

eScholarship@UMassChan

Disparate impact of risk assessment instruments: A systematic review

Item Type	Accepted Manuscript
Authors	Lawson, Spencer G.;Narkewicz, Emma L.;Vincent, Gina M.
Citation	Lawson SG, Narkewicz EL, Vincent GM. Disparate impact of risk assessment instruments: A systematic review. Law Hum Behav. 2024 Sep 30. doi: 10.1037/lhb0000582. Epub ahead of print. PMID: 39347719.
DOI	10.1037/lhb0000582
Publisher	American Psychological Association (APA)
Rights	Copyright ©American Psychological Association, 2024. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available at: https://doi.org/10.1037/lhb0000582 . Authors' accepted manuscript posted as allowed by the publisher's internet posting guidelines at https://www.apa.org/pubs/journals/resources/internet-posting-guidelines .
Download date	2024-12-23 08:38:02
Link to Item	https://hdl.handle.net/20.500.14038/53834

Disparate Impact of Risk Assessment Instruments: A Systematic Review

Spencer G. Lawson, Emma L. Narkewicz, and Gina M. Vincent

Law & Psychiatry Program, Implementation Science & Practice Advances Research Center,
Department of Psychiatry, University of Massachusetts Chan Medical School

Author Note

Spencer G. Lawson  <https://orcid.org/0000-0002-9572-169X>

Emma L. Narkewicz  <https://orcid.org/0009-0003-6199-2648>

Gina M. Vincent  <https://orcid.org/0000-0002-9315-9584>

The authors received funding from the John D. & Catherine T. MacArthur Foundation to complete this systematic review (subaward principal investigator: Gina Vincent; grant principal investigators: Sarah Desmarais & John Monahan). We have no other conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Spencer G. Lawson, PhD, University of Massachusetts Chan Medical School, Department of Psychiatry, Implementation Science & Practice Advances Research Center, Law & Psychiatry Program, 222 Maple Avenue, Shrewsbury, MA 01545, United States. Email: spencer.lawson1@umassmed.edu

Abstract

Objective: One concern about the use of risk assessment instruments in legal decisions is the potential for disparate impact by race or ethnicity. This means that one racial or ethnic group will experience harsher legal outcomes than another because of higher or biased risk estimates. We conducted a systematic review of the literature to synthesize research examining the real-world impact of juvenile and adult risk instruments on racial/ethnic disparities in legal decision making.

Hypotheses: Given the nature of research synthesis, we did not test formal hypotheses. **Method:**

Our systematic literature search as of July 2023 identified 21 articles that investigated the disparate impact of 13 risk assessment instruments on various legal outcomes. Most of these instruments were actuarial pretrial screening instruments. **Results:** Our narrative synthesis indicated that there is not strong evidence of risk instruments contributing to greater system disparity. Ten articles indicated that adopting risk instruments did not create (or exacerbate preexisting) disparities, and eight articles found that instrument use reduced disparities in legal decision making. Three articles reported evidence of disparate impact of risk instruments; only one of these studies received a strong study quality assessment score. We observed a scarcity of high-quality articles that employed what we deem to be the gold standard approach for examining the disparate impact of risk instruments (i.e., pretest-posttest design). **Conclusion:** The evidence signals that risk instruments can contribute to reductions in disparities across multiple stages of legal decision making. Yet study quality remains low and most research has been conducted on decisions during the pretrial stage. More rigorous research on disparate impact across diverse legal decision points and approaches to risk assessment is needed.

Keywords: risk assessment, fairness, disparate impact, bias, racial and ethnic disparities

Public Significance Statement

Risk assessment instruments have been criticized for widening racial and ethnic disparities in legal decision making. Current evidence overall signals that risk instruments are not resulting in harsher treatment for individuals with minoritized racial and ethnic identities. That said, future research must enhance methodological quality and examine proper implementation of risk assessment instruments.

©American Psychological Association, 2024. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. The final article is available, upon publication, at: 10.1037/lhb0000582

Accepted Version

Disparate Impact of Risk Assessment Instruments: A Systematic Review

Legal and forensic settings commonly incorporate risk screening and assessment (i.e., risk assessment instruments¹) to provide a structured determination of an individual's criminogenic risk-needs profile. These instruments offer an empirical approach to predict the likelihood of relevant outcomes (e.g., pretrial misconduct, criminal recidivism, institutional misconduct) and can help inform legal decisions, including pretrial detention, sentencing, facility security level, and supervision. In addition, risk assessment instruments can offer an informed picture of programmatic needs. Proponents of risk instruments claim that such tools promote efficiency and objectivity at key legal decision points better than unstructured clinical judgments (Desmarais et al., 2022; Milgram et al., 2015; Vincent & Viljoen, 2020). However, critics argue that risk instruments are biased and increase racial and ethnic disparities at multiple decision points across the criminal legal system (Harcourt, 2015; Pretrial Justice Institute, 2020; Starr, 2014).

In response to the growing scrutiny of risk assessment instruments, scholars have used well-established testing standards and guidelines for assessing “bias” or “fairness” in risk instruments (American Educational Research Association et al., 2014; Cleary, 1968; Skeem & Lowenkamp, 2016). Evaluations of whether an instrument is biased toward individuals with minoritized racial or ethnic identities must make the distinction between whether a tool shows evidence of test bias versus disparate impact. Test bias (i.e., predictive bias) is present when an instrument's risk score *functions* differently between groups of people. Alternatively, a risk

¹We use the term “risk assessment instruments” to encompass both risk screening and assessment tools. In decision points in which the primary concern is level of risk for criminal behavior (e.g., postarrest diversion, pretrial detention, prison classification), brief risk screening tools help “screen out” the individuals who are low risk and do not need further assessment or intervention. Alternatively, comprehensive risk-needs assessments can be conducted to determine criminogenic risk and programmatic needs, which generally involves in-depth evaluations that require specialized staff to conduct interviews and review administrative records. Risk assessments typically guide an individual's case planning and risk management (Vincent et al., 2012).

assessment instrument would have a disparate impact if there were unfair *application* of the instrument's score between groups. Current psychometric research demonstrates that only a few risk instruments contain predictive bias across race/ethnicity (e.g., Ahmed et al., 2023; Campbell & Miller, 2018; Desmarais et al., 2021). However, it remains unclear whether the use of risk instruments results in one racial or ethnic group experiencing harsher legal outcomes than another (i.e., disparate impact). The increased and often legally required use of risk assessment instruments (Monahan & Skeem, 2013; Seto, 2005; such as Indiana *Criminal Rule 26*, 2017) warrants a rigorous review and synthesis of the risk tool scholarship.

To our knowledge, only one systematic review has examined the disparate impact of risk assessment instruments and did so by examining restrictive placement decision making (Viljoen et al., 2019). The researchers determined that the evidence was insufficient to offer conclusions about disparate impact, given that most of the studies lacked scientific rigor. The present study extends the work of Viljoen and colleagues by reporting results of an updated systematic review that used narrative synthesis to examine disparate impact of risk instruments by race and ethnicity. Instead of limiting our focus solely to restrictive placement decisions, we broadened our scope to include various legal decision-making outcomes (e.g., pretrial release, length and type of sentencing, correctional programming decisions) in adult or juvenile forensic settings. This systematic review addressed the following research question: How does the adoption of risk assessment instruments impact racial and ethnic disparities at multiple stages of legal decision making?

Growth of Risk Assessment Instruments in Legal and Forensic Settings

Risk assessment instruments have proliferated across legal and forensic psychiatric settings (Grisso, 2005; Hu et al., 2017; Pretrial Justice Institute, 2019). For instance, juvenile

probation has adopted risk instruments for use in dispositional planning in all 50 U.S. states, with 42 states mandating use of a single instrument statewide (Juvenile Justice Geography, Policy, Practice & Statistics, 2020). Furthermore, most jurisdictions have implemented brief pretrial risk screening tools (Lattimore et al., 2020; Lawson et al., 2022; Pretrial Justice Institute, 2019). The idea that risk assessments can promote more informed, impartial judgments compared with assessing risk via unstructured clinical judgment alone has contributed to their widespread adoption. The use of these instruments has given professionals (e.g., pretrial officers, judges, psychologists, and psychiatrists) an opportunity to evaluate an individual's risk of engaging in continued criminal behavior or failing to comply with legal obligations. Risk instrument use also allows professionals to pinpoint areas in which risk management may be necessary (Lawson et al., 2022; Viljoen et al., 2018; Yu et al., 2023).

Risk assessment instruments consider a combination of items that cover a variety of static (i.e., historical factors such as number of prior offenses) or dynamic (i.e., changeable factors such as procriminal social networks) domains to formulate conclusions of criminogenic risk and needs. Shorter risk instruments, particularly those used for pretrial decisions, often contain only static items. Actuarial tools represent one type of instrument that requires a professional to add up item ratings to yield a total numerical score (Grove & Meehl, 1996), which corresponds to scaled risk estimates (e.g., low, moderate, high). Conversely, structured professional judgment instruments rely on a combination of well-defined and valid risk factors (structure) and a trained evaluator's discretion of their relevance to the individual's risk (judgment) to ascertain an estimate of risk level (Borum, 1996).

