as it is resistant to variation in base rates, selection ratios, and truncated distributions (Rice & Harris, 1995). The AUC ranges from 0.50 to 1.00, and describes the probability that a randomly selected domestic violence case will have a higher risk assessment score than a randomly selected non-domestic violence case. In other words, the AUC value is an index of how well a risk assessment instrument discriminates between domestic violence cases and non-domestic violence cases across all possible cut-off scores of the instrument. AUC values between 0.556 and 0.639 correspond with a small effect size, AUC values between 0.639 and 0.714 with a moderate effect size, and AUC values of 0.714 and higher with a large effect size (Rice & Harris, 2005).

Most primary studies reported on AUC values of the risk assessment instrument being studied. When other statistical information was reported, several methods were used to transform this information into AUC values. Converting Pearson's correlations (r) into AUC values was done using the formulas of Rosenthal (1994). When only the sensitivity (i.e., the proportion of domestic perpetrators who were classified as high risk) and specificity (i.e., the proportion of domestic perpetrators who were classified as low risk) were presented in the primary studies, the formula [sensitivity + specificity] / 2 was used to calculate the AUC value. However, when a 2 × 2 contingency table was given with true and false positives, and true and false negatives (see Fig. 2), this information was used to calculate the sensitivity (with the formula [[number of true positives] / [number of true positives + number of false negatives]]) and the specificity (with the formula [[number of true negatives] / [number of true negatives + number of false positives]]; Singh, 2013).

2.5. Data analyses

First, the AUC values were transformed into Pearson's correlations (using the formulas of Ruscio (2008) and Rosenthal (1994)), as an effect size in which the value zero represents a random discriminative accuracy was required for the statistical analyses. After all correlation coefficients were obtained, the r values were transformed into Fisher's z

values, as correlations are non-normally distributed (see, for instance, Lipsey & Wilson, 2001).

Next, extreme values of effect sizes and sample sizes ($Z > 3.29$ or $Z < -3.29$; Tabachnik & Fidell, 2013) were identified, as these extreme values may have excessive influence on the results of the current meta-analysis. No outliers were found in the extracted effect sizes. However, outliers with a Z value > 3.29 were found in the sample size of two studies (that produced six effect sizes). In order to reduce the influence of these large studies on our results, the original coded sample sizes were adjusted downward to the nearest sample size falling within the normal range of sample size.

Because most studies reported on more than one effect size, a traditional random effects (two-level) model was extended to a three-level random effects model (Cheung, 2014; Houben, Van den Noortgate, & Kuppens, 2015; Van den Noortgate, López-López, Marin-Martinez, & Sánchez-Meca, 2013, Van den Noortgate, López-López, Marin-Martinez, & Sánchez-Meca, 2015). A major advantage of this three-level approach is the absence of the need to aggregate or select data, implying that all relevant effect sizes can be extracted from primary studies (see also Assink & Wibbelink, 2016). This means no loss of information, a more precise estimation of (moderator) effects, and a maximum statistical power in the analyses. In our meta-analytic model, three forms of variance were taken into account: sampling variance of the observed effect sizes (Level 1), variance between effect sizes extracted from the same study (Level 2), and variance between studies (Level 3). By building this model without covariates (i.e., an intercept-only model), an overall effect can be estimated that is represented by the intercept. Further, in case of significant variation in effect sizes from the same study (i.e., level 2 variance) and/or significant variation between studies (i.e., level 3 variance), the model was extended with potential moderating variables to determine whether this variation can be explained by study, sample, design, instrument, or outcome characteristics.

The statistical software package R (version 3.5.0) and the function "rma.mv" of the metafor package (Viechtbauer, 2010) were used to build the three-level meta-analytic models. We used the syntax as described by Assink and Wibbelink (2016). In these three-level meta-analytic models, the significance and corresponding confidence intervals of the individual regression coefficients was tested using the t- and F- distributions (Knapp & Hartung, 2003). To determine the significance of the level 2 and level 3 variance, two separate log-likelihood ratio tests were performed in which the full model was compared to a model excluding either the level 2 or level 3 variance parameter. The distribution of effect sizes was considered to be heterogeneous when the variance at level 2 and/or 3 was significant, and consequently, moderator analyses were performed to identify variables that can explain this significant (level 2 or level 3) variance. All model parameters were estimated using the restricted maximum likelihood estimation method and prior to the analyses, each continuous variable was centered around its mean and dichotomous dummy variables were created for all categories of discrete variables. The log-likelihood-ratio-tests were performed one-tailed and all other tests were performed two-tailed. We considered p values < .05 as statistically significant, and p values < .10 as trend significant.

2.6. Publication Bias

A common problem in conducting a meta-analysis is that studies with nonsignificant or negative results are less likely to be published than studies with positive and significant results. This phenomenon is called publication bias and is often referred to as the 'file drawer problem' (Rosenthal, 1995). Besides publication bias, the results may be affected by other forms of bias, such as coding or selection bias. To determine whether results of the present meta-analysis were affected by (different forms of) bias, a funnel plot based trim-and-fill analysis was conducted (Duval & Tweedie, 2000a, 2000b) using the function "trimfill" of the metafor package (Viechtbauer, 2010) in the R environment (Version 3.5.0; R Core Team, 2015). A funnel plot is a scatter plot of the effect sizes against the effect size's precision (1 divided by the standard error). This method is built on the assumption that effect sizes are symmetrically distributed (in the form of a funnel) around the "true" effect size, when bias in the results is absent. In case of an asymmetric plot, the asymmetry is restored by imputing effect sizes that are estimated on the basis of existing effect sizes in the dataset. Subsequently, by means of a sensitivity analysis, an "adjusted" overall effect can be estimated, using the dataset to which the imputed effect sizes that were produced by the trim-and-fill algorithm have been added. In this way, the degree to which the results were affected by bias can be made visible.

3. Results

3.1. Descriptive characteristics, central tendency, and variability

The present study included 50 studies (k) published between 2000 and 2018 (the median publication year was 2009). In total, these studies reported on validation research of 39 different risk assessment tools for domestic violence, from which 205 effect sizes could be extracted. Each effect size represented the discriminative accuracy of a risk assessment instrument or a statistical predictive model that was used for the purpose of risk assessment. An overview of all included studies, the risk assessment tools that have been examined in these studies, and several study characteristics is presented in Table 1.

The total sample size consisted of $N = 68,855$ individuals for whom the risk for domestic violence was assessed using one of the risk assessment tools as listed in Table 1. Sample sizes in the primary studies ranged from 26 to 29,317 participants. The included studies were conducted in the USA ($k = 24$), Europe ($k = 10$), Canada ($k = 10$), Australia/New Zealand ($k = 4$), and in other countries ($k = 2$).

The statistical analyses yielded an overall effect of $z = 0.263$ (SE = 0.016), $t(204) = 16.083$, $p < .001$, which equals an AUC value of 0.647 (see Table 2). The results of the trim-and-fill-analysis suggested that bias was present in the dataset, because the distribution of effect sizes was not symmetrical. This asymmetry could also be detected by visually inspecting the funnel plot (see Fig. 3). This plot shows that effect sizes were missing to the left of the estimated mean effect, meaning that in particular small and negative effect sizes were missing in the dataset we analyzed. According to the trim-and-fill analysis, 38 effect sizes had to be imputed to the left of the estimated mean effect to restore the symmetry of the funnel. Based on the standard errors that were produced by the trim-and-fill algorithm (and the fact that some of these errors were equal), we derived that these 38 "missing" effect sizes could have been produced by 23 "missing" studies. Next, these "missing" effect sizes (with a unique effect size ID, and a study ID) were added to the dataset, so that a "corrected" overall effect could be estimated. In this way, we could determine to what extent our initially estimated overall effect ($z = 0.263$; see above) may have been influenced by bias. *Re*-estimating the overall effect produced an effect size of $z = 0.177$ (SE = 0.021), $t(242) = 8.558$, $p < .001$, equaling an AUC value of 0.599 (see Table 2). The difference between the two effects (0.263–0.177 = 0.086) is rather modest, implying that our results have not been seriously influenced by (publication) bias.

As for heterogeneity in effect sizes, the one-sided likelihood-ratio tests showed significant variance both on the second level $\chi^2(1) = 849.859$, $p < .001$ and the third level $\chi^2(1) = 53.897$, $p < .001$ of the meta-analytic model. Consequently, we proceeded to moderator analyses to examine whether characteristics of risk assessment instruments and/or studies could (partly) explain level 2 and/or level 3 variance.

Table 1
Included studies and their characteristics.

