

Criminogenic risk assessment: A meta-review and critical analysis

Punishment & Society

2021, Vol. 23(4) 578–604

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14624745211025751

journals.sagepub.com/home/pun**Seth J Prins  and Adam Reich**

Columbia University, USA

Abstract

A vast body of research underlies the ascendancy of criminogenic risk assessment, which was developed to predict recidivism. It is unclear, however, whether the empirical evidence supports its expansion across the criminal legal system. This meta-review thus attempts to answer the following questions: 1) How well does criminogenic risk assessment differentiate people who are at high risk of recidivism from those at low risk of recidivism? 2) How well do researchers' conclusions about (1) match the empirical evidence? 3) Does the empirical evidence support the theory, policy, and practice recommendations that researchers make based on their conclusions? A systematic literature search identified 39 meta-analyses and systematic reviews that met inclusion criteria. Findings from these meta-analyses and systematic reviews are summarized and synthesized, and their interpretations are critically assessed. We find that criminogenic risk assessment's predictive performance is based on inappropriate statistics, and that conclusions about the evidence are inconsistent and often overstated. Three thematic areas of inferential overreach are identified: contestable inferences from criminalization to criminality, from prediction to explanation, and from prediction to intervention. We conclude by exploring possible reasons for the mismatch between proponents' conclusions and the evidence, and discuss implications for policy and practice.

Keywords

criminal justice, criminogenic risk assessment, critical criminology, methodology, recidivism, theory

Corresponding author:

Seth J Prins, Columbia University, 722 W. 168th St. Room 521, New York, NY 10032, USA.

Email: seth.j.prins@columbia.edu

Introduction

Over the past 25 years, actuarial risk assessment of criminogenic risk factors has become an “evidence-based” policy and practice in the criminal legal system, strongly promoted within expert circles of policymakers, researchers, and practitioners (National Institute of Corrections, 2010).¹ *Criminogenic risk assessment* can be defined as (1) the use of statistical methods to predict an individual’s legal system outcomes and categorize them accordingly, purportedly to (2) manage carceral populations through efficient and effective allocation of supervision resources and, ideally, to reduce individuals’ risk through appropriate rehabilitative and social services.

The first part of this definition is about quantifying certain individual characteristics associated with, and often thought to be generative of, illegal behavior. Four of these individual characteristics (a history of antisocial behavior, antisocial personality pattern, antisocial attitudes and cognitions, and antisocial associates), have been consistently associated with recidivism, violence, and other legal system outcomes in almost any sample of people involved in the criminal legal system (Dowden and Andrews, 1999; Gendreau et al., 1996; Lipsey and Derzon, 1998). The second part of the definition is about intervening on manipulable aspects of these predictors such as attitudes, cognitions, elements of personality, and other “criminogenic” targets. Such efforts can modestly reduce recidivism rates (Andrews et al., 1990; Andrews and Dowden, 2006).

A vast body of research underlies the ascendancy of criminogenic risk assessment. As a result of its apparant success, it is moving from the back-end of the criminal legal system, where it was developed to assess the risk of recidivism, to the front-end of the system, in pre-trial processing, sentencing, and policing (Gottfredson and Moriarty, 2006; Lowenkamp and Whetzel, 2009; Storey et al., 2014; Trujillo and Ross, 2008).

The relative success of this approach to risk assessment has been interpreted as evidence that it taps into the causes of “criminal behavior” more generally, and that targeting these factors can therefore also reduce illegal behavior and correctional supervision rates overall. Indeed, an explanatory framework emerged around “the Big Four” antisocial criminogenic risk factors as fundamental to the roots of crime itself, and a model for organizing and applying this knowledge—the risk-need-responsivity model of correctional assessment and rehabilitative programming—is widely accepted and promoted (Andrews and Bonta, 2010; Bonta and Andrews, 2017; James, 2018; Serin and Lowenkamp, 2015).

Yet, with the field’s embrace and promotion of criminogenic risk assessment and the risk-need-responsivity model, its advocates make expansive claims about what it can achieve. Some proponents even argue that risk assessment should characterize the proper function of the criminal legal system itself. For example, Andrews and Bonta (2010) suggest that the prediction of illegal behavior is a central activity of the criminal legal system, because “from it stems community safety, prevention, treatment, ethics, and justice.” In addition to reducing

recidivism rates, proponents suggest that the framework might be able to improve sentencing procedures, facilitate jail diversion, reduce prison populations, help scale down mass incarceration without jeopardizing public safety, and ultimately, prevent crime altogether (Andrews et al., 2011; Clement et al., 2011; Monahan and Skeem, 2016).

The present meta-review interrogates the plausibility of such claims by attempting to answer the following questions:

1. How well does criminogenic risk assessment differentiate people who are at high risk of recidivism from those at low risk of recidivism?
2. How well do researchers' conclusions about (1) match the empirical evidence?
3. Does the empirical evidence support the theory, policy, and practice recommendations that researchers make based on their conclusions?

To date, scores of meta-analyses and systematic reviews have attempted to answer the first question, by synthesizing vast amounts of research on the predictive utility and validity of criminogenic risk factors and particular risk assessment instruments. These reviews typically conclude that the evidence supports the continued use and expansion of criminogenic risk assessment.² Concurrently, many critics have written about the scientific, cultural, and political forces that brought risk assessment to the forefront in the era of mass incarceration (e.g., Feeley and Simon, 1992; Garland, 2003), and on the ways in which risk may be gendered and racialized (Hannah-Moffat, 1999, 2004). However, these critiques have not always engaged directly with the empirical evidence thought to support criminogenic risk assessment, instead challenging the framework's premises outright.

This division of academic labor means that researchers who largely accept the premises of criminogenic risk assessment have tended to oversee empirical research, its translation to policy and practice, and assessments of its effectiveness. Critics, in turn, have tended to question or dismiss the entire endeavor without directly engaging the empirical evidence on which proponents base their claims. The present study bridges these worlds, approaching the empirical basis of criminogenic risk assessment from a theoretical perspective more skeptical than many of its current proponents.

Our purpose, in sum, is to evaluate whether what the field says about criminogenic risk assessment is consistent with what the evidence says about criminogenic risk assessment. We do this by conducting a meta-review of 39 meta-analyses and systematic reviews of the predictive performance of criminogenic risk factors, with a focus on history of antisocial behavior, antisocial attitudes and cognitions, antisocial personality, and antisocial peers. Our goal is to provide a bird's eye view of not only the empirical evidence surrounding criminogenic risk assessment, but also how the field understands and interprets that knowledge. This entails that we engage with the literature's quantitative data and methods, but also that we excavate its tacit theoretical and political assumptions.

A premise of our approach is that the way researchers mobilize concepts, language, and methods to make claims about evidence and practice can reveal hidden ontological and epistemological assumptions, and even contradictions. This is consequential if the widespread acceptance and expansion of criminogenic risk assessment is predicated on the misinterpretation or misuse of the concepts, terms, and methods associated with it. This, in turn, can have a real impact on people's lives, if scores generated from risk assessments restrict people's freedom or determine their access to health treatment or other services.

Moreover, we focus primarily on the empirical basis of criminogenic risk assessment, and the field's interpretation of it, rather than the merits of the risk-need-responsivity model, because the former is prerequisite for certain aspects of the latter. Indeed, the originators of the model acknowledge that criminogenic risk assessment was developed based on a "radical empirical approach to building theoretical understanding" (Andrews and Bonta, 2010: 132). Although they admit that this approach might be confused with "dustbowl empiricism" (Andrews and Bonta, 2010: 133), they argue that it nonetheless "lead[s] to a deeper theoretical appreciation of criminal conduct" and is "practically useful in decreasing the human and social costs of crime" (Andrews and Bonta, 2010: 133). Moreover, while the most recent iteration of the Risk-Need-Responsivity model de-emphasizes prior distinctions between risk factors based on the antisociality construct and others (Bonta and Andrews, 2017), the influence of this psychopathological conceptualization of crime and criminality—as something that emerges from within deviant or abnormal individuals, versus a social relation—looms large, as we shall see below. This meta-review analyzes, assesses, and critiques this logic.

