Do Risk Assessment Tools Help Manage and Reduce Risk of Violence and Reoffending?
A Systematic Review

Jodi L. Viljoen       Dana M. Cochrane       Melissa R. Jonnson
Simon Fraser University

Law and Human Behavior

DOI: http://dx.doi.org/10.1037/lhb0000280

This article may not exactly replicate the authoritative document published in the APA Journal.
It is not the copy of record.
Author Note:

Jodi L. Viljoen, Dana Cochrane, and Melissa Jonnson, Department of Psychology, Simon Fraser University.

This research was supported by a grant from the Simon Fraser University/Social Sciences and Humanities Research Council of Canada’s Institutional Grants Program. The opinions expressed in this article reflect the views of the authors alone. The authors would like to thank Andrew Gray for his work as a research assistant.

Correspondence concerning this article should be address to Jodi Viljoen, Department of Psychology, Simon Fraser University, Burnaby, BC V5A 1S6. Contact: jviljoen@sfu.ca
Abstract

Although it is widely believed that risk assessment tools can help manage risk of violence and offending, it is unclear what evidence exists to support this view. As such, we conducted a systematic review and narrative synthesis. To identify studies, we searched 13 databases, reviewed reference lists, and contacted experts. Through this review, we identified 73 published and unpublished studies (N = 31,551 psychiatric patients and offenders, N = 10,002 professionals) that examined either professionals’ risk management efforts following the use of a tool, or rates of violence or offending following the implementation of a tool. These studies included a variety of populations (e.g., adults, adolescents), tools, and study designs. The primary findings were as follows: (1) despite some promising findings, professionals do not consistently adhere to tools or apply them to guide their risk management efforts; (2) following the use of a tool, match to the risk principle is moderate and match to the needs principle is limited, as many needs remained unaddressed; (3) there is insufficient evidence to conclude that tools directly reduce violence or reoffending, as findings are mixed; and (4) tools appear to have a more beneficial impact on risk management when agencies use careful implementation procedures and provide staff with training and guidelines related to risk management. In sum, although risk assessment tools may be an important starting point, they do not guarantee effective treatment or risk management. However, certain strategies may bolster their utility.

Public Significance Statement

To help guide treatment and management decisions, professionals often assess offenders’ and patients’ risk of violence or offending using risk assessment tools. However, based on our review, these tools are often not used in an optimal manner. As such, they may not have as large an impact as intended.

Keywords: systematic review, risk assessment, violence, offending, risk management
Do Risk Assessment Tools Help to Manage and Reduce Risk of Violence and Reoffending?  
A Systematic Review

The field of violence risk assessment has grown rapidly. In the past several decades, researchers have developed over 400 different tools designed to assess risk of violence and offending (Singh, Desmarais, Hurducas, et al., 2014). Professionals, such as psychologists, probation officers, nurses, psychiatrists, and police, have widely adopted these tools in as many as 44 countries (Singh, Desmarais, Hurducas, et al., 2014). In addition, administrators and policymakers have created policies and, in some cases, laws mandating the use of tools (Monahan & Skeem, 2014). For instance, in the United States, at least 20 states use risk assessment tools in sentencing offenders (Starr, 2014), and at least 28 states use tools to determine if an offender should be released on parole (Harcourt, 2007). All 50 states have adopted the use of risk assessment tools in youth probation, with 34 states mandating the use of tools on a state-wide basis (Wachter, 2015). Thus, considering that over 4.5 million American adults and adolescents are placed on probation or parole each year (Kaeble & Bonczar, 2017; Livsey, 2012) and over 1.5 million Americans are incarcerated (Carson & Anderson, 2016), tools are used with many offenders.

Risk assessment tools are also commonly used in forensic psychiatric facilities (Singh, Desmarais, Hurducas, et al., 2014), general psychiatric hospitals (Higgins, Watts, Bindman, Slade, & Thornicroft, 2005), treatment programs (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010), correctional centers (Taxman, Cropsey, Young, & Wexler, 2007), and in a variety of court evaluations, such as those involving the transfer of adolescents to adult court and the civil commitment of individuals who have sexually offended (Viljoen, McLachlan, & Vincent, 2010). Extrapolating from these reports, it could be estimated that, globally, over a million risk assessments are conducted each year.

Driving the widespread adoption of risk assessment tools is a belief that such tools can help professionals to better manage and ultimately reduce risk. Broadly speaking, risk management refers to the process of planning and implementing strategies to help prevent violence and other forms of offending. It is carried out by a variety of professionals (e.g., psychologists, probation officers, nurses, police) and encompasses not only treatment (e.g., therapy), but also strategies such as supervision, case management, and placement decisions. For instance, Hart, Douglas, and colleagues (Douglas, Hart, Webster, & Belfrage, 2013; Hart et al., 2003) conceptualize risk management strategies as falling in four domains: (1) treatment such as one-on-one or group counselling, (2) supervision and control efforts such as placements in secure facilities or supervision by probation officers, (3) monitoring of risk factors, and (4) victim-safety planning strategies such as warning potential victims.

There is widespread agreement, among risk assessment researchers, that risk management is a key goal of risk assessment (Andrews & Bonta, 2010; Douglas & Kropp, 2002; Hart & Logan, 2011; Heilbrun, 1997). Some authors have pointed out that the goal of risk management may be more salient in some contexts than others (e.g., modifiable decisions in which courts have continued legal jurisdiction vs. decisions in which there is no continued jurisdiction; Heilbrun, 1997). However, other authors have asserted that risk management is the primary goal or sine qua non of risk assessment regardless of context (Hart & Logan, 2011). For instance,
Hart (1998) states, “To put it simply, the clinical task is violence prevention, not violence prediction” (p. 123). This is because predicting if someone will reoffend, in and of itself, has little value if nothing is done to manage risk. In fact, according to some authors, assessing risk without making efforts to manage it is “harmful and unethical” (Roychowdhury & Adshead, 2014, p. 81).

As a result, the field of risk assessment has evolved to focus increasingly on the management and reduction of risk, rather than solely its prediction (Bonta & Andrews, 2017). Indeed, many risk assessment tools are now “infused with the concept of risk reduction” (Skeem & Lowenkamp, 2016, p. 682). This focus on risk management is evident in tools’ manuals and descriptions. For instance, the authors of the Structured Assessment of Violence Risk in Youth (SAVRY) describe that “the objective of the SAVRY is ultimately to assist in prevention and risk reduction” (Borum, Bartel, & Forth, 2006, p. 5).

Similarly, many tool developers list specific types of risk management tasks for which their tools can be used. These include tasks such as to help judges “decide between alternative sentences” (Kropp & Gibas, 2010, p. 230), determine an “appropriate security classification” (Bonta & Andrews, 2017, p. 202), “inform future risk management and treatment plans” (Douglas, Hart, Webster, & Belfrage, 2013, p. 35), “assist in treatment and daily management” (Webster, Martin, Brink, Nicholls, & Desmarais, 2009, p. 4), and “help parole boards in making release decisions” (Bonta & Andrews, 2017, p. 202). The belief that risk assessment can help manage and reduce reoffending is also evident in some practice guidelines. For instance, the National Working Group of the National Center for State Courts encourages judges to use risk assessment tools in sentencing decisions because tools offer benefits such as “contributing to public safety” (Casey, Warrne, & Elek, 2011, p. 7). However, even though many risk assessment tools aim to help manage and reduce risk, it is unclear whether tools do, in fact, achieve this goal. As we describe below, opinions about this differ.

Tools Might Help Reduce Risk by Linking Individuals to Appropriate Interventions

Many tool developers, researchers, and practitioners consider risk assessment to be an important—or even essential—starting point to risk management (e.g., Douglas & Kropp, 2002; Vincent, Paiva-Salisbury, Cook, Guy, & Perrault, 2012). However, in more exact terms, how might the use of risk assessment tools improve risk management practices, or reduce reoffending and violence? Although the mechanism between risk assessment and risk management is not always articulated, the primary hypothesized mechanism appears to be twofold (see Bonta & Andrews, 2017). First, it is thought that tools will increase professionals’ level of adherence to the risk-need-responsivity (RNR) model of offender treatment. Second, it is believed that this will, in turn, decrease offending.

The RNR model (Andrews, Bonta, & Hoge, 1990; Bonta & Andrews, 2017) is the dominant model of offender treatment, and it has considerable research support. A number of meta-analyses have found that programs that adhere to the RNR principles are more effective in reducing reoffending than those that do not adhere to these principles (e.g., Hanson, Bourgon, Helmus, & Hodgson, 2009; Koehler, Lösel, Akoensi, & Humphreys, 2013; Lipsey, 2009; Lowenkamp, Latessa, & Holsinger, 2006; Prendergast, Pearson, Podus, Hamilton, & Greenwell,
As such, increasing adherence to this model is a reasonable goal.

The RNR model has three primary principles, namely the risk, need, and responsivity principles (Andrews & Bonta, 2010; Bonta & Andrews, 2017). According to the risk principle, high risk individuals should receive more intensive interventions than lower risk individuals. According to the need principle, individuals should receive individualized interventions that target modifiable risk factors that contribute to offending (i.e., criminogenic needs), such as substance use treatment for an offender with substance use problems. Finally, according to the responsivity principle, individuals should receive cognitive-behavioral treatment delivered in the context of a strong therapeutic relationship, with treatment being tailored to individual characteristics such as strengths and culture.

Risk assessment tools are designed to help collect information that is relevant to the RNR principles. For instance, risk assessment tools, by definition, are designed to help generate estimates of individuals’ risk levels (Skeem & Monahan, 2011). In addition, some tools are designed to identify criminogenic needs, such as antisocial attitudes, and responsivity factors, such as strengths (Bonta & Andrews, 2017). By structuring the routine collection of this information, tools might conceivably enable better match to the RNR principles (e.g., Douglas & Kropp, 2002). For instance, by systematically assessing risk level, professionals may be able to make more appropriate decisions about whether individuals can be safely managed in the community rather than in a secure facility. Also, by identifying criminogenic needs, professionals might be more likely to target those needs in treatment.

In sum, tools are viewed as a means by which to deploy interventions that are evidence-based and individually-tailored, thereby avoiding a “one size fits all” approach. In this sense, many modern risk assessment tools arose from a “rehabilitative ideal” that rejected the claim that “nothing works” and instead emphasized the importance of effective, individualized treatments rather than harsh punishments (Maurutto & Hannah-Moffat, 2007, p. 470).

The Impact of Tools May Depend on the Nature of the Tool and Subsequent Interventions

Although many researchers believe that risk assessment tools can help manage and reduce risk, some researchers have noted that the impact of tools may be contingent upon other factors, such as the type of tool (e.g., Douglas & Skeem, 2005), and whether the tool is followed through on to match individuals to appropriate interventions (e.g., Bonta, Rugge, Scott, Bourgon, & Yessine, 2008). Indeed, tools vary in terms of the types of items they include, and how items are combined to form an overall conclusion (Skeem & Monahan, 2011). Whereas some tools focus primarily on historical or static factors, such as history of offending, other tools include dynamic or modifiable factors, such as anger management difficulties (i.e., criminogenic needs; Douglas & Skeem, 2005). In addition, on some tools (i.e., actuarial tools), evaluators add up risk factors to generate a total score (Skeem & Monahan, 2011). On other tools (i.e., structured professional judgement tools), evaluators use their discretion to make a separate summary risk rating after considering the items and additional case-specific considerations. Thus, tools may differ in their utility for various risk management decisions.

For instance, many authors have asserted that tools that include dynamic risk factors will
have greater utility for treatment-planning than tools that include primarily historical factors, as dynamic factors can serve as treatment targets (e.g., Douglas & Kropp, 2002; Douglas & Skeem, 2005; Hanson & Harris, 2000; Hart & Logan, 2011; Roychowdhury & Adhhead, 2014). However, historical factors may be relevant to other aspects of risk management, such as decisions about the intensity of supervision and the need for secure placements (Hart, Webster, & Douglas, 2001; Douglas & Kropp, 2002). Some authors have also argued that structured professional judgment tools may have greater utility for treatment-planning than actuarial tools because existing structured professional judgment tools tend to place a greater focus on dynamic factors than do many actuarial tools (e.g., Hart & Logan, 2011; Roychowdhury & Adhhead, 2014). That said, there nothing inherent in actuarial tools that necessarily precludes attention to dynamic factors (Douglas, Hart, Groscup, & Litwack, 2013; Skeem & Monahan, 2011). Some actuarial tools can and do include dynamic factors.

Besides variation in item content, risk assessment tools also vary in terms of their validity, or their ability to predict reoffending. Some tools have been found to significantly predict reoffending in multiple studies, with effect sizes falling in the moderate range (Hanson & Morton-Bourgon, 2009; Olver, Stockdale, & Wormith, 2009, 2014; Singh, Grann, & Fazel, 2011; Yang, Wong, & Coid, 2010). However, other risk assessment tools have poor predictive validity, or have not yet been tested at all. Presumably, unless a tool has adequate validity in predicting reoffending, it will have limited value for risk management.

Finally, tools vary in the extent to which they include a focus on risk management. Whereas most risk assessment tools do not include explicit instructions or guidance regarding how to manage risk, some tools aim to bridge risk assessment to risk management by providing greater structure and support for risk management, such as by including case management planning forms. Such tools are sometimes referred to as “fourth generation” tools to reflect the notion that that they represent an advancement over earlier tools (Bonta & Andrews, 2017).

Not only does the utility of risk assessment depend on the nature of tool, it could also depend on how professionals use and apply tools. In particular, as is true of any type of assessment, the value of risk assessment likely lies primarily in what happens after the assessment. Although risk assessment tools may be a starting point for treatment and risk management, they are not a treatment in and of themselves. As such, it may be unrealistic to expect that using a tool will help manage risk or reduce reoffending unless: (a) appropriate treatments are, in fact, available, and (b) professionals meaningfully apply tools to match individuals to these treatments.

For this reason, Vincent, Paiva-Salisbury, et al. (2012) note that “simply adopting a risk assessment tool is a necessary but insufficient method” for improving risk management (p. 553). Similarly, as Wong and Olver (2010) describe, violence risk assessment might help reduce the likelihood of violence if it occurs “in combination with interventions and treatment” (p. 121). Herein lies a potential barrier in the pathway between risk assessment and risk management. It is unclear to what extent tools are, in fact, optimally used to guide interventions. Some researchers have observed significant gaps between risk assessments and risk management efforts (Bonta et al., 2008; Peterson-Badali, Skilling, & Haqanee, 2015).
Some Have Questioned the Value of Tools for Risk Management

Despite the debate about whether some tools may be better-suited for risk management than others, most risk assessment researchers appear to believe that tools can, in principle, aid in risk management. However, this viewpoint is not embraced by everyone (e.g., Almond, 2012; Hannah-Moffat, Maurutto, & Turnbull, 2009; Harcourt, 2007; Large, Ryan, Callaghan, Paton, & Singh, 2014; Ryan, Nielssen, Paton, & Large, 2010; Rose, 1998; Szmukler & Rose, 2012; Starr, 2014). Some critics have expressed concern that risk assessment tools, in general, are not only ineffective in managing risk, they might even cause harm to patients and offenders (e.g., Large et al., 2014).

In particular, critics have pointed out that tools tend to have high rates of false positives, meaning that many offenders who are rated as high risk do not reoffend (e.g., Ryan et al., 2010). It is thought that this ‘risk inflation’ might lead to stigma, over-incarceration and warehousing of individuals who are high risk, and a decreased focus on rehabilitation and therapeutic efforts (Almond, 2012; Fitzgibbon, 2007; Rose, 1998; Ryan et al., 2010). At the same time, there is concern that the use of risk assessment tools might lead to an inappropriate withdrawal of resources from clients who are low risk, because these individuals may be falsely viewed as not needing treatment (Large et al., 2014; Rose, 1998; Ryan et al., 2010). Thus, in these critics’ views, risk assessment tools provide a “beguiling, but flawed, rationale for the distribution of resources” (Nielssen, Ryan, & Large, 2011, p. 270). This concern is compounded by the fact that assessments can be costly and could erode resources that might otherwise be available for treatment (Nielssen et al., 2011).

