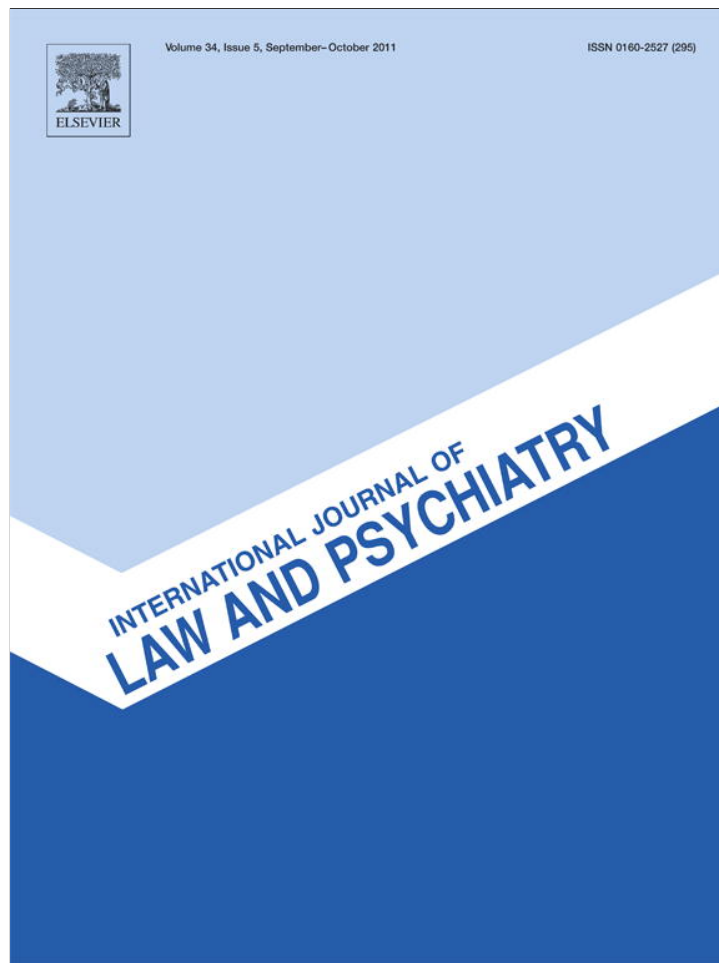


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

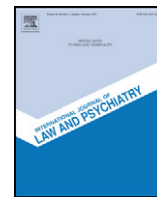
Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at [SciVerse ScienceDirect](http://www.sciencedirect.com)

International Journal of Law and Psychiatry



Freedom in paradise: Quality of conditional release reports submitted to the Hawaii judiciary

Annie H. Nguyen^a, Marvin W. Acklin^{a,*}, Kristen Fuger^a, W.N. Gowensmith^b, Lawrie A. Ignacio^a^a Argosy University Hawaii Campus, Honolulu, HI, United States^b Adult Mental Health Division, State of Hawaii, Department of Health, Honolulu, HI, United States

ARTICLE INFO

Available online 13 September 2011

Keywords:

Forensic reports
Conditional release
Release decision-making

ABSTRACT

Annually thousands of insanity acquitees are released from mental hospitals when they are no longer determined to be dangerous. This research examined quality of post-acquittal Conditional Release (CR) reports submitted to the Hawaii Judiciary. Hawaii utilizes a “three panel” system for assessing trial felony competency, criminal responsibility, and conditional release, where typically two psychologists (one Department of Health and one community-based) and one community-based psychiatrist submit independent reports to the Court. One hundred fifty CR reports were rated using a 44-item report quality measure. Interrater reliability trials indicated good to excellent agreement between quality ratings. Overall level of report quality was poor regardless of examiners’ professional discipline, employer, or board certification status. Concordance rates for CR opinions were poor. Level of agreement between the judicial determination and majority recommendations was also poor. Reasons for the poor quality and level of agreement are discussed with recommendations for report quality improvement, including standardization of procedures and use of forensic risk assessment instruments.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Following adjudication, many insanity acquitees are released into the community under Conditional Release (CR) programs. Courts and public safety officials have the responsibility to assess the degree of danger posed by insanity acquitees. Typically, risk assessments performed by forensic mental health examiners provide release-related information to the courts. Conditional release programs balance competing public interests and individual civil liberties. Conditional release is defined as “monitored community treatment, which provides structure and a process to reintegrate the mentally ill offender into society while providing public protection through the revocation process” (Bloom, Williams, & Bigelow, 1991). CR programs have also been referred to as “outpatient status” (Hafemeister, 1998). CR programs typically provide monitored psychiatric treatment as well as addressing other needs the individual might have such as housing, medication, case management (Bloom et al., 1991; Hafemeister, 1998), emergency services, nursing services and substance abuse evaluations (Bertman-Pate et al., 2004). CR programs can mandate drug testing, medication compliance, and psychotherapy (Wiederanders, 1992; Wiederanders, Bromley, & Choate, 1997). CR programs operate in many states including Hawaii, California, Florida, Louisiana, Ohio, Oregon, New York, Connecticut, Maryland, and