Author(s) (Pub. year)	*N*	Type of sample	Name of instrument	Assessment approach	Method of assessment	Source of DV outcome	Type of DV	AUC
Belfrage et al. (2011)	429	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	SCJ	Official records	Physical	0.5700
	429	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	Physical	0.6300
Belfrage and Strand (2012)	214	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	Severe	0.4567
	215	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.4996
Berk, He, and Sorenson (2005)	516	C	No name given	Domestic violence	Actuarial	Official records	General	0.5915
Bourgon and Bonta (2004)	613	V (ID)	Secondary Risk Assessment for Partner Abusers (SRA—PA)	Domestic violence	Actuarial	Official records	General	0.6100
	613	V (ID)	Primary Risk Assessment (PRA)	General violence	Actuarial	Official records	General	0.6200
Callan-Bartkiw (2012)	32	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.7200
	32	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	SCJ	Official records	General	0.7800
Cattaneo, Bell, Goodman, and Dutton (2007)	246	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Physical	0.5783
(2) Bell, Cattaneo, Goodman, and Dutton (2008)	244	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Non-physical	0.6252
Cattaneo and Goodman (2003)	96	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	General	0.6600
Chan (2014)	1114	V (D)	Chinese Risk Assessment Tool for Perpetrators (CRAT-P)	Domestic violence	Actuarial	Self-report/other	Physical	0.7600
Dayan, Fox, and Morag (2013)	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	Physical	0.6717
	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	Physical	0.6776
	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	Physical	0.6954
	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	General	0.7074
	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	Non-physical	0.7743
	81	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Victim report	Non-physical	0.7867
	1133	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Official records	General	0.5700
	1133	V (ID)	Spouse Violence Risk Assessment Inventory (SVRA-I)	Domestic violence	Actuarial	Official records	Physical	0.5800
De Ruiter (2011)	200	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.7000
Echeburúa, Fernández-Montalvo, de Corral, and López-Go-i (2009)	1081	V (ID)	Severe Intimate Violence Partner Risk Prediction Scale (SIVIPAS)	Domestic violence	Actuarial	Self-report/other	Severe	0.6465
Ferguson (2011)	446	V (ID)	Field Reassessment of the Offender Screening Tool (FROST)	General violence	Actuarial	Official records	General	0.5690
	446	V (ID)	Field Reassessment of the Offender Screening Tool (FROST)	General violence	Actuarial	Official records	General	0.6520
	573	V (ID)	Offender Screening Tool (OST)/ Field Reassessment of the Offender Screening Tool (FROST)	General violence	Actuarial	Official records	General	0.5730
	573	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	General	0.5830
Fitzgerald and Graham (2016)	7330	V (D)	No name given	Domestic violence	Actuarial	Official records	General	0.6940
	7330	C	No name given	Domestic violence	Actuarial	Official records	General	0.7010
Glass et al. (2008)	93	C	Danger Assessment revised (DA-R)	Domestic violence	Actuarial	Victim report	Physical	0.5395
Gondolf & Wernik, 2009	302	V (ID)	No name given	Domestic violence	SCJ	Victim report	Physical	0.6020
	302	V (ID)	No name given	Domestic violence	SCJ	Victim report	Severe	0.6460
	337	V (ID)	No name given	Domestic violence	SCJ	Victim report	Physical	0.5000
	337	V (ID)	No name given	Domestic violence	SCJ	Victim report	Severe	0.6620
Goodman, Dutton, and Bennett (2000)	49	V (ID)	Conflict Tactics Scale 2 (CTS2)	Domestic violence	Actuarial	Victim report	General	0.6544
	49	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	General	0.7114
Grann and Wedin (2002)	56	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6500
	83	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6300
	87	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.5900
	88	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.5200
Gray (2007)	91	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Official records	General	0.6070
	94	V (ID)	Statistical Information on Recidivism-Revised 1 (SIR-R1)	General violence	Actuarial	Official records	General	0.5900
	94	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6040
	94	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7120
	94	V (ID)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	General	0.7130
Heckert and Gondolf (2004)	499	V (ID)	Kingston Screening Instrument for Domestic Violence Offenders (K-SID)	Domestic violence	Actuarial	Victim report	General	0.5700
	499	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	General	0.7000
	499	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Victim report	General	0.6400
	499	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	General	0.6400

(*continued on next page*)

Table 1 (*continued*)

Author(s) (Pub. year)	*N*	Type of sample	Name of instrument	Assessment approach	Method of assessment	Source of DV outcome	Type of DV	AUC
	499	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	General	0.6300
Hendricks, Werner, Shipway, and Turinetti (2006)	200	V (ID)	Level of Service Inventory–Revised (LSI-R)	General violence	Actuarial	Official records	General	0.6350
Hilton et al. (2004)	100	V (D)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	General	0.5300
	589	C	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	General	0.5900
	100	V (D)	Domestic Violence Supplementary Report (DVSR)	Domestic violence	Actuarial	Official records	General	0.5900
	589	C	Domestic Violence Supplementary Report (DVSR)	Domestic violence	Actuarial	Official records	General	0.6700
	100	V (D)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.5400
	589	C	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6400
	100	V (D)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7200
Hilton, Harris, Rice, Houghton, and Eke (2008)	649	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	General	0.5964
	649	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	Physical	0.6193
	649	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Self-report/other	Severe	0.6251
	649	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	General	0.6367
	346	V (D)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Self-report/other	Severe	0.6836
	346	V (D)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	General	0.7074
	346	V (D)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	Physical	0.7375
	346	V (D)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	General	0.7558
	649	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Self-report/other	Severe	0.5792
	649	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	Physical	0.5964
	649	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	General	0.6136
	649	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	General	0.6251
	346	V (D)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Self-report/other	Severe	0.6483
	346	V (D)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.6659
	346	V (D)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	Physical	0.6954
	346	V (D)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7074
	649	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Official records	General	0.6659
	649	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Self-report/other	Severe	0.6659
	649	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Official records	General	0.7074
	649	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Official records	Physical	0.7134
	649	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6193
	649	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	Physical	0.6251
	649	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Self-report/other	Severe	0.6309
	649	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6542
	649	V (ID)	Violence Risk Appraisal Guide (VRAG)	General violence	Actuarial	Self-report/other	Severe	0.6425
	649	V (ID)	Violence Risk Appraisal Guide (VRAG)	General violence	Actuarial	Official records	General	0.6659
	649	V (ID)	Violence Risk Appraisal Guide (VRAG)	General violence	Actuarial	Official records	Physical	0.6776
	649	V (ID)	Violence Risk Appraisal Guide (VRAG)	General violence	Actuarial	Official records	General	0.6836
Hilton, Harris, Popham, and Lang (2010)	140	V (ID)	Level of Service Inventory–Ontario Revised (LSI-OR)	General violence	Actuarial	Official records	General	0.5020
	150	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.6380
Hilton and Harris (2009)	309	V (ID)	Domestic Violence Supplementary Report (DVSR)	Domestic violence	SCJ	Official records	General	0.6100
	309	V (ID)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	General	0.7400
Houry et al. (2004)	96	V (ID)	Partner Violence Screen (PVS)	Domestic violence	Actuarial	Victim report	Non-physical	0.6960
	96	V (ID)	Partner Violence Screen (PVS)	Domestic violence	Actuarial	Victim report	Physical	0.7572
Jung and Buro (2017)	190	V (ID)	Family Violence Investigative Report (FVIR)	Domestic violence	Actuarial	Official records	General	0.5700
(2) Olver and Jung (2017)	190	V (ID)	Family Violence Investigative Report (FVIR)	Domestic violence	Actuarial	Official records	General	0.5700
	289	V (ID)	Family Violence Investigative Report (FVIR)	Domestic violence	Actuarial	Official records	General	0.6000
	100	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7000
	226	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.6600
	289	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7200
	125	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6800
	198	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.7400
	289	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.7400
Koziol-McLain, Coates, and Lowenstein (2001)	405	V (ID)	No name given	Domestic violence	Actuarial	Victim report	Non-physical	0.5955
	405	V (ID)	No name given	Domestic violence	Actuarial	Victim report	Physical	0.7203

(*continued on next page*)

Table 1 (*continued*)

