

Factors related to bias in forensic psychiatric assessments in criminal matters: A systematic review

Leonardo Fernandez Meyer^{*}, Alexandre Martins Valença

Institute of Psychiatry, Federal University of Rio de Janeiro (UFRJ), Brazil

ARTICLE INFO

Keywords:

Criminal responsibility
Legal capacities
Assessment
Bias
Decision making
Opinion
Psychometric assessment
Insanity defense

ABSTRACT

Objective: Identify factors related to bias in forensic psychiatric assessments in criminal matters.

Method: Based on the PRISMA guidelines, we searched the following keywords with Boolean operators: (criminal responsibility OR legal responsibility OR neurolaw OR insanity defense) AND (forensic psychiatry OR assessment OR evaluation OR bias OR decision-making OR capacity OR psychometric). The search included publications from January 1998 to December 2019 in the English language, published in PubMed, Web of Science, Taylor & Francis, and Scopus databases.

Results: The final sample consisted of 30 articles separated into three groups: (1) legal elements and the wording of expert reports, (2) psychometric tools applied to criminal inquiries, and (3) expert forensic technique and inter-examiner agreement.

Discussion: Multiple factors for biases were identified: difficulties in equivalence between legal and psychiatric terminologies, elements of countertransference between the expert and the examinee, absence of standardization of expert evaluations, low quality of expert reports, differences in the training of professionals involved in the evaluations, use of psychometric tools, number of professionals working on the same case, and the methodology adopted. Psychometric tools developed specifically for forensic psychiatric evaluations allowed the introduction of objective parameters in expert evaluations. Special attention was found in psychometric tools structured as vignettes that allowed the detailed evaluation of legal capacities, present in the legal texts. Psychometric tools in checklist format appeared to be more susceptible to interviewer biases.

Conclusion: The control of inherent biases in forensic psychiatry assessments on criminal matters remains a current challenge, difficult to control in forensic practice. The identification, control and avoidance of them may improve the quality the forensic psychiatric expertise in criminal matters.

1. Introduction

Forensic psychiatry is a specific area of psychiatry's intersection with legal sciences (Rosner & Scott, 2017). The production of forensic medical documents (reports and expert opinions) is one of the principal tasks of forensic psychiatry (Resnick & Soliman, 2012; Rosner & Scott, 2017). However, recent studies have highlighted the poor quality of psychiatric reports submitted to criminal courts and the lack of agreement between examiners when evaluating the same case (Fuger, Acklin, Nguyen, Ignacio, & Gowensmith, 2014; Kacperska, Heitzman, Bak, Leško, & Opio, 2016; Robinson & Acklin, 2010). These findings underscore the vulnerability of the technical characteristics and scientific standards of forensic psychiatric reports, indicating a crisis of credibility in contemporary forensic psychiatry (Guivarch et al., 2017).

This scenario illustrates the complexity of adapting forensic psychiatry assessments to other medical specialties' scientific standards (Meyer et al., 2015; Rosner & Scott, 2017). The development of diagnostic tests with high sensitivity and specificity, such as laboratory and imaging tests, allows diagnostic confirmation in non-psychiatric medical specialties (Meyer et al., 2015; Rosner & Scott, 2017). The lack of these gold standard tests in psychiatry shows the limits of categorical and objective diagnosis in this specialty, in which scientific criteria are still essentially clinical (Fuchs, 2013; Jaspers, 1968; Meyer et al., 2015).

The explanation for the high disagreement between experts on the same case under analysis is complex and involves consideration of various simultaneous factors in forensic psychiatric practice (Buchanan, 2015). The way forensic psychiatric data were obtained by the expert, the report's wording, performance of structured evaluations, use of

^{*} Corresponding author at: Federal University of Rio de Janeiro (UFRJ), Avenida Venceslau Brás 71 (fundos), Botafogo, Rio de Janeiro, RJ 22290-140, Brazil.
E-mail address: lfm1205@gmail.com (L.F. Meyer).

psychometric tools (PT), the expert's experience and professional background, the demand for reports, and countertransference elements in the forensic setting are possible causes of such biases (Buchanan, 2015; El-Shenawy, 2019; Freedman & Woods, 2018; Guivarch et al., 2017; Mossière & Maeder, 2015). However, it is still not known how these variables correlate.

The biases' interference in expert evaluations poses a relevant and current technical and scientific challenge for forensic psychiatry (Freedman & Woods, 2018; Guivarch et al., 2017). The high rate of inter-examiner disagreement challenges the scientific basis of the criteria applied to forensic psychiatric reasoning and conclusions in this setting and tends (mistakenly) to equate such criteria with the expert's personal opinion or argument of authority (Fuger et al., 2014; Kacperska et al., 2016; Robinson & Acklin, 2010). This tends to equate the scientific standard of forensic psychiatry to the reasoning in the underlying legal criteria applied by judges and lawyers, whose *modus operandi* (although factual and probatory) operates by persuasion or established jurisprudence (Freedman & Woods, 2018). The judges' goal of impartiality when drafting legal rulings and sentences has been studied scientifically and remains a persistently elusive ideal, given the observation of biases that reveal elements of influenceability in the criteria adopted by the legal field (Angermeyer & Dietrich, 2006; Angermeyer & Matschinger, 2005; Mossière & Maeder, 2015). However, such biases should not be confused with the inherent biases of forensic psychiatric evaluations per se, which are the current article's focus.

The article's main objective is to verify factors related to bias in forensic psychiatric assessments in criminal matters.

2. Method

This systematic review was carried out according to PRISMA guidelines and registered the protocol Prospective Register of Systematic Reviews (PROSPERO; record no. CRD42020192777). Four electronic databases were searched PubMed, Web of Science, Taylor & Francis, and Scopus databases, over the past 20 years. The following Medical Subject Heading (MeSH) terms were searched with Booleans operators, without any keyword target: (criminal responsibility OR legal responsibility OR neurolaw OR insanity defense) AND (forensic psychiatry OR assessment OR evaluation OR bias OR decision making OR capacity OR psychometric). The search was completed by 31 December 2019. The two authors conducted the search.

2.1. Eligibility criteria

The inclusion criteria adopted were studies that mentioned in the abstract, as the main objective: (1) possible bias factors in criminal forensic assessments conducted by psychiatric and/or psychologist experts, (2) divergent conclusion in forensic assessments based on psychometric tools; (3) factors regarding forensic techniques to explain divergences between experts in the same case analyzed.

The following types of studies were included: editorials, systematic reviews, meta-analyses, original studies, and book reviews. The final sample excluded articles not related to this review's objectives and that addressed aspects pertaining to the jury trial or that exclusively biased the reasoning of judges, jurors, and lawyers. We excluded case reports and duplicate articles and only included publications in English.

2.2. Study selection and data extraction

This review was conducted in two stages. In the first stage, the title of the initial sample containing any target ($n = 7.224$ studies) were screened by the first and second authors, independently. Only studies selected by both authors were included in the next stage. Agreement between reviewers, measured by intraclass correlation coefficient (ICC), was 0.80 (95% confidence interval [95%CI] 0.76–0.85). Divergences were discussed individually by the authors.

In second stage, the abstracts of the selected studies ($n = 325$) were read by the first author, looking for the inclusion criteria. The studies selected were rechecked by the second author and, in case of agreement for eligibility, they were selected for full text analysis ($n = 65$). Then, these studies were fully read by the first author for the inclusion criteria and data extraction. Only the studies approved by both authors composed the final sample ($n = 30$). The first author was responsible for collecting the data of each article, and the second author was responsible for reviewing these data. The flowchart summarizes the article selection resulting in the final sample (Fig. 1).

3. Results

The final sample ($n = 30$) was divided into 3 groups according to each article's content in order to better organize the sample (Table 1). The groups were divided according to the principal themes addressed in the articles: legal factors, psychometric indicators in the expert evaluation, and forensic technique.

3.1. Legal factors

Group 1 included the articles that addressed legal aspects ($n = 13$) (Table 1). The main themes in this group were: difficulty in establishing equivalence between legal and psychiatric terminologies, disagreements between expert reports, possible contributions by neurosciences to expert evaluations, and quality of the reports' wording.

The concept of "capacity" was mentioned as central to forensic psychiatric evaluations (Buchanan, 2015). More than one article cited the lack of consensus in the definition of this concept (Buchanan, 2015; Lacroix, O'Shaughnessy, McNeil, & Binder, 2017; Meynen, 2012). This group also included legislations specifically developed to clarify forensic psychiatry and experts' demands (Lacroix et al., 2017; Meynen, 2012). For example, an article from Canada suggested the incorporation of specific legislation (called "Bill C-14") for criminal responsibility (CR) according to psychiatric evaluations (Table 1) (Lacroix et al., 2017). The application of other ethically-based theoretical constructs was suggested as an alternative for elucidating the legal concept of "capacity", in order to improve the technical assessment of criminal responsibility (Table 1) (Meynen, 2012).

The lack of precise equivalence between legal texts and psychiatric terminology was identified as a possible source of bias in CR evaluations (Joubert & van Staden, 2016; McSherry, 2004; Meynen, 2013; O'Sullivan, 2018). Clinical presentations with psychopathological characteristics not contemplated in legal texts, such as autism, motor automatisms (chorea-like conditions), parasomnias, and transient states of consciousness (sleep-wake cycle), were cited as challenging situations for forensic practice (Joubert & van Staden, 2016; McSherry, 2004; Meynen, 2013). These clinical presentations were described as "borderline", given the limitations in establishing equivalences between the clinical condition (or diagnosis) and legal terminologies (McSherry, 2004; Meynen, 2013).

