Forensic Mental Health Practitioners' Use of Structured Risk Assessment Instruments, Views about Bias in Risk Evaluations, and Strategies to Counteract It

Jennifer Kamorowski, Corine de Ruiter, Maartje Schreuder, Karl Ask & Marko Jelícic

Published online: 30 Mar 2021.

Submit your article to this journal ⬚

Article views: 797

View related articles ⬚

View Crossmark data ⬚

Forensic Mental Health Practitioners' Use of Structured Risk Assessment Instruments, Views about Bias in Risk Evaluations, and Strategies to Counteract It

Jennifer Kamorowski[a,b] (iD), Corine de Ruiter[a] (iD), Maartje Schreuder[a] (iD), Karl Ask[b] (iD), and Marko Jelícic[a] (iD)

[a]Department of Clinical Psychological Science, Maastricht University, Maastricht, Netherlands; [b]Department of Psychology, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

The use of structured risk assessment instruments (SRAIs) has increased significantly over the past decades, with research documenting variation between countries. The use of SRAIs, their perceived utility and potential for mitigating bias in forensic risk evaluations (FREs) was investigated in a survey of Dutch forensic mental health practitioners ($N = 110$) We found generally positive views regarding SRAI utility. Bias in FREs was of concern to respondents. We found no evidence of a *bias blind spot* (the belief that oneself is less prone to bias than peers/colleagues). SRAIs were rated as the most effective debiasing strategy, but respondents also endorsed introspection. There were few differences in beliefs about sources of bias or debiasing strategies between respondents who had bias training and those who had not, suggesting the need for development of effective strategies to mitigate bias and training related to bias in FREs.

Forensic mental health practitioners play an important role in advising courts about the treatment and management of people who suffer from a mental illness and have been accused or convicted of a crime. Forensic mental health evaluations are a critical component of legal decision making because judges tend to follow the recommendations made by a behavioral expert (Gowensmith et al., 2012, 2014; Leij et al., 2001; Messina et al., 2019). Specifically, an offender's estimated risk of violent or sexual reoffending is a key consideration in determining what conditions are necessary—and legally justified—to minimize the risk of harm to potential victims and to make the best use of limited resources (Harte & Breukink, 2010).

Use of structured risk assessment instruments in forensic risk evaluations

Research conducted over the past decades indicates that using structured approaches in forensic risk evaluations (FREs)—particularly structured risk assessment instruments (SRAIs)—increases the accuracy of predictions about the likelihood of future violent or sexual offending, as compared to *unstructured clinical judgment* (AEgisdóttir et al., 2006; Andrews et al., 2006; Dawes et al., 1989; Grove & Meehl, 1996; Hanson & Morton-Bourgon, 2009). Unstructured clinical judgment (UCJ) is a method of FRE in which the clinician uses only their experience and intuition to evaluate an examinee's risk of future violence (Grove et al., 2000). In a seminal monograph, Monahan (1981) revealed that UCJs about the likelihood of violent reoffending were correct in about one out of three cases, which lead to criticisms of UCJ as being inaccurate and unreliable (Douglas & Kropp, 2002).

Since the early 1980s, an astonishing variety of standardized and evidence-based SRAIs have been developed for evaluating the potential for different types of violence (Shepherd & Sullivan, 2017). For example, among a large-scale international survey of forensic mental health professionals ($N = 2,135$), the respondents reported using over 200 different commercially available SRAIs and more than 200 different locally-developed SRAIs (Singh et al., 2014). SRAIs provide information about the probability of recidivism, the severity of the consequences if reoffending

occurs, and whether the reoffending risk is imminent or more remote. SRAIs can also help to identify offender treatment needs that, if addressed, may reduce the risk of reoffending (Bonta & Andrews, 2007).

Actuarial risk assessment instruments are one type of SRAI and contain factors empirically related to an increased risk of violent, sexual, or general criminal recidivism (Doyle & Dolan, 2007), depending on the instrument's purpose. Risk factors are scored and combined according to an algorithm determined by the tool developers. The final risk score thereby provides a recidivism risk estimate based on the recidivism rates of groups of individuals with the same score (Doyle & Dolan, 2007). A second type of SRAIs employs *structured professional judgment* (SPJ), a method in which the evaluator considers the presence or absence of empirically-based risk factors and risk factors they deem relevant based on expertise. The evaluator uses their professional judgment to determine the relative importance of these factors to the individual case (Douglas & Kropp, 2002). Actuarial and SPJ instruments generally exhibit approximately similar predictive validity (Campbell et al., 2009; Yang et al., 2010), but the type of SRAI an evaluator chooses may depend on the setting and purpose of the risk evaluation (Brown & Rakow, 2016).

Numerous national and international surveys of forensic mental health professionals indicate that the use of SRAIs to assess the risk of violent recidivism is increasingly common in FREs (Archer et al., 2006; Hurducas et al., 2014; Kelley et al., 2020; Lally, 2003; Neal & Grisso, 2014b; Singh et al., 2014). Findings from more recent surveys also indicate that SRAI use and FRE practices can vary widely across countries (e.g., Canada: McLaughlin & Kan, 2014; Denmark: Nielsen et al., 2015; Belgium: Pham et al., 2016; Israel: Singh et al., 2019). Yet there is relatively little country-specific research about which SRAIs forensic evaluators are required to use in practice, which SRAIs they choose to use, and their perceptions about the usefulness of those tools (Hurducas et al., 2014).

Use of SRAIs in forensic risk evaluations in the Netherlands

SRAIs were introduced into the Dutch forensic mental health system in the late 1990s (de Ruiter & Hildebrand, 2007) and are now commonly used in FREs. Dutch forensic mental health experts evaluate approximately four to five thousand criminal defendants a year and produce a written report for a court (Messina et al., 2019). In the Netherlands, a person who has committed a crime and has a mental disorder can be ordered to treatment in a secure psychiatric facility, usually after time served in prison, a disposition known as *maatregel van terbeschikkingstelling* (TBS; de Ruiter & Hildebrand, 2007). Although the goal of treatment under TBS is to successfully reintegrate the offender into the community, a TBS order can entail lifelong mandatory treatment (Bogaerts et al., 2018). SRAIs are therefore used to assess offenders both for potential commitment to TBS, and for regular reviews to determine if the individual's risk has been lowered to a degree sufficient to warrant release (de Ruiter, 2016).

We are unaware of any published surveys of Dutch forensic mental health practitioners about the SRAIs they most commonly use and the perceived utility of those instruments. Two relatively recent studies published in the Dutch language reported on SRAIs commonly used in forensic settings in the Netherlands (Harte & Breukink, 2010; van Horn et al., 2016), but neither surveyed practitioners about which SRAIs they use in practice, nor the perceived utility of the SRAIs. Therefore, in the current study, we survey forensic mental health practitioners about the SRAIs they are required or choose to use and the perceived usefulness of specific SRAIs.

Bias awareness and sources of bias in forensic risk evaluations

Cognitive bias refers to systematic errors in logic or reasoning that occur outside of conscious awareness (Wilson & Brekke, 1994) and are the result of the mind's automatic processing of information based on experience and prior expectations (Tversky & Kahneman, 1974). Therefore, cognitive bias represents a threat to the objectivity and validity of a forensic mental health evaluation (Neal & Grisso, 2014a; Zapf & Dror, 2017). In addition, the reliability and accuracy of FREs depend to a large extent on the evaluator's ability to minimize the influence of cognitive bias and reach objective conclusions (MacLean et al., 2019).

A number of recently published surveys suggest that forensic mental health practitioners are aware of, and concerned about, the potential for bias in forensic mental health evaluations (Kukucka et al., 2017; Neal & Brodsky, 2016; Neal & Grisso, 2014b; Zapf et al., 2018). However, many remain skeptical about bias affecting their own work as evidenced by a *bias blind spot* (Pronin et al., 2002), that is, the belief that they are less prone to bias than their colleagues (Boccaccini et al., 2017; Kukucka et al., 2017; Neal & Brodsky, 2016; Zapf et al., 2018; Zappala et al., 2018). For

example, Zapf and colleagues (Zapf et al., 2018) surveyed 1,099 mental health practitioners who conduct forensic evaluations and just over half (52.2%) agreed that their own judgments can be influenced by cognitive bias (Zapf et al., 2018).

In order for forensic evaluators to take appropriate steps to counter bias, they must first be aware of it, have accurate perceptions about how bias operates (e.g., Kukucka et al., 2017; MacLean et al., 2019; Zapf & Dror, 2017), and accept the potential for it to affect their work (Dror, 2018; Lilienfeld et al., 2009; Wilson & Brekke, 1994). Unfortunately, it also appears that most forensic evaluators do not receive formal training about how various type of cognitive biases can impact FREs (Zapf et al., 2018). Finally, forensic evaluators need guidance about specific *effective* debiasing strategies that they can employ to increase the objectivity and validity of their conclusions (MacLean et al., 2019; Neal & Brodsky, 2016; Zapf et al., 2018).

Debiasing strategies

Research related to potential debiasing strategies is substantial in medical, business, and policy applications (Soll et al., 2015).Some progress has also been made in seeking to counteract bias that may occur in the physical forensic sciences (Dror, 2018; Dror et al., 2015; Jeanguenat et al., 2017; Kassin et al., 2013). Yet, despite indications that bias can affect forensic evaluations (e.g., Beckham et al., 1989; Boccaccini et al., 2008; Boccaccini et al., 2017; Guarnera et al., 2017; Murrie et al., 2013), research regarding forensic mental health evaluators' awareness of various types of cognitive bias and potential strategies to counteract them has only recently emerged (Dror & Murrie, 2018; MacLean et al., 2019; Neal & Brodsky, 2016; Neal & Grisso, 2014a; Zapf et al., 2018; Zapf & Dror, 2017).

In the first study to examine forensic mental health evaluators' ideas for potential debiasing strategies, Neal and Brodsky (2016) conducted interviews with 20 forensic psychologists certified by the American Board of Forensic Psychology. The interview prompts were designed to obtain information about the psychologists' awareness of the potential for bias in forensic evaluations and different strategies they believed could minimize bias in their own work. Qualitative analysis of the interviews revealed 25 strategies the participants believed may be useful to mitigate bias in forensic evaluations (see Table 4 for a complete list). In the second stage of their study, Neal and Brodsky asked 351 members of the American Psychological Association to rate the 25 strategies on their usefulness as potential bias correction measures. Overall, participants in the second stage of the study rated 22 of the strategies as useful or very useful. We sought to add to these findings by eliciting Dutch forensic evaluators' beliefs about potential sources of bias and the effectiveness of various strategies to counteract it in FREs (Neal & Brodsky, 2016; Zapf et al., 2018).

The current study

In this study, we surveyed Dutch forensic mental health evaluators regarding their use and perceived effectiveness of SRAIs for evaluating violent, sexual, criminal, and intimate partner violence risk. Based on previous research conducted by Zapf et al. (2018), we also measured Dutch evaluators' awareness about the potential for bias in forensic risk evaluations. This study further adds to the previous research conducted by Neal and Brodsky (2016) by examining evaluators' perceptions of the efficacy of various debiasing strategies to reduce the potential for cognitive bias in FREs.