Author(s) (Pub. year)	*N*	Type of sample	Name of instrument	Assessment approach	Method of assessment	Source of DV outcome	Type of DV	AUC
	405	V (ID)	No name given	Domestic violence	Actuarial	Victim report	Severe	0.8650
	405	V (ID)	No name given	Domestic violence	Actuarial	Victim report	Physical	0.5533
Kropp and Hart (2000)	102	V (ID)	Psychopathy Checklist-Revised (PCL-R) Screening version	Psychopathy	Actuarial	Official records	General	0.5541
	102	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6019
Lauria, McEwan, Luebbers, Simmons, and Ogloff (2017)	198	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	Physical	0.6800
	200	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	Non-physical	0.7200
Loinaz (2014)	40	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER) Spanish translation	Domestic violence	Actuarial	Victim report	General	0.7600
Mason and Julian (2009)	3569	V (ID)	Risk Assessment Screening Tool (RAST) (short version)	Domestic violence	Actuarial	Official records	General	0.7260
	1406	V (ID)	Risk Assessment Screening Tool (RAST) (all items)	Domestic violence	Actuarial	Official records	General	0.6020
Messing, Amanor-Boadu, Cavanaugh, Glass, and Campbell (2013)	148	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	Severe	0.6920
	148	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	Physical	0.6868
	148	C	Danger Assessment for Immigrant Women (DA-I)	Domestic violence	Actuarial	Victim report	Severe	0.8522
	148	C	Danger Assessment for Immigrant Women (DA-I)	Domestic violence	Actuarial	Victim report	Physical	0.7745
	148	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Severe	0.6375
	148	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Physical	0.6246
	148	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Severe	0.6535
	148	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Physical	0.6390
Messing, Campbell, Sullivan Wilson, Brown, and Patchell (2017)	233	V (ID)	Lethality Screen (11-item version of the DA)	Domestic violence	Actuarial	Victim report	General	0.5445
	240	V (ID)	Lethality Screen (11-item version of the DA)	Domestic violence	Actuarial	Victim report	Non-physical	0.5385
	44	V (ID)	Lethality Screen (11-item version of the DA)	Domestic violence	Actuarial	Victim report	Severe	0.5738
	28	V (ID)	Lethality Screen (11-item version of the DA)	Domestic violence	Actuarial	Victim report	Severe	0.5708
Messing, Campbell, and Snider (2017)	200	V (ID)	Danger Assessment 5 (DA-5)	Domestic violence	Actuarial	Victim report	Severe	0.6800
Moser, Campbell, and Campus (2012)	200	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7000
	200	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	Non-physical	0.7000
	174	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.7000
	26	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.6700
	200	V (ID)	P-Trait	Psychopathy	Actuarial	Official records	General	0.8000
	174	V (ID)	P-Trait	Psychopathy	Actuarial	Official records	General	0.8000
	26	V (ID)	P-Trait	Psychopathy	Actuarial	Official records	General	0.8500
	200	V (ID)	Ontario Domestic Assault Risk Assessment-Revised (R-ODARA)	Domestic violence	Actuarial	Official records	General	0.7900
	174	V (ID)	Ontario Domestic Assault Risk Assessment-Revised (R-ODARA)	Domestic violence	Actuarial	Official records	General	0.7800
	26	V (ID)	Ontario Domestic Assault Risk Assessment-Revised (R-ODARA)	Domestic violence	Actuarial	Official records	General	0.8200
Murphy, Morrel, Elliott, and Neavins (2003)	76	C	Partner Abuse Prognostic Scale (PAPS)	Domestic violence	Actuarial	Self-report/other	Severe	0.7014
	649	C	Partner Abuse Prognostic Scale (PAPS)	Domestic violence	Actuarial	Self-report/other	Physical	0.6367
	346	C	Partner Abuse Prognostic Scale (PAPS)	Domestic violence	Actuarial	Self-report/other	Physical	0.7375
	649	C	Partner Abuse Prognostic Scale (PAPS)	Domestic violence	Actuarial	Official records	General	0.6309
	649	C	Partner Abuse Prognostic Scale (PAPS)	Domestic violence	Actuarial	Self-report/other	Physical	0.6776
Rettenberger and Eher (2013)	125	V (ID)	Domestic Violence Risk Appraisal Guide (DVRAG)	Domestic violence	Actuarial	Official records	General	0.5600
	1497	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA) German version	Domestic violence	Actuarial	Official records	General	0.7100
	66	V (ID)	Psychopathy Checklist-Revised (PCL-R)	Psychopathy	Actuarial	Official records	General	0.7100
Roehl, O'Sullivan, Roehl, Webster, and Campbell (2005)	442	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	Severe	0.6280
	1307	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Official records	Physical	0.6130
	782	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	Physical	0.6350
	782	V (ID)	Danger Assessment (DA)	Domestic violence	Actuarial	Victim report	Severe	0.6700
	1307	V (ID)	DV-MOSAIC	Domestic violence	Actuarial	Official records	Physical	0.4740
	1307	V (ID)	DV-MOSAIC	Domestic violence	Actuarial	Official records	Severe	0.5250
	782	V (ID)	DV-MOSAIC	Domestic violence	Actuarial	Victim report	Severe	0.5130
	782	V (ID)	DV-MOSAIC	Domestic violence	Actuarial	Victim report	Physical	0.5130

(*continued on next page*)

Table 1 (*continued*)

Author(s) (Pub. year)	N	Type of sample	Name of instrument	Assessment approach	Method of assessment	Source of DV outcome	Type of DV	AUC
	1307	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	Physical	0.4870
	1307	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Official records	Severe	0.5670
	782	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Victim report	Physical	0.5080
	782	V (ID)	Domestic Violence Screening Inventory (DVSI)	Domestic violence	Actuarial	Victim report	Severe	0.5970
	1307	V (ID)	Kingston Screening Instrument for Domestic Violence (K-SID)	Domestic violence	Actuarial	Official records	Physical	0.5110
	1307	V (ID)	Kingston Screening Instrument for Domestic Violence (K-SID)	Domestic violence	Actuarial	Official records	Severe	0.5230
	782	V (ID)	Kingston Screening Instrument for Domestic Violence (K-SID)	Domestic violence	Actuarial	Victim report	Physical	0.5160
	782	V (ID)	Kingston Screening Instrument for Domestic Violence (K-SID)	Domestic violence	Actuarial	Victim report	Severe	0.5140
	1465	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Severe	0.6100
	1307	V (ID)	Victim's risk assessment	Victim's risk rating	–	Official records	Physical	0.5720
	1307	V (ID)	Victim's risk assessment	Victim's risk rating	–	Official records	Severe	0.5510
	782	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Physical	0.5990
Snider, Webster, O'sullivan, and Campbell (2009)	400	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	Severe	0.6300
	400	C	Danger Assessment (DA) Short version	Domestic violence	Actuarial	Victim report	Severe	0.7900
Stansfield & Williams, 2014	29,317	V (ID)	Revised Domestic Violence Screening Instrument (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6490
	29,317	V (ID)	Revised Domestic Violence Screening Instrument (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6740
	29,317	V (ID)	Revised Domestic Violence Screening Instrument (DSVI-R)	Domestic violence	Actuarial	Official records	Restrain	0.6900
Stith, Milner, Fleming, Robichaux, and Travis (2016)	142	C	Intimate Partner Physical Injury-Risk Assessment Tool (IPPI-RAT)	Domestic violence	Actuarial	Victim report	Physical	0.7800
Storey, Kropp, Hart, Belfrage, and Strand (2014)	249	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	Actuarial	Official records	General	0.7000
Svalin, Mellgren, Torstensson Levander, and Levander (2017)	65	V (ID)	Police Screening Tool for Violent Crimes (PST-VC)	General violence	SCJ	Official records	General	0.6600
	65	V (ID)	Police Screening Tool for Violent Crimes (PST-VC)	General violence	SCJ	Official records	Physical	0.6700
Svalin, Mellgren, Levander, and Levander (2018)	299	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.5700
	299	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.5500
	299	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.5400
	299	V (ID)	Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER)	Domestic violence	SCJ	Official records	General	0.5000
Ulmer (2015)	14,970	V (ID)	The Level of Service/Case Management Inventory (LS/CMI)	General violence	Actuarial	Official records	General	0.6100
	299	V (ID)	Ontario Domestic Assault Risk Assessment (ODARA)	Domestic violence	Actuarial	Official records	General	0.5700
Weisz, Tolman, and Saunders (2000)	177	V (ID)	Victim's risk assessment	Victim's risk rating	–	Victim report	General	0.7436
Williams (2012)	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6210
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	Restrain	0.7150
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6590
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6570
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.6230
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	Restrain	0.7280
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.7290
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6610
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6590
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	Restrain	0.7280
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.7310
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6650
	3569	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6630
	1406	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.7260
	400	V (ID)	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6300
Williams and Grant (2006)	66	C	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial	Official records	General	0.7100
	499	C	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6400
	782	C	Domestic Violence Screening Instrument Revised (DSVI-R)	Domestic violence	Actuarial override	Official records	General	0.6100
Williams and Houghton (2004)	125	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Victim report	Non-physical	0.5800
	125	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Victim report	Physical	0.4900
	125	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Victim report	Severe	0.6500
	66	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Victim report	Non-physical	0.5600
	97	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Official records	General	0.6100
	619	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Victim report	Severe	0.6800

(*continued on next page*)

Table 1 (*continued*)

Author(s) (Pub. year)	*N*	Type of sample	Name of instrument	Assessment approach	Method of assessment	Source of DV outcome	Type of DV	AUC
Wong and Hisashima (2008) (2) Hisashima (2008) (3) Wong (2008)	1307	V (ID)	Domestic Violence Screening Instrument (DSVI)	Domestic violence	Actuarial	Official records	General	0.6280
	196	V (ID)	Spousal Assault Risk Assessment Guide (SARA)	Domestic violence	Actuarial	Official records	General	0.6140

Note: Pub. year = year of publication; *N* = total sample size; DV = domestic violence; AUC = Area Under the ROC Curve; (2) = secondary article; (3) = tertiary article; V (ID) = validation sample (independent); C = construction sample; V (D) = validation sample (dependent); SCJ = structured clinical judgment.

Table 2
Overall effects before and after trim-and-fill analyses.

	Mean *z* (SE)	95% CI	Sig. mean z (*p*)	% var. at level 1	Level 2 variance	% Var. at level 2	Level 3 variance	% Var. at level 3	AUC-value
Overall effect before trim-and-fill	0.263 (0.016)	0.231, 0.296	< 0.001***	5.4	0.008***	49.4	0.008***	45.2	0.647
Overall effect after trim-and-fill	0.177 (0.021)	0.136, 0.217	< 0.001***	3.0	0.008***	25.0	0.023***	72.0	0.599

Note: Mean z = mean effect size (Fisher's *z*); SE = standard error; CI = confidence interval; Sig = significance; Var = variance; Level 1 variance = sampling variance of observed effect sizes; Level 2 variance = variance between effect sizes extracted from the same study; Level 3 variance = variance between studies; AUC = Area under the ROC curve.

