American Psychological Association
American Psychology-Law Society

Law and Human Behavior

ISSN: 0147-7307

2019, Vol. 43, No. 5, 397–420
http://dx.doi.org/10.1037/lhb0000344

Impact of Risk Assessment Instruments on Rates of Pretrial Detention, Postconviction Placements, and Release: A Systematic Review and Meta-Analysis

Jodi L. Viljoen, Melissa R. Jonnson, Dana M. Cochrane, and Lee M. Vargen
Simon Fraser University

Gina M. Vincent
University of Massachusetts Medical School

Objectives: Many agencies use risk assessment instruments to guide decisions about pretrial detention, postconviction incarceration, and release from custody. Although some policymakers believe that these tools might reduce overincarceration and recidivism rates, others are concerned that they may exacerbate racial and ethnic disparities in placements. The objective of this systematic review was to test these assertions. ***Hypotheses:*** It was hypothesized that the adoption of tools might slightly decrease incarceration rates, and that impact on disparities might vary by tool and context. ***Method:*** Published and unpublished studies were identified by searching 13 databases, reviewing reference lists, and contacting experts. In total, 22 studies met inclusion criteria; these studies included 1,444,499 adolescents and adults who were accused or convicted of a crime. Each study was coded by 2 independent raters using a data extraction form and a risk of bias tool. Results were aggregated using both a narrative approach and meta-analyses. ***Results:*** The adoption of tools was associated with (a) small overall decreases in restrictive placements (aggregated odds ratio [*OR*] = 0.63, $p < .001$), particularly for individuals who were low risk and (b) small reductions in any recidivism (*OR* = 0.85, $p = .020$). However, after removing studies with a high risk of bias, the results were no longer significant. ***Conclusions:*** Although risk assessment tools might help to reduce restrictive placements, the strength of this evidence is low. Furthermore, because of a lack of research, it is unclear how tools impact racial and ethnic disparities in placements. As such, future research is needed.

Public Significance Statement
Use of a risk assessment tool for pre or post-trial decisions may help reduce rates of incarceration while still protecting public safety. However, much of the available research is poor in quality. In addition, findings are inconsistent, and few studies have tested for racial and ethnic disparities. As such, there is a strong need for more rigorous research before clear conclusions can be drawn.

Keywords: risk assessment, violence, reoffending, incarceration, racial and ethnic disparities

Supplemental materials: http://dx.doi.org/10.1037/lhb0000344.supp

Risk of recidivism tools are widely used in criminal and juvenile justice settings. In some cases, these tools are used primarily to guide case management and treatment-planning. However, in other cases, tools are used to inform high stakes decisions about custodial placements. This includes front-end decisions about who to detain prior to trial, as well as later decisions about postconviction incarceration and release from prison (Monahan & Skeem, 2016). For instance, 88% of American pretrial agencies use risk tools to guide pretrial detention decisions (Pretrial Justice Institute, 2009), 20 states use them to guide sentencing decisions (Starr, 2014), and

This article was published Online First August 15, 2019.

Jodi L. Viljoen, Melissa R. Jonnson, Dana M. Cochrane, and Lee M. Vargen, Department of Psychology, Simon Fraser University; Gina M. Vincent, Law and Psychiatry Program, University of Massachusetts Medical School.

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.

One of the authors (Jodi L. Viljoen) is an author of a risk assessment tool; this tool was not included in the review. Gina M. Vincent is an author of two studies that were included in this review. However, she did not code or evaluate any of these studies. Instead, ratings were made by the other authors. We are very grateful to the researchers who responded to our requests for more information about their studies.

Correspondence concerning this article should be addressed to Jodi L. Viljoen, Department of Psychology, Simon Fraser University, Burnaby, BC V5A 1S6, Canada. E-mail: jviljoen@sfu.ca

up to 28 states use them to guide parole release decisions (Harcourt, 2007). In juvenile probation settings, close to 40 states have adopted risk tools on a state-wide basis for dispositional planning (Wachter, 2015). Furthermore, many organizations, policymakers, and scholars explicitly encourage the use of risk tools in placement decisions (e.g., American Bar Association, 2007; American Law Institute, 2014; National Association of Pretrial Services Agencies, 2004; National Research Council of the National Academy of Sciences, 2013).

Despite the anticipated benefits of risk tools, their impact on incarceration rates remains unclear. Do they decrease incarceration rates and enhance public safety, as some researchers and policymakers believe? And/or do they have unintended negative consequences such as increasing racial and ethnic disparities, as critics argue? To help answer these questions, we conducted a systematic review and meta-analysis. To set the stage for this review, we begin by discussing the relevance of risk to placement decisions.

The Role of Risk in Placement Decisions

In making decisions about whether to detain defendants prior to trial, defendants' risk to others is often a key consideration (Myburgh, Camman, & Wormith, 2015). According to recent estimates, 48 states and the District of Columbia have laws permitting courts to consider defendants' dangerousness in bail and pretrial detention hearings (Baradaran & McIntyre, 2012). Risk is also relevant to postconviction or postadjudication decisions about sentences (Monahan & Skeem, 2016). Specifically, within a utilitarian model (Bentham, 1789/2000), the goal of sentencing is to protect society; reoffense risk is important, as it relates directly to public safety (Monahan & Skeem, 2016). Risk also plays a role within limited retribution sentencing models (Monahan & Skeem, 2016). In this model, sentences should be tied to moral concerns about culpability (Morris, 1974). However, considerations of risk might be used to bump someone up or down within the range of possible penalties (Monahan & Skeem, 2016; Skeem & Lowenkamp, 2016).

Despite the relevance of risk to legal decisions, some jurisdictions do not formally or explicitly assess risk for recidivism with instruments. This does not mean that considerations of risk are averted. Instead, in such cases, judges and other legal professionals likely rely on their own subjective impressions about offenders' dangerousness to others (Tonry, 1987; Vigorita, 2003). As research has demonstrated, these subjective impressions of risk are more vulnerable to inaccuracies than judgments made using an empirically supported risk tool (Ægisdóttir et al., 2006; Hanson & Morton-Bourgon, 2009).

Some Believe Tools Will Decrease Incarceration and Enhance Public Safety

Many scholars and policymakers believe that risk tools not only improve the accuracy of risk predictions, but also minimize incarceration rates so that incarceration is only used when necessary (Austin, 2004; Elek, Warren, & Casey, 2015; Kopkin, Brodsky, & DeMatteo, 2017; Laura & John Arnold Foundation, 2014; Vincent, Guy, & Grisso, 2012). After decades of "get tough" laws, many states are now faced with inordinately high rates of incarceration, which has proven costly and unsustainable (Clear & Frost, 2014; Tonry, 2017). Thus, some states have adopted tools as part of an effort to reduce incarceration (La Vigne et al., 2014).

There are several mechanisms by which the adoption of tools could reduce placements (see van Wingerden, van Wilsem, & Moerings, 2014). First, tools might provide judges with information about modifiable factors, thereby mitigating the need for more restrictive placements; second, they might help to reclassify offenders who would otherwise be assumed to be high risk; and third, they may help judges to resist public political pressures to get tough on crime by providing them with greater justification for decisions to divert or release low risk offenders. In addition, one of the appealing features of risk tools is that they might enable more strategic decisions, wherein high-risk offenders are incarcerated but low-risk offenders are not (Laura & John Arnold Foundation, 2014). This is consistent with the risk principle of the risk-need-responsivity (RNR) model (Bonta & Andrews, 2017). For example, according to the Laura and John Arnold Foundation (n.d., p. 1), the use of tools "can help to ensure that the relatively small number of defendants who need to be in jail remain locked up—and the significant majority of individuals who can be safely released are returned to the community."

If risk tools do facilitate match to the risk principle, they might reduce incarceration without increasing reoffending (Casey, Warren, & Elek, 2011; Laura & John Arnold Foundation, 2014; Thompson, 2017). According to some authors, the use of tools might even lead to decreases in offending by helping to ensure that high risk offenders are not released prematurely without sufficient supports, and by helping to divert low risk offenders so that they avoid the harmful effects of incarceration (Austin, 2004; Casey et al., 2011). However, it is unclear what evidence supports these views and, as such, we tested this in this systematic review.

Others Believe Tools May Exacerbate Racial and Ethnic Disparities

Despite the potential benefits of risk tools, some policymakers and scholars have expressed concerns that any benefits might be "offset by costs to social justice" (see Monahan, Skeem, & Lowenkamp, 2017, p. 191). More specifically, tools might lead to more punitive sanctions for racial and ethnic minority groups, such as African Americans and Indigenous populations, who are overrepresented in justice settings (Harcourt, 2015; Holder, 2014; Maurutto & Hannah-Moffat, 2007; Petersilia & Turner, 1987; Starr, 2014). For instance, Eric Holder, the former attorney general of the United States, asserted, "Although these measures were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice" (Holder, 2014, para. 23). Legal scholar Starr (2014) argued that tools can create a scientifically rationalized guise for discrimination.

The reason for this concern is that even though tools do not directly include race or ethnicity as a consideration, people of color sometimes receive higher scores on tools than nonminorities (e.g., Skeem & Lowenkamp, 2016). For instance, people of color are more likely to experience social disadvantage and poverty, and may have fewer opportunities for education and employment, which could lead to inflated risk scores (Maurutto & Hannah-Moffat, 2007). Higher scores, in turn, could be used to justify

harsher sentences. However, although some policymakers and scholars believe that tools will exacerbate disparities, others believe that risk assessment tools are preferable to the alternative, namely unstructured decision-making (Eaglin & Solomon, 2015; Hoge, 2002; Thompson, 2017), as disparities are common even when tools are not used (e.g., Bridges & Steen, 1998; Graham & Lowery, 2004; Steffensmeier, Ulmer, & Kramer, 1998).

Before researchers can offer conclusions, more data are needed, including studies on (a) test bias (e.g., whether tools predict equally well across groups) and (b) disparate impact (i.e., whether tools lead to inequitable decisions that may be morally unfair; Skeem & Lowenkamp, 2016). As Skeem and Lowenkamp (2016) explained, even if instruments are not necessarily biased, they could nevertheless "create disparate impact" if racial and ethnic minority groups have higher average scores than nonminorities (p. 685). However, these researchers note that it seems unlikely that well-validated, unbiased instruments would create more disparate impact than the status quo (i.e., subjective decisions about risk).

Thus far, some studies have reported that, in some cases, African Americans and Indigenous people may receive higher scores than Whites on certain risk factors (e.g., Perrault, Vincent, & Guy, 2017; Skeem & Lowenkamp, 2016) or on total scores (e.g., Olver, Stockdale, & Wong, 2012; Shepherd, Luebbers, Ferguson, Ogloff, & Dolan, 2014). However, this depends on the risk instrument used. Furthermore, even though higher scores could raise the *possibility* that certain tools may increase racial and ethnic disparities in incarceration rates, comparing mean differences in scores across groups does not provide a direct test of how tools impact placement decisions. As such, in the present review, we synthesized research that tested how tools affect rates of restrictive placements for people of color.

The Impact of Tools May Depend on the Tool and Other Factors

Although some authors advocate for tools and others oppose them, tools themselves differ considerably and, thus, their impact on incarceration rates may vary. Some tools contain primarily historical or static factors, such as prior offenses; others focus on dynamic or modifiable risk factors (i.e., needs), such as substance abuse. Maurutto and Hannah-Moffat (2007) argued that dynamic measures may inadvertently lead to harsher penalties for minority groups because such measures conflate risk with rehabilitative needs. However, other researchers argue that static measures may lead to harsher penalties for minorities because static factors (e.g., offense history) are more highly correlated with race than dynamic factors (Perrault et al., 2017; Skeem & Lowenkamp, 2016; Vincent, Chapman, & Cook, 2011).

Risk tools also vary in the level of discretion they allow (Skeem & Monahan, 2011). In structured professional judgment tools, assessors do not add up scores. Instead, they make their own judgment about risk level, drawing from case-specific information and their professional opinion. In contrast, in actuarial tools, assessors sum items to create an overall score, which is often used to generate a specific numerical risk estimate (e.g., 10–20% of offenders with similar scores reoffend within a 5-year period). Hart (2011) cautioned that if professionals claim that they can identify high risk offenders with high specificity, then policymakers will, naturally, "target these people for extreme incapacitative measures" (p. 67), thereby using risk assessments to justify "draconian political decisions and social policies" (p. 67). Thus, in this review, we compared whether the impact of tools depends on factors such as the type of tool.