Legal professionals and forensic evaluators can select from a variety of risk assessment instruments, ranging from brief pretrial risk screening tools (e.g., Ohio Risk Assessment System–

Pretrial Assessment Tool [ORAS-PAT]; Latessa et al., 2010) to internationally used comprehensive risk/need assessments (e.g., Level of Service/Case Management Inventory; Andrews et al., 2004). For example, more than 400 different tools have been designed to assess one's likelihood of violence (Singh et al., 2014). A patchwork of local and state policies and procedures governs the adoption and implementation of risk instruments into case processing (Desmarais & Lowder, 2019; Vincent et al., 2012). In short, the nature and type of the tool, the policies on how the tool informs legal decision making, the level of stakeholder buy-in, and the amount of community resources will influence the tool's impact (Desmarais & Lowder, 2019; Douglas & Skeem, 2005; Viljoen et al., 2018).

Bias in Risk Assessment Instruments

Despite the expected benefits of facilitating objective and uniform decisions via risk assessment instruments, the heightened awareness of structural inequities throughout the U.S. criminal legal system has led to an influx of research and criticism of risk instruments. Specifically, scholars and advocacy groups have characterized risk tools as racially biased because they frequently incorporate unchangeable, static factors (e.g., criminal history) that reflect systemic inequities (Harcourt, 2015; Pretrial Justice Institute, 2020; Starr, 2014). Decades of research have documented evidence of racial/ethnic disparities from arrest to reentry decision-making stages (Kochel et al., 2011; Kurlychek & Johnson, 2019). Given this context, criminal history items incorporated into risk tools may compound structural racism carried over from earlier criminal legal contacts and, as a result, bias individuals—which falls disproportionately on people with minoritized racial and ethnic identities—toward higher risk estimates (Skeem et al., 2023; Vincent & Viljoen, 2020).

In the interest of assessing fairness or bias in risk instrument outcomes, the fields of criminology, psychology, law, and machine learning have advanced multiple definitions of these concepts (e.g., Ashford, Spivak, & Shepherd, 2022; Berk et al., 2021; Zottola et al., 2021). Broadly, evaluations of risk assessment instruments must make the distinction between whether an instrument shows evidence of test bias versus whether an instrument has a disparate impact (American Educational Research Association et al., 2014; Skeem & Lowenkamp, 2016). Test bias, also known as predictive bias in a risk context, is present when scores from a risk tool function differently between groups. In other words, scores should statistically relate to the outcome that the tool was designed to detect in the same way regardless of group membership. It is recommended that jurisdictions examine group differences in an instrument's predictive accuracy by comparing the areas under the curves of receiver operating characteristics and using an accepted regression-based procedure (i.e., Cleary method; Cleary, 1968; Meade & Fetzer, 2009; Mossman, 2013).

Disparate impact would be present if significant mean score differences between racial and ethnic groups on an instrument lead to unfair application of scores (or even the perception of unfair decision making; Skeem & Lowenkamp, 2016) or unequal legal outcomes, compared with baseline. The comparison with baseline is essential because a racial or ethnic disparity may already exist at any given point in the system. For the purpose of the current study, we define disparity as a group's representation at a particular point in the legal system differing substantially from the representation of other groups at that decision point. Put simply, it would be alarming if a risk instrument led to a greater proportion of sentenced Black individuals being incarcerated compared with White individuals, and this disparity was greater than the existing disparity observed prior to use of the instrument (baseline). Relative to studies of test bias, there

is considerably less research investigating the real-world impact of risk instruments on legal outcomes. To rigorously evaluate the disparate impact of risk assessment instruments, an investigation must compare the use of these tools with existing practices in the absence of risk instruments (i.e., the gold standard; Lowder et al., 2023; Vincent & Viljoen, 2020).

Empirical Research on Risk Assessments and Bias

To date, scholars have primarily conducted empirical research on the predictive accuracy of risk assessment instruments across groups (e.g., race, ethnicity; Ahmed et al., 2023; Campbell & Miller, 2018; Lawson & Lowder, 2023). The increase in predictive bias studies can be linked to a highly publicized 2016 report by *ProPublica*, which found that the Correctional Offender Management Profiling for Alternative Sanctions yielded biased assessments of risk against Black individuals on the basis of a comparison of error rates for outcomes (i.e., false positive rates and false negative rates; Angwin et al., 2016). Scholars have challenged Angwin and colleagues' methodology and conclusions (Chouldechova, 2017; Flores et al., 2016), and the report prompted an influx of research on bias in risk assessment instruments.

Overall, scholars have not found widespread predictive bias in risk instruments across diverse racial and ethnic groups. Campbell and Miller (2018) conducted a systematic review of 54 juvenile risk tool validation studies. The authors found that, of the studies that investigated predictive bias across racial and ethnic groups ($n = 24$), on average, risk assessment instruments predicted criminal recidivism similarly across groups. Meta-analytic findings, drawn from 17 distinct studies, have shown no substantive differences in the predictive accuracy of the Static-99R among Black, White, and Asian individuals charged with or convicted of sexually motivated offenses (Ahmed et al., 2023). However, the authors did find some evidence that the Static-99R might predict sexual recidivism with lower accuracy for Indigenous peoples and Hispanic

individuals. Using another meta-analysis of 11 studies examining six pretrial risk tools, Desmarais and colleagues (2021) reported, on average, comparable predictive validity estimates across racial/ethnic groups.

Despite the generally positive findings on test bias, ostensibly, what the critics are mainly concerned about is disparate impact. As Vincent and Viljoen (2020) noted, concerns about disparate impact warrant considerable attention because many risk instrument studies have found that individuals of color have significantly higher risk scores, on average, than White individuals. A mean score difference will occur on unbiased risk instruments if one group is truly more likely to recidivate, or more likely to be apprehended, than another group (i.e., true differences in recidivism risk), meaning that it is not necessarily indicative of bias. For instance, studies using the Youth Level of Service/Case Management Inventory (Hoge & Andrews, 2002) have found significant mean differences in risk scores between White youths and youths of color even though its predictive accuracy for these groups was not appreciably different (Barnes et al., 2016; Barnes-Lee & Campbell, 2020). Moreover, the mean differences in risk scores between racial/ethnic groups likely vary in magnitude across settings even on well-validated risk instruments because of systematic selection biases. For example, Black individuals are more likely than White individuals to come into contact with law enforcement and experience arrest (Carter et al., 2022; Kurlychek & Johnson, 2019; Pierson et al., 2020). Thus, the population of Black individuals who have been arrested is likely to have a greater proportion of low-risk individuals than White individuals who have been arrested.

Skeem and Lowenkamp (2016) asserted that risk instruments used in legal decision making should be both “empirically valid and perceived as morally fair across groups” (p. 685). To determine whether risk instruments are truly having a disparate impact (i.e., not morally fair),

we need to generate evidence that these instruments contribute to greater disparity in rates of incarceration (or other legal decisions) than the traditional approach of unstructured decision making (Vincent & Viljoen, 2020). Accordingly, Viljoen and colleagues (2019) conducted a systematic review of studies investigating how the adoption of risk instruments impacted rates of restrictive placements (i.e., pretrial placements, postconviction incarceration, release from secure facilities) for adult and juvenile defendants from diverse racial and ethnic backgrounds. The authors located six articles up to August 2017. Most of the articles ($k = 5$) reported a decrease in absolute rates of restrictive placements for individuals from underrepresented racial/ethnic groups following adoption of a risk instrument. Yet only two studies found a reduction in disparities, meaning the tool decreased restrictive placements more for individuals of color than for White individuals. Most of the six articles lacked scientific rigor, which led the authors to conclude that there was no strong evidence for or against risk instruments causing a disparate impact.

Clearly, research into the disparate impact of risk assessment instruments on racial/ethnic groups in legal and forensic settings has not kept pace with predictive bias studies. Increased media attention and recent scholarly works on this topic (e.g., Skeem & Lowenkamp, 2016) have prompted more recent studies examining the impact of risk instruments on injustices among individuals with minoritized racial and ethnic identities. Thus, there is a need for an updated systematic review. The current study extended the work of Viljoen and colleagues (2019) by conducting a systematic review of studies that investigated a broad scope of legal outcomes over the past two decades. We defined disparate impact studies as those examining the real-world impact of risk assessment instruments on legal outcomes across racial or ethnic groups. The

review included research on any risk assessment instrument used in adult or juvenile forensic settings as well as on disparities in any relevant legal outcome from pretrial to reentry.

Method

A systematic review typically entails the following steps: formulating a clear research question, establishing the review's scope, setting criteria for article inclusion, identifying all pertinent research, addressing bias in selected articles, and examining the articles to analyze results and draw conclusions (Higgins et al., 2019). To ensure sufficient reporting, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines and the APA Style *Journal Article Reporting Standards (JARS) - Quantitative Meta-Analysis Article Reporting Standards* (American Psychological Association, 2020). This study did not require institutional review because our data (i.e., articles) were accessed through publicly available sources (i.e., national and subject-specific bibliographic databases) or by contacting experts in the field. We collected no additional data through any interactions or interventions with human subjects.

Inclusion and Exclusion Criteria

Three primary inclusion criteria guided our systematic literature search:

1. The article's focal population of interest came from a legal setting (e.g., forensic hospital, juvenile or criminal legal system).
2. The article examined disparate impact of a risk assessment tool on a legal outcome (e.g., pretrial detention, bail, diversion, length and type of sentencing, and treatment-related referrals).
3. The article compared racial/ethnic differences before and/or after a tool was implemented or made a similar comparison using a statistical simulation with official data.

