

# Tipping the scales of justice: the role of forensic evaluations in the criminalization of mental illness

Katherine E. McCallum\*  and W. Neil Gowensmith

The Denver Forensic Institute for Research, Service, and Training (Denver FIRST), Graduate School of Professional Psychology, University of Denver, Denver, Colorado, USA

An unprecedented number of individuals with mental illness are represented in the criminal justice system. The unending growth of mentally ill populations in the justice system has led to jails and court dockets being increasingly overwhelmed with cases involving mental illness, state hospitals devoting far more beds and resources to forensic cases, and people without a criminal commitment left waiting for mental health services as forensic cases are prioritized. Although a forensic mental health evaluation is only one component of this larger system, common problems with forensic mental health evaluations can exacerbate the criminalization of persons with mental illness in many ways. This article reviews the current literature regarding issues of quality, reliability, and validity of forensic mental health evaluations, discusses the broader impact of these issues, and offers potential solutions for the field.

Received 20 May 2019; Accepted 03 July 2019; First published online 22 October 2019

**Key words:** Forensic psychology, forensic psychiatry, competency to stand trial, forensic mental health evaluation, criminalization of mental illness, sanity evaluation.

## Introduction

For psychologists or psychiatrists conducting forensic evaluations, a forensic psychological report is a work product—one of many reports they will author over the course of their careers. Many forensic evaluators conduct a large number of evaluations per year; for example, Colorado state evaluators conduct an average of 144 competency to stand trial (CST) evaluations per year.<sup>1</sup> For psychologists or psychiatrists who author a large volume of evaluations, some cases may seem routine. Evaluators may fall into a pattern in which many evaluations appear mundane and typical.

However, for the individual being evaluated, each report holds a tremendous amount of influence. Forensic evaluations cover a wide swath of psycholegal referral questions, and they carry a great deal of influence over the lives of those under evaluation. Although forensic evaluators are not triers of fact, judicial decisions are overwhelmingly correlated to opinions of forensic evaluators;<sup>2</sup> studies have shown that judges follow the opinions of evaluators in 76%–99% of cases.<sup>3–6</sup> These opinions can be far-reaching. For example, evaluations of

adjudicative competency or sanity can influence whether a defendant is temporarily detained in a correctional facility, hospital, or released to the community—and they may also ultimately lead to charges being dismissed.<sup>7</sup> Criminal responsibility evaluations can provide the tipping point between acquittal and hospital commitment versus a guilty verdict and imprisonment.<sup>8</sup> Other forensic mental health evaluations may influence whether a parent maintains custody of his/her child, a person is released from a locked facility, a teenager is tried as an adult, a confession is valid, or a plaintiff receives monetary awards. In capital cases, a forensic mental health evaluation can influence whether a defendant is executed.

In addition to the impact on the individual, forensic evaluations have systemic impact. Forensic evaluations that are not conducted within a certain time period can result in a backlog of cases. Many states are currently in the throes of federal lawsuits centered around these delays in evaluation time frames, as defendants with mental illnesses languish in county jails awaiting their evaluations.<sup>9,10</sup> However, when evaluations are conducted too quickly, emerging research shows that forensic opinions are often subject to inaccuracies.<sup>11,12</sup> Further, the quality of reports from state evaluators has been shown to be of mediocre quality in some settings.<sup>13</sup> Reports of poor quality can result in appeals or second opinion requests,

\* Address correspondence to: Katherine E. McCallum, Denver FIRST, University of Denver, Denver, CO 80220-1857, USA.  
(Email: [katemccallumphd@gmail.com](mailto:katemccallumphd@gmail.com))

compounding the backlog even more. Finally, a great deal of research demonstrates that many evaluators are biased by multiple internal and external factors.<sup>14–23</sup> Biased, unreliable, or low quality forensic mental health evaluations deteriorate the fairness of the justice system overall. Given the important systemic and individual impacts of forensic evaluations, it is critical that they are efficient, valid, reliable, and held to high standards of quality. However, an accumulating body of literature suggests that the efficiency, validity, reliability, and quality of these reports have substantial room for improvement.

