PREDICTIVE VALIDITY OF PRETRIAL RISK ASSESSMENTS

A Systematic Review of the Literature

SARAH L. DESMARAIS
SAMANTHA A. ZOTTOLA
SARAH E. DUHART CLARKE
North Carolina State University
EVAN M. LOWDER
George Mason University

Bail reform is sweeping the nation and many jurisdictions are looking to pretrial risk assessment as one potential strategy to support these efforts. This article summarizes the findings of a systematic review of research examining the predictive validity of pretrial risk assessments. We reviewed 11 studies (13 publications) examining the predictive validity of six pretrial risk assessment instruments reported in the gray and peer-reviewed literature as of December, 2018. Findings typically show good to excellent predictive validity. Differences in predictive validity for men and women were mixed and small. When it could be examined, predictive validity was generally comparable across racial/ethnic subgroups; however, three comparisons revealed notably lower, albeit still fair to good, predictive validity for defendants of color than White defendants. Findings suggest that pretrial risk assessments predict pretrial outcomes with acceptable accuracy, but also emphasize the need for continued investigation of predictive validity across gender and racial/ethnic subgroups.

Keywords: risk assessment; predictive validity; decision-making; criminal justice system; race

Across the United States, efforts are underway to maximize pretrial release rates while minimizing pretrial misconduct, including failure to appear in court and perpetration of new crime during the pretrial period. Since 2012, every state in the United States has enacted some form of pretrial legal reform (National Conference on State Legislatures, 2018), with many states implementing pretrial risk assessment instruments as part of these efforts. Briefly, pretrial risk assessment instruments are designed to forecast the likelihood of failure to appear in court and/or perpetration of new crime during the pretrial period

AUTHORS' NOTE: *Correspondence concerning this article should be addressed to Sarah L. Desmarais, Department of Psychology, North Carolina State University, Campus Box 7650, Raleigh, NC 27695; e-mail: sdesmarais@ncsu.edu.*

CRIMINAL JUSTICE AND BEHAVIOR, 201X, Vol. XX, No. X, Month 2020, 1–23.
DOI: 10.1177/0093854820932959
Article reuse guidelines: sagepub.com/journals-permissions

through consideration of items that have been shown in research to be associated with these outcomes. In this way, pretrial risk assessment instruments may provide some empirical evidence to inform pretrial decisions (Desmarais & Lowder, 2019). A recent scan of pretrial practices across the United States found that approximately two thirds of surveyed counties reported using a pretrial risk assessment instrument (Pretrial Justice Institute, 2019b). Despite their widespread implementation, pretrial risk assessment is the subject of considerable controversy.

Some of the controversy centers on whether pretrial risk assessment instruments are inherently biased because they are developed using data that reflect biased policing practices that target people of color (Eckhouse et al., 2019; Mayson, 2019). Other critiques reflect beliefs that pretrial risk assessments cannot predict pretrial outcomes, that they increase rates of detention, and that they exacerbate racial biases in pretrial decision-making (e.g., Barabas et al., 2019; Pretrial Justice Institute, 2020; Scurich & Krauss, 2020). Behind these critiques, at least in part, are analyses of risk assessments completed using a single instrument in a single jurisdiction, including ProPublica's analysis of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in Broward County, Florida (Angwin et al., 2016) and a subsequent study that concluded COMPAS assessments performed no better than lay persons at predicting recidivism (Dressel & Farid, 2018). Although these two particular studies have been criticized and rebutted in the academic literature (e.g., Bansak, 2019; Flores et al., 2016; Holsinger et al., 2018; Lin et al., 2020), they are among the most prominent sources cited in calls to abolish risk assessment instruments from the criminal justice system.

Discussion of racial bias in risk assessment—pretrial or otherwise—often conflates several issues (Desmarais & Zottola, in press), a few of which we touch on here. A first issue is whether Black and Brown people receive higher risk scores or classifications than White people. A second issue is whether Black and Brown people have higher rates of criminal behavior than others (actual or as the result of unfair police practices). A third issue is whether Black and Brown people are over classified at higher risk levels and under classified at lower risk levels relative to their actual rates of criminal behavior. It is this third issue that speaks to predictive validity. In particular, pretrial risk assessments should be able to forecast pretrial outcomes with comparable predictive validity, regardless of the base rate of offending (Lowder et al., 2019). To our knowledge, there has been no effort to date that examines and compares the predictive validity of pretrial risk assessments as a function of race/ethnicity, or gender for that matter, across instruments and outcomes. We sought to address this gap.

PRETRIAL RISK ASSESSMENT INSTRUMENTS

Pretrial risk assessment instruments were designed as a strategy to help overcome some of the limitations of human decision-making that may contribute to biased and unfair decisions in the pretrial context (Desmarais & Lowder, 2019). It is well accepted that human judgment is influenced by our personal beliefs and that increasing structure can reduce reliance on heuristics in the decision-making process, thereby producing less biased and more accurate decisions (Tversky & Kahneman, 1974). Consequently, increasing structure through the use of risk assessment instruments may contribute to more accurate and less biased pretrial decisions. Indeed, meta-analyses of research conducted over the past 60 years show that predictions of future behavior completed using statistical methods produce more accurate assessments of future violent and criminal behavior compared with unstructured human judgments (Ægisdóttir et al., 2006; Grove et al., 2000). While a similar body of work does not exist in the pretrial context specifically, findings of these studies suggest that we may anticipate such gains in the accuracy of predictions regarding future behavior during the pretrial period, as well.

The first pretrial-specific risk assessment instrument, the Vera Point Scale, was created and adopted in New York City more than 50 years ago (Ares et al., 1963). Since then, additional pretrial risk assessment instruments have been introduced, some of which we highlight here. Several pretrial risk assessment instruments were designed for use in a specific state and then adopted by other jurisdictions, such as the Virginia Pretrial Risk Assessment Instrument (VPRAI; VanNostrand, 2003) and the Ohio Risk Assessment System-Pretrial Assessment Tool (ORAS-PAT; Latessa et al., 2009). Another pretrial risk assessment instrument, the U.S. Federal Pretrial Services Risk Assessment Instrument (PTRA; Lowenkamp & Whetzel, 2009), was developed for use with individuals charged with federal offenses. Most recently, the Public Safety Assessment (PSA; VanNostrand & Lowenkamp, 2013) was created as a publicly available pretrial risk assessment instrument. The PSA is now used statewide in Arizona, Kentucky, Utah, Rhode Island, and New Jersey, as well as in major cities (e.g., Chicago, Houston, Phoenix, and Los Angeles) and smaller jurisdictions across the United States (Laura and John Arnold Foundation, 2020). Many other pretrial risk assessment instruments have been developed locally without being implemented as widely (Myburgh et al., 2015).

Pretrial risk assessment instruments are primarily comprised of risk factors (i.e., characteristics of a defendant or their environment that may increase likelihood of pretrial failure). Pretrial risk assessment instruments also may include protective factors (i.e., characteristics that may mitigate the likelihood of pretrial failure; Monahan & Skeem, 2016). Both risk and protective factors may be static (i.e., historical or otherwise unchangeable) or dynamic (i.e., able to change; Douglas & Skeem, 2005). Pretrial risk assessment instruments typically use an actuarial approach; that is, numeric item ratings are weighted and combined into a total score that is cross-referenced with a table describing outcome rates or probabilities. In practice, pretrial risk assessments should be used to inform, but not replace judicial discretion (Desmarais & Lowder, 2019). *Wisconsin v. Eric Loomis* (2016) asserts that scores produced by risk assessment instruments may not be the determinative factor in decisions of release.

PREDICTIVE VALIDITY OF PRETRIAL RISK ASSESSMENTS

The landscape of pretrial risk assessment practice and policy has changed dramatically in recent years, as jurisdictions across the nation have ramped up their pretrial reform efforts (Pretrial Justice Institute, 2019b). Although other fields have experienced long delays from research and development of new technologies to adoption in practice (Green et al., 2009), the implementation of pretrial risk assessment instruments has outpaced the peer-reviewed research in many ways. Peer-reviewed studies examining predictive validity in the context of pretrial risk assessment are relatively few and far between (Desmarais & Lowder, 2019). Instead, the body of research largely reflects the findings of local validation efforts disseminated in reports to government, nonprofits, and other outlets, including mass media. Over the past 10 years, a few systematic reviews and meta-analyses have summarized this work and set the stage for the current review (Bechtel et al., 2011, 2017; Mamalian, 2011; Myburgh et al., 2015). In the sections that follow, we briefly review the findings of the two meta-analyses, specifically, highlighting the aspects of these efforts upon which we seek to expand.

The first meta-analysis was conducted about a decade ago (Bechtel et al., 2011) and examined associations of risk factors and risk assessments with pretrial outcomes, including failure to appear, re-arrest, new crime, and any pretrial failure (i.e., one or more of the individual measures), across 13 studies. In general, the strength of association of individual factors with pretrial outcomes was low, but tended to be higher for static than dynamic factors. Global estimated effect sizes showed that pretrial risk assessments, in aggregate, demonstrated correlations moderate in size with re-arrest, failure to appear, and any pretrial failure, but not with new crime. Predictive validity was not reported by instrument. More recently, Bechtel and colleagues (2017) examined correlations between pretrial risk scores and failure to appear, pretrial arrest, or a composite measure of pretrial failure across 16 studies of various instruments. Analyses revealed a mean effect size representing “fair” validity for failure to appear and “good” validity in relation to re-arrest and any pretrial failure. Although the authors reported and compared effect sizes for each study, information on predictive validity by instrument was not provided.

