Review Manuscript

The Validity and Reliability of Available Intimate Partner Homicide and Reassault Risk Assessment Tools: A Systematic Review

TRAUMA, VIOLENCE, & ABUSE
2021, Vol. 22(1) 18-40

DOI: 10.1177/1524838018821952
journals.sagepub.com/home/tva
SAGE

Laurie M. Graham[1], Kashika M. Sahay[2], Cynthia F. Rizo[1], Jill T. Messing[3], and Rebecca J. Macy[1]

Abstract

At least one in seven homicides around the world is perpetrated by intimate partners. The danger of intimate partner homicide (IPH) associated with intimate partner violence (IPV) has led to the development of numerous IPV reassault and IPH risk assessment tools. Using 18 electronic databases and research repositories, we conducted a systematic review of IPH or IPV reassault risk assessment instruments. After review, 43 studies reported in 42 articles met inclusion criteria. We systematically extracted, analyzed, and synthesized data on tools studied, sample details, data collection location, study design, analysis methods, validity, reliability, and feasibility of use. Findings indicate that researchers in eight countries have tested 18 distinct IPH or IPV reassault risk assessment tools. The tools are designed for various professionals including law enforcement, first responders, and social workers. Twenty-six studies focused on assessing the risk of male perpetrators, although eight included female perpetrators. Eighteen studies tested tools with people in mixed-sex relationships, though many studies did not explicitly report the gender of both the perpetrators and victims/survivors. The majority of studies were administered or coded by researchers rather than administered in real-world settings. Reliable and valid instruments that accurately and feasibly assess the risk of IPH and IPV reassault in community settings are necessary for improving public safety and reducing violent deaths. Although researchers have developed several instruments assessing different risk factors, systematic research on the feasibility of using these instruments in practice settings is lacking.

Keywords

predicting domestic violence, domestic violence, cultural contexts, homicide, assessment

Intimate partner violence (IPV) includes stalking, sexual violence, physical violence, the threat of physical or sexual violence, psychological aggression or coercion, and other exertions of abusive control over a partner within the context of a romantic relationship (Black et al., 2011). Of those who are victimized by an intimate partner, a portion of victims are at heightened risk of being repeatedly and severely abused or murdered by an intimate partner (Black et al., 2011; Stöckl et al., 2013). This systematic review aims to provide a thorough overview of research on the reliability, validity, and feasibility of the use of intimate partner homicide (IPH) and IPV reassault risk assessment tools, focusing specifically on evidence useful to practitioners.

Global estimates indicate that at least one in seven homicides is committed by an intimate partner; for women, this number is one in three (Stöckl et al., 2013). In the United States, approximately 51% of female homicide victims are killed by a current or former intimate partner, compared to about 8% of male homicide victims (Jack et al., 2018). Such numbers likely underestimate the prevalence of IPH, given that reports do not always record victim–perpetrator relationships, murders by ex-partners may not be captured, and data on deaths may be of poor quality or missing (Campbell, Messing, & Williams, 2017; Stöckl et al., 2013). Thus, accurate assessment of the risk for IPH and IPV reassault is critical for improving public safety.

The Assessment of Risk

The need to screen for IPH and IPV reassault has led to both a body of research on related risk factors and the emergence of

[1] School of Social Work, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[2] Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA
[3] School of Social Work, Arizona State University, Phoenix, AZ, USA

Corresponding Author:
Laurie M. Graham, School of Social Work, University of North Carolina, Tate-Turner-Kuralt Building, 325 Pittsboro Street, CB #3550, Chapel Hill, NC 27599, USA.
Email: grahamlm@email.unc.edu

numerous risk assessment tools (Campbell et al., 2003; Campbell, Glass, Sharps, Laughon, & Bloom, 2007; Messing & Thaller, 2013; Nicholls, Pritchard, Reeves, & Hilterman, 2013). In reviewing risk factors for IPH, Campbell, Glass, Sharps, Laughon, and Bloom (2007) identified a history of IPV as the most important risk factor with up to 75% of IPH victims abused by their partner prior to homicide (Campbell et al., 2003; Moracco, Runyan, & Butts, 1998; Pataki, 1997; Sharps, Campbell, Campbell, Gary, & Webster, 2003). IPH is often a culmination of escalating IPV (Juodis, Starzomski, Porter, & Woodworth, 2014b; Stöckl et al., 2013). Thus, given the relative difficulty of identifying homicide risk, current science aims to reduce the risk of both reassault and homicide (Messing & Thaller, 2015).

We use the phrase "IPV/IPH risk assessment tools" as a term for danger and lethality assessments focused on evaluating risk for future IPH and IPV reassault. Comprised of questions that assess risk factors for IPH and IPV reassault, IPV/IPH risk assessment tools are designed to assist domestic violence advocates, law enforcement officers, nurses, social workers, and first responders in identifying individuals at risk for ongoing danger and homicide in the context of intimate partnerships (Messing & Thaller, 2013). Common perpetration risk factor domains across tools include past violent, controlling, or threatening behavior; unemployment or recent life changes; a history of mental health concerns; drug/alcohol misuse; and prior use of a weapon. IPV/IPH risk assessment tools are intended to predict criminal recidivism, IPV reassault, severe reassault, or lethality (Messing & Thaller, 2015). Many of these risk assessment tools are also intended to identify individuals with the greatest need for intervention or to mitigate the risk that has been identified (Douglas & Kropp, 2002; Messing & Thaller, 2015).

The manner in which IPV/IPH risk assessment instruments are used to inform services varies. Scholars suggest that service professionals use these tools to assist IPV survivors with decisions about self-care and safety (e.g., Campbell, Webster, & Glass, 2009), determine who among perpetrators might be suitable for entering a batterers' treatment program (e.g., Morgan & Gilchrist, 2010), and engage in risk management (e.g., Douglas & Kropp, 2002). Communities are increasingly using collaborative interventions wherein IPV/IPH risk assessment tools are used at police-involved IPV incidents to identify risk, educate survivors about available services, and/or connect survivors to a local crisis response agency (Messing & Campbell, 2016).

Psychometric Properties

Reliability and validity are common psychometric properties used to assess the consistency and accuracy of measurement tools. Reliability is concerned with the consistency of scores when repeatedly and independently measuring the same person or phenomenon under the same circumstances. Reliability is often assessed in terms of internal consistency reliability and interrater reliability, depending on the measurement approach used to structure the instrument (i.e., latent variable vs. index) and the method used to administer or complete the instrument (e.g., self-report and observer). Internal consistency reliability measures the homogeneity of items within a scale, that is, how well a group of items perform together to measure an underlying latent variable (DeVellis, 2003). Nunnally and Bernstein (1994) suggest that Cronbach's α, the reliability coefficient typically used to measure the internal consistency of an instrument, should approach .90 for clinical or practice settings. Interrater reliability examines consistency among raters and is used when multiple raters independently complete an instrument for the same people under the same circumstances. Different strategies for calculating interrater reliability include percent agreement, interclass and intraclass correlation, Pearson r, Spearman ρ, and Cohen's κ. Benchmarks for acceptable interrater reliability vary based on the statistical approach used (DeVellis, 2003).

Validity is concerned with the accuracy of an instrument or how well the instrument measures what it is intended to measure. Different forms of validity include content, construct, and criterion validity. Content validity refers to the extent to which the items on an instrument reflect all major facets of the construct that the instrument is intended to measure (Carmines & Zeller, 1979). Construct validity is concerned with the degree to which scores on an instrument are correlated with the scores from measures of theoretically related and unrelated concepts (Cronbach & Meehl, 1955). Criterion-related validity examines an instrument against some external criterion generally accepted as another indicator of the construct being measured by the target instrument (DeVellis, 2003).

A specific form of criterion-related validity is predictive validity. This validity type is of relevance for IPV/IPH risk assessment instruments because it compares a participant's score on an instrument to some criterion measured in the future (e.g., homicide, IPV reassault). Predictive validity is often assessed in terms of sensitivity and specificity. In the context of IPV/IPH reassault risk assessment instruments, sensitivity refers to the correct classification of individuals who are expected to kill or reassault their intimate partners, whereas specificity refers to the correct classification of individuals who are not expected to kill or reassault their intimate partners. Both are calculated as a proportion with higher scores reflective of greater predictive validity (Douglas, Guy, Reeves, & Weir, 2008).

Another commonly used approach to assess predictive validity, particularly of risk assessment instruments, is the receiver operating characteristic (ROC; Douglas et al., 2008; Messing & Thaller, 2013). The ROC plots sensitivity against1-specificity as a curve, with the area under the curve (AUC) indicating the probability of prediction (Douglas et al., 2008). The AUC ranges from 0 to 1.0 with 0.5 reflecting an inability to predict and higher scores reflecting better positive predictive ability. Other approaches to assessing predictive validity include hazard ratio (HR) and negative predictive value/positive predictive value (NPV/PPV). Singh (2013) provides detailed descriptions of these measures of association and

metric tools (e.g., AUC, PPV, and NPV), suggesting that violence prediction tools should ideally be able to predict accurately (i.e., calibration) and discriminate between those who will/will not be violent (i.e., discrimination).

Current Study

IPV/IPH risk assessment tools that are valid, reliable, and feasible to use are essential for targeted prevention and intervention efforts (Messing & Campbell, 2016), both nationally and internationally (Messing & Thaller, 2013; Nicholls et al., 2013). Several reviews exist on the topic of IPV/IPH risk assessment (Dutton & Kropp, 2000; Farrell, 2011; Hanson, Bourgon, &Helmus, 2007; Kropp, 2008; Messing & Thaller, 2013; Nicholls et al., 2013), each of which differs from the current review in terms of the review aims, methods used to locate and synthesize research, and/or the range of years during which reviewed articles were published. Nicholls, Pritchard, Reeves, and Hilterman (2013) conducted the most comprehensive systematic review on this topic that we could locate. This review included quantitative, peer-reviewed articles (1990–2011) that reported findings on the reliability and/or validity of IPV risk assessment tools used in Western countries. In their review, Nicholls et al. (2013) conclude that IPV risk assessment tools vary considerably in terms of the quality of the studies that evaluate the tools as well as the tools' reliability and validity. They further conclude that the instruments have not been assessed in diverse settings and did not endorse any of the identified and reviewed instruments as a gold standard for assessing IPV risk.

Given the fast growing research on IPV/IPH risk assessment tools (Messing & Thaller, 2013), our systematic review builds on prior research by updating findings regarding the psychometric properties of the most recent versions of IPV/IPH risk assessment tools through May 2015. Our review also differs from Nicholls et al. (2013) because we include studies published in nonWestern countries, include studies about additional tools, and examine gray literature.

Moreover, this review extends prior research by providing context about the populations studied, such as victim/survivor, perpetrator, and relationship characteristics; immigration status; and victim/survivor age at time of tool administration. This review also highlights the conditions under which the tools were administered, including who administers or completes the assessment tool, how, and where, as well as the feasibility of use (i.e., how practical it might be to use the tool in a real-world setting).

This review was guided by the following research questions: (a) How have reliability, validity, and feasibility of use been tested for IPV/IPH risk assessment tools, and what are the related findings? and (b) In what settings, populations, and environments have IPV/IPH risk assessment tools been tested? In addition to guiding the selection and use of IPV/IPH risk assessment tools for research and practice, findings from such a review serve to identify methodological strengths and gaps in the current literature.