Methods

To answer the three questions posed above, we conducted a systematic literature search and review to identify meta-analyses and systematic reviews that examined the predictive utility of criminogenic risk factors. (We will subsequently refer to the meta-analyses and systematic reviews as "reviews," while we will refer to the primary studies and data sources that constituted those reviews as "primary studies.") The details of our methods follow.

Inclusion criteria

Reviews were included if they were published in English language journals between 1990 and 2020, focused on a legal system outcome (e.g., recidivism or arrest), and focused on male subjects. We excluded studies of criminogenic risk assessment among women for several interrelated reasons. Sex does not appear to moderate associations between criminogenic risk factors and criminal legal system outcomes (Singh and Fazel, 2010). Yet, it was "...derived from statistical analyses of aggregate male correctional population data and...based on male-derived theories of crime" (Hannah-Moffat, 2009: 211), and thus while criminogenic risk assessment

may appear to be “gender neutral,” it may nonetheless fail to be *gender-responsive* (Hannah-Moffat, 2009, 2013). More recent efforts to incorporate gender-informed variables into the criminogenic risk framework, however, may merely reproduce gender-normative stereotypes and “neutralize gender politics and decontextualize women’s experiences” (Hannah-Moffat, 2010: 201). While these issues are critical, they are beyond the scope of the present review.

Search strategy

See the online supplement for search databases and terms. Search results were downloaded into a reference management system, de-duplicated, and titles and meta-data were screened to isolate meta-analyses and systematic reviews. Titles and abstracts of retained reviews were screened based on inclusion criteria to obtain a final sample.

Data extraction and analysis

Meta-data were compiled from the final sample of reviews. Citation information was obtained from Web of Science and Google Scholar. Select characteristics of reviews were tabulated. To answer the first question of this meta-review, we extracted and synthesized quantitative results and researchers’ conclusions and interpretations. To answer the second question, each author of the present meta-review independently rated review conclusions, to determine whether reviews deemed the evidence for the predictive utility of criminogenic risk assessment to be strong, moderate, or weak. Our inter-rater reliability, estimated with Cohen’s kappa, was 0.84, $p < 0.01$. Ratings reflect consensus scores reached after discussing disagreements. To answer the third question, we make claims based on a close reading of the reviews, from which we identify and examine recurring issues with the concepts, language, and methods mobilized by researchers in this body of work.

Results

Supplemental Figure 1 is a diagram of the flow of information through the meta-review process. The initial search yielded 12,952 records. Articles were retained if their titles or abstracts contained the terms *meta-analysis* or *review*. This reduced the number of records to 561. Titles and abstracts of these 561 reviews were read to determine whether they met inclusion criteria. The vast majority were excluded because they did not include a criminal legal system outcome. Thirty-nine meta-analyses or systematic reviews were retained for complete analysis.

Select review characteristics

Table 1 provides a description of retained reviews, and Supplemental Table 1 presents selected information from each, including disaggregated data from Table 1.

Table 1. Meta-description of included meta-analyses and systematic reviews.

Meta-description	N	%	Bibliometric analysis	Times cited
Studies included in meta-review	39		<u>Top 10 references within the reviews</u>	
Unique publications sources	25		Bonta et al. (1998)	10
Study type			Andrews et al. (1990)	8
Meta-analysis	26	65	Andrews and Bonta (2010)	8
Meta-regression	1	2.5	Gendreau et al. (1996)	8
Meta-review	1	2.5	Harris et al. (1993)	8
Systematic review	8	20	Andrews et al. (2004)	7
Narrative review	4	10	Gendreau et al. (1996)	7
Peer reviewed			Andrews et al. (1990)	6
Yes	36	92.3	Andrews and Dowden (2006)	6
No	3	7.7	Cohen (1988)	8
Year of publication			<u>Top 10 first authors cited in the reviews</u>	
1990–2000	7	17.5	Andrews DA	91
2001–2010	16	40	Bonta J	33
2011–2020	17	42.5	Gendreau P	31
Unique publication outlets	24		Hare RD	29
Top 3 publication outlets			Walters GD	21
Criminal Justice & Behavior	7		Douglas KS	14
Law and Human Behavior	4		Harris GT	13
Psychological Assessment	3		Cooke DJ	12
Top five most-cited reviews	3934	52.1*	Edens JF	12
Lipsey and Derzon (1998)	1553	20.6*	Dowden C	11
Gendreau et al. (1996)	827	10.9*		
Bonta et al. (1998)	602	8.0*		
Andrews and Dowden (2006)	585	7.7*		
Leistico et al. (2008)	367	4.9*		
Risk assessment instruments [†]				
Many	12	30.8		
Level of services inventory	4	10.3		
Psychopathy checklist	8	20.5		
Youth level of services Inventory	1	2.6		
Other	3	7.7		
Not reported	11	28.2		
Not applicable	3	7.7		
Sample characteristics				
Offenders	16	41.0		
Juvenile offenders	7	17.9		
Offenders and community	10	25.6		
Not reported	3	7.7		
Not applicable	3	7.7		
Outcome definition				
Any recidivism	13	33.3		
General recidivism	3	7.7		
Violent recidivism	3	7.7		
General/violent recidivism	4	10.3		

(continued)

Table 1. Continued.

Meta-description	N	%	Bibliometric analysis	Times cited
Any or violent recidivism	3	7.7		
Any re-arrest or re-conviction	3	7.7		
Violent or sexual reoffending	1	2.6		
Not reported	6	15.4		
Not applicable	3	7.7		

Note: Percentages are of the 39 studies included in this meta-review unless otherwise noted.

*Percentage of the 7553 total citations.

†Some studies counted in multiple categories, e.g., they reported the LSI and PCL.

Table 1 shows that the 39 reviews, two-thirds of which were meta-analyses, were published in 25 unique sources. *Criminal Justice and Behavior* and *Law and Human Behavior* published the most number of reviews (7 and 4 respectively). The vast majority of reviews were peer-reviewed (N=36, or 92.3%). Those that were not peer reviewed appeared in books or government-sponsored publications.

Collectively, reviews have been cited 7,553 times by other journals, according to Web of Science or Google Scholar. While the plurality of reviews has been cited between one and 20 times, 52.1% of the total citations can be attributed to five high-impact reviews. The plurality of reviews were published between 2011 and 2020.

Samples from primary studies in 84.5% of reviews were drawn from people who were involved with the criminal legal system (either adult or juvenile “offenders”). The outcome investigated by nearly all reviews was recidivism. However, definitions of this construct were heterogeneous: types of recidivism often were not distinguished (i.e., re-arrest, re-conviction, and technical violations were considered the same outcome), or a definition was not provided.

Supplemental Table 1 shows that primary studies from the reviews cover a half-century, from 1965–2020, and sample sizes (of combined participants from primary studies) ranged from roughly 2,400 to nearly 140,000 though many reviews did not report this information.