We did not include investigations of disparate impact via groups by other demographic characteristics (e.g., gender, age). Also, we did not include vignette studies because our systematic review focused on the real-world impact of risk assessment tools in pretrial, adjudication, or postdisposition decision making.

Literature Search

To identify relevant articles, we conducted a search for literature published between January 2000 and September 2022 in four bibliographic databases using three categories of Boolean search terms: Risk Assessment Instrument (“Risk assessment” OR “Risk instrument” OR “Risk tool” OR “Pretrial tool”) AND Racial/Ethnic Identity (“Race” OR “Racial Bias” OR “Ethnic*” OR “BIPOC” OR “Disproportionate minority contact” OR “Disparity” OR “Different*”) AND Criminal Recidivism (“Recidiv*” OR “Violence” OR “Re-offend*” OR “Reoffend*”). The search found articles in EBSCO Publishing Academic Search Complete ($n = 320$ unique articles), ProQuest Criminal Justice Database ($n = 275$ unique articles), APA PsycINFO ($n = 582$ unique articles), and ProQuest Dissertations & Theses Global ($n = 450$ unique articles). In addition, the review included the first 20 pages of Google Scholar using our Boolean search terms, yielding 198 records. After removal of duplicate records, 1,487 unique articles remained for screening.

Using Covidence (Veritas Health Innovation, 2023), all authors reviewed the first 200 titles and abstracts and held consensus meetings to maximize fidelity to inclusion and exclusion criteria. We double-screened the remaining article titles and abstracts and categorized as “include: yes” or “include: no.” Interrater reliability was moderate (Cohen’s $\kappa = 0.45$; 94.9% agreement; Landis & Koch, 1977), and all study authors discussed any conflicts until there was consensus. The final decision to include or exclude articles used consensus decision making.

This resulted in 33 articles for full text review. In addition, the researchers queried scholars for any gray literature or peer-reviewed articles that may not have been indexed in the four databases, resulting in another 31 articles and reports for full text review. We ceased contacting experts in the field in July 2023. In addition, we reviewed six articles included in Viljoen et al.'s (2019) systematic review of the impact of risk instruments on racial and ethnic disparities in restrictive placement decision making. The full text review resulted in 21 articles for inclusion. Figure 1 provides the number of articles initially identified, screened, and reviewed.

[FIGURE 1 ABOUT HERE]

Data Extraction

We developed our data extraction protocol by reviewing PRISMA guidelines, *JARS Quantitative Meta-Analysis Article Reporting Standards*, and other relevant data extraction approaches (e.g., Desmarais et al., 2021; Singh et al., 2011; Viljoen et al., 2019). We split the articles into two sets, and two authors (S.G.L. and E.L.N.) independently extracted the following information from their set of articles: (a) characteristics of the article (e.g., year of publication, publication format, study jurisdiction); (b) characteristics and content of the risk assessment tool (e.g., name of tool, number of items); (c) characteristics of the sample (e.g., sample size, racial and ethnic identities); and (d) disparate impact information (e.g., legal outcome of interest, preexisting disparities in outcomes [i.e., prior to risk tool use], summary of findings, use of *disparate impact* label). Most of the disparate impact information that coders extracted from each article represented string-based text. All authors discussed the disparate impact information, particularly the summary of findings, via consensus meetings to finalize disparate impact determinations for each article. After we completed initial data extraction, the first and last

authors (S.G.L. and G.M.V.) checked for data extraction accuracy and held additional consensus meetings to address coding queries and discrepancies.

We assessed study quality with the Quality Assessment Tool (QAT) for Quantitative Studies (Thomas et al., 2004). Raters using this tool examined study quality in eight categories (i.e., study design, analyses, withdrawals and dropouts, data collection methods, selection bias, intervention integrity, blinding as part of a controlled trial, and confounders), which generated an ordinal global quality rating (i.e., strong, moderate, and weak) for each article. Criminal legal scholars who conducted systematic reviews have used the QAT to assess salient methodological indicators (e.g., Lowder et al., 2018; Zottola et al., 2023). For the current study, all authors independently rated two articles on the QAT. We then held a consensus meeting to clarify our rating scheme and resolve rating discrepancies. The authors double-coded the remaining 19 articles at a substantial level of interrater reliability for the global QAT rating (Cohen's weighted $\kappa = 0.64$; 68.4% agreement; Landis & Koch, 1977). For each article for which the rater pair disagreed on the Global QAT rating, all three authors discussed the article until consensus was reached.

Narrative Synthesis

In our narrative synthesis of methods and findings, we tabulated characteristics of the article, sample, and risk assessment tool. Next, we synthesized the disparate impact findings from each article and identified reoccurring themes and patterns across all the articles (where k equals the number of articles). We conducted a systematic review, similar to the review by Viljoen and colleagues (2019), with narrative synthesis over a meta-analysis of studies on disparate impact of risk assessment instruments due to methodological heterogeneity and multiple study nuances (e.g., varying risk instruments, legal settings, and outcomes). Attempts to

construct a distribution of effect size estimates across results on risk assessment instruments would be an improper aggregation until more studies are available that use the same outcome measures, at the very least. Thus, we decided against operationalizing the quantitative findings on a numerical scale.

Results

Description of Included Articles

The full text review resulted in 21 articles for inclusion. As shown in Table 1, about one half (52.4%, $k = 11$) were peer-reviewed articles. Doctoral dissertations or master's theses accounted for 19.0% ($k = 4$), followed by government reports also at 19.0% ($k = 4$). In addition, there was one working paper (4.8%) and one preprint database entry (4.8%). Twelve articles (57.1%) reported findings from pretest-posttest designs, with the rest using posttest-only research designs (28.6%, $k = 6$), statistical simulations (9.5%, $k = 2$), and a randomized controlled trial with a statistical simulation (4.8%, $k = 1$). The publication years ranged from 2000 to 2023. All studies were conducted in the United States. Most articles (76.2%, $k = 16$) focused on preadjudication outcomes (e.g., pretrial release vs. detention, length of detention, bond type, charges dismissed). The remaining five articles (23.8%) examined disposition (including incarceration) or correctional programming outcomes. Five articles (23.8%) concentrated on adolescent samples in the juvenile justice system, whereas the remainder focused on the adult system.

[TABLE 1 ABOUT HERE]

In total, the articles used eight validated risk assessment instruments (see Table 1). The majority of these were pretrial risk tools, including the Public Safety Assessment ($k = 6$), ORAS-PAT ($k = 3$), Pretrial Risk Assessment ($k = 1$), and Kentucky Pretrial Risk Assessment ($k = 1$). A

few of the validated tools involved risk-needs assessment instruments (Positive Achievement Change Tool [$k = 2$], Level of Service Inventory–Revised [$k = 1$], Risk and Needs Triage [$k = 1$]), and one article used the Minnesota Screening Tool Assessing Recidivism Risk 2.0 (i.e., a correctional screening tool assessing recidivism risk). Two articles focused on homegrown instruments with unclear validity. Furthermore, three articles assessed outcomes of the Juvenile Detention Alternative Initiative, of which a pretrial detention tool is one part. This made it challenging to ascertain whether observed results were attributable to the risk tools or other Juvenile Detention Alternative Initiative efforts (e.g., detention alternative programming). All instruments being evaluated in the included articles were actuarial, with most being relatively short (≤ 13 items). No articles covered risk assessment instruments that followed a structured professional judgment approach to estimate risk. Eleven articles (52.4%) investigated disparate impact of a risk assessment instrument via comparisons between Black and White individuals, with seven articles (33.3%) adding or exclusively focusing on analyses for Hispanic or Latinx groups. Only two articles (9.5%) included other racial groups in their analyses, such as Asian and Native North American people. Several articles simply compared White individuals with “non-White” individuals or Black individuals with “non-Black” individuals (see Table 1).

Race and Ethnicity Conceptualizations

As noted above, the included articles categorized racial and ethnic groups as Black, non-Black, White, non-White, and Hispanic, among others. However, articles typically did not define these terms or discuss the classification of individuals with mixed racial or ethnic backgrounds. Furthermore, the articles did not consistently clarify whether racial or ethnic identities documented in administrative records were derived from self-reported identification or determined by criminal legal professionals who made identity assessments based on factors such

as skin color or surname. Staff may have also determined racial and ethnic categorizations based on place of birth, nationality, immigration status, or primary language. Inconsistencies in conceptualizing and operationalizing racial/ethnic minority identities in risk assessment studies are well documented (Ahmed et al., 2023; Fanniff et al., 2023).

Key Findings on Disparate Impact

Overall, findings of our systematic review showed no strong evidence of differential treatment by the legal system based on race or ethnicity as the result of risk instrument use (see Table 2). A majority of the 21 articles (85.7%, $k = 18$) reported that adopting risk instruments either did not create (or exacerbate preexisting) disparities or had beneficial impacts (i.e., reduced disparities). Out of the 18 articles that did not find a disparate impact, 10 reported that the use of risk assessment instruments did not lead to the creation of racial or ethnic disparities in criminal legal decision making, nor did it worsen existing ones (i.e., null effect). For example, Lowder and colleagues (2023) documented racial disparities in pretrial outcomes (e.g., pretrial release, bond amount, length of detention) in multiple jurisdictions prior to the use of a risk tool. Although disparities in pretrial release decisions continued to exist post-implementation of the risk tool, adoption of the risk tool did not exacerbate them, despite the risk tool showing evidence of predictive bias. In fact, the use of the risk tool relative to practice as usual increased overall rates of nonfinancial release for both Black and White defendants.