Furthermore, some critics have questioned the motives that underlie the adoption of risk assessment tools. Hannah-Moffat and colleagues (2009) argues that a primary reason that agencies use tools is not to reduce risk or protect others, but rather to protect themselves from lawsuits by creating a “paper trail” (p. 396). Similarly, Undrill (2007) describes risk assessment as a “neurotic organisational attempt to tame anxiety” (p. 294). In other words, there is concern that the use of tools puts the needs of the agency before those of the clients themselves.

Finally, critics have pointed out that despite claims that risk assessment tools help to manage risk, there is little evidence to support such assertions (e.g., Wand, 2012). For instance, in a survey of 1,937 psychiatrists, 87% of respondents endorsed the view that tools provide a false sense of security, as there is little direct evidence that tools help to reduce adverse events (Royal College of Psychiatrists, 2008). Some risk assessment researchers and tool developers have offered similar cautions. For instance, Rice, Harris, and Hilton (2010) noted that, “At this point, no one can promise that the introduction of any formal violence risk assessment (actuarial or nonactuarial) will improve real risk-related decisions. No one really knows what strategies in the application of available empirical evidence will ensure that avoidable violent recidivism and unnecessary restriction of offenders’ freedoms are simultaneously minimized” (p. 113).

The Present Study

In sum, opinions about the value of risk assessment tools for risk management efforts range considerably. Although some researchers view risk assessment tools as invaluable to risk
management, other scholars believe that such tools are ineffective at best and harmful at worst. Still others acknowledge both the promise and pitfalls of tools, noting that the value of tools for risk assessment might depend on factors such as the nature of the tool. As such, at the present time, it is unclear what conclusions are justified. Concerningly, many assertions about the utility of tools (or lack of utility) have been offered without reference to research findings.

One possible explanation for the lack of empirical grounding for these assertions is that relevant research simply does not exist, as of yet. However, another possibility is that existing research has not yet been adequately integrated into the literature due to a lack of comprehensive reviews. Indeed, although there have been numerous efforts to synthesize research on the predictive validity of risk assessment tools (e.g., Hanson & Morton-Bourgon, 2009; Olver et al., 2009, 2014; Singh et al., 2011; Yang et al., 2012), systematic reviews on the utility of tools for risk management are lacking.

To our knowledge, only one systematic review has examined the utility of risk assessment tools for risk management, and this review was not designed to focus on risk assessment per se. In particular, Gaynes et al. (2017) conducted a systematic review of strategies to reduce aggression among psychiatric patients in acute settings (e.g., staff training, medication protocols). As part of this review, they identified two randomized control trials (RCTs) which examined whether the implementation of risk assessment tools reduces aggression (i.e., Abderhalden et al., 2008; van de Sande et al., 2011). Gaynes et al. (2017) concluded that although these two studies provide some evidence that risk assessments may reduce aggression in acute psychiatric patients, the strength of this evidence is low due to limitations in research.

The goal of the present systematic review was to expand on Gaynes et al.’s (2017) review. First, rather than focusing exclusively on psychiatric patients in acute care settings, we included a range of populations (e.g., patients, offenders) and settings (e.g., jails, forensic hospitals). Second, rather than restricting our review to RCTs, our review encompassed a variety of designs (e.g., RCTs, pre-post studies). We adopted this broad lens so that we could determine the full scope of extant research. Third, instead of solely examining whether the use of risk assessment tools reduce violence, we also investigated their impact on professional practices. Certainly, violence reduction is one indicator of the success of a tool. For instance, Douglas and Kropp (2002) note that “risk assessment should be considered successful when we can demonstrate reduced rates of violence in connection with risk assessment procedures” (p. 623). However, violence reduction is not the only possible indicator of tools’ impact. If risk assessment tools do, indeed, reduce violence and offending, this relationship may be indirect, and contingent upon professionals’ risk management practices.

As such, we reviewed research on whether tools facilitate professionals’ adherence to the risk and need principles, as these are the hypothesized mechanism by which tools might help manage risk. We also examined whether professionals perceive tools as useful for risk management and whether they use tools to guide their risk management effort; tools are unlikely to be effective if professionals do not apply them or view them as useful. In other words, our goal was to understand, more thoroughly, the pathway between risk assessment and risk management. Furthermore, to develop an agenda for future research, we reviewed studies on strategies to enhance the utility of risk assessment tools for risk management, such as staff
training. Even if risk assessment tools do not have as positive an impact as desired, there may be means to improve their utility. As such, our research questions were as follows:

1. Do professionals perceive risk assessment tools to be useful for risk management?
2. When professionals conduct risk assessments with tools, do these assessments guide risk management efforts?
3. Does the use of risk assessment tools facilitate match to the risk principle?
4. Does the use of risk assessment tools facilitate match to the need principle?
5. Does the implementation of risk assessment tools reduce violence or offending?
6. What strategies help to improve the utility of risk assessment tools for risk management?

**Method**

**Design**

To examine our research questions, we chose to conduct a systematic review rather than a traditional literature review because systematic reviews are more transparent, comprehensive, and objective (Green, Johnson, & Adams, 2006). For instance, in a systematic review, authors provide a detailed account of their methodology (e.g., search terms, databases).

To synthesize findings, we used a narrative approach. An empirical synthesis (i.e., meta-analysis) was neither feasible nor appropriate because our review included a wide range of designs (e.g., RCTs, surveys), populations (e.g., offenders, patients), tools (i.e., 34 different risk assessment tools), and outcomes (e.g., supervision, dispositions; see Borenstein, Hedges, Higgins, & Rothstein, 2009; Popay et al., 2006; Ryan & Cochrane Consumers and Communication Review Group, 2013). In particular, to help determine if a meta-analysis was viable, we created a table for each research question that listed the tools in the row headings and the specific outcomes in the column headings. Most cells had only one or two studies listed, indicating that it would be difficult to meaningfully combine studies in an empirical manner. In addition to being more viable, a narrative approach enabled us to provide more detailed examples and descriptions of studies than is typical in an empirical synthesis. This was important because research in this area is at an early stage, and thus relatively little is known about these studies.

To ensure that our systematic review met best practice standards and followed relevant reporting guidelines, we followed the criteria set forth in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Liberati et al., 2009) and the Assessment of Multiple Systematic Reviews (AMSTAR) tool (Shea et al., 2007).

**Inclusion and Exclusion Criteria**

Studies were required to meet the following inclusion criteria: (1) empirical study that was published or disseminated in English; (2) included a sample of individuals who were assessed with a structured risk assessment tool in real-world practice, and/or a sample of professionals who used risk assessment tools in practice; and (3) included an outcome relevant to at least one of our research questions (e.g., perceived utility, adherence to the risk principle). We defined structured risk assessment tools as tools that included a designated list of risk factors and
an overall rating of risk level for violence or offending (Skeem & Monahan, 2011). Thus, we did not include measures of psychopathy. Also, we did not include qualitative research, as systematic reviews of qualitative studies use different methodologies, nor did we include research on the responsivity principle, as this principle encompasses a wide range of constructs (i.e., culture, trauma, mental health) and is not as well-researched as the risk and need principles. When disseminations were based on the same sample, we selected the study that was the most comprehensive and rigorous (e.g., largest sample).

Search Methods

Our search procedure is outlined in Figure 1. To select our search terms, we consulted with a research librarian and pilot tested 15 different combinations of terms. Our final set of search terms were as follows: "risk assessment" AND (violence* OR reoffense* OR recidivism OR offending*) AND ("risk management" OR utility OR useful OR "risk-need-responsivity" OR RNR or "decision-making" OR "case management"). Using these terms, we searched 13 databases, including the following academic databases: PsycINFO, PsycARTICLES, PsycBOOKS, MEDLINE, National Criminal Justice Reference Service, Criminal Justice Abstracts, Sociological Abstracts, Social Sciences Abstracts, Social Sciences Full Text, Social Services Abstracts, and Web of Science.

To reduce the likelihood of publication bias (Ioannidis, 2005; Kicinski, Springate, & Kontopantelis, 2015), we also searched databases that capture more of the gray literature (i.e., ProQuest Dissertations and Theses Global Database, Google Scholar). These searches encompassed January 1990 (i.e., an approximate date of when risk assessment tools were first developed) to the date of the search (September 30, 2016). Given that Google Scholar searches generate an unwieldy number of hits, researchers typically restrict their screening to the first 50 to 100 search records when conducting systematic reviews (Haddaway, Woodcock, Macura, & Collins, 2015). However, we adopted a more comprehensive approach and examined the first 300 Google Scholar search records. Only 15 new studies were screened in through our Google Scholar search and only one of these met inclusion criteria, which suggests that we successfully identified most of the relevant studies through the other databases. Besides searching databases, we reviewed reference lists of relevant studies and emailed 15 authors with multiple studies in this area (i.e., two or more studies included in this review); these searches were completed August 31, 2017. Included studies are listed in Appendix A.

Data Collection

To collect data, we used a six-stage process, and followed the guidelines and examples provided by the Cochrane Collaboration (e.g., Johnson, Sandford, & Tyndall, 2003; Ryan & Cochrane Consumers and Communication Review Group, 2013) and the PRISMA statement (Liberati et al., 2009). In Stage 1, the authors screened the abstracts of the 1,588 articles identified from our searches to determine if they were relevant (see Figure 1). To check the accuracy of our screening, each author screened 75 abstracts for practice; they correctly screened in each of the eligible studies. In Stage 2, the study authors examined full copies of each of the 236 screened-in abstracts to determine if the study met inclusion criteria.
In Stage 3, data from the 73 studies that met inclusion criteria was extracted using a data extraction form (described below). We also appraised each study’s limitations. To increase objectivity and replicability, each study was coded by two raters (i.e., study authors [DC, MJ] and one trained research assistant). If the dissemination did not include adequate information (e.g., the type of tool was not adequately specified), we contacted the authors for clarification.

In Stage 4, the raters discussed each study’s overall findings and study appraisal to reach consensus; the first author, who reviewed all studies, provided an opinion when consensus could not be reached. In Stage 5, the first author compiled ratings in evidence tables (i.e., tables that summarized the methodology and findings of each study; Green et al., 2006; Popay et al., 2006) and prepared narrative descriptions of the results. These tables and results were reviewed and checked by the other authors, and a second set of consensus meetings were held to finalize ratings. Finally, in Stage 6, each of the authors graded the overall strength of evidence for each research question and arrived at a consensus.

Materials

**Data Extraction Form.** Our data extraction form included approximately 150 questions, divided into sections that pertained to each of our research questions. Extracted data included: study information (e.g., publication type), sample (e.g., gender), method (e.g., design), risk assessment tool, outcome (e.g., violence), results, and follow-up information (e.g., new studies to screen). We developed this form by reviewing other data extraction forms (e.g., Guy, 2008). We then pilot tested the form with six cases, and adjusted the form accordingly. Cohen’s (1960) kappa coefficients for setting (i.e., justice setting, mental health or forensic setting), gender of sample (i.e., male only, female only, mixed), age of sample (i.e., adult, adolescent, mixed), type of professionals (i.e., clinicians, probation officers, police, other), and study design (i.e., RCT, pre-post, post-test only, survey, other) were 1.00, 0.90, 0.90, 0.89, and 0.82, respectively (n = 73 cases). Thus, agreement fell in the “almost perfect” range (i.e., > .80; Landis & Koch, 1977).

In addition to rating study characteristics, we made summary ratings for each study to help interpret findings. Interrater reliability of summary ratings were assessed using intraclass correlation coefficients ([ICCs], two-way mixed, absolute agreement, average measures; Hallgren, 2012; McGraw & Wong, 1996). For research questions 1 – 4 (i.e., perceived utility, use of tools for risk management, match to the risk and need principles), we made summary ratings of low, mixed/moderate, or high (operational definitions are provided in the Results section and in the tables; ICC = 0.92, n = 82). For research question 5 (i.e., violence/reoffending), we made summary ratings of decrease, mixed, or no change (ICC = 1.00, n = 12). For research question 6 (i.e., strategies to improve utility), we made summary ratings of improvement, some improvement, or no change (ICC = 0.89, n = 8). These ICC values were all in the “excellent” range (i.e., > .75; Cicchetti, 1994).

**Study Appraisal.** We appraised the limitations of RCTs using the Cochrane Risk of Bias tool for RCTs (Higgins et al., 2011). This tool examines various forms of bias (e.g., bias arising from the randomization process, measurement of the outcome). Raters then assessed overall risk of bias as low, medium, or high. ICCs were not calculated because only four studies were RCTs and there was limited variability in ratings. However, to check our ratings, we
We independently rated the two RCTs that were included in Gaynes et al. (2017) and then compared our ratings to Gaynes et al.’s ratings to ensure that they were consistent.

We appraised the limitations of observational studies and non-controlled intervention studies (e.g., pre-post studies) using the following items: sample, adequacy of comparison group, adherence to the tool, measurement approach ([a] description, [b] support for measurement approach). These items are based on the National Institute for Health and Care Excellence [NICE] tool (2012), but the wording was adjusted for this context (e.g., replacing the term “intervention” with the term “risk assessment tool”). To help structure ratings, we developed anchors. Each item was rated yes, possibly/partially, no or not reported, or not applicable. Then, an overall judgement was made (i.e., low, medium, or high study limitations). The ICC (mixed, absolute agreement, average measures) for overall judgment was .73 (n = 68).

Finally, we appraised survey studies using the following items from the Center for Evidence-Based Medicine’s Critical Appraisal of Surveys (N.D.): sample, response rate, and measurement approach ([a] description, [b] support for measurement approach). Once again, each item was rated yes, possibly/partially, no or not reported, or not applicable, and an overall judgement was made (i.e., low, medium, or high study limitations). The ICC for overall judgment was .92 (n = 20).

**Overall Strength of Evidence for Outcomes.** In addition to evaluating individual studies, we rated the overall strength of evidence for our research questions (e.g., whether tools facilitate match to the risk principle). To do so, we used the system developed by the Evidence-Based Practice Center of the Agency for Healthcare Research and Quality (AHRQ; Berkman et al., 2015). The AHRQ system is one of the most common methods for rating quality of evidence (Kane, Butler, & Ng, 2016), and it draws heavily from the GRADE framework (Grading of Recommendations Assessment, Development, and Evaluating Work Group; Guyatt et al., 2011). Within the AHRQ system, raters evaluate a body of research in five domains, including study limitations (the overall quality of study designs and risk of bias), consistency (whether different studies find the same pattern of results), directness (whether the evidence directly links the intervention to the outcome), precision (the degree of certainty of the estimated effect), and reporting bias (publishing or reporting findings based on their favorability). Then, they grade the overall quality of evidence as high, moderate, low, or insufficient. We did not complete AHRQ ratings for studies on perceived utility or use of tools for risk management, as the AHRQ system was not designed for survey research.

**Results**

In total, 73 studies met inclusion criteria (see Figure 1 and Appendix A). Sixteen of these studies were unpublished; nine of the unpublished studies were dissertations or theses and the remaining seven were reports by researchers, government, or other organizations. These studies included 31,551 offenders or patients, 10,002 professionals, and 34 risk assessment tools. Most tools included dynamic or modifiable factors (i.e., criminogenic needs; 76.5%, k = 26), and were validated (i.e., have been found to significantly predict violence or reoffending; 79.4%, k = 27). However, the degree of research support for tools ranged from single studies to large meta-analyses. Appendix B provides a brief description of the tools.
1. Do Professionals Perceive Risk Assessment Tools to be Useful for Risk Management?