Method

We employed two complementary search strategies: (a) a systematic database search of scholarly and gray literature and (b) a backward literature search of the references of each study included in our review. After consulting with a research librarian, we first conducted a systematic search for studies that met our prespecified inclusion criteria in the following 10 databases: PsycINFO, CINAHL Plus with Full Text, Family & Society Studies Worldwide, Health Source: Nursing/Academic Edition, PsycTESTS, Social Work Abstracts, Sociological Abstracts, Social Services Abstracts, PAIS International, and Web of Science. We also systematically searched eight repositories for gray literature: ProQuest Dissertations and Theses, Open Grey, National Institute of Justice Abstracts, VAWnet/National Resource Center on Domestic Violence Publications, Centers for Disease Control and Prevention Intimate Partner Violence Publications, World Health Organization Violence Prevention Publications and Resources, Public Safety Canada, and WorldCat Dissertations. For documents that were not available in full text in the database or repository searched, the research team e-mailed document authors and associated organizations and/or requested the item via two different universities to exhaust all potential avenues for locating full-text documents identified in searches.

In each database and repository, we used the following search terms with all possible combinations: (a) domestic violence homicide OR IPH OR femicide; AND (b) danger assessment (DA) OR risk assessment OR lethality assessment. Our database searches yielded 2,578 potentially relevant documents. After removing duplicates, we were left with 1,241 documents for title review. Figure 1 provides a flowchart depicting the review process.

Documents were excluded from our review based on specific criteria set by our research team. Documents had to (a) be written in English; (b) be a peer-reviewed journal article, dissertation, or government/nongovernmental report published during or before December 2015; (c) analyze data (i.e., empirical studies); and (d) assess the reliability, validity, and/or feasibility of the use of the most current, full-length version of at least one publicly available IPV/IPH risk assessment instrument. Systematic reviews and nonempirical articles were excluded, as were studies that tested earlier versions of IPV/IPH risk assessment instruments, did not administer a specified instrument in its entirety, and/or tested instruments that were not publicly available. Studies evaluating any form of the Level of Service Inventory were excluded as this tool was not created specifically for the prediction of future IPV (Andrews, 1982).

Upon title review, one reviewer pared down the remaining reports to 771 documents. After abstract review, this same team member reduced this number to 220 reports for more extensive text review. This reviewer then scanned the full text of each remaining article to determine whether the article used a tool pertinent to our review, which left 95 documents for full-text review. Two researchers completed full-text review of these documents, which identified 40 articles that fully met our

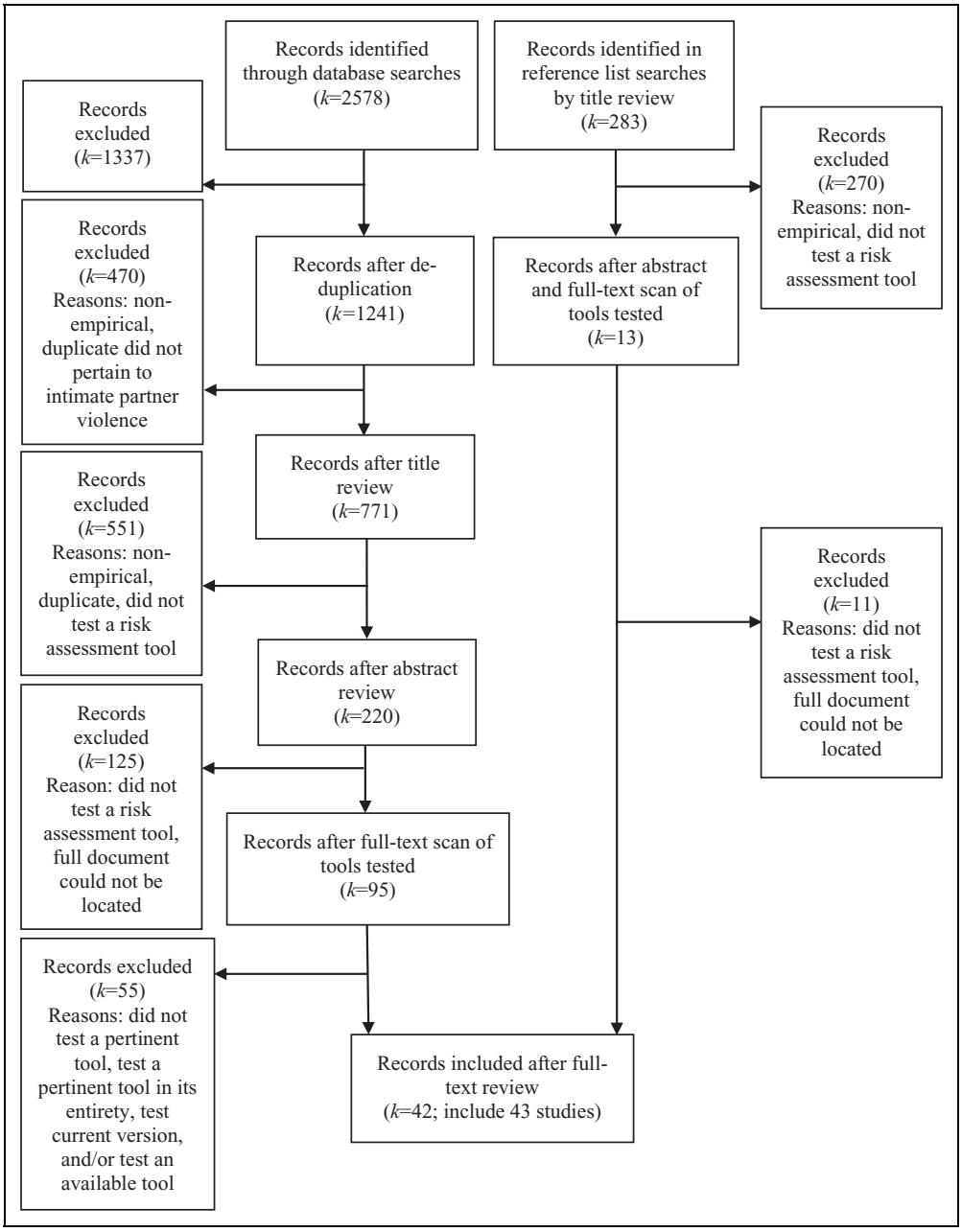

Figure 1. Flowchart of the systematic review search process.

inclusion criteria. For each of the 55 documents excluded during the full-text data review phase, coders found that the studies (a) did not test the reliability, validity, and/or feasibility of the use of an IPV/IPH risk assessment tool; (b) did not test a pertinent tool in its entirety; (c) did not test the most up-to-date version of the tool; (d) did not test a publicly available tool; or (e) could not be located. We then systematically searched the reference lists of the 40 documents that met our inclusion criteria, which led to two other relevant documents. Thus, 42 documents were included in our review, which discuss a total of 43 individual studies. Two reviewers systematically extracted data from all included studies ($k = 43$) and stored the information in a spreadsheet created prior to data extraction. Any discrepancies that arose between reviewers were resolved by reaching consensus between two or more research team members.

Results

This review includes all studies that met the inclusion criteria, regardless of whether multiple articles examined the same study data or included the same participants. This approach was taken because different articles based on the same data reported on different aspects of tool reliability and validity. A single article reports information from four studies (Dayan, Fox, & Morag, 2013), each of which had a different focus and

Table 1. Risk Assessment and Screening Tool Names and Abbreviations.

Tool Name	Abbreviation
Risk Assessment Tools Covered in this Review	
Brief Spousal Assault Form for the Evaluation of Risk	B-SAFER
Brief Spousal Assault Form for the Evaluation of Risk Version 2	B-SAFER v.2
Chinese Risk Assessment Tool for Perpetrators	CRAT-P
Chinese Risk Assessment Tool for Victims	CRAT-V
Danger Assessment-5 items	DA-5
Danger Assessment (20-item Revised Version)	DA
Danger Assessment-Immigrant	DA-I
Danger Assessment–Revised	DA-R
Domestic Violence Risk Appraisal Guide	DVRAG
Domestic Violence Screening Instrument–Revised	DVSI-R
Kingston Screening Instrument for Domestic Violence	K-SID
Lethality Screen	—
Ontario Domestic Assault Risk Assessment	ODARA
Risk Assessment Scale for Domestic Violence	RAS-DV
Severe Intimate Violence Partner Risk Prediction Scale	SIVIPAS
Spousal Assault Risk Assessment Guide	SARA
Spouse Violence Risk Assessment Inventory	SVRA-I
Secondary Risk Assessment for Partner Abusers	SRA-PA
Other Tools	
Conflict Tactics Scale Revised	CTS2
Domestic Violence Screening Instrument	DVSI
Domestic Violence Supplementary Report	DVSR
General Statistical Information on Recidivism	GSIR
Historical Clinical Risk Management-20	HCR-20
Historical part of the HCR-20	H-10
Interpersonal Behavior Survey	IBS
Hare Psychopathy Checklist–Revised	PCL-R
Hare Psychopathy Checklist: Screening Version	PCL-SV
Social Problem-Solving Inventory–Short Form	SPSI-SF
Violence Risk Appraisal Guide	VRAG

sample, and is counted independently throughout the review. Additionally, two studies had details reported in two separate documents (Study 1: Campbell, O'Sullivan, Roehl, & Webster, 2005a, 2005b; Study 2: Messing et al., 2014; Messing, Campbell, Sullivan Wilson, Brown, & Patchell, 2017). Language and categorization of types of reliability and validity reflects that used by study authors.

Risk Assessment Tools Tested

We identified 18 instruments used for IPV/IPH risk assessment (see Table 1). These tools each include between 5 and 35 items to determine the risk of IPV reassault and/or IPH and were designed for the use in various practice contexts such as clinical and criminal justice settings. The risk factors included in these scales overlap extensively with a focus on perpetrator history, criminal record, alcohol/drug use, employment status, and severity of violence perpetrated.

Specifically, this review included studies that tested the reliability, validity, and/or feasibility of use of the following 18 instruments: DA ($k = 8$), Ontario Domestic Assault Risk Assessment (ODARA; $k = 8$), Spousal Assault Risk Assessment Guide (SARA; $k = 6$), Domestic Violence Screening Instrument–Revised (DVSI-R; $k = 5$), Spouse Violence Risk Assessment Inventory (SVRA-I; $k = 4$), Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; $k = 2$), Brief Spousal Assault Form for the Evaluation of Risk Version 2 (B-SAFER v.2; $k = 2$), Domestic Violence Risk Appraisal Guide (DVRAG; $k = 2$), Chinese Risk Assessment Tool for Perpetrators (CRAT-P; $k = 1$), Chinese Risk Assessment Tool for Victims (CRAT-V; $k = 1$), Danger Assessment-5 items (DA-5; $k = 1$), Danger Assessment for Immigrant Women (DA-I; $k = 1$), Danger Assessment–Revised (DA-R; $k = 1$), Kingston Screening Instrument for Domestic Violence (K-SID; $k = 1$), Lethality Screen ($k = 1$), Risk Assessment Scale for Domestic Violence (RAS-DV; $k = 1$), Secondary Risk Assessment for Partner Abusers (SRA-PA; $k = 1$), and Severe Intimate Violence Partner Risk Prediction Scale (SIVIPAS; $k = 1$). Two versions of the B-SAFER are included because it was determined that each version could be used in different contexts, depending on the information available (e.g., access to victim information vs. no access). Overall, the ODARA and DA have been tested most frequently (see Table 2).

Table 2. Descriptive and Sample Characteristics of Reviewed Studies Assessing Intimate Partner Homicide and Intimate Partner Violence Reassault Risk.