### Efficiency

Efficiency of conducting forensic evaluations is specifically relevant to CST evaluations, the most frequently ordered mental health evaluations by criminal courts.<sup>25</sup> Courts order an estimated 25,634–51,500 CST evaluations each year nationally, varying from fewer than 50 to approximately 5,000 per year in individual states.<sup>26,27</sup> Numbers of evaluations continue to increase annually. For example, CST evaluations in Wisconsin increased 32.5% from 2010 through 2015,<sup>28</sup> while evaluations in Washington increased 76.3% from 2001 through 2012.<sup>29</sup> Colorado reported a 206% increase in the number of CST evaluations from 2005 to 2014,<sup>30</sup> and Los Angeles county reported a 273% increase from 2010 to 2015.<sup>31</sup> Unsurprisingly, this burgeoning need for CST evaluations has led to long waitlists for evaluations, as states across the nation have struggled to keep pace with the rapid growth of demand. Some states have reported waitlists of more than a year for CST evaluations to be conducted, and several other states are operating under federal or state oversight to ensure that evaluation wait times are reasonable.<sup>10,32</sup> Class action lawsuits in the states of Oregon and Washington successfully lobbied for shorter wait times for CST evaluations and access to competency restoration services.<sup>33,34</sup> Other states are grappling with consent decrees, lawsuits, and legislation on the issue (eg, Alabama, Colorado, Pennsylvania, Nevada, and California).

Some legislation has resulted in specified time frames for evaluations to be completed. The national average number of days from court order to evaluation report date is 31 days.<sup>32</sup> However, there is considerable variability in these time frames. Oregon and Maryland require CST reports within 7 days, North Carolina has a 10-day time frame for defendants awaiting evaluations in jail, Washington as a 14-day time limit, and Rhode Island mandates reports within 15 days. Alternatively, several states (Arkansas, Kansas, Missouri, Montana, among others) extend the deadline to 60 days, while 15 states have no statutorily defined time frame at all. Evaluation time frames are currently being adjusted in several states to accommodate the increasing demand for CST evaluations;

Colorado's current Consent Decree will decrease the current 28-day time frame for evaluations to 21 days in 2020.<sup>1</sup>

Decreasing these CST evaluation time frames may seem like a good solution to long waitlists for these evaluations. Hiring more evaluators to conduct evaluations more quickly certainly ensures that defendants will be evaluated more quickly. However, emerging research suggests that systems may experience unintended negative consequences if evaluations are conducted *too* quickly.

Most CST evaluators find between 20% and 40% of defendants incompetent to stand trial (IST).<sup>35</sup> However, recent data shows that the timing of CST evaluations has a substantial impact on evaluators' opinions. Hawaii data indicates that IST rates were nearly 40% higher in evaluations conducted within 7 days of the court order than those conducted beyond 7 days.<sup>12</sup> Washington shows a 50% IST rate for CST evaluations conducted within 7 days,<sup>34</sup> as does the state of Maryland.<sup>36</sup> This trend seems to be especially true for defendants with psychotic and substance use diagnoses. Defendants found IST within 7 days in the Hawaii sample were more likely to have substance-related and/or psychotic disorders. Additionally, data from a large dataset in Texas show that the rate of IST opinions rose approximately 25% in defendants with a schizophrenia-spectrum diagnosis or a substance-related disorder when evaluations were completed within 10 days of the court order.<sup>11,37</sup> It seems that quick turnaround time frames for CST evaluations may artificially inflate the numbers of IST opinions in some cases.

Aside from the effects of quick turnaround evaluation on the accuracy of individual competency opinions, these aggressive time frames may also have systemic effects. If courts and attorneys are assured that they will receive CST opinions within 1–15 days, they may in turn request them more often. Paradoxically, shorter time frames for evaluations may incentivize courts and attorneys to order them more frequently—increasing referral numbers, evaluator caseloads, and ultimately defendant wait times. No data is available currently to monitor this possibility; however, Oregon and Washington's recent drastic reductions in evaluation time frames would provide a good naturalistic opportunity to review CST evaluation referral rates, evaluator opinion rates, and systemic effects on wait times.

### Quality

The forensic evaluator has considerable influence in how information is included and presented to the court. A quality forensic evaluation is not only defined by good grammar, syntax, and readability, but also the inclusion of critical elements and a well-supported answer to the psycholegal question posed. Poorly written reports could result in a myriad of negative consequences, such as an increasing need for second opinions, compounding the problems noted above. Further, providing the trier of fact

with inaccurate or incomplete information risks an unfair trial process.