The findings of the two meta-analytic investigations advanced the scientific literature addressing the degree to which pretrial risk assessments, and their components, are associated with pretrial outcomes; however, gaps remain that merit an updated review and synthesis of the research. First, as noted earlier, the pretrial risk assessment landscape has changed in recent years, including the implementation of the PSA in dozens of jurisdictions across the United States (VanNostrand & Lowenkamp, 2013). At the time of Bechtel and colleagues’ most recent meta-analysis, only one study examining the predictive validity of assessments completed using the PSA was available for analysis. Second, these meta-analyses reported results in aggregate across instruments and defendants to present an overall state of the science, rather than examining and comparing predictive validity by instrument or within specific subgroups. As such, our understanding of the empirical evidence regarding predictive validity of pretrial risk assessments for individual instruments and within subgroups of defendants, most notably those defined by race/ethnicity and gender, is limited. Third, the authors examined correlations, which measure the direction and strength of association, but are greatly constrained by base rates.[1] Fourth, prior reviews have not examined validity of pretrial risk assessments in predicting new violent crime, which is arguably the most relevant outcome for informing pretrial decisions. In fact, statutes and guidelines generally emphasize public safety over flight risk and interpret threat to public safety narrowly as reflecting new violent criminal activity (as opposed to any new criminal activity; American Bar Association, 2007; *United States v. Salerno*, 1987). For these four reasons, a re-examination of the empirical evidence is due.

THE CURRENT REVIEW

The aims of the current review are to: (a) describe pretrial risk assessment instruments used in jurisdictions across the United States; (b) summarize the characteristics of studies that have examined their predictive validity; and (c) synthesize findings regarding their validity in predicting failure to appear in court, new criminal activity, new violent criminal activity, and technical violations during the pretrial period overall and for subgroups of defendants. Our goal is to provide a summary of the empirical evidence regarding predictive validity that not only adds to the scientific literature, but also informs decisions regarding the selection and implementation of a pretrial risk assessment instrument, if any, to support pretrial reform efforts.

METHOD

REVIEW PROTOCOL

We followed the preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) statement (Moher et al., 2015), a 17-item checklist to promote transparent and consistent reporting of our systematic review protocol and findings.

SEARCH STRATEGY

Identification of Pretrial Risk Assessment Instruments

We identified risk assessment instruments used to predict the likelihood of failure to appear in court, new crime involvement, new violent crime involvement, and technical violations during the pretrial period. We searched PsycINFO, PsycArticles, Web of Science, National Criminal Justice Reference Service Abstracts, ProQuest Dissertation & Theses electronic databases, Google Scholar, and Google using all possible combinations of the following keywords: “pretrial,” “risk assessment,” “risk assessment instrument,” “risk assessment tool,” “failure to appear,” “crime,” “criminal activity,” “bail,” “bond,” and “defendant.” We also examined the references of papers, reports, reviews, and online resources that we identified through our database searches (Bechtel et al., 2017, 2011; Bureau of Justice Assistance, n.d.; Mamalian, 2011; Myburgh et al., 2015; Pretrial Justice Institute, 2019a; Summers & Willis, 2010) and consulted with experts in the field. We limited our search to instruments whose source material (e.g., manual, calibration study) had been produced by December 31, 2018. We included instruments if they were: (a) designed to predict outcomes during the pretrial period (i.e., failure to appear, new criminal activity, new violent criminal activity, and/or technical violation) and (b) used in multiple jurisdictions in the United States. We identified 10 pretrial risk assessment instruments that met inclusion criteria:

1. Colorado Pretrial Assessment Tool (CPAT; Pretrial Justice Institute, 2012);
2. Connecticut Risk Assessment for Pretrial Decision-Making (Connecticut Decision Aid; Hedlund et al., 2005);
3. Correctional Offender Management Profile for Alternative Sanctions–Pretrial Release Risk Scale (COMPAS-PRRS; Dieterich, 2010);
4. Florida Pretrial Risk Assessment Instrument (FPRAI; Austin et al., 2011);
5. Indiana Risk Assessment System–Pretrial Assessment Tool (IRAS-PAT; Latessa et al., 2013);
6. ORAS-PAT (Latessa et al., 2009);
7. PSA (VanNostrand & Lowenkamp, 2013);
8. U.S. Federal PTRA (Lowenkamp & Whetzel, 2009);
9. Vera Point Scale (Ares et al., 1963); and
10. VPRAI (VanNostrand, 2003).

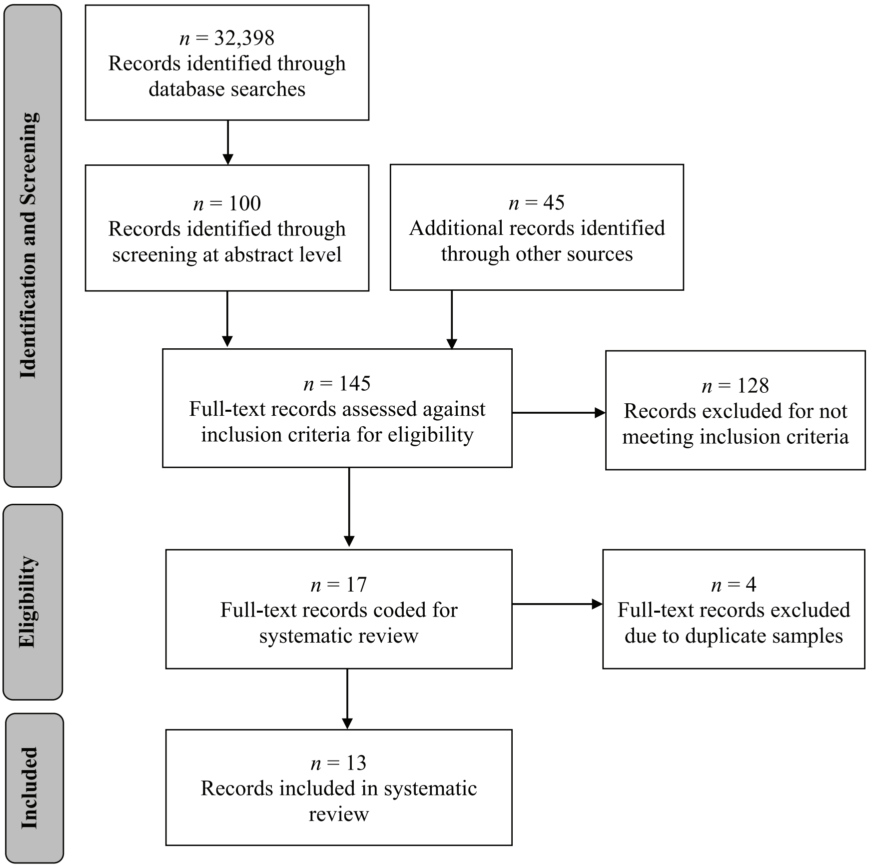

Figure 1: Results of Systematic Literature Search for Predictive Validity Studies

Identification of Predictive Validity Studies

Our next step was to identify studies that examined the predictive validity of these 10 instruments. We used the same search databases and sources listed above, this time using the full names and acronyms of the instruments as keywords. Again, we limited our search to studies that had been conducted by December 31, 2018. We included studies if they: (a) examined validity in predicting pretrial outcomes (i.e., failure to appear, new criminal activity, new violent criminal activity, and/or technical violation); (b) were reported in peer-reviewed journals, dissertations, theses, conference presentations, government or other reports available online, or book chapters; and (c) were written in English or a reliable translation was available. In the case of overlapping samples, we included unique analyses once for each sample. If samples overlapped and the same analytic technique was applied, we included the predictive validity estimate from the sample with the most participants.

We conducted our search in January, 2019, which returned an initial total of 1,998,701 hits. We reviewed 32,398 records[2] over the next several months to arrive at a final count of 11 studies across 13 publications[3] (marked with an asterisk in the reference list), including three journal articles and 10 government/technical reports (see Figure 1). Predictive validity was examined in one study each for COMPAS-PRRS, CPAT, and ORAS-PAT assessments and two or more studies each for PSA, PTRA, and VPRAI assessments. Across studies, data were collected between 1998 and 2016. The unit of analysis was at the person level for five studies, at the case level for four, and unclear for two.[4] Our search did not identify any studies of predictive validity that met our inclusion criteria and reported the necessary information for four instruments: the Connecticut Decision Aid, FPRAI, IRAS-PAT, and Vera Point Scale. These four instruments are excluded from the remainder of this review.

Data Extraction

Two of members of the research team extracted the following information from each study using a standardized data extraction protocol developed for this project (available upon request): (a) demographics of the study samples (e.g., sample size, gender, race/ethnicity, age) study design characteristics (e.g., study source, tool authorship, duplicate sample, research or practice context, length of follow-up), and the assessment process (e.g., setting, format, assessor, sources of information used to fill out assessment); (b) characteristics of the risk assessment instruments (e.g., assessment approach, number of items, types of items, and predicted outcome); and (c) predictive validity estimates or the information needed to calculate them (e.g., frequencies, failure rates) overall and as a function of gender and race/ethnicity, when reported. Table 1 provides full details on all variables for which data were extracted. There was an acceptable level of interrater reliability ($\kappa = .79$) for all studies coded by two researchers. We settled coding disagreements through discussion with the research team.