Subcategory	Group	Count
Study information		
Study location	United States	16
	Canada	12
	Sweden	5
	Israel	4
	China	3
	Austria	1
	New Zealand	1
	Spain	1
Data collection strategies[a]	Case or record review	27
	Structured interview/ questionnaire	14
	Secondary data analysis	6
	Semistructured interview	3
	Field observation	1
	Focus groups	1
Types of reliability examined/reported[a]	Interrater reliability	16
	Internal consistency reliability	11
Types of validity examined/reported[a]	Predictive validity	29
	Concurrent validity	14
	Construct validity	11
Tool administration information		
Tool[a]	ODARA	8
	DA	8
	SARA	6
	DVSI-R	5
	SVRA-I	4
	B-SAFER	2

(continued)

Table 2. (continued)

Subcategory	Group	Count
	B-SAFER v.2	2
	DVRAG	2
	CRAT-P	1
	CRAT-V	1
	DA-5	1
	DA-I	1
	DA-R	1
	K-SID	1
	Lethality Screen	1
	RAS-DV	1
	SIVIPAS	1
	SRA-PA	1
Administered/coded by[b]	Researchers	21
	Law enforcement/officers of the court/corrections staff	13
	Allied health[c]	6
	Social/shelter/hotline workers	3
Language[a]	Not reported	24
	English	15
	Spanish	5
	Cantonese	2
	French	1
	German	1
	Mandarin	1
	Putonghua	1
	Swedish	1
Outcome assessed for predictive validity[a,d]	Domestic violence/wife assault/IPV recidivism or reassault	25
	Severe/near fatal IPV	6
	Any violent crime recidivism	3
	General crime recidivism	3
	Threatened physical or sexual IPV	1
Follow-up time for predictive validity[d]	Record review/no follow-up reported[e]	14
	1 month to 1 year	7
	More than 1 year	5
Sample characteristics		
Population focus[a,f]	Wife assault/IPV perpetrators	26
	IPV victims/survivors	14
	Family violence perpetrators[g]	2
	Both victims/survivors and perpetrators	1
Mixed-/same-sex partnerships[a,f]	Not reported/unclear[h]	24
	Mixed-sex	18
	Female–female	3
	Male–male	0
Age[f]	Participants ≥18 years old (adults) only	30
	Participants <18 years old (minors) only	0
	Both minors and adults	3
	Not reported	10
Sample size[f]	0–50	4
	51–100	8
	101–250	9
	251–500	5
	501–750	7
	751–1,000	1
	1,001–2,000	2
	2,001–3,000	4
	3,001[d]	3
Sex of perpetrators[f]	>90% male	24
	65–90% male	5
	>90% female	3
	Not reported/unclear	11
Diversity[f]	Race/ethnicity/nationality/immigration status reported	29

Note. Language used in this table mirrors language used by study authors. ODARA = Ontario Domestic Assault Risk Assessment; DA = Danger Assessment; SARA = Spousal Assault Risk Assessment Guide; DVSI-R = Domestic Violence Screening Instrument–Revised; SVRA-I = Spouse Violence Risk Assessment Inventory; B-SAFER = Brief Spousal Assault Form for the Evaluation of Risk; B-SAFER v.2 = Brief Spousal Assault Form for the Evaluation of Risk Version 2; DVRAG = Domestic Violence Risk Appraisal Guide; CRAT-P = Chinese Risk Assessment Tool for Perpetrators; CRAT-V = Chinese Risk Assessment Tool for Victims; DA-5 = Danger Assessment-5 items; DA-I = Danger Assessment for Immigrant Women; DA-R = Danger Assessment–Revised; K-SID = Kingston Screening Instrument for Domestic Violence; RAS-DV = Risk Assessment Scale for Domestic Violence; SIVIPAS = Severe Intimate Violence Partner Risk Prediction Scale; SRA-PA = Secondary Risk Assessment for Partner Abusers; IPH = intimate partner homicide; IPV = intimate partner violence.

[a]Studies could be in multiple categories; count column does not add to $N = 35$. [b]Some studies compared coding scores between different types of practitioners. Studies using secondary data analysis are classified according to the way in which the data used were originally collected. [c]Allied health included nurses, family relationship counselors, correctional staff, and mental health staff. [d]These counts only include studies that assessed predictive validity. [e]This category includes studies that used case/record review in which the length of time covered in cases/records was generally unclear and/or variable. [f]For studies including key informants, counts reflect information for victims/survivors or perpetrators administered the tool. [g]Family violence perpetrators committed violence against any member of their family, including by not exclusively an intimate partner. [h]Many studies did not explicitly report the sex of the partners of the study participants. In these articles, the presumption appeared to be that readers would assume the participants to be in mixed-sex partnerships with a male perpetrator and female victim.

Research Question 1: Reliability, Validity, and Feasibility

Studies examined a variety of forms of reliability and validity, but only one study (Cairns & Hoffart, 2009) focused on feasibility. Many of the studies examined more than one psychometric property of a tool. Given the purpose of IPV/IPH risk assessment tools, the most commonly assessed psychometric property was predictive validity. Twenty-nine studies assessed the focal tool's predictive validity in some manner, though the ways in which the findings were presented varied widely across studies. Other psychometric properties assessed included inter-rater reliability ($k = 16$), concurrent validity ($k = 14$), internal consistency reliability ($k = 11$), and construct validity ($k = 11$). Findings related to all above categories of reliability and validity are available for only five instruments: the DA, DVSI-R, K-SID, ODARA, and SARA (see Table 3).

Table 3. Critical Findings Related to Reliability and Validity by Risk Assessment Tool.

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
B-SAFER Au et al. (2008) Storey, Kropp, Hart, Belfrage, and Strand (2014)	Not reported	*Concurrent*: B-SAFER scores and CTS2 subscale scores (physical aggression and physical assault) were significantly correlated[d] *Construct*: 95% of cases correctly classified[d]	*Outcome*: Repeat encounter with law enforcement for IPV-related incident AUC = .70 (*SE* = .04; 95% CI [.64, .76])
B-SAFER v.2[a] Belfrage and Strand (2008) Storey and Strand (2013)	Not reported	*Concurrent*: Ratings for individual items were significantly related to police officers' overall judgments of risk for future IPV[d] *Construct*: Reported for specific victim vulnerability factors	Not reported
CRAT-P Chan (2014)	Not reported	Not reported	*Outcome*: IPV perpetration Optimal cut point: 7.0%; sensitivity = .61; specificity = .64; overall accuracy = .64; PPV = .12; NPV = .95; AUC = .76 (95% CI [.70, .81], $p < .001$) (note that numbers reported for validation sample)
CRAT-V Chan (2012)	Not reported	Not reported	*Optimal cutoff probability* = 6.5 Sensitivity = .64; specificity = .62 *Outcome*: IPV victimization Sensitivity = .74; specificity = .68; overall accuracy = .69; PPV = .16; NPV = .97; AUC = .75 (95% CI [.69, .81], $p < .01$) (note that numbers reported for validation sample)
DA-5 Snider, Webster, O'Sullivan, and Campbell (2009)	Not reported	Not reported	*Optimal cut point*: Three of five "yes" answers on items (i.e., severe/near fatal IPV) Sensitivity = .83 (95% CI [.71, .91]); specificity = .56 (95% CI [.51, .62]); PPV = .25; NPV = .95; AUC = .79 (95% CI [.73, .85])

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
DA Campbell, O'Sullivan, Roehl, and Webster (2005a, 2005b) Campbell, Webster, and Glass (2009) Juodis, Starzomski, Porter, and Woodworth (2014a) Messing, Amanor-Boadu, Cavanaugh, Glass, and Campbell (2013) Peterson (2013) Rollins et al. (2012) Sabri et al. (2013)	*Internal consistency:* • $\alpha = .66$ (appears to be unweighted) $\alpha = .72$ (unweighted) • $\alpha = .83$ (weighted) • $\alpha = .77$ (not specified if weighted or unweighted)	*Concurrent:* Significant correlation between baseline total score and frequency weighted CTS2 severe physical abuse score ($r = .45$, $p \leq .01$); CTS2 severe sexual abuse score ($r = .30$, $p \leq .01$); victim/survivor perceived risk of future abuse ($r = .25$–$.34$, $p \leq .01$) *Construct:* Significant correlation between total score and abuse perpetration (Pearson $r = .38$, $p \leq .01$)	*Optimal cut point*: Severe IPV = total score of 14–17 Sensitivity = .75 (range for different IPV danger categories: .55–.99); specificity = .86 Summary AUC = .69–.86 • AUC =.86 (95% CI [.81, .91], $p < .001$) • AUC = .69 *Outcome*: Any IPV AUC = .69 *Outcome*: Any sexual or physical IPV • Increased danger: sensitivity = .89–.98, specificity = .21, PPV = .28–.39 • Severe danger: sensitivity = .68–.78, specificity = .49–.52, PPV = .28–.43 • Extreme danger: sensitivity = .46–.54, specificity = .67–.68, PPV = .30–.44 • AUC = .61–.67 ($p \leq .001$) *Outcome*: Severe sexual or physical IPV • Increased danger: sensitivity = .92–.98, specificity = .20–.21, PPV = .26–.28 • Severe danger: sensitivity = .73–.78, specificity = .47–.52, PPV = .28–.29 • Extreme danger: sensitivity = .49–.54, specificity = .67, PPV = .30–.31 • AUC = .63–.69 ($p \leq .001$) Tests of risk categories ($\chi^2 = 49.99$, $df = 9$, $p < .001$) • Increased danger for severe abuse a$OR = 3.7$ ($p = .056$), minor/moderate abuse a$OR = 2.7$ ($p = .05$), stalking and threatening a$OR = 2.7$ ($p = .10$) • Severe danger for severe abuse a$OR = 11.0$ ($p < .001$), minor/moderate a$OR = 2.1$ ($p = .20$), stalking/threatening a$OR = 2.7$ ($p = .12$) • Extreme danger for severe abuse a$OR = 17.2$ ($p < .001$), minor/moderate a$OR = 12.7$ ($p = .004$), stalking/threatening a$OR = 6.5$ ($p = .002$) Prediction of stalking/threatening Wald = 15.14 ($p \leq .001$), minor/moderate physical or sexual abuse Wald = 9.58 ($p \leq .01$), severe abuse = 27.64 ($p \leq .001$) The number of DA risk factors endorsed (for domestic homicide perpetrators) was not significantly associated with perpetrator, victim, or offense characteristics (all p values > .05)

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
DA-I Messing et al. (2013)	Not reported	Not reported	*Outcome*: Severe IPV AUC = .85 (weighted score); AUC is significantly larger than that of the DA weighted score (χ^2[1, N = 148] = 15.40, p < .001) than the AUC of the weighted DA score *Outcome*: Any IPV AUC = .77 (weighted score); AUC is significantly larger than that of the DA-20R weighted score (χ^2[1, N = 148] = 5.17, p < .05)
DA-R Glass et al. (2008)	*Interrater*: Qualitative; independent review, debriefing, and peer review	Not reported	DA-R significantly predicted threatened or actual physical or sexual violence at 1 month in both weighted and unweighted versions of the scale (p < .01; 17 items scored; no further statistics provided by article authors)[d]
DVRAG Hilton, Harris, Rice, Houghton, and Eke (2008) Rettenberger and Eher (2013)	*Interrater*: Two independent, masked raters for 10 random cases; interclass correlation (single measure) ranged from .89 to .95	*Concurrent*: Moderate and highly significant association with VRAG, PCL-R, SARA, DA, DVSI, and CTS2	*Outcome*: Domestic recidivism Summary: AUC = .70–.71 • AUC = .71 (SE = .03), Cohen's d = .80; significant improvement over using only the ODARA (one-tailed test, p < .05)[d] • AUC = .70 (SE = .03, 95% CI [.64, .75]), Cohen's d = .75; significant improvement over using only the ODARA (one-tailed test, p < .05) • AUC = .70 (SE = .02, 95% CI [.66, .74]), p < .001, Cohen's d = .75. • AUC = .71 (95% CI [.58, .83], p < .05) *Outcome*: General violent recidivism AUC = .70 (95% CI [.57, .83], p < .05)[d] *Outcome*: General criminal recidivism AUC = .71 (95% CI [.59, .83], p < .01)[d]