Present Study

In sum, some authors argue that risk tools could help reduce mass incarceration without jeopardizing public safety, whereas others argue that these tools may exacerbate racial disparities in sentencing. However, it is currently unclear which perspectives are accurate. Although a recent systematic review examined how risk tools impact treatment-planning and risk management (Viljoen, Cochrane, & Jonnson, 2018), that review did not examine how the adoption of tools affects overall rates of placements. As such, we conducted a systematic review and meta-analysis to test the following research questions:

1. Does the adoption of risk tools decrease restrictive placements (i.e., pretrial placements, postconviction incarceration, release from secure facilities)?
2. If so, are these findings due to confounds or study biases? Or do findings remain similar even when only the highest quality studies are examined?
3. Which factors moderate or influence the effect of tools on rates of restrictive placements (e.g., type of tool)?
4. When tools are adopted in sentencing, do rates of recidivism and violations change?
5. How does the adoption of risk tools impact racial and ethnic disparities in restrictive placements?

Our overarching aims were to inform debates about the potential benefits and costs of risk tools and create an agenda for future research.

Method

To ensure that we reported our systematic review in a thorough, rigorous, and transparent manner, we followed criteria set forth in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (Moher, Liberati, Tetzlaff, & Altman & the PRISMA Group, 2009), the AMSTAR 2 tool (A MeaSurement Tool to Assess systematic Reviews 2; Shea et al., 2017), and the Risk of Bias in Systematic Reviews tool (Whiting et al., 2016). Our review question, search strategy, inclusion/exclusion criteria, data extraction materials (e.g., risk of bias assessment), and data analytic plan were established a priori.

Step 1: Search

To identify relevant studies (published and unpublished), we searched 13 databases (e.g., Criminal Justice Abstracts, PsycINFO, ProQuest Dissertations & Theses, Google Scholar; see Figure 1) using the following terms: "risk assessment" AND (violen* OR reoffen* OR recidivism OR offen*) AND ("sentencing" or "incarceration" or "sanctions"). These searches encompassed all time

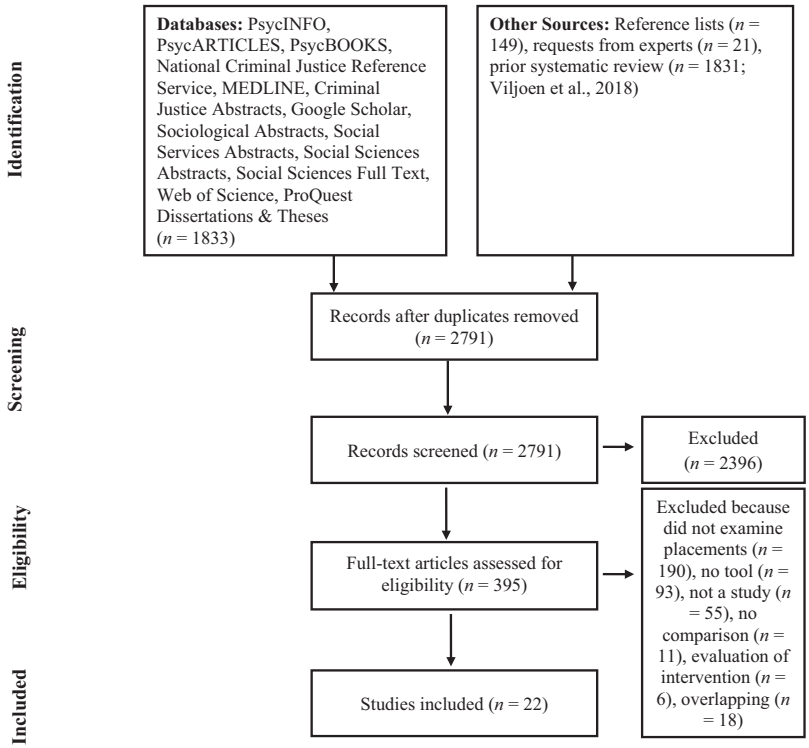

Figure 1. Search strategy.

periods up to August 31, 2017. Although researchers typically restrict Google Scholar searches to the first 50 to 100 search records (Haddaway, Woodcock, Macura, & Collins, 2015), we examined the first 300 records identified in Google Scholar. To identify additional studies, we reviewed the reference lists of included studies and contacted 24 experts (i.e., authors of included studies). In addition, we reviewed the abstracts of studies identified via a prior systematic review on the utility of risk assessment tools for risk management (see Viljoen et al., 2018).

Step 2: Screening and Eligibility Criteria

After removing duplicates via RefWorks, we identified 2,791 disseminations through the above-described searches. Two authors then reviewed the abstracts and titles to determine if they met eligibility criteria. To help ensure that our screening was reliable and accurate, they completed 25 practice cases, and correctly screened in each of the eligible studies. To be included, studies had to (a) include a sample of offenders who were assessed with a structured risk assessment tool in real-world practice, (b) include a comparison group of offenders who were not assessed with a tool, and (c) examine how the use of tools influenced restrictive placements (i.e., pretrial detention, postconviction incarceration, release). We defined structured risk assessment tools as tools that included a list of risk factors, guidelines for rating these factors, and an overall risk rating (see Skeem & Monahan, 2011). We did not restrict our review to certain types of designs, such as randomized control trials, because we expected such studies would be scarce and we wished to synthesize all available research, nor did we restrict our review based on the publication date or language (i.e., non-English studies were included in our search).

Step 3: Full Text Review

Next, we conducted a full text review of the 395 abstracts that were initially screened in. Of these, 22 studies met inclusion criteria. Included studies are marked with a star in the reference list. Most of the remaining studies did not meet the prespecified inclusion criteria ($n = 349$). For instance, upon review (and contacts with authors, as needed), we determined that some studies did not examine rates of placement or did not include a risk assessment tool ($n = 190$ and 93, respectively; see Figure 1). Also, in 11 studies, there was no comparison group, or the comparison group was already using some type of tool (e.g., Berk, 2017; Cadigan & Lowenkamp, 2011; Turner, Braithwaite, Kearney, Murphy, & Haerle, 2012). Six studies were excluded because they focused on evaluating a comprehensive initiative or intervention program rather than a tool (e.g., Schweitzer Smith, 2017). We also excluded overlapping studies ($n = 18$). When disseminations were based on the same sample and timeframe, we selected the study that was the most comprehensive (e.g., Stevenson, 2018, rather than Laura & John Arnold Foundation, 2014).

Step 4: Data Extraction and Consensus Ratings

To increase objectivity and replicability of our ratings, each of the 22 included studies was independently coded by two study authors. We then held consensus meetings to discuss disparate

ratings. When the two raters could not reach a consensus, the first author (who reviewed all studies) made a rating. Each of these raters (three graduate students, one faculty member) had prior coursework and applied experience with risk instruments. In addition, raters completed approximately 5 hr of training on the study protocol (e.g., practice cases, quizzes).

Data extraction form. Using a 56-page rating form (available upon request), raters extracted information about the study characteristics (e.g., publication type), sample, design, risk assessment tool, and results (e.g., potential moderators). When the study did not include adequate information to code an effect size, we contacted the authors for further information. Cohen's kappa coefficients for age of the sample (i.e., adult, adolescent), sample (i.e., pretrial, other), and study design (i.e., randomized control trial, comparison, pre–post) were .89, .88, and .84, respectively (n = 20). These values fell in the "almost perfect" range ($\kappa > .80$; Landis & Koch, 1977). The intraclass correlation coefficient (ICC; two-way mixed, absolute agreement, average measures; McGraw & Wong, 1996) for sample size was 1.00.

Summary ratings. Next, raters made independent summary ratings of study findings, namely, the impact of the tool on rates of restrictive placements, recidivism, and minority confinement (i.e., decreases, mixed, no change, increases). ICCs (two-way mixed, absolute agreement, average measures) were .94 for restrictive placements (n = 20) and .75 for minority confinement (n = 5). These values fell in the excellent range (i.e., ≥.75; Cicchetti, 1994). However, the ICC for recidivism was lower and fell in the fair range (.49, n = 9), possibly because reoffense type was not clearly defined. As such, we separated forms of reoffending (e.g., any, violent) and recoded outcomes. This resulted in improved ICCs (1.00 for any recidivism, violent recidivism, and violations).

Risk of bias. Finally, raters appraised the quality of studies and risk of bias with the Risk of Bias in Non-Randomized Studies—of Interventions (ROBINS-I; Sterne, Hernán, et al., 2016; Sterne, Higgins, et al., 2016). On this tool, raters examine bias in seven domains (i.e., confounding factors, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, selective reporting), and then make an overall rating of bias (i.e., Low, moderate, serious, critical, or no information). ICCs (two-way mixed, absolute agreement, average measures) fell in the excellent range for the overall rating (.85, n = 22).

Step 5: Analyses

Quantitative syntheses (i.e., meta-analyses). To synthesize our findings, we used a mixed methods approach, which included (a) a quantitative synthesis and (b) a narrative or qualitative synthesis (Gough, 2015). In our quantitative synthesis, we conducted a meta-analysis of aggregated odds ratios (OR) using Comprehensive Meta-Analysis Version 2 (Borenstein, Hedges, Higgins, & Rothstein, 2005). We used random-effect models because (a) we anticipated that the results might vary across studies, and (b) we wished to generalize findings beyond the particular studies included in the meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2010; Hedges & Vevea, 1998). To examine heterogeneity between studies, we calculated a within-group Q statistic (Q_w), which tests the presence or absence of heterogeneity, and Higgins I^2, which is interpreted as an indication of the proportion of variance due to heterogeneity (an I^2 = 25% is low, 50% is medium, and 75% is high; Huedo-Medina, Sánchez-Meca, Marín-Martínez, & Botella, 2006). In addition to performing an overall meta-analysis, we performed subgroup analyses to examine the impact of tools on pretrial detention and postconviction sentencing. If fewer than three studies were included in an aggregated effect size, we did not empirically synthesize the findings.

Narrative synthesis. Our narrative synthesis complemented our meta-analysis in two respects. First, given that many studies did not include the information necessary to include them in the meta-analysis, our narrative synthesis allowed us to draw from a broader pool of studies, thereby more fully capturing the literature. Second, it enabled us to examine more nuanced issues, such as possible confounds and moderators (Gough, 2015; Popay et al., 2006). In our narrative synthesis, we first created evidence tables, which summarized the methods and findings of each study. Then, we calculated basic descriptive statistics of our summary ratings (i.e., frequency counts), and identified themes and patterns that raters identified.

Step 6: Overall Strength of Evidence

After conducting our syntheses, three authors independently graded the overall strength of evidence for whether tools reduce placements and recidivism rates using the Agency for Health care Research and Quality system (AHRQ; Berkman et al., 2015). On the AHRQ, evaluators rate a body of research on five domains (i.e., study limitations, consistency, directness, precision, and reporting bias) and then grade the overall strength of evidence as High, Moderate, Low, or Insufficient. Each rater had prior training and experience with the AHRQ. The raters obtained unanimous agreement.

Results

Description of Included Studies

In total, 22 studies were included, with an aggregated sample size of 1,444,499 individuals who were accused of or convicted of a crime. These studies reported separate data for 30 independent sites. Half of the studies were unpublished reports that were not peer-reviewed, such as reports written by government agencies or foundations (50.0%, k = 11), and almost all studies were conducted in the United States (86.4%, k = 19). Although most studies focused on projects conducted in the 2000s (81.8%, k = 18), five studies were conducted during the 1980s or 1990s (18.2%). Over half of the studies (59.1%, k = 13) were funded by private foundations (e.g., Vera Institute for Justice), five (22.7%) by government granting agencies (e.g., U.S. Department of Justice), two (9.1%) were not funded, and two (9.1%) did not provide funding information.

Slightly over half of the studies focused on adolescent samples in the juvenile system (59.1%, k = 13), whereas the remainder focused on the adult system. Most studies focused on pretrial detention (63.6%, k = 14). However, five studies (22.7%) examined placements following conviction/adjudication and three studies (13.6%) examined release from jail or prison. Only one study (4.5%) used a randomized comparison group. Instead, most studies used a pre–post design (77.3%, k = 17); four of these pre–post

studies (18.2%) used propensity score matching to minimize group differences. Also, three studies (13.6%) had a nonrandomized comparison group, in which they compared sites that used a tool to sites that did not.

In total, 17 different risk tools were used in the studies (see Appendix A). All tools used in pretrial settings were brief screening measures (i.e., 13 items or less), which focused largely on static factors (e.g., offense history, current offense, age). In contrast, except for one measure, the tools used in studies on postconviction or release decisions were risk-needs assessment instruments, which were lengthier (i.e., 30 items or more) and contained both static and dynamic risk factors (e.g., attitudes, peers, family). Whereas all tools used in the postconviction or release decisions had evidence to support their predictive validity (100%, $k = 5$), we were unable to locate any validation studies for 36.3% of the pretrial tools ($k = 4$). On most tools (94.7%, $k = 16$), the final risk judgment was derived numerically by adding up total scores. Only one of the tools used a structured professional judgment approach (i.e., Structured Assessment of Violence Risk in Youth [SAVRY]; Borum, Bartel, & Forth, 2006). However, all tools appeared to provide some discretion in final risk judgments, such as the option to override total scores.