**Study Characteristics.** In 12 studies (six of which were unpublished), researchers surveyed or interviewed professionals about the utility of risk assessment tools for risk management ($N = 6,664$ professionals; see Table 1). These studies examined the extent to which professionals considered tools to be useful for management-related decisions, such as placement and supervision decisions (66.7%, $k = 8$), and/or treatment-related decisions, such as service referrals (33.3%, $k = 4$). Some studies involved probation or parole officers (POs) or other correctional staff (50.0%, $k = 6$), whereas other studies involved mental health clinicians (e.g., psychologists; 41.7%, $k = 5$). Also, a couple of studies involved judges (16.7%, $k = 2$). In most studies, researchers examined tools that have been found to significantly predict offending (83.3%, $k = 10$; see Appendix B). However, in three studies, the authors surveyed professionals about tools in general rather than asking about a specific tool per se. Most of the examined tools included dynamic or modifiable factors (66.7%, $k = 7$). However, in two studies, researchers examined the VRAG, a tool which focuses on historical or static factors.

**Study Appraisal.** Five studies (41.7%) were rated as having low limitations overall, whereas the remaining studies had medium (41.7%, $k = 5$) or high limitations (16.7%, $k = 2$; see Table 7). Of the specific domains that were appraised (sample representativeness, response rate, description of measure, and support for measure), the most common limitation was sample representativeness. Only four of the 12 studies (33.3%) included clear evidence to support the representativeness of their sample. This means that the results may not generalize to the population of tool users; for instance, survey respondents may have had more positive or negative views about the perceived utility of tools than typical users. In addition, few studies (33.3%, $k = 4$) included clear support for the approach that was used to measure perceived utility, such as evidence of pilot testing or expert feedback.

**Results.** In most studies (58.3%, $k = 7$), perceived utility was mixed, mixed-low, or mixed-high (see Tables 1 and 7). A rating of mixed was given when professionals rated usefulness as being just above the midpoint, such as a mean of 5 to 7 out of 10 (see Table 1 for rating criteria). For instance, in one survey, corrections staff rated the usefulness of the Y-LSI for identifying treatment needs as 5.5 on a 10-point scale, with 10 being the most positive rating (Flores, Travis, & Latessa, 2004); they rated its usefulness for making placement decisions as 5.2 out of 10. In another study, forensic staff rated the perceived utility of the VRAG for determining transfers to be 7.8 out of 10 (with 10 being very important; McKee, Harris, & Rice, 2007). However, they rated other factors, such as severity of the index offense, to be more important than the VRAG. In a couple of studies (16.7%, $k = 2$), perceived utility was low (i.e., helpfulness was rated below the midpoint). For instance, in a survey of psychiatrists, only 48% of respondents reported that tools lead to better decisions (Royal Colleges of Psychiatrists, 2008). Conversely, in a couple of studies (25.0%, $k = 3$), perceived utility was high (i.e., most participants viewed the tool as helpful). For example, in one study, 78% of clinicians reported that the VRAG was very helpful for making discharge decisions (Muheizen, 2014).

**Potential Moderators.** We did not find evidence of publication bias (i.e., wherein unpublished studies are less positive). However, there was some evidence that the perceived utility of tools varied somewhat by tool. For instance, in a large international survey, Singh,
Desmarais, Hurducas, and colleagues (2014) found that tools with the highest perceived utility for risk management were the FORTES, HKT-30, HCR-20, SAPROF, START, and VRS; these tools all include dynamic risk factors. In contrast, the tool with the lowest perceived utility was the VRAG; this tool focuses on historical factors and/or factors which are assumed to be stable. However, in other studies, tools like the VRAG were considered helpful (McKee et al., 2007; Muheizen, 2014). There were also some differences across professionals. Specifically, judges rated structured decision-making tools as being more useful for post-adjudication placement than did POs (Shook & Saari, 2007). Finally, although this was not directly tested, there was some indirect evidence that perceived utility may be higher when tools are selected or designed with staff input, and fit the specific needs of sites (e.g., are relevant to the population served at the site). For instance, in one study, an agency developed a site-specific risk assessment tool through staff consultation (Richardson, 2009); clinicians viewed the resulting tool as very useful.

**Summary.** In sum, studies revealed a range in views about the utility of tools for risk management; some professionals view these tools as useful whereas others do not (see Table 7). However, even if professionals view tools as useful, they may not adequately apply tools to their decision-making. Therefore, as a next step, we reviewed studies that examined the use of risk assessments for risk management.

2. **When Professionals Conduct Risk Assessments with Tools, Do These Assessments Guide Risk Management?**

**Study Characteristics.** We identified 14 studies (including three unpublished studies) in which researchers examined the extent to which findings from risk assessments tools guide or influence risk management efforts (see Table 2; N = 5,628 professionals). These studies examined use for management-related decisions (e.g., determining supervision level; 42.9%, k = 6), treatment-related decisions (e.g., measuring treatment progress; 35.7%, k = 5), and/or general use (e.g., developing “risk management plans,” 28.6%, k = 4). Most of the studies were surveys of professionals at agencies in which tools had been implemented (64.3%, k = 9; i.e., professionals’ self-reported use). However, in a couple of studies (35.7%, k = 5), researchers investigated the use of risk assessment tools through observational methods, such as by examining whether judges referred to tools in their written legal decisions. Many studies focused on POs or other correctional staff (64.3%, k = 9), but some studies included clinicians (21.4%, k = 3), judges (21.4%, k = 3), and/or jurors (14.3%, k = 2). In most studies, researchers examined tools which have been found to significantly predict offending and/or violence (64.3%, k = 9; see Appendix B), but in three studies, researchers examined tools in general rather than a specific tool, or did not specify the name of the tool, and thus validity could not be determined. Many of the examined tools include a focus on dynamic factors (64.3%, k = 9). However, two studies examined tools that focus on historical or static factors (i.e., Static 99, MnSOST-R).

**Study Appraisal.** Although six studies (42.9%) were rated as having low study limitations, the remaining studies had medium (35.7%, k = 5) or high limitations (21.4%, k = 3; see Table 7). Of the specific domains that were appraised, the most common limitation was with respect to research support for measurement approaches; only two of the 14 studies (14.3%) included clear evidence that the approach that they used to measure professionals’ use of tools was reliable or valid. As such, these studies may under- or over-estimate the use of tools.
**Results.** In most studies (57.1%, k = 8), the application of risk assessment tools to risk management efforts was mixed, mixed-low, or mixed-high (see Tables 2 and 7). A rating of mixed was given when helpfulness was rated as just above the midpoint (see Table 2 for rating criteria). As an example, Miller and Maloney (2013) classified 40% of tool users as bureaucratic compliers, meaning that although they completed risk/needs assessment tools for assessment purposes, they were less likely to apply these tools to guide their decision-making. Also, in a case law review of adolescent offender cases, judges placed great weight on the risk assessment in 4% of cases, and some weight on it in 58% of cases (Urquhart & Viljoen, 2014). Two studies (14.3%) reported high use of tools for risk management where in most professionals (e.g., over 70%) used the tool for risk management. For instance, Vincent, Paiva-Salisbury, et al. (2012) found that, following the implementation of the SAVRY and YLS/CMI, 80% of youth probation officers (YPOs) reported that they used the tool in service decisions and 90% used it in supervision decisions.

In contrast, in four studies (28.6%), the use of tools for risk management was low (i.e., less than half of professionals used the tool for risk management). In all but one of these studies, the use of tools was measured via observation or written records, suggesting that observation yields more modest estimates of use than do professionals’ self-reports about their use of tools. For instance, in an observational study of 69 PO–probationer interactions, POs only used the tool for case management purposes on one occasion (Viglione, Rudes, & Taxman, 2015). In another study, professionals directly referred to clients’ risk assessments in only 8% of plans (Kewley, Beech, Harkins, & Bonsall, 2015). Also, in a study of sexually violent predator trials, jurors’ perceptions of risk were not significantly associated with respondents’ Static-99 or MnSOST-R risk scores (Boccaccini, Turner, Murrie, Henderson, & Chevalier, 2013).

**Potential Moderators.** Results of published and unpublished studies were quite similar. However, a couple of studies found that the use of risk assessment tools for risk management may vary by profession (e.g., Haas & DeTardo-Bora, 2009; Shook & Saari, 2007). For instance, in one study, nurses were significantly more likely than doctors to report that risk assessments influenced their judgment or therapeutic actions (Hawley, Gale, Sivakumaran, & Littlechild, 2010). The use of tools for risk management may also vary depending on the particular task. For example, Haas and DeTardo (2009) found that 80% of professionals reported that they used the LSI-R to guide service referrals, but only 42% used it to develop reentry plans. In addition, some tools, such as the HCR-20, appear to be used for risk management more frequently than other tools (Singh, Desmarais, Hurducas, et al., 2014), possibly because professionals view them as more relevant to risk management.

**Summary.** Overall, this research indicates that although tools guide decisions in some contexts, “slippage” often occurs between assessments and risk management (Peterson-Badali, et al., 2015; see Table 7). Thus, we subsequently investigated, more specifically, how the use of tools impacts match to the risk and need principles, as this is the hypothesized mechanism by which tools are thought to improve risk management.

3. **Does the Use of Risk Assessment Tools Facilitate Match to the Risk Principle?**

**Study Characteristics.** According to the risk principle, higher risk individuals should
receive more intensive risk management strategies than lower risk individuals. We identified 36 studies (including nine unpublished studies) in which researchers examined match to the risk principle following the use of risk assessment tools; these studies included a total of 17,597 assesses (e.g., offenders, patients) and 2,507 professionals (see Table 3). Studies examined a variety of risk management decisions, including decisions about placements (e.g., incarceration, 25.0%, \( k = 9 \)), other sentencing outcomes (e.g., dangerous offender designation, 5.6%, \( k = 2 \)), supervision levels (22.2%, \( k = 8 \)), and release or discharge (16.7%, \( k = 6 \)). In addition, several studies examined decisions about the overall level or intensity of services (e.g., number of referrals, 16.7%, \( k = 6 \)) or the number of strategies included in case management or risk management plans (25.0%, \( k = 9 \)); overall intensity of services is an aspect of match to the risk principle (Bonta & Andrews, 2017).

Many studies were with offenders (80.6%, \( k = 29 \)), and approximately half were with adult samples (47.2%, \( k = 17 \)). In most studies (80.6%, \( k = 29 \)), researchers measured match by examining associations between risk scores and the intensity of risk management strategies (e.g., correlations between risk total scores and decisions to incarcerate). However, in a couple of studies (8.3%, \( k = 3 \)), researchers rated the proportion of individuals who received an acceptable intensity of risk management strategies, as defined by the researchers or policy guidelines.

Many studies examined tools which have demonstrated predictive validity (83.3%, \( k = 30 \); see Appendix B), but we were unable to identify evidence of empirical support for the predictive validity of a couple of tools (i.e., AJADMS, DSI, L17A, RAI). Most tools included dynamic or modifiable factors (80.6%, \( k = 29 \)). However, six studies examined tools that focus on static or historical factors (i.e., AJADMS, DSI, RAI, Static-99, VRAG).

**Study Appraisal.** Most studies had medium limitations (66.7%, \( k = 24 \); see Table 7). Of the specific domains that were appraised (sample representativeness, comparison group, adherence to tool, measurement of need principle), the most common limitation was a lack of comparison groups. Only two of the 36 studies (5.6%) included a comparison group of individuals who were not assessed with a tool. As such, most studies were not designed to directly test whether the use of tools increases match to the risk principle per se. Also, many studies did not include evidence of adherence to the tools (e.g., that professionals completed tools as mandated); only 14 of the studies (38.9%) fully met this criterion.

**Results.** Based on the two studies that included comparison groups, there is some evidence that match to the risk principle may be better when tools are used than when they are not (see Tables 3 and 7). Bonta and Motiuk (1990) found that low risk inmates at jails that used the LSI were more likely to be transferred to a halfway house than were low risk inmates at a jail that did not use the LSI. In addition, Vincent, Paiva-Salisbury, et al. (2012) found that YPOs were more likely to identify risk level as an important consideration in their supervision decisions following the implementation of the SAVRY and YLS/CMI rather than prior to the implementation of tools. However, their reported use of risk level in service and disposition decisions did not significantly increase following implementation.

Although the remaining 34 studies did not have a comparison group of individuals who were not assessed with a tool, they nevertheless provide descriptive information regarding match
to the risk principle following the use of a tool. In most of these studies (64.7%, $k = 22$), match to the risk principle was moderate following the use of risk assessment tools (i.e., significant associations between risk total and intervention intensity which fell in the small or moderate range; see Table 4 for rating criteria). Furthermore, in five studies (14.7%), match to the risk principle following the use of a tool was high. For instance, in one study, offenders with a life sentence were significantly less likely to be granted parole if they were rated as high risk than if they were rated as low risk with the HCR-20 and LS/CMI (Guy, Kusaj, Packer, & Douglas, 2015). Not only were high risk individuals given more restrictive penalties, on average, they were also referred to more services, and professionals recommended that they receive more risk management strategies than low risk individuals. For example, in one study, police officers recommended a median of four risk management strategies for intimate partner violence offenders who were rated as high risk on the SARA, compared to a median of two strategies for offenders rated as low risk (Belfrage et al., 2012). Whereas match to the risk principle was generally moderate or high following the use of tools, in seven studies (20.6%), it was low. For instance, in a couple of studies, VRAG scores failed to predict tribunal decisions (Hilton & Simmons, 2001; McKee et al., 2007).

**Potential Moderators.** Overall, the findings in the published and unpublished studies were similar. There was some evidence that risk assessments might have a greater influence on the behavior of individuals who conducted the risk assessments themselves (e.g., clinicians) than on the behavior of external consumers of risk assessments (e.g., review boards who were provided with clinicians’ risk assessments). For instance, in McKee et al. (2007), VRAG scores predicted psychiatrist recommendations and team recommendations, but not tribunal decisions. In other words, the influence of risk assessments may be weaker when assessors have an intermediary role rather than a direct influence on risk management.

Match to the risk principle may also vary depending on the characteristics of the individuals being assessed, such as gender. However, results are unclear. Storey and Strand (2013) found lower match to the risk principle for female than male offenders, but Singh, Desmarais, Sellers, et al. (2014) found higher match for females, and Holloway (2015) found no significant gender differences. Finally, match may vary based on the nature of the decision. For instance, whereas studies generally found a high match in planned risk management strategies, match was lower for studies that examined actual number of supervision contacts (Auditor General of Ontario, 2012) or services received (McCormick, Peterson-Badali, & Skilling, 2017). This slippage could be due to various factors; for instance, professionals may not follow through on their recommendations and/or high-risk individuals might have lower engagement and attendance.

**Summary and Overall Strength of Evidence.** Overall, the research in this area suggests moderate match to the risk principle following the use of risk assessment tools. However, based on the AHRQ system (Berkman et al., 2015), there was insufficient evidence to determine if tools improve match, as only two studies directly tested this (see Table 7). Next, we examined match to the need principle.

4. **Does the Use of Risk Assessment Tools Facilitate Match to the Need Principle?**
Study Characteristics. According to the need principle, risk management strategies should target criminogenic needs that contribute to each individual’s offending (i.e., antisocial personality pattern, procriminal attitudes, procriminal associates, substance abuse, family/marital, school/work, leisure/recreation; Bonta & Andrews, 2017). We identified 17 non-overlapping studies (including five unpublished studies) in which researchers examined match to the need principle following the use of risk assessment tools; these studies included a total of 8,202 assesseses and 1,278 professionals (see Table 4). In most studies (58.8%, \( k = 10 \)), match was rated at the stage of the case management or intervention plan (e.g., services planned), such as whether professionals mentioned in their plans that they intended to address specific needs that were identified in the risk assessment (e.g., stating that they would refer a person with substance use difficulties to substance use treatment). However, in some studies, researchers examined whether assesseses actually received these services (23.5%, \( k = 4 \)), such as whether people who were identified as having substance use difficulties obtained substance use treatment. Finally, in two studies, researchers examined audiotaped PO-offender probation sessions to determine whether POs discussed and addressed needs that were identified in the risk assessment in their probation sessions with offenders (11.8%, \( k = 2 \)). Most identified studies were with offender populations (\( n = 13 \)); only one study was with psychiatric patients. Eight studies (47.1%) focused on adolescents (i.e., 18 or younger).