Oklahoma (Bertman-Pate et al., 2004; Bloom et al., 1991; Heilbrun & Griffin, 1998; Stafford & Karpawich, 1997; Wiederanders, 1992) and other countries such as England, Wales, Scotland (Grounds, Snowden, & Taylor, 1993), the Netherlands, and Canada (Evans, 1999; Wiederanders et al., 1997), although significant differences exist from program to program.

While therapeutic jurisprudence and conditional release are recent trends in the United States, Canada has had a “strong commitment to conditional release as an instrument of public safety as well as a humanitarian measure” which extends over the past 100 years (Evans, 1999). In 1970, the United States Supreme Court’s landmark mental health law decision (*O’Connor v. Donaldson*, 1975), the Court ruled it unconstitutional to indefinitely confine a person on the basis of mental illness alone and that a non-dangerous person had a right to liberty. Two years later, the U.S. Supreme Court held that a person could only be confined when he or she presents an immediate danger to self or others (*Lessard v. Schmidt*, 1974). More recently, the Supreme Court ruled that a person must be both mentally ill and dangerous to be confined involuntarily (*Foucha v. Louisiana*, 1992). In line with these judicial determinations, conditional release programs which have emerged in the last several decades attempt to balance individual civil liberties, public safety, and the least restrictive alternatives.

The process of CR varies by jurisdiction, which may or may not include judicial review (Stafford & Karpawich, 1997). The process of revocation or re-hospitalization also depends on the jurisdiction and the particular requirements of each program. Most CR candidates

* Corresponding author at: 850 W. Hind Drive Suite 203, Honolulu, HI 96821, United States.

E-mail address: acklin@hawaii.edu (M.W. Acklin).

are single, Caucasian males in their late thirties and have a diagnosis of schizophrenia or other psychotic disorders (Stafford & Karpawich, 1997; Wiederanders et al., 1997). The average length of hospitalization before CR is granted varies by state. In Oregon, the average length of hospitalization is 18 months; in California more than 36 months (Wiederanders et al., 1997). In most cases, CR programs provide communities with substantial cost savings; for example, California's CONREP program is one-fifth of the hospitalization cost per client (Wiederanders, 1992). Exact cost savings figures are difficult to calculate, but it is generally agreed that CR programs are more cost effective and usually cost one-fourth the expense of hospitalization (Wiederanders et al., 1997). In addition to being cost effective, CR programs also reduce the length of hospitalizations (Bloom et al., 1991).

Conditional release programs reduce the intense demand for long-term predictions of dangerousness and violence risk, since individuals can be re-hospitalized without lengthy procedures. Without CR programs, individuals are likely to be hospitalized for longer periods of time due to the concern for public safety, and the difficulty in making long range predictions of dangerousness. CR programs not only protect the public, but also protect the mentally disordered offender by putting in place a mechanism for a safe, supervised transition into the community.

The number of individuals eligible for or on conditional release at any one time varies by jurisdiction. Canada's National Parole Board reports 8434 individuals on conditional release, which represents approximately .02% of the population. In Victoria, Australia, 35.5% or 8879 of 24,973 persons were granted CR in the years between 1990 and 2000 (Segal & Burgess, 2006). Connecticut's Psychiatric Security Review Board oversees the supervision and treatment of all acquittees: of 152 individuals being supervised, 22 were on conditional release (2009). In Hawaii, there are approximately 500 persons on CR; this is believed to be the largest CR program per capita in the United States (Gowensmith, 2008).