Several reviews of forensic evaluations have revealed deficits in report quality. Robinson and Acklin found that many critical components, such as historical information, collateral information, or a rationale for the forensic opinion, were not included in 150 CST reports, resulting in an overall “poor” quality rating.<sup>14</sup> A similar review of 150 conditional release reports from Hawaii found that evaluators documented informed consent in only half (52%) of reports and provided a rationale for their opinion of dangerousness in only 34.7% of reports.<sup>13</sup> Skeem, Golding, Cohn, and Berge found that CST evaluators in Utah rarely linked the defendant’s clinical presentation to the forensic opinion and rarely explained their rationale for arriving at that opinion.<sup>38</sup>

The rising demand for CST evaluations can further threaten the quality of reports by pressuring state systems to widen the pool of qualified evaluators. However, the discipline of the evaluator likely has less influence on report quality than requisite training. Originally, criminal courts only qualified psychiatrists as forensic experts before slowly including psychologists.<sup>39</sup> Despite early concerns about report quality of psychologists, no substantive differences in report quality have consistently been found between disciplines.<sup>40,41</sup> However, both psychiatry and psychology have developed an infrastructure of forensic specialty training through pre- and postdoctoral programs, dedicated subspecialty professional organizations, and high-impact forensic journals.<sup>9</sup> This training and infrastructure is crucial to for evaluators to attain minimum quality standards, as forensic evaluations are often complicated, requiring evaluators to consider difficult psychological, legal, and cultural nuances.<sup>15,17,42–44</sup> Simply expanding the pool of eligible evaluator disciplines runs the serious risk of experts offering poorly formulated opinions, unless enough forensic training and infrastructure within that discipline exists.

## Reliability

When a psychological evaluation is used in either criminal or civil court, it is expected to be objective and reliable. In admitting either psychiatric or psychological testimony, courts consider factors established by the *Daubert* decision, such as whether it is scientifically valid, whether the “theory or technique can be (and has been) tested,” and whether there is a “known potential rate of error.”<sup>45</sup> Several scholars have argued psychiatric and psychological testimony do not meet this standard, citing a lack of available base rates, errors in clinical decision making, and a range of theories too diverse and inconsistent to result in reliable opinions.<sup>23,24</sup> Of course, psychiatric and psychological testimony continued to be present in courtrooms despite these criticisms. In the time since, literature has only accumulated regarding threats to reliability of forensic opinions.

Variability between forensic evaluators is concerningly high. In a sample of 59 evaluators who conducted a total of 4,498 evaluations of legal sanity, seven evaluators opined the individual was sane in 100% of their evaluations, whereas three evaluators opined the individual was sane in 50% of their evaluations.<sup>21</sup> In a sample of 15 evaluators who each completed at least 100 CST evaluations, rates of incompetency findings ranged from 1.7% to 27.9%.<sup>20</sup> Similar discrepancies among evaluators have also been found in the use of forensic assessment instruments. For example, Boccaccini, Turner, and Murrie found that some evaluators assigned consistently higher scores on the PCL-R compared to other evaluators.<sup>46</sup>

This variability may simply be caused by differences between evaluators and the sometimes subjective nature of interpreting statutory language of mental health law. For example, one psychologist may have a higher decision threshold for determining a defendant’s “sufficient present ability to consult with his lawyer with a reasonable degree of rational understanding” (emphasis added).<sup>47</sup> Indeed, Mossman found individual differences in decision thresholds between evaluators in CST evaluations.<sup>48</sup> Though he acknowledged individual differences in feelings and beliefs may contribute to differences in decision thresholds, he also discussed several other influencing variables, such as internal and external expectations and conventions in specific agencies, knowledge of local judicial decision-making trends, or differing understandings of constructs underlying adjudicative competence.