The forensic psychiatric (or psychological) evaluation complemented with neuroscientific evidence (e.g., neuroimaging tests) was suggested as a possible technical improvement to the evaluation of legal capacity (Penney, 2012; Schleim, 2012). The lack of clear correspondence between the clinical diagnosis (the biological element) and legal terminology, indispensable for verifying legal capacities such as capacity for understanding and capacity for self-determination, could be resolved through neuroscientific evidence (the neuroanatomical element) (Penney, 2012; Schleim, 2012). The identification of neuroanatomical elements and the precise correspondence with legal capacities (when possible) would allow adding objective parameters to forensic psychiatric evaluations in this context (Penney, 2012; Schleim, 2012).

Updating traditional concepts in forensic psychiatry, such as psychopathy, was cited as a factor that could help decrease

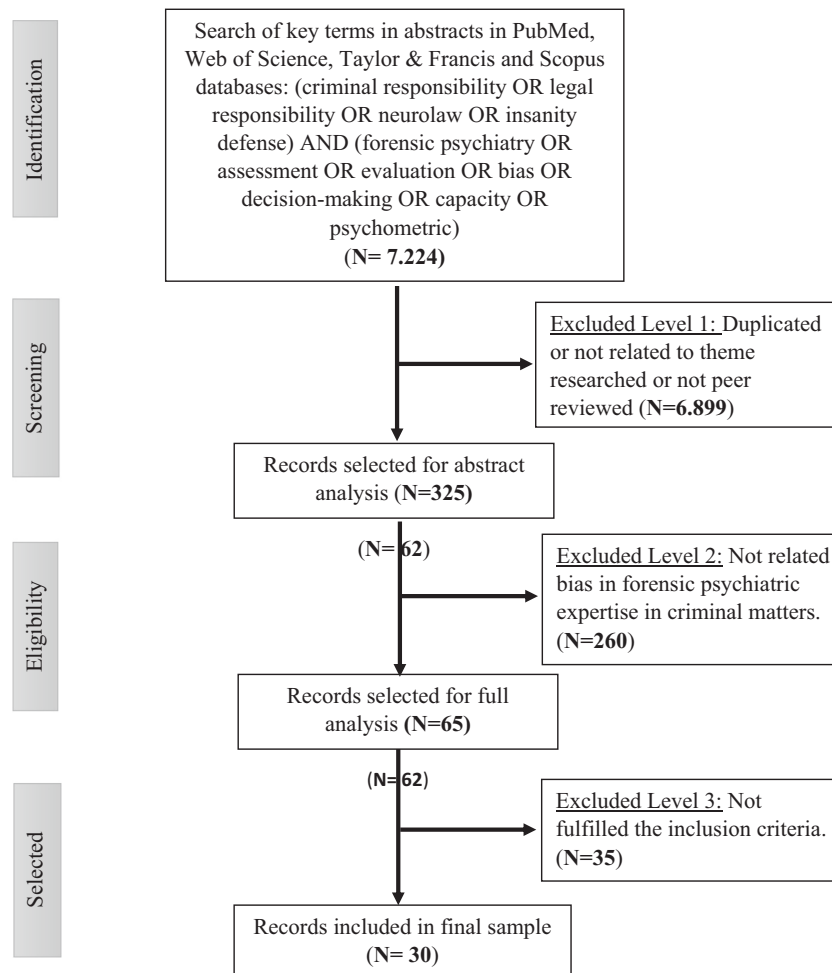


Fig. 1. Study selection PRISMA flow diagram.

countertransference biases and theoretical prejudices in forensic psychiatric evaluations (Felthous, 2010). Contemporary psychiatry acknowledges the presence of psychopathological characteristics in different clinical presentations, but without relevant repercussions for judges and lawyers or in forensic practice (Felthous, 2010). This polymorphism was adopted by the alternative model for diagnosis of personality disorders in the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM 5) (Hopwood, Thomas, Markon, Wright, & Krueger, 2012).

The low technical quality and high disagreement between expert reports submitted to criminal justice on individual cases are limitations to forensic psychiatric practice (Fuger et al., 2014; Kacperska et al., 2016). Reports on CR submitted to the courts in Hawaii (150 expert reports pertaining to 50 cases) produced by three examiners, who were psychologists and psychiatrists (at least one had to be a psychiatrist), from 2006 to 2010, were analyzed with a 43-item quality-coding instrument to verify the documents' quality standards. Each item in this instrument scores "0" when incomplete, "1" if it contains partial information, and "2" if it contains complete information (Fuger et al., 2014). Thirty-three items from the quality-coding instrument were used to compare cases of criminal insanity, in which the total score varied from 0 to 66 (Fuger et al., 2014). Cases of criminal insanity considered 37 items (four additional items referring to danger to others), with the score varying from 0 to 74. The instrument's seven other items were not useful for the comparison of expert reports (Fuger et al., 2014).

In the final sample ($n = 50$) in this same study, consensus among the three examiners was reached in 46% of the cases ($n = 23$) (Fuger et al., 2014). In 52% of the cases ($n = 26$), representing 78 reports, two

examiners agreed on the conclusions (Fuger et al., 2014). Only one case (2% of the sample) showed disagreement between all three examiners (Fuger et al., 2014). Inter-examiner agreement was considered "weak" ($ICC < 0.6$), independently of professional background (psychiatrist or psychologist) (Fuger et al., 2014). Consensus between the examiners and the court occurred in 40.8% of cases ($n = 20$) (Fuger et al., 2014). Lack of standardization of the documents, variation in forensic reasoning, and poor quality of the documents' wording favored high disagreement between the reports submitted to the courts (Fuger et al., 2014).

A Polish study analyzed agreement between expert reports ($n = 381$) pertaining to 117 cases submitted to forensic psychiatric evaluation (Kacperska et al., 2016). Of these, 68 cases were evaluated by more than one professional (forensic psychiatrists or psychologists) (Kacperska et al., 2016). Two evaluations were performed in 41 cases, three evaluations in 17 cases, four evaluations in eight cases, and five evaluations in two cases (Kacperska et al., 2016). A proportional correlation was found between the number of evaluations and the number of divergent expert reports ($r = 0.51$) (Kacperska et al., 2016). In only 36 cases (among the 68 that were evaluated more than once), the expert conclusion was not altered by subsequent evaluations (Kacperska et al., 2016). Meanwhile, among the 32 cases in which there was disagreement, in 22 cases from the total sample (18.8%) the expert conclusion was altered once, in eight cases (6.8%) the expert conclusion was altered twice, and in two cases (1.7%) the expert conclusion was altered three times (Kacperska et al., 2016). Lack of access to the content of the previous reports was identified as the main determinant of this outcome (Kacperska et al., 2016).

The use of hyperbolic clinical wording and lack of standardization of

Table 1

Factors related to bias in forensic psychiatric and psychological expertise in criminal matters: summary objectives, results and conclusions of the final sample of the articles ($n = 30$).

Author	Type of article	Objectives	Results and Conclusion
Group 1: Legal elements (total = 13)			
Buchanan (2015)	Editorial	Discuss the introduction of the concept of "capacity" for evaluations of criminal responsibility, proposed by the United Kingdom Law Commission's Discussion Paper, Criminal Liability: Insanity and Automatism.	Concept of capacity does not clarify the individual's fitness to perform a given act, failing to include fundamental elements for the evaluation of CR: a) rationally evaluate the person's conduct or the act's circumstances; b) understand the act's illegality, and c) control the person's own acts at the moment of the offense.
Felthous, A. (2010)	Review	Compare the evolution of the concept of antisocial personality (psychopathy) in psychiatry and law. Forensic psychiatric experts can help improve the evaluation of criminal responsibility, including psychopathy's personality traits among clinical symptoms verified in other mental disorders.	Concepts pertaining to psychopathy and criminal responsibility are outdated. Updating these concepts, as well as those of clinical presentations of other mental disorders, could improve judges' understanding of the impact of psychopathic personality traits on clinical responsibility.
Fuger et al. (2014)	Original (retrospective)	Compare the quality of forensic reports on criminal responsibility ($n = 150$) submitted to the criminal courts in Hawaii, through a 44-item quality coding instrument.	Low quality of forensic reports was identified, resulting from lack of the professionals' experience in evaluations of criminal responsibility, compared to evaluations of competence to stand trial. Differences in the expert's setting, lack of standardization, and examination performed on different dates are suggested as possible sources of bias.
Joubert and van Staden (2016)	Editorial	Discuss the importance of the forensic psychiatrist's evaluation in cases of automatism representative of pathological behaviors that do not meet the legal definitions' requirements.	The report's conclusion is essential for distinguishing between cases of pathological automatisms and non-pathological impulsive behaviors. The expert must determine whether the automatism meets the legal