Thirty-three of the 39 meta-analyses and systematic reviews were available in the Web of Science database, which made it possible to conduct a bibliometric analysis of their complete reference lists. The results of this analysis are presented in the second column of Table 1, which shows the top 10 cited references and top 10 cited first authors. Andrews (91 citations) and Bonta (33 citations), the creators and owners of the Level of Services Inventory, and their students or frequent co-authors (e.g., Dowden, 11 citations and Gendreau, 31 citations) were among the top-cited authors and were authors of the top-cited references.

How well does criminogenic risk assessment differentiate people who are at high risk of recidivism from those at low risk of recidivism?

Table 2 presents meta-analytic effect size estimates and other predictive performance indicators from the sample of reviews for the four “antisocial” criminogenic risk factors for recidivism. Most reviews reported findings in terms of either weighted point-biserial correlation coefficients or Cohen’s *d* statistics, both of which were typically referred to as “effect sizes.”

For studies that reported correlation coefficients, the range of mean effect size estimates for history of antisocial behavior was 0.06 – 0.35, for antisocial attitudes 0.16 – 0.2, for antisocial personality 0.18 – 0.31, and for antisocial peers 0.18 – 0.27. The range of estimates for demographic characteristics such as sex, racialized group membership, and education/employment status was 0.05 – 0.26. The magnitude of point-biserial correlations are difficult to interpret because it depends on the coefficient itself and the prevalence of the outcome (an issue we will discuss below). However, a heuristic is that coefficients of 0.1, 0.3, and 0.5 are small, medium, and large, respectively (Rice and Harris, 2005). Thus, reviews tended to find small to medium effect sizes.

Also in Table 2, for studies that reported weighted mean Cohen’s *d*, the range of estimates for history of antisocial behavior was 0.32 – 0.57, for antisocial attitudes 0.23 – 0.51, for antisocial personality 0.42 – 0.6, and for antisocial peers 0.39 – 0.41. For demographic characteristics, the range was 0.16 – 0.44. Cohen’s *d* is easier to interpret, as it does not depend on the prevalence of the outcome. Cohen’s *d* can be interpreted as the proportion of a standard deviation difference between two groups. Cohen’s heuristic for small, medium, and large effects is 0.2, 0.5, and 0.8, respectively (Rice and Harris, 2005). Reviews reporting Cohen’s *d* thus tended to find small to medium effect sizes.

Other meta-analyses reported weighted mean estimates for particular instruments overall. Table 2 shows that the correlation coefficient effect size estimates for the Level of Services Inventory ranged from 0.06 – 0.6, and for the Psychopathy Checklist, 0.26 – 0.28. Factor 2 of the Psychopathy Checklist, which measures antisocial characteristics, anger/aggression, and impulsivity, had a stronger effect size (0.29 – 0.32) than Factor 1, which measures callous, unemotional, and narcissistic traits (0.15 – 0.18).

A small number of meta-analyses calculated the mean area under the Receiver Operating Characteristic curve (ROC-AUC). This statistic represents the probability that a randomly chosen individual who has recidivated would be ranked as having higher criminogenic risk than a randomly chosen individual who had not recidivated. Schwalbe (2007), calculated an ROC-AUC of 0.64 from a meta-analysis of 28 different risk assessment instrument validation studies. Whittington and colleagues (2013) found a mean ROC-AUC of 0.69 from 65 studies. In a meta-analysis of 23 samples using the Level of Services Inventory and the Psychopathy Checklist, Fazel and colleagues (2012) found a mean ROC-AUC for recidivism of 0.66, a sensitivity of 0.4 (the probability that someone was

Table 2. Meta-analytic effect sizes and other performance indicators for criminogenic risk factors and general recidivism.

Study	History of antisocial behavior	Antisocial attitudes	Antisocial personality	Antisocial peers	Demographics	LSI total	PCL total	PCL factor 1	PCL factor 2
Cohen's <i>d</i>									
Asscher et al. (2011)	0.32	0.37	0.42						
Bonta et al. (2014)	0.5	0.51	0.56		0.17 – 0.42				
Gardner et al. (2015)		0.23 – 0.31							
Gutierrez et al. (2013)	0.44	0.36	0.51	0.41	0.16 – 0.43				
Leistico et al. (2008)							0.55	0.38	0.6
Mokros et al. (2014)						0.29 – 0.76			
Wilson and Gutierrez (2013)	0.57	0.39	0.6	0.39					
Correlation coefficients									
Bonta et al. (1998)	0.08		0.07		0.12				
Desmarais, Johnson and Singh						0.24 – 0.36			
Campbell et al. (2009)									
Cottle et al. (2001)	5 instruments, 0.06 – 0.35	0.22 – 0.32		0.2	0.03 – 0.23				
Edens et al. (2007)						0.25	0.27	0.18	0.29
Gendreau et al. (1996)	0.22	0.16	0.19	0.27	0.06 – 0.18				
Gendreau et al. (1996)	0.18	0.18	0.18	0.18	0.05 – 0.16				
Lipsey and Derzon (1998)	0.09 – 0.27			0.04 – 0.43	0.09 – 0.26				
Olver et al. (2014)	0.28	0.19	0.31	0.22	0.12 – 0.24				
Olver et al. (2009)						0.29			
Pusch and Holtfreter (2018)						0.32	0.28		
Schwalbe (2008)						0.02 – 0.7			
						0.32 – 0.4			

(continued)

assessed as high-risk given that they recidivated), a specificity of 0.8 (the probability that someone was assessed as low-risk given that they did not recidivate), a positive predictive value of 0.52 (the probability that someone will recidivate given that they were assessed as high-risk), and a negative predictive value of 0.76 (the probability that someone will not recidivate given that they were assessed as low-risk).

Eighteen of the reviews, or roughly 46%, tested for heterogeneity in meta-analytic results as a function of study characteristics such as sample composition (male/female, white/racialized group), study design (cross-sectional, longitudinal), source of risk assessment coding (interview/files), publication status (published/unpublished), etc. In general, these reviews found moderate to high degrees of heterogeneity that were attributable to the above characteristics. Seven reviews, or roughly 18%, discussed the quality of their primary studies. Four of these considered study design to be a proxy for quality, and as a result two included only prospective, longitudinal designs (Bonta et al., 1998, 2014). Two assessed whether design moderated meta-analytic results. One of these found that design had no effect on results (Andrews and Dowden, 2006), and one found that prospective studies were more likely to obtain statistically significant results than cross-sectional studies (Whittington et al., 2013). One study found that coder-rated quality of the outcome variable was positively associated with effect size (Lipsey and Derzon, 1998). Eight reviews mentioned publication bias and 6 (15%) tested for it, and found that the likelihood of publication bias was low. This is consistent with Singh and Fazel's (2010) meta-review, which found that only a quarter of reviews assessed for publication bias, which likely biases results in favor of positive significant findings.

How well do conclusions about criminogenic risk assessment's performance match the empirical evidence?

Supplemental Table 2 paraphrases the primary conclusions of the reviews. Roughly 37% of the reviews concluded that evidence for predictive performance was strong, 37% concluded it was moderate, 13% concluded it was weak or that results should be interpreted cautiously, and 13% did not draw explicit conclusions.

Thus, while over a third of the reviews judged the predictive performance of criminogenic risk assessment to be weak to moderate, over a third of the reviews deemed it to be strong. All but one meta-analysis drew these conclusions based on point-biserial correlations, Cohen's *d*, or ROC-AUC. The vast majority relied on the former two statistics, which do not quantify predictive performance.

Measures of "effect" versus measures of prediction/classification. Most reviews used the language of "effect size" in describing point-biserial correlations or Cohen's *d*. This confuses and conflates the language and goals of causal inference with the language and goals of prediction. Moreover, there are a number of major, well-understood problems with the use of point-biserial correlations and Cohen's *d*

even as measures of *effect*, including their dependence on the marginal distribution of the independent variable, arbitrary features of study design, and sampling variability (e.g., Cumming, 2013, 2014; Greenland et al., 1986).