*** $p < .001$.

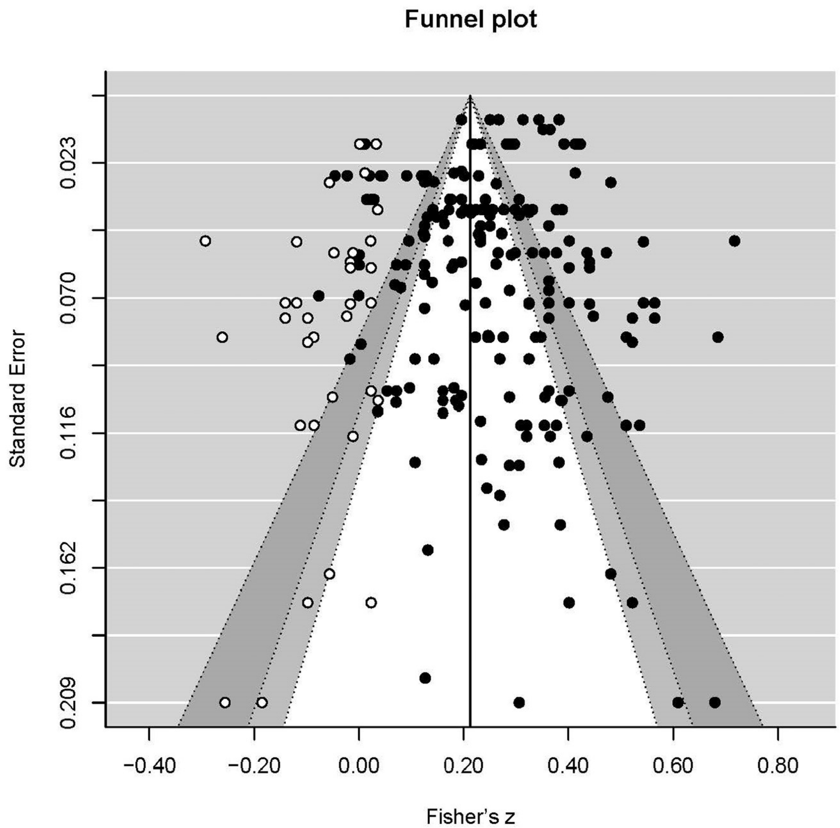

Fig. 3. Funnel Plot. *Note:* Contour enhanced funnel plot with the standard error on the y-axis and Fisher's *z* on the x-axis. The black dots denote the observed effect sizes, and the white dots denote the filled effect sizes. The solid vertical line represents the overall mean effect. From inside to outside, the dashed lines limit the 90%, 95%, and 99% pseudo confidence interval regions.

3.2. Moderator analyses

Each potential moderator of interest was examined in a bivariate model. The results of these analyses can be found in Table 3 in which potential moderators are classified into instrument characteristics and study design characteristics.

3.2.1. Instrument characteristics

The results showed no significant moderating effect for risk assessment approach. Tools specifically designed for predicting domestic violence (AUC = 0.647) did not perform significantly better than assessments based on victim ratings (AUC = 0.637) or risk assessment instruments designed for general/violent recidivism (AUC = 0.638). In searching and coding primary studies, we also found studies in which domestic violence was predicted with screening tools for psychopathology, such as the Psychopathy Checklist-Revised (PCL-R). The mean predictive value of these psychopathy tools for predicting domestic violence (AUC = 0.684) was not significantly different from the other approaches.

Second, we examined whether any specific domestic violence instrument or screening tool for psychopathology performed better than general or violent recidivism instruments. Most of the instruments performed better than chance, with the exception of the CTS2, DV-MOSAIC, and FVIR. The following instruments performed better than the reference category (general and violent recidivism instruments): the CRAT-P (AUC = 0.760; trend significant), DVRAG (AUC = 0.715), IPPA-RAT (AUC = 0.780; trend significant), ODARA (AUC = 0.690) and P-trait (AUC = 0.772). In addition, the KSID (AUC = 0.555) performed significantly worse than the reference category.

Further, a significant moderating effect was found for the risk assessment method. The mean effect size of actuarial tools (AUC = 0.657) was higher than the mean effect size of structured clinical judgment tools (AUC = 0.580). No significant difference was found between the mean effect size of actuarial tools and actuarial tools with an override option. No significant moderating effect was found for instrument length (i.e., number of items).

3.2.2. Study design characteristics

A significant moderating effect was found for the focus of the instrument. The mean effect size of tools predicting the onset of domestic violence (AUC = 0.744) was higher than the mean effect size of tools predicting the recurrence of domestic violence (AUC = 0.643). Further, a significant moderating effect was found for the sample used for validation. Both the mean effect size of tools validated in a construction sample (AUC = 0.705) and the mean effect size of tools validated in a dependent validation sample (AUC = 0.691) were higher than the mean effect size of tools validated in an independent validation sample (AUC = 0.635). None of the other study design characteristics (i.e., study design, follow-up length, type of domestic violence that is predicted in follow-up measurements, source of follow-up, publication year, percentage of males in the sample, mean age of the sample, and percentage of cultural minorities in the sample) significantly moderated the overall predictive accuracy of risk assessment tools.

4. Discussion

This meta-analysis investigated the predictive validity of risk assessment tools for domestic violence, and whether and how this was influenced by instrument and/or study characteristics. Overall, a significant medium effect was found (AUC = 0.647), indicating a moderate predictive accuracy of risk assessment tools. Unfortunately, previous review studies on predicting domestic violence did not report an overall accuracy to which we can compare our currently found overall predictive accuracy. However, meta-analyses on risk assessment tools in (juvenile) justice settings show similar overall predictive accuracies. For example, Schwalbe (2007) found a mean AUC of 0.640 for juvenile justice risk assessment tools, and Fazel, Singh, Doll, and Grann (2012) found a mean AUC of 0.660 for juvenile and adult risk assessment tools.

Table 3
Results of moderator analyses (bivariate models).