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Kacperska et al. (2016)	Original (retrospective)	Verify the quality of psychiatrists' and psychologists' reports on criminal responsibility produced from 2005 to 2010 by a Polish forensic institute.	requirements for a possible case of insanity defense. The study showed high inter-examiner disagreement on the same cases submitted to evaluation of criminal responsibility. The improvement of inter-interviewer reliability is essential for improving the scientific standards applied to evaluations of criminal responsibility and for the creation of standardized forensic reports. These are key factors to decrease court costs and identify individuals standing criminal trial and who need treatment.
Lacroix et al. (2017)	Review	Discuss forensic psychiatric aspects of the Canadian criminal code and potential impact of inclusion in Bill C-14 (The Not Criminally Responsible Reform Act) in psychiatric evaluations of criminal responsibility.	The impact from changes in Bill C-14 is still unknown, such as the inclusion of the term the term "high-risk" to refer to offenders considered more prone for criminal recidivating in the forensic report. The lack of empirical results with the suggested changes and criticisms from society and mental health experts were decisive for its lack of incorporation into the criminal law.
McSherry (2004)	Editorial	Discuss aspects of clinical presentations not addressed by the law, such as dissociation in individuals undergoing heavy stress or trauma that present transient behavior changes or automatisms in individuals standing criminal trial.	Transient psychiatric presentations secondary to stress, dissociative states, or automatisms are complex situations in forensic psychiatry. Their contextualization can facilitate the expert's understanding of the situation when assessing criminal responsibility in each individual case. Assessment of the setting that caused the transient psychiatric presentation in each case can be the most appropriate starting point for evaluating

(continued on next page)

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Meynen (2012)	Editorial	Verify the applicability of ethical/moral parameters (Susan Wolf's account) to concepts pertaining to criminal responsibility and forensic practice.	criminal responsibility in these cases. The approach to ethical and moral parameters can benefit forensic practice, as a possibility for complementing traditional evaluation of legal capacities.
Meynen (2013)	Editorial	Verify how neuroscience research can contribute to improving the standard of forensic psychiatric evaluations of criminal responsibility by: approaching such reports as the evaluation of the influence of mental disorders on the defendant's decision-making process; and verify the scientific progress in the correlation between mental disorders and decision-making process.	The legal texts display little understanding of how a mental disorder can impair an individual's decisions. The adoption of concepts pertaining to mental disorders and patients' decision-making process can help improve the equivalence between forensic psychiatry and legal terminology in psychiatric examinations of criminal responsibility.
O'Sullivan O. (2018)	Editorial	Discuss aspects of autism spectrum disorder and psychiatric evaluations of criminal responsibility. Explore how autism spectrum disorder's clinical characteristics can affect capacity for understanding and capacity for self-determination.	Legal texts lack concepts pertaining to the cognitive limitations of individuals with autism spectrum disorder diagnosis. The forensic psychiatrist's work is indispensable in elucidating these cases in evaluations of criminal responsibility. Updating the legislations or drafting new legal provisions that address aspects pertaining to autism spectrum disorder diagnosis can contribute to the forensic psychiatric evaluation in these cases.
Penney (2012)	Editorial	Verify how the neurosciences can contribute to the evaluation of cases of pathological impulsive behavior in evaluations of criminal responsibility by: empirically demonstrating impairment to capacity for self-determination in individuals with preserved capacity for understanding; and	The neurosciences allow an approach between pathological impulsive behavior and capacity for understanding and capacity for self-determination through: neuroimaging tests, approach to the capacity for discernment between situations

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Schleim (2012)	Editorial	Discuss the influence of new neuroscience in the legal understanding of free will and the neurological basis of individuals' antisocial or criminal behavior in expert evaluations in forensic psychiatry.	demonstrating how legislation can address this phenomenon. legally defined as "right" versus "wrong", and psychometric tools used by experts in the examination setting. The application of new neuroscientific knowledge pertaining to antisocial behavior in forensic psychiatry and the legal understanding of criminal responsibility and violence risk assessment is premature. The incorporation of these factors by forensic psychiatric examinations and legal reasoning requires more research on the topic.
Slovenko (1999)	Editorial	Discuss concepts pertaining to the diagnosis of mental disorders as a prerequisite for criminal responsibility evaluations and establishing equivalence between psychiatric terminology and prevailing legal terminology.	In the scope of forensic psychiatry, the definition of "mental disorder" depends on legal definitions and interpretations of the subject. It is not possible to circumscribe this concept exclusively within clinical elements or guidelines in diagnostic classification manuals such as the DSM.
Group 2: Psychometric instruments (n = 7)			
Advokat et al. (2012)	Original (retrospective)	Compare individuals submitted to psychiatric treatment with or without clinical recovery, according to competence to stand trial (n = 43) through the following psychometric tools: BPRS, GCCT, MMSE, REALM, GAF.	The groups considered fit and unfit according to competence to stand trial, in the initial and final evaluations, presented results with statistical significance in the GCCT. The fit group in competence to stand trial presented significant improvement in psychotic symptoms and were discharged significantly sooner (7.7 ± 8.6 months), compared to the unfit group in competence to stand trial (17.9 ± 8.6 months). The groups did not show significant results

(continued on next page)

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Akinkunmi (2002)	Original	Adapt and validate MacCAT-CA for use in England and Wales, called MacCAT-FP in that jurisdiction, to be applied in individuals with and without mental disorders who were defendants in criminal cases. Clinical variables and IQ were analyzed, respectively, with BPRS and IQ tests.	with the other variables. MacCAT-FP was applied to two groups of prison inmates (with mental disorders and controls) ($n = 45$) and presented satisfactory internal consistency between two interviewers ($k = 0.77$). Participants without diagnosis of mental disorder (prison inmates) scored higher on MacCAT-FP and BPRS, when compared to the group with mental disorders (admitted to hospital units). IQ tests did not differ between the groups. However, MacCAT-FP and BPRS did not show satisfactory indices (ROC curve) for distinguishing between the groups. Psychoticism and withdrawal subscales scored higher in the group considered unfit to stand trial. Depression subscale scored higher in the fit to stand trial group. The MacCAT-FP, once validated for research purposes, may be developed for clinical application. RSCR was applied retrospectively to $n = 1187$ cases. Inter-interviewer reliability was performed in 10 cases with six researchers. RSCR showed distinct scores for individuals unfit to stand trial, partially fit, and fit to stand trial, respectively, of 9.66 (± 5.11), 26.54 (± 5.21), and 40.08 (± 7.90), controlling for other variables (type of offense, diagnosis). The instrument was effective in 88.9%, with Cronbach's alpha of 0.93 ($r = 0.5-0.89$). Inter-rater reliability with kappa = 0.785 ($p < 0.001$).
Cai et al. (2014)	Original	Validate a PT for evaluation for criminal responsibility assessment (RSCR) to the criminal legal context in China.	

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Hilton et al. (2016)	Original (retrospective)	Verify the use of PTs (VRAG, PCL-R, and HCR-20) by experts in the evaluation of the progression of confinement for patients submitted to compulsory psychiatric treatment in hospital units.	Patient files and medical records were reviewed for 63 patients submitted to psychiatric hospital treatment from 2009 to 2012 in Ontario, Canada. VRAG, PCL-R, and HCR-20 were cited, respectively, in 77%, 80%, and 30% of the cases of detention; cases of transference were cited, respectively, in 67%, 66%, and 7%. VRAG, PCL-R, and HCR-20 were applied, respectively, to 12.6%, 21.9%, and 31.3%, in cases of detention; cases of transference were applied, respectively, to 4.2%, 15%, and 15.5%. VRAG was used more in initial evaluations, given the impossibility of verifying the benefit of therapeutic interventions with its items. HCR-20 was applied more to subsequent evaluations due to the possibility of variation in the score on its items. Internal validation ($n = 20$) and external validation ($n = 100$) were performed with AlaSATCom, whose final version consisted of 10 items. The overall reliability of the psychometric tool showed satisfactory indices ($r = 0.949$), as well as inter-examiner agreement ($k = 0.602-1.0$). The psychometric tool showed agreement with the team in charge of the case in 96% of the cases. The official records for 23 cases were reviewed, containing two evaluations with PCL-R. Inter-examiner reliability (prosecution and defense) was lower ($k = 0.39$) than in the research setting and in validation studies of PCL-R.
Hooper et al. (2005)	Original	Develop a psychometric tool called AlaSATCom to support the expert's conclusion on progressive easing of compulsory treatment for patients (considered unfit to stand trial) under long-term compulsory institutionalization.	
Murrie et al. (2008)	Original (retrospective)	Verify inter-examiner reliability (prosecution and defense) in the application of PCL-R to defendants accused of sexual crimes.	

(continued on next page)

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Rogers and Sewell (1999)	Original	Verify the validity of the theoretical construct of R-CRAS and its six domains (simulation, organicity, psychiatric, cognitive, behavioral, and relational) through the combination and reanalysis of data from different research centers with this psychometric tool (obtained in 1984 and 1986).	This demonstrates the examiner's partiality in completing the PCL-R, in contexts with interviewer partiality. 413 cases were reanalyzed through 76 different protocols, performed by two independent researchers. The R-RAS domains were tested with the parameters applied by the jurisdictions covered by the ALI standard. The instrument's internal consistency had an alpha coefficient of 0.84. The mean alpha coefficient was 0.62. No inter-rater reliability between examiners.
Group 3: Inter-examiner disagreement and forensic technique (n = 10)			
Golonka AD. (2016)	Original (retrospective)	Verify the interference of clinical elements related to alcohol consumption (with or without comorbidity with mental disorder) in conclusions of forensic reports on criminal responsibility conducted in Poland.	Analysis of forensic reports by psychiatrists and psychologists. Tendency towards expert conclusion of partial criminal responsibility in cases of alcohol dependence syndrome. Presence of psychotic disorder secondary to alcohol tended towards absence of criminal responsibility. Cases of acute intoxication (not correlated with alcohol dependence) tended towards expert conclusion of criminal responsibility.
Gowensmith et al. (2017)	Original (retrospective)	Verify the degree of agreement between forensic examiners on the conditional release of patients submitted to compulsory hospital treatment, considered unfit to stand trial in the evaluation of criminal responsibility.	175 forensic reports were reviewed (referring to 62 cases) drafted by three examiners. Inter-examiner agreement was 53.2% of the cases. The judges agreed with the examiners in 79.3% of the evaluations. In cases of consensus between the examiners, only 34.5% of the cases were readmitted to hospital; in cases with disagreement, 71.4% were readmitted.
	Meta-analysis		