Method

Participants

We aimed to obtain a representative sample of Dutch mental health practitioners who conduct forensic risk evaluations. To be eligible to participate, respondents had to be a mental health professional who works in the Netherlands and conducts risk evaluations. Out of 154 respondents who began the survey, 44 did not complete it (71.4% completion rate). We excluded incomplete surveys from our analyses. Therefore, our sample was comprised of 110 respondents with current or previous experience conducting risk evaluations. About 60.9% of respondents were women and approximately 39.1% men. Most had obtained a Master's degree ($n = 77$; 70.0%) and 25 had a doctoral degree (22.7%). Six respondents held a Bachelor's degree (5.5%), and two had a 2-year degree (1.8%).

Experience in mental health settings among respondents ranged from 0 to 46 years ($M = 18.9$, $SD = 11.8$, $Mdn = 18$). Respondents worked in a variety of settings, including forensic psychiatric hospitals (5.5%), private practice (14.5%), mental health clinics (9.1%), and one person indicated working in a hospital setting. Half of respondents selected the option of "other" for their work environment, with most indicating that they worked in some type of clinical or correctional setting (e.g., forensic outpatient clinic, prison psychiatric center or detention facility), and

several who worked solely as a court-appointed evaluator or consultant.

Procedure

This study was reviewed and approved by the Ethics Review Committee for XXX (reference 185_06_11_2017). Data collection took place between late March and early June of 2018. Participants were recruited via advertisements in the newsletter of the *Nederlands Register Gerechtelijk Deskundigen* (NRGD; Dutch Registry of Court Experts) and of the *Expertisecentrum Forensische Psychiatrie* (EFP; Expertise Center for Forensic Psychiatry). The NRGD is a registry of forensic professionals created by statute in 2010. Forensic experts who wish to register with the NRGD are evaluated based on field-specific requirements before they are approved (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). The EFP is a professional organization that facilitates cooperation between researchers and practitioners in the field of forensic psychiatry in the Netherlands (Expertisecentrum Forensische Psychiatrie, n.d.).

We also recruited participants via social media with Facebook posts in groups of former forensic psychology Master's students from XXX University, and via personal invitations by e-mail to professional contacts of one of the authors. This same author also shared the survey announcement on their professional LinkedIn page. The survey was also advertised on KNAPP, which is an online site dedicated to forensic psychiatric care in the Netherlands, and permits messages to be posted to members to facilitate collaboration and knowledge-sharing (https://www.knapp-efp. nl). In return for completing the study, respondents were offered an opportunity to enter a raffle for a chance to win a voucher in the amount of €50 from an online shopping site. The winner was randomly selected from those who completed the survey and provided their e-mail address.

It is difficult to provide an accurate estimate of the number of potential survey respondents we reached. There were 489 forensic mental health experts listed on the NRGD website at the time we conducted the survey and all of them were sent the NRGD newsletter with the study announcement. To NRGD behavioral experts specializing in either adult or juvenile forensic psychology or psychiatry whose e-mail addresses were available online, we sent a personal invitation by e-mail ($n = 270$) and two follow-up reminders. We have no information about how many potential respondents may have viewed the EFP newsletter online. We estimate approximately 400 people eligible to participate in the survey were reached by Facebook and LinkedIn. Finally, there are approximately 1,880 members on the KNAPP website, although it is unknown how many of them were eligible to participate in the survey. There is also overlap between NRGD experts, professionals who receive the EFP newsletter, the author's professional network, and users of the KNAPP website. We were unable to determine the number of participants obtained from the individual platforms we used for advertisement of the study. Thus, the exact response rate is unknown.

Survey

The survey was conducted online using Qualtrics and was available to respondents in both English and Dutch. We used the cookies-based Qualtrics feature to prevent respondents from taking the survey multiple times. We opted not to provide a back button in order to prevent respondents from changing their answers after they completed a page.

Respondents were provided with a brief description of the survey and acknowledged their consent to participate. The survey was comprised of five major sections as outlined below.

Demographics, FRE experience, general frequency of SRAI use, concerns about errors, and consulting about FREs

In the first section, we collected demographic information regarding gender, education level, years of experience in mental health settings, and current work environment. We asked respondents to specify how many forensic risk evaluations they had performed over the past two years. Next, respondents were asked to indicate on a 5-point Likert-type scale how useful they thought SRAIs are in conducting risk evaluations (1 = *not at all*, 2 = *slightly*, 3 = *moderately*, 4 = *very*, 5 = *extremely*). We also asked how often an SRAI was used when conducting a risk evaluation (1 = *never*, 2 = *sometimes*, 3 = *about half the time*, 4 = *most of the time*, 5 = *always*).

Next, respondents were asked to rate their concerns about the possibility of false positives and false negatives on two separate 5-point Likert scales (1 = *not at all*, 2 = *a little*, 3 = *a moderate amount*, 4 = *a lot*, to 5 = *a great deal*). In the survey, we defined a false negative as occurring when an individual is classified as having a low risk to reoffend, when in reality he/she has a high risk. We defined a false positive as occurring when an individual is classified as having a

high risk to reoffend, when in reality, he/she has a low risk.

The next two questions related to respondents' frequency of consulting with third parties about FREs (1 = *never* to 5 = *always*). We defined consulting in the survey as "seeking advice about the evaluation." We also asked respondents to specify with whom they consulted from a list of options (*colleagues, supervisor, other treatment provider(s) of the evaluee, prison/jail staff, evaluee's family members, probation/parole officer(s), or other*). Respondents could select multiple options, and if *other* was selected, we asked them to specify in free-text.

Use of specific SRAIs and usefulness ratings
In the subsequent section, respondents were asked which SRAIs they were required by their employer or jurisdiction to use, and which SRAIs they chose to use. We created a list of commonly-used SRAIs in the Netherlands (see Table 1 for complete list); respondents could select multiple SRAIs and also provide free-text responses. For the SRAIs that respondents indicated they were required or chose to use, they rated the usefulness of each separately on a 5-point Likert scale (1 = *not at all useful* to 5 = *extremely useful*).

Cognitive bias concerns in FREs
We asked respondents about their views regarding cognitive bias in forensic risk evaluations (FREs). The extent to which respondents thought cognitive bias is a problem in FREs was rated on a 5-point Likert scale (1 = *not at all*, 2 = *a little*, 3 = *a moderate amount*, 4 = *a lot*, 5 = *a great deal*). We defined cognitive bias in the survey as an error in reasoning, evaluating, remembering, or processing information. Respondents then indicated *yes* or *no* in response to questions about whether they had received any specific training about cognitive bias in FREs and whether they had ever been concerned about bias in an FRE conducted by (a) themselves or (b) someone else.

Ratings of effectiveness of potential debiasing strategies
In the next section, respondents were asked to rate on a 5-point Likert scale the efficacy of each of the 25 debiasing strategies (Neal & Brodsky, 2016) to reduce the potential that cognitive bias will influence an evaluator's judgment about future violence risk (1 = *not effective at all* to 5 = *extremely effective*). In contrast to Neal and Brodsky, we did not give participants an option to provide a rating of *unsure*, but used *moderately effective* as the midpoint of our scale. We were

interested in respondents' *beliefs* about how effective the strategies are, not whether the strategies are effective in reality. A response of *unsure* would have made it impossible to distinguish between a respondent who was unsure about whether the debiasing strategy was effective and a respondent who was unsure about their own opinion.

Potential sources of bias in FREs
Finally, respondents indicated on a 5-point Likert scale (1 = *strongly disagree* to 5 = *strongly agree*) the extent to which they believed that various situations have the potential to bias an evaluator when conducting an FRE. Seven items were derived from Zapf et al. (2018) survey. The statements included situations such as whether an evaluator's prior beliefs affect how they analyze a case or their ultimate opinion about the case, whether making a conscious effort to set aside prior beliefs can reduce the likelihood the evaluator will be influenced by them, and whether evaluators know in advance what conclusion they are expected to reach in a case and whether that affects their conclusion (see Table 3 for the complete list).

One difference that should be noted between the current study and Zapf et al.'s study is that we eliminated questions about the effects of irrelevant contextual information (Zapf et al., 2018, p. 5, Table 2, #5-7). Despite the fact that research suggests irrelevant contextual information can result in bias and errors in other areas of forensic science (Dror, 2012, Dror, 2018; Dror et al., 2006, Dror et al., 2015; Kukucka & Kassin, 2014; Nakhaeizadeh et al., 2014), there is debate about what type of information should be considered irrelevant contextual information in these fields (Curley et al., 2019, 2020; Dror & Murrie, 2018; Gardner et al., 2019; Thompson et al., 2020). Further, we were unable to locate any agreed-upon definition of exactly what constitutes irrelevant contextual information specifically for FREs. Therefore, we opted not to include items based on this concept.

Data analysis
All analyses were conducted using IBM SPSS software version 25.

Results

Experience conducting FREs and consulting about FREs

In the two years preceding the survey, the majority of respondents (59%) conducted 10 or more forensic risk

Table 1. Risk assessment instruments: required and optional users.

Risk assessment instrument	Instrument type	Prescribed use	Required users n (%)	Optional users n (%)
Static-99 (Hanson & Thornton, 1999) or *Static-99R* (Helmus et al., 2012)	Actuarial	Sexual reoffending risk	59 (53.6%)	10 (9.1%)
Historical Clinical Risk Management-20 (HCR-20V2; Webster et al., 1997; HCR-20^{V3}; Douglas et al., 2014)	SPJ	Violence risk	53 (48.2%)	14 (12.7%)
Historische, Klinische, Toekomstige-30 (HKT-30; Werkgroep Pilotstudy Risicotaxie, 2002) or HKT-R (Spreen et al., 2014)	SPJ	Violence risk	52 (47.3%)	7 (6.4%)
Structured Assessment of PROtective Factors (SAPROF; de Vogel et al., 2009, 2012)	SPJ	Protective factors	52 (47.3%)	22 (20.0%)
Psychopathy Checklist Revised (PCL-R; Hare, 2003) or the *Psychopathy Checklist: Screening Version* (PCL:SV; Hart et al., 1995)	Actuarial	Psychopathic traits	49 (44.5%)	18 (16.4%)
OTHER (Free text entry)	N/A	N/A	39 (35.5%)	43 (39.1%)
Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2002)	SPJ	Youth (12-18 years) violence risk	18 (16.4%)	5 (4.5%)
Short-Term Assessment of Risk and Treatability (START; Webster et al., 2004)	SPJ	Violence risk	10 (9.1%)	2 (1.8%)
None	N/A	N/A	9 (8.2%)	5 (4.5%)
Brief Spousal Assault Form for the Evaluation of Risk (B-SAFER; Kropp et al., 2005)	Actuarial	Spousal assault risk	8 (7.3%)	2 (1.8%)
Violence Risk Appraisal Guide (VRAG; Harris, Rice, & Quinsey, 1993)	Actuarial	Violence risk	1 (0.9%)	1 (0.9%)
Recidive Inschattingsschalen [Recidivism Assessment Scales] (RISC; Adviesbureau Van Montfoort & Reclassering Nederland, 2004)	Actuarial	General risk of antisocial behavior, criminogenic needs	1 (0.9%)	1 (0.9%)
Violence Risk Screening-10 (V-RISK-10; Hartvig et al., 2007)	Actuarial	Violence risk screening	0 (0.0%)	0 (0.0%)
Violence Risk Scale (VRS; Wong & Gordon, 2006)	Actuarial	Violence risk	0 (0.0%)	1 (0.9%)
Classification of Violence Risk (COVR; Monahan et al., 2006)	Actuarial	Violence risk in community	0 (0.0%)	0 (0.0%)
Forensic Operationalized Therapy/Risk Evaluation System (FOTRES; Urbaniok, 2007)	SPJ	Recidivism risk and treatment progress	0 (0.0%)	0 (0.0%)
Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 1998)	Actuarial	Sexual reoffending risk	0 (0.0%)	0 (0.0%)