Tested moderator variables	# Studies	# ES	Intercept/mean z (95% CI)	β_1(95% CI)	Mean AUC	F (df1, df2)[a]	p[b]	Level 2 variance	Level 3 variance
Overall effect	50	205	0.263 (0.231, 0.296)		0.647				
Instrument characteristics									
Assessment approach						1.119 (3, 201)	0.342	0.008***	0.008***
Domestic violence instruments (RC)	45	166	0.264 (0.231, 0.297)***		0.647				
General or violence instruments	8	14	0.248 (0.174, 0.322)***	−0.016 (−0.088, 0.057)	0.638				
Victim's risk ratings	7	15	0.246 (0.175, 0.316)***	−0.018 (−0.087, 0.050)	0.637				
Instrument for psychopathic traits	5	10	0.332 (0.247, 0.418)***	0.068 (−0.014, 0.150)	0.684				
Specific instrument						2.945 (23, 181)	< 0.001***	0.006***	0.008***
General or violence instruments (RC)	8	14	0.261 (0.191; 0.331)***		0.645				
DVSI/DVSI-R	8	37	0.198 (0.141; 0.256)***	−0.063 (−0.141; 0.016)	0.611				
B-SAFER	5	9	0.173 (0.058; 0.288)***	−0.088 (−0.222; 0.047)	0.597				
CRAT-P	1	1	0.481 (0.241; 0.720)***	0.220 (−0.030; 0.469)+	0.760				
CTS2	1	1	0.254 (−0.109; 0.617)	−0.007 (−0.376;0.362)	0.642				
DA	10	23	0.282 (0.224; 0.340)***	0.021 (−0.060; 0.102)	0.657				
DV-MOSAIC	1	4	0.080 (−0.023; 0.182)	−0.181 (0.299; −0.063)**	0.545				
DVRAG	3	6	0.391 (0.294; 0.487)***	0.130 (0.024; 0.235)*	0.715				
FVIR	1	3	0.081 (−0.064; 0.226)	−0.180 (−0.335; −0.025)*	0.546				
IPPI-RAT	1	1	0.522 (0.236; 0.808)***	0.261 (−0.034; 0.555)+	0.780				
ODARA/R-ODARA	10	22	0.343 (0.278; 0.409)***	0.082 (−0.001; 0.166)+	0.690				
PCL-R	4	7	0.327 (0.239; 0.415)***	0.066 (−0.033; 0.165)	0.681				
PVS	1	2	0.415 (0.164; 0.667)**	0.154 (−0.107; 0.415)	0.727				
RAST	1	2	0.297 (0.088; 0.507)**	0.036 (−0.184; 0.257)	0.665				
SARA	10	21	0.256 (0.193; 0.318)***	−0.005 (−0.088; 0.077)	0.643				
DVSR	2	3	0.267 (0.131; 0.404)***	0.006 (−0.143; 0.155)	0.649				
KSID	2	5	0.097 (0.003; 0.191)*	−0.164 (−0.275; −0.053)**	0.555				
P-Trait	1	3	0.506 (0.336; 0.676)***	0.245 (0.066; 0.424)**	0.772				
PAPS	1	5	0.325 (0.108; 0.542)**	0.064–0.164; 0.291)	0.680				
SRA-PA	1	1	0.220 (0.007; 0.434)*	−0.041 (−0.254; 0.173)	0.623				
SVRA-I	1	8	0.273 (0.080; 0.466)**	0.012 (−0.193; 0.217)	0.652				
SIVIPAS	1	1	0.263 (0.023; 0.502)	0.002 (−0.248; 0.251)	0.647				
Victim's rating	7	15	0.203 (0.133; 0.273)***	−0.058 (−0.151; 0.034)	0.614				
No name	4	11	0.271 (0.166; 0.376)***	0.010 (−0.116; 0.136)	0.651				
Assessment method						4.006 (2, 187)	0.020*	0.008***	0.008***
Actuarial (RC)	43	163	0.282 (0.247, 0.318)***		0.657				
Structured clinical judgment	7	15	0.142 (0.049, 0.235)**	−0.140 (−0.238, −0.042)**	0.580				
Actuarial with override	2	12	0.273 (0.186, 0.361)***	−0.009 (−0.093, 0.075)	0.652				
Number of items	49	203	0.263 (0.230, 0.295)***	−0.001 (−0.002, 0.001)	–	0.772 (1, 201)	0.381	0.008***	0.008***
Study design characteristics									
Focus of the assessment						4.286 (1, 203)	0.040*	0.008***	0.007***
Recurrence (perpetrators of DV) (RC)	48	202	0.257 (0.225, 0.289)***		0.643				
Occurrence (high risk group/general population)	2	3	0.448 (0.269, 0.627)***	0.191 (0.009, 0.373)*	0.744				
Study design						0.297 (1, 203)	0.586	0.008***	0.008***
Prospective (RC)	33	127	0.257 (0.216, 0.297)***		0.643				
Retrospective	17	78	0.275 (0.221, 0.330)***	0.019 (−0.049, 0.087)	0.653				
Sample used for validation						8.220 (2, 202)	< 0.001***	0.007***	0.008***
Validation (independent) (RC)	42	173	0.242 (0.207, 0.276)***		0.635				
Validation (dependent)	4	14	0.346 (0.275, 0.418)***	0.105 (0.036, 0.173)**	0.691				
Construction	9	18	0.373 (0.299, 0.447)***	0.131 (0.054, 0.209)**	0.705				
Follow-up (in months)	47	198	0.263 (0.228, 0.298)***	−0.000 (−0.002, 0.001)	–	0.407 (1, 196)	0.524	0.008***	0.008***
Type of follow-up (type of domestic violence predicted)						1.698 (4, 200)	0.152	0.008***	0.008***
General/combined (RC)	36	115	0.263 (0.226, 0.300)***		0.647				
Physical and sexual	16	42	0.243 (0.192, 0.293)***	−0.020 (−0.073, 0.033)	0.636				
Non-physical	8	10	0.252 (0.164, 0.340)***	−0.010 (−0.100, 0.079)	0.641				
Severe/near fatal	12	34	0.290 (0.236, 0.344)***	0.027 (−0.029, 0.083)	0.661				

(*continued on next page*)

Table 3 (*continued*)

Tested moderator variables	# Studies	# ES	Intercept/mean z (95% CI)	β_1(95% CI)	Mean AUC	F (df1, df2)[a]	p[b]	Level 2 variance	Level 3 variance
Violation of restraining order	2	4	0.359 (0.253, 0.465)***	0.096 (−0.004, 0.197)+	0.698				
Source of follow-up measurement						1.644 (2, 202)	0.196	0.008***	0.008***
Official records (RC)	33	132	0.249 (0.212, 0.286)***		0.639				
Partner reports	18	60	0.296 (0.248, 0.345)***	0.048 (−0.004, 0.100)+	0.664				
Self-reports/other	4	13	0.248 (0.170, 0.326)***	−0.001 (−0.075, 0.073)	0.638				
Publication year	50	205	0.263 (0.230, 0.295)***	0.003 (−0.004, 0.009)	–	0.640 (1, 203)	0.425	0.008***	0.008***
Percentage males in samples	40	179	0.256 (0.221, 0.292)***	−0.057 (−0.241, 0.126)	–	0.384 (1, 177)	0.536	0.007***	0.008***
Mean age of sample	31	183	0.261 (0.221, 0.302)***	−0.004 (−0.018, 0.010)	–	0.329 (1, 136)	0.567	0.005***	0.009***
Percentage minorities in samples	28	99	0.271 (0.225, 0.316)***	−0.004 (−0.013, 0.005)	–	0.778 (1, 97)	0.380	0.006***	0.009***

Note. # Studies = number of studies # ES = number of effect sizes; mean z = mean effect size (z); CI = confidence interval; RC = reference category; β1 = estimated regression coefficient; mean AUC = mean effect size expressed in an AUC value; AUC = area under the ROC curve; df = degrees of freedom; Level 2 variance = variance between effect sizes extracted from the same study; Level 3 variance = variance between studies.
[a] Omnibus test of all regression coefficients in the model.
[b] p-Value of the omnibus test.
+ $p < .1$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

In their meta-analysis, Van der Put et al. (2017) also found a medium effect size for risk assessment instruments used to predict child maltreatment (AUC = 0.681). The results of the trim-and-fill-analysis suggested that bias was present in the current dataset, and therefore a “corrected” overall effect was estimated, resulting in an AUC value of 0.599. Because there are several methodological shortcomings regarding the trim-and-fill method (see limitations section), this adjusted AUC value should not be interpreted as the true effect size, but only as an indicator of (possible) bias in the effect sizes that were synthesized.

Moderator analyses revealed a number of significant moderators. In line with our expectations, we found a higher mean effect size for actuarial tools (AUC = 0.657) than for structured clinical tools (AUC = 0.580). In other words, we found that actuarial tools had a better discriminative accuracy than clinical methods. The previous meta-analysis on risk assessment tools for domestic violence (Hanson e.a., 2007) also found a lower predictive accuracy for structured clinical tools than for actuarial tools and victim ratings, however this difference was not significant. Meta-analyses on the performance of risk assessment tools in other disciplines, such as criminal justice, forensic mental health, and clinical psychology, also found that actuarial methods outperform clinical methods (Aegisdottir et al., 2006; Dawes et al., 1989; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009; Hilton et al., 2006). There are two explanations for this finding. First, the mathematical features of actuarial methods ensure not only that solely variables with predictive value are part of an actuarial risk assessment instrument, but also that these variables are weighted in accordance with their independent contribution to the outcome of interest (Dawes et al., 1989). Earlier studies showed that it is difficult for professionals to accurately predict an outcome of interest using their clinical judgment, because professionals are unable to focus on the most important factors nor to properly weigh the observed risk factors (Dawes et al., 1989). Second, the reliability of actuarial tools is higher than that of clinical methods, and hence the actuarial prediction is more consistent and accurate (e.g., Dawes et al., 1989; Gambrill & Shlonsky, 2000). This can be explained by the fact that risk factors in actuarial prediction are scored according to a fixed algorithm, meaning that professionals use the same objective scoring rules, regardless of the expertise of the professional. On the other hand, scoring risk factors in clinical methods is done subjectively (e.g., Dawes et al., 1989; Gambrill & Shlonsky, 2000). In contrast to our expectations, no significant difference was found in predictive accuracy between actuarial tools (AUC = 0.657) and actuarial tools with an override option (AUC = 0.652).

We did not find a significant difference in predictive accuracy between different approaches to risk assessment, which is in line with previous research findings (Hanson et al., 2007). Tools that are specifically designed for predicting domestic violence (AUC = 0.647) did not perform significantly better than risk assessments based on victim ratings (AUC = 0.637), tools designed for general/violent criminal recidivism (AUC = 0.638), and tools developed for screening psychopathology (AUC = 0.684). Hanson and colleagues concluded in their meta-analysis that the absence of evidence concerning the superiority of any of these approaches is likely due to limited available research. Since the publication of their meta-analysis, a lot of new studies had been published on the predictive validity of domestic violence tools. In fact, the current meta-analysis included 50 studies, whereas only 18 studies were included in the former. Moreover, contrary to Hanson and colleagues who extracted one effect size per study, we performed a three-level meta-analysis making it possible to extract multiple effect sizes from the same study, which accordingly increases the statistical power in the analysis and the precision of estimated effects.

Although we found no significant difference in discriminative accuracy between the different approaches to risk assessment, it cannot be stated that these approaches are interchangeable in clinical practice. At group level, the different assessment approaches perform equally well, but at the individual case level, one of the approaches may outperform the other. General risk assessment tools consist of risk factors that are relevant for predicting both domestic violence and general (violent) delinquency, such as substance abuse, unemployment, antisocial personality, and anger and hostility, whereas specialized tools mainly consist of risk factors that are specifically relevant for predicting domestic violence, such as negative attitudes towards women, instability of partner relations, and barriers that keep victims away from help or care (Dutton & Kropp, 2000). If a general risk assessment tool instead of a specialized risk assessment tool is used in cases with one or more of these specific risk factors, individual clients may be wrongly assessed as a low-risk offender. To avoid cases with a low criminogenic risk, but a high domestic violence risk "falling through the cracks", we recommend using both types of risk assessment instruments in clinical practice.