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
DVSI-R Buchanan (2009) Gerstenberger and Williams (2013) Stansfield and Williams (2014) Williams (2012) Williams and Grant (2006)	*Internal consistency:* • $\alpha = .51$ • $\alpha = .75$ *Interrater*: • 80% agreement standard (no further statistics provided by article authors) • Significant correlation between ODARA completed with probation officer and ODARA coded by researcher: Pearson r [96] = .84 (p < .000)	*Concurrent*: Total score significantly correlated with ODARA (Pearson's r [96] = .62, p < .000) *Construct*: Principal components analysis supported factorial validity of the tool	*Outcome*: Recidivism (i.e., re-arrests for family violence offenses, protective order/restraining order violations, court order violations) AUC = .62–.73[b,e] *Outcome*: Repeat IPV AUC = .71 (95% CI [.70, .72])[e] *Outcome*: IPV recidivism[d] Optimal cut point: score = 7.5; AUC = .68 (*SE* = .06); sensitivity = .58; specificity = .34 Total risk score is a significant predictor of rearrest after controlling for other factors: HR = 1.56 (*SE* = .05)[e] AUC = .65 (95% CI [.64, .66])[e] Total risk score is a significant predictor of protective/restraining order violation: HR = 1.82 (*SE* = .09)[e] AUC = .69 (95% CI [.68, .70])[e] Higher total numeric DVSI-R was statistically significantly related to increased odds of IPV recidivism (*OR* = 1.09, *SE* = .01, p < .001) compared to lower total numeric DVSI-R scores[e] Point-biserial correlation: rpb [93] = .201, p = .05)[d]

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
K-SID Campbell et al. (2005a, 2005b)	Internal consistency: $\alpha = .04$	*Concurrent*: Significant correlation between baseline total score and frequency weighted CTS2 severe physical abuse score ($r = .29$, $p \leq .01$); CTS2 severe sexual abuse score ($r = .01$, $p \leq .01$); victim/survivor perceived risk of future abuse ($r = .14$–$.18$, $p \leq .01$) *Construct*: Significant correlation between total score and abuse perpetration (Pearson $r = .13$, $p \leq .01$)	*Outcome*: Any sexual or physical IPV • Moderate risk: sensitivity = .65–.66, specificity .42, PPV = .33–.53 • High risk: sensitivity = .30–.32, specificity = .75–.76, PPV = .37 • Very high risk: sensitivity = .26–.27, specificity = .78, PPV = .36 • AUC = .51–.55 (ns), .61 ($p \leq .001$) *Outcome*: Severe sexual or physical IPV • Moderate risk: sensitivity = .66–.67, specificity = .41, PPV = .19–.22 • High risk: sensitivity = .26–.28, specificity = .74, PPV = .19–.20 • Very high risk: sensitivity = .24–.25, specificity = .77, PPV = .18–.20 • AUC = .52–.54 (ns), .62 ($p \leq .01$) Tests of risk categories ($\chi^2 = 13.57$, $df = 9$, $p = .14$) • Medium risk a*OR* for severe abuse = 1.2 ($p = .68$), minor/moderate a*OR* = 1.2 ($p = .66$), stalking/threatening a*OR* = 1.0 ($p = .99$) • High risk for severe abuse a*OR* = 1.6 ($p = .60$), minor/moderate = 4.9 ($p = .06$), stalking/threatening = 2.0 ($p = .45$) • Very high risk for severe abuse a*OR* = 1.2 ($p = .62$), minor/moderate = 3.2 ($p = .007$), stalking/threatening = 2.2 ($p = .04$) Prediction of stalking/threatening Wald = .03 (ns), minor/moderate = 2.06 (ns), severe abuse = .34 (ns)
Lethality Screen Messing et al. (2014) Messing, Campbell, Sullivan Wilson, Brown, and Patchell (2017)[b]	Not reported	*Concurrent*: • Significantly related to DA *Construct:* • Study participant prediction of partner likely to seriously hurt them in the next year is significantly correlated with Lethality Screen scores	*Outcome*: Near-fatal IPV Cut point: "high" danger on lethality screen; sensitivity = .93; specificity = .21; NPV = .96; PPV = .13 *Outcome*: Severe violence Cut point: "high" danger on lethality screen; sensitivity = .93; specificity = .22; NPV = .93; PPV = .22 *Outcome*: Any IPV Sensitivity = .87; specificity = .22; NPV = .80; PPV = .32 *Outcome*: Intimate partner abuse Sensitivity = .84; specificity = .24; NPV = .48; PPV = .64

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
ODARA Buchanan (2009) Eke, Hilton, Harris, Rice, and Houghton (2011) Hilton and Harris (2009) Hilton, Harris, and Holder (2008) Hilton, Harris, Rice, et al. (2008) Hilton, Harris, Popham, and Lang (2010) Hilton et al. (2004) Hilton, Popham, Lang, and Harris (2014) Rettenberger and Eher (2013)	Internal consistency: • $\alpha = .48$ • $\alpha = .65$ *Interrater*: • Assessed for data transcription; high interrater reliability • For ODARA scoring: ICC = .90 (*SE* = .48) • For ODARA score with two independent raters for 10 cases, ICC = .94 (absolute agreement and single measures)	*Concurrent*: • Total score significantly correlated with DA, SARA, and DVSR; ($r = .43$–$.60$, $p < .01$) • No evidence of concurrent validity with CTS2, IBS, and SPSI-SF[d] • Total score significantly correlated with DVSI-R (Pearson's r [96] = .62, $p < .000$) *Construct*: • ODARA significantly associated with perpetrator's use of severe violence, sexual assault, 5-point injury scale, presence of potentially lethal acts (i.e., stabbing, slashing, strangulation), and prior medical treatment for assault ($r = .19$–$.26$, $p < .01$) • Wife assault recidivists had significantly higher ODARA scores than other assault recidivists Principal components analysis supported factorial validity of the tool	*Outcome*: IPV recidivism Cut point: score = 4; sensitivity = .59; specificity = .79; PPV = .54; NPV = .82 *ODARA total score and outcome:* • Domestic recidivism: AUC = .71 (95% CI [.58, .84], $p < .05$); AUC = .65 (*SE* = .03, 95% CI [.59, .71]), Cohen's $d = .55$[d] • Violent crime recidivism: AUC = .69 (95% CI [.57, .82], $p < .05$)[d] • General criminal recidivism: AUC = .66 (95% CI [.53, .79], $p < .05$)[d] Using Cox regression modeling over 8 years, ODARA score predicted domestic violence recidivism and any recidivism for full sample AUC = .64 (*SE* = .09; 95% CI [.54, .73])[d] Overall predictive accuracy for full sample: AUC = .67 (*SE* = .03; 95% CI [.61, .73]), Cohen's $d = 0.6$ Predictive validity based on different cut points on the ODARA scale: AUC = .77 (*SE* = .02; 95% CI [.73, .81]); Cohen's $d = 1.1$; PPV = .30–.72; NPV = .70–.96 ODARA score and IPV recidivism based on almost 9 years of follow-up (female offenders):[d] AUC = .72 (95% CI [.50, .94]); Cohen's $d = .92$ *Outcome:* IPV recidivism[d] Optimal cut point: score = 6.5; point-biserial correlation analyses: rpb [93] = .30 ($p = .003$); AUC = .72 (*SE* = .07); sensitivity = .68; specificity = .35
RAS-DV Gulati (2000)[d]	Internal consistency: $\alpha = .76$	Not reported	Not reported

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
SARA Belfrage et al. (2012) Callan-Bartkiw (2012) Grann and Wedin (2002) Kropp and Hart (2000) Rud, Skiling, and Nonemaker(2011) Wong and Hisashima (2008)	*Internal consistency* (total score): • Mean Inter-Item Correlation (MIC) = 0.1 • Cronbach's $\alpha = .65$ • Cronbach's $\alpha = .78$ *Interrater*: • For perpetrators with multiple police contacts ($n = 93$): ○ Total score stability between first and second contact: ICC = .76 ○ Summary risk rating stability between first and second contact: ICC = .45 • Total scores (with 2 raters on 18 cases): ICC = .85, F(2,18), $p < .01$ ○ Average Cohen's κ for all items = .58 • Interview and case history versus case history: ICC = 0.84 • Three methods in a single study: rater with interview notes ICC = .29, rater with audio recording ICC = .84, rater with interview notes and audio recording ICC = .36	*Concurrent*: • Total score significantly correlated with PCL: SV ($r = .43$, $p < .001$) but not with GSIR and VRAG[e] • Total score significantly correlated with PCL-R score, HCR-20, and VRAG ($r = .33$–$.59$, $p < .01$)[d] • Predicted recidivism within 1 year after release with other scores:[d] ○ PCL-R scores: AUC = .71 (95% CI [.60, .80]) ○ H-10: AUC = .68 (95% CI [.57, .78]) ○ VRAG: AUC = .75 (95% CI [.65, .84]) • Significantly correlated with DVSI (Pearson's $r = .54$, $p < .01$) and LSI-R (Pearson's $r = .43$, $p < .01$). Significantly correlated with DVSI (Spearman's $\rho = .67$, $p < .001$) *Construct*: • In comparing recidivist and nonrecidivist assaulters, recidivistic assaulters did not have statistically significant higher PCL-SV scores as compared to recidivists ($p = .07$),[b,e] • Recidivists' total score was on average 3.79 points higher than that of nonrecidivists ($p < .001$) • Significantly correlated with measures of domestic violence conviction and general criminal conviction over 8 years ($p < .001$) • Higher proportion of high-risk perpetrators rearrested (32.4%) compared to low-risk perpetrators (18.7%) • *Convergent and divergent validity*: Significantly correlated with measures of general violence ($r = .97$, $p < .01$) and IPV ($r = .96$, $p < .01$); summary risk rating $r = .63$ ($p < .01$); DVSI ($r = .74$, $p < .01$); nonsignificantly correlated with nonviolence CTS2 negotiation subscale ($r = .02$, ns), CTS2 physical assault subscale ($r = .19$, ns); CTS2 psychological aggression subscale ($r = .23$, ns); CTS2 sexual coercion subscale ($r = -.02$, ns); CTS2 injury subscale ($r = .29$, ns)	Cut point = 19 Sensitivity = .82; specificity = .50 Recidivism and total scores on the SARA correlation had a positive correlation (rpb = .18; $p < .001$) *Outcome*: IPV recidivism (i.e., having some additional IPV-related police contact) and SARA total score[d] • Statistically significant differences in SARA total score among recidivists and nonrecidivists, recidivist $M = 18$, $SD = 9.31$; nonrecidivist $M = 11.67$, $SD = 5.40$; $t(34) = -2.18$, $p < .05$; $\eta^2 = .12$ • AUC = .63 ($SE = .03$) • AUC = .72 (95% CI [.53, .90], $p < .05$) Recidivist partner violence at various time points with total score:[d] • AUC (6 months) = .52 (95% CI [.41, .63], ns) • AUC (1 year) = .59 (95% CI [.48, .70], ns) • AUC (2 years) = .63 (95% CI [.51, .73], $p < .01$) • AUC (5 years) = .65 (95% CI [.51, .77], $p < .01$) *Outcome*: Domestic violence recidivism (i.e., "abuse or assault of a household or family member, terroristic threatening, harassment, temporary restraining order [TRO] and/or a violation of a TRO, violation of a protective order; Wong & Hisashama, 2008, p. 3) Statistically significant relative risk probabilities: ○ "41% probability that [domestic violence]-related re-arrests (recidivism risk) will decline, if offenders are assessed downward from high to low risk levels" (Wong & Hisashama, 2008, p. 5) ○ AUC = .61 ($p < .05$) ○ "39% chance of making a classification error" (Wong & Hisashama, 2008, p. 6) *Outcome*: general recidivism (i.e., "probation revocation, general nonhousehold/nonfamily member felonies, property crimes, drug arrests, criminal contempt of court; Wong & Hisashama, 2008, p. 3) • "33% probability that general re-arrests will decline, as offenders are assessed downward from high to low risk levels" (Wong & Hisashama, 2008, p. 5) • AUC = .63 ($p < .01$)