We followed the Risk Assessment Guidelines for Evaluation of Efficacy (RAGEE; Singh et al., 2015), a 50-item checklist detailing the information that should be included in papers reporting on predictive validity of risk assessments, to describe study quality. We elected to use the RAGEE over other measures for two reasons. First, the RAGEE was developed and validated through a Delphi process and reflects expert consensus regarding information that should be reported in papers describing research on the predictive validity of risk assessments (Singh et al., 2015). Second, recent methodological discussions have emphasized reporting study design features that may affect the interpretation of findings rather than reporting a particular scale value or rating (Hohn et al., 2019; Widman et al., 2020).

Data Analysis

We summarized sample, study design, and instrument characteristics using measures of central tendency, when appropriate. We then summarized the item type and content of the pretrial risk assessment instruments included in our review. Next, we extracted frequencies and predictive validity estimates for total scores and risk levels across instruments, where possible. Predictive validity was assessed for failure to appear, new criminal activity, new violent criminal activity, and technical violations, as well as combinations of these outcomes. Extracted estimates included the area under the receiver operating characteristic curve (AUC), correlation coefficient (r), the odds ratio (OR), and the Dispersion Index for Risk (DIF-R), which were the indices most commonly reported. Briefly, AUC represents the probability that a randomly selected defendant who engaged in one of the pretrial outcomes would have received a higher risk rating than a randomly selected defendant who experienced pretrial success; r represents the direction and strength of association between risk rating and pretrial failure; OR represents the ratio of the odds of a lower risk rating in those who experienced pretrial success to the odds of a higher risk rating in those who experienced pretrial failure; and DIF-R represents the extent to which risk classifications produce reasonably sized groups of defendants with maximally different rates of pretrial failure (Silver et al., 2000; Singh, 2013).

We first examined predictive validity of total scores and risk levels by calculating, when necessary, and comparing AUC values. The AUC is a suitable metric because its values are not influenced by base rates, which are anticipated to differ across groups and studies, to the same degree as other effect sizes (e.g., correlations; Smith, 1996). Moreover, AUCs are the most commonly reported effect size measure both in studies included herein and in risk assessment research more generally (Singh et al., 2013). For three studies, AUC values were not reported, but could be calculated from correlations and chi-square values (see Rice & Harris, 2005). We imputed missing standard error values from confidence intervals and *p* values (Higgins & Green, 2011). Although DIF-R has gained in popularity as a measure of predictive validity, it was only reported in two studies we identified. Furthermore, there are no formulae for estimating its standard errors and its values are dependent upon base rates and sample size. For these reasons, we excluded DIF-R from further review.

We also examined predictive validity of the risk levels via proportional ORs, which represents a strategy for pooling and comparing heterogeneous ORs across studies of multiple tests that used different cut-offs for response categories (Siadaty & Shu, 2004). Briefly, the proportional ORs assume that the effect associated with each dichotomization of an ordered categorical variable, such as risk levels, is the same; for example, that the increased likelihood of pretrial failure from low to moderate/high risk is the same as from low/moderate to high risk. First, we calculated ORs based upon the risk level, outcome frequencies, and sample size data extracted from each study. Second, we collapsed the number of risk levels to three (Singh et al., 2011).[5] Third, we ran two logistic regressions for each instrument and pretrial outcome to examine the assumption of proportionality. In the first logistic regression we compared risk level 1 (i.e., "low" risk) to Levels 2 and 3 (i.e., "moderate" and "high" risk). In the second logistic regression we compared risk Levels 1 and 2 (i.e., "low" and "moderate" risk) versus Level 3 (i.e., high "risk"). Visual examination and statistical comparison of meta-analyzed ORs for each study and pretrial outcome indicated that the assumption of proportionality was met in all cases. For the fourth and final step, we calculated proportional odds models examining the association between the 3-level risk variables for each instrument and each pretrial outcome. (Data and results available upon request.)

When these extraction and computational steps resulted in multiple AUCs or proportional odds for an outcome, instrument, or subgroup of interest, we computed an estimate of the average effect size using random effects model (Riley et al., 2011). When multiple effect sizes were reported in a study, we averaged the effects at the study level prior to meta-analysis to remove bias from correlated outcomes (Cooper, 1998). Due to the small number of studies, we do not statistically compare AUCs or proportional odds across outcomes, instruments, or subgroups. Instead, we describe effect sizes in relation to their practical significance, or more accurately, interpretation of the strength of association between the pretrial risk assessments and outcomes. Anchored to J. Cohen's d (1988), AUC values <.55 were considered poor, .55–.63 fair, .64–.71 good, and .71–1.00 excellent (Rice & Harris, 2005). Based upon the calculations of Chen and colleagues (2010), proportional ORs <1.50 were considered poor, 1.50–2.99 fair, 3.00–4.99 good, and 5.00 or greater excellent.

RESULTS

CHARACTERISTICS AND CONTENT OF INSTRUMENTS

Most instruments were relatively short, ranging in length from seven items in the ORAS-PAT to 14 items in the COMPAS-PRRS, with an average of 10.17 (SD = 2.64) items per instrument. The type and content of items included in the six pretrial risk assessment instruments are summarized in Supplemental Table S1 (available in the online version of this article). All instruments were comprised of risk factors, both static and dynamic, to the exclusion of protective factors. All instruments included items that represented at least three of four possible content domains (i.e., demographic, personal/social, criminal justice, and clinical). The COMPAS-PRRS and the PTRA included items across all content domains. All instruments included items assessing some form of criminal history, though the operational definitions varied. The CPAT was the only instrument that did not include an item assessing history of failure to appear. Four of the six instruments included items that queried employment status; such items were not included in the CPAT or the PSA. In fact, the PSA did not include items assessing personal/social or clinical characteristics. The CPAT, ORAS-PAT, and VPRAI did not include items assessing demographic characteristics.

STUDY AND SAMPLE CHARACTERISTICS

Aggregate summaries of the characteristics of the studies and samples included in our review can be found in Table 1. Further information on each study is available in Supplemental Table S2 (available in the online version of this article). Pretrial risk assessments included in our review were completed by professionals, such as pretrial service staff, in five studies and by research staff in two studies; in the other four studies it was not clear who had completed the assessments. Information used to complete the assessments was typically obtained via a combination of review of official or administrative records, interviews with defendant, and interviews with others. Most studies were conducted by the authors of the pretrial risk assessment instrument under investigation. For four instruments—COMPAS-PRRS, CPAT, ORAS-PAT, and PTRA—all studies were completed by an author of the instrument. Only two instruments—PSA and VPRAI—were examined in studies completed by researchers who did not create the instrument; however, in both cases, the developers of the instrument commissioned the research. In just over one third of the studies, assessments were completed as part of routine practice following implementation; in the other studies, assessments were completed for the purpose of research (see Table 1).

Three studies used a prospective design, while six studies used a retrospective design; the research design was not clearly described in the other two. The average length of follow-up was reported in two studies (M = 11.50 months, SD = 0.71). All studies included failure to appear and new arrests during the pretrial period as outcome measures; two studies examined new arrest for a violent crime and five studies included technical violations. The types of offenses included in the operational definition of new arrest included felonies or misdemeanors in four studies; two studies also included traffic offenses. The study of COMPAS-PRRS assessments (Dieterich, 2010) focused on felony offenses, specifically. Ten studies reported frequencies and eight studies reported AUC values, which were the two most frequently reported statistics (see Table 1). Only one type of statistic was reported for the COMPAS-PRRS and CPAT assessments: AUC values and frequencies, respectively. Studies of the other four instruments reported multiple different statistics. On average, studies

TABLE 1: Summary of Assessment, Study, and Sample Characteristics

Characteristic	Group	Frequency
Assessment process		
Risk assessor	Researcher	2
	Professional	5
	Not reported/unclear	4
Source of information	Official records	5
	Interview with defendant	5
	Interview with others	1
	Not reported/unclear	4
Assessment format	Assessor-completed, computer format	4
	Self-administered, format not specified	1
	Assessor-administered, format not specified	5
	Not reported/unclear	2
Study context	Research	7
	Practice/implementation	4
Study characteristics		
Publication type[a]	Journal article	3
	Government/technical report	10
Tool authorship	Author of tool was study author	9
	Study author was not author of tool	2
Study jurisdiction	Single site/jurisdiction	2
	Multisite/jurisdiction	9
Duplicate or overlapping sample	Yes	6
	No	5
Temporal design	Prospective	3
	Retrospective	6
	Not reported/unclear	2
Length of follow-up in months (k = 2)	*M* (*SD*)	11.50 (0.71)
Predictor(s) tested	Individual item scores	2
	Total scores	7
	Subscale scores	1
	Risk levels	8
	Not reported/unclear	1
Pretrial failure outcome measured	Failure to appear	11
	New criminal activity	11
	Violent offending	2
	Violation or breach of conditions	5
Source of outcome data	Official records	5
	Not reported/unclear	6
Statistics reported	Frequencies	10
	Correlations	4
	Chi square	2
	Odds ratios	3
	AUC	8
	DIF-R	2
RAGEE items reported (of 50 items)	*M* (*SD*)	25.00 (4.78)
Sample characteristics		
Sample characteristics calculated on	Full sample, no attrition	6
	Final sample, after attrition	2
	Not reported/unclear	3
Sample size at assessment (k = 11)	*M* (*SD*)	61,229.09 (79,563.05)
Male defendants (k = 9)	*M* (*SD*)	30,189.64 (39,676.04)
White defendants (k = 9)	*M* (*SD*)	28,581.60 (44,853.75)
Black defendants (k = 8)	*M* (*SD*)	10,660.40 (10,477.57)
Latinx defendants (k = 4)[b]	*M* (*SD*)	5,596.08 (9,687.37)
Other race defendants (k = 9)[c]	*M* (*SD*)	1,615.40 (2,266.47)
Age (in years) at assessment (k = 6)	*M* (*SD*)	33.22 (2.39)

Note. Some category counts add to more than 11 studies because some categories are not mutually exclusive. AUC = area under the receiver operating characteristic curve; DIF-*R* = Dispersion Index for Risk; RAGEE = Risk Assessment Guidelines for Evaluation of Efficacy; ORAS-PAT = Ohio Risk Assessment System–Pretrial Assessment Tool; VPRAI = Virginia Pretrial Risk Assessment Instrument.