Further, and confirming our expectations, we found higher mean effect sizes in construction samples (AUC = 0.705) and in dependent validation samples (AUC = ,691) than in independent validation samples (AUC = 0.635). This result is in line with the notion that predictive validity estimates reported for construction samples are commonly inflated, as "overfitting" data is a common problem for models built and tested in construction samples (see also the Introduction). An important limitation of the literature is that only very few tools have been validated in multiple independent samples and therefore, the reliability of predictive validity estimates is generally unknown. It may be possible that high estimates of the predictive value of risk assessment tools decline when multiple estimates are averaged over independent samples. Our result showed that cross-validated tools are vulnerable to random sampling error when construction and validation samples are randomly selected from the same sample.

Last, a significantly larger effect size was found for tools predicting the onset of domestic violence (AUC = 0.744) compared to tools predicting the recurrence of domestic violence (AUC = 0.643). A possible explanation for this finding is that predictive models benefit from greater variation in the prevalence of risk factors in general populations. To our knowledge, this variable has not been previously meta-analytically examined for instruments predicting domestic violence. We did not find a moderating effect of the length of the instrument, as the predictive validity of risk assessment tools did not vary by the number of items included in the risk assessment tools. So, tools with a relatively small number of items were equally accurate in predicting domestic violence as tools of longer length. Further, the predictive validity was not dependent on the study design, even though a prospective design is considered to be superior to a retrospective design (Caldwell et al., 1988). The predictive validity of tools examined in prospective studies was lower (AUC = 0.643) than that of tools examined in retrospective studies (AUC = 0.653), but this difference was not significant.

4.1. Clinical implications

Several implications for clinical practice can be derived from our results. Overall, a medium significant effect was found, indicating a moderate predictive accuracy of risk assessment tools for predicting domestic violence. This result shows that it is important to use risk assessment tools, especially because *unstructured* clinical judgment is widely recognised to be flawed, due to lower transparency, reliability and predictive validity (see, for example, Dorsey, Mustillo, Farmer, & Elbogen, 2008; Munro, 1999; Van der Put et al., 2017). Moreover, it is important to select and implement an instrument that showed at least moderate predictive accuracy and has been validated in at least two studies, which applies to the instruments DA, DVRAG, ODARA, PCL-R, SARA, and DVSR.

Furthermore, our results show that actuarial tools are preferable to clinical tools, because actuarial tools make a better distinction between high-risk and low-risk cases. However, bringing risk assessment of domestic violence to a higher level requires an improvement of actuarial tools. Actuarial tools in their current form are limited in their ability to guide case planning, because they do not always identify the full range of risk factors necessary for effective intervention planning (Schwalbe, 2008; Shlonsky & Wagner, 2005). Most actuarial tools that are currently being used are brief tools derived from multivariate statistical techniques, that mainly assess static risk factors. Therefore, these tools are particularly suitable for the purpose of risk assessment (predicting domestic violence to determine intervention urgency and intensity), but not for the purpose of needs assessment (identifying targets of interventions, bases on dynamic risk factors, in order to individualize case planning). Actuarial tools for domestic violence should therefore be further developed and strengthened by distinguishing between risk and needs assessment, and by integrating risk assessment with case management.

Moderator analyses revealed that the onset of domestic violence can be better predicted than the recurrence of domestic violence, which stresses the importance of early detection and prevention of domestic violence. Our review showed that the predictive accuracy of currently available screening tools (tools assessing the onset of domestic violence) is sufficient to justify their use in assessing risks for domestic violence in both high risk and general populations. Currently, the most commonly employed prognostic process is assessing the risk of *recurrence* of domestic violence. Given the relatively good performance of screening tools, it seems fruitful to invest time, money, and resources in developing and strengthening preventive strategies for domestic violence.

4.2. Limitations

Several limitations need to be discussed. First, the reliability of predictive validity estimates is generally unknown, because only very few tools have been validated in multiple independent samples. For empirically derived actuarial tools, ongoing replication studies are required to determine whether estimations of predictive validity are robust to random sampling variation. Second, studies were included regardless of their methodological quality in order to analyze a representative sample of the literature. To address this limitation, possible sources of within- and between-study heterogeneity were examined, including features of methodological quality, such as sample size, prospective or retrospective design, and length of follow-up. Although multiple potential moderating variables were examined, it is possible that other study design, sample, and instrument characteristics contribute to effect size variation, which we did not investigate. For example, the clinical background of the professionals who administered the tools was not examined, whereas previous research showed that this may be an important moderator of the predictive validity of risk assessment tools (Aegisdóttir et al., 2006). Studies generally do not report on such potentially important moderators. Third, there are several methodological difficulties regarding the trim and fill method. Nakagawa and Santos (2012) mentioned that this method has originally been designed for meta-analyses in which independence of effect sizes can be assumed. Further, the performance of the trim and fill method is limited when effect sizes prove to be heterogeneous (Peters, Sutton, Jones, Abrams, & Rushton, 2007; Terrin, Schmid, Lau, & Olkin, 2003), and moreover, the application of the trim and fill method could mean adding and adjusting for non-existent effect sizes in response to funnel plots that are asymmetrical, simply because of random variation (Egger, Davey-Smith, & Altman, 2001). Despite these shortcomings of the trim and fill method, there is no best method for detecting and handling missing data in meta-analysis, and therefore, the results of the trim and fill method should be interpreted with caution. In the present study, we only used the trim-and-fill method to calculate an adjusted overall effect, that was not interpreted as the true underlying effect. A final limitation is that the moderator analysis in which mean predictive accuracies of individual instruments were compared was based on a substantial number of categories, and a low number of studies and effect sizes. This may imply an inflated Type 2 error rate as well as insufficient statistical power for detecting true differences in predictive accuracy between instruments. The low number of effect sizes that was analyzed underlines the need for more research on the predictive validity of individual instruments.

5. Conclusion

The present meta-analysis aimed to examine the general predictive validity of risk assessment tools for domestic violence and to identify characteristics that influence this predictive validity. An important finding in this study was that the discriminative accuracy of actuarial tools outperforms the discriminative accuracy of structured clinical judgment tools. Because actuarial risk assessment tools are often optimized for predicting (domestic) violence, it is important that these tools are further developed into well performing instruments for violence prevention and risk management. One important way of improvement is to extend actuarial tools with a broad array of dynamic risk factors. Assessing these modifiable factors are essential in formulating proper clinical hypotheses, and in identifying targets for interventions with the aim to reduce the risk for (the recurrence of) domestic violence.

Conflict of interest

None of the authors of this review have any conflicts of interest to declare. Further, because this research did not involve human subjects, it did not require consent or assent forms that needed approval by an ethics committee.

References[1]

Aegisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382.

Alhabib, S., Nur, U., & Jones, R. (2010). Domestic violence against women: Systematic review of prevalence studies. *Journal of Family Violence, 25*(4), 369–382.

Arkins, B., Begley, C., & Higgins, A. (2016). Measures for screening for intimate partner violence: A systematic review. *Journal of Psychiatric and Mental Health Nursing, 23*(3–4), 217–235. https://doi.org/10.1111/jpm.12289.

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*, 154–174. https://doi.org/10.20982/tqmp.12.3.p154.

*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. *Behavioral Sciences & the Law, 30*(4), 420–430. https://doi.org/10.1002/bsl.2019.

*Belfrage, H., Strand, S., Storey, J. E., Gibas, A. L., Kropp, P. R., & Hart, S. D. (2011). Assessment and management of risk for intimate partner violence by police officers using the spousal assault risk assessment guide. *Law and Human Behavior, 36*, 60–67. https://doi.org/10.1037/h0093948.

*Bell, M. E., Cattaneo, L. B., Goodman, L. A., & Dutton, M. A. (2008). Assessing the risk of future psychological abuse: Predicting the accuracy of battered women's predictions. *Journal of Family Violence, 23*(2), 69–80. https://doi.org/10.1007/s10896-007-9128-5.

*Berk, R. A., He, Y., & Sorenson, S. B. (2005). Developing a practical forecasting screener for domestic violence incidents. *Evaluation Review, 29*(4), 358–383. https://doi.org/10.1177/0193841X05275333.

*Bourgon, G., & Bonta, J. (2004). Risk assessment for general assault and partner abusers (user report no. 2004-04). Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-ssssmnt-gnrl/rsk-ssssmnt-gnrl-eng.pdf.

Bowen, E. (2011). An overview of partner violence risk assessment and the potential role of female victim risk appraisals. *Aggression and Violent Behavior, 16*(3), 214–226. https://doi.org/10.1016/j.avb.2011.02.007.

Caldwell, R. A., Bogat, G. A., & Davidson, W. S., II (1988). The assessment of child abuse potential and the prevention of child abuse and neglect: A policy analysis. *American Journal of Community Psychology, 16*(5), 609–624.

*Callan-Bartkiw, U. (2012). Risk for intimate partner violence: An investigation of the psychometric properties of the spousal assault risk assessment guide in a New Zealand population (Doctoral dissertation). Retrieved from https://mro.massey.ac.nz/bitstream/handle/10179/4658/02_whole.pdf.