[TABLE 2 ABOUT HERE]

Eight articles (38.1%) presented evidence showing that the application of risk assessment instruments had advantageous effects on observed disparities in outcomes, such as the use of nonfinancial bonds, pretrial detention rates, and correctional programming decisions. A majority of the articles that documented a beneficial impact of tool adoption represented pretrial decision

points (87.5%, $k = 7$); most of these articles reported reductions in disparities relative to baseline. For instance, Sloan (2020) found that before the use of the ORAS-PAT, non-White and Hispanic defendants were 10% less likely to be released on nonfinancial bond and 40% more likely to be detained during the pretrial period than White defendants. Yet the adoption of the ORAS-PAT decreased racial and ethnic disparities in offers of nonfinancial bonds by judges and pretrial detention. A couple of studies used simulations, reporting that reductions in disparities would have occurred had a risk assessment instrument been followed. For the study in a correctional setting, Duwe (2021) found that if there had been greater adherence to risk-based decision making via the Minnesota Screening Tool Assessing Recidivism Risk 2.0 in the assignment of correctional programming, it could have contributed to more equitable risk reduction outcomes on release across racial groups compared with the status quo, particularly for Black and American Indian individuals.

Three articles reported evidence for disparate impact of risk assessment instruments. Lowder and colleagues (2019) found that the disparate impact was not in the hypothesized direction; the Level of Service Inventory–Revised contributed to White probationers with low-risk scores receiving longer sentences than Black probationers with similar risk scores. The other two articles showed heightened racial disparities, with Black individuals facing harsher pretrial release and sentencing decisions relative to White individuals because of risk instrument use (Schaefer & Hughes, 2019; Stevenson & Doleac, 2019).

Study Quality

Most articles received a moderate (33.3%, $k = 7$) or weak (38.1%, $k = 8$) rating on the QAT (see Table 2). Six (28.6%) articles received a strong QAT score. Overall, articles tended to be methodologically weak with respect to not reporting group differences before use of the risk

instrument, not controlling for or testing potential covariates, and/or sample selection biases (e.g., the jurisdiction did not administer the risk instrument to all eligible individuals at a particular decision point).

Out of the 18 articles that found no evidence of disparate impact as the result of risk screening or assessment, most were rated on the QAT as moderate (38.9%, $k = 7$) or weak (33.3%, $k = 6$), whereas five (27.8%) earned a strong QAT score. Among the five articles rated as strong quality by the QAT, Duwe (2021) was the only one that used a statistical simulation study design, which documented a beneficial impact. Out of the remaining higher quality articles, two employed pretest-posttest designs. The other two articles with strong QAT scores used research designs that were exclusively posttest, meaning that they reported findings on the basis of data collected after the adoption of a risk instrument. Overall, we found that high-quality articles using what we consider to be the gold standard research design for investigating disparate impact of risk instruments (i.e., namely, pretest-posttest comparisons) were scarce.

Only one out of the three articles reporting a disparate impact received a strong QAT score (i.e., Lowder et al., 2019), which found the opposite disparate impact effect. The two articles reporting negative outcomes for Black defendants were rated on the QAT as weak (i.e., Schaefer & Hughes, 2019; Stevenson & Doleac, 2019). These weak quality ratings stemmed from limitations of posttest-only designs, problems with the representativeness of the target population, and unanswered questions about the model building approach (e.g., selection of covariates).

Discussion

The increased controversy around the use of risk assessment instruments to inform legal decisions has led some jurisdictions and previous advocates to denounce risk tools (Pretrial

Justice Institute, 2020). The problem with efforts to eliminate risk instrument use is that, on the whole, these instruments have led to small reductions in restrictive placements (Viljoen et al., 2019), use of pretrial detention (Desmarais et al., 2022), formal processing (as opposed to diversion), and supervision intensity for those sentenced to probation (Vincent et al., 2016). To date, evidence from meta-analytic reviews (Desmarais et al., 2021; Olver et al., 2014; Singh et al., 2011) indicates that most risk instruments have reasonable accuracy for predicting criminal behavior, and there is a lack of consistent evidence to suggest that risk instruments exacerbate the longstanding racial or ethnic disparities present in legal systems (Kochel et al., 2011; Kurlychek & Johnson, 2019).

To intervene on this debate, criminal legal scholars and professionals should routinely test risk instruments used at every stage in the legal continuum for both predictive bias and disparate impact (Skeem & Lowenkamp, 2016). Unfortunately, disparate impact has not received the same level of attention as predictive bias. We conducted this systematic review to determine whether there is sufficient evidence to support or discourage use of risk assessment instruments as a method for promoting fairness and objectivity during legal decision making. There has been a noteworthy increase in disparate impact studies since Viljoen et al.'s (2019) systematic review, which included articles to mid-2017. Notably, a majority of the articles ($k = 16$, 76.2%) in the current review were published subsequent to the work of Skeem and Lowenkamp (2016), who played a pivotal role in elevating the concept of disparate impact of risk instruments among forensic psychologists, criminologists, and criminal legal scholars.

Summary of Findings

This rigorous review of the risk screening and assessment literature revealed no strong or consistent evidence that risk instruments are having widespread disparate impact by racial/ethnic

groups or worsening existing disparities in the legal system. To the contrary, there are more studies showing positive or null effects. We identified 21 articles spanning more than two decades that covered a range of key legal outcomes, including pretrial release versus detention, length of detention, dismissal of charges, and sentencing outcomes. Of the 21 articles, only three found that the specific risk assessment instrument contributed to disparate impact. One of these articles found that non-Latino White probationers received longer sentences than Black probationers at low levels of risk (Lowder et al., 2019), which is the opposite of the hypothesized disparate impact effect. The two articles reporting a disparate impact for Black defendants either did not address whether the disparities were worsened following implementation of the risk assessment instrument (Schaefer & Hughes, 2019) or raised some unanswered questions about their statistical model specification (Stevenson & Doleac, 2019).

Thirty-eight percent of the included articles ($k = 8$) reported that the respective risk instrument had a positive impact on reducing disparities or would have had a positive impact had the instrument been put into practice (for statistical simulation studies; DeMichele et al., n.d.; Duwe, 2021). The remaining articles ($k = 10$) found that risk assessment instruments had null effects or no impact on existing disparities. Out of the 10 articles, seven indicated that racial/ethnic disparities in legal outcomes were evident and remained so before and/or after the use of risk assessment instruments, despite no evidence of disparate impact. Some scholars asserted that these disparities in outcomes might have been mitigated if the risk instrument had been more closely followed (Lehmann et al., 2020; Marlowe et al., 2020). Of note, the Public Safety Assessment—one of the most commonly used pretrial screening instruments—showed no significant disparate impact in five out of six articles that assessed it. We advise caution in interpreting this trend due to the limited number of articles, some of which are of low quality. In

sum, there is more consistent evidence supporting the use of risk instruments in practice than there is evidence indicating that risk instruments cause harm to individuals with minoritized racial or ethnic identities.

Limitations

Our conclusion on the disparate impact of risk assessment instruments comes with a few caveats. First, almost one half of the disparate impact research to date has been published outside of peer-reviewed journals, which constrained the number of relevant articles retrieved by searching bibliographic databases. QAT scores indicated that more than one third of the articles had weak scientific rigor. The evidence base contains limited high-quality articles using what we consider to be the gold standard research design for investigating disparate impact (i.e., pre-post comparisons). Even among the extant pretest-posttest designs, few articles used robust statistical controls to balance groups (e.g., propensity matching, sufficiently controlling for covariates). Many articles did not adequately communicate whether significant baseline differences that may impact legal decisions existed between racial/ethnic samples, such as arrest history or risk level. Some articles also had serious sample selection biases (e.g., the responsible agency appeared to selectively administer risk assessments versus using them universally at a specific decision point, or studies required the defendant to consent to the risk assessment). Many articles either did not record or did not disclose the percentage of cases that should have undergone a risk assessment but were missed. To adequately determine whether a new procedure results in disparate impact, researchers must include or control for all the cases at this decision point that did not receive the procedure as intended.

A second caveat to the conclusions of this review is the limited range of racial and ethnic groups studied in relation to disparate impact, which has been primarily constrained to

comparisons of Black and White individuals. The lack of disparate impact research on individuals with Hispanic or Latinx ethnicity and Indigenous peoples represents a significant gap, especially because these groups are overrepresented in legal settings in the United States and internationally. Despite representation of international research on test bias with diverse groups (e.g., Ashford, Spivak, Ogloff, & Shepherd, 2022; Gutierrez et al., 2013; Muir et al., 2020), our systematic review failed to locate any international studies on disparate impact.

A final caveat is that most of the articles focused on pretrial-related outcomes and a narrow type of risk instrument (e.g., brief pretrial risk screening tools), with a few examining other decision-making stages or using more comprehensive risk assessments. Whereas most of the articles focused on pretrial instruments with primarily static risk factors, six articles evaluated assessment instruments (e.g., Positive Achievement Change Tool, Level of Service Inventory–Revised) or pretrial instruments (i.e., Pretrial Risk Assessment) that include dynamic risk items. It is unlikely that instruments with dynamic items are having different impacts. Indeed, the results of articles with these instruments were comparable with the instruments with mostly static items (e.g., most had null effects). Only one article on a risk assessment containing dynamic risk items reported a disparate impact, which may be attributed to implementation issues (Lowder et al., 2019). Furthermore, our systematic literature search did not yield any disparate impact studies that used structured professional judgment instruments (e.g., the Historical-Clinical-Risk Management-20; Douglas et al., 2014) or assessments administered by clinicians instead of legal system staff. The extent to which structured professional judgment instruments have a disparate impact on legal decision making is an open empirical question. Of note, most studies to date on structured professional judgment instruments examining predictive accuracy by race and/or

racial differences in proportions at each risk level have not reported race-related test bias (see Munoz et al., 2020; Vincent & Viljoen, 2020).