Question 1: Does the Adoption of Risk Tools Decrease Restrictive Placements (i.e., Pretrial Placements, Postconviction Incarceration, Release From Secure Facilities)?

Based on our narrative review and coding of the full set of 22 studies, 68.2% of the included studies found that the use of tools was associated with decreases in restrictive placements at some phase of the proceedings (see Table 1 for a study-by-study description of findings and Table 2 for a summary of results). The results of published, peer-reviewed studies (e.g., in academic journals) and unpublished studies were similar; 72.7% ($k = 8$) of published studies reported decreases in placements compared to 63.6% ($k = 7$) of unpublished studies. In addition, the results of studies with juveniles and adults were similar; 69.2% of juvenile studies ($k = 9$) and 66.7% of adult studies ($k = 6$) reported decreases in placements.

Although only 13 studies (with 21 separate effects) contained the necessary statistical information to be included in the meta-analysis (e.g., sample size, effect size), the meta-analysis yielded similar results as our narrative review. The aggregated random-effect *OR* was significant, but small (Chen, Cohen, & Chen, 2010; Chinn, 2000), and indicated that when tools were used, offenders were 63% as likely to receive a placement (see Table 3 and the online supplementary materials for forest plots). However, heterogeneity was high (Huedo-Medina et al., 2006).

As such, we examined whether the impact of tools might vary depending on the phase of sentencing. Overall, 64.3% of the studies that examined pretrial placements found that the adoption of tools was associated with a decrease in placements, as did 60.0% of the studies that examined postconviction placements, and 100% of the studies that examined release from custody (see Table 2). Based on a meta-analysis of the available results, offenders were about half as likely to receive pretrial detention when tools were used (aggregated *OR* = .52; see Table 3). However, the results for postconviction placements were nonsignificant, and it was not possible to meta-analyze results for studies on release from custody because only one study reported the necessary information.

Question 2: Can These Decreases in Placements Be Explained by Confounds or Biases?

Although we found modest decreases in rates of restrictive placements, we wished to examine whether this finding could be due, in part, to biases. This was important because even though some studies were very rigorous, over half of studies (59.1%, $k = 13$) were rated as having a "serious" risk of bias on the ROBINS-I (see Appendix B). Out of the domains evaluated with the ROBINS-I, the most common source of serious risk of bias was confounding factors (45.5%, $k = 10$). In particular, most studies did not match offenders in the tool and no-tool groups on characteristics such as age or offense history. As such, lower placement rates could be due to group differences. For instance, if the group assessed with a tool had fewer high-risk offenders than the group not assessed with a tool, then the lower rates of placements could be due to this lower risk level rather than the tool. In addition, even though rates of incarceration have declined in the United States over the past decade (Carson, 2018), and these historical trends might thus explain the observed decreases in placements, few studies accounted for this possibility. Another common bias arose from cointerventions; 31.8% of studies ($k = 7$) were rated as having a serious risk of this type of bias. For instance, several studies were conducted as part of the Juvenile Detention Alternatives Initiative (JDAI; Annie E. Casey Foundation, 2017). Though tools are a "centerpiece" of this initiative (Maloney & Miller, 2015), the JDAI includes other strategies to reduce detention, such as community collaboration and enhanced alternatives to detention (Mendel, 2014). As such, it is difficult to determine if reductions in placements were due to the adoption of tools or these other strategies.

Given these potential biases, we removed studies that had a serious risk of bias and reran our analyses with the remaining nine studies (16 separate effects; see Table 3 for a list of these studies). In contrast to the overall findings presented above, only 55.6% of the higher quality studies ($k = 5$) found reductions in restrictive placements, and the aggregated *OR* was no longer statistically significant ($p = .122$; see Table 3). However, most of the data that could be meta-analyzed focused on postconviction placements, and these studies found inconsistent results. For instance, in a rigorous study that used propensity-score matching, van Wingerden et al. (2014) found that incarceration rates were lower when the Recidivism Assessment Scale was used in sentencing than when it was used after sentencing when placement decisions had already been made. In another rigorous study with propensity score matching, Vincent, Guy, Perrault, and Gershenson (2016) found that, following implementation of the SAVRY or Youth Level of Service/Case Management Inventory (Hoge & Andrews, 2002), postadjudication placements decreased at two of the six sites, remained similar at three sites, and increased at one site.

As such, according to our rating on the AHRQ (Berkman et al., 2015), the overall strength of evidence that risk tools reduce restrictive placements is low because (a) the results were attenuated after removing studies with a serious risk of bias, and (b) the magnitudes of the effects were inconsistent (e.g., heterogeneity

Table 1
Does the Adoption of Risk Tools Decrease Restrictive Placements?

Authors, year (state, country)	Sample (gender)	Risk tool (assessors)	Other initiatives	Design	Results	Summary
Pretrial						
Annie E. Casey Foundation, 2017 (USA)	>284,887 adolescents, 164 sites (M/F)	RAI (user N.R.)	JDAI	Pre–post	Decrease in annual detention admissions of 49%; decrease in average daily detention population of 43%	Less restrictive
Bazemore, 1993 (FL, USA)	Approx. 3,000 adolescents (M/F)	RAI (detention staff)	Detention criteria, etc.	Pre–post	Decrease in centers overcapacity from 80% to 38% of centers; decrease in daily detention population from approx. 1,500 to 1,250	Less restrictive
Cooprider, 2009 (IL, USA)	Adults (*n* and gender N.R.)	Lake County Pretrial RAI (pretrial officer)	—	Pre–post	Increase in release without bond (from 16% in 2005 to 24% in 2007)	Less restrictive
Feyerherm, 2000 (OR, USA)	18,788 adolescents (M/F)	Multnomah RAI (detention staff)	JDAI	Pre–post	Decrease in pretrial detention (from 18% to 9%)	Less restrictive
Fratello et al., 2011 (NY, USA)	5,173 adolescents (M/F)	New York City RAI (POs)	Alternatives to detention	Pre–post	Decrease in use of detention (from 32% to 24%)	Less restrictive
Goldkamp & Gottfredson, 1985 (PA, USA)	1,800 adults (M/F)	Philadelphia Bail Guidelines (detention staff)	—	RCT	No change in overall use of pretrial detention but were more likely to release lower risk defendants	**No change**
Maloney & Miller, 2015 (NJ, USA)	1,432 adolescents (M/F)	RAI for NJ (detention staff)	JDAI	Pre–post (matched)	Decrease in detention (from 67% to 40%)	Less restrictive
Orlando, 1999 (FL, USA)	Adolescents (*n* N.R.)	Cook County RAI, etc. (detention staff)	JDAI	Pre–post	In some cases, detention decreased but in one case, unexpected initial increases in detention due to poor validation or amendments (i.e., Cook County; see also Bishop & Griset, 2001)	Mixed
Puzzanchera et al., 2012 (PA, USA)	>2,098 adolescents (M/F)	Allegheny DAI (detention staff or YPO)	JDAI	Pre–post	Decrease in detention from ~21% in 2007 to ~15% in 2009, but detention was declining even before tool was implemented	Less restrictive
Schwartz et al., 1991 (FL, USA)	20,227 adolescents (M/F)	Tool N.R. (N.R.)	Alternatives to detention	Pre–post, comparison	Decrease in secure detention by 22% (the rest of the state had a 6% decrease)	Less restrictive
Simpson, 2010 (LA, USA)	202 adolescents (gender N.R.)	Rapides Parish DSI (detectives)	—	Pre–post	Small non-significant reduction in detention (22 youth detained posttool vs. 27 youth pretool)	No change
Stevenson, 2018 (KY, USA)	1,030,732 adults (M/F)	Kentucky Pretrial RAI, PSA (pretrial staff)	—	Pre–post	Release initially increased by 4% after tool mandated by law but then reverted to usual practices	**Mixed**

(*table continues*)

Table 1 (*continued*)

Authors, year (state, country)	Sample (gender)	Risk tool (assessors)	Other initiatives	Design	Results	Summary
Toborg et al., 1984 (DC, USA)	34,291 adults (M/F)	DC Pretrial RAI	—	Pre–post	Increase in unrestricted releases from 1% to 12% but overall release did not change	**Less restrictive**
VanNostrand, 2016 (OH, USA)	34,763 adults (M/F)	PSA (user N.R.)	—	Pre–post	Increase in pretrial detention (from 17% to 23%)	More restrictive
Postconviction or adjudication						
Guy et al., 2015 (MS, USA)	110 adolescents (M/F)	SAVRY (service counsellors)	—	Comparison (matched)	Fewer placements *following adjudication* when tool was used (0% vs. 5%); however, placements *over the follow-up* did not vary between groups	**Less restrictive**
van Wingerden et al., 2014 (NL)	6,118 adults/ adolescents (M/F)	RISc (POs)	—	Comparison (matched)	Decrease in detention (from 66% to 61%), particularly for high and medium risk groups	**Less restrictive**
Vincent et al., 2016 (USA)—six sites	1,694 adolescents (M/F)	SAVRY YLS/ CMI (YPOs)	—	Pre–post (matched)	Decrease in placements at *adjudication* at 1 of 6 sites; decrease in placements *over follow-up* at 2 of 6 sites and increase at 1 site	**Mixed**
Vincent & Perrault, 2018 (AR, USA)—four sites	754 adolescents (M/F)	SAVRY (YPOs)	—	Pre–post (matched)	No change in detention disposition at the 4 sites (however were more likely to be diverted at 2 of the 4 sites); any postdisposition placements increased at 2 of 4 sites	**Mixed (more restrictive in 2/4 sites)**
Virginia Sentencing Commission, 2012 (VA, USA)	Adults (*n* N.R., M/F), 6 pilot sites	Nonviolent Risk Assessment (POs)	—	Pre–post, comparison	Increase in diversion increased by ~30% in sites using tool vs. 4% for sites no using tool	Less restrictive
Release						
Barnes-Ceeney, 2013 (NJ, USA)	445 adolescents (M/F)	SAVRY (psychologists)	—	Pre–post, case series	When tool used, were more likely to be released early (i.e., 1.71 times *less* likely to max out sentence)	**Less restrictive**
Bonta & Motiuk, 1987 (Canada)	378 adults (male)	LSI (N.R.)	—	Pre–post	For low-scoring inmates, transfer to halfway house was higher when tool used (59% vs. 32%)	Less restrictive
Bonta & Motiuk, 1990 (Canada)	580 adults (male)	LSI (classification staff)	—	Comparison	For low risk inmates, release to halfway house was higher when tool used (51% vs. 16%); no difference higher risk inmates	**Less restrictive**

Note. AR = Arkansas; DAI = Detention Assessment Instrument; DC = District of Columbia; DSI = detention screening instrument; FL = Florida; M/F = male/female; IL = Illinois; KY = Kentucky; LA = Louisiana; LSI = Level of Service Inventory; MS = Mississippi; NE = Nebraska; NJ = New Jersey; NL = Netherlands; N.R. = not reported; OH = Ohio; OR = Oregon; PA = Pennsylvania; PO = probation officer; PSA = Public Safety Assessment; RAI = risk assessment instrument; RISc = Recidivism Assessment Scales; SAVRY = Structured Assessment of Violence Risk in Youth; USA = United States of America; VA = Virginia; YPO = youth probation officer. Studies with low or moderate overall risk of bias are bolded to indicate that more weight should be given to these studies; the remaining studies have serious risk of bias. The summary ratings are defined as follows: Less restrictive = all or most analyses indicated a decrease in restrictive placements; Mixed = studies showed an inconsistent pattern of results; No change = all or most analyses indicated that restrictive placements did not significantly change; More restrictive = all or most analyses indicated an increase in restrictive placements.