But one issue in particular warrants further examination: the point-biserial correlation coefficient depends on the prevalence of the outcome, which was frequently not reported in the reviews or the primary studies that constituted them. Of greater concern is that a large number of reviews made conversions among correlation coefficients, Cohen's d , and ROC-AUC, in order to implement meta-analytic procedures, using methods for this conversion that are sensitive to outcome prevalence. However, these reviews rarely reported the outcome prevalence estimates used in conversions or acknowledged that commonly cited tabular conversion charts assume an outcome prevalence of 50%. Using a 50% prevalence, or base rate, can overestimate the correlation coefficient if the true base rates are lower or higher. This is relevant because a study of nearly 68,000 people released from prisons in 2005, randomly sampled to represent the roughly 401,000 people released from prisons that year in 30 states, found that average recidivism rates are appreciably higher than 50% (Alper et al., 2018). The proportion of people who were re-arrested within three, six, and nine years of release was 68%, 79%, and 83% respectively (Alper et al., 2018).

Supplemental Figure 2 demonstrates the instability of point-biserial correlations converted from Cohen's d , as a function of outcome prevalence and the magnitude of d . This plot was developed using the standard conversion formula from Rice and Harris (2005). For various magnitudes of Cohen's d (curved lines), an outcome prevalence (x-axis) of 50% results in the maximum point-biserial r (y-axis). As outcome prevalence decreases or increases from 50%, the point-biserial r decreases. The potential for serious bias revealed in this figure—that the true magnitudes of correlations are likely lower than reported in the reviews—has been comprehensively discussed in the psychology literature (McGrath and Meyer, 2006).

Even if point-biserial correlation coefficients and Cohen's d were described and interpreted not as effects, but purely for prediction, they do not convey some important information relevant to answering the first, technical question of this meta-review, about how well criminogenic risk assessment differentiates people who are at high risk of recidivism from those at low risk of recidivism. Only one meta-analysis (Fazel et al., 2012) presented measures that provide this information: sensitivity, specificity, positive predictive value, and negative predictive value. This review found that criminogenic risk assessments were better at identifying people at low risk for recidivism than people at high risk for recidivism, i.e., negative predictive values were high. They argued, however, that positive predictive values were unacceptably low: only 52% of individuals judged to be moderate to high risk went on to commit any offense (virtually equivalent to flipping a coin).

Furthermore, one of the meta-analyses reviewed here found that the Receiver Operator Characteristic curve was defined incorrectly in 27.8% of studies, and the Area Under the Curve statistic was defined in only 34% of studies, and, when it

was defined, the definition was incorrect 37.5% percent of the time (Singh et al., 2013). Of greater concern, the estimated Area Under the Curve values were only interpreted in one-third of the studies, and was interpreted accurately in only 12.5% of these.

Thus, while empirical indicators provide relatively consistent magnitudes for the association between criminogenic risk factors and recidivism, the most commonly used statistics do not directly answer the first question regarding criminogenic risk assessment's ability to distinguish people at high vs. low risk of recidivism. And because the most common statistic—the point-biserial correlation coefficient—is unstable relative to outcome prevalence, even those measures were likely inflated: of the 17 reviews that presented correlation coefficients, only three explicitly stated that they collected information about outcome prevalence from their primary studies. Five others mentioned the issue of sensitivity to outcome prevalence, but did not state whether they had information on true base rates from primary studies or made assumptions about outcome prevalence. The one meta-analysis that reported positive and negative predictive values found that risk assessments were good at correctly identifying people at low risk of recidivism, but virtually no better than chance at identifying people at high risk of recidivism. The technical performance of criminogenic risk assessment has thus been interpreted inconsistently, and arguably inappropriately, by the framework's proponents.

Does the empirical evidence support the theory, policy, and practice recommendations that researchers make based on their conclusions?

In this section, we analyze how the reviews talk about risk assessment and illegal behavior more broadly, and assesses whether they make inferences that are supported by the data. Three themes are identified: contestable inferences from criminalization to criminality, contestable inferences from prediction to explanation, and contestable inferences from prediction to intervention.

Contestable inferences from criminalization to criminality. Reviews tended to conflate *exposure to the criminal legal system* with *illegal behavior*. This occurred with both the outcome (recidivism) and predictors (criminogenic risks). For the outcome, reviews tended to conflate the causes of re-arrest, re-conviction, or the revocation of probation or parole with the causes of recidivism resulting from new crimes. Indeed, 50% of reviews used heterogeneous definitions of recidivism or did not report a definition of recidivism. There are two broad categories of situation that can result in recidivism: new illegal offenses and technical violations of the terms of community supervision, e.g., missing an appointment with a parole officer. Most technical violations are not instances of illegal behavior (Council of State Governments Justice Center, 2019), and there is often great discretion among individual community corrections officers and agencies about which technical violations are pursued (Jones and Kerbs, 2007). Thus, incident illegal behavior is sufficient but not necessary for recidivism.

The heterogeneity of recidivism definitions reflects the heterogeneity among risk assessment instruments used to predict recidivism. In their review, Desmarais and colleagues (2016) found that of 19 risk assessment instruments validated in U.S. correctional settings, 31% of validation studies defined recidivism as a new arrest, 13% as re-conviction, 10% as reincarceration, and 4% as technical violations. Importantly, the definition of recidivism influences the predictive performance of risk assessment instruments. For example, the Level of Services Inventory was found to be a valid predictor of recidivism in roughly half as many studies when the definition was re-arrest versus reincarceration (Vose et al., 2008).

Only two of the meta-analyses and systematic reviews acknowledged the difference between exposure to the criminal legal system and illegal behavior. The remainder of the reviews took for granted that legal system outcomes were the result of agential behaviors that emerged from within deviant individuals (e.g., Bonta et al., 2014).

Recidivism can be the result of an individual's own behaviors, the proclivities of their supervision officer, or institutional policies and customs, and the causal mechanisms for recidivism are not uniform across these scenarios. For example, impulsivity may be one of many mechanisms for committing a new robbery, but family or employment problems may be the mechanism for missing a mandated treatment session. And the disposition of a community corrections officer might supersede both of these mechanisms in some circumstances.

As Schwalbe (2008) notes in his review, none of this is important if the goal of criminogenic risk assessment is purely prediction:

As statistical prediction devices, actuarial risk assessments do not assume an underlying causal process related to recidivism. Rather, they count risk factors irrespective of the specific factors that may or may not be present for an individual case. (pp. 1368–1369)

But for *explaining* crime or illegal behavior, and *reducing* risk, enumerating the correct mechanisms of recidivism is paramount.

An analogous problem arises with criminogenic predictor constructs, which also conflate illegal behavior with exposure to the criminal legal system. Only two reviews recognized the conceptual and empirical distance between illegal behavior and exposure to the criminal legal system, both within the context of racialized disparities. In the first, Wilson and Gutierrez (2013) compared the predictive ability of the Level of Services Inventory among Aboriginal versus non-Aboriginal “offenders” in Canada, and found effect modification of Aboriginal status and risk score: high-risk Aboriginals and non-Aboriginals had the same probability of recidivism, but low-risk Aboriginals had a higher probability of recidivism than low-risk non-Aboriginals. The authors characterized this finding as an “underclassification” of low-scoring Aboriginals. But a more critical interpretation is that low-risk Aboriginals were subject to a lower threshold of policing, arrest, and sentencing, i.e., they were victims of racialized discrimination. Similarly, in a

review of studies that compared risk assessments for ethnic minority and white offenders in the United Kingdom, Raynor and Lewis (2011) found that ethnic minorities consistently had significantly lower risk scores, but received the same sentences as higher-risk white offenders. The authors attributed this finding to racialized discrimination in the British criminal legal system.