More studies of risk instruments used to inform other legal contexts (e.g., dispositions, case planning, prison classification, and release from incarceration) are also needed. Paroling authorities' decisions on release or revocations, for instance, represent a distinct intercept point that could benefit from the use of decision guideline tools, but there is limited research on bias in risk instrument use during conditional release (i.e., parole) decisions. Of note, our systematic review could examine the disparate impact of risk assessment instruments with only qualifying articles or available research for review. Although our evidence challenges the assumption that all risk assessment instruments are inherently biased, it is crucial to recognize that disparate impact has not yet been tested with all risk instruments in existence or at all decision points.

Implications for the Field and Future Directions

Along with consistent evidence of the predictive accuracy of risk instruments across diverse racial and ethnic groups and small reductions in restrictive placements linked to risk instrument use (e.g., Viljoen et al., 2019), our findings suggest that there is stronger evidence supporting the use of risk instruments than opposing it. Although 48% of articles ($k = 10$) indicated null effects of risk instrument use, the number of articles showing a decrease in disparities and a reduction in overall harsher legal outcomes (38.1%; $k = 8$) was greater compared with the number of articles (14.3%; $k = 3$) reporting the opposite findings. In fact, several articles indicated that disparities in final legal decisions occurred, but these disparities could have been avoided if the risk instrument had been properly used or adopted.

The impact of a risk assessment instrument, or any new evidence-based practice, is not merely a function of the validity of the practice. Clearly, whether these practices are

implemented to fidelity at a point that permits use in key decisions is equally, if not more, important (e.g., Viglione et al., 2015; Viljoen & Vincent, 2024; Vincent et al., 2012). Indeed, Lowder et al. (2023) found that use of a pretrial risk tool that demonstrated predictive bias by race still did not lead to racial disparities in pretrial decisions. We recommend that agencies, in addition to evaluating the chosen risk instrument for predictive bias and disparate impact, also address implementation issues (e.g., judicial and attorney training about their appropriate use). Whether a particular risk instrument can be tied to potential downstream disparities rests on whether the instrument is actually used to inform (not dictate) legal decisions.

This systematic review sheds light on multiple gaps in our current knowledge about disparate impact of risk instruments, leading to recommendations for future research. Only five of the 21 articles identified explicitly used the term *disparate impact*, which demonstrates a lack of a taxonomy or structured vocabulary in this area of research. An uncontrolled vocabulary and limitations of bibliographic databases (e.g., the journal's selection of abstracting and indexing databases, word limits for titles and abstracts) may have contributed to the limited number of relevant articles found via database searching (Cooper et al., 2019). It would be helpful for the risk assessment field to develop a common nomenclature for this type of work to make disparate impact research findable, accessible, interoperable, and reusable (Wilkinson et al., 2016). Given the legal and practical implications of this work, we strongly recommend that researchers disseminate findings through briefs or other materials that criminal legal professionals are likely to access. There is a need for more disparate impact studies with a higher level of scientific rigor, particularly in the areas of dispositions or sentencing and correctional outcomes with diverse racial/ethnic groups. We also recommend that future research do the following:

- Sample selection issues must be communicated, such as how many individuals never received the risk instrument. Sample selection biases represent an implementation concern that clearly affects whether any resulting disparities in decisions can be attributed to the risk tool. Eligible defendants who did not receive an expected risk instrument should be included in analyses or as a matched control group.
- Pre-post study designs should provide the rate of disparities prior to an instrument's adoption and any significant differences between the racial/ethnic samples after adoption that may affect legal decisions (e.g., gender, length of prior record).
- Pre-post study designs ideally should implement strong statistical controls such as propensity score matching or an equivalent.
- Scholars should assess how well the legal outcomes under investigation followed the risk level of the respective risk assessment instrument, or at least speak to implementation issues, such as judge and staff confidence in the instrument or adherence to its administration for all eligible defendants or sentenced individuals.
- Agencies should conduct pilot tests of the risk assessment instrument and regularly reevaluate its impact on legal decisions, considering the potential for progress as well as setbacks in the implementation process.

Conclusion

In sum, our review indicates that there is sufficient evidence at this time to support the use of risk assessment instruments, especially actuarial pretrial screening instruments. Risk instruments can serve as a potential strategy to help jurisdictions gain the anticipated benefits of these instruments without exacerbating, and possibly even reducing, existing racial and ethnic disparities. In this respect, more than one third of included articles found that risk assessment

instruments helped decrease preexisting racial and ethnic disparities in legal decision making. Nevertheless, the methodological quality of research must be enhanced, and further investigation is needed at points beyond the pretrial period and with additional approaches to risk assessment (e.g., structured professional judgment). Indeed, risk assessment approaches and tools are diverse, and not all risk instruments are equally effective in supporting broader reform goals in a jurisdiction, in part because of implementation issues. If the decision makers are not reviewing or properly using the instruments in their decisions, risk instruments will not have an effect. It is important to view risk assessment instruments as integral components of a multistage approach aimed at building racial and ethnic equity in the criminal legal system.

Accepted Version

References

- Ahmed, S., Lee, S. C., & Helmus, L. M. (2023). Predictive accuracy of Static-99R across different racial/ethnic groups: A meta-analysis. *Law and Human Behavior, 47*(1), 275–291. <https://doi.org/10.1037/lhb0000517>
- Albright, A. (2019). *If you give a judge a risk score: Evidence from Kentucky bail decisions* [Working paper]. https://thelittledataset.com/about_files/albright_judge_score.pdf
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Psychological Association. (2020). *JARS quantitative meta-analysis article reporting standards*. <https://apastyle.apa.org/jars/quant-table-9.pdf>
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *User's manual of the Level of Service/Case Management Inventory (LS/CMI): An offender assessment system*. Multi-Health Systems.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Ashford, L. J., Spivak, B. L., Ogloff, J. R. P., & Shepherd, S. M. (2022). The cross-cultural fairness of the LS/RNR: An Australian analysis. *Law and Human Behavior, 46*(3), 214–226. <https://doi.org/10.1037/lhb0000486>
- Ashford, L. J., Spivak, B. L., & Shepherd, S. M. (2022). Racial fairness in violence risk instruments: A review of the literature. *Psychology, Crime & Law, 28*(9), 911–941. <https://doi.org/10.1080/1068316X.2021.1972108>

Barnes, A. R., Campbell, N. A., Anderson, V. R., Campbell, C. A., Onifade, E., & Davidson, W.

S. (2016). Validity of initial, exit, and dynamic juvenile risk assessment: An examination across gender and race/ethnicity. *Journal of Offender Rehabilitation, 55*(1), 21–38.

<https://doi.org/10.1080/10509674.2015.1107004>

Barnes-Lee, A. R., & Campbell, C. A. (2020). Protective factors for reducing juvenile reoffending: An examination of incremental and differential predictive validity. *Criminal Justice and Behavior, 47*(11), 1390–1408. <https://doi.org/10.1177/0093854820952115>

Berk, R., Heidari, H., Jabbari, S., Kearns, M., & Roth, A. (2021). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research, 50*(1), 3–44.

<https://doi.org/10.1177/0049124118782533>

Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines, and training. *American Psychologist, 51*(9), 945–956.

<https://doi.org/10.1037/0003-066X.51.9.945>

Campbell, C., & Miller, W. (2018). A review of the validity of juvenile risk assessment across race/ethnicity. In H. Pontell (Ed.), *Oxford research encyclopedia of criminology and criminal justice*. Oxford University Press.

<https://oxfordre.com/criminology/display/10.1093/acrefore/9780190264079.001.0001/acrefore-9780190264079-e-345>

Carter, T., Knode, J., & Wolfe, S. E. (2022). *Exploring racial/ethnic disparities in Michigan State*

Police traffic stops using the veil of darkness methodology (pp. 1–33). Michigan

Statistical Analysis Center, School of Criminal Justice, Michigan State University.

<https://www.michigan.gov/msp/>

/media/Project/Websites/msp/reports/2022_Traffic_Stop_Report.pdf?rev=513ad7479b1547769af86b02568c1081

- Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, 5(2), 115–124.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.
- Criminal Rule 26, 94S00-1701-MS-5 (Indiana Supreme Court 2017).
<https://www.in.gov/courts/rules/criminal/>
- DeMichele, M., Silver, I. A., & Labrecque, R. M. (n.d.). *Smart decarceration: Are pretrial assessments an effective strategy to detain fewer people and reduce arrests?* [Working paper].
- Desmarais, S. L., & Lowder, E. M. (2019). *Pretrial risk assessment tools: A primer for judges, prosecutors, and defense attorneys*. Safety and Justice Challenge, John D. and Catherine T. MacArthur Foundation. <http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf>
- Desmarais, S. L., Monahan, J., & Austin, J. (2022). The empirical case for pretrial risk assessment instruments. *Criminal Justice and Behavior*, 49(6), 807–816.
<https://doi.org/10.1177/00938548211041651>
- Desmarais, S. L., Zottola, S. A., Duhart Clarke, S. E., & Lowder, E. M. (2021). Predictive validity of pretrial risk assessments: A systematic review of the literature. *Criminal Justice and Behavior*, 48(4), 398–420. <https://doi.org/10.1177/0093854820932959>

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014).