Note. SRAIs are arranged in descending order of number of required users.

evaluations ($Mdn = 12$; range of 0–250), seven (6.4%) of whom had conducted 100 or more. Forty-six respondents (41.8%) had performed 10 or more pre-sentencing evaluations, and six (5.5%) had conducted 100 or more. Twenty-two respondents (20.0%) had performed at least one inpatient TBS evaluation in the past two years, 11 (10.0%) of whom had performed 10 or more. Fifteen (13.6%) respondents had conducted TBS extension evaluations within the previous two years, eight (7.3%) of whom had done 10 or more of these evaluations.

Nearly all respondents ($n = 104$; 94.5%) said they consulted with others about FREs at least some of the time. Over half (57.3%) said they consulted with others *always* or *most of the time*. More than two-thirds of respondents (69.1%) indicated they consult with colleagues and a substantial minority (41.8%) said they consulted with another treatment provider of the examinee. More than one-third consult with either the examinee's probation/parole officer or prison staff (29.1% and 6.4%, respectively), while less than 10% consult with the examinee's family members (9.1%).

Frequency of SRAI use and required and optional use of specific SRAIs

Almost all respondents indicated they used an SRAI for risk evaluations *always* or *most of the time* ($n = 107$; 97.3%). Only one respondent indicated that they never used an SRAI when conducting a risk evaluation, one said they did about half the time, and one said they did sometimes. The respondents also mostly agreed that SRAIs are useful in conducting risk evaluations, with 14 (12.7%) rating them as extremely useful and 69 (62.7%) rating them as very useful. However, 19 respondents (17.3%) rated SRAIs as moderately useful, and eight (7.3%) rated them as only slightly useful. Most of our respondents were required to use a specific SRAI, with only about 8% indicating they were not required to do so. Five respondents chose not to use any SRAIs; however, it is unknown whether they chose not to use any specific SRAIs because they used only those that were required.

Many respondents indicated they were both required and choose to use the same SRAIs. In other words, these respondents were in the required and

optional users group and provided duplicate ratings for the same SRAI. We therefore created two conceptually distinct groups: *required users* and *optional users*. If respondents indicated they were required, or both required and chose to use the SRAI, we considered them required users. If respondents indicated only that they chose to use the SRAI, we considered them optional users (see Tables 1 and 2).

The Historical Clinical Risk Management-20 (HCR-20V2; Webster et al., 1997; HCR-20V3; Douglas et al., 2014) and the Historische, Klinische, Toekomstige-30 (HKT-30; Werkgroep Pilotstudy Risicotaxie, 2002) or its successor, the HKT-R (Spreen et al., 2014), were the two most commonly reported required SRAIs (48.2% and 47.3%, respectively). The HCR-20 is an SPJ tool designed to assess psychosocial functioning as it relates to violence risk among adults and includes 10 historical, five clinical, and five risk management items (Douglas et al., 2014). The HKT-30 is a Dutch-language SPJ tool that was created in the Netherlands for the evaluation of violence risk in forensic psychiatric settings. The HKT-30 has a structure similar to the HCR-20, and consists of 11 historical, 13 clinical and dynamic, and six risk management items rated on a 5-point scale (Werkgroep Pilotstudy Risicotaxie, 2002).

Similarly, nearly half (47.3%) of our respondents were required to use the Structured Assessment of PROtective Factors for violence risk (SAPROF; de Vogel, de Ruiter, Bouman, & de Vries Robbé, 2009; 2012). The SAPROF is designed to be used in conjunction with an SPJ SRAI and comprises 17 factors considered to be protective against violent behavior (de Vries Robbé et al., 2013). Each protective factor is rated on a three-point scale (0 = *clearly absent*, 1 = *somewhat present*, 2 = *clearly present*). An Integrative Final Risk Judgment is obtained by integrating information about the protective factors assessed with the SAPROF and the risk factors measured with an SPJ risk assessment tool (de Vries Robbé et al., 2013).

Nearly half of respondents (44.5%) were required to use the Psychopathy Checklist-Revised (PCL-R; Hare, 2003) or the Psychopathy Checklist: Screening Version (PCL:SV; Hart, Cox, & Hare, 1995). The PCL-R and PCL:SV are psychometric tools to assess an individual's level of psychopathy. Although the PCL-R and PCL:SV are not violence risk assessment instruments *per se*, psychopathic traits are linked with an increased risk of violent recidivism (Barbaree et al., 2001; Hawes et al., 2013; Lanterman et al., 2014). Therefore, the PCL-R or PCL:SV is commonly used in combination with SRAIs to assess the risk of violent or sexual reoffending.

To estimate sexual reoffending risk, more than half of respondents (53.6%) were required to use the Static-99 (Hanson & Thornton, 1999) or the Static-99R (Helmus et al., 2012). The Static-99R is an actuarial tool comprised of 10 static, historical risk factors. Although not specifically listed as a choice, two respondents reported they were required to use an SPJ tool, the *Sexual Violence Risk-20* (SVR-20; Boer et al., 1997) to assess sexual reoffending risk. The SVR-20 consists of 20 items related to three domains: psychosocial adjustment, sexual offenses, and future plans (Rettenberger et al., 2017).

A number of participants indicated by free-text response that they used a combination of the Static-99R, the STABLE-2007 (Hanson et al., 2007), and the ACUTE-2007 (Hanson et al., 2007) to evaluate the risk of sexual reoffending. Some were required to use the STABLE (Hanson et al., 2007) or the ACUTE (Hanson et al., 2007). The STABLE-2007 is used in sexual offending recidivism risk assessment and contains 13 risk factors related to the offender's ability to regulate his sexual behavior (Hanson et al., 2015). The ACUTE-2007 is a measure of dynamic risk factors and contains seven items related to sexual reoffending risk that can change over the short-term (Hanson et al., 2007).

Perceived usefulness of most frequently used SRAIs

An independent samples *t*-test revealed that the only significant difference between required and optional users with respect to their usefulness ratings of the five most commonly used SRAIs (i.e., HCR-20, HKT-30/R, SAPROF, PCL-R/SV, and Static-99/99R) were found for the Static-99/99R. The optional users rated the Static-99/99R as significantly less useful ($M = 3.90$, $SD = 0.32$) than required users ($M = 4.31$, $SD = 0.75$), $t(30.02) = 2.90$, $p = .007$, Hedges' $g = 0.58$ (medium effect), 95% CI [-0.10, 1.26]. Mean usefulness ratings of the five most commonly used SRAIs for required and optional users are presented in Table 2, as are the results of the independent samples *t*-tests.

Concerns about cognitive bias and errors in FREs

Overall, respondents rated cognitive bias as a *moderate* problem in FREs in general ($M = 3.32$, $SD = 0.75$). A McNemar's test revealed no significant difference in the proportion of respondents who were concerned about bias in an FRE conducted by someone else compared to those concerned about bias in an FRE

Table 2. Group mean and differences in usefulness ratings of most frequently used SRAIs.

Risk assessment instrument	Required users M (SD)	Optional users M (SD)	t	p	Hedges' g [95% CI]
HCR-20	3.98 (0.72)	3.93 (0.83)	$t(65) = 0.24$	.815	0.07 [-0.52, 0.66]
HKT-30/R	3.92 (0.88)	3.57 (0.98)	$t(57) = 0.98$	.332	0.39 [-0.40, 1.18]
SAPROF	3.71 (0.85)	3.91 (0.75)	$t(72) = -0.95$	.347	0.24 [-0.26, 0.74]
PCL-R/SV	3.88 (1.03)	3.89 (0.83)	$t(65) = -0.04$	.967	0.01 [-0.64, 0.66]
Static-99/99R[a]	4.31 (0.75)	3.90 (0.32)	$t(30.02) = 2.90$	.007	0.57 [-0.10, 1.26]

Note. SRAIs were rated on a Likert scale from 1 = not at all useful to 5 = extremely useful.
[a]Welch's t-test (unequal variance between groups).

conducted by themselves, $p = .388$. A paired samples t-test revealed that respondents were significantly more concerned about the possibility of a false negative ($M = 3.07$, $SD = 1.00$) than a false positive ($M = 2.85$, $SD = 0.81$) in a risk evaluation, $t(109) = -2.38$, $p = .019$, Hedges' $g_{avg} = 0.31$ (small effect), 95% CI [0.05, 0.57].

Potential sources of bias in FREs

Similar to the approach of Zapf and colleagues (2017), we asked respondents about situations that may have the potential to bias evaluators when they are conducting risk evaluations. Mean ratings, response frequencies, and modal responses among our respondents for each level of agreement regarding potential sources of bias in forensic evaluations are presented in Table 3.

A substantial majority (83.6%) of our respondents did not agree that cognitive bias is *less* of a problem in forensic psychology than in other forensic sciences (e.g., fingerprint analysis, hair matching). They generally agreed (57.2%) that evaluators sometimes know what conclusions they are expected to reach and that this affects their conclusions (61.8%). Most of our respondents agreed that an evaluator's prior beliefs and expectations can affect how they analyze a case (90.9%) or their ultimate opinion on the case (90.0%). Yet, a substantial majority (71.8%) agreed that if an evaluator makes a conscious effort to set aside their prior beliefs and expectations, they are less likely to be influenced by them.

Perceived effectiveness of debiasing strategies

We asked our respondents to rate how effective they thought the 25 strategies identified in Neal and Brodsky (2016) are in reducing the potential for cognitive bias in a risk evaluation. The mean effectiveness ratings for all debiasing strategies are presented in Table 4. We calculated the sample mean rating across all six strategies that have been suggested in the literature as potentially effective in mitigating bias, and the

sample mean rating for the 19 strategies identified by Neal and Brodsky (2016) as ineffective or not specifically suggested previously. Only introspection has been shown to be ineffective, because people are often unaware of the existence of a particular biasing stimulus, that their response is influenced by the stimulus, or both (Nisbett & Wilson, 1977; Pronin & Kugler, 2007; Wilson & Brekke, 1994).