Findings such as these reveal that because crime is viewed as emerging from within deviant or abnormal individuals, criminogenic risk assessments struggle to account for distortions in the purported “signal” of individual differences that are in fact due to socio-structural “noise.” In fact, whether or not a person will be re-arrested or re-convicted is influenced by factors that have nothing to do with their criminogenic risk profiles, such as the way the criminal legal system targets their racialized social position.

Indeed, criminogenic risk assessment avoids altogether basic questions about which behaviors are considered crimes and whether behaviors that are deemed criminal are treated differentially across time, space, and groups of people. Story (2016: 10) clarifies this difference between criminality and criminalization:

While criminality is understood to be a state of objective deviance located in the individual, to be criminalized is to be subjectified as well as subjugated by the coercions of law enforcement and the criminal justice system, both of which are highly malleable relative to changes in laws, policy, and institutional dictates. . . .

The point is not that criminogenic risk instruments may contain racialized, gendered, or other sorts of biases, but rather that, even if they do not, they may still perform unevenly across groups if they attempt to map onto individuals the discriminatory operations of the criminal legal system. Calibrating individual-level risk items for the sole purpose of reducing the uneven performance of risk assessments across racialized groups, as Wilson and Gutierrez (2013) suggest, without addressing structural and institutional sources of discrimination and disparities, thus becomes a normative rather than technical solution. While it might make risk assessments “perform better” in a predictive sense, such recalibration would likely serve to mask, and reproduce, the structural and institutional discrimination that caused the instrument’s underperformance in the first place.

Instead, most reviews implied that the question *Why do some people engage in illegal behavior more than others?* is the same as the question *Why does the criminal legal system target some people more than others?* This conflation was sometimes made rather consciously:

The risk principle of case classification relates not to the retributive or deterrent aspects of justice but to the objective of reduced reoffending through rehabilitative programs. Let justice be done and let the just penalty be set, the just obligations be established, and the just decisions be made. The risk principle of human service becomes relevant when, in that just context, interest extends to public protection through the delivery of human services. (Andrews and Dowden, 2006: 90)

In other words, advocates of criminogenic risk assessment take as a premise that the criminal legal system is just. If there are unjust distortions, they are not the concern of criminogenic risk assessment because they belong to the system as a whole. But if, in practice, risk assessment reflexively reinscribes systemic injustice under a guise of scientific objectivity, the intellectual and moral indifference implied by the above quotation becomes untenable.

Contestable inferences from prediction to explanation. The outcome in nearly all of the reviews was recidivism, and roughly 74% provided a definition of this outcome. However, many reached conclusions that were not restricted to recidivism, but also to crime or illegal behavior more broadly. As noted above and in Table 1, 58% of the reviews drew on primary studies that had samples made up exclusively of juvenile and adult “offenders.” Most of these discussed their theoretical orientation and findings in a way that strongly suggested their results tapped into the origins of crime or illegal behavior, and that predictors of recidivism might explain the onset and duration of illegal behavior. For example (emphases added):

GPCSL [General Personality and Cognitive Social Learning theory] proposes that **the causes of crime are to be found** within the individual and his/her social learning environment. (Bonta et al., 2014: 279)

The general findings of the current meta-analysis are consistent with broad social psychological perspectives of **criminal behavior**. (Bonta et al., 1998: 138)

The Big Four and Central Eight underpin a general personality and cognitive social learning theory of **criminal behavior** that provides an **explanatory model of the origin and continuation of criminal conduct**, and informs methods for predicting, **reducing**, managing, and **preventing** criminal behavior. (Olver et al., 2014: 157)

The LSI was developed from a general personality and social psychological perspective of crime (Andrews & Bonta, 2003), embodied in the Big Four **covariates of criminal conduct**—antisocial attitudes, antisocial associates, antisocial personality, and a history of antisocial behavior (the constellation is sometimes referred to as the Central Eight, with the inclusion of the needs areas leisure and recreation, family and marital, substance abuse, and employment and education). These covariates are linked to **the origin of criminal behavior** (and are hence called criminogenic needs), and services directed toward these areas of risk and need might reduce antisocial behavior. (Olver et al., 2009: 331)

These quotations show that many reviews motivated their analyses with a *theory of crime* or *theory of criminal behavior*, although reviews focused on studies of *recidivism*, in which individuals were already involved in the criminal legal system.

The problem with conflating the predictors, let alone causal explanations, for the onset of illegal behavior or exposure to the legal system with causal explanations for recidivism has long been recognized (e.g., *asymmetric causation*,

Uggen and Piliavin, 1998). Yet, few reviews dealt directly with the implications of generalizing from their legal system sampling frames to individuals not involved in the system, and thus made the extension from recidivism to “crime” or onset of illegal behavior without clear intention or justification. One exception is a thoughtful explanation in Cottle and colleagues (Cottle et al., 2001), regarding why their meta-analysis would focus only on recidivism and not initial offending:

It is not feasible to make meaningful assumptions about predictors of reoffending behavior based on predictors found to be associated with first-time delinquency... . [S]tudies examining recidivism risk factors typically are based on more homogeneous samples of adolescents already identified as delinquent. Therefore, variables significantly associated with reoffending behavior in juveniles are not necessarily useful in initially distinguishing between adolescents who will or will not become delinquents.

Nevertheless, slippage from what the evidence says about recidivism prediction to what research says about the onset, duration, and origins of illegal behavior appears in nearly half of the reviews analyzed here.

Contestable inferences from prediction to intervention. Even if criminogenic risk assessment correctly predicted recidivism, correct prediction does not imply effective intervention; this is true even if predictive risk factors are manipulable (Greenland, 2005; Hernán and VanderWeele, 2011; Pearl, 2014). Accurately predicting the effects of interventions is not possible without the identification of causal mechanisms (Schwartz et al., 2016). Yet, proponents of criminogenic risk assessment switch from talking about recidivism prediction to talking about recidivism reduction without directly engaging with causation—their emphasis on manipulable risk factors merely assumes it. Below is a sample of quotations that illustrate this question-begging (emphases added):

The importance of these dynamic risk factors is that, **in addition to being predictive** of criminal behavior, they can **serve as targets for treatment** programming. Treatments that successfully address these dynamic risk factors or criminogenic needs are associated with reduced recidivism. (Bonta et al., 2014: 280)

Changes in dynamic factors achieved through treatment that are subsequently linked to reductions in recidivism are known as criminogenic needs. (Dowden and Brown, 2002: 243)

Moreover, **the design of effective offender treatment programs is highly dependent on knowledge of the predictors of recidivism** (p. 575)...Dynamic risk factors, or what Andrews and Bonta commonly refer to as criminogenic needs (e.g., antisocial cognitions, values, and behaviors), are mutable and thus serve as the appropriate targets for treatment (Gendreau et al., 1996: 575)

Although the prediction of adult criminal recidivism is important and interesting, some have argued (Douglas & Kropp, 2002), and we concur, that **the ultimate purpose of risk assessment should be the prevention as opposed to the prediction** of criminal recidivism. (Olver et al., 2009: 346)

This theory argues that interventions should **target for change empirically established predictors of recidivism** (such as antisocial peers, antisocial attitudes, and antisocial personality. (p.23). . . Given the fact that the LSI includes a number of dynamic items, a reduction in an offender's total LSI score should occur after the offender has received treatment services appropriate for his or her risk. . . (Vose et al., 2008: 27)

Even if we granted that criminogenic risk assessment's manipulable risk factors were indeed causal, research evaluating correctional interventions suggests that these ostensibly causal effects do not equal potential intervention effects. While a complete review of the correctional intervention literature is beyond the scope of this analysis, it is worth briefly noting that this literature does not clearly corroborate the causal assumptions in the preceding quotations. Numerous analyses of the effectiveness of interventions that target criminogenic risk factors to reduce recidivism tend to find small to moderate effects and have not confirmed hypotheses about mechanisms of action (Andrews and Dowden, 2006; Lowenkamp et al., 2006). In fact, intervention effects are significantly larger when programs are combined with other services, such as mental health counseling, employment and vocational training, and educational programs (Landenberger and Lipsey, 2005). There is very little evidence that recidivism reduction is achieved by reducing "antisocial" criminogenic risk factors per se, rather than more general therapeutic and social service outcomes combined with real improvements in the material conditions of people's lives. The assumptive transition, then, in many of the reviews analyzed here, from risk prediction to risk reduction, is not supported by the data.