All of the studies examined tools which include dynamic or modifiable factors (100%, \( k = 17 \); see Appendix B). For instance, eight studies included versions of the LSI (e.g., LSI-R, YLS/CMI). In addition, most of the included tools have been found to significantly predict offending and/or violence (82.4%, \( k = 14 \)). In two studies, the name of the tool was not reported or the researchers examined tools in general rather than a specific tool.

Study Appraisal. Most studies had medium (64.7%, \( k = 11 \)) or high (29.4%, \( k = 5 \)) limitations (see Table 7). The most common limitation was a lack of comparison groups. None of the studies had a comparison group of individuals who were not assessed with a risk assessment tool. Also, given the lack of standardized approaches to measure match to the need principle, researchers often developed their own measurement approaches. Many of these measurement approaches had limited or no research support (64.7%, \( k = 11 \)), and descriptions of measures were often vague or missing (64.7%, \( k = 11 \)). For instance, in many studies, researchers rated match as adequate or inadequate using criteria that they developed (70.6%, \( k = 12 \)); inferential statistics were rarely used (11.8%, \( k = 2 \)). In addition, there often was little information about adherence to tools, such as whether professionals received training on the risk assessment tool (64.7%, \( k = 11 \)).

Results. In nearly half of the studies (41.2%, \( k = 7 \)), overall match to the need principle was low (e.g., less than half of needs were addressed; see Tables 4 and 7). In one study, for instance, only 1 to 5% of youths with high scores in Y-LSI family, education, drug, or peer domains received interventions in these domains (Flores et al., 2004). In another study, an average of 35% of youths’ needs were addressed in probation services (Vieira, Skilling, & Peterson-Badali, 2009). In almost all remaining studies (52.9%, \( k = 9 \)), match to the need principle was “moderate” following the use of risk assessment tools. However, we defined moderate fairly leniently to mean slightly more than half of needs [50-69%] were addressed (see Table 4 for rating criteria). For instance, in one study that was rated as showing moderate match,
intervention plans addressed 52% of youths’ critical needs on the START:AV (Singh, Desmarais, Sellers, et al., 2014). There was only one study (5.9%) in which match to the need principle study was rated as high, and that study relied on POs’ self-report. In that study, POs self-reported that they frequently targeted needs in their case management efforts (Miller & Maloney, 2013).

**Potential Moderators.** The results of published and unpublished studies were comparable. However, there was some evidence that rates of match to the need principle may be higher in initial decisions, such as intervention plans, than in later decisions, such as the actual receipt of services. For instance, Peterson-Badali et al. (2015) found that, when youth had needs in areas such as family or education/employment, clinicians commonly mentioned these needs in the assessment reports that they wrote for POs. However, after receiving these reports, POs addressed only 31% of youths’ needs through the provision of services.

Match to the need principle may also vary depending on how the need principle is defined. For instance, some researchers interpreted over-servicing as a poor match to the need principle (i.e., providing low-risk individuals more services than needed; Luong & Wormith, 2011), whereas other researchers focused solely on underservicing (i.e., providing high-risk individuals fewer services than needed; Peterson-Badali et al., 2015). Similarly, some researchers rated any intervention that targeted a need as an adequate match (e.g., court order to refrain from substance use; Flores et al., 2004), whereas other researchers rated a match as adequate only if the intervention was adequate in quality and intensity (i.e., regular attendance at an evidence-based substance use program; Peterson-Badali et al., 2015). Finally, some needs may be addressed more commonly than others. However, patterns differed considerably across studies. For example, in some studies, rates of match to family needs was over 70% (Holloway, 2015; Luong & Wormith, 2011). In other research, it was 10% or less (Flores et al., 2004).

**Summary and Overall Strength of Evidence.** Overall, research suggests limited match to the need principle following the use of risk assessment tools. However, it is unclear if this is better, worse, or the same as what it would have been had a tool not been used, as none of the studies included a comparison group. As such, based on the AHRQ system (Berkman et al., 2015), the strength of evidence to determine if tools improve match was deemed insufficient (see Table 7). Whereas our findings, thus far, have focused on the impact of tools on professionals’ behaviors, our next question focused on the impact of tools on assesses’ behaviors, namely whether tools reduce violence and offending among people in mental health and justice settings.

5. **Does the Implementation of Risk Assessment Tools Reduce Violence or Offending?**

**Study Characteristics.** After removing overlapping studies, 12 studies examined the association between risk assessment tools and either violence (91.7%, k = 11) and/or general offending (33.3%, k = 4). Most studies were published (83.3%, k = 12). Studies included a total of 7,350 patients or offenders (see Table 5). Many of the studies were with general psychiatric or forensic psychiatric populations (66.7%, k = 8), primarily adults (66.7%, k = 8). Some studies had rigorous designs. Specifically, three studies (25.0%) used cluster RCTs, wherein wards were randomly assigned to an intervention group (i.e., a risk assessment tool was used), or a comparison group (i.e., a tool was not used). The remaining studies used a non-randomized
design, such as a pre-post design (66.7%, \( k = 8 \)), or a non-randomized comparison group design (8.3%, \( k = 1 \)). However, two of these non-randomized studies used propensity score matching to help ensure that the intervention and comparison groups were similar in characteristics such as race, gender, and offense history (i.e., Guy, Vincent, Grisso, & Perrault, 2015; Vincent, Guy, Perrault, & Gershenson, 2016). Nearly all studies examined tools which include dynamic or modifiable factors (91.7%, \( k = 11 \); see Appendix B), but one study examined the PSA, a tool which has only historical factors. In addition, most of the included tools have been found to significantly predict offending and/or violence (83.3%, \( k = 10 \)). However, we were unable to find support for the predictive validity of two tools (i.e., CANVAS, PSA). The findings from the cluster RCTs and non-randomized designs are presented separately, consistent with guidelines (Berkman et al., 2015).

A. Cluster RCTs.

**Study Appraisal.** All of the cluster RCT studies were rated as having medium risk of bias (see Table 7). A common limitation was that, in each of the studies, professionals who conducted the risk assessment also collected data on violence or reoffending, thus creating the possibility of bias. Adherence to the tool was often unclear or inadequate. For instance, in Troquete et al. (2013) only 65% of patients who were supposed to be assessed with the START were assessed. In addition, despite randomization, the groups were not fully comparable in two studies. For example, in Abderhalden et al. (2008), patients in the wards that implemented the BVC-CH had significantly higher baseline levels of aggression than the comparison wards.

**Findings.** Each of the cluster RCTs involved tools that have support for their predictive validity (see Appendix B). In two of the three cluster RCTs (66.7%), the use of the risk assessment tool was associated with declines in aggressive incidents (see Tables 5 and 7). Both of these studies used the BVC, either alone (Abderhalden et al., 2008) or as part of a larger battery of tools (van de Sande et al., 2011). However, in one RCT (33.3%), findings were non-significant (Troquete et al., 2013). Given that this study found low adherence to the tool (i.e., the START), the authors reanalyzed results with only those who received the START. However, this did not alter the findings and, as such, the results do not appear fully attributable to poor adherence.

**Potential Moderators.** In two of the cluster RCTs, the tool was implemented alongside other risk management strategies (e.g., guidelines, policies). Although it was not directly tested, these strategies may have influenced results. Specifically, in Abderhalden et al. (2008), nurses were provided with a list of preventive measures for patients who scored above 7 on the BVC (e.g., relaxation exercise, one-to-one observation) and encouraged to consult with a multidisciplinary team for patients who scored above 10. Although this study reported decreases in aggressive incidents, it is unclear if this positive result would have been achieved had these guidelines not been in place. In Troquete et al. (2013), a process of joint care planning was introduced at the same time the START was implemented, wherein patients and case managers jointly developed a plan to address agreed upon treatment targets. The authors noted that the joint care plan may have led to a greater focus on treatment needs that may not be directly relevant to the reduction of violence, which could, in turn, have contributed to their null results.
Summary and Overall Strength of Evidence. Based on the AHRQ system (Berkman et al., 2015), there is insufficient evidence to draw a conclusion regarding whether tools reduce violence or offending because existing findings are mixed (see Table 7). Although RCTs should be given more weight than non-randomized designs (Oxford Centre for Evidence-based Medicine [OCEBM] Levels of Evidence Working Group, 2011), we also reviewed findings from non-randomized studies to determine if these studies yield similar results ($k = 9$).

B. Non-Randomized Studies.

Study Appraisal. Most of the non-randomized studies had moderate (55.6%, $k = 5$) limitations (see Table 7). Of the specific domains assessed, the most common limitation was with respect to comparison groups. Although all studies had comparison groups, these comparison groups were less than ideal (i.e., seven of the nine studies [77.8%] did not fully meet this criteria). For instance, studies generally did not test if the baseline risk levels of the pre- and post-samples differed, thus creating a possible confound; if individuals in the post-implementation sample were lower risk than those in the pre-implementation sample, it may appear as though the tool led to decreased offending when, in fact, this finding was due to differences in initial risk levels.

Findings. In two studies (22.2%), rates of violence or reoffending decreased following the implementation of a risk assessment tool (see Tables 5 and 7). For instance, in a study by the Laura and John Arnold Foundation (2014), rates of pretrial crime decreased by 15% following the implementation of a brief pretrial risk assessment tool, the PSA. No inferential tests were provided, and we were unable to find evidence for the predictive validity of this tool. Also, this study did not control for confounds. Thus, it is possible that the decrease in pretrial crime may be due to other factors, such as general decreases in crime rates, rather than the tool per se. Indeed, it may not be realistic to expect that a non-validated nine-item tool, focused on offense history, could directly reduce crime in and of itself.

In six studies (66.7%), there was no overall change in violence or reoffending when a risk assessment tool was implemented (i.e., violence or offending did not significantly change in all or almost all analyses). For instance, Vincent, Guy, Perrault and Gershenson (2016) conducted a pre-post study (using propensity score matching) at six adolescent probation offices. At five of the six sites, new petitions (i.e., charges) did not significantly decrease following the implementation of the SAVRY or YLS/CMI, even though these tools have been well-validated and include an emphasis on dynamic factors (see Appendix B).

In one study (11.1%), the findings were mixed. In particular, Guy, Vincent, Grisso, and Perrault (2015) compared sites that used the SAVRY (i.e., experimental condition) to sites that did not use any risk assessment tool (i.e., comparison condition). They found that adolescents in the experimental condition were less likely than those in the comparison condition to receive a new petition for a violent offense. However, there were no significant differences in new petitions for any offense, non-violent offenses, status offenses, or probation violations.

Potential Moderators. There was some evidence that risk assessment tools may have a larger impact in settings in which pre-existing base rates of violence are high. Specifically, in
one study, the implementation of a risk assessment protocol (i.e., the Alert System) was associated with larger initial reductions in violence in higher risk departments (e.g., psychiatry, emergency) than in lower risk departments (Kling, Yassi, Smailes, Lovato, & Koehoorn, 2011). Although not tested directly, the impact of risk assessment tools may also depend on the availability of appropriate treatment. As some authors noted, even if an individual was assessed to be high risk, it did not guarantee that they received the necessary treatment (Kling et al., 2011; Troquete et al., 2013). Finally, the impact of tools may be affected by jurisdictional policies and how offending is measured. For instance, Vincent et al. (2016) found that, at one site, new adjudications (i.e., convictions) significantly increased following the implementation of a tool, although new petitions (i.e., charges) did not. The authors explained that this might be due to a new district attorney, whose platform was to toughen sanctions for crime (Vincent, Guy, et al., 2012). In other words, the increase in new adjudications could reflect increased pressure to adjudicate adolescents rather than changes in actual rates of reoffending.

Summary and Overall Strength of Evidence. Similar to findings from the RCT studies, there was insufficient evidence from non-randomized studies to conclude that risk assessment tools reduce violence and offending (see Table 7). In sum, our results thus far indicate that use of risk assessment tools does not consistently improve assessees’ outcomes, nor does it consistently improve professionals’ risk management practices. Thus, as a final research question, we examined strategies that may help to rectify gaps between risk assessment and risk management.


Study Characteristics. We identified eight published studies (N = 2,617 offenders or patients, and 102 professionals; see Table 6), in which researchers examined strategies to improve the ability of risk assessment tools to either (a) facilitate adherence to the risk and/or need principles (75.0%, k = 6), and/or (b) enhance their effectiveness in reducing violence or offending (25.0%, k = 2). Three studies examined training (i.e., tools alone vs. tools plus training), three studies examined risk management guidelines (i.e., tools alone vs. tools plus guidelines), and two studies examined the impact of implementation strategies (e.g., policies). Five studies focused on POs or YPOs, whereas the other studies involved police and nurses. Most studies used a pre-post design (62.5%, k = 6). However, one study (12.5%) used an RCT design and two studies (25.0%) used a vignette design wherein professionals developed case plans based on vignettes. All tools included in these studies had support for their predictive validity, and included dynamic or modifiable factors (see Appendix B).

Study Appraisal. Most identified studies were rated as having medium (62.5%, k = 5) limitations. The most common limitation was with respect to the quality of comparison groups (only three studies fully met this criteria), as it was often unclear whether the comparison groups were sufficiently similar in variables such as risk level.

Findings. Of the three studies that examined the impact of training, training was associated with clear or partial improvements in risk management in two studies (66.7%, see Tables 6 and 7). For instance, Bonta et al. (2011) found that POs who completed an intensive
training program (i.e., Strategic Training in Community Supervision; STICS) spent more time discussing criminogenic needs, in general, and attitudes, in particular, with probationers. Although they were less likely than POs in the control group to discuss education/employment, within the STICS program, attitudes are prioritized as being more important to address than other needs. Furthermore, POs in the STICS program showed better intervention skills and their clients were less likely to reoffend. Storey, Gibas, Reeves and Hart (2011) found improved match to the need principle, but not the risk principle, following an eight-day course on risk management. In contrast to these generally positive findings, Needham et al. (2004) found that a five-day training course did not significantly enhance the ability of the BVC to reduce rates of aggressive incidents.

Of the three studies that examined the impact of risk management guidelines, guidelines were associated with some improvements in risk management in two studies (66.7%). Specifically, Bosker and colleagues compared results before and after guidelines were added to a risk assessment tool called the RISC (Bosker, Witteman, Hermans, & Heij, 2015; Bosker & Witteman, 2016). They found that guidelines improved consistency in decision-making about risk and needs, and led to more appropriate levels of planned supervision. In contrast, Daffern et al. (2009) found that guidelines did not enhance the ability of the DASA to reduce aggression.

Finally, two studies from a large, multisite study found that higher quality implementation of tools was associated with improved match to the risk principle. For instance, Vincent, Guy, et al. (2012) found that, after professionals completed standard SAVRY training, there were no significant associations between risk level and decisions; it was only after policies and protocols were in place that risk level significantly predicted POs’ management decisions.

Potential Moderators. Although research is this area is preliminary, one factor that may moderate the effectiveness of strategies, such as training, is staff engagement (Daffern et al., 2009). For instance, Bonta et al. (2008) found that POs who participated in ongoing support activities were more likely to benefit from the STICS program. Specifically, they were less likely than those with poor participation to focus on offenders’ noncriminogenic needs.

Summary. We did not complete an overall AHRQ strength of evidence rating for this question because this question encompassed a range of strategies (e.g., training, guidelines). However, these preliminary findings suggest that training, guidelines, and implementation efforts might improve match to the risk and need principles (see Table 7). That said, these approaches do not necessarily enhance the ability of tools to reduce violence and offending.