Strategies previously suggested in scientific literature include the following: critically examining conclusions, consulting with colleagues, taking time to think about the evaluation information rather than immediately writing the report, receiving explicit didactic training about objectivity, exposure to the importance of objectivity through reading professional literature, and using structured evaluation methods. A paired samples t-test revealed that respondents gave significantly higher ratings to strategies that have been suggested in the literature ($M = 3.47$, $SD = 0.57$) than to those identified by Neal and Brodsky as not having been specifically suggested previously ($M = 3.09$, $SD = 0.59$), $t(109) = 10.51$, 95% CI [0.31, 0.45], $p < .001$, Hedges' $g_{avg} = 0.65$ (medium effect), 95% CI [0.50, 0.80].

Effects of cognitive bias training

A minority of respondents (27%) indicated that they had received training on cognitive bias. An independent samples t-test revealed no significant difference between respondents who had received bias training ($M = 3.37$, $SD = 0.77$) and those who had not ($M = 3.30$, $SD = 0.75$) with regard to the extent to which they believed cognitive bias is a problem in FREs, $t(108) = -0.41$, Hedges' $g = 0.09$, 95% CI [-0.33, 0.51]. Overall, this finding indicates respondents think cognitive bias is a moderate problem in FREs, regardless of whether they have had cognitive bias training or not.

We also examined whether training about cognitive bias in forensic evaluations affected ratings of the debiasing strategies. A Welch's t-test revealed that practitioners who had some training related to bias

Table 3. Mean ratings and frequencies of beliefs about potential sources of cognitive bias in forensic risk evaluations.

Item	M (SD)	Strongly disagree n (%)	Somewhat disagree n (%)	Neither agree nor disagree n (%)	Somewhat agree n (%)	Strongly agree n (%)
An evaluator's prior beliefs and expectations can affect how s/he goes about analyzing a forensic case	4.15 (0.74)	1 (0.9%)	4 (3.6%)	5 (4.5%)	**67 (60.9%)**	33 (30.0%)
An evaluator's prior beliefs and expectations can affect his/her ultimate opinion about a forensic case	4.12 (0.71)	1 (0.9%)	3 (2.7%)	7 (6.4%)	**70 (63.6%)**	29 (26.4%)
An evaluator who makes a conscious effort to set aside his or her prior beliefs and expectations is less likely to be influenced by them	3.75 (0.85)	4 (3.6%)	3 (2.7%)	24 (21.8%)	**65 (59.1%)**	14 (12.7%)
When evaluators know what they are expected to find, it affects the conclusions they reach	3.53 (1.01)	6 (5.5%)	11 (10.0%)	25 (22.7%)	**55 (50.0%)**	13 (11.8%)
Evaluators sometimes know what conclusion they are expected to reach	3.49 (1.05)	7 (6.4%)	10 (9.1%)	30 (27.3%)	**48 (43.6%)**	15 (13.6%)
An experienced evaluator is less likely than a new evaluator to be influenced by prior beliefs/expectations	2.38 (1.09)	27 (24.5%)	**37 (33.6%)**	25 (22.7%)	19 (17.3%)	2 (1.8%)
Cognitive bias is generally less of a problem in forensic psychology than in other forensic sciences (e.g., fingerprint analysis; ballistics; hair matching, etc.)	1.69 (0.87)	**57 (51.8%)**	35 (31.8%)	14 (12.7%)	3 (2.7%)	1 (0.9%)

Note. Items are presented according to mean ratings in descending order. Modal responses are shown in bold.

gave significantly higher effectiveness ratings ($M = 4.10$, $SD = 0.61$) to "taking personal responsibility to continue learning after completing formal training and education" as a debiasing strategy than those who had not received such training ($M = 3.71$, $SD = 0.83$), $t(71.02) = -2.68$, $p = .009$, Hedges' $g = 0.50$ (medium effect), 95% CI [0.08, 0.93]. We are unaware of research that tests whether this strategy is effective in counteracting cognitive bias.

Discussion

Risk evaluation practices and use of SRAIs may vary by country, by type of legal system (i.e., adversarial vs. inquisitorial), and other system factors, such as recommendations by professional organizations, organizational or statutory requirements to use (specific) SRAIs, and regulations governing recognition as an expert in the legal system (McLaughlin & Kan, 2014). In addition, awareness and education about the potential for cognitive bias in forensic risk evaluations and potentially useful debiasing strategies may also vary between countries. Therefore, the purpose of this survey was threefold: (a) to identify which SRAIs are commonly used by forensic mental health evaluators who conduct FREs in the Netherlands and their perceptions of the usefulness of those SRAIs; (b) to gain insight into evaluators' views about potential sources of bias when conducting FREs; and (c) to examine evaluators' concerns about cognitive bias and their views of potential strategies to diminish bias.

Frequency of use and usefulness ratings of SRAIs

Similar to previous surveys (Hurducas et al., 2014; Neal & Brodsky, 2016; Zapf et al., 2018), respondents in the current study were quite experienced in their respective fields and averaged approximately 18 years of experience in mental health settings. The majority of our respondents had conducted more than 10 risk assessments in the two years preceding our survey, with a small proportion having conducted 100 or more.

Our respondents reported a very high rate of SRAI use, as nearly all of them indicated that they use an SRAI always or most of the time. Further, the vast majority of respondents were required to use a specific SRAI. These high rates of SRAI use likely reflect professional recommendations in the Netherlands that urge the use of SRAIs in forensic assessments (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). Our results are in line with findings in other recent international surveys suggesting the increasing use of SRAIs (Neal & Grisso, 2014b; Singh et al., 2014). For example, Neal and Grisso (2014b) reported that 96.9% of their respondents reported using an SRAI for sexual offender risk assessment and 89.0% for violence risk assessment. Similarly, among an international sample of mental health professionals from Europe, Singh and colleagues (2014) reported that over the 12 months preceding the survey, approximately 63% of respondents used SRAIs to conduct risk assessments.

Table 4. Means and Frequencies of Effectiveness Ratings for Debiasing Strategies (N = 110).

Strategy	M (SD)	Effectiveness Ratings n (%)				
		Not at all	Slightly	Moderately	Very	Extremely
Using structured evaluation methods	4.00 (0.89)	1 (0.9%)	5 (4.5%)	22 (20.0%)	**47 (42.7%)**	35 (31.8%)
Taking personal responsibility to continue learning after completing formal training and education	3.82 (0.79)	0 (0.0%)	6 (5.5%)	28 (25.5%)	**56 (50.9%)**	20 (18.2%)
Investigating all relevant data before forming an opinion[a]	3.75 (0.82)	1 (0.9%)	6 (5.5%)	30 (27.3%)	**56 (50.9%)**	17 (15.5%)
Basing conclusions and opinions on sound data[b]	3.74 (0.82)	0 (0.0%)	9 (8.2%)	28 (25.5%)	**56 (50.9%)**	17 (15.5%)
Consulting with colleagues about issues of potential bias	3.73 (0.80)	1 (0.9%)	6 (5.5%)	30 (27.3%)	**58 (52.7%)**	15 (13.6%)
Avoiding advocacy	3.50 (0.97)	1 (0.9%)	18 (16.4%)	33 (30.0%)	**41 (37.3%)**	17 (15.5%)
Fostering a continuing commitment to objectivity	3.45 (0.88)	1 (0.9%)	15 (13.6%)	39 (35.5%)	**44 (40.0%)**	11 (10.0%)
Being an active consumer of scientific knowledge	3.45 (0.96)	4 (3.6%)	15 (13.6%)	29 (26.4%)	**52 (47.3%)**	10 (9.1%)
Critically examining conclusions (e.g., considering alternative hypotheses)	3.38 (0.82)	2 (1.8%)	11 (10.0%)	**47 (42.7%)**	43 (39.1%)	7 (6.4%)
Limiting the scope of the inquiry and report to the referral question	3.36 (0.98)	4 (3.6%)	16 (14.5%)	38 (34.5%)	**40 (36.4%)**	12 (10.9%)
Receiving explicit didactic training about objectivity	3.32 (0.97)	3 (2.7%)	20 (18.2%)	36 (32.7%)	**41 (37.3%)**	10 (9.1%)
Taking time to think about evaluation information rather than immediately writing the report	3.29 (0.94)	5 (4.5%)	15 (13.6%)	40 (36.4%)	**43 (39.1%)**	7 (6.4%)
Resisting allegiance effects	3.27 (.99)	4 (3.6%)	18 (16.4%)	**44 (40.0%)**	32 (29.1%)	12 (10.9%)
Continuous introspection about personal biases	3.26 (0.81)	1 (0.9%)	17 (15.5%)	**49 (44.5%)**	38 (34.5%)	5 (4.5%)
Attending to wording choice in reports to edit out value-laden language	3.19 (0.92)	4 (3.6%)	19 (17.3%)	**46 (41.8%)**	34 (30.9%)	7 (6.4%)
Exposure to the importance of objectivity through reading professional literature	3.12 (0.99)	6 (5.5%)	23 (20.9%)	**40 (36.4%)**	34 (30.9%)	7 (6.4%)
Taking careful notes during an evaluation[c]	3.11 (1.04)	8 (7.3%)	21 (19.1%)	**40 (36.4%)**	33 (30.0%)	8 (7.3%)
Disengaging emotionally from cases	3.07 (0.92)	5 (4.5%)	23 (20.9%)	**45 (40.9%)**	33 (30.0%)	4 (3.6%)
Examining patterns of personal decision-making (e.g. agreement with referral party preferences)[d]	3.05 (0.96)	8 (7.3%)	18 (16.4%)	**49 (44.5%)**	30 (27.3%)	5 (4.5%)
Restricting conclusions and opinions to scientific foundations	2.89 (1.02)	11 (10.0%)	24 (21.8%)	**47 (42.7%)**	22 (20.0%)	6 (5.5%)
Intentionally controlling existing bias	2.83 (0.92)	9 (8.2%)	26 (23.6%)	**54 (49.1%)**	17 (15.5%)	4 (3.6%)
Observing others who manage their personal biases successfully	2.60 (1.02)	18 (16.4%)	32 (29.1%)	**38 (34.5%)**	20 (18.2%)	2 (1.8%)
Developing a sense of pride in one's professional identity	2.45 (1.20)	**31 (28.2%)**	26 (23.6%)	**31 (28.2%)**	16 (14.5%)	6 (5.5%)
Accepting referrals only for cases in which bias is unlikely	2.14 (0.97)	35 (31.8%)	35 (31.8%)	30 (27.3%)	10 (9.1%)	0 (0.0%)
Limiting empathy and rapport in forensic cases	1.87 (1.06)	**56 (50.9%)**	23 (20.9%)	22 (20.0%)	7 (6.4%)	2 (1.8%)

Note. Items are presented according to mean ratings arranged in descending order. Modal effectiveness ratings are shown in bold. Debiasing strategies identified by Neal and Brodsky (2016) as appearing as suggestions in scientific literature are in italics.
[a, b, c, d] Additional strategies we identified as suggested in the literature about FREs.