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
Guarnera and Murrie (2017)		Analyze the results of original studies on inter-examiner reliability in evaluations of competence to stand trial and criminal responsibility, based on published studies.	59 studies were identified, with the final sample consisting of 17 studies: nine assessed competence to stand trial and eight assessed criminal responsibility. Inter-examiner agreement in the final sample varied from 57% to 100% (k = 0.28–1.0). Independent examiners showed k = 0.49 (95% CI: 0.40–0.58) in evaluations of competence to stand trial and k = 0.41 (95% CI: 0.29–0.53) for criminal responsibility. Lack of detailed information on the variables limited the comparison of results.
Höglund et al. (2009)	Original (retrospective)	Compare conclusions by forensic psychiatrists (n = 30), forensic psychologists (n = 30), social workers (n = 45), and nurses (n = 45) from five different health services on: (a) how 12 distinct diagnoses of mental disorders can affect criminal responsibility, (b) evaluation of criminal responsibility in five clinical vignettes, and (c) list additional factors that each health profession considers indispensable for their expert conclusions.	The diagnosis of mental disorder alone was a factor for the expert conclusion of criminal responsibility. There was low agreement between professionals on the method adopted for evaluation. Schizophrenia was the diagnosis most closely associated with decreased criminal responsibility, followed by diagnoses of dementia and mental retardation. Psychoactive substance disorder negatively influenced the decrease in criminal responsibility. Borderline personality, antisocial, and narcissistic traits also negatively influenced the decrease in criminal responsibility. Socioeconomic variables contributed to the expert conclusion.
Kalis and Meynen (2014)	Editorial	Discuss the relevance of theoretical decision-making models and the repercussions of their application to forensic	The theoretical model from clinical psychiatry based on decision-making (option generation

(continued on next page)

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
		evaluations of criminal responsibility.	→ option selection → action initiation) could contribute to the forensic psychiatrist's arguments, with demonstration of forensic psychiatric reasoning in evaluation of criminal responsibility.
Kois et al. (2017)	Original	Compare the experts' conclusions (competent-sane, competent-insane, incompetent-sane, and incompetent-insane) in combined and uncombined evaluations of competence to stand trial and criminal responsibility, in defendants referred for criminal psychiatric examination.	Reports were reviewed that had been submitted to the state courts of Virginia and New York from 1990 to 2005 ($n = 2751$). Defendants with diagnosis of organic mental disorder or developmental disorder were more likely to be considered competent-sane. Those with psychotic disorders were 7 times more likely to be considered competent-insane, while those with organic mental and developmental disorders were 3 times more likely. Psychoactive substance use had little influence on the conclusion between sane versus insane.
Large et al. (2009)	Original	Compare agreement among multiple examiners on the same case, working with the same party in the case or opposing parties, in forensic psychiatric evaluations of competence to stand trial and criminal responsibility conducted in Australia.	Official records analyzed from 110 cases from 2005 to 2010 with two or more forensic reports. 270 expert reports were collected (226 by 30 psychiatrists and 44 by 15 psychologists), of which 122 were produced by the prosecution and 148 by the defense. Types of offenses were: 30 charges of murder or attempted murder (referred to as homicide offenses), 35 charges of serious injury or assault, 14 charges of sexual assault, 12 serious property offenses, 10 drug cases, and the remaining eight cases included fraud, kidnapping, arson, and firearms offenses. Agreement levels between

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
			experts for the same party to the case for evaluations of competence to stand trial and criminal responsibility were, respectively, $k = 0.471$ (0.142–0.690) and $k = 0.644$ (0.315–0.973). Between experts for opposing parties in the cases, the agreement values for competence to stand trial and criminal responsibility were, respectively, $k = 0.293$ (0.134–0.451) and $k = 0.508$ (0.295–0.720). The main criteria applied to evaluations on release of patients submitted to compulsory psychiatric treatment are: (a) diagnosis of mental disorder and (b) clinical evidence of danger to others. Evaluations prior to the 1990s did not present documentation or objective parameters to back the expert's opinion. Evaluation of the risk of repeat offense and the use of psychometric tools (diagnostic and violence risk assessment) in expert conclusions were incorporated since the 1990s and 2000s. Scientific developments in the area led to improvement in compulsory psychiatric treatment, the expert's concern with future risk of violence, and greater standardization of the forensic psychiatric elements considered in the report. The main factors considered in treatment progression, independently of the decade, were treatment adherence and response and presence of substance disorder. The factors considered by experts in the
McDermott BE. et al. (2008)	Original	Verify the criteria applied by psychiatrists to justify the release of patients submitted to compulsory psychiatric institutionalization, especially if the criteria changed according to the period (decade) in which the evaluation was performed.	
Niveau and Sozonets (2001)	Original (retrospective)	Verify factors considered by experts in evaluations of	

(continued on next page)

Table 1 (continued)

Author	Type of article	Objectives	Results and Conclusion
		criminal responsibility and whether they changed depending on the period analyzed (1973–1974 and 1997–1998).	conclusions differed. There was no significant difference in factors in cases considered unfit to stand trial. There was a statistically significant difference in the psychiatric diagnosis performed, according to the period analyzed.
Stredny et al. (2012)	Original	Compare agreement between psychiatrists and psychologists on the same team concerning the release of patients from hospital custody.	Psychiatrists and psychologists agreed on 78% of cases, considered moderate agreement ($k = 0.47$). Psychiatrists and psychologists recommended hospitalization in 70.6% and 71.3% of cases, respectively. There was slight disagreement on the items considered by each professional group. For psychiatrists, the most determinant factors for hospitalization were: history of suicide attempt, lack of social support, and unemployment. For psychologists, the most determinant factors for hospitalization were: history of suicide attempt, use of weapons, and employment problems.

AlaSATCom = Alabama Structured Assessment of Treatment Completion; BPRS = Brief Psychiatric Rating Scale; DSM = Diagnostic and Statistical Manual of Mental Disorders; GAF = Global Assessment of Functioning; GCCT = Georgia Court Competency Test; HCR-20 = Historical, Clinical and Risk Management 20; IQ = intelligence quotient; MacCAT-CA = MacArthur Competence Assessment Tool-Criminal Adjudication; MacCAT-FP = MacArthur Competence Assessment Tool-Fitness to Plead; MMSE = Mini-Mental State Examination; PCL-R = Psychopathic Checklist Revised; REALM = Rapid Estimate of Adult Literacy in Medicine; R-CRAS = Rogers Criminal Responsibility Scale; RSCR = Rating Scale of Criminal Responsibility for mentally disordered offenders.

the expert reports and of equivalence between clinical and legal terminology were identified as factors for inter-examiner disagreement on the same cases, as well as between the court ruling and the forensic psychiatric conclusion (Slovenko, 1999). The use of non-scientific expressions or those denoting clinical severity (e.g., “mental defect” versus “mental disorder”) tended to favor court rulings of compulsory treatment and disagreement between examiners (Slovenko, 1999).

3.2. Psychometric tools in forensic evaluation

This group included studies that used PTs in forensic psychiatric evaluations (Table 1). The PTs in this group can be subdivided into three subgroups: PTs specifically developed for forensic psychiatry (called

“forensic assessment techniques”), PTs based on clinical psychiatric elements with high relevance for forensic psychiatry (“forensically relevant assessment techniques”), and PTs largely applied in clinical psychiatric assessments useful in forensic psychiatry (“clinical assessment techniques”) (Advokat, Guidry, Burnett, Manguno-Mire, & Thompson, 2012; Cai et al., 2014; Hilton, Simpson, & Ham, 2016; Murrie, Boccaccini, Johnson, & Janke, 2008; Rogers, Seman, & Clark, 1986; Rogers & Sewell, 1999; Rosner & Scott, 2017).

In our review, the subgroup “forensic assessment techniques” included PTs applied to the assessment of psychopathological legal capacities (Rosner & Scott, 2017). PTs to evaluate competence to stand trial (CST) and CR were included in this subgroup (Akinkunmi, 2002; Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999). CST assesses whether the defendant displays discernment concerning his or her legal status and the trial’s consequences at that moment (cross-sectional evaluation) (Akinkunmi, 2002). When clinical elements for discernment are impaired, the defendant is submitted to compulsory psychiatric treatment until obtaining satisfactory clinical improvement in order to be prosecuted (Advokat et al., 2012; Akinkunmi, 2002). Criminal responsibility assesses whether the individual, at the time the offense was committed, was able to comprehend the act’s illegality and to practice self-determination based on this comprehension (retrospective evaluation) (Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999).

CST assessments are typically employed in Anglo-Saxon jurisdictions (Akinkunmi, 2002). An example of a PT developed for this purpose and widely applied in the United States is the MacArthur Competence Assessment Tool-Criminal Adjudication, which is structured in clinical vignettes with questions and answers focused on legal competencies (Akinkunmi, 2002). This PT was also validated in the United Kingdom, where it was called MacArthur Fitness to Plead (Akinkunmi, 2002). This cross-cultural validation did not find statistical significance in the comparison of the instrument’s score with the expert conclusion (Akinkunmi, 2002). However, the study highlighted the inherent difficulty in psychometric evaluation of psychopathological constructs pertaining to legal competencies without considering clinical elements (e.g., presence of delusions and hallucinations) or criminal dynamics (use of weapons in the crime, type of crime), as with PTs structured in checklists. An example of this kind of PT in our sample is the Georgia Court Competence Test (GCCT), which is structured as a 21-item checklist divided into sections (Advokat et al., 2012). The sections evaluate the defendant’s capacity to visually perceive the representation of a hearing room, receive the defense attorney’s assistance, and answer questions aimed at verifying potential malingering or conscious simulation (Table 1) (Advokat et al., 2012).