[a]There are 13 publication types because we drew information from a peer-reviewed paper and a technical report that reported on the same study for the ORAS-PAT, and the results for one of the VPRAI studies was published across two separate technical reports. [b]In most studies, race/ethnicity were recorded in one variable; race and ethnicity were coded separately in just one study. [c]Includes Asian, Native American, Pacific Islander, unidentified races, and in one study, Black defendants.

reported information for 25 of the 50 RAGEE items, ranging from a low of 20 items for studies of PSA (VanNostrand & Lowenkamp, 2013) and PTRA (Lowenkamp & Whetzel, 2009) assessments to a high of 34 items for studies of ORAS-PAT (Latessa et al., 2009) and PTRA (Cohen & Lowenkamp, 2019) assessments.

Across studies, the average analytic sample size was 55,431 (*SD* = 73,376.63; range = 452–200,583). When reported, the mean defendant age was 33.22 years (*SD* = 2.39) and the majority of defendants (74.8%) were male. For studies in which information on race/ethnicity was reported, just over half of the assessments were completed on White defendants (59.1%) and close to one third on Black defendants (31.1%), with 15.4% completed on defendants identified as Latinx[6] (see Table 1). Outcomes were reported within subsamples for three instruments across four studies: PSAs by gender and race (DeMichele et al., 2018); PTRA by gender, race, and ethnicity (Cohen & Lowenkamp, 2019); and VPRAI assessments by gender, race, and income (Danner et al., 2015, 2016; VanNostrand, 2003).

PREDICTIVE VALIDITY

Table 2 presents the AUC values for total and subscale scores predicting pretrial outcomes. AUC values were reported or could be calculated for pretrial risk assessments completed using four instruments: the ORAS-PAT, PSA, PTRA, and VPRAI. Generally speaking, only one AUC value was available for each instrument and outcome, with the exception of the PTRA (see Table 2). Across instruments and outcomes, predictive validity was typically good, and sometimes excellent. AUC values ranged from .644 for VPRAI assessments predicting the combined measure of new criminal activity, failure to appear, and/or technical violations to .730 for PTRA predicting technical violations (see Table 2). PTRA demonstrated the highest AUC values across all outcomes.

AUC values for total and subscale scores within subgroups defined by gender and race/ethnicity were provided for PSA and PTRA assessments predicting new criminal activity and new violent criminal activity, as well as for PSAs predicting failure to appear. On the whole, predictive validity was comparable across subgroups; even when AUC values differed slightly between groups, they were still indicative of good predictive validity. For instance, AUC values for PSAs were slightly higher for men but still good for women in the prediction of new criminal activity (men = .653 vs. women = .637), whereas the opposite was true for PTRAs (men = .670 vs. women = .690). As another example, AUC values for PSAs were slightly higher for defendants of color but still good for White defendants in the prediction of new criminal activity (defendants of color = .659 vs. White defendants = .647), while the opposite was true for PTRAs (defendants of color = .663 vs. White defendants = .675). There were two instances in which the AUC values suggested differing levels of predictive validity, both involving PSAs: AUC values for new violent criminal activity and technical violations indicated good predictive validity for White defendants (AUCs = .666 and .655), but fair predictive validity for defendants of color (AUCs = .631 and .612). That said, the absolute differences were quite small: .035 and .039, respectively.

Table 3 presents the AUC values for risk levels predicting pretrial outcomes. AUC values were reported or could be calculated for risk assessments completed using five instruments: the COMPAS-PRRS, ORAS-PAT, PSA, PTRA, and VPRAI. With the exception of the VPRAI, only one AUC value was available for each instrument and outcome (see Table 3). Across instruments and outcomes, predictive validity varied from fair to excellent, with AUC values ranging from .620 for PSA risk levels predicting the combined measure of new

TABLE 2: Areas Under the Curve for Total and Subscale Scores Predicting Pretrial Outcomes

	Gender														
										Race/ethnicity					
	All defendants			Male defendants			Female defendants			White defendants			Defendants of color		
Outcome	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)
NCA															
All instruments	2	.664	[0.643, 0.685]	2	.665	[0.644, 0.686]	2	.675	[0.645, 0.705]	3	.671	[0.655, 0.687]	3	.662	[0.593, 0.731]
PSA	1	.650	—	1	.653	—	1	.637	—	1	.647	—	1	.659	—
PTRA	1	.680	[0.660, 0.690]	1	.670	[0.650, 0.700]	1	.690	[0.660, 0.710]	2	.675	[0.657, 0.693]	2	.663	[0.583, 0.743]
NVCA															
All instruments	2	.673	[0.649, 0.698]	2	.668	[0.634, 0.702]	2	.658	[0.616, 0.700]	3	.674	[0.644, 0.705]	3	.656	[0.571, 0.741]
PSA	1	.664	—	1	.654	—	1	.657	—	1	.666	—	1	.631	—
PTRA	1	.690	[0.660, 0.720]	1	.690	[0.640, 0.740]	1	.660	[0.590, 0.740]	2	.682	[0.639, 0.726]	2	.671	[0.563, 0.780]
FTA															
All instruments	2	.655	[0.639, 0.671]	—	—	—	—	—	—	—	—	—	—	—	—
PSA	1	.646	—	1	.642	—	1	.655	—	1	.655	—	1	.612	—
PTRA	1	.670	[0.650, 0.690]	—	—	—	—	—	—	—	—	—	—	—	—
TV															
All instruments	2	.710	[0.680, 0.739]	—	—	—	—	—	—	—	—	—	—	—	—
PTRA	1	.730	[0.720, 0.740]	—	—	—	—	—	—	—	—	—	—	—	—
VPRAI	1	.688	[0.674, 0.703]	—	—	—	—	—	—	—	—	—	—	—	—
NCA or FTA															
All instruments	5	.676	[0.659, 0.693]	—	—	—	—	—	—	—	—	—	—	—	—
ORAS-PAT	1	.675	—	—	—	—	—	—	—	—	—	—	—	—	—
PSA	1	.638	—	—	—	—	—	—	—	—	—	—	—	—	—
PTRA	3	.689	[0.684, 0.695]	—	—	—	—	—	—	—	—	—	—	—	—
NCA or FTA or TV															
All instruments	5	.690	[0.652, 0.727]	—	—	—	—	—	—	—	—	—	—	—	—
PTRA	3	.721	[0.712, 0.729]	—	—	—	—	—	—	—	—	—	—	—	—
VPRAI	2	.644	[0.599, 0.689]	—	—	—	—	—	—	—	—	—	—	—	—

Note. When multiple effect sizes were coded for a given predictor and outcome, the mean effect sizes were estimated using the random effects model—not reported and could not be calculated. *k* = number of effect sizes; AUC = area under the receiver operating characteristic curve; 95% CI = 95% confidence interval; NCA = New Criminal Activity; PSA = Public Safety Assessment; PTRA = Pretrial Risk Assessment Instrument; NVCA = New Violent Criminal Activity; FTA = Failure to Appear; TV = technical violation; ORAS-PAT = Ohio Risk Assessment System-Pretrial Assessment Tool; VPRAI = Virginia Pretrial Risk Assessment Instrument.