Discussion

Although risk assessment tools are thought to be an important starting point to risk management, it is unclear whether research evidence exists to support this view. As such, we conducted a systematic review of the research. Although we identified more relevant studies than anticipated \( k = 73 \), research on the utility of tools for risk management is nevertheless scarce in comparison to research on the predictive validity of tools. Furthermore, our ability to draw conclusions from this research is limited for several reasons. First, much of the existing evidence on the impact of tools on professional practices is indirect. For instance, the vast majority of studies on adherence to the risk and need principles did not directly compare
professionals’ risk management practices with and without a tool.

Second, many studies had medium or high limitations or possible confounds, as this type of research is challenging to conduct. For instance, adherence to risk assessment tools was often poor or unclear, which created a significant confound; it is difficult to meaningfully judge if tools can reduce violence if tools are not being used or are not used properly.

Finally, findings for some outcomes (e.g., violence and offending) were inconsistent. This suggests that important moderators could influence the effectiveness of risk assessment tools. However, at this point, it unclear which variables may act as moderators, as few studies directly tested potential moderators. As an example, although many researchers have hypothesized that some tools may be better suited for risk management than others, for the most part, this has not yet been directly tested. As such, it is not possible to draw conclusions. With these caveats in mind, our key findings are discussed below.

Key Findings

Professionals’ Reported Varying Opinions About the Utility of Tools. If professionals do not “buy-in” to tools or perceive them as useful, they may not adequately utilize them. Thus, as an initial step, we examined professionals’ attitudes towards risk assessment tools. We found that, although some professionals held positive views about tools, in many studies, professionals had mixed views about the utility of tools for risk management (e.g., treatment-planning, placement decisions). This is not particularly surprising; professionals often feel reluctant to adopt new assessment and intervention approaches even when these approaches have strong research support (Jensen-Doss & Hawley, 2010; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman, 2013). Furthermore, manuals and training for risk assessment tools often focus on how to complete item ratings rather than how to apply the tool to risk management efforts. As such, professionals’ questions about the utility of tools may be understandable.

Use of Tools for Risk Management Was Mixed. Not only did professionals have mixed views about the utility of tools for risk management, in most of the identified studies, the use of risk assessment tools for risk management was mixed. Specifically, although some professionals reported that they relied on tools to guide their risk management decisions (e.g., decisions about services or placements), others reported that they did not use tools, even when employers mandated their use. As such, these findings illustrate that risk assessments do not necessarily flow through to risk management efforts. Slippage might be more likely to occur when risk assessors do not have direct control over risk management decisions, but instead act as intermediaries to decision-makers (e.g., judges). In such cases, the application of tools to risk management may depend not only on evaluators’ use of tools, but also on whether subsequent decision-makers perceive tools as useful and relevant. In addition, future research should test whether the application of tools to risk management might differ by tool.

Adherence to the Risk Principle Was Moderate Following the Use of Tools. Despite the mixed application of tools to risk management efforts overall, match to the risk principle was moderate following the use of risk assessments tools. In a number of studies, high-risk individuals were referred to more services than low-risk individuals. They were also more likely
to receive secure placements. However, most studies did not have a comparison group of individuals who did not receive risk assessments. As such, it is difficult to determine if such findings are attributable to the use of the tool; some research suggests that high-risk individuals receive more intensive risk management strategies than low-risk individuals even when a risk assessment tool is not used (e.g., Campbell et al., 2015; Crocker, Braithwaite, Côté, Nicholls, & Seto, 2011). To address this possible confound, two studies directly compared match to the risk principle when a tool is and is not used (i.e., Bonta & Motiuk, 1990; Vincent, Paiva-Salisbury et al., 2012). These studies provide some initial evidence that tools may improve match. For instance, in one study, jails that used the LSI transferred a greater proportion of low-risk inmates to halfway houses compared to a jail that did not use the LSI (Bonta & Motiuk, 1990). This research is important as it suggests that, contrary to critics’ concerns, tools might sometimes help reduce restrictive sanctions for individuals who are unlikely to reoffend. However, further research is needed, particularly studies with comparison groups in which risk assessment tools were and were not used.

**Adherence to the Need Principle Was Limited Following the Use of Tools.** Contrary to the positive findings relating to the risk principle, match to the need principle appeared limited following the use of risk assessment tools (match was rated as mixed or low in all but one study). This means that many of offenders’ and patients’ needs remained unaddressed even when risk assessment tools were used. This could indicate that professionals are not paying adequate attention to risk assessments when they are making decisions about services. Alternatively, these low rates of overall match could occur because professionals opt to focus on only a couple “high impact” needs at a given time, as it may not be feasible to simultaneously target all needs (Haqanee et al., 2015). Another possibility is that low rates of match occur because services to address needs are simply not available (Haqanee, Peterson-Badali, & Skilling, 2015). Clearly, identifying needs has limited value if there are no viable means by which to address these needs. Finally, low match to the needs principle may, in part, arise from limited compliance; offenders and patients may not necessarily attend or engage in the services to which they are referred.

**Evidence on Whether Tools Reduce Violence and Offending Was Inconsistent.** In light of the above findings, it is perhaps not surprising that evidence on whether tools reduce violence and offending was inconsistent. Although two RCTs found that the use of the BVC resulted in decreases in violence, another RCT did not find significant changes in violence or other criminal incidents when another tool, the START, was implemented. In addition, although two pre-post studies found that the implementation of risk assessment tools was associated with decreases in violence or offending, the bulk of pre-post studies (k = 7) did not. Thus, at present, there is insufficient evidence to conclude that tools reduce violence or offending.

One possible explanation for these findings is that it may be unrealistic to expect risk assessment tools to directly reduce violence or offending. As an analogy, would assessing whether someone is depressed directly reduce depression? Probably not. Instead, it may depend largely on what happens after the assessment. The same may be true for risk assessment tools; tools might be effective only if they enhance the likelihood that individuals receive appropriate, empirically-supported interventions.

Another possible explanation for the modest findings is that the effectiveness of tools
might vary by factors such as the setting, population, or tool. For instance, the two cluster RCT studies that found positive results (Abderhalden et al., 2008; van de Sande et al., 2011) were both conducted in acute psychiatric wards. This could suggest that tools may have a more positive impact in acute psychiatric settings than in other settings, perhaps because it is a more controlled environment. However, this may not entirely explain results because some studies in controlled psychiatric settings found null results (e.g., Gunenc, O’Shea, & Dickens, 2017).

It is also possible that some tools may be more effective than other tools. For instance, the two RCTs with positive results included the BVC, a brief tool that includes many dynamic or changeable risk factors. Again, though, this may not fully explain results because other studies in which other validated dynamic risk assessment tools were implemented (e.g., BVC-R, DASA, SAVRY, START, YLS/CMI) failed to find significant decreases in violence or offending.

Training, Guidelines, and Implementation Quality May Enhance the Utility of Tools. Overall, our results suggest that there is a disconnect between the theory of risk assessment and what actually happens in real-world practice (Peterson-Badali et al., 2015). However, although research is limited, preliminary evidence suggests that it may be possible to enhance professionals’ risk management practices by combining the use of risk assessment tools with approaches such as risk management training (e.g., Bonta et al., 2011) and structured risk management guidelines (e.g., Bosker & Witteman, 2016). Though such approaches may not directly reduce violence and offending, they have been found to improve match to the risk and need principles, thus providing potential avenues by which to enhance the utility of tools. Sound implementation practices, such as policies and protocols to guide the use of tools, are also critical (e.g., Vincent et al., 2016). For instance, to assist agencies, Vincent, Guy, and Grisso (2012) have developed a guidebook for implementing risk assessment tools.

Limitations of Systematic Review

In interpreting our findings, several limitations of this review are important to note. Although we systematically searched 13 databases, reviewed reference lists, and contacted experts, our review likely missed some studies, such as studies written in languages other than English. In addition, although we attempted to summarize our findings with terms such as low, mixed, or high, definitions of such terms are somewhat subjective by nature. As such, to increase transparency and objectivity, we provided operational definitions of our summary terms, and summarized study findings in more detail using evidence tables (see Tables 1 – 6). In addition, two independent raters coded each study and we conducted consensus ratings. Another limitation of this systematic review is that there is a lack of appropriate tools for appraising risk of bias in risk assessment studies. As such, we drew items from other tools, and adapted the wording for this context. However, the approach that we used to appraise observational and survey studies was brief and, as such, our review likely failed to capture some relevant study limitations. Finally, although we examined differences in general patterns of results across published and unpublished studies, it is difficult to evaluate publication bias in narrative reviews.

Implications for Practice and Research

Our findings suggest that risk assessment tools are not sufficient to guarantee sound risk
management practices or reductions in violence (e.g., Peterson-Badali et al., 2015). Thus, researchers and tool developers should be careful to not overstate the potential value of risk assessment for risk management. Likewise, professionals, agencies, and policy-makers should not rely on risk assessment tools as their sole or primary risk management strategy. Instead, agencies who use risk assessment tools should work to build staff buy-in, regularly monitor adherence, and ensure that they are providing effective treatment, rather than funneling all their resources into assessment.

That said, even though risk assessment tools have limitations, they remain a best available practice. Although tools may not reduce violence or offending in and of themselves, there is no evidence that alternative approaches, such as assessing risk via unstructured clinical judgement or not assessing risk at all, do so either. For instance, one study found that when tools are not used, judges relied on unstructured risk assessments in making sentencing decisions, even though these assessments did not accurately predict violence (Lodewijks, Doreleijers, & de Ruiter, 2008). Indeed, numerous studies indicate that tools can help to predict violence and reoffending (e.g., Olver et al., 2014; Singh et al., 2011), and outperform unstructured approaches (Ãegisdóttir et al., 2006; Hanson & Morton-Bourgon, 2009).

Based on this review, there are number of important areas for future research. First, many studies have lacked appropriate comparison groups, making it impossible to determine if tools improve practices per se. As such, there is a strong need for further research with comparison groups of individuals who were not assessed with a risk assessment tool, including studies with both mental health and justice populations. Second, to determine if certain tools may have a more beneficial impact on risk management than other tools, head-to-head comparisons of tools are needed. Third, adherence to risk assessment tools appears to be poor in some cases, making it difficult to evaluate the impact of tools. As such, research should routinely measure and report adherence. Fourth, given that the utility of tools for risk management likely depends heavily on what happens after the risk assessment, researchers examine the pathway between risk assessment and risk management, such as by developing and testing conceptual models. Finally, researchers should create and evaluate approaches to improve the utility of tools for risk management, such as training initiatives, structured risk management guidelines, and quality improvement or audit systems.

Conclusion

Overall, this review suggests that even though risk assessment tools may be a starting point for risk management, they are not sufficient in and of themselves (Vincent, Paiva-Salisbury, et al., 2012). Although some studies found positive results, indicating that tools might help achieve better match to the risk principle or even reductions in violence in some circumstances, the findings also revealed that “there is no guarantee that the results of these protocols flow through to front line service provision” (Haqane et al., 2015, p. 39). For risk assessment researchers, the results of this review may be somewhat disappointing. Even though agencies pay “lip service” to using risk assessment to guide their risk management efforts (Skeem, 2013, p. 302), there is often “a lack of follow through between the assessment and case management” (Bonta et al., 2008, p. 266).

To ensure that risk assessment instruments are optimally used and do not degenerate into merely a bureaucratic exercise (Webster, 2011), further efforts are needed. In particular, rather
than focusing exclusively on predictive validity studies and the development of new tools, researchers need to pay greater attention to how tools are applied to guide real-world decisions, such as by testing the pathways between risk assessment and risk management, identifying areas of slippage, and developing strategies to facilitate the ability of risk assessments to translate into better risk management efforts. Such initiatives are essential to ensuring that the potential value of risk assessment is more fully realized.
References


Bosker, J., & Witteman, C. (2016). Finding the right focus: Improving the link between...


Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments


Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine, 2*(8), 696-701. doi:10.1371/journal.pmed.0020124


Laura and John Arnold Foundation (2014). *Results from the First Six Months of the Public Safety Assessment – Court™ in Kentucky*. Laura and John Arnold Foundation.


### Table 1
Do Professionals Perceive Risk Assessment Tools to be Useful for Risk Management?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample</th>
<th>Risk Tool</th>
<th>Design</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonta et al., 2005 (CA)</td>
<td>482 judges, POs, lawyers</td>
<td>Tools in general</td>
<td>Mail survey, interviews</td>
<td>Judges/POs rated usefulness of risk assessments in presentencing reports as $M = 8.1 - 9.1$ (of 10)</td>
<td>**</td>
<td>High</td>
</tr>
<tr>
<td>Flores et al., 2004 (USA)</td>
<td>195 corrections staff</td>
<td>Y-LSI</td>
<td>Mail survey</td>
<td>Rated usefulness for placements as $M = 5.2$ (of 10) and usefulness for identifying treatment needs as $M = 5.5$</td>
<td>**</td>
<td>Mixed</td>
</tr>
<tr>
<td>Flores, 2013 (USA)</td>
<td>185 corrections staff</td>
<td>YLS/CMI</td>
<td>Mail survey</td>
<td>Rated usefulness for placement decisions, identifying treatment needs, and case planning as $M = 5.3 - 5.7$ (of 10)</td>
<td>**</td>
<td>Mixed</td>
</tr>
<tr>
<td>Guy et al., 2014 (USA)</td>
<td>71 YPOs</td>
<td>SAVRY YLS/CMI</td>
<td>Interviews</td>
<td>63 – 71% rated SAVRY as very helpful for service, disposition, and supervision decisions; 41 – 50% rated YLS/CMI as very helpful for these decisions</td>
<td>**</td>
<td>Mixed-High</td>
</tr>
<tr>
<td>Mair et al., 2016 (UK)</td>
<td>180 POs</td>
<td>OASys</td>
<td>Interviews</td>
<td>36% stated that tool helps produce better supervision plans</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>McKee et al., 2007 (CA)</td>
<td>157 forensic staff</td>
<td>VRAG</td>
<td>In-person survey</td>
<td>Rated importance of tool in transfer decisions as $M = 7.8$ (of 10); tools seen as less important than index offense, etc.</td>
<td>**</td>
<td>Mixed</td>
</tr>
<tr>
<td>Muheizen, 2014 (USA)</td>
<td>433 clinicians</td>
<td>VRAG</td>
<td>Online survey</td>
<td>78% rated tool as very helpful to discharge decisions</td>
<td>***</td>
<td>High</td>
</tr>
<tr>
<td>Richardson, 2009 (UK)</td>
<td>45 clinical staff</td>
<td>SCJ:Risk</td>
<td>In-person survey</td>
<td>72% rated tool as useful for informing clinical practice, 78% rated it as useful for risk management plan</td>
<td>***</td>
<td>High</td>
</tr>
<tr>
<td>Royal College of Psychiatrists, 2008 (UK)</td>
<td>1937 psychiatrists</td>
<td>Tools in general</td>
<td>Online survey</td>
<td>48% reported that tools lead to better decisions</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>Schneider et al., 1996 (USA)</td>
<td>179 POs</td>
<td>WRNA</td>
<td>Mail survey</td>
<td>37% stated that tool helps supervision decisions/caseload management; 53% indicated it helps high risk offenders receive supervision</td>
<td>***</td>
<td>Low-Mixed</td>
</tr>
<tr>
<td>Shook &amp; Saari, 2007 (USA)</td>
<td>665 POs and judges</td>
<td>Tools in general</td>
<td>Survey</td>
<td>Rated usefulness for pretrial detention, placement, and release decisions as $M = 3.06 - 3.71$ (of 5); 36% judges/41% POs stated that tools ensure appropriate placement</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Singh et al., 2014 (international)</td>
<td>2135 clinicians, 44 countries</td>
<td>Various tools</td>
<td>Online survey</td>
<td>Rated usefulness for risk management as $M = 3.2 - 4.5$ (of 6); results varied by tool</td>
<td>***</td>
<td>Mixed – High</td>
</tr>
</tbody>
</table>