In addition, in the Netherlands, a strong emphasis is placed on using an SPJ approach to FREs (Nederlands Instituut voor Forensische Psychiatrie en Psychologie, 2018). Our findings indicate that SPJ tools—specifically, the HCR-20 and the HKT-30/R—are the SRAIs most commonly used among Dutch forensic evaluators. In fact, most of our respondents reported that they are required to use one of these two tools. Regardless of whether they were required or optional users, evaluators rated the HCR-20 and the HKT-30/R as moderately to highly useful. About two-thirds of our respondents also used the SAPROF and rated it as moderately to very useful.

It is common practice for forensic evaluators to use a psychopathy measure when performing a (sexual) violence risk assessment. Psychopathy is generally considered relevant in psycholegal contexts because of its relationship with risk for general criminal, violent, and sexual recidivism (DeMatteo et al., 2020). Nearly half of our respondents reported that they were required to use a psychopathy measure (PCL-R or PCL:SV) and a substantial majority of these required users rated the measure as very or extremely useful. Similarly, in their survey of forensic psychologists, Viljoen and colleagues (2010) reported that nearly 65% of respondents who

conducted any type of adult risk assessment used the PCL-R or PCL:SV.

Interestingly, the majority of our respondents indicated they were required or chose to use an actuarial (rather than SPJ) SRAI—the Static-99/99R—for evaluating the risk of sexual reoffending. This is not entirely unexpected as the Static-99/R is one of the most commonly used SRAIs for estimating the likelihood of sexual recidivism (Archer et al., 2006; Chevalier et al., 2015; Kelley et al., 2020; Neal & Grisso, 2014b). Furthermore, previous surveys have reported that actuarial SRAIs are used more often than SPJ SRAIs for evaluating sexual recidivism risk (Kelley et al., 2020; Neal & Grisso, 2014b).

Despite its widespread use and popularity, recent field studies indicate that the Static-99R can result in significant overestimates of sexual recidivism risk (Boccaccini et al., 2017). The Static-99R has also been criticized because it considers only historical risk factors that cannot be changed with treatment (Cauley, 2007; Craig et al., 2005). The Risk-Needs-Responsivity (RNR; Bonta & Andrews, 2007) model of rehabilitation for people who have committed criminal offenses indicates that addressing treatment needs requires the assessment of dynamic factors that can be targets for treatment (Bonta & Andrews, 2007; Mann et al., 2010). In fact, some respondents in our survey wrote a free-text response that said they used the Static-99/99R in combination with the ACUTE-2007/STABLE-2007 to measure static and dynamic factors and treatment needs. Some respondents also wrote that they use the Sexual Violence Risk-20 (SVR-20; Boer et al., 1997), which is the most commonly used SPJ instrument for evaluating sexual recidivism risk.

Concerns about bias and errors in forensic risk evaluations

Respondents to our survey were significantly more concerned about the possibility of a false negative than a false positive outcome in their risk evaluations. This finding suggests that Dutch forensic evaluators, like evaluators elsewhere (Bonta & Motiuk, 1990), may tend to err on the side of caution (i.e., being more averse to the potential outcome of improperly classifying someone as low risk who then reoffends than improperly classifying someone as high risk who would not have reoffended). Furthermore, erring on the side of caution can increase the likelihood of false positives, the consequence of which is that people are unjustly deprived of their freedom and limited mental health and correctional resources are needlessly wasted (Bonta & Motiuk, 1990; Harris, 2006). Of course, false positives and false negatives both carry harmful consequences and neither error is desirable. Yet it is much easier to identify a false negative because the crime the person commits is likely to come to the attention of law enforcement, and potentially, the media. On the other hand, a false positive is unlikely to be identified because the absence of reoffending is likely to be attributed to incapacitation and treatment, not a misclassification of the individual's risk.

We were also interested in how often and with whom our respondents consulted regarding FREs. We defined consulting for our respondents as "seeking advice about the evaluation," thereby leaving open the possibility of accounting for collateral interviews. Previous studies indicate that risk judgments can be more accurate than individual evaluator ratings by using a consensus method (de Vogel & de Ruiter, 2006; Huss & Zeiss, 2004; McNiel et al., 2000), and this method is commonly used in the Netherlands when conducting FREs (Harte & Breukink, 2010). Our findings confirm that it is common practice for Dutch evaluators to seek advice from others about an FRE *always* or *most of the time*. In addition, over 40% of our respondents indicated that they consult with other treatment providers of the evaluee who is the subject of the FRE.

Beyond consultations with colleagues, we also note that our respondents indicated they "consult" with other parties about FREs, including probation and parole officers, the evaluee's family, and prison staff. Because of how we defined "consulting" in our survey, the responses elicited from our respondents likely include inter-professional consultations as well as interviews with collateral informants. In fact, collateral sources of information appear fairly common in FREs. Neal and Grisso (2014b) reported that their international respondents ($N = 434$) conducted collateral interviews, both with other professionals (54.5% for FREs for violence and 25.0% for FREs for sexual violence) and with nonprofessionals (27.0% and 35.2%, respectively).

Guidelines from The Netherlands Institute of Forensic Psychiatry and Psychology indicate forensic evaluators should include relevant information in their reports, but the decision about relevant and irrelevant information is left to the subjective opinion of the evaluator (de Ruiter & Kaser-Boyd, 2015). On the one hand, collateral interviews may provide evaluators with important information needed to conduct a complete FRE. On the other hand, there is a concomitant

risk that the evaluator will be exposed to potentially biasing, irrelevant contextual information (Zapf & Dror, 2017). There is a growing body of empirical studies providing evidence that exposure to task-irrelevant contextual information can bias forensic evaluators (Dror, 2012, Dror, 2018; Dror et al., 2006, Dror et al., 2015; Kukucka & Kassin, 2014; Nakhaeizadeh et al., 2014) and should be avoided. In fact, avoiding potentially biasing, irrelevant information may be one of the most effective strategies to prevent bias before it occurs (Dror et al., 2015; Gardner et al., 2019; National Research Council, 2009; Wilson & Brekke, 1994).

Potential sources of bias in forensic evaluations

Similar to Zapf et al. (2018) findings, our respondents were inclined to agree that an evaluator's prior beliefs and expectations can affect how they analyze a forensic case and formulate their ultimate opinion, suggesting some awareness of confirmation bias among our respondents (Kassin et al., 2013). Nevertheless, less than half of our respondents viewed bias as a problem in FREs. This contrasts with nearly 86% of forensic evaluators in Zapf et al.'s survey who expressed concern about cognitive bias in forensic evaluations as a whole, and about 79% of whom said they were concerned about cognitive bias in their specific domain of forensic evaluations.

Furthermore, a substantial majority of our respondents agreed that a conscious effort to set aside prior beliefs or expectations makes it less likely an evaluator will be influenced by them. Yet, conscious efforts to set aside prior beliefs and expectations are unlikely to be effective in eliminating bias, because bias operates outside of awareness (Wilson & Brekke, 1994). Furthermore, even when an evaluator is aware of the potential for preexisting motivations and emotions to affect their evaluation, efforts to counteract them are not necessarily effective (Kassin et al., 2013).

Perhaps our respondents were not as concerned about bias in FREs because they commonly used SRAIs (de Ruiter, 2016; de Ruiter & Hildebrand, 2007). As applied to FREs, SRAIs may help reduce the effects of bias, although a growing body of evidence suggests this is not always the case (Chappell et al., 2013; Gowensmith & McCallum, 2019; Guay & Parent, 2018; Guy et al., 2014; Miller & Maloney, 2013; Murrie et al., 2008, 2009; Murrie & Balusek, 2008; Schmidt et al., 2016; Shepherd & Sullivan, 2017; Storey et al., 2012; Wormith et al., 2012). Therefore, evaluators should still remain aware of the potential for bias if they are to take steps to effectively minimize its effects on FREs (Dror, 2018; Lilienfeld et al., 2009; Wilson & Brekke, 1994).

Perceived effectiveness of debiasing strategies

Neal and Brodsky (2016) identified six debiasing strategies that have previously been suggested as potentially effective debiasing techniques for use in forensic evaluations. These strategies are: using structured methods to gather and analyze data (Croskerry et al., 2013; Graber et al., 2012; Neal & Grisso, 2014b; Zapf & Dror, 2017), consulting with colleagues (Croskerry et al., 2013; Graber et al., 2012), critically examining conclusions (e.g., considering alternative hypotheses; Galinsky et al., 2000; Galinsky & Moskowitz, 2000; Grisso, 2010; Lord et al., 1984; Mumma & Wilson, 1995; Soll et al., 2015; Zapf & Dror, 2017), receiving explicit didactic training about objectivity (Bridge & Marić, 2019; Croskerry et al., 2013; Graber et al., 2012; Soll et al., 2015), taking time to think about the evaluation before writing the report (Croskerry et al., 2013; Lilienfeld et al., 2009), and reading professional literature about the importance of objectivity (Croskerry et al., 2013). Our respondents gave significantly higher effectiveness ratings to the six debiasing strategies with some empirical support than they did to the remaining 19 strategies. However, we note that the empirical support for several of these 'effective' debiasing strategies is derived from medical research related to reducing errors and improving accuracy in clinical and diagnostic decision-making. Therefore, we cannot say with certainty that the debiasing strategies suggested in other fields will be effective in mitigating bias in FREs (Fischhoff, 1982; Soll et al., 2015).

In addition to the six strategies identified by Neal and Brodsky (2016) as having been suggested in scientific literature, we think there are at least four others (from the 25 listed) that have been suggested as potentially effective in countering bias specific to forensic evaluations. These four strategies are: investigating all relevant data before forming an opinion (Grisso, 2010; Zapf & Dror, 2017), basing conclusions and opinions on sound data (Grisso, 2010; Zapf & Dror, 2017), taking careful notes during an evaluation (Arkes, 1981; Borum et al., 1993; Mumma & Wilson, 1995), and examining patterns of personal decision-making (e.g., agreement with referral party preferences; comparing one's decisions over time to base rates; Brodsky, 2013; DeClue & Rice, 2016; Gowensmith & McCallum, 2019; Murrie & Balusek, 2008; Murrie & Warren, 2005; Parker, 2016). However, we note that

very few strategies to mitigate bias have been empirically tested. Therefore, field and task-specific research regarding the effectiveness of potential strategies to mitigate bias in FREs would be of significant benefit for the discipline, forensic examinees, and legal decision-makers.