A now well-established threat to reliability is bias of the evaluator toward the side that retained his/her services.<sup>15,17,19</sup> This “adversarial allegiance” effect is present even when evaluators score objective, structured risk assessments designed to mitigate subjective bias.<sup>49</sup> Earlier research has suggested evaluators’ opinions are also influenced by the fees they earn.<sup>23,24,50</sup> Additionally, research has found discrepancies in evaluator opinions related to the racial and/or ethnic background of the defendant.<sup>16,18,22</sup> Clearly, these threats to the reliability of forensic opinions must be addressed if evaluations are to reach the highest standards of objectivity and neutrality. The reliability of evaluator opinions is surprisingly low across nearly all psycholegal referral questions,<sup>51</sup> most likely for many of the reasons articulated previously. However, this poor reliability can cause differential and undue harm to certain defendants. Defendants of color, for example, should not be routinely found IST more often than Caucasian defendants; unfortunately, some scholars have shown this to be true in some samples.<sup>16,18</sup> If race or skin color is truly a differential factor for some CST evaluators, then the notion of objective and reliable opinions is clearly tainted. The final forensic opinion (and subsequent judicial adjudication) should not depend on race, skin color, fees, the individual evaluator, or any other idiosyncratic factors. However, if the realities of implicit and explicit bias—as well as other factors related

to low reliability across evaluators—are not adequately addressed, then some subpopulations of persons being forensically evaluated may be at risk for injustice.

### Validity

The above reliability concerns undoubtedly undermine the validity of forensic evaluations. Unfortunately, research evaluating the accuracy of forensic opinions in the field is substantially lacking because the ground truth is often unknown.<sup>52</sup> Forensic evaluators typically never know if they “got it right” with their opinion, but some research has attempted to measure the accuracy of forensic psychologists. In 2010, Mossman et al. found impressively high accuracy in five forensic evaluators asked to review redacted court reports and provide CST opinions.<sup>53</sup> However, the evaluators provided opinions on a graded scale, rather than a binary one typical for most CST evaluations (competent or not). Though the finding may be encouraging to some, the incredible variability of forensic evaluator opinions summarized above suggest that validity is likely much poorer in the field. Agreement between evaluators intuitively seems to lead to greater accuracy and evidence seems to support this. In Hawaii, when three evaluators evaluated the same defendant for conditional release and unanimously opined release was appropriate, most evaluators “got it right”; about 74% accurately predicted those who would be rehospitalized within three years. Interestingly, agreement of evaluators in their predictions was important. In cases in which three evaluators independently opined in favor of release, only 29.6% of defendants were re-hospitalized—compared to a 71.4% rehospitalization rate in cases in which evaluators disagreed as to the person’s readiness for release.<sup>54</sup>

### Corrective strategies

Forensic evaluations play an important role in the due process of adjudicating criminal cases with mental health components. Poor forensic standards can unwittingly collude with discriminatory social and judicial practices, such that certain defendants undergoing poor evaluation practices may be more likely to be found competent, sane, dangerous, and so on. It is critical that forensic evaluations provide the court with objective, comprehensive, and accurate information so that the court can utilize the data and opinions in a just manner. Although we have outlined several potential areas in which evaluations can be misused, we now turn to potential mechanisms to mitigate against these threats to evaluation reliability, validity, and quality.

#### *Ensuring high-quality evaluations*

One possible strategy to improve quality and reliability is training and certification of forensic evaluators.

Approximately half of the states in the US have implemented formal certification processes for psychologists to conduct CST evaluations.<sup>40</sup> In Hawaii, forensic psychologists and psychiatrists who attended a three-day certification training subsequently showed improvement in the quality and reliability of their CST evaluations.<sup>14,55</sup> However, much more research examining the outcomes of these training is needed to truly assess effectiveness.<sup>9</sup> The use of a peer-reviewed evaluation report system can also be used to identify evaluation areas or specific evaluators that need improvement. Although little empirical evidence exists as to the incremental utility of standardized training and peer reviews of reports, it seems safe to assume that high-quality training, and maintenance of high standards among evaluators and their work, will guard against the misuse of forensic evaluations in many contexts.

#### *Evaluation parameters*

The amount of time allotted to complete forensic evaluations appears to matter. Completing evaluations too quickly may lead to inflated rates of incompetent findings among CST evaluations.<sup>11,12,38</sup> However, completing evaluations too slowly can lead to defendants experiencing unconstitutional wait times in jail.<sup>9</sup> To strike a balance, The National Judicial College recommends that CST evaluations be conducted within 15 to 30 days of the initial court order.<sup>56</sup> Although more research is needed to fully assess the optimal time frame of forensic evaluations, the current evidence suggests 15 to 30 may be ideal.<sup>9</sup> In addition to time frames, other conditions of evaluations may affect quality. The ideal testing environment for psychological testing (quiet, uninterrupted time, enough working space, few distractions) is often inaccessible in correctional facilities. Although flexibility is important, evaluators must be willing to require minimally acceptable assessment standards within correctional facilities, lest their evaluation results be potentially tainted.