Historical-Clinical-Risk Management-20, Version 3 (HCR-20V3): Development and overview. *International Journal of Forensic Mental Health, 13*(2), 93–108.

<https://doi.org/10.1080/14999013.2014.906519>

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11*(3), 347–383.

<https://doi.org/10.1037/1076-8971.11.3.347>

Duwe, G. (2021). Applying the risk principle to optimize accuracy and equity in correctional risk assessment: Results from a simulation. *International Journal of Offender Therapy and Comparative Criminology, 65*(13–14), 1473–1495.

<https://doi.org/10.1177/0306624X20986523>

Fanniff, A. M., York, T., & Gutierrez, R. (2023). Developing consensus for culturally informed forensic mental health assessment: Experts' opinions on best practices. *Law and Human Behavior, 47*(3), 385–402. <https://doi.org/10.1037/lhb0000531>

Feyerherm, W. H. (2000). Detention reform and over-representation: A successful synergy.

Corrections Management Quarterly, 4(1), 44–51.

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to “Machine bias: There’s software used across the country to predict future criminals. And it’s biased against Blacks.” *Federal Probation, 80*(2), 38–46.

Grisso, T. (2005). Why we need mental health screening and assessment in juvenile justice programs. In T. Grisso, G. M. Vincent, & D. Seagrave (Eds.), *Mental health screening and assessment in juvenile justice* (pp. 3–21). Guilford Press.

- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293–323. <https://doi.org/10.1037/1076-8971.2.2.293>
- Gutierrez, L., Wilson, H. A., Rugge, T., & Bonta, J. (2013). The prediction of recidivism with aboriginal offenders: A theoretically informed meta-analysis. *Canadian Journal of Criminology and Criminal Justice*, 55(1), 55–99. <https://doi.org/10.3138/cjccj.2011.E.51>
- Harcourt, B. E. (2015). Risk as a proxy for race: The dangers of risk assessment. *Federal Sentencing Reporter*, 27(4), 237–243. <https://doi.org/10.1525/fsr.2015.27.4.237>
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions* (2nd ed.). The Cochrane Collaboration and John Wiley & Sons Ltd.
- Hoge, R. D., & Andrews, D. A. (2002). *The Youth Level of Service/Case Management Inventory: User's manual*. Multi-Health Systems.
- Hu, C., KiDeuk, K., & Mohr, E. (2017). *National scan of policy and practice in risk assessment* (Policy Brief No. 2017-01). The Risk Assessment Clearinghouse. <https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/pb-scan-of-practice.pdf>
- Imai, K., Jiang, Z., Greiner, D. J., Halen, R., & Shin, S. (2023). Experimental evaluation of algorithm-assisted human decision-making: Application to pretrial public safety assessment*. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 186(2), 167–189. <https://doi.org/10.1093/jrssa/qnad010>
- Juvenile Justice Geography, Policy, Practice & Statistics. (2020). *Juvenile justice practices: Risk assessment*. <http://www.jjgps.org/juvenile-justice-services#risk-assessment?year=2020>

- Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of suspect race on officers' arrest decisions. *Criminology*, *49*(2), 473–512. <https://doi.org/10.1111/j.1745-9125.2011.00230.x>
- Kurlychek, M. C., & Johnson, B. D. (2019). Cumulative disadvantage in the American criminal justice system. *Annual Review of Criminology*, *2*(1), 291–319. <https://doi.org/10.1146/annurev-criminol-011518-024815>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. <https://doi.org/10.2307/2529310>
- Lapcevich, H. (2020). *Does a validated risk assessment reduce the likelihood of an individual recidivating as well as bond disparity in court (Publication No. 28678920)* [Master's thesis, Youngtown State University]. ProQuest Dissertations Publishing.
- Latessa, E. J., Lemke, R., Makarios, M., Smith, P., & Lowenkamp, C. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, *74*(1), 16–23.
- Lattimore, P. K., Tueller, S., Levin-Rector, A., & Witwer, A. (2020). The prevalence of local criminal justice practices. *Federal Probation*, *84*(1), 28–37.
- Lawson, S. G., & Lowder, E. M. (2023). Differential prediction by race in IRAS-PAT assessments: An application of debiasing strategies. *Justice Quarterly*, *40*(4), 451–477. <https://doi.org/10.1080/07418825.2022.2086481>
- Lawson, S. G., Rising, S. J., Grommon, E., & Lowder, E. M. (2022). Pretrial risk assessment tool adoption and pretrial operations. In C. S. Scott-Hayward, J. E. Copp, & S. Demuth (Eds.), *Handbook on pretrial justice* (Vol. 6, pp. 276–295). Routledge & Kegan Paul.

- Lehmann, P. S., Meldrum, R. C., & Greenwald, M. A. (2020). Upward departures from structured recommendations in juvenile court dispositions: The intersection of race, ethnicity, and gender. *Justice Quarterly*, 37(3), 514–540.
<https://doi.org/10.1080/07418825.2018.1531143>
- Lowder, E. M., Diaz, C. L., Grommon, E., & Ray, B. R. (2023). Differential prediction and disparate impact of pretrial risk assessments in practice: A multi-site evaluation. *Journal of Experimental Criminology*, 19, 561–594. <https://doi.org/10.1007/s11292-021-09492-9>
- Lowder, E. M., Morrison, M. M., Kroner, D. G., & Desmarais, S. L. (2019). Racial bias and LSI-R assessments in probation sentencing and outcomes. *Criminal Justice and Behavior*, 46(2), 210–233. <https://doi.org/10.1177/0093854818789977>
- Lowder, E. M., Rade, C. B., & Desmarais, S. L. (2018). Effectiveness of mental health courts in reducing recidivism: A meta-analysis. *Psychiatric Services*, 69(1), 15–22.
<https://doi.org/10.1176/appi.ps.201700107>
- Lowenkamp, C., DeMichele, M., & Warren, L. K. (2020). *Replication and extension of the Lucas County PSA project* (pp. 1–84). RTI International.
- 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(5), 900–927. <https://doi.org/10.1080/07418825.2013.863961>
- Marlowe, D. B., Ho, T., Carey, S. M., & Chadick, C. D. (2020). Employing standardized risk assessment in pretrial release decisions: Association with criminal justice outcomes and racial equity. *Law and Human Behavior*, 44(5), 361–376.
<https://doi.org/10.1037/lhb0000413>

- Meade, A. W., & Fetzner, M. (2009). Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods, 12*(4), 738–761. <https://doi.org/10.1177/1094428109331487>
- Milgram, A., Holsinger, A. M., Vannostrand, M., & Alsdorf, M. W. (2015). Pretrial risk assessment: Improving public safety and fairness in pretrial decision making. *Federal Sentencing Reporter, 27*(4), 216–221. <https://doi.org/10.1525/fsr.2015.27.4.216>
- Monahan, J., & Skeem, J. L. (2013). Risk redux: The resurgence of risk assessment in criminal sentencing. *Federal Sentencing Reporter, 26*(3), 158–166.
- Moore, L. D. (2011). *The influence of risk assessment instruments on racial/ethnic disparities in the sentencing of juvenile offenders (Publication No. 3483589)* [Doctoral dissertation, The Florida State University]. ProQuest Dissertations Publishing.
- Mossman, D. (2013). Evaluating risk assessments using receiver operating characteristic analysis: Rationale, advantages, insights, and limitations. *Behavioral Sciences & the Law, 31*(1), 23–39. <https://doi.org/10.1002/bsl.2050>
- Muir, N. M., Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., & Rogers, B. J. (2020). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) with Indigenous and Caucasian female and male adolescents on probation. *Psychological Assessment, 32*(6), 594–607. <https://doi.org/10.1037/pas0000816>
- Munoz, C. G., Perrault, R. T., & Vincent, G. M. (2020). Probation officer assessments of risk when the youth look different: Contributions of structured professional judgment to concerns about racial bias. *Youth Violence and Juvenile Justice, 19*(2), 206–226. <https://doi.org/10.1177/1541204020954264>

- 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*(1), 156–176. <https://doi.org/10.1037/a0035080>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal, 372*, Article 71. <https://doi.org/10.1136/bmj.n71>
- Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour, 4*(7), 736–745. <https://doi.org/10.1038/s41562-020-0858-1>
- Pretrial Justice Institute. (2019). *Scan of pretrial practices* (No. 2019-10-02KN). <https://www.pretrial.org/files/resources/scanofpretrialpractices.pdf>
- Pretrial Justice Institute. (2020). *The case against pretrial risk assessment instruments*. [https://www.pretrial.org/resources/the-case-against-pretrial-risk-assessments#:~:text=Pretrial%20risk%20assessment%20instruments%20\(RAIs,or%20commit%20a%20violent%20crime](https://www.pretrial.org/resources/the-case-against-pretrial-risk-assessments#:~:text=Pretrial%20risk%20assessment%20instruments%20(RAIs,or%20commit%20a%20violent%20crime)
- Puzzanchera, C., Knoll, C., Adams, B., & Sickmund, M. (2012). *Allegheny County detention screening study*. National Center for Juvenile Justice. http://www.ncjj.org/pdf/MFC/FINAL_Allegheny%20Detention%20Assessment.pdf