TABLE 3: Areas Under the Curve for Risk Levels Predicting Pretrial Outcomes

				Gender						Race/ethnicity						
	All defendants			Male defendants			Female defendants			White defendants			Defendants of color			
Outcome	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*	AUC	(95% CI)	*k*
NCA																
All instruments	2	.626	[0.617, 0.634]	—	—	—	—	—	—	—	—	—		—	—	—
PSA	1	.630	—	—	—	—	—	—	—	—	—	—		—	—	—
VPRAI	1	.621	—	—	—	—	—	—	—	—	—	—		—	—	—
FTA																
All instruments	2	.631	[0.619, 0.643]	—	—	—	—	—	—	—	—	—		—	—	—
PSA	1	.640	—	—	—	—	—	—	—	—	—	—		—	—	—
VPRAI	1	.622	—	—	—	—	—	—	—	—	—	—		—	—	—
TV																
VPRAI	1	.655	—	—	—	—	—	—	—	—	—	—		—	—	—
NCA or FTA																
All instruments	4	.669	[0.633, 0.705]	—	—	—	—	—	—	—	—	—		—	—	—
COMPAS-PRRS	1	.715	—	—	—	—	—	—	—	—	—	—		—	—	—
ORAS-PAT	1	.650	—	—	—	—	—	—	—	—	—	—		—	—	—
PSA	1	.620	—	—	—	—	—	—	—	—	—	—		—	—	—
PTRA	1	.692	[0.687, 0.696]	—	—	—	—	—	—	—	—	—		—	—	—
NCA or FTA or TV																
All instruments	4	.663	[0.621, 0.705]	—	—	—	—	—	—	—	—	—		—	—	—
PTRA	1	.726	[0.722, 0.729]	—	—	—	—	—	—	—	—	—		—	—	—
VPRAI	3	.642	[0.613, 0.671]	2	.657	[0.630, 0.686]	2	.665	[0.630, 0.700]	2	.678	[0.648, 0.709]		2	.637	[0.540, 0.735]

Note. When multiple effect sizes were coded for a given predictor and outcome, the mean effect sizes were estimated using the random effects model—not reported and could not be calculated. *k* = number of effect sizes; AUC = area under the receiver operating characteristic curve; 95% CI = 95% confidence interval; NCA = New Criminal Activity; PSA = Public Safety Assessment; VPRAI = Virginia Pretrial Risk Assessment Instrument; FTA = Failure to Appear; TV = technical violation; COMPAS = Correctional Offender Management Profiling for Alternative Sanctions; ORAS-PAT = Ohio Risk Assessment System-Pretrial Assessment Tool; PTRA = Pretrial Risk Assessment Instrument; NVCA = New Violent Criminal Activity.

criminal activity and/or failure to appear to .726 for PTRA risk levels predicting the combined measure of new criminal activity, failure to appear, and/or technical violations. PSA risk levels demonstrated the highest AUC values for new criminal activity and failure to appear, while COMPAS-PRRS risk levels demonstrated the highest AUC value for the combined measure of new criminal activity and/or failure to appear. No AUC values for risk levels were reported or could be calculated for new violent criminal activity.

AUC values for risk levels within subgroups defined by gender and race/ethnicity were available only for VPRAI assessments predicting the combined measure of new criminal activity, failure to appear, and/or technical violations. Although the AUC values differed slightly across subgroups, they indicated good predictive validity overall (see Table 3).

Table 4 presents the proportional ORs for risk levels predicting pretrial outcomes. Proportional ORs could be calculated for risk assessments completed using five instruments (i.e., CPAT, ORAS-PAT, PSA, PTRA, and VPRAI). There were one or two proportional ORs for each instrument and outcome (see Table 4). Predictive validity of the risk levels varied from fair to excellent, with proportional ORs ranging from 2.267 for VPRAI assessments predicting the combined measure of new criminal activity and/or failure to appear to 5.282 for CPAT assessments predicting failure to appear. CPAT assessments also demonstrated the highest proportional ORs for new criminal activity and the combined measure of new criminal activity and/or failure to appear, while PTRAs demonstrated the highest proportional ORs for new violent criminal activity and the combined measure of new criminal activity, failure to appear, and/or technical violations (see Table 4).

Proportional ORs could be calculated within subgroups defined by gender and race/ethnicity for PSA and PTRA risk levels predicting new criminal activity and new violent criminal activity, for PSA risk levels predicting failure to appear, and for VPRAI risk levels predicting any pretrial misconduct (defined both ways). More often than not, the proportional ORs suggested comparable predictive validity across subgroups, even if the exact values differed slightly. For example, the proportional ORs for VPRAI risk levels were slightly higher for men (2.602) than for women (2.526) in the prediction of the combined measure of new criminal activity and/or failure to appear and/or technical violation, whereas the reverse was found for PSA risk levels predicting new violent criminal activity (men = 2.744 vs. women = 2.938). As another example, the proportional ORs for PTRA risk levels were slightly higher for White defendants (2.888) than defendants of color (2.717) in the prediction of new criminal activity, but the absolute value of the difference is very small (i.e., 0.171).

There were a handful of instances in which the proportional ORs suggested differing levels of predictive validity as a function of gender and race/ethnicity. Comparing results for male and female defendants, there is no consistent pattern in the directionality of the differences. For example, PTRA risk levels demonstrated good validity in predicting new criminal activity among women (3.012) and only fair validity in predicting this outcome among men (2.880), but the trend was reversed for the prediction of new violent criminal activity (men = 3.090 vs. women = 2.379). In contrast, when differences were seen in the magnitude of proportional ORs for White defendants and defendants of color, they usually demonstrated greater predictive validity among White defendants. An exception to this pattern was seen for the PTRA risk levels predicting new violent criminal activity for which predictive validity was slightly greater among defendants of color (3.070) than White defendants (2.953). In most cases, the absolute differences in the proportional OR values for

TABLE 4: Proportional Odds Ratios for Risk Levels Predicting Pretrial Outcomes

				Gender						Race/ethnicity					
	All defendants			Male defendants			Female defendants			White defendants			Defendants of color		
Outcome	*k*	POR	(95% CI)	*k*	POR	(95% CI)	*k*	POR	(95% CI)	*k*	POR	(95% CI)	*k*	POR	(95% CI)
NCA															
All instruments	6	2.673	[2.379, 3.004]	2	2.744	[2.596, 2.900]	2	2.707	[2.222, 3.299]	3	3.100	[2.652, 3.623]	3	2.730	[2.593, 2.874]
CPAT	1	3.068	[2.240, 4.203]	—	—	—	—	—	—	—	—	—	—	—	—
PSA	2	2.459	[2.091, 2.891]	1	2.699	[2.604, 2.797]	1	2.461	[2.324, 2.606]	1	3.516	[3.400, 3.636]	1	2.722	[2.532, 2.927]
PTRA	1	2.967	[2.819, 3.122]	1	2.880	[2.616, 3.171]	1	3.012	[2.700, 3.361]	2	2.888	[2.678, 3.113]	2	2.717	[2.459, 3.003]
VPRAI	2	2.774	[1.719, 4.476]	—	—	—	—	—	—	—	—	—	—	—	—
NVCA															
All instruments	2	3.075	[2.860, 3.306]	2	2.795	[2.557, 3.056]	2	2.763	[2.289, 3.334]	3	3.023	[2.757, 3.314]	3	2.658	[2.355, 2.999]
PSA	1	2.964	[2.714, 3.236]	1	2.744	[2.490, 3.023]	1	2.938	[2.386, 3.617]	1	3.042	[2.743, 3.373]	1	2.305	[1.944, 2.733]
PTRA	1	3.323	[2.924, 3.776]	1	3.090	[2.464, 3.875]	1	2.379	[1.696, 3.337]	2	2.953	[2.414, 3.612]	2	3.070	[2.587, 3.644]
FTA															
All instruments	6	2.512	[2.308, 2.735]	—	—	—	—	—	—	—	—	—	—	—	—
CPAT	1	5.282	[3.701, 7.539]	—	—	—	—	—	—	—	—	—	—	—	—
PSA	2	2.459	[2.247, 2.690]	1	2.517	[2.449, 2.607]	1	2.702	[2.579, 2.831]	1	2.711	[2.633, 2.792]	1	1.751	[1.648, 1.860]
PTRA	1	2.587	[2.355, 2.841]	—	—	—	—	—	—	—	—	—	—	—	—
VPRAI	2	2.172	[1.880, 2.509]	—	—	—	—	—	—	—	—	—	—	—	—
NCA or FTA															
All instruments	6	2.920	[2.511, 3.396]	—	—	—	—	—	—	—	—	—	—	—	—
CPAT	1	4.333	[3.204, 5.859]	—	—	—	—	—	—	—	—	—	—	—	—
ORAS-PAT	1	3.831	[2.108, 6.962]	—	—	—	—	—	—	—	—	—	—	—	—
PSA	1	2.429	[2.360, 2.500]	—	—	—	—	—	—	—	—	—	—	—	—
PTRA	2	3.073	[2.831, 3.336]	—	—	—	—	—	—	—	—	—	—	—	—
VPRAI	1	2.267	[1.858, 2.767]	1	3.540	[2.862, 4.378]	1	3.751	[2.374, 5.926]	—	—	—	—	—	—
NCA or FTA or TV															
All instruments	4	3.240	[2.767, 3.794]	—	—	—	—	—	—	—	—	—	—	—	—
PTRA	2	3.910	[3.645, 4.195]	—	—	—	—	—	—	—	—	—	—	—	—
VPRAI	2	2.532	[2.253, 2.846]	1	2.602	[2.338, 2.895]	1	2.526	[2.102, 3.036]	1	2.898	[2.548, 3.297]	1	2.280	[1.994, 2.607]

Note. When multiple effect sizes were coded for a given predictor and outcome, the mean effect sizes were estimated using the random effects model—not reported and could not be calculated. k = number of effect sizes; POR = proportional odds ratio; 95% CI = 95% confidence interval; NCA = New Criminal Activity; CPAT = Colorado Pretrial Assessment Tool; PSA = Public Safety Assessment; PTRA = Pretrial Risk Assessment Instrument; VPRAI = Virginia Pretrial Risk Assessment Instrument; NVCA = New Violent Criminal Activity; FTA = Failure to Appear; TV = technical violation; ORAS-PAT = Ohio Risk Assessment System–Pretrial Assessment Tool.

defendants of color and White defendants were quite small. In three comparisons, however, the proportional ORs for PSA risk levels predicting failure to appear, new criminal activity, and new violent criminal activity were markedly higher for White defendants than defendants of color (see Table 4).