(continued)

Table 3. (continued)

Tool and References	Reliability	Concurrent and Construct Validity	Predictive Validity
SIVIPAS Echeburúa, Fernández-Montalvo, de Corral, and López-Goñi (2009)	Internal consistency: • Total sample (n = 1,081): Cronbach's α = .71 • Severe perpetrators (n = 269): Cronbach's α = .69 • Control group (n = 812): Cronbach's α = .66	*Construct*: Found statistically significant differences between severe and nonsevere perpetrators on the total assessment score	Cut point: "High" category (score = 10; i.e., predicting severe IPV) Sensitivity = .48; specificity = .81
SRA-PA Bourgon and Bonta (2004)	Not reported	Not reported	*Outcomes*: Domestic violence-related recidivism (i.e., arrest) • Pearson r's between SRA-PA scores and each outcome were statistically significant • Pattern seen for male offenders and not female offenders like due to small sample of female offenders • For male offenders, statistically significant domestic violence recidivism rates at extreme risk levels were found: high-risk recidivism rate = 20.8%, medium risk = 12.1%, low risk = 9.3% *Outcome*: domestic violence recidivism AUC = .61 (95% CI [.54–.68])
SVRA-I Dayan, Fox, and Morag (2013)[c]	*Interrater*: Significant correlation between 3 independent raters for 19 cases (r = .68–.75, p < .01)	Concurrent: • Social workers' and clinical criminal professionals' risk assessment scores and SVRA-I significantly correlated (r = .45–.48, p < .01)[d] • Investigator's subjective risk scores and SVRA-I scores significantly correlated[d]	Not reported

Note. We report information available in each study, and all values are rounded to two decimal places. IPH = intimate partner homicide; IPV = intimate partner violence; ICC = intraclass correlation; *OR* = odds ratio; CI = confidence interval; HR = hazard ratio; AUC = area under the receiver operating characteristic curve; PPV = positive predictive value; NPV = negative predictive value; ns = nonsignificant; B-SAFER = Brief Spousal Assault Form for the Evaluation of Risk; CTS2 = Conflict Tactics Scale Revised; B-SAFER v.2 = Brief Spousal Assault Form for the Evaluation of Risk Version 2; CRAT-P = Chinese Risk Assessment Tool for Perpetrators; CRAT-V = Chinese Risk Assessment Tool for Victims; DA-5 = Danger Assessment-5 items; DA = Danger Assessment; DA-I = Danger Assessment for Immigrant Women; DA-R = Danger Assessment–Revised; DVRAG = Domestic Violence Risk Appraisal Guide; VRAG = Violence Risk Appraisal Guide; PCL-R = Psychopathy Checklist–Revised; SARA = Spousal Assault Risk Assessment Guide; DVSI = Domestic Violence Screening Instrument; ODARA = Ontario Domestic Assault Risk Assessment; DVSI-R = Domestic Violence Screening Instrument–Revised; K-SID = Kingston Screening Instrument for Domestic Violence; ODARA = Ontario Domestic Assault Risk Assessment; RAS-DV = Risk Assessment Scale for Domestic Violence; SARA = Spousal Assault Risk Assessment Guide; GSIR = General Statistical Information on Recidivism; PCL-SV = Psychopathy Checklist-Screen Version; SIVIPAS = Severe Intimate Violence Partner Risk Prediction Scale; SRA-PA = Secondary Risk Assessment for Partner Abusers; SVRA-I = Spouse Violence Risk Assessment Inventory.
[a]Both Versions 1 and 2 of the B-SAFER are included in this review because it was determined that each version could be used in difference contexts, depending on the information available (e.g., access to victim information vs. no access). [b]This article was made available in 2015 as an advance online publication. [c]This article includes four studies that focus on the SVRA-I. [d]A study with sample size ≤100 or validity finding from a study with sample size ≤100. [e]A study with sample size ≥2,000 or validity finding from a study with sample size ≥2,000.

Internal consistency reliability. When reported ($k = 11$), internal consistency was assessed using Cronbach's α, which ranged from .04 to .83. This information was reported for the DA ($\alpha = .66$–.83), DVSI-R ($\alpha = .51$ and .75), K-SID ($\alpha = .04$), ODARA ($\alpha = .48$ and .65), RAS-DV ($\alpha = .76$), SARA ($\alpha = .65$ and .78), and SIVIPAS (severe perpetrators $\alpha = .69$, control group $\alpha = .66$, and total sample $\alpha = .71$).

Interrater reliability. One study employed only qualitative processes for examining interrater reliability, involving multiple reviewers who independently coded instruments and narratively discussed the results. However, most studies ($k = 15$) used statistical approaches to calculate interrater reliability correlations, including intraclass correlation (ICC), Pearson's r, and tests of significance. Quantitative interrater reliability data were available for the DVRAG (ICC range = .89–.95), DVSI-R (at least .80 agreement and $r = .84$), ODARA (ICC range = .90–.94), SARA (ICC range = .29–.84), and SVRA-I (r range = .68–.75). Interrater reliability correlations of .75 or above are considered high.

Concurrent validity. Concurrent validity was reported for the following instruments ($k = 14$): B-SAFER, B-SAFER v.2, DA, DVRAG, DVSI-R, K-SID, Lethality Screen, ODARA, SARA, and SVRA-I. For the purposes of this review, we focused on the total score of the focal tool rather than subparts of the scale. The Revised Conflict Tactics Scale (CTS2; $k = 4$), DA ($k = 3$), and DVSI ($k = 3$), followed by the SARA ($k = 2$), Violence Risk Appraisal Guide ($k = 2$), and Psychopathy Checklist–Revised or Screen Version ($k = 2$ and $k = 1$, respectively), were most commonly used as a gold standard for assessing the concurrent validity of the focal tool. Additionally, the DVSI-R, Domestic Violence Supplementary Report (DVSR), Interpersonal Behavior Survey, Social Problem Solving Inventory–Short Form, General Statistical Information on Recidivism, Historical Clinical Risk Management-20, Level of Service Inventory–Revised, ODARA, victim/survivor perceptions of risk, and expert professional judgments were used as comparison tools ($k = 1$ for each). Seven studies compared a focal tool to multiple gold standards. For example, the DA was significantly associated with DVRAG, Lethality Screen, and ODARA scores. Whereas most focal tools were significantly related to the comparison gold standard tools, the ODARA and SARA evidenced mixed findings. For instance, the CTS2 was found to be significantly associated with B-SAFER and DVRAG scores, but the ODARA total score had no evidence of concurrent validity with the CTS2.

Construct validity. Information was available for the B-SAFER, B-SAFER v.2, DA, DVSI-R, K-SID, Lethality Screen, ODARA, SARA, and SIVIPAS ($k = 11$). Where possible, we focused on total scale scores rather than specific risk factors. Construct validity was reported in many ways. The most common strategy was comparing known groups (e.g., batterers vs. nonbatterers, recidivists vs. nonrecidivists, and severe vs. nonsevere perpetrators) on their risk assessment scores. Other strategies included comparing victim/survivor prediction of risk and perpetrators' abusive behaviors (e.g., assault, injury, and potentially lethal acts). The one study examining the construct validity of the B-SAFER v.2 did not examine the total assessment score and instead focused on specific vulnerability findings, and the one study examining DVSI-R and ODARA construct validity stated that the principal components analysis conducted supported this form of validity for these tools. Otherwise, almost all studies found that the focal instrument was (a) significantly related to victim/survivor prediction of risk, (b) significantly related to perpetrators' abusive behaviors, (c) significantly associated with known group membership, or (d) used to correctly classify a high percentage of cases (see Table 3).

Predictive validity. Predictive validity findings ($k = 29$) were available for 15 of the 18 tools: B-SAFER, CRAT-P, CRAT-V, DA-5, DA, DA-I, DA-R, DVRAG, DVSI-R, K-SID, Lethality Screen, ODARA, SARA, SIVIPAS, and SRA-PA. Wide variability existed in the outcomes assessed and, in some studies, more than one outcome was examined (see Table 2). These outcomes included (a) IPV-related recidivism or reassault (e.g., victimization, perpetration, repeat encounter with law enforcement, new offenses, protective/restraining orders, and rearrest; $k = 25$), (b) severe/near fatal IPV ($k = 6$), (c) any violent crime recidivism ($k = 3$), (d) general crime recidivism ($k = 3$), and/or (e) threatened physical or sexual IPV ($k = 1$). Relatedly, few studies ($k = 7$) analyzed victim/survivor-reported data on IPV (as opposed to analyzing criminal justice-reported incidents). When reported, the length of time between risk assessment and measurement of predictive outcomes ranged from 1 month to 8 years. However, many studies did not report information on time to follow-up or employed record review in which the length of time covered was generally unclear and/or variable across records/cases ($k = 14$).

AUC. The AUC was reported for 12 instruments: the B-SAFER ($k = 1$), CRAT-P ($k = 1$), CRAT-V ($k = 1$), DA-5 ($k = 1$), DA ($k = 3$), DA-I ($k = 1$), DVRAG ($k = 2$), DVSI-R ($k = 4$), K-SID ($k = 1$), ODARA ($k = 7$), SARA ($k = 4$), and SRA-PA ($k = 1$). These statistics were computed for different outcomes and with various follow-up periods; as such, AUC statistics varied, ranging from 0.51 (K-SID) to 0.86 (DA), are not directly comparable. The AUC statistics for the B-SAFER, CRAT-V, and DVRAG were significantly related to each outcome of interest and ranged from 0.70 to 0.71. For the DVRAG, outcomes included IPV recidivism, general violent crime recidivism, and general crime recidivism. AUC statistics for the B-SAFER focused on repeat encounters with law enforcement for IPV-related incidents, and calculations for the CRAT-V focused on IPV victimization. Significant AUC statistics were also reported for the CRAT-P (IPV perpetration AUC = 0.76), DA-5 (severe/near fatal IPV AUC = 0.79), DA-I (IPV AUC = 0.77, severe IPV AUC = 0.85), K-SID (any sexual/physical IPV AUC = 0.51–0.55, severe sexual/physical IPV AUC = 0.52–0.54), and SRA-PA (domestic violence recividism AUC

= 0.61). The reported AUCs for the DVSI-R were significantly related to the outcomes of interest (range = 0.62–0.73) and varied based on outcome. AUC statistics ranged from 0.64 to 0.77 for the ODARA, 0.61 to 0.86 for the DA, and 0.52 to 0.72 for the SARA and varied based on the amount of time between assessment and outcome. The DA-5, DA-I, and RAS-DV samples were used for scale development and testing.