Table 2
Narrative Synthesis: Summary Ratings of Study Findings

Outcomes	Summary of findings							
	Decrease		Mixed		No change		Increase	
	%	*k*	%	*k*	%	*k*	%	*k*
Restrictive placements								
Overall placements ($k = 22$)	68.2	15	18.2	4	9.1	2	4.5	1
Pretrial placements ($k = 14$)	64.3	9	14.3	2	14.3	2	7.1	1
Postconviction placements ($k = 5$)	60.0	3	40.0	2	.0	0	.0	0
Release ($k = 3$)	100.0	3	.0	0	.0	0	.0	0
Recidivism								
Any recidivism ($k = 10$)	20.0	2	.0	0	80.0	8	.0	0
Violent recidivism ($k = 5$)	40.0	2	.0	0	60.0	3	.0	0
Violations ($k = 8$)	25.0	2	.0	0	62.5	5	12.5	1
Racial and ethnic disparities								
Placements of minorities ($k = 6$)	83.3	5	.0	0	16.7	1	.0	0
Overrepresentation and disparities ($k = 5$)	40.0	2	20.0	1	20.0	1	20.0	1

Note. k = number of studies. The summary of findings is defined as follows: Decrease = studies that showed decreases in restrictive placements, recidivism, or disparities; Mixed = studies that showed mixed results; No change = studies that showed no significant changes in placements, recidivism, or disparities; Increase = studies that showed increases in placements, recidivism, or disparities.

was high). Given the heterogeneity in findings, we identified potential moderators next.

Question 3: Which Factors Moderate the Effect of Tools on Restrictive Placements?

Risk level. According to the risk principle of the RNR model, tools should decrease placements to a greater extent for people who are low risk compared to those who are high risk (see Bonta & Andrews, 2017). Our results were consistent with this principle. Of the six studies that reported rates of placements separately by risk level, all but one found reductions in placements for youth or adults who were low risk (Bonta & Motiuk, 1987, 1990; Fratello, Salsich, & Modulescu, 2011; Stevenson, 2018; van Wingerden et al., 2014; cf. Barnes-Ceeney, 2013). In contrast, the impact of tools

Table 3
Meta-Analysis: Impact of Tools on Restrictive Placements and Recidivism

Outcomes	*k*	Random-effect models					Heterogeneity		
		OR_w	95% CI		Z	*p*	*Q*	*p*	I^2
Restrictive placements									
Overall placements[a]	21	.63	.48	.82	−3.47	.001	1443.04	<.001	98.61
Excluding studies with serious risk of bias[b]	16	.70	.44	1.10	−1.55	.122	737.64	<.001	97.97
Pretrial placements[c]	8	.52	.36	.75	−3.45	.001	1398.65	<.001	99.50
Placements following sentencing/adjudication[d]	12	.86	.59	1.26	−.77	.445	31.90	.001	65.52
Recidivism									
Any recidivism[e]	17	.85	.73	.97	−2.33	.020	81.84	<.001	80.45
Only studies in which placements decreased[f]	8	.93	.80	1.08	−.93	.353	47.90	<.001	85.39
Excluding studies with serious risk of bias[g]	15	.90	.79	1.02	−1.68	.093	50.11	<.001	72.06
Violent recidivism[h]	12	.70	.49	1.00	−1.96	.050	29.50	.002	62.71
Violations[i]	11	1.03	.82	1.28	.23	.815	18.39	.049	45.63

Note. k = number of effect sizes that were aggregated; ORw = weighted odds ratio; CI = confidence interval. See the online supplementary materials for forest plots.
[a] Overall placements: Bonta and Motiuk (1990); Feyerherm (2000); Fratello et al. (2011); Goldkamp and Gottfredson (1985); Guy et al. (2015); Maloney and Miller (2015); Schwartz et al. (1991); Stevenson (2018); Toborg et al. (1984); VanNostrand (2016); van Wingerden et al. (2014); Vincent et al. (2016—six sites), Vincent & Perrault (2018—four sites). [b] Overall—excluding studies with serious bias: Bonta and Motiuk (1990); Goldkamp and Gottfredson (1985); Guy et al. (2015); Stevenson (2018); Toborg et al. (1984); van Wingerden et al. (2014); Vincent et al. (2016—six sites), Vincent & Perrault (2018—four sites). [c] Pretrial placements: Feyerherm (2000); Fratello et al. (2011); Goldkamp and Gottfredson (1985); Maloney & Miller (2015); Schwartz et al. (1991); Stevenson (2018); Toborg et al. (1984); VanNostrand (2016). [d] Placements following sentencing/adjudication: Guy et al. (2015); van Wingerden et al. (2014); Vincent et al. (2016—six sites), Vincent & Perrault (2018—four sites). [e] Any recidivism: Bonta & Motiuk (1987); Fratello et al. (2011); Goldkamp and Gottfredson (1985); Guy et al. (2015); Stevenson (2018); Toborg et al. (1984—felony and misdemeanor cases), Vincent et al. (2016—six sites), Vincent & Perrault (2018—four). [f] Any—only studies in which placements decreased: Bonta & Motiuk (1987); Fratello et al. (2011); Guy et al. (2015); Stevenson (2018); Toborg et al. (1984); Vincent et al. (2016—Site 1 and Site 3). [g] Any—excluding studies with serious bias: Goldkamp and Gottfredson (1985); Guy et al. (2015); Stevenson (2018); Toborg et al. (1984); Vincent et al. (2016—six sites), Vincent & Perrault (2018—four sites). [h] Violent recidivism: Guy et al. (2015); Stevenson (2018); Vincent et al. (2016—six sites), Vincent & Perrault (2018—four sites). [i] Violations: Goldkamp and Gottfredson (1985); Guy et al. (2015); Stevenson (2018); Vincent et al. (2016—four sites), Vincent & Perrault (2018—four sites). For Vincent et al. (2016) and Vincent and Perrault (2018), we examined detention/commitment/placement dispositions.

on youth or adults who were high risk was mixed. In two studies, placements for high-risk defendants increased when tools were used (Fratello et al., 2011; Stevenson, 2018). In one study, it did not change (Bonta & Motiuk, 1990), and in two studies placements decreased slightly. For example, Barnes-Ceeney (2013) found that when high-risk youth were assessed with the SAVRY, it reduced the likelihood that they would max out their sentence, possibly because service providers adopted a more proactive approach in reducing risk (see also van Wingerden et al., 2014).

Evaluator adherence. Even when tools were implemented, some professionals did not routinely use them. For instance, in a multisite study, Vincent et al. (2016) found that, at one site, only 42% of eligible youth were assessed with a risk assessment tool, whereas completion rates at other sites were as high as 100%. Clearly, the adoption of tools is unlikely to reduce placements if professionals are not using tools as mandated or, in other words, when implementation quality is poor. Consistent with this, Vincent et al. (2016) found that sites with high completion rates were more likely to find reductions in placements than those with fair or poor completion rates.

Legal decision-makers' consideration of tools. In several studies, researchers noted that the impact of risk tools on placements depended heavily on how much legal decision-makers bought in to tools. Stevenson (2018) found that although tools initially resulted in a 4% increase in release rates, this impact eroded over time as judges returned to their earlier practices (see also Goldkamp & Vîlcicã, 2009). Furthermore, in several studies, researchers noted that legal decision-makers tended to be more conservative and restrictive than tools (i.e., Goldkamp & Gottfredson, 1985; Puzzanchera, Knoll, Adams, & Sickmund, 2012; Virginia Sentencing Commission, 2012; cf. Simpson, 2010). For instance, in one study, judges agreed with the tool most of the time (Goldkamp & Gottfredson, 1985). However, when judges departed from the tool, they tended to suggest more restrictive rather than less restrictive pretrial release decisions.

Tools. Although researchers hypothesize that different tools may differentially affect placement rates, the included studies did not provide much relevant data. Given that static tools were used in different contexts than dynamic tools (i.e., pretrial detention vs. postconviction sentencing), it was not possible to meaningfully compare how these types of tools impacted placement rates. Although two studies examined whether changing from one tool to another tool affected placement rates, those studies did not compare static versus dynamic tools either. Specifically, Guy, Vincent, Grisso, and Perrault (2015) found that switching from a homegrown dynamic tool to another dynamic tool, the SAVRY, did not alter rates of out-of-home placements. Similarly, Stevenson (2018) found that switching from the Kentucky Pretrial Risk Assessment Instrument to another brief static tool, the Public Safety Assessment, did not alter placement rates.

Preexisting rates of placements. In some studies, researchers found that tools were more likely to reduce placements if sites had high preexisting placements prior to adopting a tool, than if sites already had low placement rates. Specifically, Vincent et al. (2016) found that, after adopting a risk tool and adhering to relevant policies, placement rates decreased in sites that initially had high placement rates (46–47% to 31–33%). In contrast, placement rates increased at one site that initially placed very few youth (from 8% to 21%). However, even after this increase, this site still fell below the national average rate for placements. Subsequent studies in different states found the same trend (Guy et al., 2015; Vincent & Perrault, 2018).

Political climate. In two studies, researchers noted that political climate affected the impact of tools. For instance, following a highly publicized case in Florida in which an adolescent allegedly murdered a British tourist (Orlando, 1999; see also Bishop & Griset, 2001), the courts broadened criteria for detention, and apparently adjusted the criteria on their risk assessment instrument. As Bishop and Griset (2001, p. 27) wrote:

> [I]ronically, the RAI [risk assessment instrument], whose initial development had earlier advanced the cause of detention reform, now stood as an obstacle to reducing the detention population admitted through intake. Its screening criteria were broad, and it was not a scientifically valid prediction instrument.

Researchers in Philadelphia found a similar pattern of results (Goldkamp & Vîlcicã, 2009; see also Goldkamp & Gottfredson, 1985). Although the tool initially led to increases in pretrial release of low risk defendants, as the political climate changed, the rate of overrides became very high, and as a result, detention increased.

Question 4: When Tools Are Adopted, Do Rates of Recidivism or Violations Change?

Ten of the studies in this review (45.5%) examined how the adoption of tools impacted rates of any recidivism, violent recidivism, and/or violations (e.g., failures to appear, technical violations such as curfew breaches or failed drug tests). In most cases, researchers measured recidivism by examining arrest rates (60.0%, $k = 6$; see Table 4). However, in the remaining studies they examined petitions or reincarceration. Two studies used fixed follow-up periods of 60 or 90 days, and three studies used variable follow-up periods of approximately 12 to 18 months. The remaining five studies (50.0%) did not report follow-up lengths.

For sites in which restrictive placements decreased, the adoption of tools did not lead to increases in recidivism or violations (see Table 4). However, the adoption of tools did not consistently predict reductions in recidivism or violations either. According to our ratings of the full set of studies, only 20.0% of studies found reductions in any recidivism, 25.0% found reductions in violations, and 40.0% found reductions in violent recidivism (see Table 2). When we meta-analyzed studies that included the necessary information, the adoption of risk tools was associated with small but significant reductions in any recidivism, but there were no significant changes in violent recidivism or violations ($ps = .050$ and .815, respectively; see Table 3).

As a next step, we examined whether these results remained the same after removing studies that were potentially biased. Given that none of the studies on violent recidivism or violations had a serious risk of bias, it was not necessary to remove studies and reanalyze results for those outcomes. However, of the studies that examined any recidivism, four studies were rated as having a serious risk of bias (40.0%) on the ROBINS-I. In those studies, the authors failed to measure offending appropriately or to control for differences in the follow-up length between the tool and no-tool groups (by using a fixed follow-up period or survival analyses). As an example, although the Laura and John Arnold Foundation (2014) originally reported reductions in reoffending immediately

Table 4
When Tools Are Adopted, Do Rates of Recidivism or Failure to Appear Change?