Even with the ideal evaluation parameters in place, proficient evaluators are paramount in ensuring improved report quality. Creating and maintaining an infrastructure of training and certification of forensic evaluators is a critical step. Additionally, reasonable workloads and competitive salaries will also bolster a workforce of capable evaluators.

#### *Identifying and mitigating potential bias*

Employing strategies to identify and mitigate against internal and external evaluation biases is important in ensuring optimal levels of evaluation reliability and validity. Self-monitoring should be a systematic, consistent process for evaluators. However, not all self-monitoring processes are created equal; selecting effective mechanisms is critical. Humans are simply not adept at monitoring their own biases through introspection.<sup>57</sup> Evaluators, like all people,

are not only prone to the effects of bias but also to the “bias blind spot”—an insidious phenomenon in which people do not recognize their own biases.<sup>58</sup> Therefore, simply looking in the metaphorical mirror to check one’s biases is rarely effective.

Instead, following social psychology’s tenets, evaluators should use objective, measurable data to guide the identification and mitigation of biases.<sup>59</sup> For example, evaluators could collect and analyze data from their own evaluations to assess measurable outcomes. Brodsky suggested practitioners can use such a database to measure their own objectivity.<sup>60,61</sup> Some practitioners who have engaged in such self-studies have found surprising and illuminating results.<sup>16,62</sup> Further, if self-monitoring were a standard of practice and data were aggregated across evaluators and settings, we would have far richer information about base rates of opinions in relation to a myriad of potentially significant factors (i.e. geographic locations, relationships to attorneys, fees, workplace environments).

Such an approach can be an especially valuable tool in mitigating the criminalization of mental illness in the forensic evaluation context. If robust data is captured on all forensic evaluations submitted to the court (even, for example, within a state-employed forensic evaluator pool), analyses could illuminate areas of potential discrimination or differentiation. Certain jurisdictions might be found to refer a high number of spurious cases. Specific evaluators may have unreasonable thresholds for psycholegal criteria. Perhaps defendants from certain races may be found more dangerous than those from other races. Perhaps fees, the referring attorney, or the setting in which evaluations are conducted lead to differential rates of opinions. This sort of data would be tremendously informative in ensuring that systemic biases are identified and addressed in a practical and tangible way, optimizing the reliability and validity of the forensic evaluation process. In doing so, the evaluation process would be less likely to be an unintentional contributor to the criminalization of persons with mental illness.

## Conclusions

An unprecedented number of people with mental illness are being funneled to the criminal court in order to access mental health care. Jails and court dockets are increasingly overwhelmed with cases involving mental illness, state hospitals devote far more beds and resources to forensic cases, and people without a criminal commitment are increasingly left waiting for mental health services as forensic cases are prioritized. Each component of the forensic mental health process likely plays a role in maintaining this trend. However, some forensic evaluators appear to operate in a figurative vacuum, assuming that their roles have little impact on the criminalization of people with mental illness; after

all, evaluators do not arrest people, raise forensic referral questions, order evaluations, or make any final judicial decisions. However, poorly conducted forensic evaluations can indeed exacerbate the criminalization of persons with mental illness in many ways. Evaluations that suffer from poor reliability, low quality, or evaluator biases can extend criminal commitments unnecessarily, cause delays in the resolution of the case, or posit inaccurate opinions. Research is consistent that evaluation reliability, quality, validity, and accuracy all have room for improvement. In addition, forensic evaluations are inevitably vulnerable to errors and bias, as is all human decision making. However, when the stakes for individuals and systems are so high, forensic evaluators should always strive for the highest standards. As a field and as individual evaluators, this will involve examining ourselves and our field objectively, ensuring that the standards for our evaluators and our field are as high as possible.

## Disclosures

Katherine McCallum and W. Neil Gowensmith report that they have developed an app for mobile phones that allows forensic evaluators to track base rates of forensic opinions. One of the uses of the app is to identify and mitigate against bias. In the current article, the authors encourage practicing forensic evaluators to keep track of their base rates, variables that could affect opinions, and so on. Although the authors do not highlight their app in the current article, and although several options for identifying and mitigating bias were provided, Dr. McCallum and Dr. Gowensmith both state that some readers might equate the recommendation that evaluators monitor their work for bias with the authors’ financial interest in the mobile app. To be clear, their ultimate goal is to encourage evaluators to monitor their work reliably and carefully, regardless of methodologies used.