Sensitivity, specificity, PPV, and NPV. Sensitivity and specificity were reported for 10 instruments based on a cut point: CRAT-P, CRAT-V, DA-5, DA, DVSI-R, K-SID, Lethality Screen, ODARA, SARA, and SIVIPAS. Sensitivity ranged from .48 for the SIVIPAS (i.e., 48% of those who severely reassaulted their partners were identified using a cutoff score of 10) to a high of .93 for the Lethality Screen (i.e., 93% of near fatal IPV was identified). Sensitivity of the other tested instruments included (a) .58 for the DVSI-R, (b) .59–.68 for the ODARA, (c) .61 for the CRAT-P, (d) .64 for the CRAT-V, (e) .75 for the DA, (f) .82 for the SARA, and (g) .83 for the DA-5. Specificity also varied across instruments, from a low of .21–.24 for the Lethality Screen to a high of .86 for the DA. Additional specificity estimates included (a) .34 for the DVSI-R, (b) .35– .79 for the ODARA, (c) .50 for the SARA, (d) .56 for the DA-5, (e) .62–.68 for the CRAT-V, (f) .64 for the CRAT-P, and (g) .81 for the SIVIPAS. Some studies also reported sensitivity and specificity by risk category (e.g., high risk, for recidivism in criminal settings). PPV and NPV were less common measures of predictive ability. These statistics were reported for five instruments: the CRAT-P (PPV = .12; NPV = .95), CRAT-V (PPV = .16; NPV = .97), DA-5 (PPV = .25; NPV = .95), Lethality Screen (PPV = .13–.64; NPV = .48–.96), and ODARA (PPV = .30–.72; NPV = .70–.96).

Other analyses. HRs were presented for the DVSI-R based on time to recidivism. Correlations, adjusted odds ratios, and/or relative risk probabilities were presented for the DA, DVSI-R, K-SID, ODARA, SARA, and SRA-PA. Details on these analyses, which were uncommon, are presented in Table 3.

Feasibility of use. One study (Cairns & Hoffart, 2009) directly assessed the feasibility of using an IPV/IPH risk assessment tool, specifically the DA, in a real-world setting. Researchers conducted semistructured interviews with staff working in women's shelters in Alberta, Canada, who administered the DA to a total of 509 survivors of IPV. Study findings regarding successes and challenges with the DA indicated (a) the calendar portion of the DA helped shelter staff better support survivors as it increased their understanding of the suvivors' risk for IPV/IPH, (b) survivors tended to minimize their abuse for reasons including fear of repercussions from the child welfare system, (c) survivors with low literacy levels had difficulty comprehending the DA questions, (d) and the DA may not be applicable for aboriginal women entering women's shelters (i.e., violence experienced not always perpetrated by an intimate partner). The study also found that the following innovations to DA administration may be useful in shelter settings: (a) individual completion of the DA calendar while in the presence of a group of survivors and (b) completion of routine items on an intake form prior to DA completion.

Research Question 2: Settings, Populations, and Environments Tested

Study location and tool administration. The risk assessment tools identified were tested in eight countries. Most studies were conducted in the United States ($k = 16$) and Canada ($k = 12$), followed by Sweden ($k = 5$), Israel ($k = 4$), China ($k = 3$), Austria ($k = 1$), New Zealand ($k = 1$), and Spain ($k = 1$). Across 48.8% ($k = 21$) of the studies, risk assessment tools were administered or coded by researchers; the remaining 51.2% ($k = 22$) used service professionals to administer the tools. Specifically, some tools were administered or coded by law enforcement/officers of the court/corrections staff (B-SAFER, B-SAFER v.2, SARA, SIVIPAS, SRA-PA, and SVRA-I; $k = 13$), allied health professionals (i.e., nurses, family relationship counselors, correctional staff, and mental health staff; B-SAFER, DVSI-R, ODARA, and SVRA-I; $k = 6$), and social/shelter/hotline workers (DA, RAS-DV, and SVRA-I; $k = 3$). Several tools were administered or coded by researchers only: DVRAG ($k = 2$), CRAT-P ($k = 1$), CRAT-V ($k = 1$), DA-5 ($k = 1$), DA-I ($k = 1$), DA-R ($k = 1$), K-SID ($k = 1$), and Lethality Screen ($k = 1$).

Instruments were tested in at least eight different languages, with some studies using a single tool in multiple languages. These languages were English (B-SAFER, DA, DA-5, DA-I, DVSI-R, K-SID, Lethality Screen, ODARA, and SARA; $k = 15$), Spanish (DA, DA-I, and K-SID; $k = 5$), Cantonese (CRAT-P and CRAT-V; $k = 2$), French (ODARA; $k = 1$), German (ODARA; $k = 1$), Mandarin (CRAT-P; $k = 1$), Putonghua (CRAT-V; $k = 1$), and Swedish (B-SAFER; $k = 1$). Twenty-four studies did not explicitly report the language in which the instrument was administered.

Data collection strategies. Most studies used information from IPV perpetrators or records of perpetrators' data ($k = 26$), and 14 studies analyzed information from IPV victims/survivors or records of their data. Two studies included family violence perpetrators more broadly (i.e., perpetrators who committed violence against any member of their family), and one study analyzed data from both IPV victims/survivors and perpetrators.

Several studies used multiple data collection methods. The majority of studies employed retrospective case and/or record review to collect data on participants ($k = 27$; i.e., reviewing individuals' criminal or other type of case file information). Fourteen studies used structured interviews/questionnaires to collect data directly from victims/perpetrators, six conducted secondary data analyses of existing data sets, three used semi-structured interviews, one used field observation, and one used both focus groups and interviews.

Sample size and characteristics. The sample size in the reviewed studies ranged from 19 to 29,317, with a median sample size of 254 participants. Overwhelmingly, the identified studies that explicitly reported victim/survivor and perpetrator gender focused on assessing IPV/IPH risk among mixed-sex intimate partnerships ($k = 18$) and male violence perpetration ($k = 29$). Only three studies included same-sex intimate partnerships, all of which were female–female partnerships, and many studies did not report victim/survivor and/or perpetrator genders ($k = 24$). Most studies ($k = 30$) included only adult participants (18 years or older). Three studies included both adolescent and adult participants, and no studies included only adolescent participants.

Twenty-nine studies reported categories for race/ethnicity and/or immigration status of perpetrators and/or victims/survivors. These varied widely and included categories such as race, ethnicity, country of origin, aboriginal, immigration status, and nationality. In the United States, researchers primarily described samples by racial/ethnic categories such as White, Latino/Latina, Black/African American, Asian, "minority," and "ethnic minority." A study in Canada focused on Caucasian and aboriginal status. Other studies, primarily outside of the United States, referred to individuals according to national or foreign-born status.

Discussion

Given the critical need to prevent IPH, it is imperative that service providers and first responders are equipped with IPV/IPH risk assessment instruments that are valid, reliable, easy to use, and appropriate for their setting and population of interest. This review aimed to detect and synthesize research examining the validity, reliability, and feasibility of use of current IPV/IPH risk assessment tools, as well as the settings, populations, and environments in which these tools have been tested. We identified 43 studies examining 18 different IPV/IPH risk assessment tools. Review findings demonstrate that most instruments had only been examined by one or two studies with current versions of the ODARA, DA, SARA, DVSI-R, and SVRA-I being the most studied instruments. The DA, DVSI-R, K-SID, ODARA, and SARA are the only instruments with information regarding internal consistency and interrater reliability, as well as construct, concurrent, and predictive validity. Table 4 highlights key findings and implications.

Table 4. Summary Table With Implications for Practice, Policy, and Research.

Practice
- The review results offer first responders and service providers an overview of existing IPV/IPH risk assessment instruments, the psychometric strength of these instruments, and the populations and settings in which the instruments have been tested. We encourage professionals to use the strongest psychometric instrument available that has been developed and tested with their target population and setting.
- IPV/IPH risk assessment instruments are not meant to be interventions but instead should be used to connect survivors in high-risk situations to needed services and interventions. As such, first responders and service providers seeking to implement an IPV/IPH risk assessment instrument in their setting need to determine a plan for addressing high-risk cases.

Policy
- Local and state governments that suggest or require the use of IPV/IPH risk assessment in the criminal justice system should understand the psychometric properties of such instruments, choose the instrument best suited to their settings and populations, and ensure that policy specifies reasonable intervention in high risk cases.
- Funding is needed to ensure that first responders and service professionals have access to and receive training on how to use IPV/IPH screening tools appropriately and have the resources to respond to cases identified as high risk.

Research
- Standardized reporting practices and guidelines are needed to advance research on the psychometric properties of IPV/IPH risk assessment tools to improve comparability of results across studies and tools.
- Future research is needed to assess the reliability, validity, and feasibility of IPV/IPH risk assessment tools across diverse samples, particularly related to perpetrator gender and relationship composition (e.g., same-sex couples, people who identify as LGBTQI+, and female perpetrators). Research is also needed to test these tools in non-Western countries and in languages other than English.
- There is a critical need for research examining the feasibility of using these tools in clinical and practice settings. Similarly, there is a need for more testing focused on feasibility of use in real-world settings with practitioner administration.

Psychometric Properties of IPV/IPH Risk Assessment Tools

The reviewed studies analyzed and presented the psychometric properties of IPV/IPH risk assessment tools in various ways. Reliability was most commonly assessed in terms of interrater reliability, which was generally high across studies. Internal consistency reliability was reported infrequently and with rather low Cronbach αs. No study achieved optimal internal consistency reliability thresholds suggested by the research literature (i.e., .90+ for practitioner use). This form of reliability indicates the extent to which the risk items on a tool are related to one another. To understand the relevance of this measure of reliability, it is important to consider the assumption about the theoretical conceptualization of risk that underlies the use of the test. If risk is conceptualized as a latent variable or an underlying agent that causes the items in a scale to take on certain values (DeVellis, 2003), then it would be expected that the items are intercorrelated. Alternatively, if risk is conceptualized as an induced variable wherein the items collectively determine the level of risk and are not indicators of a causal latent variable, then presenting internal consistency reliability could be misleading because we would not expect high intercorrelation among items (Bollen, 1989; DeVellis, 2003). It may also be that clusters of items are intercorrelated due to an underlying causal agent, such as coercive control, that

is an indicator of risk but is not risk itself (Myhill & Hohl, 2016). Future IPV/IPH risk assessment research should focus on better delineating the function and form of risk.

Predictive validity was commonly examined as the primary aim of IPV/IPH risk assessment tools is to predict the likelihood of future assault or homicide. Some measures of predictive ability were rather low (e.g., SIVIPAS sensitivity =.48; Lethality Screen specificity =.21–.24; SARA AUC =0.52–0.65). Although predictive validity is typically assessed in terms of having both high sensitivity and specificity (Douglas, Yeomans, & Boer, 2005), depending on the intended use of the IPV/IPH risk assessment instrument, a practitioner might value one form of predictive validity over another. For example, when an instrument is used to engage high-risk survivors in a brief, low-cost intervention (e.g., the Lethality Screen), it might be more important to identify any potential danger (high sensitivity), even if it means including survivors as high risk who will not experience violence in the follow-up period (low specificity). On the other hand, practitioners interested in identifying high-risk survivors for a costly, resource-intensive intervention might prefer an instrument with more balanced sensitivity and specificity (Messing & Campbell, 2016).

There were mixed findings related to concurrent validity (e.g., ODARA and SARA were found to be significantly related to some gold standard instruments, but not others). Notably, limited consensus exists on the gold standard that IPV/IPH risk assessment tools should be tested against. In this review, numerous different "gold standards" were tested including professional judgment, victim/survivor perception of risk, the CTS2, and other IPV/IPH risk assessment instruments (e.g., DA, SARA, DVSI, DVSR), as well as instruments designed to measure psychopathic tendencies, interaction styles that may lead to conflict, risk of violence or general/violent recidivism, and social problem-solving. The broad range of gold standards used suggests that the field is challenged in conceptualizing risk. That is, it is not possible to identify to what risk should be related to until researchers specify *what risk is*. For example, physical violence as measured by the CTS2 may capture the risk posed by abusers who use severe violence but may not capture risk posed by abusers who use coercive control and threats. Likewise, comparing an IPV risk assessment to an IPH risk assessment may yield poor concurrent validity because not all IPV and IPH risk factors overlap (Messing & Thaller, 2015). Given that research has indicated that risk assessment is more accurate than professional judgment (Campbell et al., 2009; Hilton et al., 2004), using professional judgment as a gold standard may be misleading. This study's findings should be interpreted in light of the challenges faced both in conceptualizing risk and determining standards by which to assess risk instruments.