Authors, year (country)	Sample (gender)	Risk tool (assessors)	Recidivism (follow-up length)	Design	Results	Summary		
						Any recidivism	Violent recidivism	Violations
Bonta & Motiuk, 1987 (Canada)	378 adults (male)	LSI (N.R.)	Reincarceration (N.R)	Pre–post	Reincarceration did not differ for tool vs. no-tool groups (14% vs. 8%, p = ns)	No change	—	—
Cooprider, 2009 (IL, USA)	Adults (n is N.R.)	Lake County Pretrial RAI (pretrial officer)	Arrests (N.R.)	Pre–post	FTA decreased (17% to 10%); violations decreased (32% to 28%); arrests were similar (4% to 8%)	No change	—	Decrease
Fratello et al., 2011 (NY, USA)	5,173 adolescents (M/F)	New York City RAI (POs)	Arrests (N.R.)	Pre–post—alternatives to detention	Rearrests while case was pending significantly decreased (26% to 18%)	Decrease	—	—
Goldkamp & Gottfredson, 1985 (PA, USA)	1,800 adults (M/F)	Philadelphia Bail Guidelines (detention staff)	Arrests (90 days)	RCT	Rearrests were similar for tool vs. no-tool groups (10% vs. 11%), as were FTAs (13% vs. 12%)	**No change**	—	**No change**
Guy et al., 2015 (MS, USA)	110 adolescents (M/F)	SAVRY (service counsellors)	Petitions (M = 344 days)	Comparison, matched	Any new petitions did not differ for tool vs. no tool groups (38% vs. 50%) nor did violations (13% vs. 17%); violent petitions were lower (2% vs. 22%)	**No change**	**Decrease**	**No change**
Stevenson, 2018 (KY, USA)	1,030,732 adults (M/F)	Kentucky tool and PSA (pretrial staff)	Arrests (60 days)	Pre–post	Violent rearrests were similar for tool vs. no-tool groups (~.5%—.6%) as were any pretrial arrests (8% vs. 7.3%); FTA was higher for tool group 10% vs. 8%	**No change**	**No change**	**Increase**
Toborg et al., 1984 (DC, USA)	34,291 adults (M/F)	DC Pretrial Services Risk Assessment	Arrests (N.R.)	Pre–post	Pretrial arrests did not change for tool vs. no-tool groups (20.7% vs. 19.4%), nor did FTAs (~16%)	**No change**	—	**No change**
VanNostrand, 2016 (OH, USA)	48,807 adults (M/F)	PSA (N.R.)	Pretrial arrests (N.R.)	Pre–post	FTA decreased (41% to 29%), as did any recidivism (20% to 10%) and violent arrests (5% to 3%)	Decrease	Decrease	Decrease
Vincent et al., 2016 (USA)—six sites	1,694 adolescents (M/F)	SAVRY, YLS/CMI (YPOs)	Petitions (M = 18 months)	Pre–post, matched	Any and violent petitions didn't change at 5 of 6 sites but decreased at one site; violation petitions did not change	**No change at 5/6 sites**	**No change at 5/6 sites**	**No change**
Vincent & Perrault, 2018 (AR, USA)—four sites	754 adolescents (M/F)	SAVRY (YPOs)	Petitions (Mdn = 11–13.5 months)	Pre–post (matched)	Any petitions decreased at 1 of 4 sites, violent petitions decreased at 1 site, and violations did not change at any site	**No change at 3/4 sites**	**No change at 3/4 sites**	**No change**

Note. AR = Arkansas; DC = District of Columbia; IL = Illinois; KY = Kentucky; LSI = Level of Service Inventory; M = mean; M/F = male/female; MS = Mississippi; N.R. = not reported; ns = nonsignificant; NY = New York; OH = Ohio; PA = Pennsylvania; PO = probation officer; PSA = Public Safety Assessment; RAI = risk assessment instrument; SAVRY = Structured Assessment of Violence Risk in Youth; USA = United States of America; YLS/CMI = Youth Level of Service/Case Management Inventory; YPO = youth probation officer. Studies with low or moderate overall risk of bias are bolded to indicate that more weight should be given to these studies; the remaining studies have serious risk of bias. The summary ratings are defined as follows: Decrease = all or most analyses indicated a decrease in offending and/or violations; Mixed = studies showed an inconsistent pattern of results; No change = all or most analyses indicated that offending and/or violations did not significantly change; Increase = all or most analyses indicated an increase in the offending and/or violations.

following the adoption of the Public Safety Assessment, Stevenson (2018) reanalyzed the data and concluded that this was an artifact caused by delays in case processing.

When we excluded studies with a serious risk of bias, the results were attenuated. The adoption of risk assessment tools was no longer associated with significant reductions in any recidivism ($p = .093$; see Table 3). As such, the strength of evidence that the adoption of risk tools reduces rates of any recidivism was rated as low on the AHRQ. In addition, there was insufficient evidence to conclude that tools reduce violent recidivism or violations, as none of those results reached significance ($ps = .050$ and .815, respectively; see Table 3).

Question 5: How Does the Adoption of Risk Tools Impact Racial and Ethnic Disparities in Restrictive Placements?

Only six of the studies in this review (27.3%) reported findings on how the adoption of tools impacted rates of restrictive placements for defendants from racial and ethnic minority groups. All six of these studies focused on pretrial detention and used brief static tools that focused on offense history. In five studies ($k = 5$; 83.3%), absolute rates of restrictive placements were lower for people of color following the adoption of the tool (see Tables 2 and 5). These decreases ranged from a nonsignificant decrease of 6% (Simpson, 2010) to a sizable decrease of 57% (Feyerherm, 2000).

Even though the use of tools was associated with decreases in absolute rates of restrictive placements, the more important question is whether tools decrease placements more for Whites than for people of color. Such a pattern could indicate an exacerbation of preexisting disparities. In two studies, disparities decreased following the adoption of a tool (see Table 5). For instance, Feyerherm (2000) found that admission rates decreased 57% for African American youth and 41% for White youth following the adoption of the Multnomah County Risk Assessment Instrument (i.e., the interaction between race and the tool was significant). Furthermore, this effect remained even after the authors controlled for other variables in analyses (e.g., offense history). This reduction in the overrepresentation of African American youth may have occurred because the risk tool used in that study was designed to avoid racial bias. For example, the authors described that rather than rating the presence of intact family structure, the tool examined the presence of a responsible adult.

In one study (Maloney & Miller, 2015), the adoption of a risk tool had a similar impact on placement rates for White, African American, and Hispanic youth (i.e., the interaction was nonsignificant). Finally, in two studies, researchers found mixed results or increases in disparities. For instance, a large report concluded that although the JDAI initiative was associated with reduced rates of detention for both youth of color and white youth, these reductions were larger for white youth (Annie E. Casey Foundation, 2017).

All but one of the studies that examined restrictive placements among minority groups were rated as having a Serious risk of bias on the ROBINS-I (see Table 5). For instance, four studies (66.7%) were part of the JDAI initiative. As such, it is difficult to determine whether any observed findings are due to the tool or other JDAI initiatives (e.g., alternatives to detention). The only study that did not have a serious risk of bias was Stevenson (2018), which reported mixed results. Stevenson found that, prior to the implementation of legislation that mandated the use of a risk tool, White defendants were two percentage points more likely than Black defendants to receive nonfinancial release. After this legislation, White defendants were 10 percentage points more likely than Black defendants to receive nonfinancial release. However, based on post hoc analyses, the authors concluded that this increased racial gap could be partially due to regional differences. In addition, the racial gap was halved once factors such as gender, age, and current charge were controlled. As such, given that high quality studies were scarce, and the results were mixed, the evidence on how risk tools impact racial and ethnic disparities was rated as insufficient on the AHQR.

Discussion

To help inform debates about the impact of risk tools on restrictive placements, we conducted a systematic review and meta-analysis. Given that much of the research in this area was in the form of unpublished reports, we systematically searched 13 databases of published and unpublished sources, hand-searched reference lists, and contacted experts. Although our review captured 22 studies with 1,444,499 defendants and offenders from 30 independent sites, many of the studies failed to match tool and no-tool groups on key characteristics (e.g., offense history) or control for historical trends, such as decreases in incarceration rates over time. In addition, in some studies, other initiatives were implemented at the same time as tools (e.g., alternatives to detention programs), making it difficult to determine if the results were due to the tool or these other initiatives. Furthermore, 40.9% of included studies did not contain the necessary statistical information to include in a meta-analysis (despite efforts to obtain such information from study authors).

As such, to provide a more comprehensive synthesis of findings, we conducted both a meta-analysis of the subset of studies that could be empirically synthesized, as well as a narrative review of the full set of studies. We also tested whether results remained the same after removing studies with a serious risk of bias. Overall, the meta-analysis provided a similar pattern of results as the narrative review, providing some confirmation of the findings. However, because results were attenuated after controlling for study limitations, only modest and tentative conclusions can be drawn. Also, given that most of the included studies were conducted in the United States, it is unclear whether the findings generalize to other countries. With these caveats in mind, key findings are discussed.

Key Findings

Although some researchers and policymakers have hypothesized that the adoption of tools might reduce rates of incarceration (e.g., Laura & John Arnold Foundation, 2014), we found tenuous results. When we examined the full set of studies (regardless of their quality), the adoption of risk tools appeared to be associated with small but significant reductions in restrictive placements. Specifically, when tools were used, fewer defendants were placed in detention prior to trial, and more inmates were released from custodial centers. However, results varied between studies, and we did not find significant reductions in postconviction placements. Moreover, when we removed studies with a serious risk of bias, the findings were no longer significant. As such, the overall strength of evidence that tools reduce placements is low.

Table 5
How Does the Adoption of Risk Assessment Tools Impact Racial and Ethnic Disparities in Restrictive Placements?

Authors, year (country)	Sample (gender)	Risk tool (assessors)	Other programs/ initiatives	Design	Results	Summary	
						Absolute rates minority placements	Overrepresentation and disparities
Annie E. Casey Foundation, 2017 (USA)	>284,887 adolescents, 164 sites (M/F)	RAI (user N.R.)	JDAI	Pre–post	Detention decreased 44% for youth of color and 59% for White youth	Decreased	Increased
Feyerherm, 2000 (OR, USA)	18,788 adolescents (M/F)	Multnomah RAI (detention staff)	JDAI	Pre–post	Detention decreased 60% for Asian youth, 57% for Black youth, 41% for Hispanic youth, 55% for Native American youth, 51% for Caucasian youth, and 52% for minorities overall[a]	Decreased	Decreased (in general)
Maloney & Miller, 2015 (NJ, USA)	1,432 adolescents (M/F)	RAI (intake staff)	JDAI	Pre–post (matched)	Detention decreased at similar rates for White, Black, and Hispanic youth (interaction was non-significant)	Decreased	No change
Puzzanchera et al., 2012 (PA, USA)	>2,098 adolescents (M/F)	Allegheny DAI (detention staff or YPO)	JDAI	Pre–post	Detention decreased ~36% for Black youth and ~32% for White youth (2007 vs. 2009)[a]	Decreased	Decreased
Simpson, 2010 (LA, USA)	202 adolescents (gender N.R.)	Rapides Parish DSI (detectives)	—	Pre–post	Detention admission rate decreased 6% for Black youth (non-significant)[a]	No change	—
Stevenson, 2018 (KY, USA)	1,030,732 adults (M/F)	Kentucky tool and PSA (pretrial staff)	—	Pre–post	Larger increase in non-financial pretrial release for Whites than Blacks, widening racial gap from 2% to 10%, but effect reduced once regional differences, etc. accounted for	**Decreased**	**Mixed**

Note. DAI = Detention Assessment Instrument; DSI = detention screening instrument; JDAI = Juvenile Detention Alternatives Initiative; KY = Kentucky; M/F = male/female; LA = Louisiana; NJ = New Jersey; N.R. = not reported; OR = Oregon; PA = Pennsylvania; PSA = Public Safety Assessment; RAI = risk assessment instrument; USA = United States of America; YPO = youth probation officer. Studies with low or moderate overall risk of bias are bolded to indicate that more weight should be given to these studies; the remaining studies have serious risk of bias.
[a] Calculated percentage differences with the following formula: % change = [(new % − old %)/old %] × 100.

There are several possible explanations for these modest findings. First, the impact of tools on placement rates may be attenuated by implementation problems (Stevenson, 2018; Vincent et al., 2016). Even when agencies adopted tools, evaluators did not always complete required risk assessments due to lack of buy-in, and judges did not always place much weight on tools in their decision-making. Second, even when tools are implemented properly, they may not be powerful enough to reduce placements, especially in postconviction sentencing decisions in which judges have many different factors to consider. Instead, if the goal is to reduce placements, tools may need to be accompanied by a larger package of initiatives such alternatives to detention programs. Third, tools might have a limited impact on overall placement rates because, based on the RNR model, tools might decrease restrictive placements for people who present a low risk of recidivism but not those who present a high risk (e.g., van Wingerden et al., 2014). In other words, their impact may depend on the composition and risk level of the sample, as well as existing placement rates (Vincent et al., 2016). Finally, the impact of tools on placements might vary by tool. For instance, some researchers hypothesize that tools with dynamic factors may be more likely to reduce placement rates than static tools (Kopkin et al., 2017). Unfortunately, however, no studies directly compared dynamic and static tools, and as such, it is not possible to offer conclusions at this point.

The results of our systematic review confirmed that recidivism rates did not increase following the adoption of a risk assessment tool even when incarceration rates decreased. Prior research has found that incarceration is not an effective method to reduce recidivism (Nagin, Cullen, & Jonson, 2009). Our findings similarly illustrate that it is possible to reduce incarceration rates without increasing recidivism. However, although recidivism did not *increase*, we did not find clear and consistent evidence that the use of tools led to a significant *decrease* in recidivism. In most studies, rates of any recidivism, violent recidivism, and violations did not significantly change following the adoption of risk tools. In addition, in the meta-analysis, reductions in recidivism were not significant after removing studies with a serious risk of bias. As such, the strength of evidence that tools reduce recidivism is low. A prior systematic review also reported modest and mixed findings on whether the adoption of tools decreases recidivism rates (Viljoen et al., 2018).