Our review included two studies with PTs developed for the evaluation of CR: the Rogers Criminal Responsibility Assessment Scales (R-CRAS) and the rating scale of criminal responsibility for mentally disordered offenders (Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999). Both are structured on checklists completed by the examiner with clinical impressions of the forensic examination and prior data, such as official records of the crime or medical documents. The Rogers Criminal Responsibility Assessment Scales feature a 25-item inventory (clinical data related to the examiner’s opinion), subdivided into six cardinal domains: patient reliability, organicity, psychopathy (at the time of the crime), cognitive control, behavioral control, and causal nexus (Rogers et al., 1986; Rogers & Sewell, 1999). The item scores vary from 0 to 5 or 6. A score of 0 is assigned when there is no available information to complete the item. A score of 1 means absence of the characteristic. Scores of 2 or more indicate presence of the characteristic with increasing and directly proportional severity (Rogers et al., 1986; Rogers & Sewell, 1999).

The validation studies on the Rogers Criminal Responsibility Assessment Scale found that the six cardinal domains of the instrument share some clinical items (Rogers & Sewell, 1999). The main clinical items with statistical significance shared by the majority of the domains was: observable bizarre behavior, awareness of criminality, level of

verbal coherence, examiner's assessment of patient's self-control, and relationship of loss of control to psychosis. These results point to clinical similarities between the domains and limitations in their individualization (Rogers & Sewell, 1999). The comparison of the R-CRAS result to the expert's opinion showed statistical significance ($p < 0.01$) (Table 1).

The rating scale of criminal responsibility for mentally disordered offenders includes various non-psychopathological items (Cai et al., 2014). This PT consists of 18 items: criminal motivation; aura before the crime; inducement to crime; time, space, object, and tool selected for the crime; emotion during the crime; shirking after the crime; concealing the truth during the interview; malingering; understanding the nature of the crime; understanding the consequence of the crime; impairment of life skills; impairment to learning at work; impaired insight; impaired judgment of reality; and impaired self-control (Cai et al., 2014). Each item's score varies from 0 to 4, according to absence (score 0) or presence of the characteristic (score of 4 for the highest degree) (Cai et al., 2014). The rating scale of criminal responsibility for mentally disordered offenders was validated retrospectively through expert reports and showed statistically significant results ($p < 0.01$) (Table 1).

In our review, the subgroup of forensically relevant assessment techniques includes psychometric instruments to evaluate the risk of violence. These tools assess the probabilistic risk of violent behavior in an individual (with or without mental disorder) who has previously committed an offense (prospective evaluation) (Hilton et al., 2016; Murrie et al., 2008). Studies to assess the risk of violence have applied different PTs, such as: Historical Clinical Risk Management-20 (HCR-20), Violence Risk Appraisal Guide (VRAG), Psychopathic Checklist Revised (PCL-R), and Alabama Structured Assessment of Treatment Completion for Insanity Acquittes (AlaSATcom) (Table 1) (Hilton et al., 2016; Hooper, McLearn, & Barnett, 2005; Murrie et al., 2008). In common, all these PTs assess the risk of violent behavior in individuals with diagnosis of mental disorder or psychopathic personality traits.

Specifically, the PCL-R was mentioned as a source of disagreement among examiners in the same case (Murrie et al., 2008). The comparison of the PCL-R scores by two independent examiners representing opposing parties in a court case (prosecution versus defense) revealed high disagreement (Murrie et al., 2008). Most of the results showed large discrepancies, with disagreement greater than 6.0 points in more than 60% of the cases (total 23). There was agreement between the examiners' total scores in two cases (Murrie et al., 2008). The PCL-R scores in studies on inter-examiner disagreement in violence risk assessment differed from those in a validation study on this PT (Table 1).

Finally, the group of clinical assessment techniques featured PTs widely used in clinical practice, such as the Brief Psychiatric Rating Scale (BPRS), Mini-Mental State Examination (MMSE), and Global Assessment of Functioning (GAF) (Advokat et al., 2012). The PTs in the subgroups "forensically relevant assessment techniques", and "clinical assessment techniques" were mainly used to support clinical decisions to discharge patients from psychiatric hospitals or for verification of defendants' competence to stand trial. The use of these PTs in the expert reports varied from simply citing them to performing a complete application of the PT with a total score (Table 1).

3.3. Forensic technique

This group included the articles that addressed topics on forensic technique, such as: disagreement between examiners on the same case, factors capable of introducing biases in forensic techniques, and the adaptation of theoretical models from clinical psychiatry to forensic psychiatry. Violence risk assessment was the most common type of study in this group (Table 1).

Agreement between three examiners on discharging patients from compulsory psychiatric hospitalization showed varied results, generally with insufficient levels of agreement (Gowensmith, Murrie, Boccaccini, & McNichols, 2017). In one study, the three examiners were unanimous on 53.2% of the cases ($r = 0.35$) (Gowensmith et al., 2017). In a 3-year

period, patients that underwent violence risk assessment showed lower rates of hospital readmission (29.6%) when the examiners were unanimous, compared to hospital readmission when there was disagreement (70.4%) (Table 1) (Gowensmith et al., 2017).

The opinions of psychiatrists and psychologists in releasing mentally ill offenders from temporary custodial treatment was compared in another study (Stredny, Parker, & Dibble, 2012). Psychiatrists and psychologists frequently recommend hospitalization of patients in the following situations: history of suicide attempts or self-inflicted injury; family or other psychosocial issues; lack of structured activities in the community; and individuals admitted to hospital directly from prison (Stredny et al., 2012). Psychiatrists were more likely to recommend hospitalization of patients with substance use disorders. Meanwhile, psychologists were more likely to recommend hospitalization of patients that used weapons when committing the offense (Stredny et al., 2012).

In our sample, the reliability of PTs developed for violence risk assessment (like PCL-R) had lower rates than in their validation studies (Table 1) (Murrie et al., 2008). Disparate scores on the application of PCL-R by different examiners in the same case indicate the effect of biases on the scores with these PTs, for example, in relation to the type of crime committed (Murrie et al., 2008).

The decade in which the forensic evaluation was performed determined the variation in the criteria applied to violence risk assessment for patients under compulsory hospitalization (McDermott et al., 2008). The most prevalent factors in all the decades studied were: adherence and response to the proposed treatment, history of psychoactive substance use, and risk of violence (McDermott et al., 2008). These factors were also found in a meta-analysis on the topic, together with: type of crime committed by the defendant and consent to the forensic examination, presence of psychiatric diagnosis, the expert's experience and professional background (psychiatrist or psychologist), availability of alternative sources of information, prevailing jurisdiction, methodology, and conditions of the evaluation (use of PTs, simultaneous or sequential examiners) (Guarnera & Murrie, 2017). The biases pertaining to the examiner's professional background (field attendants, nurses, psychiatrists, and psychologists), tested with clinical vignettes in the context of CR, showed significant differences in the methodology adopted by each professional category and in the evaluation of the clinical elements in the capacity for self-determination (Höglund, Levander, Anckarsäter, & Radovic, 2009).

The type of forensic psychiatric or psychological evaluation performed (CST, CR, and violence risk assessment) was identified as an additional factor for inter-examiner disagreement (Guarnera & Murrie, 2017; Höglund et al., 2009; Kois, Wellbeloved-Stone, Chauhan, & Warren, 2017; Large, Nielssen, & Elliott, 2009). In jurisdictions where CST and CR coexist, the simultaneous performance of both exams resulted in different conclusions from those when the two tests were performed at different moments (Höglund et al., 2009). Inherent elements in each type of evaluation (CST and CR) can affect each examination's methodology and objectives, generating biases for both conclusions (Höglund et al., 2009). In CST examinations, experts acting for different sides (defense versus prosecution) showed low agreement ($r = 0.293$) when compared to experts working for the same party ($r = 0.41$) (Kois et al., 2017). Moderate levels of agreement ($r = 0.51$) were seen in CR examinations and in those with diagnosis of schizophrenia in the examinee (Kois et al., 2017). There was higher agreement between examiners in homicide cases (CST) and with the presence of diagnosis of schizophrenia (CR) (Kois et al., 2017). A meta-analysis compared the agreement between the conclusions of expert reports on CST and CR on the same cases and obtained similar results. The study showed important heterogeneity in the results and weak levels of inter-examiner agreement on the same forensic case (Table 1) (Large et al., 2009).

The mental disorder's diagnostic group was identified as a source of bias in the forensic expert's conclusion (Golanka, 2016; Höglund et al., 2009). Diagnoses of organic and psychotic mental disorders were more frequent in individuals considered incompetent in CST evaluations

(Höglund et al., 2009). Defendants who had committed their offense under the influence of psychoactive substances were more likely to be considered competent in CST evaluations and to be convicted (Höglund et al., 2009). The presence of a diagnosis of alcohol use disorder or alcohol intoxication at the time of the offense contributed to the defendant's being considered fit to stand to trial according to expert assessments performed by psychiatrists and psychologists in Poland (Golonka, 2016). However, these variables in the presence of serious psychiatric comorbidity (psychotic disorder and mental retardation) did not interfere in the expert conclusion concerning CST (Golonka, 2016).