Summarizing validity findings was challenging because of methodological differences across studies. Like concurrent validity, predictive validity was assessed in various ways (e.g., PPV, NPV, AUC, sensitivity/specificity), using many different outcomes, and ranging substantially in length of time between IPV/IPH assessment tool administration and outcome measurement. Given the difficulty summarizing and interpreting reliability and validity findings, standardizing reporting practices and guidelines on the psychometric properties of IPV/IPH risk assessment tools could improve comparability of results across studies. Such a recommendation echoes and underscores prior calls for the standardization of predictive validity reporting protocols (Singh, Desmarais, & Van Dorn, 2013). For such a consensus to emerge, researchers must first address the broader questions about the nature of IPV/IPH risk noted previously.

Feasibility of IPV/IPH Risk Assessment Tools

We identified only one study that specifically provided information regarding the feasibility of using an IPV/IPH risk assessment tool in a real-world setting. Further, almost half of the studies examined the reliability and/or validity of a given tool based on researcher completion via retrospective case analysis or record review as opposed to tool administration in a real-time setting by service providers. It is imperative that future research investigate the psychometric properties of IPV/IPH risk assessment tools administered by service providers in real-world settings and the feasibility of typical providers' appropriate and successful use of these tools.

Populations, Settings, and Environments of IPV/IPH Risk Assessment Tool Testing

Although this review found that the identified tools had been examined in several different countries and languages, research was primarily conducted in North America and in English. Future research should examine the administration of IPV/IPH risk assessment in non-Western countries and languages other than English (Messing & Thaller, 2015). When determining what tool would be most appropriate for a given setting, professionals should ensure the tool has been tested in the target respondent's primary language.

Most of the studies included in this review gathered information solely from IPV perpetrators or perpetrator records. Focusing on data that are recorded by law enforcement officials or other systems rather than provided by victims/survivors has significant implications for current knowledge. Given the well-established underreporting of IPV to authorities (e.g., Tjaden & Thoennes, 2000), relying on official criminal justice data biases our understanding of existing instruments by focusing on a likely unrepresentative subsample of IPV situations. Thus, instruments tested soley with criminal justice data might have limited generalizability. Likewise, it might not be feasible in certain practice settings, such as social work advocacy settings, to collect information from perpetrators. Tools should ideally be tested and found valid and reliable with the respondent who will be providing this information in each practice setting.

The included studies generally examined IPV/IPH risk assessment tools in the context of mixed-sex relationships with a male perpetrator, and many studies did not report on the gender of either the victim/survivor or perpetrator. The lack of reporting on gender and the almost exclusive focus on

mixed-sex partnerships with male perpetrators reinforce the dominant narrative of IPV as only (as opposed to primarily) a problem of male violence against female partners (Blosnich & Bossarte, 2009). Future research is needed to assess the reliability, validity, and feasibility of IPV/IPH risk assessment tools across diverse samples including among (a) same-sex couples, (b) people who identify as lesbian, gay, bisexual, transgender, queer, intersex, and additional identities (LGBTQI+), and (c) female perpetrators.

Choosing and Using an IPV/IPH Risk Assessment Instrument

Overall, this review suggests that many of the available IPV/IPH risk assessment instruments have some evidence to support their reliability and validity. Practitioners working with IPV victims/survivors and/or perpetrators must consider a range of factors—not just reliability and validity—when determining which IPV/IPH risk assessment tool is appropriate for their practice setting. These factors can broadly be understood as the fit between the context in which the instrument has been developed and tested and the context in which the practitioner wishes to use the instrument. Another important consideration is the length of the instrument. The IPV/IPH risk assessment instruments identified in this review ranged from 5 to 35 items; practitioners must determine the feasibility of administering an instrument given the amount of time required to complete it.

The majority of research on IPV/IPH risk assessment focuses narrowly on the prediction of future violence rather than on the use of risk assessment as a violence prevention strategy (Douglas & Kropp, 2002). Some risk assessment instruments, like the Lethality Screen, were specifically designed as part of a risk-informed collaborative intervention (Messing & Campbell, 2016; Messing et al., 2017). Similarly, in the United Kingdom, there is a long history of Multi-Agency Risk Assessment Conferences and Independent Domestic Violence Advocacy schemes that utilize information from IPV/IPH risk assessment to enhance victim/survivor safety and offender accountability (Robinson, 2007). Whatever the application and context, practitioners should use IPV/IPH risk assessment within an evidence-based practice framework wherein risk assessment is the best evidence of future violence, and both practitioner expertise and client self-determination are used to contextualize risk scores and develop client-specific strategies to mitigate risk (Messing, 2019). Risk-informed safety planning, wherein safety information is provided based on overall risk score and targeting specific risk factors (e.g., gun ownership, stalking behavior), is an important aspect of the DA, for example (Campbell, 2001). Other risk assessments are used to determine criminal justice interventions such as conditions of release (e.g., DVSI-R, ODARA). Risk is dynamic and should be reassessed to understand the risk posed at a particular time (e.g., upon separation). In other words, IPV/IPH risk assessment is a process, not an end goal.

Limitations and Conclusions

Review findings should be considered in light of study limitations. Despite a comprehensive search strategy, it is possible that relevant studies were not identified or located and included in this review. Furthermore, by limiting the review to studies available in English, we might have missed relevant studies published in other languages. Additionally, it is possible that we might have overlooked or misinterpreted information presented in the included studies. To address this potential concern, data were systematically extracted by two reviewers using a standard review form with a thorough process for addressing discrepancies.

Despite the limitations noted above, this review offers important findings for communities and professionals considering IPV/IPH risk assessment by providing an overview of the validity, reliability, and feasibility of use of available IPV/IPH risk assessment tools in their most current version and summarizing information concerning the populations, settings, and environments in which these tools have been tested. In addition, this review notes the gaps in the current research, while providing practitioners with evidence concerning the psychometric properties of IPV/IPH risk assessment tools. Informed by this study's findings, we invite policy makers and funders to ensure that practitioners have access to appropriate IPV/IPH risk assessment tools and related trainings on identifying and responding to high-risk cases.

Acknowledgments

We would like to thank the Caroline H. and Thomas S. Royster Fellowship awarded by the University of North Carolina at Chapel Hill which supported, in part, Laurie Graham's time and effort, and the Chancellor's Fellowship awarded by the University of North Carolina at Chapel Hill that supported, in part, Kashika Sahay's time and effort for this research.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Laurie M. Graham https://orcid.org/0000-0003-3318-8591

References

*References marked with an asterisk indicate studies included in the meta-analysis.

Andrews, D. A. (1982). *The Level of Supervision Inventory (LSI): The first follow-up (report)*. Toronto, Canada: Ontario Ministry of Correctional Service.

*Au, A., Cheung, G., Kropp, R., Yuk-chung, C., Lam, G. L., & Sung, P. (2008). A preliminary validation of the Brief Spousal Assault

Form for the Evaluation of Risk (B-SAFER) in Hong Kong. *Journal of Family Violence*, *23*, 727.

*Belfrage, H., & Strand, S. (2008). Structured spousal violence risk assessment: Combining risk factors and victim vulnerability factors. *International Journal of Forensic Mental Health*, *7*, 39–46.

*Belfrage, H., Strand, S., Storey, J. E., Gibas, A. L., Kropp, P. R., & Hart, S. D. (2012). 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.

Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., . . . Stevens, M. R. (2011). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 summary report*. Atlanta, GA: Centers for Disease Control and Prevention.

Blosnich, J. R., & Bossarte, R. M. (2009). Comparisons of intimate partner violence among partners in same-sex and opposite-sex relationships in the United States. *American Journal of Public Health*, *99*, 2182–2184.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

*Bourgon, G., & Bonta, J. (2004). *Risk assessment for general assault and partner abusers*. Ottawa, Canada: Public Safety and Emergency Preparedness Canada.

*Buchanan, K. (2009). *Risk assessment and spousal violence: Predictive validity and cultural applicability* (Unpublished doctoral dissertation). University of Regina, Regina, Saskatchewan, Canada.

*Cairns, K., & Hoffart, I. (2009). *Keeping women alive: Assessing the danger*. Edmonton, Canada: Alberta Council of Women's Shelters.

*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* (Unpublished doctoral dissertation). Massey University, Wellington, New Zealand.

Campbell, J. C. (2001). Safety planning based on lethality assessment for partners of batterers in intervention programs. *Journal of Aggression, Maltreatment & Trauma*, *5*, 129–143.

Campbell, J. C., Glass, N., Sharps, P. W., Laughon, K., & Bloom, T. (2007). Intimate partner homicide review and implications of research and policy. *Trauma, Violence, & Abuse*, *8*, 246–269.

Campbell, J. C., Messing, J. T., & Williams, K. R. (2017). Prediction of homicide of and by battered women. In J. C. Campbell & J. T. Messing (Eds.), *Assessing dangerousness: Domestic violence offenders and child abusers*. New York, NY: Springer.

*Campbell, J. C., O'Sullivan, C., Roehl, J., & Webster, D. W. (2005a). *Intimate partner violence risk assessment validation study: The RAVE study* (Final Report to the National Institute of Justice; NCJ 209731-209732). Washington, DC: Bureau of Justice Statistics.

*Campbell, J. C., O'Sullivan, C., Roehl, J., & Webster, D. W. (2005b). *Intimate partner violence risk assessment validation study: The RAVE study practitioner summary and recommendations: Validation of tools for assessing risk from violent intimate partners* (Report to the National Institute of Justice; NCJ 209732). Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

*Campbell, J. C., Webster, D. W., & Glass, N. (2009). The danger assessment validation of a lethality risk assessment instrument for intimate partner femicide. *Journal of Interpersonal Violence*, *24*, 653–674.

Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C. R., Campbell, D. W., Curry, M. A., . . . Wilt, S. A. (2003). Assessing risk factors for intimate partner homicide. *National Institute of Justice Journal*, *250*, 14–19.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment* (Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 17). Thousand Oaks, CA: Sage.

*Chan, K. L. (2012). Predicting the risk of intimate partner violence: The Chinese risk assessment tool for victims. *Journal of Family Violence*, *27*, 157–164.

*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*, 500–516.

Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

*Dayan, K., Fox, S., & Morag, M. (2013). Validation of spouse violence risk assessment inventory for police purposes. *Journal of Family Violence*, *28*, 811–821.

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Douglas, K. S., Guy, L. S., Reeves, K. A., & Weir, J. (2008). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography*. Retrieved from http://kdouglas.files.wordpress.com/2006/04/annotate10-24nov2008.pdf

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

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior*, *32*, 479–510.

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

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

*Eke, A. W., Hilton, N. Z., Harris, G. T., Rice, M. E., & Houghton, R. E. (2011). Intimate partner homicide: Risk assessment and prospects for prediction. *Journal of Family Violence*, *26*, 211–216.

Farrell, H. M. (2011). Batterers: A review of violence and risk assessment tools. *Journal of the American Academy of Psychiatry and the Law Online*, *39*, 562–574.