Individuals who are conditionally released and have sufficient levels of supervision and support do not pose a significant danger to the public, whereas unconditionally released post-acquittal individuals are more likely to re-offend at higher rates than their supervised counterparts (Wiederanders, 1992). A study of 44 acquittees deemed NGRI had a 25% rehospitalization rate and a 5% rearrest rate over two years (Cavanaugh & Wasyliw, 1985). Other studies have found rehospitalization rates of 11% (Silver, 1983) and 18% (Rogers & Wettstein, 1984). A study of 295 NGRI acquittees had re-hospitalization rates that varied by year across a four-year period from 1980 to 1983 that were 78%, 62%, 31%, and 20% respectively (Rogers & Wettstein, 1984). Persons released on CR in Hawaii have a 13% rehospitalization rate and an 11% rearrest rate (Gowensmith et al., *in press*).

Conditional release revocation rates vary by program and can be as low as 20% (Stafford & Karpawich, 1997; Wiederanders, 1992) to 34% (Bertman-Pate et al., 2004) to as high as 49% in Oregon and California and 55% in New York (Wiederanders et al., 1997). A study of NGRI acquittees by Lamb, Weinberger, and Gross (1988) found a 48% conditional release revocation rate after five years. Comparisons between conditionally and unconditionally released persons suggest that CR programs may reduce the number of serious or violent offenses and arrests (Wiederanders, 1992). In a meta-analysis by Wiederanders et al. (1997) of three state programs, there was a 10% re-arrest rate. Other studies have re-arrest rates of 13% (Rogers, Bloom, & Manson, 1986), 22% (McGreevy, Steadman, Dvoskin, & Dollard, 1991), and 24% (Silver, 1983).

Violence risk assessments are a critical aspect of court-ordered forensic evaluations given the need to protect public safety (Goldstein, 2003; Heilbrun, 2010). An important component of release decision-making is the prediction, assessment, and management of risk, particularly violence risk (Heilbrun, 1991). Petrila (2007) identifies four primary types of risk assessments: 1) unstructured; 2) structured professional

judgment; 3) anamnestic risk assessment, and 4) actuarial risk assessment. Petrila proposes that long-term assessment of risk should be based on formal probabilistic assessment since unstructured risk decisions produce low reliability. Structured risk decisions produce higher reliability (Poythress, Bonnie, Monahan, Otto, & Hoge, 2002). Unstructured risk assessments tend to grossly overestimate violence risk (Kozol, Boucher, & Garofolo, 1972), underestimate the risk of violence in women (Lidz, Mulvey, & Gardner, 1993), have poor sensitivity, but good specificity (Kozol et al., 1972; Lidz et al., 1993; Thornberry & Jacoby, 1979). Clinicians who rely on unstructured clinical violence risk assessments make predictions that are only slightly better than chance (Mossman, 1994).

Minimizing errors in release decision-making is important in order to reduce error: the number of false positive (prediction of high risk when no violence occurs) and false negative (prediction of low risk when violence occurs) predictions (Webster & Hucker, 2003). False positive predictions potentially can potentially impinge upon the civil liberties of the mentally disordered offender, while false negative predictions may place public safety at risk. Studies have found differing false-positive rates, ranging from 65% (Kozol et al., 1972) to 86% (Thornberry & Jacoby, 1979). These high false positive rates likely indicate that more people have been wrongly considered dangerous, but without comprehensive follow up and tracking of individuals who are released, it is virtually impossible to ascertain exact base rates of recidivism and violence. The establishment of accurate base rates can be difficult due to changing diagnostic categories, inaccurate diagnoses, changes in the law, and undetected offenses (Webster & Hucker, 2003). Despite the difficulty related to obtaining accurate base rates, they must be considered in the release decision-making process (Conroy & Murrie, 2007; Monahan, 1981; Webster & Hucker, 2003). Clinicians should consider not only the type of violent behavior that might occur, but the likelihood that it will occur as well.

According to Elbogen and Tomkins (2000), risk assessment should include three distinct components: 1) actuarial instruments such as the Violence Risk Appraisal Guide (VRAG; Quinsey, Rice, Harris, & Cormier, 2006); 2) anamnestic risk assessment highlighting idiographic information; and 3) clinical input from the individual being assessed. The dynamic relationship between stable and static factors contributes to risk and recidivism, the reason why both actuarial and clinical judgment should be used in tandem (Douglas, 2009; Webster & Hucker, 2003). Others have argued that actuarial methods are always superior to clinical judgment (Grove & Meehl, 1996; Harris, Rice, & Quinsey, 2008; Meehl, 1957 research cited at <http://www.mhcop.on.ca>).