- Redcross, C., Henderson, B., Miratrix, L., & Valentine, E. (2019). *Evaluation of pretrial justice system reforms that use the Public Safety Assessment: Effects in Mecklenburg County, North Carolina* (Brief 2, pp. 1–41). MDRC Center for Criminal Justice Research.
https://www.mdrc.org/sites/default/files/PSA_Mecklenburg_Brief2.pdf
- Schaefer, B. P., & Hughes, T. (2019). Examining judicial pretrial release decisions: The influence of risk assessments and race. *Criminology, Criminal Justice, Law & Society*, 20(2), 47–58.
- Seto, M. C. (2005). Is more better? Combining actuarial risk scales to predict recidivism among adult sex offenders. *Psychological Assessment*, 17(2), 156–167.
<https://doi.org/10.1037/1040-3590.17.2.156>
- Simpson, T. P. (2010). *Do objective measures reduce the disproportionate rates of minority youth placed in detention: Validation of a risk assessment instrument? (Publication No. 1117)* [Doctoral dissertation, University of New Orleans]. ScholarWorks.
- Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., & Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13(3), 193–206.
<https://doi.org/10.1080/14999013.2014.922141>
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980

- participants. *Clinical Psychology Review*, 31(3), 499–513.
<https://doi.org/10.1016/j.cpr.2010.11.009>
- Skeem, J. L., & Lowenkamp, C. T. (2016). Risk, race, and recidivism: Predictive bias and disparate impact. *Criminology*, 54(4), 680–712. <https://doi.org/10.1111/1745-9125.12123>
- Skeem, J. L., Montoya, L., & Lowenkamp, C. (2023). Understanding racial disparities in pretrial detention recommendations to shape policy reform. *Criminology & Public Policy*, 22(2), 233–262. <https://doi.org/10.1111/1745-9133.12620>
- Sloan, C. W. (2020). *Essays on the economics of crime and discrimination* [Doctoral dissertation, Texas A&M University]. OAKTrust.
- Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, 66, 803–872.
- Stevenson, M. (2018). Assessing risk assessment in action. *Minnesota Law Review*, 58, 303–384.
- Stevenson, M., & Doleac, J. L. (2019). Algorithmic risk assessment in the hands of humans. *SSRN Electronic Journal*, 1–71. <https://doi.org/10.2139/ssrn.3489440>
- Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 1(3), 176–184.
<https://doi.org/10.1111/j.1524-475X.2004.04006.x>
- Veritas Health Innovation. (2023). *Covidence systematic review software* [Computer software]. www.covidence.org
- Viglione, J. L., Rudes, D. S., & Taxman, F. S. (2015). Misalignment in supervision: Implementing risk/needs assessment instruments in probation. *Criminal Justice and Behavior*, 42(3), 263–285. <https://doi.org/10.1177/0093854814548447>

- 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*(3), 181–214. <https://doi.org/10.1037/lhb0000280>
- Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., & Vincent, G. M. (2019). Impact of risk assessment instruments on rates of pretrial detention, postconviction placements, and release: A systematic review and meta-analysis. *Law and Human Behavior, 43*(5), 397–420. <https://doi.org/10.1037/lhb0000344>
- Viljoen, J. L., & Vincent, G. M. (2024). Risk assessments for violence and reoffending: Implementation and impact on risk management. *Clinical Psychology: Science and Practice, 31*(2), 119–131. <https://doi.org/10.1111/cpsp.12378>
- Vincent, G. M., Guy, L. S., & Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*. John D. and Catherine T. MacArthur Foundation. https://modelsforchange.net/publications/346/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*(6), 683–696. <https://doi.org/10.1037/lhb0000214>
- Vincent, G. M., & Viljoen, J. L. (2020). Racist algorithms or systemic problems? Risk assessments and racial disparities. *Criminal Justice and Behavior, 47*(12), 1576–1584. <https://doi.org/10.1177/0093854820954501>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers,

- R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yu, R., Molero, Y., Lichtenstein, P., Larsson, H., Prescott-Mayling, L., Howard, L. M., & Fazel, S. (2023). Development and validation of a prediction tool for reoffending risk in domestic violence. *JAMA Network Open*, 6(7), Article e2325494. <https://doi.org/10.1001/jamanetworkopen.2023.25494>
- Zottola, S. A., Crozier, W. E., Ariturk, D., & Desmarais, S. L. (2023). Court date reminders reduce court nonappearance: A meta-analysis. *Criminology & Public Policy*, 22(1), 97–123. <https://doi.org/10.1111/1745-9133.12610>
- Zottola, S. A., Desmarais, S. L., Lowder, E. M., & Duhart Clarke, S. E. (2021). Evaluating fairness of algorithmic risk assessment instruments: The problem with forcing dichotomies. *Criminal Justice and Behavior*, 49(3), 1–22. <https://doi.org/10.1177/00938548211040544>

Table 1*Descriptions of Included Articles*

Author(s), year (location)	Study year(s)	Publication format	Design	Sample size	Racial/ ethnic groups compared ^a	Outcome type	Risk tool
Albright, 2019 (KY, USA)	2009–2013	Government report	Pre-post	383,080	Black, White	Bond type	KPRA
DeMichele et al., n.d. (USA)	2017–2018	Working paper	Simulation	28,188	Black, White	Pretrial release	PSA
Duwe, 2021 (MN, USA)	2017–2019	Peer reviewed	Simulation	9,529	African American, White, American Indian, Asian, Hispanic	Correctional program assignment	MnSTARR 2.0
Feyerherm, 2000 (OR, USA)	1990–1991 & 1996– 1997	Peer reviewed	Pre-post	18,788	Black, White, Asian, Indian, and Hispanic	Pretrial detention	JDAI RAI
Imai et al., 2023 (WI, USA)	2017–2019	Peer reviewed	Blended randomized controlled trial with simulation	1,891	Non-White, White	Bond type & amount	PSA
Lapcevich, 2020 (OH, USA)	2017 & 2019	Dissertation/ thesis	Pre-post	588	Minority (White Hispanic or Latino, Black, unknown/other)	Bond type & amount	ORAS-PAT
Lehmann et al., 2020 (FL, USA)	2012–2015	Peer reviewed	Posttest	56,913	Black, White, Hispanic	Above- guidelines disposition	PACT

Author(s), year (location)	Study year(s)	Publication format	Design	Sample size	Racial/ ethnic groups compared ^a	Outcome type	Risk tool
Lowder et al., 2023 (IN, USA)	2015–2018	Peer reviewed	Pre-post (nonequivalent control groups)	7,204	Black, White	Bond type & amount, pretrial release, pretrial detention length	ORAS-PAT ^b
Lowder et al., 2019 (KS, USA)	2003–2015	Peer reviewed	Posttest	11,585	Black, White	Sentence length	LSI-R
Lowenkamp et al., 2020 (OH, USA)	2012–2018	Government report	Pre-post (matched)	40,200	Black, non-Black	Pretrial release	PSA
Maloney & Miller, 2015 (NJ, USA)	2007–2009	Peer reviewed	Pre-post (matched)	1,432	African American, White, Hispanic	Pretrial detention	JDAI RAI
Marlowe et al., 2020 (MS, USA)	2015–2016	Peer reviewed	Posttest	521	African American, Caucasian	Pretrial detention length, deferred adjudication, charge dismissal, entry of conviction, not guilty verdict	RANT
Moore, 2011 (FL, USA)	2006–2008	Dissertation/ thesis	Posttest	26,681	Black, White, Latino/a	Disposition type	PACT
Puzzanchera et al., 2012 (PA, USA)	2007–2010	Government report	Pre-post	2,098	Black, White	Pretrial detention	JDAI RAI
Redcross et al., 2019	2012–2015	Government report	Pre-post	93,950	Black, White	Restrictiveness of release	PSA

Author(s), year (location)	Study year(s)	Publication format	Design	Sample size	Racial/ ethnic groups compared ^a	Outcome type	Risk tool
(NC, USA)						conditions, pretrial detention, case dismissal	
Schaefer & Hughes, 2019 (KY, USA)	2014–2017	Peer reviewed	Posttest	25,614	Black, White	Bond type	PSA
Simpson, 2010 (LA, USA)	2008	Dissertation/ thesis	Pre-post	202	Black	Pretrial detention	Homegrown RAI
Skeem et al., 2023 (USA)	2015–2019	Peer reviewed	Posttest	149,816	Black, White	Pretrial detention recommendation	PTRA
Sloan, 2020 (TX, USA)	2011–2014	Dissertation/ thesis	Pre-post	143,092	Non-White, Hispanic, White	Bond type, pretrial detention	ORAS-PAT
Stevenson, 2018 (KY, USA)	2009–2016	Peer reviewed	Pre-post	1,030,732	Black, White	Bond type, pretrial release	PSA
Stevenson & Doleac, 2019 (VA, USA)	2001–2004	Preprint database entry	Pre-post	58,744	Black, White	Incarceration, sentence length	Homegrown RAI

Note. KPRA = Kentucky Pretrial Risk Assessment; PSA = Public Safety Assessment; MnSTARR = Minnesota Screening Tool Assessing Recidivism Risk; JDAI = Juvenile Detention Alternative Initiative; RAI = Risk Assessment Instrument; ORAS-PAT = Ohio Risk Assessment System–Pretrial Assessment Tool; PACT = Positive Achievement Change Tool; LSI-R = Level of Service Inventory–Revised; RANT = Risk and Needs Triage; PTRA = Pretrial Risk Assessment.