Campbell, J. C. (2002). Health consequences of intimate partner violence. *The Lancet, 359*(9314), 1331–1336.

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 (report no. 2013–05).* New Brunswick, Canada: Centre for Criminal Justice Studies.

*Cattaneo, L. B., Bell, M. E., Goodman, L. A., & Dutton, M. A. (2007). Intimate partner violence victims' accuracy in assessing their risk of re-abuse. *Journal of Family Violence, 22*, 429–440. https://doi.org/10.1007/s10896-007-9097-8.

*Cattaneo, L. B., & Goodman, L. A. (2003). Victim-reported risk factors for continued abusive behavior: Assessing the dangerousness of arrested batterers. *Journal of Community Psychology, 31*(4), 349–369. https://doi.org/10.1002/jcop.10056.

*Chan, K. L. (2014). Assessing the risk of intimate partner violence in the Chinese population the Chinese risk assessment tool for perpetrator (CRAT-P). *Violence Against Women, 20*(5), 500–516. https://doi.org/10.1177/1077801214535107.

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211–229. https://doi.org/10.1037/a0032968.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674.

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

*De Ruiter, C. (2011). Risicotaxatie bij relationeel geweldplegers in de praktijk: De B-SAFER [risk assessment for relational perpetrators in practice: The B-SAFER]. *GZ-psychologie, 3*(7), 24–33. https://doi.org/10.1007/s41480-011-0071-y.

De Vogel, V., De Ruiter, C., Hildebrand, M., Bos, B., & Van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3*(2), 149–165.

Dorsey, S., Mustillo, S. A., Farmer, E. M. Z., & Elbogen, E. (2008). Caseworker assessments of risk for recurrent maltreatment: Association with case-specific risk factors and re-reports. *Child Abuse & Neglect, 32*, 377–391.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29*(5), 617–658.

Dutton, D. G., & Kropp, P. R. (2000). A review of domestic violence risk tools. *Trauma, violence, & abuse, 1*(2), 171–181.

Duval, S., & Tweedie, R. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–99. https://doi.org/10.2307/2669529.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

*Echeburúa, E., Fernández-Montalvo, J., de Corral, P., & López-Go-i, J. J. (2009). Assessing risk markers in intimate partner femicide and severe violence: A new assessment instrument. *Journal of Interpersonal Violence, 24*(6), 925–939. https://doi.org/10.1177/0886260508319370.

Egger, M., Davey-Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629.

Fazel, S., Singh, J. P., 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. *BMJ, 345*, e4692.

*Ferguson, J. (2011). Risk assessment for domestic violence offenders: Predicting probation outcomes (Doctoral dissertation). Retrieved from https://repository.asu.edu/attachments/56467/content/Ferguson_asu_0010E_10382.pdf.

*Fitzgerald, R., & Graham, T. (2016). Assessing the risk of domestic violence recidivism. *Crime and Justice Bulletins, 189*, 1–12.

Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review, 22*, 813–837.

*Glass, N., Perrin, N., Hanson, G., Bloom, T., Gardner, E., & Campbell, J. C. (2008). Risk for reassault in abusive female same-sex relationships. *American Journal of Public Health, 98*(6), 1021–1027. https://doi.org/10.2105/AJPH.2007.117770.

*Gondolf, E. W., & Wernik, H. (2009). Clinician ratings of batterer treatment behaviors in predicting reassault. *Journal of Interpersonal Violence, 24*(11), 1792–1815. https://doi.org/10.1177/0886260508325493.

*Goodman, L. A., Dutton, M. A., & Bennett, L. (2000). Predicting repeat abuse among arrested batterers: Use of the danger assessment scale in the criminal justice system. *Journal of Interpersonal Violence, 15*(1), 63–74. https://doi.org/10.1177/088626000015001005.

*Grann, M., & Wedin, I. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime & Law, 8*, 5–23. https://doi.org/10.1080/10683160208401806.

*Gray, A. L. (2007). Assessing risk for intimate partner violence: A cross-validation of the ODARA and DVRAG within a sample of incarcerated offenders (Doctoral dissertation). Retrieved from https://curve.carleton.ca/system/files/etd/72e3aeb0-7550-4f38-9f47-d50f33431895/etd_pdf/6c6aacb2023c96d125559e194b77f097/gray-assessingriskforintimatepartnerviolenceacrossvalidation.pdf.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal(mechanical, algorithmic) prediction procedures. *Psychology, Public Policy, and Law, 2*(2), 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19.

Hanson, R. K., Bourgon, G., & Helmus, L. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis.* Ottawa, Ontario: Public Safety Canada.

Hanson, R. K., Helmus, L., & Bourgon, G. (2007). *The validity of risk assessments for intimate partner violence: A meta-analysis.* Ottawa, Ontario: Public Safety Canada (pp. 2007-07).

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological*

[1] References marked with an asterisk (*) were included in the meta-analysis.