In some ways, the lack of consistent reductions in recidivism is not particularly surprising. The aim of brief pretrial risk tools is not to decrease recidivism per se, but rather to decrease unnecessary incarceration of low-risk defendants without increasing recidivism. In addition, recidivism reduction may be difficult to achieve in less than three years from the adoption of risk assessments (Flores, Lowenkamp, Holsinger, & Latessa, 2006), and because most studies in our review examined only short-term recidivism, they may not have captured longer-term changes. Finally, these findings suggest that risk tools are unlikely to have an impact on recidivism if they are not paired with a RNR approach and quality services and programming to reduce an individual's risk (Vincent et al., 2016).

Even if the use of tools in sentencing has certain benefits, one of the major concerns is that they might exacerbate racial and ethnic disparities in placements (e.g., Maurutto & Hannah-Moffat, 2007). Unfortunately, our review found that research is insufficient to offer conclusions. Only six of the 22 studies included in this review reported results on how the adoption of tools impacted disparities, and all but one of these studies had a serious risk of bias. Furthermore, these studies found variable results. In two studies, placements decreased more for White defendants than defendants of color, thereby increasing disparity (Annie E. Casey Foundation, 2017; Stevenson, 2018). Conversely, in two studies, the opposite effect occurred wherein placements decreased more for African Americans than for Whites, thereby decreasing disparity (Feyerherm, 2000; Puzzanchera et al., 2012). Thus, these findings could suggest that the impact of tools on disparity may depend on the tool and context.

Implications for Research

One of the primary conclusions of this systematic review is that we need better research to determine how tools impact placement and recidivism rates, particularly studies that use rigorous designs such as randomized trials, staggered designs, and propensity score matched studies. However, this type of research is challenging to conduct. Many agencies have already implemented risk tools, making it difficult to find appropriate comparison groups. As such, in addition to conducting field studies, researchers could use carefully controlled experimental designs, such as case vignette studies, to examine how tools influence judges' placement decisions when other factors are held constant. In addition, when agencies adopt tools for the first time or switch from one tool to another, researchers can take advantage of these naturally occurring experiments to test how these changes alter placement rates or recidivism.

To ensure that this research is valid and credible, it is critical that researchers carefully attend to possible confounds and biases. Placement rates can be affected by numerous factors, such as whether incarceration rates are already decreasing and whether professionals are adhering to tools. As such, researchers should measure implementation level outcomes (e.g., fidelity to tools), and take steps to address potential biases in their design and analyses. For instance, to accurately test how tools impact recidivism rates, researchers should control for the length of time at risk for recidivism and time spent incarcerated.

Given that many advocates and critics of risk assessment have strong opinions about the impact of risk assessment tools, researchers should take steps to ensure that their own views do not jeopardize their objectivity. Rather than adopting a mindset that their job is to promote the value of tools, researchers should carefully test both potential benefits of tools as well as unintended effects, such as the possibility that tools may exacerbate racial and ethnic disparities. In addition, rather than making overly simplistic generalizations, such as concluding that tools are either good or bad, researchers should test more nuanced questions such as, Do certain tools exacerbate disparities in confinement rates, and, if so, which tools and under what circumstances? Are tools more or less likely to create disparities than the alternative approach, namely intuitive judgments about risk? To deter the possibility of selective reporting, namely, the tendency to report findings that confirm researchers' own hypotheses, researchers should ensure that their data analytic choices are transparent and determined prior to initiating the study.

Implications for Policy and Practice

Although we found that tools might help reduce restrictive placements in some cases, our results highlight that agencies should not develop unrealistic expectations that tools are a panacea. In and of themselves, tools likely have only a modest impact on placement rates and recidivism. To have a strong and sustainable impact, tools need to be implemented well with adequate staff and stakeholder buy-in, appropriate policies, and routine quality assurance practices (Bonta, Bourgon, Rugge, Gress, & Gutierrez, 2013; Vincent et al., 2016). For instance, agencies should provide judges, probation officers, and other users with training on the RNR model and on how to use risk assessments in placement decisions.

Prior to adopting a tool, agencies should pilot test the tool, and then continue to periodically reevaluate its use (Vincent et al., 2012). This reevaluation is important because agencies can experience a combination of both "moving forward and slipping back" (Bazemore, 1993, p. 41). According to some authors, without ongoing reevaluation, risk tools might potentially even "become a straitjacket that binds the juvenile justice system to inappropriate use of detention" (Bishop & Griset, 2001, p. 42). As we found through this review, some agencies are already making efforts to evaluate the impact of tools on placement decisions, which is commendable. However, much of this work consisted of brief unpublished reports that did not control for possible confounds. As such, agencies should work toward increasing the rigor of their research such as by pairing with academic researchers. Agencies should also take steps to disseminate their findings, including both positive and negative results. This willingness to identify and learn from challenges captures the spirit of evidence-based practice; evidence-based practice is not a one-shot implementation of a tool but instead, a commitment to ongoing review and refinement (Stevenson, 2018).

In sum, our review indicates that although risk assessment tools are not a remedy to overincarceration, they might potentially help to reduce restrictive placements without increasing recidivism. In this respect, tools may help balance public safety and offenders' liberty while presumably decreasing costs to the system. However, research is scarce, and many studies are poor in quality. Furthermore, it remains to be seen whether any potential benefits of tools come at a cost to social justice, and if so, under what circumstances. As such, researchers and policymakers need to invest greater efforts into rigorously investigating these important questions.

References

*References marked with an asterisk indicate studies included in the systematic review.

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., . . . Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34,* 341–382. http://dx.doi.org/10.1177/0011000005285875

American Bar Association. (2007). *ABA standards for criminal justice: Pretrial release* (3rd ed.). Chicago, IL: Author. Retrieved from https://www.americanbar.org/content/dam/aba/publications/criminal_justice_standards/pretrial_release.authcheckdam.pdf

American Law Institute. (2014). *Model penal code (Tentative Draft No. 3)*. Philadelphia, PA: Author.

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

*Annie E. Casey Foundation. (2017). *Juvenile detention alternatives (JDAI at 25): Insights from the annual results report*. Baltimore, MD. Retrieved July 2, 2019, from http://www.aecf.org/m/resourcedoc/aecf-jdaiat25-2017.pdf

Austin, J. (2004). The proper and improper use of risk assessment in corrections. *Federal Sentencing Reporter, 16,* 194–199. http://dx.doi.org/10.1525/fsr.2004.16.3.194

Austin, J., Ocker, R., & Bhati, A. (2010). *Kentucky Pretrial Risk Assessment Instrument validation*. Retrieved from http://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=60b06cf8-f956-d6f1-d07f-a426f0465846

Baradaran, S., & McIntyre, F. L. (2012). Predicting violence. *Texas Law Review, 90,* 497–553.

*Barnes-Ceeney, K. (2013). *Negotiating labyrinths of risk: The impact and implementation of a structured violence risk assessment instrument in juvenile parole* (Unpublished doctoral dissertation). City University of New York, New York, NY.

*Bazemore, G. (1993). Formal policy and informal process in the implementation of juvenile justice reform. *Criminal Justice Review, 18,* 26–45. http://dx.doi.org/10.1177/073401689301800104

Bentham, J. (2000). An introduction to the principles of morals and legislation. In J. H. Burns (Ed.), *The collected works of Jeremy Bentham (electronic version)*. Charlottesville, VA: InteLex Corp. (Original work published 1789)

Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology, 13,* 193–216. http://dx.doi.org/10.1007/s11292-017-9286-2

Berkman, N. D., Lohr, K. N., Ansari, M. T., Balk, E. M., Kane, R., McDonagh, M., . . . Chang, S. (2015). Grading the strength of a body of evidence when assessing health care interventions: An EPC update. *Journal of Clinical Epidemiology, 68,* 1312–1324. http://dx.doi.org/10.1016/j.jclinepi.2014.11.023

Bishop, D. M., & Griset, P. L. (2001). *Replicating detention reform: Lessons from the Florida Initiative*. Baltimore, MD: Annie E. Casey Foundation.

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). New York, NY: Routledge.

Bonta, J., Bourgon, G., Rugge, T., Gress, C., & Gutierrez, L. (2013). Taking the leap: From pilot project to wide-scale implementation of the Strategic Training Initiative in Community Supervision (STICS). *Justice Research and Policy, 15,* 17–35. http://dx.doi.org/10.3818/JRP.15.1.2013.17

*Bonta, J., & Motiuk, L. L. (1987). The diversion of incarcerated offenders to correctional halfway houses. *Journal of Research in Crime and Delinquency, 24,* 302–323. http://dx.doi.org/10.1177/0022427887024004006

*Bonta, J., & Motiuk, L. L. (1990). Classification to half-way houses: Quasi-experimental evaluation. *Criminology, 28,* 497–506. http://dx.doi.org/10.1111/j.1745-9125.1990.tb01336.x

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis version 2*. Engelwood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1,* 97–111. http://dx.doi.org/10.1002/jrsm.12

Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.

Bridges, G. S., & Steen, S. (1998). Racial disparities in official assessments of juvenile offenders: Attributional stereotypes as mediating mechanisms. *American Sociological Review, 63,* 554–570. http://dx.doi.org/10.2307/2657267

Cadigan, T. P., & Lowenkamp, C. T. (2011). Implementing risk assessment in the federal pretrial services system. *Federal Probation, 75,* 30–34.

Carson, E. A. (2018). *Prisoners in 2016*. Washington, DC: U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Casey, P. M., Warren, R. K., & Elek, J. K. (2011). *Using offender risk and needs assessment information at sentencing: Guidance for courts from a National Working Group*. Williamsburg, VA: National Center for State Courts.

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics Simulation and Computation, 39,* 860–864. http://dx.doi.org/10.1080/03610911003650383

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19,* 3127–3131. http://dx.doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290. http://dx.doi.org/10.1037/1040-3590.6.4.284

Clear, T. R., & Frost, N. A. (2014). *The punishment imperative: The risk and failure of mass incarceration in America*. New York, NY: New York University Press.

*Cooprider, K. (2009). Pretrial risk assessment and case classification: A case study. *Federal Probation, 73,* 12–15.

Dedel, K., & Davies, G. (2007). *Validating Multnomah County's Juvenile Detention Risk Assessment Instrument*. Portland, OR: Multnomah County Department of Community Justice. Retrieved from https://multco.us/file/29611/download

Eaglin, J., & Solomon, D. (2015). *Reducing racial and ethnic disparities in jails: Recommendations for local practice*. New York, NY: Brennan Center for Justice at New York University School of Law.

Elek, J. K., Warren, R. K., & Casey, P. M. (2015). *Using risk and needs assessment information at sentencing: Observations from ten jurisdictions*. Williamsburg, VA: National Center for State Courts.

*Feyerherm, W. H. (2000). Detention reform and over-representation: A successful synergy. *Corrections Management Quarterly, 4,* 44–51.

Flores, A. W., Lowenkamp, C. T., Holsinger, A. M., & Latessa, E. J. (2006). Predicting outcome with the Level of Service Inventory-Revised: The importance of implementation integrity. *Journal of Criminal Justice, 34,* 523–529. http://dx.doi.org/10.1016/j.jcrimjus.2006.09.007

Fratello, J., Salsich, A., & Modulescu, S. (2011). *Juvenile detention reform in New York City: Measuring risk through research*. New York, NY: Vera Institute of Justice.

Goldkamp, J. S. (1979). *Two classes of accused: A study of bail and detention of American justice*. Cambridge, MA: Ballinger.

Goldkamp, J. S., & Gottfredson, M. R. (1985). *Policy guidelines for bail: An experiment in court reform*. Philadelphia, PA: Temple University Press.

Goldkamp, J. S., & Vîlcicã, E. R. (2009). Judicial discretion and the unfinished agenda of American bail reform: Lessons from Philadelphia's evidence-based judicial strategy. In A. Sarat (Ed.), *New perspectives on crime and criminal justice* (pp. 115–157). Bingley, UK: Emerald Group Publishing Limited. http://dx.doi.org/10.1108/S1059-4337(2009)0000047007

Gough, D. (2015). Qualitative and mixed methods in systematic reviews. *Systematic Reviews, 4,* 181. http://dx.doi.org/10.1186/s13643-015-0151-y

Graham, S., & Lowery, B. S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior, 28,* 483–504. http://dx.doi.org/10.1023/B:LAHU.0000046430.65485.1f

*Guy, L. S., Vincent, G. M., Grisso, T., & Perrault, R. (2015). *Advancing use of risk assessment in juvenile probation: Technical report submitted to the Office of Juvenile Justice and Delinquency Prevention*. Worcester: University of Massachusetts Medical School. Retrieved from https://www.ncjrs.gov/pdffiles1/ojjdp/grants/249155.pdf

Haddaway, N. R., Woodcock, P., Macura, B., & Collins, A. (2015). Making literature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology, 29,* 1596–1605. http://dx.doi.org/10.1111/cobi.12541

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21,* 1–21. http://dx.doi.org/10.1037/a0014421

Harcourt, B. E. (2007). *Against prediction: Profiling, policing, and punishing in an actuarial age*. Chicago, IL: University of Chicago Press.