**Note.** The summary ratings were: Low (participants rated helpfulness as being below the midpoint [e.g., less than half believed the tool aided in risk management]), Mixed (participants rated helpfulness as being just above the midpoint [e.g., 5-7 out of 10, with 10 being the most positive]), High (most participants rated tool as helpful [e.g., mean of >7 out of 10, with 10 being most positive]). The study appraisal ratings were: * (high limitations [<50% on the study appraisal form]), ** (medium limitations [51 – 79% on the study appraisal form]), *** (low limitations [>80% on the
study appraisal form)). CA = Canada; N.R. = not reported; PO = probation or parole officer; \(M\) = mean; U.K. = United Kingdom; USA = United States of America; YPO = youth probation officer. Sample size refers to the sample used for the relevant analyses.
## Table 2
When Professionals Conduct Risk Assessments with Tools, Do These Assessments Guide Risk Management Efforts?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample</th>
<th>Risk Tool</th>
<th>Design</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Self-Reported Use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flores et al., 2004 (USA)</td>
<td>195 corrections staff</td>
<td>Y-LSI</td>
<td>Mail survey</td>
<td>57% used tool to identify treatment goals, 71% used it to gauge treatment effectiveness, 80% used it to develop case plans, 86% used it to inform supervision decisions</td>
<td>**</td>
<td>Mixed – High</td>
</tr>
<tr>
<td>Flores, 2013 (USA)</td>
<td>185 corrections staff</td>
<td>YLS/CMI</td>
<td>Mail survey</td>
<td>82% used tool to develop case plans</td>
<td>**</td>
<td>High</td>
</tr>
<tr>
<td>Haas &amp; DeTardo-Bora, 2009 (USA)</td>
<td>94 CMs, POs, counsellors</td>
<td>LSI-R</td>
<td>Mail survey</td>
<td>42% used tool to develop reentry plans, 67% to assess treatment progress, 78% to develop treatment plan, 79% to set supervision level, and 80% for referrals</td>
<td>**</td>
<td>Mixed – High</td>
</tr>
<tr>
<td>Hawley et al., 2010 (UK)</td>
<td>300 nurses, doctors, etc.</td>
<td>Tools in general</td>
<td>In-person survey</td>
<td>27% used tool less than half the time, 21% used it half the time, and 52% used it more than half the time</td>
<td>*</td>
<td>Mixed</td>
</tr>
<tr>
<td>Krysik &amp; LeCroy, 2002 (USA)</td>
<td>80 POs, judges</td>
<td>NCCD tool†</td>
<td>Survey</td>
<td>36% of POs used tool to make court recommendations, 42% of judges said tool played a significant role in decisions</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>Miller &amp; Maloney, 2013 (USA)</td>
<td>1087 community corrections staff</td>
<td>Tools in general</td>
<td>Email survey</td>
<td>48% completed the tool and applied it to decision-making, 40% completed the tool but didn’t fully apply it to decisions, and 12% neither completed nor applied the tool</td>
<td>**</td>
<td>Low – Mixed</td>
</tr>
<tr>
<td>Shook &amp; Sarri, 2007 (USA)</td>
<td>665 POs, judges, etc.</td>
<td>Tools in general</td>
<td>Survey</td>
<td>Judges: 45-48% routinely used risk and needs assessments in dispositions; POs: 64-72% routinely used risk and needs assessment in dispositions</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Singh et al., 2014 (international)</td>
<td>2135 clinicians, 44 countries</td>
<td>Various tools</td>
<td>Online survey</td>
<td>69% used tools to develop risk management plans; in 35% of known cases, plans were not implemented</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Vincent, Paiva-Salisbury et al., 2012 (USA)</td>
<td>88 YPOs, 23 managers</td>
<td>SAVRY, YLS/CMI (YPOs)</td>
<td>Interview</td>
<td>80% used tools in service decisions, 77% in disposition decisions, and 90% in supervision decisions</td>
<td>***</td>
<td>High</td>
</tr>
<tr>
<td><strong>B. Observational Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boccaccini et al., 2013 (USA)</td>
<td>299 jurors</td>
<td>Static-99, MnSOST-R</td>
<td>In-person</td>
<td>No significant relationship between risk scores and jurors’ perception of offenders’ reoffense risk</td>
<td>***</td>
<td>Low</td>
</tr>
<tr>
<td>Kewley et al., 2015 (UK)</td>
<td>216 plans by POs</td>
<td>RM2000, OASys</td>
<td>File review</td>
<td>Many risk factors identified weren’t mentioned in the plans; 8% of management plans referred to risk assessment</td>
<td>**</td>
<td>Low</td>
</tr>
<tr>
<td>Study</td>
<td>Participants/Settings</td>
<td>Risk Management Tool(s)</td>
<td>Method</td>
<td>Findings</td>
<td>Limitations</td>
<td>Study Type</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>--------------------------------</td>
<td>----------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>Turner et al., 2015</td>
<td>462 jurors - USA; Static-99, MnSOST-R - In-person survey</td>
<td></td>
<td></td>
<td>Perceived influence of tools in decisions was $M = 4.32 - 4.98$ (of 6); tools had lower influence than most other factors (e.g., offenders' testimony perceived remorse)</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Urquhart &amp; Viljoen,</td>
<td>50 young offender cases with risk assessment tool - International</td>
<td>SAVRY, YLS/CMI - Case law review</td>
<td></td>
<td>46% of cases – judges directly referred to tool in decisions and 30% of cases – did not refer to tool but mentioned risk; 4% of cases – judges placed great weight on tool and 58% of cases – placed some weight on tool</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Viglione et al., 2015</td>
<td>69 sessions with adult probationers - USA; Tool N.R.† POs</td>
<td></td>
<td></td>
<td>POs referred to tool in 3% of interactions, and used tool for case planning in 1% of interactions</td>
<td>*</td>
<td>Low</td>
</tr>
</tbody>
</table>

**Note.** † No support for the predictive validity of the tool could be found. The summary ratings were: Low (less than half of participants used the tool for risk management), Mixed (just above half used the tool for risk management [e.g., 50 – 70%]), High (most participants used the tool for risk management [e.g., > 70%]). The study appraisal ratings were: * (high limitations [<50% on the study appraisal form]), ** (medium limitations [51 – 79% on the study appraisal form]), *** (low limitations [>80% on the study appraisal form]). CM = case manager; $M = mean$; N.R. = not reported; PO = probation or parole officer; UK = United Kingdom; USA = United States of America.
Table 3
Does the Use of Risk Assessment Tools Facilitate Match to the Risk Principle?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample (gender of evaluatees)</th>
<th>Risk Tool (assessor)</th>
<th>Design</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Studies Comparing Tool vs. No Tool</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonta &amp; Motiuk, 1990 (CA)</td>
<td>580 incarcerated adults (M)</td>
<td>LSI (staff)</td>
<td>Non-randomized comparison</td>
<td>In the jails that used tool, inmates with low risk totals were more likely to be transferred to a halfway house (51% vs. 16% at a control jail)</td>
<td>**</td>
<td>Improvement</td>
</tr>
<tr>
<td>Vincent, Paiva-Salisbury et al., 2012 (USA)</td>
<td>88 YPOs, 23 managers</td>
<td>SAVRY, YLS/CMI (YPOs)</td>
<td>Pre-post interviews</td>
<td>After implementation, YPOs were more likely to spontaneously identify risk as important for their supervision decisions but not for their services or disposition decisions</td>
<td>***</td>
<td>Some Improvement</td>
</tr>
<tr>
<td>B. Studies with No Comparison Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Placement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeGue et al., 2008 (USA)</td>
<td>220 young offenders (M/F)</td>
<td>YLS/CMI (clinicians)</td>
<td>File review</td>
<td>Risk total predicted out-of-home placement recommendations ($t = 7.07$)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Gebo, 2002 (USA)</td>
<td>46 incarcerated young offenders (M/F)</td>
<td>RAI† (N.R.)</td>
<td>File review</td>
<td>Risk score predicted dispositions (i.e., commitment) more strongly in district vs. family court (62% vs. 29%)</td>
<td>**</td>
<td>Low – Moderate</td>
</tr>
<tr>
<td>Guy et al. (2015)</td>
<td>539 youth on probation (M/F)</td>
<td>SAVRY, JAG (YPOs)</td>
<td>File review</td>
<td>SAVRY summary risk ratings and JAG total scores generally predicted out-of-home placements</td>
<td>***</td>
<td>Moderate</td>
</tr>
<tr>
<td>Jung et al., 2015 (CA)</td>
<td>165 adult forensic outpatients (M/F)</td>
<td>LSI, HCR-20 (psych.)</td>
<td>File review</td>
<td>HCR-20 and LSI total predicted incarceration ($\eta^2 = .10 - .11$) but not length of incarceration or probation</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Perrault et al., 2012 (USA)</td>
<td>64 YPOs</td>
<td>YLS/CMI, SAVRY (POs)</td>
<td>Interviews</td>
<td>90% of YPOS said they used risk level to inform placement decisions</td>
<td>**</td>
<td>High</td>
</tr>
<tr>
<td>Simpson, 2010 (USA)</td>
<td>202 young offenders (gender N.R.)</td>
<td>DSI† (police)</td>
<td>Pre-post</td>
<td>Risk total predicted detention decision ($r = .32$)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Vincent et al., 2016 (USA)</td>
<td>847 youth on probation (M/F)</td>
<td>SAVRY (YPOs)</td>
<td>File review</td>
<td>Risk level predicted placements at 5 of 6 sites</td>
<td>***</td>
<td>High</td>
</tr>
<tr>
<td>2. Other Sentencing Decisions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blais &amp; Forth, 2014 (CA)</td>
<td>120 adult inmates (gender N.R.)</td>
<td>VRAG, etc. (clinicians)</td>
<td>File review</td>
<td>Risk total (VRAG, Static-99, SORAG) did not predict dangerous offender designation (AUCs = .50 – .51)</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>Hseih, 2016 (USA)</td>
<td>2137 adults with sexual offenses (M)</td>
<td>Static-99 etc. (N.R.)</td>
<td>File review</td>
<td>Risk total inversely predicted sentence length but once other factors controlled, positively predicted sentence length</td>
<td>*</td>
<td>Moderate</td>
</tr>
<tr>
<td>3. Supervision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Auditor General of Ontario, 2012 (CA)  
Young offenders (n and gender N.R.)  
Tool N.R. (YPOs)  
File review  
“Almost half” of youth met with POs less than required based on risk level  
Low

Bonta et al., 2008 (CA)  
154 adults/youth on probation (M/F)  
PRA (POs)  
File review  
Risk total predicted # of probation contacts for adults (r = .22) but not for youth (r = .09)  
Low – Moderate

Bosker et al., 2013b (NL)  
44 POs, 173 plans  
RISc-2 (POs)  
Vignettes, wrote plans  
Low agreement about intensity of supervision when given risk assessment results (kappa = .20)  
Low

Guy et al. (2015)  
248 youth on probation (M/F)  
SAVRY, JAG (YPOs)  
File review  
JAG and SAVRY scores were used to assign initial level of supervision, but risk ratings did not consistently predict supervision contacts received (at Mississippi site)  
Low-Moderate

Luong & Wormith, 2011 (CA)  
192 youth on probation (M/F)  
Youth LSI-SK (POs)  
File review  
Risk total predicted supervision frequency (r = .66)  
High

Maupin, 1993 (USA)  
280 youth on parole (gender N.R)  
AJADMS† (POs)  
File review  
Risk classification predicted contacts/month (R² = .12) but inconsistent, with some findings in the opposite direction  
Low – Moderate

Vincent et al., 2016 (USA)  
847 youth on probation (M/F)  
SAVRY (YPOs)  
File review  
Risk level predicted supervision level at 4 of 5 sites  
High

4. Release or Discharge

Folino et al., 2004 (AR)  
65 adult inmates & forensic patients (M)  
HCR-20, VRAG (N.R.)  
File review  
Tool ratings were not correlated with judicial release decisions  
Low

Guy et al., 2015 (USA)  
5187 adult offenders, life sentence (M/F)  
HCR-20, LS/CMI (psyc.)  
File review  
HCR-20 and LS/CMI totals predicted parole release decisions (d = 1.09 and 0.97, respectively)  
High

Hilton & Simmons, 2001 (CA)  
187 forensic inpatients (gender N.R.)  
VRAG (staff)  
File review  
Risk total did not predict tribunal decisions (r = .06), team recommendations (r = .01), or clinician testimony (r = -.02)  
Low

McKee et al., 2007 (CA)  
104 forensic inpatients (gender N.R.)  
VRAG (N.R.)  
File review  
Risk total predicted psychiatrist recommendations (r = .23) & team recommendations (r = .21) but not tribunal decisions  
Low – Moderate

Morrisey et al., 2014 (UK)  
532 adult psychiatric patients (M/F)  
HCR-20 (team)  
File review  
Risk total did not predict discharge but clinical scale total did  
Low – Moderate

Muheizen, 2014 (USA)  
433 clinicians  
VRAG  
Vignettes  
78% less likely to recommend discharge if received high risk vignette (compared to low risk or no tool)  
High

5. Overall Level or Intensity of Services

Campbell, 2015 (USA)  
2739 youth on probation (M/F)  
YLS/CMI (court officers)  
File review  
Risk total predicted number of program referrals (t = 12.20)  
Moderate

Guy et al. (2015)  
539 youth on probation (M/F)  
SAVRY, JAG (YPOs)  
File review  
SAVRY summary ratings and JAG totals sometimes predicted number of referrals but did not predict services completed  
Low
### 6. Overall Level or Intensity of Intervention or Case Management Plans

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Description</th>
<th>Methodology</th>
<th>Findings</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>McCormick et al., 2017 (US)</td>
<td>232 presentencing youth (M/F)</td>
<td>VRAG</td>
<td>Risk total did not predict intensity of services (i.e., number of needs addressed in services) while on probation</td>
<td>Low</td>
</tr>
<tr>
<td>Perrault et al., 2012 (US)</td>
<td>64 YPOs</td>
<td>YLS/CMI, SAVRY (POs)</td>
<td>73% of POS said they used risk level to inform service referrals</td>
<td>High</td>
</tr>
<tr>
<td>Vincent et al., 2016 (US)</td>
<td>847 youth on probation (M/F)</td>
<td>SAVRY (YPOs)</td>
<td>Risk level predicted number of service referrals at 4 out of 6 sites</td>
<td>Moderate</td>
</tr>
<tr>
<td>Perrault et al., 2012 (USA)</td>
<td>64 YPOs</td>
<td>Interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vincent et al., 2016 (US)</td>
<td>847 youth on probation (M/F)</td>
<td>File review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosker et al., 2013a (NL)</td>
<td>300 adults on probation (M/F)</td>
<td>RISC-2 (POs)</td>
<td>Risk total predicted intensity of planned interventions (r = .22)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Dyck, 2016 (CA)</td>
<td>100 diverted youth (M/F)</td>
<td>YLS/CMI (staff)</td>
<td>After using tool, 60% of cases adhered to risk principle in case plans</td>
<td>Moderate</td>
</tr>
<tr>
<td>Holloway, 2015 (USA)</td>
<td>147 YPOs</td>
<td>Vignettes</td>
<td>30% of plans were appropriate intensity, 59% had some limitations, and 11% had substantial limitations</td>
<td>Low – Moderate</td>
</tr>
<tr>
<td>Kewley et al., 2015 (UK)</td>
<td>216 adult probationers with sexual offenses (M)</td>
<td>RM 2000, OASys (practitioner)</td>
<td>Plans for very high risk probationers included more support than plans for high risk probationers (83% vs. 53%)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Singh et al., 2014 (USA)</td>
<td>120 detained youth (M/F)</td>
<td>START:AV (CM)</td>
<td>Risk total did not predict number of planned interventions for full sample (r = .12) but it did for girls (r = .44)</td>
<td>Low – Moderate</td>
</tr>
<tr>
<td>Storey et al, 2014 (SE)</td>
<td>249 adult IPV offenders (M)</td>
<td>B-SAFER (police)</td>
<td>Risk total predicted number of recommended management strategies (r = .43)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Storey &amp; Strand, 2013 (SE)</td>
<td>52 adult IPV offenders (F)</td>
<td>B-SAFER (police)</td>
<td>Risk total did not predict number of recommended management strategies</td>
<td>Low</td>
</tr>
<tr>
<td>Storey &amp; Strand, 2017 (SE)</td>
<td>867 adult IPV offenders (M/F)</td>
<td>B-SAFER (police)</td>
<td>Victim vulnerability scores predicted number of risk management strategies recommended (if victim was female)</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