DISCUSSION

With bail reform sweeping the nation, many jurisdictions are looking to pretrial risk assessment instruments as one potential—and controversial—strategy to support their efforts. We conducted a systematic review of research examining the validity with which six pretrial risk assessment instruments used in multiple jurisdictions across the United States forecast key pretrial outcomes overall and by race/ethnicity and gender, when possible. In doing so, we provide the most up-to-date information on the predictive validity of pretrial risk assessments by instrument and by outcome, and the first synthesis and comparison of findings across studies regarding predictive validity for defendants of color and White defendants, as well as for men and women. We also present the first application of a risk assessment-specific checklist, the RAGEE, to examine the quality of studies included in systematic reviews of the predictive validity of risk assessments. In addition, we introduce a new method for pooling and comparing ORs across studies of risk assessment instruments that use different cutoffs and risk levels: proportional ORs. Although this analytic strategy has been used in other fields, we are not aware of prior reviews that have used it in the context of risk assessment research.

SUMMARY OF FINDINGS

Overall, findings of our review generally showed good to excellent predictive validity across outcomes for these six pretrial risk assessment instruments, with effect sizes in keeping with those seen in reviews of risk assessments used in other criminal justice contexts. Also consistent with those reviews, our findings did not reveal one pretrial risk instrument that produced assessments with the greatest predictive validity overall. Instead, our findings suggest that instruments produce pretrial risk assessments with somewhat greater predictive validity for some outcomes compared with others, but also that conclusions regarding the predictive validity of a given instrument may differ depending upon the predictors, outcomes, and effect sizes examined. For total scores, AUC values were highest for the PTRA. There was greater variability in the AUC values for the risk levels, with PSA risk levels producing slightly higher AUC values than VPRAI risk levels for new criminal activity and failure to appear, while COMPAS-PRRS risk levels produced the highest AUC values for the combined measure of new criminal activity and/or failure to appear and PTRA risk levels for the combined measure of new criminal activity, failure to appear, and/or technical violations. Looking at the proportional ORs, the CPAT risk levels performed especially well in predicting new criminal activity and failure to appear. PTRA risk levels demonstrated the highest proportional ORs for new violent criminal activity. PSA risk levels demonstrated good validity for new violent criminal activity, as well.

A primary aim of this review was to provide empirical evidence regarding the predictive validity of pretrial risk assessments across subgroups of defendants defined by race/ethnicity and gender. Only four studies of three instruments (i.e., the PSA, PTRA, and VPRAI) reported results for racial/ethnicity subgroups and they defined those groups in different ways. The comparisons we could make suggested comparable predictive validity of PSA, PTRA, and VPRAI assessments for defendants of color and White defendants in general, with a few important exceptions. We found lower levels of predictive validity among defendants of color than White defendants, particularly in the proportional ORs for PSA risk levels. Such findings not only raise potential concerns regarding the use of pretrial risk assessment instruments in practice, but also raise questions regarding the source and implications of these differences. Specifically, the pattern of results observed for the predictive validity of PSAs among defendants of color and White defendants may reflect the fact that PSAs are completed solely based upon record review and rely more heavily on criminal history and prior failures to appear in court compared with the other instruments. On one hand, removing the step of interviewing defendants eliminates barriers to implementation that renders the PSA appealing to many jurisdictions, such as the staffing, space, and time required to conduct the defendant interviews (VanNostrand & Lowenkamp, 2013). On the other hand, variables that reflect prior contact with the criminal justice system have come under particular scrutiny due to their susceptibility to biased policing practices and prosecutorial decisions, as well as recordkeeping errors (Eckhouse et al., 2019; Lowder et al., 2019; Mayson, 2019).

All PSAs examined herein were drawn from Kentucky and generalizability to other jurisdictions is unknown. PSAs conducted in a different jurisdiction with different input data may very well produce a different pattern of results. Beyond identifying whether racial bias exists, research is needed to understand why there are differences in predictive validity (Schmidt et al., 2020). Although the subject of much speculation, this issue has not received much empirical attention (Shepherd & Lewis-Fernandez, 2016). Furthermore, racial bias in pretrial decisions made without risk assessment results is well documented, as are the gains in accuracy produced by structured methods for predicting future violent and criminal behavior. Consequently, we may find greater racial bias in Kentucky pretrial decisions in the absence of the PSA or another pretrial risk assessment instrument (Desmarais, 2020). Indeed, a recent meta-analytic review found reductions in pretrial detention rates associated with the use of pretrial risk assessments for both Black and White defendants (Viljoen et al., 2019). In aggregate, findings to date do not provide sufficient empirical evidence to conclude one way or another whether pretrial risk assessments systematically demonstrate racial bias. Studies investigating predictive bias remains a high priority avenue for future research.

Only a few studies provided the data necessary to compare predictive validity as a function of gender and, when they did, there was no consistent pattern of findings. Most often, we saw slightly higher levels of predictive validity among assessments of women than men, including for PTRAs predicting new criminal activity, PSAs predicting new violent criminal activity and failure to appear, and VPRAI assessments predicting the combined measure of new criminal activity and/or failure to appear. However, predictive validity was slightly greater among men than women for PTRAs predicting new violent criminal activity and VPRAI assessments predicting the combined measure of new criminal activity, failure to appear, and/or technical violations. For most of these comparisons, the absolute value of the differences were so small as to suggest similar levels of predictive validity for men and women. In other words, based upon the findings of this review, we also do not find compelling evidence to suggest that pretrial risk assessments demonstrate predictive bias as a function of gender. At the same time, the number of women in U.S. jails continues to rise

(Kajstura, 2019) and some work suggests there may be benefits to examining gender-specific factors in pretrial assessment (Gehring & Van Voorhis, 2014). More research on the predictive validity of pretrial risk assessments as a function of gender is needed.

Whether pretrial risk assessments completed using the six instruments included in our review demonstrate consistency across assessors was not addressed in the extant literature. Yet, increasing consistency in pretrial decision-making is one of the goals of current reform efforts (National Task Force on Fines, Fees, and Bail Practices, 2019). As such, the lack of information on the interrater reliability represents a limitation that should be addressed in future evaluations of pretrial risk assessment instruments. Interrater reliability is relevant to any assessment that requires rating or coding items, even without an interview (Douglas et al., 2012), as reliability is a necessary criteria for validity (Douglas et al., 2011; Gottfredson & Moriarty, 2006). Even if a large proportion of the items are completed using official records, there may be differences in how that information is interpreted by assessors and/or errors that occur in the coding. The RAGEE statement (Singh et al., 2015) specifies that interrater reliability must be reported or an explanation for lack of reporting must be given.

The lack of information on interrater reliability was not the only way in which many of the reviewed studies fell short of standards for conducting and reporting risk assessment validation studies. Individual studies generally reported on about half to two thirds of the 50 items included in the RAGEE statement. For instance, the length of follow-up data collection at the person or case level was reported in only two studies; a couple of other studies reported a total period during which data were collected overall. Follow-up length or time at risk was not controlled for in any analyses. Moreover, none of the studies accounted for pretrial conditions that could have affected outcomes. As a final example, other than race/ethnicity and gender, sample characteristics (e.g., behavioral health, socioeconomic status) were rarely provided. At least one of the reports included in our review (DeMichele et al., 2018) has now been published in a peer-reviewed outlet (DeMichele et al., 2020). As more studies are published in peer-reviewed outlets, we may see greater adherence to accepted methodological practices, reporting guidelines, and testing standards (American Educational Research Association et al., 2014; Douglas et al., 2011; Singh et al., 2015).

Finally, many studies used arrests to operationalize the measurement of new (violent) criminal activity. Reliance on arrests—as both predictor and outcome—has been criticized in the debate surrounding the use of risk assessment instruments to inform pretrial decision-making (e.g., Barabas et al., 2019). People of color are more likely to be arrested for behaviors compared with White people (Kochel et al., 2011), yet arrests remain a common metric against which criminal justice decision-makers operationalize threat to public safety. Efforts are needed to move both the research and criminal justice system away from this focus on re-arrest to potentially less biased measures, such as convictions. Using convictions, however, is not a perfect solution due to bias in the winnowing process from arrest through prosecution, as well as in case disposition and sentencing (e.g., Kutateladze et al., 2014; Stolzenberg et al., 2013). Another strategy is to use filed charges, which could reduce false positives associated with arrest records and false negatives associated with convictions (Spohn & Holleran, 2002), but may still be subject to biases. One more strategy is to include self-report, which has been shown to increase accuracy in the measurement of violence and crime (Crisanti et al., 2003; Johnson et al., 2019), but is time consuming and may have limited feasibility. Given the strengths and limitations of different strategies, future research should examine predictive validity across multiple operational definitions of pretrial outcomes using data collected from multiple sources.

LIMITATIONS

The methodology of our review limits its findings in at least three ways. First, our intent was to present a comprehensive review of the validation research available on a sample of pretrial risk assessment instruments. For this reason, we used an inclusive selection strategy. One of our key findings is that studies examining the predictive validity of pretrial assessments often did not adhere to the methodological practices and reporting standards that are accepted in the field. As such, exclusion of studies that did not meet certain quality standards would likely have eliminated many sources from our review. Due to the small number of studies, we did not statistically compare results, but rather compared the effect sizes descriptively. Second, we strove to identify and include all studies examining the predictive validity of pretrial risk assessments that are commonly used in multiple jurisdictions. There may be studies that met inclusion criteria but were missed by our search strategy or that were produced after our review cutoff date. We welcome efforts that build upon our findings as new research becomes available. Third, we were only able to examine predictive validity findings as a function of race/ethnicity for four studies and our comparisons across subgroups were limited by the ways in which studies defined those groups, including grouping all people of color together. Given the widespread implementation of pretrial risk assessment instruments, continued evaluation of fairness and racial bias, specifically, is needed.