^a Racial and ethnic identity align with the labels used in the article. ^b The Indiana Risk Assessment System–Pretrial Assessment Tool represents the RAI under investigation in the article, which is based on the ORAS-PAT.

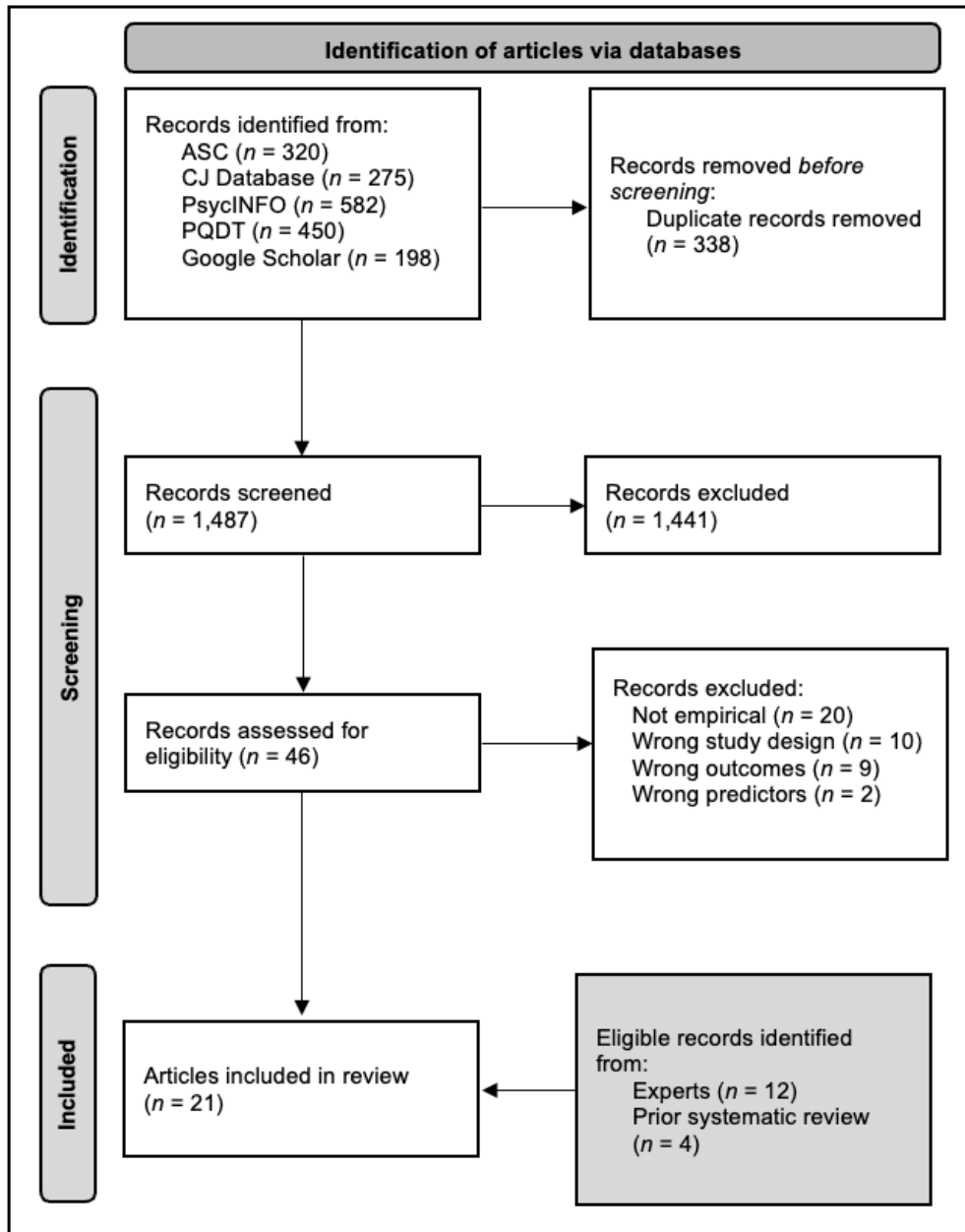
Table 2*How Does the Adoption of Risk Assessment Instruments Impact Racial and Ethnic Disparities in Justice Decision Making?*

Author(s), year	QAT rating	Preexisting racial/ethnic disparities	Results	Disparate impact term used
<i>Articles showing null effects</i>				
Lehmann et al., 2020	Strong	NA (posttest design)	Black youths had a higher likelihood of receiving a more severe disposition than recommended by the juvenile RAI-informed Disposition Recommendation Matrix compared with White youths, suggesting that deviations from the matrix are contributing to the racial disparities. No disparity in receiving an above-guidelines disposition between Hispanic and White youths.	No
Lowder et al., 2023	Strong	Yes	Despite disparities in pretrial decisions, the effect of RAI adoption on pretrial decisions did not differ across Black and White defendants.	Yes
Skeem et al., 2023	Strong	NA (posttest design)	Racial disparities in RAI pretrial detention recommendations most noticeable in situations in which there is substantial officer discretion, operating primarily through one institutionalized factor: criminal history.	Yes
Lowenkamp et al., 2020	Moderate	No	Pretrial release rates decreased similarly for Black and non-Black defendants after RAI adoption, but RAI adoption did not have a significant effect on the odds of release between Black and non-Black defendants.	No
Maloney & Miller, 2015	Moderate	No	The effect of juvenile RAI adoption on detention decision making did not differ across racial and ethnic groups.	No
Moore, 2011	Moderate	NA (posttest design)	The RAI effect on disposition decision making (i.e., commitment or community disposition) did not differ across racial and ethnic groups.	No
Redcross et al., 2019	Moderate	Yes	RAI adoption had, on average, little effect on racial disparities in pretrial detention or case dismissals; no evidence of racial disparities in bail setting pre- or postimplementation of RAI.	No

Author(s), year	QAT rating	Preexisting racial/ethnic disparities	Results	Disparate impact term used
Stevenson, 2018	Moderate	Yes	RAI adoption increased nonfinancial release for both White and Black defendants, but there were larger gains for White defendants. Similar trend observed for the likelihood of being released within 3 days of booking. Racial gaps postimplementation explained by regional differences.	Yes
Imai et al., 2023	Weak	No	The effect of RAI adoption on bond decision making (i.e., signature or cash bond) did not differ across White and non-White men.	No
Marlowe et al., 2020	Weak	NA (posttest design)	RAI resulted in comparable risk scores for both African American and Caucasian defendants, but African American defendants experienced longer periods of pretrial detention and were less likely to receive diversion opportunities compared with Caucasian defendants, which may be attributed to economic status or neighborhood stability. No racial disparities in charge dismissal, entry of conviction, or not guilty verdict.	No
<i>Articles showing beneficial effects</i>				
Duwe, 2021	Strong	Yes	Simulated RAI-based policy could equitably improve correctional programming decisions, particularly for African American and American Indian individuals.	No
Sloan, 2020	Strong	Yes	RAI adoption increased the likelihood of release on nonfinancial bonds and decreased the chance of pretrial detention for non-White and Hispanic defendants compared with White defendants.	No
Albright, 2019	Moderate	Yes	RAI adoption increased nonfinancial bonds for both Black and White defendants, but there was a larger increase for White defendants. Racial gap explained by tool deviations (i.e., overrides) from judges.	No
DeMichele et al., n.d.	Moderate	Yes	Simulated RAI-based release process could improve observed release decisions, but there were greater forecasted gains for White compared with Black defendants.	No
Feyerherm, 2000	Weak	Yes	Juvenile RAI adoption decreased the overall use of pretrial detention and reduced disparities between White and non-White youths.	No

Author(s), year	QAT rating	Preexisting racial/ethnic disparities	Results	Disparate impact term used
Lapceвич, 2020	Weak	NA (pre-post within-group comparison)	For minority individuals, RAI adoption increased the release on own recognizance and supervised release type of initial bond and decreased monetary bond.	No
Puzzanchera et al., 2012	Weak	Yes	Juvenile RAI adoption led to comparable decreases in detention rates for both Black and White youths (i.e., the proportion of youths recommended for detention out of the total RAI assessments conducted), but significant disparities were observed for the youths who were referred to the detention center.	No
Simpson, 2010	Weak	Insufficient data (pre-post within-group comparison)	The number of Black youths detained was similar pre- and postimplementation of juvenile RAI, but there was evidence to suggest that RAI override use may have helped Black youths avoid further legal involvement.	No
<i>Articles showing disparate impact</i>				
Lowder et al., 2019	Strong	NA (posttest design)	RAI increased length of sentence for White probationers—particularly those at low risk—compared with Black probationers.	Yes
Schaefer & Hughes, 2019	Weak	NA (posttest design)	RAI—particularly moderate or high-risk scores—increased the odds of receiving a financial bond for Black defendants compared with White defendants.	No
Stevenson & Doleac, 2019	Weak	Yes	RAI adoption had little effect on racial disparities in felony sentencing statewide, but courts that used RAI most frequently experienced an increase in racial disparities.	Yes

Note. NA = not applicable; RAI = Risk Assessment Instrument.

Figure 1*Flow Diagram of Systematic Literature Search*

Note. ASC = EBSCO Publishing Academic Search Complete; CJ Database = ProQuest Criminal Justice Database; PQDT = ProQuest Dissertations & Theses Global.