### 7. Other Strategies

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Description</th>
<th>Methodology</th>
<th>Findings</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belfrage &amp; Strand, 2012 (SE)</td>
<td>216 adult IPV offenders (M)</td>
<td>B-SAFER (police)</td>
<td>Police reported that they took more protective actions for individuals they rated as an imminent risk (τ = .30)</td>
<td>Moderate</td>
</tr>
<tr>
<td>Miller &amp; Maloney, 2013 (USA)</td>
<td>1087 POs</td>
<td>Variety of tools (POs)</td>
<td>POs reported that the extent to which they make decisions that correspond with risk level is $M = 4.5$ (of 6)</td>
<td>High</td>
</tr>
<tr>
<td>Trujillo &amp; Ross, 2008 (AU)</td>
<td>501 IPV offenders (M/F; age N.R.)</td>
<td>L17A† (police)</td>
<td>Risk total predicted whether action taken and intervention ordered (OR = 1.64 – 2.21), but not whether charges laid</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

**Note.** † No support for the predictive validity of the tool could be found. The summary ratings were: Low (nonsignificant associations between risk total and intervention intensity, or many cases showed inadequate match to the risk principle [e.g., nearly half]), Moderate (significant associations between risk total and intervention intensity which fell in the small or moderate range [$r < .50$], or most cases showed adequate match to the risk.
principle [e.g., 60-75%]), High (large and significant associations between risk total and intervention intensity, or almost all cases [e.g., >75%] showed adequate match to the risk principle), Improvements (improvements in all or most indicators of match to the risk principle), Some Improvements (improvements in some but not all of the indicators of match to the risk principle). The study appraisal ratings were: * (high limitations [<50% on the study appraisal form]), ** (medium limitations [51 – 79% on the study appraisal form]), *** (low limitations [>80% on the study appraisal form]). AU = Australia; AR = Argentina; CA = Canada; CM = case manager; F = female; IPV = intimate partner violence; M = male; M/F = male/female; N.R. = not reported; POs = parole or probation officer; Psyc = psychologists; SE = Sweden; UK = United Kingdom; USA = United States of America; YPO = youth probation officer.
Table 4
*Does the Use of Risk Assessment Tools Facilitate Match to the Need Principle?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample (gender of evaluatees)</th>
<th>Risk Tool (assessor)</th>
<th>Design</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Studies Comparing Tool vs. No Tool</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. Studies with No Comparison Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Needs Mentioned in Intervention or Case Management Plans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditor General of Ontario, 2012 (CA)</td>
<td>Young offenders (n and gender N.R.)</td>
<td>Tool N.R.† (YPOs)</td>
<td>File review</td>
<td>&gt;50% of case plans were missing goals for at least one high risk factor; 20% did not have goals for any high risk factor</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>Bosker et al., 2013a (NL) ‡</td>
<td>300 adults on probation (M/F)</td>
<td>RISc-2 (POs)</td>
<td>File review</td>
<td>46% of needs were adequately addressed in plans; ranged from 0% (friends) to 72% (cognitive skills)</td>
<td>**</td>
<td>Low</td>
</tr>
<tr>
<td>Bosker et al., 2013b (NL)</td>
<td>44 POs</td>
<td>RISc-2 (POs)</td>
<td>Vignettes, wrote plans</td>
<td>Agreement about the relevance of criminogenic needs fell in moderate range (kappa = .43)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Dyck, 2016 (CA)</td>
<td>100 diverted youth (M/F)</td>
<td>YLS/CMI (officer)</td>
<td>File review</td>
<td>33% of needs were addressed in plans; ranged from 22% (criminal friends) to 47% (personality/behavior)</td>
<td>**</td>
<td>Low</td>
</tr>
<tr>
<td>Holloway, 2015 (USA)</td>
<td>147 young offenders (M/F)</td>
<td>YLS/CMI (YPOs)</td>
<td>Vignettes, wrote plans</td>
<td>60% of needs were addressed in plans; ranged from 37% (leisure/recreation) to 82% (family/parenting)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Kewley et al., 2015 (UK)</td>
<td>216 adult probationers with sexual offenses (M)</td>
<td>OASys, RM2000 (practitioner)</td>
<td>File review</td>
<td>An average of 3.60 needs were identified in risk assessment but only 1.52 needs were addressed in plans</td>
<td>**</td>
<td>Low</td>
</tr>
<tr>
<td>Luong &amp; Wormith, 2011 (CA)</td>
<td>192 youth on probation (M/F)</td>
<td>Youth LSI-SK (POs)</td>
<td>File review</td>
<td>63% of needs were addressed in plans; ranged from 30% (education/employment) to 76% (family/parenting); needs in some domains, but not all, predicted domain-specific services</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Sen et al., 2015 (UK)</td>
<td>88 adult psychiatric patients (M/F)</td>
<td>HCR-20 (team)</td>
<td>File review</td>
<td>Case management plans were rated as $M = 3.9$ (of 8) on extent to which they addressed risk factors</td>
<td>*</td>
<td>Moderate</td>
</tr>
<tr>
<td>Singh et al., 2014 (USA)</td>
<td>120 detained youth (M/F)</td>
<td>START:AV (CMs)</td>
<td>File review</td>
<td>52% of critical needs were addressed in plans; ranged from 0% (e.g., coping) to 95% (substance use)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Young et al., 2006 (USA)</td>
<td>3910-4592 young offenders (gender N.R.)</td>
<td>Intake Screen† (intake staff)</td>
<td>File review</td>
<td>28% of those with education needs received referral; 82% of those with substance abuse needs received referral</td>
<td>*</td>
<td>Moderate</td>
</tr>
<tr>
<td>2. Needs Addressed via Services Received</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Flores et al., 2004 (USA)</td>
<td>1679 young offenders (M/F)</td>
<td>Y-LSI</td>
<td>File review</td>
<td>1 – 5% of youth with needs in family, peer, education, or drug domains received treatment in these areas</td>
<td>*</td>
<td>Low</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Methods</td>
<td>Findings</td>
<td>Study Appraisal Rating</td>
<td>Summary Rating</td>
<td>Study Appraisal Rating Comment</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-------------</td>
<td>---------</td>
<td>--------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>----------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Haas et al. (2007) (USA)</td>
<td>348 incarcerated adults (M/F)</td>
<td>LSI-R (staff)</td>
<td>File review</td>
<td>53% of those with substance use needs received services; 35% with education/employment needs received services</td>
<td>**</td>
<td>Low – Moderate</td>
</tr>
<tr>
<td>Peterson-Badali et al., 2015 (CA)</td>
<td>148 youth on probation (M/F)</td>
<td>YLS/CMI (clinicians)</td>
<td>File review</td>
<td>Clinicians mentioned identified needs (e.g., family) in reports, but services weren’t received (i.e., 40% of youth had no needs addressed); on average, 31% of needs were addressed</td>
<td>**</td>
<td>Low</td>
</tr>
<tr>
<td>Vieira et al., 2009 (CA)</td>
<td>122 young offenders (M/F)</td>
<td>YLS/CMI (clinicians)</td>
<td>File review</td>
<td>35% of youths’ needs addressed while on probation</td>
<td>*</td>
<td>Low</td>
</tr>
</tbody>
</table>

3. Needs Discussed in Probation Sessions

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Methods</th>
<th>Findings</th>
<th>Study Appraisal Rating</th>
<th>Summary Rating</th>
<th>Study Appraisal Rating Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bonta et al., 2008 (CA)</td>
<td>154 adults/youth probation sessions (M/F)</td>
<td>PRA (POs)</td>
<td>Audiotaped sessions</td>
<td>Mean % of identified needs discussed in session with adult offenders was 54% a; ranged from 0% (academic/vocational) to 90% (family/marital)</td>
<td>**</td>
<td>Moderate</td>
</tr>
<tr>
<td>Bonta et al., 2011 (CA)</td>
<td>75 probation sessions (M/F) from 19 POs b</td>
<td>PRA (POs)</td>
<td>Audiotapes</td>
<td>POs spent 43% of session time talking about criminogenic needs and 49% of time on non-criminogenic needs</td>
<td>***</td>
<td>Low – Moderate</td>
</tr>
</tbody>
</table>

4. Needs Addressed Based on Professionals’ Self-Report

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Size</th>
<th>Methods</th>
<th>Findings</th>
<th>Study Appraisal Rating</th>
<th>Summary Rating</th>
<th>Study Appraisal Rating Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miller &amp; Maloney, 2013 (USA)</td>
<td>1087 POs</td>
<td>Tools in general (POs)</td>
<td>Survey</td>
<td>POs reported that the extent to which they target needs identified by tool is $M = 4.6$ (of 6)</td>
<td>**</td>
<td>High</td>
</tr>
</tbody>
</table>

Note. † No support for the predictive validity of the tool could be found. The summary ratings were: Low (overall, less than half of cases showed adequate match to the need principle, or less than half of needs were addressed), Moderate (overall, slightly more than half of cases [50-69%] showed adequate match to the need principle, or slightly more than half of needs were addressed), High (overall, most cases [>70%] showed adequate match to the need principle or most needs were addressed). The study appraisal ratings were: * (high limitations [<50% on the study appraisal form]), ** (medium limitations [51 – 79% on the study appraisal form]), *** (low limitations [>80% on the study appraisal form]). a Average match was calculated as follows: (total number of needs addressed in intervention plan/total number of needs identified by risk assessment tool); b This study also examined whether training improves match; those results are described in Table 6. CA = Canada; CM = case manager; M = male; M/F = male/female; NL = Netherlands; N.R. = not reported; PO = parole or probation officer; UK = United Kingdom; USA = United States of America; YPO = youth probation officer.
Table 5
Does the Implementation of Risk Assessment Tools Reduce Violence or Offending?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample (gender)</th>
<th>Risk Tool (assessors)</th>
<th>Design</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. RCTs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abderhalden et al., 2008 (CH)</td>
<td>2364 adult psy. inpatients (M/F)</td>
<td>BVC + risk management (nurses)</td>
<td>Cluster RCT</td>
<td>Wards using tool showed greater decline in severe aggressive incidents (adjusted risk reduction = 41%) but they had higher initial rates of aggression</td>
<td>**</td>
<td>Decrease</td>
</tr>
<tr>
<td>Troquete et al., 2013 (NL)</td>
<td>632 adult forensic outpatients (M/F)</td>
<td>START + joint plan (CMs)</td>
<td>Cluster RCT</td>
<td>No significant difference in violence/criminal incidents in wards using tool than in control wards</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td>van de Sande et al., 2011 (NL)</td>
<td>597 adult psy. inpatients (M/F)</td>
<td>BVC, etc. (nurses/drs.)</td>
<td>Cluster RCT</td>
<td>Wards using tool showed greater decline in aggressive incidents than controls (relative risk ratio = -68%)</td>
<td>**</td>
<td>Decrease</td>
</tr>
<tr>
<td><strong>B. Non-Randomized Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belfrage et al., 2004 (SE)</td>
<td>47 adult inmates (M)</td>
<td>HCR-20 (N.R.)</td>
<td>Pre-post</td>
<td>64% decrease in violent incidents following implementation of tool, but risk scores did not decrease</td>
<td>**</td>
<td>Decrease</td>
</tr>
<tr>
<td>Bhui et al., 2001 (UK)</td>
<td>324 psy. patients (age, gender N.R.)</td>
<td>CANVAS† (clinicians)</td>
<td>Pre-post</td>
<td>No significant reduction in violent incidents (after taking into account cohort effect)</td>
<td>*</td>
<td>No change</td>
</tr>
<tr>
<td>Daffern et al., 2009 (UK)</td>
<td>51 adult forensic inmates (M)</td>
<td>DASA &amp; HCR-20 (nurses)</td>
<td>Pre-post</td>
<td>No significant reduction in aggression after tool implemented</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td>Gunenc et al., 2017 (UK)</td>
<td>50 adult psy. patients (M)</td>
<td>START (clinicians)</td>
<td>Pre-post</td>
<td>No significant reduction in physical or verbal aggression</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td>Guy et al., 2015 (MS, USA)</td>
<td>208 youth on probation (M/F)</td>
<td>SAVRY (service counselors)</td>
<td>Comparison (matched)</td>
<td>No significant differences in nonviolent, status, or probation violation offenses, but fewer violent offenses</td>
<td>***</td>
<td>Mixed</td>
</tr>
<tr>
<td>Kling et al., 2011 (CA)</td>
<td>807 adult psy. patients (M/F)</td>
<td>Alert System (nurses)</td>
<td>Pre-post</td>
<td>Violence initially decreased in hospital overall but then increased (results were non-significant)</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td>Laura &amp; John Arnold Foundation, 2014 (USA)</td>
<td>Defendants from 120 countries (M/F, n &amp; age N.R.)</td>
<td>PSA† (judges)</td>
<td>Pre-post</td>
<td>After tool implemented, state had a 15% reduction in pretrial crime but did not control for cohort effect, report on statistical significance, or provide methodology</td>
<td>*</td>
<td>Decrease</td>
</tr>
<tr>
<td>Needham et al., 2004 (CH)</td>
<td>576 adult psy. inpatients (M/F)</td>
<td>BVC-R (nurses)</td>
<td>Pre-post</td>
<td>No significant reduction in aggressive incidents or attacks against persons, but positive results on one ward</td>
<td>**</td>
<td>No change (overall)</td>
</tr>
<tr>
<td>Vincent et al., 2016 (USA)</td>
<td>1,694 youth on probation (M/F)</td>
<td>SAVRY &amp; YLS/CMI (YPOs)</td>
<td>Pre-post (matched)</td>
<td>New petitions did not change at 5 of 6 sites; new petitions decreased at one site; new adjudications increased at one site</td>
<td>***</td>
<td>No change (for most sites)</td>
</tr>
</tbody>
</table>
Note. † No support for the predictive validity of the tool could be found. The summary ratings were: Decrease (violence or offending decreased in all or almost all analyses), No change (violence or offending did not significantly change in all or almost all analyses), Mixed (no clear overall pattern of results). The study appraisal ratings were: * (high limitations/risk of bias on RoB tool or study appraisal form for non-randomized studies), ** (medium limitations/risk of bias), *** (low limitations/risk of bias). AU = Australia; CA = Canada; CH = Switzerland; CM = case manager; drs. = doctors; M = male; M/F = male/female; MS = Mississippi, NL = Netherlands; N.R. = not reported; psyc. = psychiatric; SE = Sweden; UK = United Kingdom; USA = United States of America; YPO = youth probation officer.
Table 6
What Strategies Help to Improve the Utility of Risk Assessment Tools for Risk Management?