Discussion

We know a great deal about which individual-level factors are associated with recidivism. However, criminogenic risk assessment 1) does a poor to modest job differentiating among people at high versus low risk, 2) its predictive performance is often misinterpreted and overstated, and 3) many inferences drawn from its empirical evidence base are not supported by the data. Our findings suggest that we know comparatively little about criminogenic risk assessment's actual predictive performance, in terms of false positives, false negatives, and other metrics derived from these measures. We know even less about how, and to what effect, decisions about sensitivity, specificity, and positive and negative predictive values are implemented and evaluated in the field, only that these metrics are poorly understood by researchers and practitioners in the rare cases they are even considered.

The slippage identified in the preceding sections suggests that the state of evidence does not warrant claims that criminogenic risk assessment's "theoretical and empirical base...should be disseminated widely for purposes of enhanced crime *prevention* throughout the criminal legal system and beyond..." (Andrews et al., 2011 emphasis added). Existing evidence does not speak to its efficacy beyond tertiary prevention. In order for such claims to be evidence-based, the methodological, definitional, and inferential problems discussed above must be systematically addressed. A complete causal model that elaborates the structural- and individual-level antecedents, confounders, and mediators of criminogenic risk factors must be subjected to explicit hypothesis testing in appropriate samples.

One reason this has not already happened may be the radical empirical approach that forms the foundation of criminogenic risk assessment. That is to say, because the theory was developed to fit the data, rather than proposed *a priori* and subjected to empirical confirmation, competing explanations were not subjected to rigorous hypothesis testing. Other reasons may include prior theoretical commitments and a lack of attention to sample construction and comparison groups. For example, Andrews and Bonta (2010: 79, 93), have argued that it is a "myth" that the "roots of crime are buried deep in structural inequality." They go on to cite the results of many of the meta-analyses reviewed here, arguing that social factors such as socioeconomic status are demonstrably weaker predictors of recidivism than criminogenic risk factors. Yet this does not appear to be the case: of the nine studies that provided estimates for so-called "demographic" risk factors, roughly 56% found "effect sizes" equal to or greater than the criminogenic risk factors. Table 2 shows that demographic risk did not perform much worse (and sometimes performed better) than antisocial characteristics in their association with recidivism. This is notable because we would not expect a factor like socioeconomic status to be strongly associated with *anything* in a sample where it does not vary appreciably, and the vast majority of people targeted by mass criminalization and mass incarceration are low-income.

What might explain the mismatch between the empirical evidence and proponents' conclusions about it?

Above we have suggested that many researchers seem to overstate the predictive utility of criminogenic risk assessment in relation to the empirical evidence on which they base their claims. One possible explanation for this mismatch is that the authors of these more optimistic reviews may not be neutral arbiters of the studies they examine—both because they are often also the authors of the studies they review, and because they have financial interests in the instruments on which these studies are based. To explore this hypothesis, we conducted a post-hoc bibliometric analysis of all references cited in our sample of reviews with R package Bibliometrix (Aria and Cuccurullo, 2017), as well as a co-citation network analysis of the reviews and their analyzed studies, using R package igraph (Csardi and Nepusz, 2006).

For 35 of the 39 meta-analyses and systematic reviews, authors indicated which references were analyzed as part of review procedures, or provided lists of these primary studies in appendices or supplemental materials. We created a directed network of the relationships between the reviews and their primary studies. Supplemental Figure 3 displays this network in two layouts, with red nodes representing reviews that judged the predictive utility of criminogenic risk factors to be strong, blue nodes representing reviews that judged it to be weak, and grey nodes representing analyzed studies. The size of the grey nodes is proportional to the number of reviews that cite them.

These networks suggest that there are two distinct clusters of reviews, each of which tends to cite a group of primary studies that the other cluster mostly ignores, although there is some overlap. Moreover, each cluster tends to correspond to a different ideological position about the performance of criminogenic risk assessment: those reviews that deem the predictive utility of criminogenic risk factors to be strong tend to co-cite a similar body of studies that is distinct from the studies cited by the reviews that deem the predictive utility of criminogenic risk factors to be weak.

What characterizes the cluster of reviews that are most bullish about the predictive utility of criminogenic risk assessment? One key feature of this cluster of reviews is the involvement of the developers of a particular risk instrument, or their students and frequent collaborators. Andrews, Bonta, Dowden, Gendreau, and Wormith were authors on 73% of the reviews that judged predictive performance to be strong. Three of the five most-cited reviews (overall) included combinations of the Level of Service Inventory's creators or their students or co-authors. When we restrict the bibliometric sample to the reviews that involve these authors, we find that 17 of the top 20 primary studies cited in those reviews were authored or co-authored by Andrews, Bonta, Dowden, or Gendreau. This degree of self-citation suggests a rather insular field that is largely self-refereed. Furthermore, Andrews, Bonta, and Wormith have a proprietary interest in the Level of Services Inventory and receive royalties on sales of the instrument from its publisher, Multi-Health Systems. Conflicts of interest such as this were disclosed in only two of the nine reviews involving these authors.

Implications for policy and practice

In theory, risk assessment in the criminal legal system might productively be used to focus resources on the people most in need of support and social institutions most in need of change. But it is difficult to imagine how it might live up to this promise without radical changes, from its conceptual underpinnings to its development, implementation, and evaluation. At the very least, as the public begins to take greater notice of criminogenic risk assessment, often opposing it on ethical as well as scientific grounds (Angwin et al., 2016; Barry-Jester et al., 2015; Smith, 2016), it is incumbent upon researchers to be clear about its scientific versus political content. This is because the perceived empirical superiority of criminogenic

risk assessment lends the appearance of scientific objectivity to the selection and prioritization of risk factors, their scoring and weighting, and their tuning and revision, belying the political and value-laden decisions inherent in all data generating and modeling endeavors (O'Neil, 2016).

One way to address the theoretical and empirical overreach demonstrated above might be to democratize and de-privatize criminogenic risk assessment. This would entail: (1) making criminogenic risk assessment instruments open source and free; (2) providing open access to scoring, coding, and statistical modeling procedures; (3) providing open access to de-identified calibration and validation data; and (4) requiring jurisdictions to collect data on, and report, false positives and false negatives.