Assessment, 21, 1–21.
*Heckert, D. A., & Gondolf, E. W. (2004). Battered women's perceptions of risk versus risk factors and instruments in predicting repeat reassault. *Journal of Interpersonal Violence, 19*(7), 778–800. https://doi.org/10.1177/0886260504265619.
Helmus, L., & Bourgon, G. (2011). Taking stock of 15 years of research on the Spousal Assault Risk Assessment Guide (SARA): A critical review. *International Journal of Forensic Mental Health, 10*(1), 64–75. https://doi.org/10.1080/14999013.2010.551709.
*Hendricks, B., Werner, T., Shipway, L., & Turinetti, G. J. (2006). Recidivism among spousal abusers: Predictions and program evaluation. *Journal of Interpersonal Violence, 21*(6), 703–716. https://doi.org/10.1177/0886260506287310.
Hilton, N. Z., & Harris, G. T. (2005). Predicting wife assault: A critical review and implications for policy and practice. *Trauma, Violence, & Abuse, 6*(1), 3–23.
Hilton, N. Z., Harris, G. T., & Rice, M. E. (2006). Sixty-six years of research on the clinical versus actuarial prediction of violence. *The Counseling Psychologist, 34*(3), 400–409.
*Hilton, N. Z., & Harris, G. T. (2009). How nonrecidivism affects predictive accuracy: Evidence from a cross-validation of the Ontario Domestic Assault Risk Assessment (ODARA). *Journal of Interpersonal Violence, 24*(2), 326–337. https://doi.org/10.1177/0886260508316478.
*Hilton, N. Z., Harris, G. T., Popham, S., & Lang, C. (2010). Risk assessment among incarcerated male domestic violence offenders. *Criminal Justice and Behavior, 37*(8), 815–832. https://doi.org/10.1177/0093854810368937.
*Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An indepth actuarial assessment for wife assault recidivism: The domestic violence risk appraisal guide. *Law and Human Behavior, 32*(2), 150–163. https://doi.org/10.1007/s10979-007-9088-6.
*Hilton, N. Z., Harris, G. T., Rice, M. E., Lang, C., Cormier, C. A., & Lines, K. J. (2004). A brief actuarial assessment for the prediction of wife assault recidivism: The Ontario Domestic Assault Risk Assessment. *Psychological Assessment, 16*(3), 267–275. https://doi.org/10.1037/1040-3590.16.3.267.
*Hisashima, J. (2008). Validation study of the Domestic Violence Screening Instrument (DVSI). Retrieved from http://icis.hawaii.gov/wp-content/uploads/2013/07/DVSI-Validation-2003-2007-Jan-2008.pdf.
Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141*, 901–930. https://doi.org/10.1037/a0038822.
*Houry, D., Feldhaus, K., Peery, B., Abbott, J., Lowenstein, S. R., Al-Bataa-De-Montero, S., et al. (2004). A positive domestic violence screen predicts future domestic violence. *Journal of Interpersonal Violence, 19*(9), 955–966. https://doi.org/10.1177/0886260504267999.
*Jung, S., & Buro, K. (2017). Appraising risk for intimate partner violence in a police context. *Criminal Justice and Behavior, 44*(2), 240–260. https://doi.org/10.1177/0093854816667974.
Kilvinger, F., Rossegger, A., Urbaniok, F., & Endrass, J. (2012). Risk assessment for domestic violence. *Fortschritte der Neurologie Psychiatrie, 80*(6), 312–319. https://doi.org/10.1055/s-0031-1273200.
Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17), 2693–2710. https://doi.org/10.1002/sim.1482.
*Koziol-McLain, J., Coates, C. J., & Lowenstein, S. R. (2001). Predictive validity of a screen for partner violence against women. *American Journal of Preventive Medicine, 21*(2), 93–100. https://doi.org/10.1016/S0749-3797(01)00325-7.
Kropp, P. R. (2008). Intimate partner violence risk assessment and management. *Violence and Victims, 23*(2), 202–220.
*Kropp, P. R., & Hart, S. D. (2000). The Spousal Assault Risk Assessment (SARA) Guide: Reliability and validity in adult male offenders. *Law and Human Behavior, 24*(1), 101–118. https://doi.org/10.1023/A:1005430904495.
*Lauria, I., McEwan, T. E., Luebbers, S., Simmons, M., & Ogloff, J. R. (2017). Evaluating the Ontario domestic assault risk assessment in an Australian frontline police setting. *Criminal Justice and Behavior, 44*(12), 1545–1558. https://doi.org/10.1177/0093854817738280.
Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousands Oaks, CA: Sage.
*Loinaz, I. (2014). Typologies, risk and recidivism in partner-violent men with the B-SAFER: A pilot study. *Psychology Crime & Law, 20*(2), 183–198. https://doi.org/10.1080/1068316X.2013.770854.
*Mason, R. L., & Julian, R. D. (2009). *Analysis of the Tasmania Police Risk Assessment Screening Tool (RAST) (report no. 2009–11).* Sandy Bay, Australia: Tasmanian Institute of Law Enforcement Studies, University of Tasmania.
Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence, 28*(7), 1537–1558. https://doi.org/10.1177/0886260512468250.
*Messing, J. T., Amanor-Boadu, Y., Cavanaugh, C. E., Glass, N., & Campbell, J. C. (2013). Culturally competent intimate partner violence risk assessment: Adapting the danger assessment for immigrant women. *Social Work Research, 37*(3), 263–275. https://doi.org/10.1093/swr/svt019.
*Messing, J. T., Campbell, J., Sullivan Wilson, J., Brown, S., & Patchell, B. (2017). The lethality screen: The predictive validity of an intimate partner violence risk assessment for use by first responders. *Journal of Interpersonal Violence, 32*(2), 205–226. https://doi.org/10.1177/0886260515585540.
*Messing, J. T., Campbell, J. C., & Snider, C. (2017). Validation and adaptation of the Danger Assessment-5: A brief intimate partner violence risk assessment. *Journal of Advanced Nursing, 73*(12), 3220–3230. https://doi.org/10.1111/jan.13459.
*Moser, A. E., Campbell, M. A., & Campus, S. J. (2012). Validation and expansion of the Ontario Domestic Assault Risk Assessment (ODARA) instrument: An early warning system (Doctoral dissertation). Retrieved from https://www.unb.ca/saintjohn/ccjs/_resources/pdf/odarapoliceresponse2012.pdf.
Munro, E. (1999). Common errors of reasoning in child protection work. *Child Abuse & Neglect, 23*, 745–758 Olver, M.E., Stochdale.
Murphy, C. M., Morrel, T. M., Elliott, J. D., & Neavins, T. M. (2003). A prognostic indicator scale for the treatment of partner abuse perpetrators. *Journal of Interpersonal Violence, 18*(9), 1087–1105. https://doi.org/10.1177/0886260503254515.
Nakagawa, S., & Santos, E. S. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*(5), 1253–1274.
Nicholls, T. L., Pritchard, M. M., Reeves, K. A., & 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.
*Olver, M. E., & Jung, S. (2017). Incremental prediction of intimate partner violence: An examination of three risk measures. *Law and Human Behavior, 41*(5), 440–453. https://doi.org/10.1037/lhb0000251.
Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562. https://doi.org/10.1002/sim.2889.
R Core Team (2015). *R: A language and environment for statistical computing.* Vienna, Austria: Foundation for Statistical Computing. Retrieved from http://www.Rproject.org/.
Rabin, R. F., Jennings, J. M., Campbell, J. C., & Bair-Merritt, M. H. (2009). Intimate partner violence screening tools: A systematic review. *American Journal of Preventive Medicine, 36*(5), 439–445. https://doi.org/10.1016/j.amepre.2009.01.024.
*Rettenberger, M., & Eher, R. (2013). Actuarial risk assessment in sexually motivated intimate-partner violence. *Law and Human Behavior, 37*(2), 75–86. https://doi.org/10.1037/b0000001.
Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737–748. https://doi.org/10.1037/0022-006X.63.5.737.
Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7.
*Roehl, J., O'Sullivan, C., Roehl, J., Webster, D. W., & Campbell, J. C. (2005). *Intimate partner violence risk assessment validation study: The RAVE study (final report to the National Institute of Justice; NCJ 209731–209732).* Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.
Rosenthal, R. (1994). In H. Cooper,, & L. V. Hedges, (Eds.). *Parametric measures of effect size* (pp. 39). New York, NY: Sage.
Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*(2), 183.
Ruscio, J. (2008). A probability-based measure of effect size: Robustness to base rates and other factors. *Psychological Methods, 13*(1), 19–30. https://doi.org/10.1037/1082-989X.13.1.19.
Schwalbe, C. S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior, 31*(5), 449–462.
Schwalbe, C. S. (2008). Strengthening the integration of actuarial risk assessment with clinical judgment in an evidence based practice framework. *Children and Youth Services Review, 30*(12), 1458–1464.
Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review, 27*(4), 409–427.
Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law, 31*(1), 8–22.
*Snider, C., Webster, D., O'sullivan, C. S., & Campbell, J. (2009). Intimate partner violence: Development of a brief risk assessment for the emergency department. *Academic Emergency Medicine, 16*(11), 1208–1216. https://doi.org/10.1111/j.1553-2712.2009.00457.x.
*Stansfield, R., & Williams, K. R. (2014). Predicting family violence recidivism using the DVSI-R: Integrating survival analysis and perpetrator characteristics. *Criminal Justice and Behavior, 41*(2), 163–180. https://doi.org/10.1177/0093854813500776.
*Stith, S. M., Milner, J. S., Fleming, M., Robichaux, R. J., & Travis, W. J. (2016). Intimate partner physical injury risk assessment in a military sample. *Psychology of Violence, 6*(4), 529–541. https://doi.org/10.1037/a0039969.
Storey, J. E., Kropp, P. R., Hart, S. D., 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. *Criminal Justice and Behavior, 41*(2), 256–271. https://doi.org/10.1177/0093854813503960.
*Svalin, K., Mellgren, C., Levander, M. T., & Levander, S. (2018). Police employees' violence risk assessments: The predictive validity of the B-SAFER and the significance of protective actions. *International Journal of Law and Psychiatry, 56*, 71–79. https://doi.org/10.1016/j.ijlp.2017.09.001.
*Svalin, K., Mellgren, C., Torstensson Levander, M., & Levander, S. (2017). Assessing and managing risk for intimate partner violence: Police employees' use of the police screening tool for violent crimes in Scania. *Journal of Scandinavian Studies in Criminology and Crime Prevention, 18*(1), 84–92. https://doi.org/10.1080/14043858.2016.1260333.
Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1–26. https://doi.org/10.1111/1529-1006.001.
Tabachnik, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Allyn and Bacon.
Tapp, J., & Moore, E. (2016). Risk assessments for dating violence in mid to late adolescence and early adulthood. *Criminal Behaviour and Mental Health, 26*(4), 278–292. https://doi.org/10.1002/cbm.2013.
Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22*, 2113–2126. https://doi.org/10.1002/sim.

*Ulmer, J. C. (2015). The Ontario Domestic Assault Risk Assessment (ODARA): A validation and comparison study for an Oregonian law enforcement agency (Doctoral dissertation). Retrieved from http://digitalcommons.georgefox.edu/cgi/viewcontent.cgi?article=1176&context=psyd
Van den Noortgate, W., López-López, J. A., Marin-Martinez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*(2), 576–594. https://doi.org/10.3758/s13428-012-0261-6.
Van den Noortgate, W., López-López, J. A., Marin-Martinez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods, 47*(4), 1274–1294. https://doi.org/10.3758/s13428-014-0527-2.
Van der Put, C. E., Assink, M., & Boekhout van Solinge, N. F. B. (2017). Predicting child maltreatment: A meta-analysis of the predictive validity of risk assessment instruments. *Child Abuse & Neglect, 73*, 71–88.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the meta for package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03.
*Weisz, A. N., Tolman, R. M., & Saunders, D. G. (2000). Assessing the risk of severe domestic violence. *Journal of Interpersonal Violence, 15*(1), 75–90. https://doi.org/10.1177/088626000015001006.
*Williams, K. R. (2012). Family violence risk assessment: A predictive cross-validation study of the Domestic Violence Screening Instrument Revised (DVSI-R). *Law and Human Behavior, 36*(2), 120–129. https://doi.org/10.1037/h0093977.
*Williams, K. R., & Grant, S. R. (2006). Empirically examining the risk of intimate partner violence: The Revised Domestic Violence Screening Instrument (DVSI-R). *Public Health Reports, 121*(4), 400–408. https://doi.org/10.1177/003335490612100408.
*Williams, K. R., & Houghton, A. B. (2004). Assessing the risk of domestic violence reoffending: A validation study. *Law and Human Behavior, 28*(4), 437–455. https://doi.org/10.1023/B:LAHU.0000039334.59297.f0.
*Wong, T. (2008). Spousal Assault Risk Assessment (SARA) validation study, state of Hawaii, 2004-2007. Retrieved from http://icis.hawaii.gov/wp-content/uploads/2013/07/SARA-Validation-2004-2007-Jan-2008.pdf.
*Wong, T., & Hisashima, J. (2008). *Domestic violence exploratory study on the DVSI and SARA, State of Hawaii, 2003–2007 (ICIS Technical Report No. 1).* Hawaii: Hawaii State Department of Health, Interagency Council on Intermediate Sanctions.