Jackson (1997) recommended examining six domains when assessing violence risk: 1) risk factors (historical, clinical, demographic); 2) strength of the evidence; 3) individual and group membership; 4) legal status; 5) type of behavior; and 6) time frame of the projected risk for the behavior to occur. Webster and Hucker (2003) suggest focusing on a few relevant factors and the interaction between those factors such as Axis I and Axis II diagnosis, substance abuse, and early maladjustment.

Conroy and Murrie (2007) have suggested a different model for assessing risk. Clinicians should begin by increasing the specificity of the referral questions. Next, clinicians should consider base rates, risk factors, and nomothetic and idiographic data in the assessment of violence. Finally, the communication of risk should be done in an ethical, "precise, honest, and easy-to-understand manner" (p. 26). Ultimately, release decisions and risk management recommendations must be made within the context of applicable laws by evaluators taking into consideration statutory criteria and ethical requirements of their profession (Webster & Hucker, 2003).

In Hawaii, the court relies on mental health professionals to perform forensic evaluations, including a "dangerousness assessment," before a post-acquittal offender can be conditionally released. In felony cases, the court calls on a panel of three examiners—two psychologists and

one psychiatrist—to independently evaluate the individual and submit a report documenting procedures, findings, and opinions. National studies have demonstrated wide variability in the quality of written forensic reports (Borum, 1994; Hecker & Steinberg, 2002; Heilbrun & Collins, 1995; Nicholson & Norwood, 2000; Skeem & Golding, 1998). Research into the quality of forensic reports in general has consistently found deficiencies in the quality of the reports regardless of type, jurisdiction, or professional writing the reports (Acklin et al., 2005; Christy, Douglas, Otto, & Petrila, 2004; Heilbrun & Collins, 1995; Petrella & Poythress, 1983; Robinson, 2007; Robinson & Acklin, 2010). In particular, the literature reveals problems related to inadequate documentation, failure to use collateral data, failure to connect clinical to psycholegal impairments, lack of rationale for opinion, and inappropriate or irrelevant psychological test usage.

Previous research on quality of competency to stand trial (CST) reports found that at least 98% of CST examinations included two sources of information; most commonly, records from the Adult Probation Division file, medical records, and arrest reports (Robinson, 2007). Psychological testing was used in a minority of examinations (37%) and only 7% used a forensic instrument (FAIs: Acklin et al., 2005). More recently, Robinson (2007) found that 18% of evaluations on CST used a forensic assessment instrument and 13% used psychological testing. Despite increased use of testing and FAIs, the examiners infrequently linked clinical status to legal impairments (Skeem, Golding, Berge, & Cohn, 1998). Evaluators infrequently explained the rationale for their decisions or recommendations (Skeem & Golding, 1998). Despite the critical importance of a rationale in the forensic opinion, 49% of criminal responsibility evaluations and 63% of CST in Hawaii failed to include a rationale for the psycholegal opinion (Acklin et al., 2005). Conditional release evaluations fared better with 80% having a rationale for the psycholegal opinion. In 87% of CR evaluations, there was a statement about the evaluatee's dangerousness and recommendations on how to manage the risk (Acklin et al., 2005). Research in Hawaii found significant agreement between CST opinions for a majority of forensic evaluators and judges with concordance rates of approximately 90% (Acklin et al., 2005; Robinson & Acklin, 2010) and unanimous agreement between all three evaluators and judicial determination to be 66% (Robinson & Acklin, 2010).

2. Method

2.1. Sampling procedures

Data for the current study was collected from archival records at the First Circuit Court of Hawaii in Honolulu, which covers the entire island of Oahu (approximately 900,000 residents). These records are accessible to the public, housed at the courthouse, and are available for examination by request. The current study strictly utilized public, archived data for which there was no researcher–participant interaction. The evaluations reviewed for this study were conducted on adults seeking post-acquittal conditional release from Hawaii State Hospital, the state's primary public forensic psychiatric facility.