*Gerstenberger, C. B., & Williams, K. R. (2013). Gender and intimate partner violence: Does dual arrest reveal gender symmetry or asymmetry? *Journal of Interpersonal Violence*, *28*, 1561–1578.

*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*, 1021–1027.

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

*Gulati, G. (2000). *Development of a risk assessment scale for domestic violence* (Unpublished doctoral dissertation). Spalding University, Louisville, KY, United States.

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

*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*, 326–337.

*Hilton, N. Z., Harris, G. T., & Holder, N. (2008). Actuarial assessment of violence risk in hospital-based partner assault clinics. *Canadian Journal of Nursing Research*, *40*, 56–70.

*Hilton, N. Z., Harris, G. T., Popham, S., & Lang, C. (2010). Risk assessment among incarcerated male domestic violence offenders. *Criminal Justice and Behavior*, *37*, 815–832.

*Hilton, N. Z., Harris, G. T., Rice, M. E., Houghton, R. E., & Eke, A. W. (2008). An in-depth actuarial assessment for wife assault recidivism: The domestic violence risk appraisal guide. *Law and Human Behavior*, *32*, 150–163.

*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*, 267–275.

*Hilton, N. Z., Popham, S., Lang, C., & Harris, G. T. (2014). Preliminary validation of the ODARA for female intimate partner violence offenders. *Partner Abuse*, *5*, 189–203.

Jack, S. P. D., Petrosky, E., Lyons, B. H., Blair, J. M., Ertl, A. M., Sheats, K. J., & Betz, C. J. (2018). Surveillance for violent deaths: National violent death reporting system, 27 states, 2015. *MMWR Surveillance Summaries*, *67*, 1–32.

*Juodis, M., Starzomski, A., Porter, S., & Woodworth, M. (2014a). A comparison of domestic and non-domestic homicides: Further evidence for distinct dynamics and heterogeneity of domestic homicide perpetrators. *Journal of Family Violence*, *29*, 299–313.

Juodis, M., Starzomski, A., Porter, S., & Woodworth, M. (2014b). What can be done about high-risk perpetrators of domestic violence? *Journal of Family Violence*, *29*, 381–390.

Kropp, P. R. (2008). Intimate partner violence risk assessment and management. *Violence and Victims*, *23*, 202.

*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*, 101–118.

Messing, J. T. (2019). Risk-informed intervention: Using intimate partner violence risk assessment within an evidence based practice framework. *Social Work*, *64*, 103–112.

*Messing, J. T., Amanor-Boadu, Y., Cavanaugh, C. E., Glass, N. E., & Campbell, J. C. (2013). Culturally competent intimate partner violence risk assessment: Adapting the danger assessment for immigrant women. *Social Work Research*, *37*, 263–275.

Messing, J. T., & Campbell, J. (2016). Informing collaborative interventions: Intimate partner violence risk assessment for front line police officers. *Policing: A Journal of Policy & Practice*, *10*, 328–340.

*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*, 205–226.

*Messing, J. T., Campbell, J., Wilson, J. S., Brown, S., Patchell, B., & Shall, C. (2014). *Police departments' use of the lethality assessment program: A quasi-experimental evaluation*. Washington, DC: National Institute of Justice.

Messing, J. T., & Thaller, J. (2013). The average predictive validity of intimate partner violence risk assessment instruments. *Journal of Interpersonal Violence*, *28*, 1537–1558.

Messing, J. T., & Thaller, J. (2015). Intimate partner violence risk assessment: A primer for social workers. *British Journal of Social Work*, *45*, 1804–1820.

Moracco, K. E., Runyan, C. W., & Butts, J. D. (1998). Femicide in North Carolina. *Homicide Studies*, *2*, 422–446.

Morgan, W., & Gilchrist, E. (2010). Risk assessment with intimate partner sex offenders. *Journal of Sexual Aggression*, *16*, 361–372.

Myhill, A., & Hohl, K. (2016). The "golden thread": Coercive control and risk assessment for domestic violence. *Journal of Interpersonal Violence*, *34*(21–22), 4477–4497. doi: 10.1177/0886260516675464

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*, 76–168.

Nunnally, J. C., & Bernstein, I. H. P. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Pataki, G. (1997). *Intimate partner homicides in New York State*. Albany: State of New York.

*Peterson, K. (2013). Learned resourcefulness, danger in intimate partner relationships, and mental health symptoms of depression and PTSD in abused women. *Issues in Mental Health Nursing*, *34*, 386–394.

*Rettenberger, M., & Eher, R. (2013). Actuarial risk assessment in sexually motivated intimate-partner violence. *Law and Human Behavior*, *37*, 75–86.

Robinson, A. L. (2007). MARACs: Reducing repeat victimization among high-risk victims of domestic violence: The benefits of a coordinated community response in Cardiff, Wales. *Violence Against Women*, *12*, 761–788.

*Rollins, C., Glass, N. E., Perrin, N. A., Billhardt, K. A., Clough, A., Barnes, J., . . . Bloom, T. L. (2012). Housing instability is as strong a predictor of poor health outcomes as level of danger in an abusive relationship: Findings from the SHARE study. *Journal of Interpersonal Violence*, *27*, 623–643.

*Rud, J. D., Skiling, N., & Nonemaker, D. (2011). *DOCCR validation of two domestic violence risk instruments: Domestic Violence Screening Instrument (DVSI) & Spousal Abuse Risk Assessment (SARA)*. St. Paul, MN: Hennepin County Department of Community Corrections and Rehabilitation, Office of Planning, Policy, and Evaluation.

*Sabri, B., Stockman, J. K., Bertrand, D. R., Campbell, D. W., Callwood, G. B., & Campbell, J. C. (2013). Victimization experiences, substance misuse, and mental health problems in relation to risk for lethality among African American and African Caribbean women. *Journal of Interpersonal Violence*, *28*, 3223–3241.

Sharps, P., Campbell, J. C., Campbell, D., Gary, F., & Webster, D. (2003). Risky mix: Drinking, drug use and homicide. *NIJ Journal*, *250*, 9–13.

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, *31*, 8–22.

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law*, *31*, 55–73.

*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*, 1208–1216.

*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*, 163–180.

Stöckl, H., Devries, K., Rotstein, A., Abrahams, N., Campbell, J., Watts, C., & Moreno, C. G. (2013). The global prevalence of intimate partner homicide: A systematic review. *The Lancet*, *382*, 859–865.

*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*, 256–271.

*Storey, J. E., & Strand, S. (2013). Assessing violence risk among female IPV perpetrators: An examination of the B-SAFER. *Journal of Aggression, Maltreatment & Trauma*, *22*, 964–980.

Tjaden, P. G., & Thoennes, N. (2000). *Extent, nature, and consequences of intimate partner violence (NCJ 181867)*. Washington, DC: National Institute of Justice.

*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*, 120–129.

*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*, 400–408.

*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, HI: Hawaii State Department of Health.

Authors' Biographies

Laurie M. Graham, MSW, received her master's degree from the School of Social Work, University of North Carolina at Chapel Hill. As a practitioner, she has worked with various community and governmental agencies to provide support services for survivors of gender-based violence (GBV) and violence prevention programming, most recently as the Programs Director for the Orange County Rape Crisis Center. As a recipient of the 5-year Caroline H. and Thomas S. Royster Fellowship, she is pursuing her doctorate at the School of Social Work, University of North Carolina at Chapel Hill, where she has taught an experiential service-learning course and focuses her research activities on sexual violence prevention, human trafficking, and intervention and prevention strategies for survivors of GBV more broadly. She has a particular interest in developing and evaluating cross-cutting approaches to prevent various forms of violent behavior. She received the 2013 Peer Support Award from the North Carolina Coalition Against Sexual Assault for being the lead author on a manual concerning best practices in developing and coordinating support group programs for survivors of sexual violence.

Kashika M. Sahay, MPH, PhD, has a PhD in maternal and child health with a minor in biostatistics from University of North Carolina at Chapel Hill. She has a master's in epidemiology from Emory University with a focus on vulnerable populations. Her dissertation focused on family planning patterns and gendered relationships between married couples in urban Nigeria. She has advanced training in statistical analysis including longitudinal and correlated data analysis. She has worked with domestic violence and sexual assault agencies to conduct electronic needs assessments for survivors of violence. She is passionate about women's health, social justice, and gender equity. She has a commitment to multidisciplinary mixed methods research. In 2013, she was awarded the Chancellor's Fellowship by the University of North Carolina at Chapel Hill to complete her PhD. She has served as a leader within the student arm of the American Public Health Association and Women in Statistics conferences. She has presented more than 15 technical and peer-reviewed presentations at various conferences and stakeholder meetings. She currently works as a contractor for the Centers for Disease Control and Prevention. Her focus is on practical and actionable research that has timely implications for policy and practice.

Cynthia F. Rizo, MSW, PhD, is an assistant professor at the School of Social Work, University of North Carolina at Chapel Hill. She has practice experience in providing services to survivors of intimate partner violence and their children. She has worked on a number of projects in the area of gender-based violence, including intimate partner violence, human trafficking, and sexual assault. Her primary research focus consists of developing and evaluating interventions for particularly vulnerable intimate partner violence survivors including Latinas, immigrants, and system-involved survivors. She is currently working on a project to develop school-based sex trafficking content for students as well as protocols that schools can use to connect at-risk youth and victims to needed community services.

Jill T. Messing, MSW, PhD, is an associate professor in the School of Social Work at Arizona State University. She specializes in intimate partner violence risk assessment. She has evaluated the predictive validity of several forms of the Danger Assessment, including the Danger Assessment-5 items and the Lethality Screen. She has created risk assessment instruments for the use in risk-informed collaborative interventions including the Danger Assessment for Law Enforcement and the Arizona intimate Partner Risk Assessment Instrument System. She is on a research team with Dr. Jacquelyn Campbell that is adapting the Danger Assessment for the use with immigrant, refugee, and Native American victims of intimate partner violence. She is also conducting the first U.S. evaluation of the Ontario Domestic Assault Risk Assessment. As a social worker, she is particularly interested in the use of risk assessment in evidence-based practice and with the development and testing of innovative interventions for victims of intimate partner violence. She was the principal investigator on the National Institute of Justice funded *Oklahoma Lethality Assessment Study*, an examination of the effectiveness of the Lethality Assessment Program, a collaborative police-social service response to intimate partner violence. She is also a coinvestigator on two studies examining the utility of internet-based decision aids for women in abusive relationships funded by the National Institutes of Health.

Rebecca J. Macy, MSW, PhD, is the L. Richardson Preyer Distinguished Chair for Strengthening Families at the School of Social Work, University of North Carolina (UNC) at Chapel Hill, where she has taught courses in mental health, trauma and violence, social work practice, and statistics. She joined the UNC faculty in 2002, after receiving her doctoral degree in social welfare from the University

of Washington in Seattle. She is currently the editor-in-chief for the *Journal of Family Violence*. She has 15 years of experience conducting community-based studies that focus on intimate partner violence, sexual violence, and human trafficking. She has dedicated her research efforts to violence prevention and to improving services for violence survivors. To find the most effective and feasible strategies, she has conducted investigations in various community settings in collaboration with survivors, service providers, and policy makers. Her completed research includes pilot and field testing, process and qualitative evaluations, and quasi-experimental studies. Currently, she is coleading a randomized study of a community-based intervention to prevent sexual violence perpetration. She has received funding for her research from foundations, federal agencies, and state government. She has published 70 peer-reviewed articles, book chapters, and invited commentaries on these topics and given more than 100 peer-reviewed and invited research presentations at national and international venues. The rigor of her research and its benefit to practice has been recognized with awards from both the Office of the University of North Carolina Provost and the Orange County Rape Crisis Center.