## REFERENCES:

1. Center for Legal Advocacy v. Bicha et al, *U.S. District Court of Colorado, Case No. 1:2011cv02285* (2011).
2. Redding RE, Murrie DC. Judicial Decision Making about Forensic Mental Health Evidence. In: A. M. Goldstein, ed. *Forensic Psychology: Emerging Topics and Expanding Roles*. Hoboken, NJ: John Wiley & Sons Inc; 2007:683–707.
3. Acklin MW, Fuger K, Gowensmith WN. Examiner agreement and judicial consensus in forensic mental health evaluations. *J Forensic Psychol Pract*. 2015;15(4):318–343. doi:10.1080/15228932.2015.1051447.
4. Gowensmith WN, Murrie DC & Boccaccini MT. Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law Hum Behav*. 2012;36(2):130–139. doi:10.1037/h0093958.

5. Zapf PA, Hubbard KL, Cooper VG, *et al.* Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *J Forensic Psychol Pract.* 2004;**4**(1):27–44. doi:10.1300/J158v04n01\_02.
6. Cruise KR, Rogers R. An analysis of competency to stand trial: an integration of case law and clinical knowledge. *Behav Sci Law.* 1998; **16**(1):35–50. doi:10.1002/(SICI)1099-0798(199824)16:1<35::AID-BSL292>3.0.CO;2-4.
7. Mowen E. Charges dismissed in crash which claimed family. *Register Herald.* July 7, 2018. [www.registerherald.com/news/22514/charges-dismissed-in-crash-which-claimed-family](http://www.registerherald.com/news/22514/charges-dismissed-in-crash-which-claimed-family). Accessed April 19, 2019.
8. Eaton E. Man who killed Elmendorf police chief found not guilty by reason of insanity. *San Antonio Express News.* July 2, 2018. <http://www.expressnews.com/news/local/article/Man-who-killed-Elmendorf-police-chief-found-not-13044383.php>. Accessed April 19, 2019.
9. Gowensmith WN. Resolution or resignation: The role of forensic mental health professionals amidst the competency services crisis. *Psychol Public Policy Law.* 2019;**25**(1):1–14. doi:10.1037/law0000190.
10. Locklair B. Due process problems with civil commitment of incompetent defendants: The current round of litigation and the next. Paper presented at: The American Psychology-Law Society Annual Conference; 2016, Atlanta, GA.
11. Bryson CN, Boccaccini MT, Gowensmith WN, *et al.* Does time matter in competency to stand trial evaluations? Paper presented at: The annual meeting of the American Psychology-Law Society; March 2019; Portland, OR.
12. Gowensmith WN, Metroz H, Bratcher J. The impact of timing on competency to stand trial evaluations. Paper presented at: The annual meeting of the American Psychology-Law Society; March 2016; Atlanta, GA.
13. Nguyen AH, Acklin MW, Fuger K, *et al.* Freedom in paradise: Quality of conditional release reports submitted to the Hawaii judiciary. *Int J Law Psychiatry.* 2011;**34**(5):341–348. doi:10.1016/j.ijlp.2011.08.006.
14. Robinson R, Acklin MW. Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary. *Int J Law Psychiatry.* 2010;**33**:131–137. doi:10.1016/j.ijlp.2010.03.001.
15. Boccaccini, MT, Chevalier CS, Murrie DC, *et al.* Psychopathy Checklist—Revised use and reporting practices in sexually violent predator evaluations. *Sex Abuse.* 2017;**29**(6):592–614. doi:10.1177/1079063215612443.
16. Parker G. Come see the bias inherent in the system! *J Am Acad Psychiatry Law.* 2016;**44**(4):411–414.
17. Chevalier CS, Boccaccini MT, Murrie DC, *et al.* Static-99R reporting practices in sexually violent predator cases: Does norm selection reflect adversarial allegiance? *Law Hum Behav.* 2015;**39**(3):209–218. doi:10.1037/lhb0000114.
18. McCallum KE, MacLean N, & Gowensmith WN. The impact of defendant ethnicity on the psycho-legal opinion of forensic evaluators. *Int J Law Psychiatry.* 2014;**39**:6–12. doi:10.1016/j.ijlp.2015.01.015.
19. Murrie DC, Boccaccini MT, Guarnera LA, *et al.* Are forensic experts biased by the side that retained them? *Psychol Sci.* 2013;**24**(10):1889–1897. doi:10.1177/0956797613481812.
20. Murrie DC, Boccaccini M, Zapf PA, *et al.* Clinician variation in findings of competence to stand trial. *Psychol Public Policy Law.* 2008;**14**:177–193. doi:10.1037/a0013578.