Limitations and conclusions

There are a number of limitations to our findings that relate to our sample. First, we note the relatively small sample size of our study ($N = 110$), although the sample size is in line with previous similar surveys conducted in other countries (e.g., Archer et al., 2006; Hill & Demetrioff, 2019; Viljoen et al., 2010). Second, practitioners who chose to participate in the survey may have done so because they possess more awareness and/or concern about the potential for bias than evaluators who chose not to participate, thus introducing (self-)selection bias. Third, eight of our respondents had not attained a graduate degree, which may have limited the types of risk assessment tools they were qualified or trained to use. Finally, we are aware of one previous international survey by Singh and colleagues (2014) that differentiated between forensic psychologists' and forensic psychiatrists' use of violence SRAIs. Their findings indicated that forensic psychologists may use violence SRAIs more frequently than forensic psychiatrists. However, we did not ask our respondents to specify their discipline. Future researchers may consider explicitly differentiating the use and perceived utility of SRAIs between these disciplines and others (e.g., social workers, nurses). For these reasons, our sample may not be representative of all forensic mental health practitioners who conduct FREs in the Netherlands.

In addition, the actual response rate to our survey is unknown, and any conclusions about the broader population of forensic mental health practitioners in the Netherlands are therefore tentative. Furthermore, we do not know the context in which our respondents are performing FREs: for example, are they conducting FREs for the court, for the purpose of treatment planning and risk management, for decisions related to patient restrictions? The context in which an FRE is undertaken may have an effect on SRAI use and how concerns about potential bias are managed. Therefore, our findings may not generalize across all forensic contexts and purposes for which FREs are conducted.

Our survey results are also limited with respect to the use of sexual recidivism SRAIs. Unfortunately, we failed to include SPJ instruments for sexual violence, such as the SVR-20, in our survey list of SRAI options. Still, a number of respondents included the SVR-20 in free text, but we cannot exclude the possibility that more respondents would have selected the SVR-20 if it had been listed as an option.

Finally, we cannot rule out the possibility of socially desirable responding. Although the survey was completed anonymously, and participants were assured of anonymity, it is possible that some people were concerned about negative portrayals of their profession, their work, or their employer. Therefore, it is possible that there is a disparity between what our respondents say they do in practice and what they actually do.

The findings from this study also point to potential areas for future research. For example, we did not ask participants about "irrelevant contextual information" as in Zapf et al.'s (2017) study, because of the lack of consensus about what constitutes irrelevant contextual information in risk evaluations. A recent survey of forensic scientists ($N = 189$) from several forensic disciplines (biology, pattern evidence, chemistry, and crime scene investigation) is illustrative of the challenges in obtaining agreement regarding what constitutes irrelevant contextual information in their tasks (Gardner et al., 2019). Gardner and colleagues reported that among crime scene investigators, a substantial majority agreed that a description of the evidence, how it was collected, the type of offense, and a synopsis of the case were essential to their tasks. Conversely, 25% said that a suspect's statement or confession was essential, whereas greater than half said they would review that information if it was available (58.3% and 66.7%, respectively). A small minority agreed that the suspect's statement or confession was irrelevant (16.7% and 8.3%, respectively).

Similar to crime scene investigation, the process of a forensic risk evaluation involves gathering information and evidence. Information that is deemed to be relevant in that process is likely to vary considerably between cases and individual experts. Given the potential for irrelevant contextual information to introduce bias into FREs (Neal & Saks, 2016; Zapf & Dror, 2017), surveying forensic mental health professionals about their views of what information is irrelevant to the FRE task seems to be an important avenue for future research. Only within the past decade or so have researchers begun to investigate the role that cognitive bias may play in FREs (Camilleri et al., in press; Charman, 2013). Furthermore, many potentially effective debiasing techniques have not been

empirically tested with respect to forensic evaluations, or specifically FREs. This point is worth noting because not all debiasing strategies are appropriate and/or effective across domains or tasks (Soll et al., 2015). It is not necessarily the case that a debiasing strategy that is effective in countering one type of bias (e.g., confirmation bias) will be effective in countering a different type of bias (Fischhoff, 1982). Therefore, direct investigations on the effect structured evaluation methods on countering bias in FREs may shed light on the debiasing potential of these methods.

Conflict of interest

We have no conflicts of interest to disclose.

Funding

ORCID

Jennifer Kamorowski http://orcid.org/0000-0001-6990-6044
Corine de Ruiter http://orcid.org/0000-0002-0135-9790
Maartje Schreuder http://orcid.org/0000-0003-2004-701X
Karl Ask http://orcid.org/0000-0001-5093-5902
Marko Jelícic http://orcid.org/0000-0002-7966-0540

Data availability statement

The data that support the findings of this study are available from the corresponding author, JK, upon reasonable request.

References

Adviesbureau Van Montfoort & Reclassering Nederland. (2004). *RISc versie 1.0. Recidive Inschattings Schalen.* Handleiding [RISc version 1.0. Risk Assessment Scales. Manual]. Flevodruk.

AEgisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34(3), 341–382. https://doi.org/10.1177/0011000005285875

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime & Delinquency*, 52(1), 7–27. https://doi.org/10.1177/0011128705281756

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment*, 87(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07

Arkes, H. R. (1981). Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology*, 49(3), 323–330. https://doi.org/10.1037/0022-006X.49.3.323

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28(4), 490–521. https://doi.org/10.1177/009385480102800406

Beckham, J. C., Annis, L. V., & Gustafson, D. J. (1989). Decision making and examiner bias in forensic expert recommendations for not guilty by reason of insanity. *Law and Human Behavior*, 13(1), 79–87. https://doi.org/10.1007/BF01056164

Boccaccini, M. T., Chevalier, C. S., Murrie, D. C., & Varela, J. G. (2017). Psychopathy Checklist-Revised Use and Reporting Practices in Sexually Violent Predator Evaluations. *Sexual Abuse : a Journal of Research and Treatment*, 29(6), 592–614. https://doi.org/10.1177/1079063215612443

Boccaccini, M. T., Rice, A. K., Helmus, L. M., Murrie, D. C., & Harris, P. B. (2017). Field validity of Static-99/R scores in a statewide sample of 34,687 convicted sexual offenders. *Psychological Assessment*, 29(6), 611–623. https://doi.org/10.1037/pas0000377

Boccaccini, M. T., Turner, D. B., & Murrie, D. C. (2008). Do some evaluators report consistently higher or lower PCL-R scores than others? Findings from a statewide sample of sexually violent predator evaluations. *Psychology, Public Policy, and Law*, 14(4), 262–283. https://doi.org/10.1037/a0014523

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence.* British Columbia Institute Against Family Violence.

Bogaerts, S., Spreen, M., ter Horst, P., & Gerlsma, C. (2018). Predictive validity of the HKT-R risk assessment tool: Two and 5-year violent recidivism in a nationwide sample of Dutch forensic psychiatric patients. *International Journal of Offender Therapy and Comparative Criminology*, 62(8), 2259–2270. https://doi.org/10.1177/0306624X17717128

Bonta, J., Andrews, D. A. (2007). *Risk-Need-Responsivity model for offender assessment and rehabilitation* (No. 2007–06). https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/rsk-nd-rspnsvty/rsk-nd-rspnsvty-eng.pdf

Bonta, J., & Motiuk, L. L. (1990). Classification to halfway houses: A quasi-experimental evaluation. *Criminology*, 28(3), 497–506. https://doi.org/10.1111/j.1745-9125.1990.tb01336.x

Borum, R., Otto, R., & Golding, S. (1993). Improving clinical judgment and decision making in forensic evaluation. *The Journal of Psychiatry & Law*, 21(1), 35–76. https://doi.org/10.1177/009318539302100104

Borum, R., Bartel, P., & Forth, A. E. (2002). *Manual for the structured assessment of violence risk in youth (SAVRY).* University of South Florida.

Bridge, C., & Marić, M. (2019). Cognitive biases in forensic science training and education. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and Fake News in Education*. Information Age Publishing.

Brodsky, S. L. (2013). The pull to affiliate and allegiance effects. In *Testifying in court: Guidelines and maxims for the expert witness*. (2012-19379-039; 2nd ed., pp. 161–165). American Psychological Association. https://doi.org/10.1037/14037-039

Brown, B., & Rakow, T. (2016). Understanding Clinicians' Use of Cues When Assessing the Future Risk of Violence: A Clinical Judgement Analysis in the Psychiatric Setting. *Clinical Psychology & Psychotherapy*, 23(2), 125–141. https://doi.org/10.1002/cpp.1941

Camilleri, A., Abarno, D., Bird, C., Coxon, A., Mitchell, N., Redman, K., Sly, N., Wills, S., Silenieks, E., Simpson, E., & Lindsay, H. (in press). A risk-based approach to cognitive bias in forensic science. *Science & Justice : Justice*, 59(5), 533–543. S1355030618303472. https://doi.org/10.1016/j.scijus.2019.04.003

Campbell, M. A., French, S., & 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

Cauley, D. R. (2007). Risk assessment of civilly committed sexually violent predators: Static versus dynamic orientations. *Forensic Examiner*, 16(1), 18–24.

Chappell, A. T., Maggard, S. R., & Higgins, J. L. (2013). Exceptions to the rule? Exploring the use of overrides in detention risk assessment. *Youth Violence and Juvenile Justice*, 11(4), 332–348. https://doi.org/10.1177/1541204012470849

Charman, S. D. (2013). The forensic confirmation bias: A problem of evidence integration, not just evidence evaluation. *Journal of Applied Research in Memory and Cognition*, 2(1), 56–58. https://doi.org/10.1016/j.jarmac.2013.01.010

Chevalier, C. S., Boccaccini, M. T., Murrie, D. C., & Varela, J. G. (2015). Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law and Human Behavior*, 39(3), 209–218. https://doi.org/10.1037/lhb0000114

Craig, L. A., Browne, K. D., Stringer, I., & Beech, A. (2005). Sexual recidivism: A review of static, dynamic and actuarial predictors. *Journal of Sexual Aggression*, 11(1), 65–84. https://doi.org/10.1080/13552600410001667733

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 2: Impediments to and strategies for change. *BMJ Quality & Safety*, 22(Suppl 2), ii65–ii72. https://doi.org/10.1136/bmjqs-2012-001713

Curley, L. J., Munro, J., Lages, M., MacLean, R., & Murray, J. (2019). Assessing cognitive bias in forensic decisions: A review and outlook. *Journal of Forensic Sciences*, https://doi.org/10.1111/1556-4029.14220

Curley, L. J., Munro, J., Lages, M., MacLean, R., & Murray, J. (2020). Authors' response. *Journal of Forensic Sciences, Advance Online Publication*, https://doi.org/10.1111/1556-4029.14273

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science (New York, N.Y.)*, 243(4899), 1668–1674. https://doi.org/10.1126/science.2648573

de Ruiter, C. (2016). Risk assessment in the Netherlands. In J. P. Singh, S. Bjørkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment*. (pp. 270–279). Oxford University Press.

de Ruiter, C., & Hildebrand, M. (2007). Risk assessment and treatment in Dutch forensic psychiatry. *Netherlands Journal of Psychology*, 63(4), 152–160. https://doi.org/10.1007/BF03061078

de Ruiter, C., & Kaser-Boyd, N. (2015). *Forensic psychological assessment in practice: Case studies.* (1st ed.). Routledge. https://doi.org/10.4324/9781315723310

de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law*, 12(3), 321–336. https://doi.org/10.1080/10683160600569029

de Vries Robbé, M., de Vogel, V., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry & Psychology*, 24(4), 440–457. https://doi.org/10.1080/14789949.2013.818162

de Vogel, V., de Ruiter, C., Bouman, Y. H. A., & de Vries Robbé, M. (2009). SAPROF. *Guidelines for the assessment of protective factors for violence risk*. Forum Educatief.

de Vogel, V., de Ruiter, C., Bouman, Y. H. A., & de Vries Robbé, M. (2012). SAPROF. *Guidelines for the assessment of protective factors for violence risk* (2nd ed.). Forum Educatief.