CR reports produced in Switzerland in the 1970s and 1990s were analyzed to verify determinant factors for experts' conclusions (Niveau & Sozonets, 2001). Conclusions that reported a diagnosis of substance use disorder in forensic psychiatric assessment were more prevalent than those reporting diagnoses of other mental disorders in both decades (Niveau & Sozonets, 2001). The increase in conclusions using clinical and forensic psychometric tools was greater in the 1990s, alongside an increase in the incidence of diagnoses of psychotic and mood disorders (depressive episode and recurrent depressive disorder) in expert reports (Niveau & Sozonets, 2001). In the 1970s, diagnoses of personality disorders were more prevalent. Socioeconomic differences limited the scope of comparison of these results (Niveau & Sozonets, 2001). Offenders with substance use disorders were more likely to be considered competent to stand trial, when compared to those with depressive and psychotic disorders (Niveau & Sozonets, 2001).

The adoption of theoretical models applied to clinical psychiatry (such as decision-making) has been suggested as an alternative for forensic psychiatry in order to facilitate equivalence between legal and clinical terminologies (Guarnera & Murrie, 2017; Kalis & Meynen, 2014). The adoption of clinical models has been suggested as a possible technical improvement in forensic psychiatric practice (Guarnera & Murrie, 2017; Kalis & Meynen, 2014).

4. Discussion

Our results show that disagreement is common between experts in the evaluation of the same criminal case and possibly involves multiple causal factors. Inter-examiner disagreement in criminal evaluations is actually the rule, which justifies a critical analysis of scientific and technical criteria adopted by forensic psychiatry. Grouping the biases according to their characteristics is useful for verifying the kind type of interference they generate and the ways to avoid them (Table 1).

Legal elements were chosen as the starting point to answer the questions that were raised, given the ultimate purpose of forensic psychiatry: to assist judges, jurors, and lawyers on psychiatric matters (Rosner & Scott, 2017). More than one article in our sample discussed the development of specific legal texts to address the needs of forensic psychiatry as a way of avoiding biases and elucidating the case (Buchanan, 2015; Felthous, 2010; Joubert & van Staden, 2016; Lacroix, O'Shaughnessy, McNeil, & Binder, 2017; McSherry, 2004; Meynen, 2012). Improvement of the legal definition of "competence" applied to forensic psychiatry, updating the legal terminologies with greater interface with this matter, and discussion between judges, lawyers, and forensic psychiatrists are possibilities for decreasing the interference of legal biases in forensic psychiatric reports on criminal cases (Buchanan, 2015; Felthous, 2010; Joubert & van Staden, 2016; Lacroix, O'Shaughnessy, McNeil, & Binder, 2017; McSherry, 2004; Meynen, 2012). However, such modifications should primarily serve the legal sciences, given the existence of other factors (not pertaining to forensic psychiatry) involved in legal reasoning, and that should be independent from other areas of knowledge (Rosner & Scott, 2017).

Our review highlighted the existence of clinical presentations that are pertinent to forensic psychiatry and are not addressed by legal terminology (Joubert & van Staden, 2016; McSherry, 2004; O'Sullivan, 2018; Penney, 2012; Schleim, 2012). Certain medical conditions, due to their syndromic or etiopathogenic characteristics, are not covered in

legal texts and may generate confusion in the expert report (Joubert & van Staden, 2016; McSherry, 2004; O'Sullivan, 2018; Penney, 2012; Schleim, 2012). The law requires the presence of a positive psychiatric diagnosis, the biological element of the biopsychological criterion, as the prerequisite for considering the defendant not guilty by reason of insanity (Rosner & Scott, 2017). However, certain clinical conditions lack an identifiable neurobiological component and/or do not fit the diagnosis of a psychiatric illness (Joubert & van Staden, 2016; O'Sullivan, 2018; Penney, 2012; Rosner & Scott, 2017; Schleim, 2012).

The psychopathologic specificities of automatisms (verbal tics, Tourette syndrome, and ballistic or chorea-like movements), parasomnias (somnambulism and sleep-wake cycle), and autism spectrum disorders are examples of clinical presentations not addressed by most criminal legislations (McSherry, 2004). In autism spectrum disorders, the impairments to interpersonal and environmental interaction are symptoms potentially capable of jeopardizing the performance of legal capacities, especially in more complex daily situations (Joubert & van Staden, 2016). States of somnambulism or involuntary movements during sleep do not represent mental disorders and may not meet legal requirements (biological element) for being considered mentally ill (McSherry, 2004; Rosner & Scott, 2017). The difficulty in the classification and identification of these clinical conditions, determined by specific neurobiological and psychogenic elements, illustrates the complexity of expert evaluations in these cases.

We contend that complementing the prevailing legal texts with clinical elements not contemplated by the law serves forensic psychiatry's principal purpose. Specific legal texts prioritizing forensic psychiatry elements at the expense of legal aspects may raise difficulties for judges when drafting sentences and creating jurisprudence. The prioritization of forensic psychiatric concepts to the detriment of those originating in the legal sciences may compromise the development of legal reasoning, which has its own methodology. In addition, the constant updating of legal texts may be counterproductive to the *modus operandi* of legal sciences, which require stability in the development of jurisprudence and rulings, essential for legal case flow.

The production of legal reports is a key point in the interface between forensic psychiatry and the legal sciences, given the reports' essential role in legal proceedings (evidence) in an area of knowledge outside the judge's expertise. The low quality of these legal documents and high disagreement between documents on the same case illustrate the difficulty of forensic psychiatry in defining its standard scientific criteria, and thus of establishing itself as a science (Fuger et al., 2014; Kacperska et al., 2016; Robinson & Acklin, 2010).

Our results point to the lack of standardization, and of a common theoretical direction for forensic psychiatry in the development of precise expert reasoning, and the elaboration of legal reports (Fuger et al., 2014; Kacperska et al., 2016; Slovenko, 1999). Expert's countertransference issues and difficulty in accessing clinical and official records are additional elements that contributed to disagreements between expert reports on the same case. Another factor was the use of colloquial expressions or unnecessarily overstating mental disorders' severity (Slovenko, 1999).

The progressive distancing between the concepts of psychopathy applied to criminal justice and in contemporary psychiatry appeared in our results as a potential source of examiner bias on the same case submitted to forensic examination (Felthous, 2010; Murrie, Boccaccini, Johnson, & Janke, 2008). Recent advances in the diagnosis of personality disorders (alternative model of DSM 5), including antisocial personality disorder, represent important scientific strides that have still not been incorporated into forensic psychiatry and legal sciences (Hopwood et al., 2012).

We found a similar situation with PCL-R in forensic practice. One of the main contributions by this PT has been to provide an accurate definition of psychopathy, currently applied in forensic psychiatry, and to its psychometric evaluation (PCL-R) (Felthous, 2010; Murrie, Boccaccini, Johnson, & Janke, 2008). However, our results point to the

limitation of PCL-R in forensic practice, given the interference from the examiner's countertransference biases related to the type of offense committed or to the type of psychiatric diagnosis (Murrie et al., 2008). Lack of training, limited practical experience with this PT, and sampling bias in other studies may explain the difference between the results with this PT in the validation study and in studies of forensic expert settings.

The development of specific PTs (forensic assessment techniques) for verifying LCs was a significant achievement in forensic psychiatric evaluations of CST and CR (Advokat et al., 2012; Akinkunmi, 2002; Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999; Rosner & Scott, 2017). The adoption of a psychometric indicator tends to decrease interferences from biases and to standardize the expert conclusion, attenuating the subjectivity of the evaluation. The principal difference between the PTs developed for the evaluation of LCs in our results involves their structuring, although both are norm-based tests (Advokat et al., 2012; Akinkunmi, 2002; Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999; Rosner & Scott, 2017). R-CRAS and RSCR are structured on psychometric inventories (checklists), while MacArthur Criminal Adjudication and MacArthur Fitness to Plead are structured as clinical vignettes. These PTs appear to be methodologically more adequate for the evaluation of LCs than violence risk assessment tools, since they exclusively consider the individual's psychopathological constructs (Akinkunmi, 2002). In our opinion, LCs are analyzed mainly through the evaluation of specific psychopathological constructs, while other variables (clinical characteristics and criminal dynamics) are complementary to the forensic psychiatric reasoning. The prioritization of the latter (rather than the former) may potentially bias the forensic reasoning through idiosyncratic elements of the examiner, clinical prejudices, or biases related to the specific offense.

The absence of a theoretical model providing a scrutinized basis for the psychopathological constructs of legal capacities assessed by PTs such as R-CRAS and RSCR may be a methodological limitation. The inclusion of varied items, although related to clinical factors and the criminal dynamics and pertinent to the forensic examination, limit the examiner in scrutinizing the necessary psychopathological constructs for examination of CST and CR (determined by law) and can generate countertransference bias (Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999). PTs structured in checklist format can be methodologically more adequate for screening severe cases that require a more focused expert evaluation, besides standardizing relevant elements of forensic psychiatry pertaining to the examination. MacCAT-CA and MacCAT-FP are PTs in vignette format whose methodology focuses on the individual's psychopathological performance related to LCs (Hoge et al., 1997; Otto et al., 1998).