Hawaii Revised Statutes sections 704–404, 704–411, and 704–415 require three concurrent and independent examinations to be completed in CR evaluations involving felony charges (Hawaii Revised Statutes, 2003). One examiner must come from a designated examiner within the State of Hawaii Department of Health (DOH); in this study, all DOH examiners were psychologists working for the Adult Mental Health Division (AMHD). By statute, the other two examiners are community-based, independent mental health practitioners who must be “certified” by the DOH. At least one of these independent examiners must be a psychiatrist, and the other may be either a psychiatrist or psychologist. In this study, all of the “third” independent examiners were psychologists. Upon receiving a court order requiring an examination for readiness for CR, each examiner is required to

complete an independent examination and submit a written report to the court.

A list of CR evaluations was provided by the Forensic Services Branch of the State of Hawaii's Adult Mental Health Division. Fifteen reports were selected for use in inter-rater training trials to establish coding reliability. Fifty cases were selected for review, comprising 150 reports (100 from community-based examiners and 50 from state-employed examiners). Inclusion criteria were three independent FHHE reports and a judicial determination, or a motion to withdraw the request for CR. Only cases from the island of Oahu were included in the study. Examiners included five community-based psychiatrists, 11 community-based doctoral level psychologists, including one board certified forensic psychologist, and 7 employees of the Courts and Corrections branch of the Department of Health, including one board certified forensic psychologist.

2.2. Measures

A survey instrument based on nationally-derived quality standards was used to examine the CR reports in this study. Data was coded using an objective coding system originally designed by Sanschagrín (2006) which was modified for application to the current study's sample. The research replicated a previous quality study of CST reports (Robinson & Acklin, 2010). The quality measure covered six report elements: 1) Data; 2) Legal/Ethical; 3) Historical; 4) Practical; 5) Clinical Assessment/Diagnostic; and 6) Opinion Rationale. We examined overall quality of reports with a specific interest in the usage of psychological testing and forensic assessment instruments, agreement between examiners, and agreement between examiner opinions and judicial determinations.

The survey instrument was comprised of 44 items that assessed report comprehensiveness and quality. The items were equally weighted and summed to produce a total quality score (QC). Each item was coded 0, 1, or 2. The survey had a range of 0 to 68 points if CR was not recommended and 0 to 74 points possible if CR was recommended. A score of 2 was the highest possible score for an item and was given if the evaluator included the information in the report and if the information was considered to be complete. A score of 1 was given if the evaluator included information, but the information was deemed incomplete. A score of 0 is given if the evaluator failed to include the information in the report. The total QC was calculated by dividing the total score for each evaluation by the maximum possible score for a report in that category and converting that number into a percentage score. The QC score ranged from 0 to 100. Report quality criterion was set at 80%.

2.3. Procedure

Collection and coding of data occurred in two parts. First, the principle investigator (AN) trained a graduate research assistant (KF) on the quality rating measure, including content domains and item ratings. After initial training, the principal investigator and research assistant independently coded 10 reports (Hecker & Steinberg, 2002; Sanschagrín, 2006). Ratings were analyzed using Cohen's κ (Cohen, 1988) to ascertain if a sufficient level of inter-rater agreement had been achieved. Both item and summary scores were analyzed to determine the level of agreement. Results of the reliability analysis demonstrated that initial reliability was “excellent” (Cicchetti, 1994) with $k = .75$, $p < .001$ for all ten reports. Perfect agreement ($k = 1.0$) was achieved on 16 items (numbers 1, 3, 5, 8, 9, 10, 13.a, 13.c, 13.d, 13.e, 18, 25, 26, 28, 37, 43). The principal investigator and research assistant reviewed discrepant ratings and discussed them to enhance accuracy and understanding of the coding criteria. Five additional reports were independently coded and the analysis was re-run. Agreement between raters reached “excellent” levels $k = .82$, $p < .001$ (Cicchetti, 1994). These

initial reports were used for training purposes only and the data was not included in the main data analysis.

Following training to criterion, the second phase of data collection and coding was conducted. The principal investigator and research assistant collected data for 50 case files (150 reports). To protect against rater drift (Haynes, 1978), every tenth report was jointly scored by the principal investigator and the research assistant. Cohen's *kappa* was computed to check agreement between raters. The resulting *kappas* ranged from .81 to .87 with a mean $k = .84$, $SD = .02$, $p < .01$, for the jointly coded reports, indicating "excellent" agreement (Cicchetti, 1994).