DeClue, G., & Rice, A. (2016). Florida's released "Sexually Violent Predators" are not "high risk. *Open Access Journal of Forensic Psychology*, 8, 22–51.

DeMatteo, D., Hart, S. D., Heilbrun, K., Boccaccini, M. T., Cunningham, M. D., Douglas, K. S., Dvoskin, J. A., Edens, J. F., Guy, L. S., Murrie, D. C., Otto, R. K., Packer, I. K., & Reidy, T. J. (2020). Statement of concerned experts on the use of the Hare Psychopathy Checklist—Revised in capital sentencing to assess risk for institutional violence. *Psychology, Public Policy, and Law*, 26(2), 133–144. https://doi.org/10.1037/law0000223

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior*, 29(5), 617–658. https://doi.org/10.1177/009385402236735

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20^{v3}): Development and overview. *International Journal of Forensic Mental Health*, 13(2), 93–108. https://doi.org/10.1080/14999013.2014.906519

Doyle, M., & Dolan, M. (2007). Standardized risk assessment. *Psychiatry*, 6(10), 409–414. https://doi.org/10.1016/j.mppsy.2007.07.004

Dror, I. E. (2012). Expectations, contextual information, and other cognitive influences in forensic laboratories. *Science & Justice*, 52(2), 132. https://doi.org/10.1016/j.scijus.2012.03.005

Dror, I. E. (2018). Biases in forensic experts. *Science (New York, N.Y.)*, 360(6386), 243 https://doi.org/10.1126/science.aat8443

Dror, I. E., & Murrie, D. C. (2018). A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law*, 24(1), 11–23. https://doi.org/10.1037/law0000140

Dror, I. E., Charlton, D., & Péron, A. E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. *Forensic Sci Int*, 156(1), 74–78. https://doi.org/10.1016/j.forsciint.2005.10.017

Dror, I. E., Thompson, W. C., Meissner, C. A., Kornfield, I., Krane, D., Saks, M., & Risinger, M. (2015). Letter to the Editor- Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making. *Journal of Forensic Sciences*, 60(4), 1111–1112. https://doi.org/10.1111/1556-4029.12805

Expertisecentrum Forensische Psychiatrie. (n.d.). About EFP. efp.nl/en/about-efp (accessed February 28, 2021).

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* (pp. 422–444). Cambridge University Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Counterfactuals as behavioral primes: Priming the simulation heuristic and consideration of alternatives. *Journal of Experimental Social Psychology*, 36(4), 384–409. https://doi.org/10.1006/jesp.1999.1409

Galinsky, A. D., Moskowitz, G. B., & Skurnik, I. (2000). Counterfactuals as self-generated primes: The effect of prior counterfactual activation on person perception judgments. *Social Cognition*, 18(3), 252–280. https://doi.org/10.1521/soco.2000.18.3.252

Gardner, B. O., Kelley, S., Murrie, D. C., & Dror, I. E. (2019). What do forensic analysts consider relevant to their decision making? *Science & Justice : journal of the Forensic Science Society*, 59(5), 516–523. https://doi.org/10.1016/j.scijus.2019.04.005

Gowensmith, W. N., & McCallum, K. E. (2019). Mirror, mirror on the wall, who's the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations. *South African Journal of Psychology*, 49(2), 165–176. https://doi.org/10.1177/0081246319835117

Gowensmith, W. N., Bryant, A. E., & Vitacco, M. J. (2014). Decision-making in post-acquittal hospital release: How do forensic evaluators make their decisions? *Behavioral Sciences & the Law*, 32(5), 596–607. https://doi.org/10.1002/bsl.2135

Gowensmith, W. N., Murrie, D. C., & Boccaccini, M. T. (2012). Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior*, 36(2), 130–139. https://doi.org/10.1037/h0093958

Graber, M. L., Kissam, S., Payne, V. L., Meyer, A. N. D., Sorensen, A., Lenfestey, N., Tant, E., Henriksen, K., LaBresh, K., & Singh, H. (2012). Cognitive interventions to reduce diagnostic error: A narrative review. *BMJ Quality & Safety*, 21(7), 535–557. https://doi.org/10.1136/bmjqs-2011-000149

Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology*, 2, 102–115.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293

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(1), 19–30. https://doi.org/10.1037//1040-3590.12.1.19

Guarnera, L. A., Murrie, D. C., & Boccaccini, M. T. (2017). Why do forensic experts disagree? Sources of unreliability and bias in forensic psychology evaluations. *Translational Issues in Psychological Science*, 3(2), 143–152. https://doi.org/10.1037/tps0000114

Guay, J. P., & Parent, G. (2018). Broken legs, clinical overrides, and recidivism risk: An analysis of decisions to adjust risk levels with the LS/CMI. *Criminal Justice and Behavior*, 45(1), 82–100.

Guy, L. S., Nelson, R. J., Fusco-Morin, S. L., & Vincent, G. M. (2014). What do juvenile probation officers think of using the SAVRY and YLS/CMI for case management, and do they use the instruments properly? *International Journal of Forensic Mental Health*, 13(3), 227–241. https://doi.org/10.1080/14999013.2014.939789

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

Hanson, R. K., & Thornton, D. (1999). *Static 99: Improving actuarial risk assessments for sex offenders.* (No. 1999–02). Solicitor General Canada.

Hanson, R. K., Harris, A. J. R., Scott, T.-L., Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (Corrections Research User Report No. 2007-05). Public Safety Canada. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf

Hanson, R. K., Helmus, L.-M., & Harris, A. J. R. (2015). Assessing the risk and needs of supervised sexual offenders: A prospective study using STABLE-2007, Static-99R, and Static-2002R. *Criminal Justice and Behavior*, 42(12), 1205–1224. https://doi.org/10.1177/0093854815602094

Hare, R. D. (2003). *Manual for the revised psychopathy checklist* (2nd ed.). Multi-Health Systems.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, 20(4), 315–335.

Hart, S. D., Cox, D. N., & Hare, R. D. (1995). *The Hare psychopathy checklist: Screening version.* Multi-Health Systems.

Harris, P. M. (2006). What community supervision officers need to know about actuarial risk assessment and clinical judgment. *Federal Probation*, 70(2), 8–14.

Harte, J. M., & Breukink, M. D. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie [Objectivity or apparent security? Quality, possibilities and limitations of instruments for risk assessment. *Tijdschrift Voor Criminologie*, 52(1), 52–72.

Hartvig, P., Ostberg, B., Alfarnes, S., Moger, T. A., Skjønberg, M., & Bjørkly, S. (2007). *Violent risk screen 10 (V-RISK-10).* www.forensicpsychiatry

Hawes, S. W., Boccaccini, M. T., & Murrie, D. C. (2013). Psychopathy and the combination of psychopathy and sexual deviance as predictors of sexual recidivism: Meta-analytic findings using the Psychopathy Checklist-

Revised. *Psychological Assessment*, 25(1), 233–243. https://doi.org/10.1037/a0030391

Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse : a Journal of Research and Treatment*, 24(1), 64–101. https://doi.org/10.1177/1079063211409951

Hill, D., & Demetrioff, S. (2019). Clinical-forensic psychology in Canada: A survey of practitioner characteristics, attitudes, and psychological assessment practices. *Canadian Psychology/Psychologie Canadienne*, 60(1), 55–63. https://doi.org/10.1037/cap0000152

Hurducas, C. C., Singh, J. P., de Ruiter, C., & Petrila, J. (2014). Violence risk assessment tools: A systematic review of surveys. *International Journal of Forensic Mental Health*, 13(3), 181–192. https://doi.org/10.1080/14999013.2014.942923

Huss, M. T., & Zeiss, R. A. (2004). Clinical assessment of violence from inpatient records: A comparison of individual and aggregate decision making across risk strategies. *International Journal of Forensic Mental Health*, 3(2), 139–147. https://doi.org/10.1080/14999013.2004.10471203

Jeanguenat, A. M., Budowle, B., & Dror, I. E. (2017). Strengthening forensic DNA decision making through a better understanding of the influence of cognitive bias. *Science & Justice : Justice*, 57(6), 415–420. https://doi.org/10.1016/j.scijus.2017.07.005

Kassin, S. M., Dror, I. E., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solutions. *Journal of Applied Research in Memory and Cognition*, 2(1), 42–52. https://doi.org/10.1016/j.jarmac.2013.01.001

Kelley, S. M., Ambroziak, G., Thornton, D., & Barahal, R. M. (2020). How do professionals assess sexual recidivism risk? An updated survey of practices. *Sexual Abuse : a Journal of Research and Treatment*, 32(1), 3–29. https://doi.org/10.1177/1079063218800474

Kropp, P. R., Hart, S. D., & Belfrage, H. (2005). *The brief spousal assault form for the evaluation of risk (BSAFER): User manual*. ProActive ReSolutions, Inc.