Another limitation to PTs structured as checklists is the possibility of direct interference from interviewer biases in their completion. In these cases, the score is more dependent on the interviewer's evaluation than on the examinee's psychopathological performance (Cai et al., 2014; Rogers et al., 1986; Rogers & Sewell, 1999). Data on disagreement in the application of PCL-R support this statement (Murrie et al., 2008). Additionally, the possibility of scoring a psychometric tool retrospectively, without a direct examinee's evaluation, can be another factor of potential bias (as RSCR validation study). Meanwhile, PTs like the GCCT can display limited applicability to the extent that they were designed to meet the needs of specific legal texts in specific jurisdictions (Advokat et al., 2012). However, similar jurisdictions in this matter can also benefit from these psychometric tools.

In our opinion, ideally, specific PTs for the evaluation of legal capacities (forensic assessment techniques) should focus on the single assessment of pertinent psychopathological constructs in order to preserve the nature of the object under investigation. The inclusion of variables extraneous to psychopathology, such as clinical elements or criminal dynamics, can bias the forensic evaluation. Importantly, the legal capacities appearing in legal texts correspond to psychopathological constructs, which are conceptualized by classical psychopathology (phenomenological method). That is, they are of a conceptual and

immaterial (phenomenological) nature, and it is essential to distinguish them from characteristically factual or objective variables (clinical and criminal dynamics). These should complement the examination of LCs, allowing an exact, impartial, and direct equivalence with the legal texts.

The most frequent use of PTs in our sample was in reports on violence risk assessment, such as with PCL-R, HCR-20, VRAG, and AlaSATcom. These were used mainly to back the decision by experts to release patients from compulsory inpatient psychiatric treatment. The PTs developed for clinical psychiatry (BPRS, MEEM, and GAF) appeared in the final sample, but only in one article (Advokat et al., 2012). The use of these PTs (clinical assessment techniques) can add objective parameters to the expert evaluation, but given their clinical purpose, they may include elements that are not pertinent to forensic psychiatry and thus introduce unintended biases.

A similar reasoning can be applied to the incorporation of theoretical models from clinical psychiatry for forensic purposes (Kalis & Meynen, 2014). For example, decision-making models are useful in evaluating impulsive behaviors, but should not be confused with the inherent concepts of legal capacity (such as the capacity for self-determination), which have their own foundations in forensic psychopathology (Kalis & Meynen, 2014). Rather, the incorporation of theoretical models developed specifically for forensic psychiatry should prevail in order to avoid clinical biases in the expert evaluations.

The presence of preconceived theoretical elements involving experts' countertransference was frequent in the articles comprising group 3 in our sample (Golonka, 2016; Guarnera & Murrie, 2017; Höglund et al., 2009; Kois, Wellbeloved-Stone, Chauhan, & Warren, 2017; Large, Nielssen, & Elliott, 2009; McDermott, Gerbasi, Quanbeck, & Scott, 2005; Niveau & Sozonets, 2001). The diagnosis of substance use disorders and states of acute intoxication (regardless of the presence of psychiatric diagnosis) while the offense was being committed were clinical conditions that favored the conclusion of criminal responsibility (Niveau & Sozonets, 2001). Meanwhile, individuals that were assessed as having decreased criminal responsibility were more common in cases of diagnoses that are classically considered severe (schizophrenia and depression). Substance use disorders also favored the conclusion of CST in CR evaluations reported in our sample (Golonka, 2016; Niveau & Sozonets, 2001). This variable may indicate that contexts related to psychoactive substances were present in most of the crimes studied (Golonka, 2016; Niveau & Sozonets, 2001). Our results suggest that substance use disorders may be a potential source of bias in the expert's conclusion, because they appear frequently in forensic settings and can negatively impact the examiner's countertransference against the examinee, especially in the presence of a comorbid mental disorder without clear-cut symptoms.

Factors related to the historical moment of the expert evaluation and new scientific knowledge in psychiatry (such as the development of PTs and diagnostic classification manuals) modified and improved forensic psychiatric technique (Golonka, 2016). Modifications of the diagnostic models for personality disorders (alternative model of DSM 5) are recent and will require further studies to verify their effect on forensic psychiatry (Hopwood et al., 2012).

Studies with high agreement between examiners were infrequent and displayed limitations in their methods and practical execution (i.e., simultaneous interview) (Murrie et al., 2008). Those with medium to low agreement had independent interviewers, used multiple and non-standardized data sources, and took place in diverse evaluation settings. The high number of expert evaluations of insanity defense in the same case contributed to the increase in disagreement between examiners (Kacperska et al., 2016). The presence of experts with different professional backgrounds contributed to the disagreement between examiners on the same case (Fuger et al., 2014; Murrie et al., 2008). Investigations that performed independent sequential evaluations with more than one expert showed higher rates of disagreement, when compared to those performing simultaneous evaluations (Fuger et al., 2014; Kacperska et al., 2016; Murrie et al., 2008). The multiplicity of

factors related to the investigation's setting limited the comparison of results in our sample. Importantly, reports performed by more than one examiner with the same professional background may represent a relevant criterion for the scientific standards of psychiatric examinations in criminal cases, since both can (simultaneously and mutually) help decrease biases and elements of countertransference.

The final sample in our review only included articles pertaining to criminal offenses. The results cannot be generalized to other areas of forensic psychiatry, such as civil and labor law, given the difference between the factors addressed in these settings and in criminal law. The different keywords used in studies on this topic may have limited the scope of our results, although we searched a total of 11 keywords (method item). The exclusion of articles in languages other than English may also have limited the final sample.

5. Conclusion

The variety of sources of biases in forensic psychiatric assessments in criminal matters reveals the theme's complexity and the difficulty in decreasing their interference. Our results provide guidance for this undertaking. The key points are detailed knowledge of inherent psychopathological concepts for legal capacities, standardization of the examination, the use of psychometric indicators developed specifically for forensic psychiatry (particularly those that exclusively assess legal capacities), and concise drafting of the expert report. Other key points are preferential use of well-established scientific terms, avoidance of jargon and buzzwords in the expert report, and simultaneous evaluations by professionals with the same background and with experience in the area. All these factors represent potential bias in forensic psychiatric assessments. The identification, control and avoidance of them may improve the quality of forensic reports submitted to the Court, making it possible to assist judges in producing adequate rulings and sentences.

Authors declarations of interest

None.