Harcourt, B. E. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter, 27,* 237–243. http://dx.doi.org/10.1525/fsr.2015.27.4.237

Hart, S. D. (2011). Complexity, uncertainty, and the reconceptualization of violence risk assessment. In R. Abrunhosa (Ed.), *Victims and offenders: Chapters on psychology and law* (pp. 57–59). Brussels, Belgium: Politeia.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. http://dx.doi.org/10.1037/1082-989X.3.4.486

Hildebrand, M., Hol, A. M., & Bosker, J. (2013). Predicting probation supervision violations. *Psychology, Public Policy, and Law, 19,* 114–125. http://dx.doi.org/10.1037/a0028179

Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29,* 380–396. http://dx.doi.org/10.1177/0093854802029004003

Hoge, R. D., & Andrews, D. A. (2002). *The Youth Level of Service/Case Management Inventory*. Toronto, Canada: Multi-Health Systems.

Holder, E. (2014, August 1). Attorney General Eric Holder speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference. *The United States Department of Justice*. Retrieved from https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods, 11,* 193–206. http://dx.doi.org/10.1037/1082-989X.11.2.193

Kleiman, M., Ostrom, B. J., & Cheesman, F. L., II. (2007). Using risk assessment to inform sentencing decisions for nonviolent offenders in Virginia. *Crime & Delinquency, 53,* 106–132. http://dx.doi.org/10.1177/0011128706294442

Kopkin, M. R., Brodsky, S. L., & DeMatteo, D. (2017). Risk assessment in sentencing decisions: A remedy to mass incarceration? *Journal of Aggression, Conflict and Peace Research, 9,* 155–164. http://dx.doi.org/10.1108/JACPR-06-2016-0232

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. http://dx.doi.org/10.2307/2529310

Laura and John Arnold Foundation. (2014). *Results from the First Six Months of the Public Safety Assessment—Court™ in Kentucky*. Houston, TX: Author. Retrieved from http://www.ncsl.org/documents/cj/PSACourtKentucky6MonthReport.pdf

Laura and John Arnold Foundation. (n.d.). *Public safety assessment: Risk factors and formula*. Houston, TX: Author.

La Vigne, N., Bieler, S., Cramer, L., Ho, H., Kotonias, C., Mayer, D., . . . Samuels, J. (2014). *Justice Reinvestment Initiative state assessment report*. Washington, DC: Urban Institute. Retrieved from https://www.urban.org/sites/default/files/publication/22211/412994-Justice-Reinvestment-Initiative-State-Assessment-Report.PDF

*Maloney, C., & Miller, J. (2015). The impact of a risk assessment instrument on juvenile detention decision-making: A check on "perceptual shorthand" and "going rates"? *Justice Quarterly, 32,* 900–927. http://dx.doi.org/10.1080/07418825.2013.863961

Maurutto, P., & Hannah-Moffat, K. (2007). Understanding risk in the context of the Youth Criminal Justice Act. *Canadian Journal of Criminology and Criminal Justice, 49,* 465–491. http://dx.doi.org/10.3138/cjccj.49.4.465

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30–46. http://dx.doi.org/10.1037/1082-989X.1.1.30

McKay, P., Hollist, D., Coolidge, J., Delano, W., Greenwood, I., King, M., . . . Doyle, D. (2014). *The Montana Pre-Adjudicatory Detention Risk Assessment Instrument*. Missoula: University of Montana, Missoula. Retrieved from https://mbcc.mt.gov/Portals/130/Data/Reports/RAI_DRAI_Comparison.pdf

Mendel, R. (2014). *Juvenile Detention Alternatives Initiative Progress Report*. Baltimore, MD. Retrieved from http://www.aecf.org/m/resourcedoc/aecf-2014JDAIProgressReport-2014.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6,* e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097

Monahan, J., & Skeem, J. L. (2016). Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology, 12,* 489–513. http://dx.doi.org/10.1146/annurev-clinpsy-021815-092945

Monahan, J., Skeem, J., & Lowenkamp, C. (2017). Age, risk assessment, and sanctioning: Overestimating the old, underestimating the young. *Law and Human Behavior, 41,* 191–201. http://dx.doi.org/10.1037/lhb0000233

Morris, N. (1974). *The future of imprisonment*. Chicago, IL: University of Chicago Press.

Myburgh. J.-E., Camman, C., & Wormith, J. S. (2015, November). *Review of pretrial risk assessment and factors predicting pretrial release failure*. Saskatchewan, Canada: Centre for Forensic Behavioral Science and Justice Studies, University of Saskatchewan.

Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. *Crime and Justice, 38,* 115–200. http://dx.doi.org/10.1086/599202

National Association of Pretrial Services Agencies. (2004). *Standards on pretrial release* (3rd ed.). Washington, DC: National Association of Pretrial Services Agencies. Retrieved from https://info.nicic.gov/nicrp/system/files/napsa-standards-2004.pdf

National Research Council of the National Academy of Sciences. (2013). *Reforming juvenile justice: A developmental approach*. Washington, DC: The National Academies Press.

Olver, M. E., Stockdale, K. C., & Wong, S. C. P. (2012). Short and long-term prediction of recidivism using the youth level of service/case management inventory in a sample of serious young offenders. *Law and Human Behavior, 36,* 331–344. http://dx.doi.org/10.1037/h0093927

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36,* 329–353. http://dx.doi.org/10.1177/0093854809331457

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the level of service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26,* 156–176. http://dx.doi.org/10.1037/a0035080

*Orlando, F. (1999). *Controlling the front gates: Effective admissions, policies, and practices. Pathways to juvenile detention reform* (Vol. 3). Baltimore, MD: Annie E. Casey Foundation. Retrieved from http://www.aecf.org/m/resourcedoc/AECF-ControllingtheFrontGates-1999.pdf

Ostrom, R. J., Kleiman, M., Cheesman, I. I. F., Hansen, R. M., & Kauder, N. B. (2002). *Offender risk assessment in Virginia: A three-stage evaluation*. Williamsburg, VA: National Center for State Courts and the Virginia Criminal Sentencing Commission.

Perrault, R. T., Vincent, G. M., & Guy, L. S. (2017). Are risk assessments racially biased?: Field study of the SAVRY and YLS/CMI in probation. *Psychological Assessment, 29,* 664–678. http://dx.doi.org/10.1037/pas0000445

Petersilia, J., & Turner, S. (1987). *Guideline-based justice: The implications for racial minorities*. Los Angeles, CA: The RAND Corporation. Retrieved from http://www.rand.org/pubs/reports/R3306.html

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., . . . Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews*. London, UK: Institute for Health Research. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf

Pretrial Justice Institute. (2009). *Survey of pretrial services programs*. Retrieved from https://university.pretrial.org/viewdocument/survey-of-pretrial-s

*Puzzanchera, C., Knoll, C., Adams, B., & Sickmund, M. (2012). *Allegheny County detention screening study*. Pitsburg, VA: National Center for Juvenile Justice. Retrieved from http://www.ncjj.org/pdf/MFC/FINAL_Allegheny%20Detention%20Assessment.pdf

*Schwartz, I. M., Barton, W., & Orlando, F. (1991). Keeping kids out of secure detention: The misuse of juvenile detention has a profound impact on child welfare. *Public Welfare, 49,* 20–26.

Schweitzer Smith, M. (2017). *Reinventing juvenile justice: Examining the effectiveness of the targeted reclaim initiative* (Unpublished doctoral dissertation). University of Cincinnati, Cincinnati, OH.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., . . . Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal*. Advance online publication. http://dx.doi.org/10.1136/bmj.j4008

Shepherd, S. M., Luebbers, S., Ferguson, M., Ogloff, J. R. P., & Dolan, M. (2014). The utility of the SAVRY across ethnicity in Australian young offenders. *Psychology, Public Policy, and Law, 20,* 31–45. http://dx.doi.org/10.1037/a0033972

*Simpson, T. P. (2010). *Do objective measures reduce the disproportionate rates of minority youth placed in detention: Validation of a risk assessment instrument?* (Unpublished doctoral dissertation). University of New Orleans, New Orleans, LA.

Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology: An Interdisciplinary Journal, 54,* 680–712. http://dx.doi.org/10.1111/1745-9125.12123

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science, 20,* 38–42. http://dx.doi.org/10.1177/0963721410397271

Starr, S. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review, 66,* 803–872.

Steffensmeier, D., Ulmer, J., & Kramer, J. (1998). The interaction of race, gender, and age in criminal sentencing: The punishment cost of being young, black, and male. *Criminology, 36,* 763–798. http://dx.doi.org/10.1111/j.1745-9125.1998.tb01265.x

Steinhart, D. (2006). *Juvenile detention risk assessment*. Baltimore, MD: Annie E. Casey Foundation.

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., . . . Higgins, J. P. T. (2016). ROBINS-I: A tool for

assessing risk of bias in non-randomised studies of interventions. *British Medical Journal, 355,* 1–7. http://dx.doi.org/10.1136/bmj.i4919

Sterne, J. A. C., Higgins, J. P. T., Elbers, R. G., & Reeves, B. C. (2016). *Risk of bias in non-randomized studies of interventions (ROBINS-I): Detailed guidance*. Retrieved from https://www.bristol.ac.uk/media-library/sites/social-community-medicine/images/centres/cresyda/ROBINS-I_detailed_guidance.pdf

*Stevenson, M. (2018). Assessing risk assessment in action. *Minnesota Law Review, 103,* 303–384.

Thompson, C. (2017). *Myths and factors: Using risk and need assessments to enhance outcomes and reduce disparities in the criminal justice system*. Washington, DC: National Institute of Corrections, Community Corrections Collaborative Network. Retrieved from https://info.nicic.gov/nicrp/system/files/032859.pdf

*Toborg, M. A., Yezer, A. M., Tseng, P., & Carpenter, B. L. (1984, June). *Pretrial release assessment of danger and flight: Method makes a difference*. McLean, VA: Lazer Management Group.

Tonry, M. (1987). Prediction and classification: Legal and ethical issues. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making* (pp. 367–413). Chicago, IL: University of Chicago Press. http://dx.doi.org/10.1086/449140

Tonry, M. (2017). Making American sentencing just, humane, and effective. *Crime and Justice, 46,* 441–504. http://dx.doi.org/10.1086/688456

Turner, S., Braithwaite, H., Kearney, L., Murphy, A., & Haerle, D. (2012). Evaluation of the California Parole Violation Decision-Making Instrument (PVDMI). *Journal of Criminal Justice, 35,* 269–295.

VanNostrand, M. (2016, August). *Assessing the impact of the public safety assessment: Public safety, court appearance, and jail population, Lucas County, OH*. Tampa, FL: Luminosity, Inc.

*van Wingerden, S., van Wilsem, J., & Moerings, M. (2014). Pre-sentence reports and punishment: A quasi-experiment assessing the effects of risk-based pre-sentence reports on sentencing. *European Journal of Criminology, 11,* 723–744. http://dx.doi.org/10.1177/1477370814525937

Vigorita, M. S. (2003). Judicial risk assessment: The impact of risk, stakes, and jurisdiction. *Criminal Justice Policy Review, 14,* 361–376. http://dx.doi.org/10.1177/0887403403253722

Viljoen, J. L., Cochrane, D. M., & Jonnson, M. R. (2018). Do risk assessment tools help manage and reduce risk of violence and reoffending? A systematic review. *Law and Human Behavior, 42,* 181–214. http://dx.doi.org/10.1037/lhb0000280

Vincent, G. M., Chapman, J., & Cook, N. E. (2011). Risk-needs assessment in juvenile justice: Predictive validity of the SAVRY, racial differences, and the contribution of needs factors. *Criminal Justice and Behavior, 38,* 42–62. http://dx.doi.org/10.1177/0093854810386000

Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*. Chicago, IL: MacArthur Foundation, Models for Change. Retrieved from http://njjn.org/uploads/digital-library/Risk_Assessment_in_Juvenile_Justice_A_Guidebook_for_Implementation.pdf

*Vincent, G. M., Guy, L. S., Perrault, R. T., & Gershenson, B. (2016). Risk assessment matters, but only when implemented well: A multisite study in juvenile probation. *Law and Human Behavior, 40,* 683–696. http://dx.doi.org/10.1037/lhb0000214

Vincent, G. M., & Perrault, R. (2018). *Risk Assessment and Behavioral Health Screening Project: Technical Report Submitted to the Office of Juvenile Justice and Delinquency Prevention*. Worcester: University of Massachusetts Medical School.