2.4. Hypotheses

Based on previous studies of community-based forensic reports (Nicholson & Norwood, 2000; Robinson & Acklin, 2010; Skeem & Golding, 1998), it was hypothesized that: 1) CR reports would fail to achieve 80% of the total possible QC score; 2) examiners employed by the Department of Health would produce higher quality reports than community-based evaluators; 3) community-based psychiatrists and community-based psychologists would not differ in report quality; 4) reports written by board-certified examiners would demonstrate higher quality than reports written by non-board certified evaluators; 5) psychologists would utilize empirically-validated assessment instruments more frequently than psychiatrists; 6) concordance of agreement for release decisions between pairs of examiners would reach levels of agreement considered "good" in clinical psychology ($kappa \geq .60$; Cicchetti, 1994); and 7) concordance of agreement between at least two examiners and the judicial CR determination would reach levels considered "good" in clinical psychology ($kappa \geq .60$; Cicchetti, 1994).

2.5. Data analysis

Inter-rater or interobserver agreement can be assessed in several ways including interobserver reliability coefficients (Patten, 2002), percentages and proportions, or by using Cohen's *kappa*. Cohen's *kappa* is the method that corrects for the agreement that occurs by chance (Warner, 2008). There are several approaches to the interpretation of *kappa*, including Grove, Andreasen, McDonald-Scott, Keller, and Shapiro (1981), Landis and Koch (1977) and Rietveld and van Hout's (1993). Previous research on report quality in Hawaii used Landis and Koch's interpretation of *kappa* (Acklin et al., 2005; Robinson, 2007). This study used Cicchetti's (1994) interpretation of *kappa* for training and rater drift trials since it could be used for interpreting intraclass correlation coefficients (ICC). ICC is a common measure of agreement in the behavioral sciences that can account for the variance of an observation or rating due to between-subject differences (McGraw & Wong, 1996; Shrout & Fleiss, 1979). ICC simplifies the assessment of reliability when more than two raters are present. ICC was used in addition to *kappa* to examine the degree of agreement between FMHE on release decision and between judges and FMHE.

According to Cicchetti (1994) higher *kappa* or ICC values indicate higher degrees of agreement and lower values indicate "poor" agreement. *Kappa* values or ICC of less than zero indicate agreement no better than chance. Values from 0.0 to 0.40 indicate "poor" agreement and values from .41 to .60 indicate "fair" agreement. Values of .61 to .74 indicate "good" levels of agreement. Values exceeding .75 indicate "excellent" levels of agreement.

Each report was evaluated by the quality instrument developed by Sanschagrin (2006). An item was coded 0 if the component was missing from the report; 1 if it was partially present; and 2 if the item was complete. After the report was evaluated, a total score was calculated, and a quality coefficient (QC) was calculated. Quality coefficients were calculated by dividing the total score for each evaluation by the maximum possible to arrive at a percentage score. This study

had a sample size of 50 cases (150 reports) with adequate statistical power ($1 - \beta$) = .80, $p < .05$.

3. Results

Quality of CR reports was mediocre across the board. QC scores ranged from 32.35 to 75.86, with a mean quality score of 53.22 ($SD = 8.35$, median = 52.94). None of the 150 reports achieved the 80% quality criterion. Data ($M = 82.87$, $SD = 8.35$) and Practical elements had the highest QC scores ($M = 91.56$, $SD = 14.54$). Informed consent was documented in only 52% of the reports. Historical elements were often entirely missing from reports. Report % for example, social history was absent from 93.3% of the reports and family history was absent in 66% of reports. Reports commonly included the defendant's psychiatric diagnosis (96%). Over half the reports did not explain how the psychiatric condition was a factor in the defendant's capacity for release (56.7%) or how the defendant's diagnosis might impact risk for violence or recidivism (52%). Examiners indicated in 97 (64.7%) reports that the defendant might pose a danger to himself, a statutorily-mandated item, but they did not provide information about what factors or circumstances might heighten or mitigate risk. In only 52 cases (34.7%) did the FMHE provide a complete rationale for the opinion on the defendant's dangerousness (Grisso, 2003). In 90 reports (60%), examiners provided a complete rationale for the CR recommendation (Grisso, 2003). In 80 reports (53.3%), examiners identified conditions under which the defendant could be safely released into the community. However, only 11 (7.3%) reports included a comprehensive justification for the conditions of release.

Overall report quality did not differ between community-based psychiatrists and psychologists. Reports prepared by community-based psychologists had quality scores that ranged from 35.14 to 75.68 ($M = 53.12$, $SD = 7.86$). Reports prepared by community-based psychiatrists had quality scores that ranged from 36.76 to 63.51 ($M = 53.05$, $SD = 6.65$). Reports submitted by community-based psychologists and psychiatrists did not differ in quality, $t(98) = .05$, $p = .96$.