CONCLUSION

Despite these limitations, this review sheds light on the predictive validity of six widely used pretrial risk assessment instruments overall and across various subgroups of defendants and outcomes. Like many reviews of risk assessment instruments before this one, our findings do not identify one instrument that produces the most (or least) accurate pretrial risk assessments. To some extent, our findings suggest a tradeoff between efficiency and predictive validity: Pretrial risk assessments produced using instruments that comprised more items and that required defendant interviews often outperformed those produced using shorter instruments that could be completed solely based on records. Findings also suggest that pretrial risk assessments forecast outcomes with acceptable predictive validity. However, findings underscore the need for investigations of predictive validity across racial/ethnic subgroups and vis-à-vis new violent criminal activity. Beyond predictive validity, more research is needed on the impact of pretrial risk assessments, toward the reform goals of more equitable and less carceral pretrial decisions.

ORCID iDs

Sarah L. Desmarais https://orcid.org/0000-0003-4976-1276

Evan M. Lowder https://orcid.org/0000-0002-5855-2479

SUPPLEMENTAL MATERIAL

Supplemental Tables S1 and S2 are available in the online version of this article at http://journals.sagepub.com/home/cjb

NOTES

1. The magnitude of a correlation is greatly reduced for dichotomous variables that have a base rate of less than 50% and pretrial outcomes have base rates well below this level (see Babchishin & Helmus, 2016, for further discussion).

2. In keeping with practices used in prior reviews, we restricted our review of Google results to the first 300 hits of each search (see, for example, Bramer et al., 2017; Haddaway et al., 2015; Piasecki et al., 2018; Viljoen et al., 2019).

3. Results from the one study of ORAS-PAT assessments were published across a paper and a report and, similarly, results for one of the VPRAI studies were published across two separate technical reports.

4. Because it was not possible to identify and exclude repeated assessments from the data available to us, and given that all instruments included at least some dynamic factors that could change over time, we treat assessments themselves as the unit of analysis, acknowledging that we may have multiple assessments for one individual.

5. The PSA uses a binary flag for risk of new violent criminal activity; we used the raw scores to create a three-level variable for analysis.

6. In keeping with this journal's publication guidelines, we use the term Latinx here. However, studies included in our review reported on Hispanic ethnicity, specifically.

REFERENCES

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

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. *Counseling Psychologist*, *34*, 341–382. https://doi.org/10.1177/0011000005285875

American Bar Association. (2007). *Standards for criminal justice: Pretrial release* (3rd ed.). https://www.americanbar.org/groups/criminal_justice/publications/criminal_justice_section_archive/crimjust_standards_pretrialrelease_toc/

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ares, C., Rankin, A., & Sturz, H. (1963). The Manhattan bail project: An interim report on the pre-trial use of pre-trial parole. *New York University Law Review*, *38*, 67–95.

Austin, J., Bhati, A., Jones, M., & Ocker, R. (2011). *Florida pretrial risk assessment instrument*. The JFA Institute. https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=58add716-5e41-eba3-0a3d-f9298f7e1a54

Babchishin, K. M., & Helmus, L. M. (2016). The influence of base rates on correlations: An evaluation of proposed alternative effect sizes with real-world dichotomous data. *Behavior Research Methods*, *48*, 1021–1031. https://doi.org/10.3758/s13428-015-0627-7

Bansak, K. (2019). Can nonexperts really emulate statistical learning methods? A comment on "the accuracy, fairness, and limits of predicting recidivism." *Political Analysis*, *27*, 370–380. https://doi.org/10.1017/pan.2018.55

Barabas, C., Dinakar, K., & Doyle, C. (2019, July 17). Opinion: The problems with risk assessment tools. https://www.nytimes.com/2019/07/17/opinion/pretrial-ai.html

Bechtel, K., Holsinger, A. M., Lowenkamp, C. T., & Warren, M. J. (2017). A meta-analytic review of pretrial research: Risk assessment, bond type, and interventions. *American Journal of Criminal Justice*, *42*, 443–467. https://doi.org/10.1007/s12103-016-9367-1

Bechtel, K., Lowenkamp, C. T., & Holsinger, A. (2011). Identifying the predictors of pretrial failure: A meta-analysis. *Federal Probation*, *75*, 78–87.

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Systematic Reviews*, *6*(1), 245. https://doi.org/10.1186/s13643-017-0644-y

Bureau of Justice Assistance. (n.d). *Public safety risk assessment clearinghouse: Landscape*. https://psrac.bja.ojp.gov/selection/landscape

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. https://doi.org/10.1080/03610911003650383

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

*Cohen, T. H., & Lowenkamp, C. (2019). Revalidation of the Federal Pretrial Risk Assessment instrument (PTRA): Testing the PTRA for predictive biases. *Criminal Justice and Behavior*, *46*, 234–260. https://doi.org/10.1177/0093854818810315

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed). SAGE.

Crisanti, A., Laygo, R., & Junginger, J. (2003). A review of the validity of self-reported arrests among persons with mental illness. *Current Opinion in Psychiatry*, *16*, 565–569. https://doi.org/10.1097/00001504-200309000-00013

*Danner, M. J. E., VanNostrand, M., & Spruance, L. M. (2015, August). *Risk-based pretrial release recommendation and supervision guidelines: Exploring the effect on officer recommendations, judicial decision-making, and pretrial outcome*. Luminosity.

*Danner, M. J. E., VanNostrand, M., & Spruance, L. M. (2016, November). *Race and gender neutral pretrial risk assessment, release recommendations, and supervision: VPRAI and praxis revised*. Luminosity.

DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., & Comfort, M. (2020). Public Safety Assessment: Predictive utility and differential prediction by race in Kentucky. *Criminology & Public Policy*, *19*, 409–431. https://doi.org/10.1111/1745-9133.12481

*DeMichele, M., Baumgartner, P., Wenger, M., Barrick, K., Comfort, M., & Misra, S. (2018). *The public safety assessment: A re-validation and assessment of predictive utility and differential prediction by race and gender in Kentucky*. https://doi.org/10.2139/ssrn.3168452

Desmarais, S. L. (2020). The role of risk assessment in the criminal justice system: Moving beyond a return to the status quo. *Harvard Data Science Review*, *2*(1). https://hdsr.mitpress.mit.edu/pub/60jfy7hm

Desmarais, S. L., & Lowder, E. M. (2019, February). *Pretrial risk assessment tools: A primer for judges, prosecutors, and defense attorneys*. John D. and Catherine T. MacArthur Foundation. http://www.safetyandjusticechallenge.org/resource/pretrial-risk-assessment-tools-a-primer-for-judges-prosecutors-and-defense-attorneys/

Desmarais, S. L., & Zottola, S. A. (in press). Violence risk assessment: Current status and contemporary issues. *Marquette Law Review*.

*Dieterich, W. (2010). *Kent County Pretrial Services outcomes study: Developing and testing the COMPAS Pretrial Release Risk Scale*. Northpointe.

Douglas, K. S., Otto, R., Desmarais, S. L., & Borum, R. (2012). Clinical forensic psychology. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology* (Vol. 2, pp. 213–244). John Wiley.

Douglas, K. S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law*, *11*, 347–383. https://doi.org/10.1037/1076-8971.11.3.347

Douglas, K. S., Skeem, J. L., & Nicholson, E. (2011). Research methods in violence risk assessment. In B. Rosenfeld & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 325–346). John Wiley.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Eckhouse, L., Lum, K., Conti-Cook, C., & Ciccolini, J. (2019). Layers of bias: A unified approach for understanding problems with risk assessment. *Criminal Justice and Behavior*, *46*, 185–209. https://doi.org/10.1177/0093854818811379

Flores, A. W., Bechtel, K., & Lowenkamp, C. T. (2016). False positives, false negatives, and false analyses: A rejoinder to Machine Bias. *Federal Probation*, *80*, 38–46.

Gehring, K. S., & Van Voorhis, P. (2014). Needs and pretrial failure: Additional risk factors for female and male pretrial defendants. *Criminal Justice and Behavior*, *41*, 943–970. https://doi.org/10.1177/0093854814538022

Gottfredson, D. S., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency*, *52*, 178–200. https://doi.org/10.1177/0011128705281748

Green, L. W., Ottoson, J. M., Garcia, C., & Hiatt, R. A. (2009). Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annual Review of Public Health*, *30*, 151–174. https://doi.org/10.1146/annurev.publhealth.031308.100049

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE*, *10*(9), e0138237. https://doi.org/10.1371/journal.pone.0138237

Hedlund, J., Cox, S. M., Hines, M., Carollo, J., & Dwyer, L. (2005). *Development and validation of an assessment for pretrial conditional release*. Court Support Services Division, Judicial Branch, State of Connecticut.

Higgins, J. P. T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions version 5.1.0* (Updated March 2011). The Cochrane Collaboration. www.handbook.cochrane.org

Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2019). Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Frontiers in Psychology*, *9*, 2667. https://doi.org/10.3389/fpsyg.2018.02667

Holsinger, A. M., Lowenkamp, C. T., Latessa, E., Serin, R., Cohen, T. H., Robinson, C. R., . . . VanBenschoten, S. W. (2018). A rejoinder to Dressel and Farid: New study finds computer algorithm is more accurate than humans at predicting arrest and as good as a group of 20 lay experts. *Federal Probation*, *82*, 50–55.