There should be no profit motive (or paywall blocking access) to the design, dissemination, and evaluation of risk assessments used to make claims about public safety, deprive people of freedom, enable or remove their access to limited treatment and social service resources, or otherwise limit or expand their life chances. In addition to transparency in the constitutive components of risk, the way in which these items are prioritized, weighted, and scored should be public and reproducible. Like certain data stored in the National Archive of Criminal Justice Data, deidentified data collected by jurisdictions using criminogenic risk assessments should be publicly available, with proper privacy protections. Jurisdictions that use criminogenic risk assessments should be required to collect data on and report sensitivity, specificity, positive predictive values, and negative predictive values on a regular basis. While the calibration of these performance measures of course has technical components, the moral and political dimensions of *misclassification* should be subject to the same public dialogue that informs other jurisprudential and penal norms.

Limitations

The present meta-review is limited in the following ways: First, it is of course possible that there was human error in implementing systematized procedures for screening reviews and extracting data. However, our procedures were designed to minimize this risk. Second, the primary aim of this meta-review was not to quantify a synthesis of findings across reviews, but rather to conduct critical, narrative analysis. Thus, despite being firmly grounded in quantitative methods, this review reflects the subjectivities, inherent biases, conceptual orientation, and political and normative perspectives of the authors. Its findings should thus be understood in that context. Finally, this meta-review is constrained by the methodological deficits of its constituent reviews.

Conclusion

As the criminogenic risk assessment expands at the same time that the criminal legal system slowly inches toward the precipice of reform, it is essential that we are

clear about what the evidence does and does not say, in order to resist the hubris of overreach and to prevent the production or reproduction of harmful, unintended consequences. Targeted, strategic, and theory-driven research on the mechanisms of prediction and successful interventions—both individual and structural—is paramount as the field moves forward.


Acknowledgments

The authors thank Drs. Sharon Schwartz, Bruce Link, and Lisa Bates for invaluable comments on earlier drafts of this manuscript. SJP also thanks Jennifer Skeem for their participation on his dissertation committee. This work was supported by the National Institute of Mental Health (T32-MH-13043) and National Institute on Drug Abuse (T32-DA-37801 and K01-DA045955).

Authors' note

References for all meta-analyses and systematic reviews analyzed in this meta-review are available in the online supplement.

ORCID iD

Seth J Prins  <https://orcid.org/0000-0003-2622-614X>

Notes

1. Not all actuarial risk assessments focus on criminogenic risk factors, and not all criminogenic risk assessments are actuarial. This meta-review, however, concerns the framework of actuarial criminogenic risk assessment.
2. One exception is a collection of meta-analyses and systematic reviews that casts doubt on criminogenic risk assessment's methodological rigor and predictive utility (Desmarais et al., 2016; Fazel et al., 2012; Singh et al., 2013; Singh and Fazel, 2010). The force of this research, though, is (appropriately) directed at unpacking the first question above, with only cursory attention to the second and third.

Supplemental Material

Supplemental material for this article is available online.

References

- Alper M, Deruose MR and Markman J (2018) *2018 Update on Prisoner Recidivism: A 9-Year Follow-up Period (2005–2014)*. NCJ250975, *Special Report*. Washington, DC: Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.
- Andrews DA, Bonta J and Wormith JS (2004) *Level of Service/Case Management Inventory (LS/CMI): An offender assessment system: User's guide*. Toronto, Canada: Multi-Health Systems.
- Andrews DA and Bonta J (2010) *The Psychology of Criminal Conduct*. 5th ed. Albany, NY: LexisNexis, Anderson Publishing.

- Andrews DA, Bonta J and Wormith JS (2011) The risk-need-responsivity (RNR) model. Does adding the good lives model contribute to effective crime prevention? *Criminal Justice and Behavior* 38(7): 735–755.
- Andrews DA and Dowden C (2006) Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology* 50(1): 88–100.
- Andrews DA, Zinger I, Hoge RD, et al. (1990) Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology* 28(3): 369–404.
- Andreas M, Vohs K, Habermeyer E (2014) Psychopathy and violent reoffending in german-speaking countries: a meta-analysis. *European Journal of Psychological Assessment* 30(2): 117. <https://doi.org/10.1027/1015-5759/a000178>.
- Angwin J, Larson J and Kirchner L (2016) Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. Available at: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 9 July 2019).
- Aria M and Cuccurullo C (2017) Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11(4): 959–975.
- Barry-Jester AM, Casselman B and Goldstein D (2015) Should prison sentences be based on crimes that haven't been committed yet? Available at: <https://fivethirtyeight.com/features/prison-reform-risk-assessment/> (accessed 9 July 2019).
- Bonta J and Andrews D (2017) *The Psychology of Criminal Conduct*. 6th ed. New York: Routledge.
- Bonta J, Blais J and Wilson HA (2014) A theoretically informed meta-analysis of the risk for general and violent recidivism for mentally disordered offenders. *Aggression and Violent Behavior* 19(3): 278–287.
- Bonta J, Law M and Hanson K (1998) The prediction of criminal and violent recidivism among mentally disordered offenders: a Meta-analysis. *Psychological Bulletin* 123(2): 123–142.
- Campbell MA, French S and Gendreau P (2009) The prediction of violence in adult offenders: a meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior* 36(6): 567–590. <https://doi.org/10.1177/0093854809333610>
- Clement M, Schwarzfeld M and Thompson M (2011) *The National Summit on Justice Reinvestment and Public Safety: Addressing Recidivism, Crime, and Corrections Spending*. New York, NY: Council of State Governments Justice Center.
- Cottle C, Lee RJ and Heilbrun K (2001) The prediction of criminal recidivism in juveniles. *Criminal Justice and Behavior* 28(3): 367–394.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Council of State Governments Justice Center (2019) *Confined and Costly: How Supervision Violations Are Filling Prisons and Burdening Budgets*. New York, NY: Council of State Governments Justice Center.
- Csardi G and Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems* 1695: 1–9.
- Cumming G (2013) Cohen's d needs to be readily interpretable: Comment on Shieh (2013). *Behavior Research Methods* 45(4): 968–971.

- Cumming G (2014) The new statistics: Why and how. *Psychological Science* 25(1): 7–29.
- Desmarais SL, Johnson KL and Singh JP (2016) Performance of recidivism risk assessment instruments in U.S. correctional settings. *Psychological Services* 13(3): 206–222.
- Dowden C and Andrews DA (1999) What works in young offender treatment: A meta-analysis. *Forum on Corrections Research* 11(2): 21–24.
- Dowden C and Brown SL (2002) The role of substance abuse factors in predicting recidivism: A meta-analysis. *Psychology, Crime & Law* 8(3): 243–264.
- Edens JF, Campbell JS and Weir JM (2007) Youth psychopathy and criminal recidivism: a meta-analysis of the psychopathy checklist measures. *Law and Human Behavior* 31(1): 53–75. <https://doi.org/10.1007/s10979-006-9019-y>
- Fazel S, Singh JP, Doll H, et al. (2012) Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: Systematic review and meta-analysis. *BMJ (Clinical Research ed.)* 345: e4692–e4692.
- Feeley M and Simon J (1992) The new penology: Notes on the emerging strategy of corrections and its implications. *Criminology* 30(4): 449–474.
- Garland D (2003) The rise of risk. In: Ericson RV and Doyle A (eds) *Risk and Morality*. Toronto: University of Toronto Press.
- Gendreau P, Little T and Goggin C (1996) A meta-analysis of the predictors of adult offender recidivism: What works. *Criminology* 34(4): 575–608.
- Gottfredson SD and Moriarty LJ (2006) Statistical risk assessment: Old problems and new applications. *Crime & Delinquency* 525(1): 178–200.
- Greenland S (2005) Epidemiologic measures and policy formulation: Lessons from potential outcomes. *Emerging Themes in Epidemiology* 2: 5.
- Greenland S, Schlesselman JJ and Criqui MH (1986) The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology* 123(2): 203–208.
- Hannah-Moffat K (1999) Moral agent or actuarial subject: Risk and Canadian women's imprisonment. *Theoretical Criminology* 3(1): 71–94.
- Hannah-Moffat KV (2004) Gendering risk at what cost: Negotiations of gender and risk in Canadian women's prisons. *Feminism & Psychology* 14(2): 243–249.
- Hannah-Moffat K (2009) Gridlock or mutability: Reconsidering “gender” and risk assessment. *Criminology & Public Policy* 8(1): 209–219.
- Hannah-Moffat K (2010) Sacrosanctor flawed: Risk. *Accountability and Gender – Responsive Penal Politics* 22(2): 25.
- Hannah-Moffat K (2013) Actuarial sentencing: An “unsettled” proposition. *Justice Quarterly* 30(2): 270–296.
- Harris GT, Rice ME and Quinsey VL (1993) Violent recidivism of mentally disordered offenders. *Criminal Justice and Behavior* 20(4): 315–335. <https://doi.org/10.1177/0093854893020004001>
- Hernán MA and VanderWeele TJ (2011) Compound treatments and transportability of causal inference. *Epidemiology (Cambridge, Mass.)* 22(3): 368–377.
- James N (2018) *Risk and Needs Assessment in the Federal Prison System*. 7-5700 R44087. Washington, DC: Congressional Research Service.