Kukucka, J., & Kassin, S. M. (2014). Do confessions taint perceptions of handwriting evidence? An empirical test of the forensic confirmation bias. *Law and Human Behavior*, 38(3), 256–270. https://doi.org/10.1037/lhb0000066

Kukucka, J., Kassin, S. M., Zapf, P. A., & Dror, I. E. (2017). Cognitive bias and blindness: A global survey of forensic science examiners. *Journal of Applied Research in Memory and Cognition*, 6(4), 452–459. https://doi.org/10.1016/j.jarmac.2017.09.001

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice*, 34(5), 491–498. https://doi.org/10.1037/0735-7028.34.5.491

Lanterman, J. L., Boyle, D. J., & Ragusa-Salerno, L. M. (2014). Sex offender risk assessment, sources of variation, and the implications of misuse. *Criminal Justice and Behavior*, 41(7), 822–843. https://doi.org/10.1177/0093854813515237

Leij, J. B. J. v d., Jackson, J. L., Malsch, M., & Nijboer, J. F. (2001). Residential mental health assessment within Dutch criminal cases: A discussion. *Behavioral Sciences & the Law*, 19(5-6), 691–702. https://doi.org/10.1002/bsl.465

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science : a Journal of the Association for Psychological Science*, 4(4), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243. https://doi.org/10.1037/0022-3514.47.6.1231

MacLean, N., Neal, T. M. S., Morgan, R. D., & Murrie, D. C. (2019). Forensic clinicians' understanding of bias. *Psychology, Public Policy, and Law*, 25(4), 323–330. https://doi.org/10.1037/law0000212

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse : a Journal of Research and Treatment*, 22(2), 191–217. https://doi.org/10.1177/1079063210366039

McLaughlin, J. L., & Kan, L. Y. (2014). Test usage in four common types of forensic mental health assessment. *Professional Psychology: Research and Practice*, 45(2), 128–135. https://doi.org/10.1037/a0036318

McNiel, D. E., Lam, J. N., & Binder, R. L. (2000). Relevance of interrater agreement to violence risk assessment. *Journal of Consulting and Clinical Psychology*, 68(6), 1111–1115. https://doi.org/10.1037/0022-006X.68.6.1111

Messina, E., Ferracuti, S., Nicolò, G., Ruggeri, M., Kooijmans, T., & Meynen, G. (2019). Forensic psychiatric evaluations of defendants: Italy and the Netherlands compared. *International Journal of Law and Psychiatry*, 66, 101473 https://doi.org/10.1016/j.ijlp.2019.101473

Miller, J., & Maloney, C. (2013). Practitioner compliance with risk/needs assessment tools: A theoretical and empirical assessment. *Criminal Justice and Behavior*, 40(7), 716–736. https://doi.org/10.1177/0093854812468883

Monahan, J. (1981). *The clinical prediction of violent behavior*. U. S. Department of Health and Human Services.

Monahan, J., Steadman, H. J., Appelbaum, P. S., Grisso, T., Mulvey, E. P., Roth, L. H., Robbins, P. C., Banks, S., & Silver, E. (2006). The classification of violence risk. *Behavioral Sciences and the Law*, 24(6), 721–730. https://doi.org/10.4135/9781412959537.n35

Mumma, G. H., & Wilson, S. B. (1995). Procedural debiasing of primacy/anchoring effects in clinical-like judgments. *Journal of Clinical Psychology*, 51(6), 841–853. https://doi.org/10.1002/1097-4679(199511)51:6<841::AID-JCLP2270510617>3.0.CO;2-K

Murrie, D. C., & Balusek, K. (2008). Forensic assessment of violence risk in adversarial proceedings: Pursuing objectivity and avoiding bias. *Journal of Forensic Psychology Practice*, 7(4), 141–153. https://doi.org/10.1300/J158v07n04_06

Murrie, D. C., & Warren, J. I. (2005). Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Professional Psychology: Research and Practice*, 36(5), 519–524. https://doi.org/10.1037/0735-7028.36.5.519

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A., & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychol Sci*, 24(10), 1889–1897. https://doi.org/10.1177/0956797613481812

Murrie, D. C., Boccaccini, M. T., Johnson, J. T., & Janke, C. (2008). Does interrater (dis)agreement on Psychopathy

Checklist scores in sexually violent predator trials suggest partisan allegiance in forensic evaluations? *Law and Human Behavior*, 32(4), 352–362. https://doi.org/10.1007/s10979-007-9097-5

Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law*, 15(1), 19–53. https://doi.org/10.1037/a0014897

Nakhaeizadeh, S., Dror, I. E., & Morgan, R. M. (2014). Cognitive bias in forensic anthropology: Visual assessment of skeletal remains is susceptible to confirmation bias. *Science & Justice : journal of the Forensic Science Society*, 54(3), 208–214. https://doi.org/10.1016/j.scijus.2013.11.003

National Research Council (2009). (Ed.). *Strengthening forensic science in the United States: A path forward.* National Academies Press.

Neal, T. M. S., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, 22(1), 58–76. https://doi.org/10.1037/law0000077

Neal, T. M. S., & Grisso, T. (2014a). The cognitive underpinnings of bias in forensic mental health evaluations. *Psychology, Public Policy, and Law*, 20(2), 200–211. https://doi.org/10.1037/a0035824

Neal, T. M. S., & Grisso, T. (2014b). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, 41(12), 1406–1421. https://doi.org/10.1177/0093854814548449

Neal, T. M. S., & Saks, M. J. (2016). *The science of science offers solutions to cognitive bias in forensic psychology.* American Psychology-Law Society Annual Conference.

Nederlands Instituut voor Forensische Psychiatrie en Psychologie (2018). NIFP-richtlijn: Ambulant forensisch psychologisch onderzoek en rapportage in het strafrecht. https://www.nifp.nl/binaries/richtlijn-ambulant-forensisch-onderzoek-en-rapportage-in-strafzaken_tcm106-351968.pdf

Nielsen, L. H., Mastrigt, S. v., Otto, R. K., Seewald, K., Ruiter, C. d., Rettenberger, M., Reeves, K. A., Rebocho, M. F., Pham, T. H., Mei Yee Ho, R., Grann, M., Godoy-Cervera, V., Folino, J. O., Doyle, M., Desmarais, S. L., Condemarin, C., Arbach-Lucioni, K., & Singh, J. P. (2015). Violence risk assessment practices in Denmark: A multidisciplinary national survey. *Scandinavian Journal of Forensic Science*, 21(2), 103–110. 0(0). https://doi.org/10.1515/sjfs-2015-0003

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259. [Database] https://doi.org/10.1037/0033-295X.84.3.231

Parker, G. (2016). Come see the bias inherent in the system!. *The Journal of the American Academy of Psychiatry and the Law*, 44(4), 411–414.

Pham, T., Ducro, C., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald,

K., Otto, R. K., & Singh, J. P. (2016). Enquête internationale sur les pratiques d'évaluation du risque de violence: Présentation des données belges [International survey of violence risk assessment practices: Presentation of Belgian data. *Annales Médico-Psychologiques, Revue Psychiatrique*, 174(7), 539–543. https://doi.org/10.1016/j.amp.2015.10.018

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43(4), 565–578. https://doi.org/10.1016/j.jesp.2006.05.011

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28(3), 369–381. https://doi.org/10.1177/0146167202286008

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* American Psychological Association.

Rettenberger, M., Rice, M. E., Harris, G. T., & Eher, R. (2017). Actuarial risk assessment of sexual offenders: The psychometric properties of the Sex Offender Risk Appraisal Guide (SORAG). *Psychological Assessment*, 29(6), 624–638. https://doi.org/10.1037/pas0000390

Risicotaxie, W. P. (2002). *Findings of a nationwide pilot study on the HKT-30.* Ministerie van Justitie.

Schmidt, F., Sinclair, S. M., & Thomasdóttir, S. (2016). Predictive validity of the Youth Level of Service/Case Management Inventory with youth who have committed sexual and non-sexual offenses: The utility of professional override. *Criminal Justice and Behavior*, 43(3), 413–430. https://doi.org/10.1177/0093854815603389

Shepherd, S. M., & Sullivan, D. (2017). Covert and implicit influences on the interpretation of violence risk instruments. *Psychiatry, Psychology, and Law : An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, 24(2), 292–301. https://doi.org/10.1080/13218719.2016.1197817

Singh, J. P., Assy, R., & Serpa, K. I. (2019). Violence risk assessment practices in Israel: A preliminary survey investigation. *Journal of Aggression, Conflict and Peace Research*, 11(2), 116–126. https://doi.org/10.1108/JACPR-05-2018-0358

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Nielsen, L. H., Pham, T. H., Rebocho, M. F., Reeves, K. A., Rettenberger, M., de Ruiter, C., Seewald, K., & Otto, R. K. (2014). International perspectives on the practical application of violence risk assessment: A global survey of 44 countries. *International Journal of Forensic Mental Health*, 13(3), 193–206. https://doi.org/10.1080/14999013.2014.922141

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making.* (pp. 924–951). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118468333.ch33

Spreen, M., Brand, E., Ter Horst, P., & Bogaerts, S. (2014). *Handleiding en Methodologische Verantwoording HKT-R, Historische, Klinische en Toekomstige – Revisie [Guidelines*

and Methodological Research of the HKT-R, Historical, Clinical and Future – Revision]. Dr. van Mesdag kliniek.

Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. Sexual Abuse : a Journal of Research and Treatment, 24(3), 289–302. https://doi.org/10.1177/1079063211423943

Thompson, W. C. (2020). Assessing cognitive bias in forensic decisions: A review and outlook. Journal of Forensic Sciences, 65(2), 666–667. Curley, L. J., Munro, J., Lages, M., MacLean, R., & Murray, J. https://doi.org/10.1111/1556-4029.14272

Urbaniok, F. (2007). FOTRES: Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System [FOTRES: Forensic Operationalized Therapy/Risk Evaluation System]. Zytglogge.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, New Series, 185(4157), 1124–1131.

van Horn, J., Eisenberg, M., & Uzieblo, K. (2016). Risicotaxatie in de Nederlandse ambulante forensische geestelijke gezondheidszorg. Tijdschrift Voor Psychiatrie, 58, 583–592.

Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. Assessment, 17(3), 377–395. https://doi.org/10.1177/1073191109359587

Webster, C. D., Douglas, K. S., Eaves, D. C., & Hart, S. D. (1997). HCR-20: Assessing risk for violence (version 2). Mental Health Law and Policy Institute.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). Manual for the short-term assessment of risk and treatability (START). St. Joseph's Healthcare, Hamilton and British Columbia Mental Health and Addiction Services.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. Psychological Bulletin, 116(1), 117–142. https://doi.org/10.1037/0033-2909.116.1.117

Wong, S. C. P., & Gordon, A. (2006). The validity and reliability of the violence risk scale: A treatment-friendly violence risk assessment tool. Psychology, Public Policy, and Law, 12(3), 279–309. https://doi.org/10.1037/1076-8971.12.3.279

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The predictive validity of a general risk/needs assessment inventory on sexual offender recidivism and an exploration of the professional override. Criminal Justice and Behavior, 39(12), 1511–1538. https://doi.org/10.1177/0093854812455741

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. Psychological Bulletin, 136(5), 740–767. https://doi.org/10.1037/a0020473

Zapf, P. A., & Dror, I. E. (2017). Understanding and mitigating bias in forensic evaluation: Lessons from forensic science. International Journal of Forensic Mental Health, 16(3), 227–238. https://doi.org/10.1080/14999013.2017.1317302

Zapf, P. A., Kukucka, J., Kassin, S. M., & Dror, I. E. (2018). Cognitive bias in forensic mental health assessment: Evaluator beliefs about its nature and scope. Psychology, Public Policy, and Law, 24(1), 1–10. https://doi.org/10.1037/law0000153

Zappala, M., Reed, A. L., Beltrani, A., Zapf, P. A., & Otto, R. K. (2018). Anything you can do, I can do better: Bias awareness in forensic evaluators. Journal of Forensic Psychology Research and Practice, 18(1), 45–56. https://doi.org/10.1080/24732850.2017.1413532