21. Murrie DC & Warren JI. Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Prof Psychol.* 2005;**36**:519–524. doi:10.1037/0735-7028.36.5.519.
22. Hagen MA. *Whores of the Court: The Fraud of Psychiatric Testimony and the Rape of American Justice.* New York, NY, US: Harper Collins Publishers; 1997.
23. Bergeron ML. The use of psychiatric expertise in the forensic context: Balm or blunder. *Windsor Yearb Access Justice.* 1994;**14**:221.
24. Faust D, Ziskin J. The expert witness in psychology and psychiatry. *Science.* 1988;**241**(4861):31–35. doi:10.1126/science.3291114.
25. Melton GB, Petrila J, Poythress NG, *et al.* *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers* (3rd ed.). New York, NY: The Guilford Press; 2007.
26. Warren JI, Chuahan P, Kois L, *et al.* Factors influencing 2260 opinions of defendants' restorability to adjudicative competency. *Psychol Public Policy Law.* 2013;**19**:498–508.
27. Fitch WL. *Forensic mental health services in the United States (Report No. 3, HHSS2834200IT).* Alexandria, VA: National Association of State Mental Health Program Directors; 2014.
28. Wisconsin Department of Health Services. *Annual Report (Report No. P-00568).* Madison, WI: Wisconsin Department of Health Services; 2015.
29. Joint Legislative and Audit Review Committee. *State of Washington Final Report: Competency to Stand Trial, Phase II (Report No. 14-1).* Olympia, WA: Joint Legislative and Audit Review Committee; 2014.
30. Colorado Department of Human Services. Needs analysis: Current status, strategic positioning, and future planning. 2015. <http://www.colorado.gov/cs/Satellite?c=Page&childpagename=CDHS-BehavioralHealth/CBONLayout&cid=1251662741340&pagename=CBONWrapper>. Accessed September 13, 2015.
31. Sewall A. L.A. County supervisors order report on unexplained surge in mental competency cases. *Los Angeles Times;* March 8, 2016. <https://www.latimes.com/local/lanow/la-me-ln-mental-competency-cases-20160308-story.html>. Accessed April 15, 2019.
32. Gowensmith WN, Murrie DC, Packer IK. *Report in Response to the Trueblood v. State Washington's Department of Social and Health Services.* State of Washington: Office of Attorney General; 2015.
33. Oregon Advocacy Center v. Mink. No. 02-35530, 9th Cir.; 2003.
34. Trueblood v. State of Washington Department of Human and Social Services. No. 2:2014cv01178, Washington Western District Court; 2015.
35. Pirelli G, Gottdiener WH, Zapf PA. A meta-analytic review of competency to stand trial research. *Psychol Public Policy Law.* 2011; **17**(1):1–53. doi:10.1037/a0021713.
36. Roskes, personal communication, December 12, 2014
37. Bryson CN, Boccaccini MT, Gowensmith WN, *et al.* Time matters in competency to stand trial evaluations. Poster presented at: The Annual Meeting of the American Psychology-Law Society; March 2018; Memphis, TN.
38. Skeem JL, Golding SL, Cohn NB, *et al.* Logic and reliability of evaluations of competence to stand trial. *Law Hum Behav.* 1998;**22**(5):519–547. doi:10.1023/A:1025787429972.
39. Gowensmith WN, Pinals DA, & Karas AC. States' standards for training and certifying evaluators of competency to stand trial. *J Forensic Psychol Pract.* 2015;**15**(4):295–317. doi:10.1080/15228932.2015.1046798.
40. Petrella RC & Poythress NG. The quality of forensic evaluations: An interdisciplinary study. *J Consult Clin Psychol.* 1983;**51**(1):76–85. doi:10.1037/0022-006X.51.1.76.
41. Poythress NG, Otto RK, & Heilbrun K. Pretrial evaluations for criminal courts: Contemporary models of service delivery. *J Ment Health Adm.* 1991; **18**:198–208.
42. Kois L, Pearson J, Chauhan P, *et al.* Competency to stand trial among female inpatients. *Law Hum Behav.* 2013;**37**(4):231–240. doi:10.1037/lhb0000014
43. Mossman D, Noffsinger SG, Ash P, *et al.* AAPL practice guideline for the forensic psychiatric evaluation of competence to stand trial. *J Am Acad Psychiatry Law.* 2007;**35**:S3–S72.