- Jones M and Kerbs JJ (2007) Probation and parole officers and discretionary decision-making: Responses to technical and criminal violations. *Federal Probation* 71(1): 9–15.
- Kennealy PJ, Skeem JL, Walters GD, et al. (2010) Do core interpersonal and affective traits of PCL-R psychopathy interact with antisocial behavior and disinhibition to predict violence? *Psychological Assessment* 22(3): 569–580. <https://doi.org/10.1037/a0019618>
- Landenberger N a and Lipsey MW (2005) The positive effects of cognitive–behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology* 1(4): 451–476.
- Leistico A-MR, Salekin RT, DeCoster J, et al. (2008) A large-scale meta-analysis relating the Hare measures of psychopathy to antisocial conduct. *Law and Human Behavior* 32(1): 28–45. <https://doi.org/10.1007/s10979-007-9096-6>
- Lipsey MW and Derzon JH (1998) Predictors of Violent or Serious Delinquency in Adolescence and Early Adulthood: A Synthesis of Longitudinal Research. *Serious & Violent Juvenile Offenders: Risk Factors and Successful Interventions* Published: 1999. Thousand Oaks: SAGE Publications, Inc. pp. pages: 86–106. Print ISBN: 9780761920403 Online ISBN: 9781452243740. DOI: <http://dx.doi.org/10.4135/9781452243740.n6> (accessed 31 December 2013).
- Lowenkamp CT, Latessa EJ and Smith P (2006) Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. *Criminology Public Policy* 5(3): 575–594.
- Lowenkamp CT and Whetzel J (2009) The development of an actuarial risk assessment instrument for U.S. pretrial services. *Federal Probation* 73(2): 33–36.
- McGrath RE and Meyer GJ (2006) When effect sizes disagree: The case of r and d. *Psychological Methods* 11(4): 386–401.
- Monahan J and Skeem JL (2016) Risk assessment in criminal sentencing. *Annual Review of Clinical Psychology* 12: 489–513.
- National Institute of Corrections (2010) *A Framework for Evidence-Based Decision Making in Local Criminal Justice Systems*. Washington, DC: Prepared for the National Institute of Corrections by the Center for Effective Public Policy, Pretrial Justice Institute, Justice Management Institute, and The Carey Group.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Books.
- Olver ME, Stockdale KC and Wormith JS (2009) Risk assessment with young offenders: A meta-analysis of three assessment measures. *Criminal Justice and Behavior* 36(4): 329–353.
- Olver ME, Stockdale KC and Wormith JS (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.
- Pearl J (2014) Is scientific knowledge useful for policy analysis? A peculiar theorem says: No. *Journal of Causal Inference* 2(1): 109–112.
- Pusch N and Holtfreter K (2018) Gender and risk assessment in juvenile offenders: a meta-analysis. *Criminal Justice and Behavior* 45(1): 56–81. <https://doi.org/10.1177/0093854817721720>
- Raynor P and Lewis S (2011) Risk-need assessment, sentencing and minority ethnic offenders in Britain. *The British Journal of Social Work* 41(7): 1357–1371.

- Rice ME and Harris GT (2005) Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior* 29(5): 615–620.
- Schwalbe CS (2007) Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior* 31(5): 449–462.
- Schwalbe CS (2008) A meta-analysis of juvenile justice risk assessment instruments. Predictive validity by gender. *Criminal Justice and Behavior* 35(11): 1367–1381.
- Schwartz S, Gatto NM and Campbell UB (2016) Causal identification: A charge of epidemiology in danger of marginalization. *Annals of Epidemiology* 26(10): 669–673.
- Serin RC and Lowenkamp CT (2015) Selecting and using risk and need assessments. Drug Court Practitioner Fact Sheet Vol. X No. 1, National Drug Court Institute.
- Simourd L and Andrews DA (1994) Correlates of delinquency: a look at gender differences. *Forum on Corrections Research* 6(1): 26–31.
- Singh JP, Desmarais SL and Van Dorn RA (2013) Measurement of predictive validity in violence risk assessment studies: A second-order systematic review. *Behavioral Sciences & the Law* 31(1): 55–73.
- Singh JP and Fazel S (2010) Forensic risk assessment. *Criminal Justice and Behavior* 37(9): 965–988.
- Smith M (2016) In Wisconsin, a backlash against using data to foretell defendants' futures. *The New York Times*, 22 June. Available at: <https://nyti.ms/2mAxdm2> (accessed 4 June 2021).
- Storey JE, Kropp PR, Hart SD, et al. (2014) Assessment and management of risk for intimate partner violence by police officers using the brief spousal assault form for the evaluation of risk. *Criminal Justice and Behavior* 41(2): 256–271.
- Story B (2016) The prison in the city: Tracking the neoliberal life of the 'million dollar block'. *Theoretical Criminology* 20(3): 257–276.
- Trujillo MP and Ross S (2008) Police response to domestic violence. *Journal of Interpersonal Violence* 23(4): 454–473.
- Uggen C and Piliavin I (1998) Asymmetrical causation and criminal desistance. *The Journal of Criminal Law and Criminology* 88(4): 1399–1422.
- Vose B, Cullen FT and Smith P (2008) The empirical status of the level of service inventory. *Federal Probation* 72(3): 22–29.
- Whittington R, Hockenfull JC, McGuire J, et al. (2013) A systematic review of risk assessment strategies for populations at high risk of engaging in violent behaviour: Update 2002–8. *Health Technology Assessment* 17(50): i–128.
- Wilson HA and Gutierrez L (2013) Does one size fit all? A meta-analysis examining the predictive ability of the level of service inventory (LSI) with aboriginal offenders. *Criminal Justice and Behavior* 41(2): 196–219.

Seth J Prins is an Assistant Professor of Epidemiology and Sociomedical Sciences at Columbia University. His work concerns the collateral public health consequences of mass incarceration and the school-to-prison pipeline, and how the division and structure of labor influence mental health.

Adam Reich is an Associate Professor of Sociology at Columbia University, and a faculty affiliate at Columbia's Interdisciplinary Center for Innovative Theory and Empirics (INCITE). He is the author of four books, the most recent of which is *Working for Respect: Community and Conflict at Walmart* (Columbia, 2018), which he co-authored with Peter Bearman.