References

- Advokat, C. D., Guidry, D., Burnett, D. M. R., Manguno-Mire, G., & Thompson, J. W. (2012). Competency restoration treatment: Differences between defendants declared competent or incompetent to stand trial. *The Journal of the American Academy of Psychiatry and the Law*, 40(1), 89–97. <https://doi.org/10.29158/JAAPL.003819-19>.
- Akinkunmi, A. A. (2002). The MacArthur competence assessment tool-fitness to plead: A preliminary evaluation of a research instrument for assessing fitness to plead in England and Wales. *The Journal of the American Academy of Psychiatry and the Law*, 30(4), 476–482. PMID: 12539898. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12539898>.
- Angermeyer, M. C., & Dietrich, S. (2006). Public beliefs about and attitudes towards people with mental illness: a review of population studies. *Acta Psychiatrica Scandinavica*, 113, 163–179. <https://doi.org/10.1111/j.1600-0447.2005.00699.x>.
- Angermeyer, M. C., & Matschinger, H. (2005). The stigma of mental illness in Germany: a trend analysis. *International Journal of Social Psychiatry*, 51(3), 276–284. <https://doi.org/10.1177/0020764005057390>.
- Buchanan, A. (2015). Who needs capacity? *International Journal of Law and Psychiatry*, 40, 1–5. <https://doi.org/10.1016/j.ijlp.2015.04.001>.
- Cai, W., Zhang, Q., Huang, F., Guan, W., Tang, T., & Liu, C. (2014). The reliability and validity of the rating scale of criminal responsibility for mentally disordered offenders. *Forensic Science International*, 236, 146–150. <https://doi.org/10.1016/j.foresci.2013.12.018>.
- El-Shenawy, O. E. (2019). Traditional psychological tests usage in forensic assessment. *Forensic, Legal & Investigative Sciences*, 3, 1–5. <https://doi.org/10.24966/flis-733x/100020>.
- Felthous, A. R. (2010). Psychopathic disorders and criminal responsibility in the USA. *European Archives of Psychiatry and Clinical Neuroscience*, 260, 137–141. <https://doi.org/10.1007/s00406-010-0136-8>.
- Freedman, D., & Woods, G. W. (2018). The developing significance of context and function: Neuroscience and law. *Behavioral Sciences & the Law*, 36(4), 411–425. <https://doi.org/10.1002/bsl.2351>.
- Fuchs, T. (2013). The phenomenology and development of social perspectives. *Phenomenology and the Cognitive Sciences*, 12(4), 655–683. <https://doi.org/10.1007/s11097-012-9267-x>.
- Fuger, K. D., Acklin, M. W., Nguyen, A. H., Ignacio, L. A., & Gowensmith, W. N. (2014). Quality of criminal responsibility reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 37(3), 272–280. <https://doi.org/10.1016/j.ijlp.2013.11.020>.
- Golonka, A. D. (2016). Effects of alcohol on the offender's sanity-Issues of criminal law and psychiatry in light of findings of research. *International Journal of Law and Psychiatry*, 46, 68–73. <https://doi.org/10.1016/j.ijlp.2016.02.016>.
- Gowensmith, W. N., Murrie, D. C., Bocaccini, M. T., & McNichols, B. J. (2017). Field reliability influences field validity: Risk assessments of individuals found not guilty by reason of insanity. *Psychological Assessment*, 29(6), 786–794. <https://doi.org/10.1037/pas0000376>.
- Guarnera, L. A., & Murrie, D. C. (2017). Field reliability of competency and sanity opinions: A systematic review and meta-analysis. *Psychological Assessment*, 29(6), 795–818. <https://doi.org/10.1037/pas0000388>.
- Guivarch, J., Piercecchi-Marti, M. D., Glezer, D., Murdymootoo, V., Chabannes, J. M., & Poinso, F. (2017). Is the French criminal psychiatric assessment in crisis? *International Journal of Law and Psychiatry*, 51, 33–41. <https://doi.org/10.1016/j.ijlp.2017.01.002>.
- Hilton, N. Z., Simpson, A. I., & Ham, E. (2016). The increasing influence of risk assessment on forensic patient review board decisions. *Psychological Services*, 13(3), 223–231. <https://doi.org/10.1037/ser0000068>.
- Hoge, S. K., Bonnie, R. J., Poythress, N., Monahan, J., Eisenberg, M., & Feucht-Haviar, T. (1997). The MacArthur adjudicative competence study: Development and validation of a research instrument. *Law and Human Behavior*, 21(2), 141–179. <https://doi.org/10.1023/A:1024826312495>.
- Höglund, P., Levander, S., Anckarsäter, H., & Radovic, S. (2009). Accountability and psychiatric disorders: How do forensic psychiatric professionals think? *International Journal of Law and Psychiatry*, 32(6), 355–361. <https://doi.org/10.1016/j.ijlp.2009.09.004>.
- Hooper, J. F., McLearn, A. M., & Barnett, M. E. (2005). The Alabama structured assessment of treatment completion for insanity Acquittes (The AlaSATcom). *International Journal of Law and Psychiatry*, 28(6), 604–612. <https://doi.org/10.1016/j.ijlp.2004.09.007>.
- Hopwood, C. J., Thomas, K. M., Markon, K. E., Wright, A. G. C., & Krueger, R. F. (2012). DSM-5 personality traits and DSM-IV personality disorders. *Journal of Abnormal Psychology*, 121(2), 424–432. <https://doi.org/10.1037/a0026656>.
- Jaspers, K. (1968). The phenomenological approach in psychopathology. *The British Journal of Psychiatry: the Journal of Mental Science*, 114(516), 1313–1323. <https://doi.org/10.1192/bjp.114.516.1313>.
- Joubert, P. M., & van Staden, C. W. (2016). Behaviour that underpins non-pathological criminal incapacity and automatism: Toward clarity for psychiatric testimony. *International Journal of Law and Psychiatry*, 49, 10–16. <https://doi.org/10.1016/j.ijlp.2016.04.007>.
- Kacperska, I., Heitzman, J., Bak, T., Leško, A. W., & Opio, M. (2016). Reliability of repeated forensic evaluations of legal sanity. *International Journal of Law and Psychiatry*, 44, 24–29. <https://doi.org/10.1016/j.ijlp.2015.08.028>.
- Kalis, A., & Meynen, G. (2014). Mental disorder and legal responsibility: The relevance of stages of decision making. *International Journal of Law and Psychiatry*, 37(6), 601–608. <https://doi.org/10.1016/j.ijlp.2014.02.034>.
- Kois, L., Wellbeloved-Stone, J. M., Chauhan, P., & Warren, J. I. (2017). Combined evaluations of competency to stand trial and mental state at the time of the offense: An overlooked methodological consideration? *Law and Human Behavior*, 41(3), 217–229. <https://doi.org/10.1037/lhb0000236>.
- Lacroix, R., O'Shaughnessy, R., McNiel, D. E., & Binder, R. L. (2017). Controversies concerning the Canadian not criminally responsible reform act. *The Journal of the American Academy of Psychiatry and the Law*, 45(1), 44–51. Retrieved from <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85014765655&partnerID=40&md5=cbaef4c650c9d5a5f5dbdeeb0d0db8f8>.
- Large, M., Nielsen, O., & Elliott, G. (2009). Reliability of psychiatric evidence in serious criminal matters: Fitness to stand trial and the defence of mental illness. *The Australian and New Zealand Journal of Psychiatry*, 43(5), 446–452. <https://doi.org/10.1080/00048670902817745>.
- McDermott, B. E., Gerbas, J. B., Quanbeck, C., & Scott, C. L. (2005). Capacity of forensic patients to consent to research: The use of the MacCAT-CR. *The Journal of the American Academy of Psychiatry and the Law*, 33(3), 299–307. doi:33/3/299.
- McDermott, B. E., Scott, C. L., Busse, D., Andrade, F., Zozaya, M., & Quanbeck, C. D. (2008). The conditional release of insanity acquittes: three decades of decision-making. *Journal of the American Academy of Psychiatry and the Law*, 36(3), 329–336. PMID: 18802180.
- McSherry, B. (2004). Criminal responsibility, “fleeting” states of mental impairment, and the power of self-control. *International Journal of Law and Psychiatry*, 27(5), 445–457. <https://doi.org/10.1016/j.ijlp.2004.06.002>.
- Meyer, L. F., Taborda, J. G. V., da Costa, F. A., Galego Soares, A. L. A., Mecler, K., & Valença, A. M. (2015). Phenomenological aspects of the cognitive rumination construct. *Trends in Psychiatry and Psychotherapy*, 37(1). <https://doi.org/10.1590/2237-6089-2014-0025>.
- Meynen, G. (2012). An ethical framework for assessments of criminal responsibility: Applying Susan Wolf's account of sanity to forensic psychiatry. *International Journal of Law and Psychiatry*, 35(4), 298–304. <https://doi.org/10.1016/j.ijlp.2012.04.011>.
- Meynen, G. (2013). A neurolaw perspective on psychiatric assessments of criminal responsibility: Decision-making, mental disorder, and the brain. *International Journal of Law and Psychiatry*, 36(2), 93–99. <https://doi.org/10.1016/j.ijlp.2013.01.001>.
- Mossière, A., & Maeder, E. M. (2015). Defendant mental illness and juror decision-making: A comparison of sample types. *International Journal of Law and Psychiatry*, 42–43, 58–66. <https://doi.org/10.1016/j.ijlp.2015.08.008>.
- Murrie, D. C., Bocaccini, 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>.
- Niveau, G., & Sozonets, E. (2001). Criminal responsibility assessment in Switzerland: Changes and continuity. *European Psychiatry*, 16(8), 483–490. [https://doi.org/10.1016/S0924-9338\(01\)00610-1](https://doi.org/10.1016/S0924-9338(01)00610-1).
- O'Sullivan, O. (2018). Autism spectrum disorder and criminal responsibility: Historical perspectives, clinical challenges and broader considerations within the criminal justice system. *Irish Journal of Psychological Medicine*, 35(4), 333–339. <https://doi.org/10.1017/ipm.2017.13>.
- Otto, R. K., Poythress, N. G., Nicholson, R. A., Edens, J. F., Monahan, J., Bonnie, R. J., ... Eisenberg, M. (1998). Psychometric properties of the MacArthur competence assessment tool—criminal adjudication. *Psychological Assessment*, 10(4), 435–443. <https://doi.org/10.1037/1040-3590.10.4.435>.
- Penney, S. (2012). Impulse control and criminal responsibility: Lessons from neuroscience. *International Journal of Law and Psychiatry*, 35(2), 99–103. <https://doi.org/10.1016/j.ijlp.2011.12.004>.
- Resnick, P. J., & Soliman, S. (2012). Planning, writing, and editing forensic psychiatric reports. *International Journal of Law and Psychiatry*, 35(5–6), 412–417. <https://doi.org/10.1016/j.ijlp.2012.09.019>.
- Robinson, R., & Acklin, M. W. (2010). Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 33(3), 131–137. <https://doi.org/10.1016/j.ijlp.2010.03.001>.
- Rogers, R., Seman, W., & Clark, C. R. (1986). Assessment of criminal responsibility: Initial validation of the R-CRAS with the M'Naghten and GBMI standards. *International Journal of Law and Psychiatry*, 9(1), 67–75. [https://doi.org/10.1016/0160-2527\(86\)90017-8](https://doi.org/10.1016/0160-2527(86)90017-8).
- Rogers, R., & Sewell, K. W. (1999). The R-CRAS and insanity evaluations: A re-examination of construct validity. *Behavioral Sciences & the Law*, 17(2), 181–194. [https://doi.org/10.1002/\(SICI\)1099-0798\(199904/06\)17:2<181::AID-BSL338>3.0.CO;2-4](https://doi.org/10.1002/(SICI)1099-0798(199904/06)17:2<181::AID-BSL338>3.0.CO;2-4).
- Rosner, R., & Scott, C. (Eds.). (2017). *Principles and practice of forensic psychiatry*. Boca Raton: CRC Press. <https://doi.org/10.4324/9781315381480>.
- Schleim, S. (2012). Brains in context in the neurolaw debate: The examples of free will and “dangerous” brains. *International Journal of Law and Psychiatry*, 35(2), 104–111. <https://doi.org/10.1016/j.ijlp.2012.01.001>.
- Slovenko, R. (1999). The mental disability requirement in the insanity defense. *Behavioral Sciences & the Law*, 17(2), 165–180. [https://doi.org/10.1002/\(SICI\)1099-0798\(199904/06\)17:2<165::AID-BSL337>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-0798(199904/06)17:2<165::AID-BSL337>3.0.CO;2-Y).
- Stredny, R. V., Parker, A. L. S., & Dibble, A. E. (2012). Evaluator agreement in placement recommendations for insanity acquittees. *Behavioral Sciences & the Law*, 30, 297–307. <https://doi.org/10.1002/bsl.1995>.