<table>
<thead>
<tr>
<th>Authors, Year (country)</th>
<th>Sample (gender)</th>
<th>Risk Tool (assessor)</th>
<th>Design</th>
<th>Outcome(s)</th>
<th>Results</th>
<th>Study Appraisal</th>
<th>Summary of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training (i.e., Tool Alone vs. Tool Plus Training)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonta et al., 2011 (CA)</td>
<td>143 adults on probation (M/F)</td>
<td>LSI-R (POs)</td>
<td>RCT</td>
<td>Need principle, offending</td>
<td>POs in STICS program were more likely than controls to discuss needs &amp; attitudes, but less likely to discuss education/employment; clients were less likely to reoffend</td>
<td>***</td>
<td>Improvement</td>
</tr>
<tr>
<td>Needham et al. (2004)</td>
<td>576 psych. inpatients</td>
<td>BVC-R (nurses)</td>
<td>Pre-post</td>
<td>Violence</td>
<td>No significant reduction in aggressive incidents or attacks in the tool + training condition vs. the tool alone condition</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td>Storey et al., 2011 (CA)</td>
<td>73 police, etc.</td>
<td>SARA, SAM, HCR-20 (police, etc.)</td>
<td>Pre-post, vignettes</td>
<td>Risk/need principles</td>
<td>After training, match between needs and planned management strategies increased; no change in number of strategies matched to risk level</td>
<td>***</td>
<td>Some improvement</td>
</tr>
<tr>
<td><strong>Risk Management Guidelines (i.e., Tool Alone vs. Tool Plus Guidelines)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosker et al., 2015 (NL)</td>
<td>29 POs</td>
<td>RISc2, RISc3 (POs)</td>
<td>Pre-post, vignettes</td>
<td>Risk/need principles</td>
<td>When tool + guidelines, agreement about planned intensity of supervision increased but was still poor; agreement about needs to target increased</td>
<td>**</td>
<td>Some improvement</td>
</tr>
<tr>
<td>Bosker &amp; Witteman, 2016 (NL)</td>
<td>579 adult probationers (M/F)</td>
<td>RISc2, RISc3 (POs)</td>
<td>Pre-post</td>
<td>Risk/need principles</td>
<td>When tool + guidelines, plans were more likely to be appropriate intensity, but mean match to needs did not improve</td>
<td>**</td>
<td>Some improvement</td>
</tr>
<tr>
<td>Daffern et al., 2009 (AU)</td>
<td>51 adult forensic inpatients (M)</td>
<td>DASA, HCR-20 (nurses)</td>
<td>Pre-post</td>
<td>Violence</td>
<td>No change in rates of aggression for tool + guidelines vs. tool alone</td>
<td>**</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Quality of Implementation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vincent, Guy, et al., 2012 (USA)</td>
<td>410 young offenders, (M/F)</td>
<td>SAVRY (YPOs)</td>
<td>Pre-post, files</td>
<td>Risk principle</td>
<td>After full implementation (i.e., policies, protocols) vs. standard SAVRY training, match to risk principle increased</td>
<td>***</td>
<td>Improvement</td>
</tr>
<tr>
<td>Vincent et al., 2016 (USA)</td>
<td>1,694 youth on probation (M/F)</td>
<td>SAVRY, YLS/CMI (YPOs)</td>
<td>Pre-post, files</td>
<td>Risk principle</td>
<td>Sites with better implementation (i.e., completed risk assessments) had better match to risk principle</td>
<td>***</td>
<td>Improvement</td>
</tr>
</tbody>
</table>

Note. The summary ratings were: Improvement (analyses indicated that the strategy was associated with improvement); Some Improvement (some but not all analyses indicated that the strategy was associated with improvement); No change (all or almost all of analyses indicated that the strategy was not associated with improvement). The study appraisal ratings were: * (high limitations [e.g., <50% on the study appraisal form]), ** (medium limitations [e.g., 51 – 79% on the study appraisal form]), *** (low limitations [e.g., >80% on the study appraisal form]). AU = Australia; CA = Canada; M = male; M/F = male/female; NL = Netherlands; N.R. = not reported; POs = parole or probation officer; STICS = Strategic Training in
Community Supervision; USA = United States of America; YPO = youth probation officer.
# Table 7
## Overall Summary of Findings

<table>
<thead>
<tr>
<th>Study Appraisal</th>
<th>Summary of Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limitations</td>
<td>Low</td>
</tr>
<tr>
<td>Low Limitations</td>
<td>k</td>
</tr>
<tr>
<td>Medium Limitations</td>
<td>k</td>
</tr>
<tr>
<td>High Limitations</td>
<td>k</td>
</tr>
<tr>
<td>Perceived Utility ($k = 12$)</td>
<td>5</td>
</tr>
<tr>
<td>Use for Risk Management ($k = 14$)</td>
<td>6</td>
</tr>
<tr>
<td>Match to Risk Principle ($k = 36$)</td>
<td></td>
</tr>
<tr>
<td>a. Comparison group ($k = 2$)</td>
<td>1</td>
</tr>
<tr>
<td>Ne</td>
<td>4</td>
</tr>
<tr>
<td>b. No comparison group ($k = 34$)</td>
<td></td>
</tr>
<tr>
<td>Match to Need Principle ($k = 17$)</td>
<td></td>
</tr>
<tr>
<td>a. Comparison group ($k = 0$)</td>
<td>–</td>
</tr>
<tr>
<td>b. No comparison group ($k = 17$)</td>
<td>1</td>
</tr>
<tr>
<td>Violence/Reoffending ($k = 12$)</td>
<td></td>
</tr>
<tr>
<td>a. RCTs ($k = 3$)</td>
<td>–</td>
</tr>
<tr>
<td>b. Non-Randomized ($k = 9$)</td>
<td>2</td>
</tr>
<tr>
<td>Strategies to Improve Utility ($k = 8$)</td>
<td></td>
</tr>
<tr>
<td>a. Training ($k = 3$)</td>
<td>1</td>
</tr>
<tr>
<td>b. Guidelines ($k = 3$)</td>
<td>–</td>
</tr>
<tr>
<td>c. Implementation Quality ($k = 2$)</td>
<td>2</td>
</tr>
</tbody>
</table>

*Note.* If the results were mixed, low-mixed, or high-mixed, they were coded as mixed. If the results were moderate, low-moderate, or high-moderate, they were coded as moderate.
**Databases**
PsycINFO, PsycARTICLES, PsycBOOKS, National Criminal Justice Reference Service, MEDLINE, Criminal Justice Abstracts, Google Scholar, Sociological Abstracts, Social Services Abstracts, Social Sciences Abstracts, Social Sciences Full Text, Web of Science, ProQuest Dissertations & Theses ($n = 1831$)

**Other Sources**
- Reference lists ($n = 75$)
- Requests from experts ($n = 16$)

---

**Figure 1.** Search strategy and phases of review.
Appendix A

List of Included Studies


39. Laura and John Arnold Foundation (2014). Results from the first six months of the public safety assessment – Court™ in Kentucky. Laura and John Arnold Foundation.


56. Simpson, T. P. (2010). Do objective measures reduce the disproportionate rates of minority youth placed in detention: Validation of a risk assessment instrument? (Ph.D., University of


### Appendix B

#### Description of Included Tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>Risk Prediction</th>
<th>Population</th>
<th>Focus on Modifiable or Dynamic Items (i.e., Needs)</th>
<th>Support for Predictive Validity?</th>
<th>Description of Predictive Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. AJADMS <em>(Arizona Juvenile Aftercare Decision-Making System; Ashford et al., 1987)</em></td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>No</td>
<td>No</td>
<td>No information could be identified</td>
</tr>
<tr>
<td>2. Alert System <em>(authors N.R.)</em></td>
<td>Inpatient violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>Sensitivity = .71, specificity = .94 for violent incidents <em>(Kling et al., 2006)</em></td>
</tr>
<tr>
<td>3. B-SAFER <em>(Brief Spousal Assault Form for the Evaluation of Risk; Kropp et al., 2005, 2010)</em></td>
<td>Intimate partner violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>AUC = .70 and .65 for total scores &amp; summary risk estimates respectively <em>(Storey, Kropp, Hart, Belfrage, &amp; Strand, 2014)</em></td>
</tr>
<tr>
<td>4. BVC <em>(Brøset Violence Checklist; Woods &amp; Almvik, 2002)</em></td>
<td>Short-term inpatient violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>AUC = .77 for any inpatient aggression <em>(Chu, Daffern, &amp; Ogloff, 2013)</em></td>
</tr>
<tr>
<td>5. CANVAS <em>(Clinical Assessment of Need, Violence Appraisal System; Guite et al., 1998)</em></td>
<td>Short-term inpatient violence</td>
<td>Not clear</td>
<td>Yes</td>
<td>No</td>
<td>No information could be identified</td>
</tr>
<tr>
<td>6. DASA <em>(Dynamic Appraisal of Situational Aggression; Ogloff &amp; Daffern 2006)</em></td>
<td>Short-term inpatient violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>AUC = .74 for all aggressive incidents <em>(Vojt, Marshall, &amp; Thomson, 2010)</em></td>
</tr>
<tr>
<td>7. DSI <em>(Rapides Parish Juvenile Detention Screening Instrument; authors N.R.)</em></td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>No</td>
<td>No</td>
<td>DSI did not significantly predict offending <em>(Simpson, 2010)</em></td>
</tr>
<tr>
<td>8. FOTRES <em>(Forensisch Operationalisiertes Therapie- und Risiko-Evaluations-System; Urbanio, 2007)</em></td>
<td>General recidivism</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>AUC = .81 and .76 for total score and summary risk rating respectively for repeat offending <em>(Rossegger et al., 2011)</em></td>
</tr>
<tr>
<td>9. HCR-20 <em>(Historical, Clinical, Violence)</em></td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
<td>AUC = .71 for violent recidivism</td>
<td></td>
</tr>
<tr>
<td>Risk Management-20; Webster et al., 1997</td>
<td>10. HKT-30 (Historical Clinical Future-30; Taskforce Risk Assessment Forensic Psychiatry, 2002)</td>
<td>Violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Intake Screen (Intake Screen for Risk and Needs; authors N.R.)</td>
<td>11. Intake Screen for Risk and Needs</td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>JAG (Juvenile Assessment Generic; authors N.R.)</td>
<td>12. JAG (Juvenile Assessment Generic; authors N.R.)</td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Domestic violence</td>
<td>13. L17A (Family Violence Risk Assessment and Management Report; authors N.R.)</td>
<td>Domestic violence</td>
<td>Adults</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>LSI-R (Level of Service Inventory-Revised; Andrews &amp; Bonta, 1995)</td>
<td>14. LSI-R (Level of Service Inventory-Revised; Andrews &amp; Bonta, 1995)</td>
<td>General recidivism</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MnSOST-R (Minnesota Sex Offender Screening Tool-Revised; Epperson et al., 1998)</td>
<td>15. MnSOST-R (Minnesota Sex Offender Screening Tool-Revised; Epperson et al., 1998)</td>
<td>Sexual reoffense risk</td>
<td>Adult sex offenders</td>
<td>No (mainly static)</td>
<td>Yes</td>
</tr>
<tr>
<td>NCCD Instrument (National Council on Crime and Delinquency; Wagner et al., 1994)</td>
<td>16. NCCD Instrument (National Council on Crime and Delinquency; Wagner et al., 1994)</td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>No</td>
<td>Yes (some)</td>
</tr>
<tr>
<td>OASys (Offender Assessment System; Home Office, 2002)</td>
<td>17. OASys (Offender Assessment System; Home Office, 2002)</td>
<td>General recidivism &amp; violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PRA (Primary Risk Assessment; Bonta et al., 2008)</td>
<td>18. PRA (Primary Risk Assessment; Bonta et al., 2008)</td>
<td>General recidivism</td>
<td>Adults/Adolescents</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PSA (Public Safety Assessment; Laura &amp; John Arnold Foundation, 2014)</td>
<td>19. PSA (Public Safety Assessment; Laura &amp; John Arnold Foundation, 2014)</td>
<td>General recidivism &amp; violence</td>
<td>Adults</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Test Description</td>
<td>Risk Type</td>
<td>Age Group</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>---</td>
<td>------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>20.</td>
<td>RISC (Recidivism Risk Assessment Scale, Version 2 and 3; Bosker et al., 2013)</td>
<td>General recidivism</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>21.</td>
<td>RM2000 (Risk Matrix, 2000; Thornton et al., 2003)</td>
<td>Sexual &amp; violent reoffense risk</td>
<td>Adult sex offenders</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>22.</td>
<td>RAI (Risk Assessment Instrument; authors N.R.)</td>
<td>Not specified</td>
<td>Adolescents</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>23.</td>
<td>SAM (Stalking Assessment and Management; Kropp et al., 2008)</td>
<td>Stalking &amp; violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>24.</td>
<td>SARA (Spousal Assault Risk Assessment; Kropp et al., 1994)</td>
<td>Intimate partner violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>25.</td>
<td>SAPROF (Structured Assessment of Protective Factors; de Vogel et al., 2007)</td>
<td>Protective factors for violence</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>26.</td>
<td>SAVRY (Structured Assessment of Violence Risk in Youth; Borum et al., 2006)</td>
<td>Violent recidivism</td>
<td>Adolescents</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>27.</td>
<td>SCJ:Risk (Structured Clinical Judgement: Risk Assessment Scheme; Hogue &amp; Allen, 2006)</td>
<td>Inpatient violence to self and others</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>28.</td>
<td>START (Short-Term Assessment of Risk and Treatability; Webster et al., 2009)</td>
<td>Violence, self-harm, etc.</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>29.</td>
<td>START:AV (Short-Term Assessment of Risk and Treatability: Adolescent Version; Nicholls et al., 2010)</td>
<td>Violence, victimization, etc.</td>
<td>Adolescents</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>30.</td>
<td>Static-99 (Hanson &amp; Thornton, 1999)</td>
<td>Sexual &amp; violent reconviction</td>
<td>Adult male sex offenders</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>31.</td>
<td>VRAG (Violence Risk Appraisal Guide; Quinsey et al., 1998)</td>
<td>Violent recidivism</td>
<td>Adults</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>No.</td>
<td>Tool Description</td>
<td>Routine Use</td>
<td>Application</td>
<td>Predictive Validity</td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>32.</td>
<td>VRS (Violence Risk Scale; Wong &amp; Gordon, 2009)</td>
<td>Violent recidivism</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>33.</td>
<td>WRNA (Wisconsin Risk/Needs Assessment Tools; Baird et al., 1979)</td>
<td>General recidivism</td>
<td>Adults</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>34.</td>
<td>YLS/CMI (Youth Level of Service/Case Management Inventory; Hoge &amp; Andrews, 2002), Y-LSI (Youth Level of Service Inventory; Andrews et al., 1984), Youth LSI-SK (LSI Saskatchewan Youth Edition; Andrews et al., 2001)</td>
<td>General recidivism</td>
<td>Adolescents</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

**Note.** AUC = area under the receiver operating curve, \( r = \) correlation, \( d = \) Cohen’s \( d \). In describing predictive validity, we first searched for meta-analyses. If meta-analyses were not available, we examined studies by independent researchers who were not authors of the tool. If independent studies were not available, we described studies by the tools’ authors. Although Singh et al. (2014) surveyed professionals about a variety of different tools, we only referenced the tools described in the text or tables (i.e., FORTES, HKT-30, HCR-20, SAPROF, START, VRS, and VRAG). When tools had multiple versions, we grouped our description together. Interpretative guidelines for predictive validity of tools (Rice & Harris, 2005): \( d = < .50 \) (small), \( .50-.80 \) (medium), \( > .80 \) (high); AUCs: \( < .64 \) (small), \( .64-.71 \) (medium), \( > .71 \) (high); \( r \) (assuming a 50% base rate) = \( < .24 \) (small), \( .24-.37 \) (medium), \( > .37 \) (high).
Appendix C

Reference List for Tools

Note. This reference list includes all of the references and studies that are cited in Appendix B.