44. Pinals DA, Tillbrook CE, & Mumley DL. Practical application of the MacArthur competence assessment tool—criminal adjudication (MacCAT-CA) in a public sector forensic setting. *J Am Acad Psychiatry Law* 2006;**34**:179–188.
45. Daubert v. Merrell Dow Pharmaceuticals, Inc., 405 U.S. 597 (1993).
46. Boccaccini MT, Turner D, & Murrie DC. Do some evaluators report consistently higher or lower psychopathy scores than others? Findings from a statewide sample of sexually violent predator evaluations. *Psychol Public Policy Law*. 2008;**14**:262–283. doi:10.1037/a00114523.
47. Dusky v. United States, 362 U.S. 402 (1960).
48. Mossman D. When forensic examiners disagree: Bias, or just inaccuracy? *Psychol Public Policy Law*. 2013;**19**(1):40–55. doi:10.1037/a0029242.
49. Murrie DC, Boccaccini MT, Turner D, et al. Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychol Public Policy Law*. 2009;**15**:19–53. doi:10.1037/a0014897.
50. Callahan LA, & Silver E. Factors associated with the conditional release of persons acquitted by reason of insanity: A decision tree approach. *Law Hum Behav*. 1998;**22**:147–163. doi:10.1023/A:1025790003139.
51. Gowensmith WN, Murrie DC, Boccaccini MT. Field reliability of competence to stand trial opinions: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law Hum Behav*. 2012;**36**(2):130–139. doi:10.1037/h0093958.
52. Dror IE, Murrie DC. A hierarchy of expert performance applied to forensic psychological assessments. *Psychol Public Policy Law*. 2018;**24**(1):11–23. doi:10.1037/law0000140.
53. Mossman D, Bowen MD, Vanness DJ, et al. Quantifying the accuracy of forensic examiners in the absence of a “gold standard.” *Law Hum Behav*. 2010;**34**(5):402–417. doi:10.1007/s10979-009-9197-5.
54. Gowensmith WN, Murrie DC, Boccaccini MT, et al. Field reliability influences field validity: Risk assessments of individuals found not guilty by reason of insanity. *Psychol Assess*. 2017;**29**(6):786–794. doi:10.1037/pas0000376.
55. Gowensmith WN, Sledd M, & Sessarego S. The impact of stringent certification standards on forensic evaluator reliability: Further analysis. Paper presented at: The Annual Meeting of the American Psychology–Law Society; March 2015; San Diego, CA.
56. National Judicial College. Mental Competencies: Best practices manual. 2011–2012. <http://www.mentalcompetency.org/pdf/BP-Model.pdf>. Accessed September 13, 2015.
57. Gowensmith WN & McCallum KE. Mirror, mirror on the wall, who’s the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations. *S Afr J Psychol*. 2019;1–12. doi:10.1177/0081246319835117.
58. Pronin E, Kugler, MB. Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *J Exp Soc Psychol*. 2007;**43**(4):565–578.
59. Neal T, Grisso T. The cognitive underpinnings of bias in forensic mental health evaluations. *Psychol Public Policy Law*. 2014;**20**(2):200–211. doi:10.1037/a0035824.
60. Brodsky SL. *Testifying in Court: Guidelines and Maxims for the Expert Witness*. Washington, DC: American Psychological Association; 1991.
61. Brodsky SL. *The Expert Witness*. Washington, DC: American Psychological Association; 1999.
62. Gowensmith WN, McCallum KE, Nadkarni L, et al. Monitoring potential bias within a forensic evaluation agency. Paper presented at: The Annual Meeting of the American Psychology–Law Society; March 2019; Portland, OR.