Johnson, K. L., Desmarais, S. L., Tueller, S. J., & Van Dorn, R. A. (2019). Methodological limitations in the measurement and statistical modeling of violence in adults with mental illness. *International Journal of Methods in Psychiatric Research*, *28*, e1776. https://doi.org/10.1002/mpr.1776

Kajstura, A. (2019, October). *Women's mass incarceration: The whole pie 2019*. Prison Policy Initiative. https://www.prisonpolicy.org/reports/pie2019women.html

Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of suspect race on officers' arrest decisions. *Criminology*, *49*, 473–512. https://doi.org/10.1111/j.1745-9125.2011.00230.x

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, *52*, 514–551. https://doi.org/10.1111/1745-9125.12047

*Latessa, E. J., Lemke, R., Makarios, M., & Smith, P. (2010). The creation and validation of the Ohio Risk Assessment System (ORAS). *Federal Probation*, *74*, 16–22.

Latessa, E. J., Lovins, B., & Makarios, M. (2013, April). *Validation of the Indiana Risk Assessment System: Final report*. University of Cincinnati. https://www.in.gov/judiciary/iocs/files/prob-risk-iras-final.pdf

*Latessa, E. J., Smith, P., Lemke, R., Makarios, M., & Lowenkamp, C. (2009). *Creation and validation of the Ohio Risk Assessment System: Final report*. University of Cincinnati. https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/ORAS_Final_Report.pdf

Laura and John Arnold Foundation. (2020). *Map: PSA sites*. https://www.psapretrial.org/about

Lin, Z. J., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. *Science Advances*, *6*, eaaz0652. https://doi.org/10.1126/sciadv.aaz0652

*Lovins, B., & Lovins, L. (2016). *Validation of a pretrial risk assessment tool*. Correctional Consultants. https://www.crj.org/assets/2017/07/6_Riverside_Validation_Final_Report_5-3-16.pdf

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*, 210–233. https://doi.org/10.1177/0093854818789977

*Lowenkamp, C. T., & Whetzel, J. (2009). Federal pretrial risk assessment. *Federal Probation*, *76*, 33–36.

Mamalian, C. A. (2011, March). *State of the science of pretrial risk assessment*. Pretrial Justice Institute.

Mayson, S. G. (2019). Bias in, bias out. *Yale Law Journal*, *128*, 2218–2300. https://www.yalelawjournal.org/article/bias-in-bias-out

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., & Petticrew, M., . . . PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Review*, *4*, 1. https://doi.org/10.1186/2046-4053-4-1

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

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

National Conference on State Legislatures. (2018, April). *Trends in pretrial release: State legislation updated*. http://www.ncsl.org/portals/1/ImageLibrary/WebImages/Criminal%20Justice/pretrialEnactments_2017web_v02.pdf

National Task Force on Fines, Fees, and Bail Practices. (2019). *Bail reform: A practical guide based on research and experience*. National Center for State Courts. https://www.ncsc.org/~/media/Files/PDF/Topics/Fines%20and%20Fees/format%20revisions/Bail%20reform%20guide%203%2012%2019.ashx

Piasecki, J., Waligora, M., & Dranseika, V. (2018). Google search as an additional source in systematic reviews. *Science and Engineering Ethics*, *24*, 809–810. https://doi.org/10.1007/s11948-017-0010-4

*Pretrial Justice Institute. (2012). *The Colorado Pretrial Risk Assessment Tool* (CPAT)

Pretrial Justice Institute. (2019a). *Pretrial assessment: Specific tools & validation studies*. https://university.pretrial.org/libraryup/topics/assessment/assessment-studies

Pretrial Justice Institute. (2019b). *Scan of pretrial practices 2019*. https://university.pretrial.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=24bb2bc4-84ed-7324-929c-d0637db43c9a&forceDialog=0

Pretrial Justice Institute. (2020). *Updated position on pretrial risk assessment tools*. https://www.pretrial.org/wp-content/uploads/Risk-Statement-PJI-2020.pdf

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's d and r. *Law and Human Behavior*, *29*, 615–620. https://doi.org/10.1007/s10979-005-6832-7

Riley, R. D., Higgins, J. P., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *British Medical Journal*, *342*, d549. https://doi.org/10.1136/bmj.d549

Schmidt, S., Heffernan, R., & Ward, T. (2020). Why we cannot explain cross-cultural differences in risk assessment. *Aggression and Violent Behavior*, *50*. https://doi.org/10.1016/j.avb.2019.101346

Scurich, N., & Krauss, D. A. (2020). Public's views of risk assessment algorithms and pretrial decision making. *Psychology, Public Policy, and Law*, *26*, 1–9. https://doi.org/10.1037/law0000219

Shepherd, S. M., & Lewis-Fernandez, R. (2016). Forensic risk assessment and cultural diversity: Contemporary challenges and future directions. *Psychology, Public Policy, and Law*, *22*, 427–438. https://doi.org/10.1037/law0000102

Siadaty, M. S., & Shu, J. (2004). Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC Medical Research Methodology*, *4*, 27. https://doi.org/10.1186/1471-2288-4-27

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior*, *27*, 733–764. https://doi.org/10.1177/0093854800027006004

Singh, J. P. (2013). Predictive validity performance indicators in violence risk assessment: A methodological primer. *Behavioral Sciences & the Law*, *31*, 8–22. https://doig.org/10.1002/bsl.2052

Singh, J. P., Desmarais, S. L., & Van Dorn, R. A. (2013). Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law*, *31*, 55–73. https://doi.org/10.1002/bsl.2053

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*, 499–513. https://doi.org/10.1016/j.cpr.2010.11.009

Singh, J. P., Yang, S., & Mulvey, E. P. (2015). Reporting guidance for violence risk assessment predictive validity studies: The RAGEE Statement. *Law and Human Behavior*, *39*, 15–22. https://doi.org/10.1037/lhb0000090.supp

Smith, W. R. (1996). The effects of base rate and cutoff point choice on commonly used measures of association and accuracy in recidivism research. *Journal of Quantitative Criminology*, *12*, 83–111. https://doi.org/10.1007/BF02354472

Spohn, C., & Holleran, D. (2002). The effect of imprisonment on recidivism rates of felony offenders: A focus on drug offenders. *Criminology*, *40*, 329–358. https://doi.org/10.1111/j.1745-9125.2002.tb00959.x

Stolzenberg, L., D'Alessio, S. J., & Eitle, D. (2013). Race and cumulative discrimination in the prosecution of criminal defendants. *Race and Justice*, *3*, 275–299. https://doi.org/10.1177/2153368713500317

Summers, C., & Willis, T. (2010, October). *Pretrial Risk Assessment: Research Summary*. Bureau of Justice Assistance, U. S. Department of Justice. https://www.bja.gov/Publications/PretrialRiskAssessmentResearchSummary.pdf

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. https://doi.org/10.1126/science.185.4157.1124

United States v. Salerno, 481 U.S. 739, 755 (1987).

*VanNostrand, M. (2003). *Assessing risk among pretrial defendants in Virginia: The Virginia Pretrial Risk Assessment Instrument*.

*VanNostrand, M., & Lowenkamp, C. T. (2013). *Assessing pretrial risk without a defendant interview*. Arnold Foundation. https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/LJAF_Report_no-interview_FNL.pdf

*VanNostrand, M., Rose, K. J., & Weibrecth, K. (2011, March). *In pursuit of legal and evidence-based pretrial release recommendations and supervision*. Virginia Department of Criminal Justice Services. https://www.dcjs.virginia.gov/sites/dcjs.virginia.gov/files/publications/corrections/pursuit-legal-and-evidence-based-pretrial-release-recommendations-and-supervision.pdf

Viljoen, J. L., Jonnson, M. R., Cochrane, D. M., Vargen, L. 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*, 397–420. https://doi.org/10.1037/lhb0000344

Widman, L., Evans, R., & Choukas-Bradley, S. (2020). More meta-analytical explorations on adolescent sexual health interventions are needed—Reply. *JAMA Pediatrics*, *174*, 390–391. https://doi.org/10.1001/jamapediatrics.2019.5894

Wisconsin v. Eric Loomis, 881 N.W.2d 749 (Wis. 2016).

Sarah L. Desmarais, PhD is a professor in the Department of Psychology and director of the Center for Family and Community Engagement at North Carolina State University. Her research focuses on the assessment and treatment of risks and needs associated with criminal behavior, interpersonal violence, and terrorism.

Samantha A. Zottola, MA, is a PhD student in the Department of Psychology at North Carolina State University. Her research examines evidence-based practices during the pretrial period. Samantha will graduate in the summer of 2020 and begin work as a Postdoctoral Research Scholar for the Center for Family and Community Engagement at North Carolina State University.

Sarah E. Duhart Clarke, MA, is a PhD student in the Department of Psychology at North Carolina State University and a public health analyst at RTI International. Her recent work has focused on pretrial practices, indigent defense systems, and the opioid epidemic.

Evan M. Lowder, PhD, is an assistant professor in the Department of Criminology, Law and Society at George Mason University. Her research interests include strategies to reduce offending and improve behavioral health outcomes among justice-involved adults, with special focus on risk assessment and diversion programs.