*Virginia Sentencing Commission. (2012). *2012 Annual Report*. Richmond: Virginia Sentencing Commission. Retrieved from http://www.vcsc.virginia.gov/reports.html

Wachter, A. (2015). *Statewide risk assessment in juvenile probation: JJGPS StateScan*. Pittsburgh, PA: National Center for Juvenile Justice.

Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., . . . the ROBIS group. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology, 69,* 225–234. http://dx.doi.org/10.1016/j.jclinepi.2015.06.005

Appendix A

Risk Assessment Tools

Tool name	Population	# items	Focused on static factors	Focused on dynamic factors	Example items	Validity for prediction of offending
Pretrial						
Allegheny DAI (see Puzzanchera, Knoll, Adams, & Sickmund, 2012)	Juveniles	9	✓	—	Most serious alleged offense, prior findings, supervision status	No known validation study.
Cook County RAI (Orlando, 1999)	Juveniles	8	✓	—	Most serious offense, past findings, current case status	No known validation study.
DC Pretrial RAI (Toborg, Yezer, Tseng, & Carpenter, 1984)	Adults	—	Unclear	Unclear	Items could not be obtained	Tool was more accurate in predicting FTA than violence (Toborg et al., 1984).
Kentucky Pretrial RAI (Austin, Ocker, & Bhati, 2010)	Adults	13	✓	—	Current charge, prior violence, drug/alcohol history	Tool predicted FTA and pretrial arrest (Austin et al., 2010).
Lake County Pretrial RAI (Cooprider, 2009)	Adults	—	✓	—	Items could not be obtained; modelled after Virginia Pretrial Risk Assessment Instrument	No known validation study.
Multnomah County RAI (Orlando, 1999)	Juveniles	7	✓	—	Most serious offense, legal status, prior offense	Tool predicted FTA/ new offense but some items were not predictive (Dedel & Davies, 2007).
New York City RAI (Fratello, Salsich, & Modulescu, 2011)	Juveniles	10	✓	—	Open warrant, school attendance, prior arrest	Selected factors that predicted FTA and rearrest (Fratello et al., 2011), but no known independent validation study.
Philadelphia Bail Guidelines (Goldkamp & Gottfredson, 1985)	Adults	8	✓	—	Offense category, recent arrests, FTA, age	Selected factors were validated using an independent sample (Goldkamp, 1979).
PSA (Laura and John Arnold Foundation, n.d.)	Adults	9	✓	—	Age, violent offense, pending charge, prior felony	Tool predicted rearrests (Laura and John Arnold Foundation, 2014).

(*Appendices continue*)

Tool name	Population	# items	Focused on static factors	Focused on dynamic factors	Example items	Validity for prediction of offending
Rapides Parish Juvenile DSI (authors N.R.)	Juveniles	7	✓	—	Most serious current offense, criminal history, FTA	Predicted rearrests at three months but not at six months (Simpson, 2010).
RAIs for JDAI sites	Juveniles	~8–10	✓	—	Vary by jurisdiction but focus on offenses (Steinhart, 2006)	Some studies have found small, significant results (McKay et al., 2014), but studies are rare (Steinhart, 2006).
RAI for NJ (see Maloney & Miller, 2015)	Juveniles	7	✓	—	Number of current charges, prior adjudications, prior FTA	No known validation study.
Postconviction and release						
LSI (Andrews, 1982)	Adults	54	—	✓	Criminal history, family/marital, alcohol/drug	Moderate effect sizes for general and violent recidivism (Olver, Stockdale, & Wormith, 2014).
Nonviolent risk assessment (Worksheet D; Ostrom, Kleiman, Cheesman, Hansen, & Kauder, 2002)	Adults	11	✓	—	Gender, age, marital status, prior offenses	Tool predicted new arrests (Kleiman, Ostrom, & Cheesman, 2007).
RISc	Adults/ juveniles	61	—	✓	Offense history, drug abuse, education/employment	Moderate effects for violations, including reoffending (Hildebrand, Hol, & Bosker, 2013).
SAVRY (Borum, Bartel, & Forth, 2006)	Juveniles	30	—	✓	History of violence, peer delinquency, strong social support	Moderate effect sizes for violent and general recidivism (Olver, Stockdale, & Wormith, 2009).
YLS/CMI (Hoge & Andrews, 2002)	Juveniles	42	—	✓	Prior/current offenses, family/parenting, substance abuse	Moderate effect sizes for general and violent recidivism (Olver et al., 2014).

Note. ✓ = yes; X = no; DAI = detention assessment instrument; DSI = detention screening instrument; FTA = failure to appear; RAI = risk assessment instrument; LSI = Level of Service Inventory; PSA = Public Safety Assessment; RISc = Recidivism Risk Assessment Scales; SAVRY = Structured Assessment of Violence Risk in Youth; YLS/CMI = Youth Level of Service/ Case Management Inventory.

Appendix B

Risk of Bias Ratings

Studies	Risk of bias in ROBINS–I Domains							Overall risk of bias
	Confounding	Selection	Classification of interventions	Deviations from interventions	Missing data	Measurement of outcomes	Selective reporting	
Pretrial								
Annie E. Casey Foundation (2017)	Serious–did not control for confounds (e.g., historical trends)	Moderate–included 164 sites, limited information on pretool sample	Serious–tool and conditions not described so we referred to other publications	Serious–results might be due to alternatives to detention rather than tool	Serious–164/197 sites reported data, pretool data unclear	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Bazemore (1993)	Serious–did not control for confounds (e.g., historical trends)	Low–appeared to include all eligible cases	Low–samples appeared to be mutually exclusive	Serious–results could be due to changes in detention criteria, changes in tool	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Cooprider (2009)	Serious–did not control for confounds (e.g., risk level)	Low–appeared to include all eligible cases	Low–clear point when risk tool was implemented	No information–possible changes in supervision	No information	No information–appeared to use official records	No information–no specified data analytic plan	Serious
Feyerherm (2000)	Moderate–controlled for covariates, but detention was declining even before tool	Low–appeared to include all eligible cases (population data)	Low–clear whether or not assessed with tool, groups are mutually exclusive	Serious–results might be due to alternatives to detention	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Fratello et al. (2011)	Serious–did not control for confounds (e.g., risk level, trends)	Moderate–comparison group is from a brief period, excluded 2007	Low–clear whether assessed with tool	Serious–results might be due to alternatives to detention	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Goldkamp & Gottfredson (1985)	Low–used random assignment, stratified quota sampling	Low–used combination of stratified and consecutive sampling	Low–randomly assigned judges to prevent contamination of groups	Low–no other interventions or initiatives seem to have occurred	Low–minimal missing data	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Low
Maloney & Miller (2015)	Moderate–matched on extensive variables but did not examine historical trends	Low–appeared to include all eligible cases, sample selection clear	Low–group were clearly defined and mutually exclusive	Serious–results might be due to alternatives to detention) rather than tool	No information	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Serious

(*Appendices continue*)

Studies	Risk of bias in ROBINS–I Domains							Overall risk of bias
	Confounding	Selection	Classification of interventions	Deviations from interventions	Missing data	Measurement of outcomes	Selective reporting	
Orlando (1999)	Serious–did not control for confounds (e.g., risk level, historical trends)	No information	No information	Serious–results might be due to alternatives to detention rather than tool	No information	Low–used official records, same approach across conditions	Moderate–did not test minority confinement in all sites	Serious
Puzzanchera et al. (2012)	Serious–did not control for confounds, detention was declining even before tool	Low–all years reported in graph seems to be population level	Low–clear point when tool was implemented	Low–no other interventions or initiatives were reported	No information	Low–used official records, same approach across conditions	Moderate–did not present planned analyses on detention rates	Serious
Schwartz, Barton, and Orlando (1991)	Moderate–had comparison group to address historical trends, did not compare group differences	Low - appeared to include all eligible cases	Moderate–unclear when tool and other interventions started	Serious–results might be due to alternatives to detention rather than tool	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Simpson (2010)	Serious–did not control for confounds (e.g., risk level, historical trends)	Moderate–pretool sample consisted only of detained youth	Serious–many youth in tool condition did not get tool (i.e., 18 of 22)	Moderate–forms were repeatedly revised	Moderate–19% were missing tool, no info for comparison group	Low–used official records, same approach across conditions	Moderate–clear analytic plan but statistics were not always reported	Serious
Stevenson (2018)	Moderate–didn't test group differences but ran residual analyses to rule out changes	Low–appeared to include all eligible cases	Moderate–some people in the pretool group were receiving tools but less often	Low–some legal changes but no other initiatives	No information	Low–used official records, same approach across conditions	Moderate–did not present certain statistical analyses (but provided results on request)	Moderate
Toborg et al. (1984)	Low–ruled out important group differences and judicial trends	Low–included all eligible cases, group time periods were equivalent	Low–groups were clearly defined (clear implementation date)	Low–no other interventions or initiatives seem to have occurred	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Low
VanNostrand (2016)	Serious–did not control for confounds (e.g., risk level, historical trends)	Low–appeared to include all eligible cases	Low–groups were clearly defined (clear implementation date)	Low–no other interventions or initiatives were reported	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious

(*Appendices continue*)

Studies	Risk of bias in ROBINS–I Domains							Overall risk of bias
	Confounding	Selection	Classification of interventions	Deviations from interventions	Missing data	Measurement of outcomes	Selective reporting	
Postconviction								
Guy, Vincent, Grisso, and Perrault (2015)	Low–matched on extensive variables	Low–used full sample of consecutive cases	Low–separate sites	Low–no other interventions or initiatives seem to have occurred	Low–minimal missing data	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses inplan	Low
van Wingerden et al. (2014)	Low–matched on variables and samples were from same time period so cohort effects unlikely	Low–selected participants using official records	Low–used clear records to determine whether tool was conducted pretrial	Low–samples are from same time period, did not appear to be differences in interventions	Moderate–1/3 of data was excluded; no comparison of cases with missing data	Low–used official records, same approach across conditions	Low–analytic plan was clearly specified and had appropriate rationale	Moderate
Vincent, Guy, Perrault, and Gershenson (2016)	Low–matched on extensive variables, historical trends unlikely	Low–used a combination of consecutive and random sampling	Low–pre- and postgroups were mutually exclusive	Low–policy changes linked to tool but no other co-interventions	Low–sample was generated based on complete cases	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Low
Vincent & Perrault (2018)	Low–matched on extensive variables, historical trends unlikely	Low–used a complete sample of cases	Low–pre- and postgroups were clearly defined	Low–policy changes linked to tool but no other co-interventions	Low–sample was generated based on complete cases	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Low
Virginia Sentencing Commission (2012)	Serious–did not control for confounds (e.g., risk level, historical trends)	Low–used a full sample (consecutive) at 6 pilot sites	Serious–some offenders in tool group didn't receive tool	Low–no other interventions or initiatives seem to have occurred	No information	Low–used official records, same approach across conditions	No information–no specified data analytic plan	Serious
Release								
Barnes-Ceeney (2013)	Moderate–tested for group differences but did not examine historical trends	Moderate–inclusion criteria wasn't clear	Low–clear whether assessed with tool	Low–no other interventions or initiatives seem to have occurred	Moderate–missing data on risk factors	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Moderate

(*Appendices continue*)

Studies	Risk of bias in ROBINS–I Domains							Overall risk of bias
	Confounding	Selection	Classification of interventions	Deviations from interventions	Missing data	Measurement of outcomes	Selective reporting	
Bonta & Motiuk (1987)	Serious–groups may have differed because pregroup was collected in summer	Low–used consecutive sampling, sampling periods differed in length	Moderate–LSI scores were available in pretool group but were instructed not to use them	Low–no other interventions or initiatives seem to have occurred	No information	Low–used official records, same approach across conditions	Low–clear data analytic plan, carried out analyses in plan	Serious
Bonta & Motiuk (1990)	Moderate–controlled for historical bias, compared groups but procedures not described	Low–used all inmates in the three jails during the study period	Low–samples were clearly defined (used different jails for groups)	Low–policy changes linked to tool but no other interventions that could explain results	Low–missing data not discussed but based on *n*'s missing data seems unlikely	Low–used official records, same measurement approach across conditions	No information–no specified data analytic plan	Moderate

Note. Overall risk of bias is rated as follows (Sterne, Hernán, et al., 2016, p. 4): low ("the study is comparable to a well performed randomized trial"), moderate ("the study provides sound evidence for a nonrandomized study but cannot be considered comparable to a well performed randomized trial"), serious ("the study has some important problems"), and critical ("the study is too problematic to provide any useful evidence and should not be included in any synthesis").

Received December 26, 2018
Revision received July 3, 2019
Accepted July 8, 2019 ■