Report quality was compared between community-based and DOH psychologists. None of the reports by either group scored above the 80% quality criterion. Reports prepared by community-based psychologists had quality scores that ranged from 35.14 to 75.68 ($M = 53.12$, $SD = 7.86$). Reports prepared by DOH psychologists had QC scores that ranged from 32.35 to 71.62 ($M = 53.50$, $SD = 10.29$). Report quality between the two groups did not differ, $t(91.67) = .21$, $p = .83$. Report quality was compared between board-certified psychologists (American Board of Professional Psychology) and non-board-certified psychologists. Report quality scores by board-certified psychologists ($n = 16$) had quality scores that ranged from 32.35 to 65.63 ($M = 46.84$, $SD = 11.24$). Reports written by non-board-certified psychologists ($n = 84$) had scores that ranged from 35.14 to 75.68 ($M = 54.54$, $SD = 8.16$). Reports by non-board certified psychologists obtained higher QC scores, $t(18.13) = 2.61$, $p < .05$.

When examining psychological test usage in the total sample ($N = 150$) some type of assessment instrument was used in 23.3% of the reports ($n = 35$). Cognitive testing was administered in 22 reports (14.7%) and personality testing was administered in 1 case (0.7%). Forensic Assessment Instruments (FAIs) were completed in only 13 reports (8.6%). There was no significant difference in test usage between psychologists ($N = 100$, $M = .28$, $SD = .53$) and psychiatrists ($N = 50$, $M = .20$, $SD = .40$), $t(148) = -.93$, $p = .35$. When types of testing were disaggregated, psychologists administered cognitive testing more frequently than psychiatrists. When the comparisons between psychiatrists and psychologists were broken down by psychologist type (community-based versus DOH), differences emerged. Psychiatrists ($M = .20$, $SD = .40$) used assessment instruments more often than community-based psychologists ($M = .04$, $SD = .20$), $t(71.24) = 2.51$, $p < .05$.

Agreement on the CR recommendation among evaluators was poor. In 40% of reports ($n = 20$ cases), all three-panel members reached CR recommendation consensus. In 46% of reports ($n = 23$ cases), two of the three examiners reached CR recommendation agreement. Thus, in 86% ($n = 43$) of the cases at least two of the three evaluators made similar CR recommendations. In 14% ($N = 7$) of the cases, there was complete disagreement among examiners. For all seven cases in this group one examiner recommended for CR, another recommended against CR, and the third evaluator did not report an opinion. DOH psychologists and community-based psychiatrists agreed in 60% of cases ($n = 30$ cases). Their level of agreement was “poor” ($k = .27$, $p < .01$). Community-based and DOH psychologists agreed in 46% of cases ($n = 23$). Level of agreement in their release recommendations was “poor” ($k = .10$, $p = .34$). Community-based psychiatrists and psychologists agreed in 60% of cases ($n = 30$). Their level of agreement was “poor” ($k = .21$, $p = .04$). Level of agreement among all three FMHE was very low, $ICC = .068$, $p = .19$, $F = 1.23$, $df = (49, 98)$, 95% CI: $-.08$ to $.25$, reflecting a “poor” level of agreement. In summary, agreement of CR recommendations among the evaluators was poor.

Concerning CR judicial determination, the court and all three examiners reached consensus in 39% of cases ($n = 19$). In an additional 29% of cases ($n = 14$), judicial determination was concordant with at least two of the examiners. Thus, for 67.3% of cases ($n = 33$), the court agreed with at least two of the evaluators on release decisions. This level of agreement is “poor” ($k = .26$, $p = .07$, $ICC = .30$, $p < .05$; 95% CI: $-.02$ to $.53$). In 14% of cases ($N = 7$), examiners had no consensus on release decisions (one recommended CR, one recommended against CR, one did not present a recommendation). Consequently, in these cases judges agreed with only one evaluator.

Conversely, in 16% of cases ($n = 8$), the court went against the majority opinion of at least two of the evaluators. Judges and community-based psychiatrists reached agreement on release decisions in 42 cases (85.7%), for a “good” level of agreement ($k = .66$, $p < .001$). Judges and community-based psychologists reached consensus in 28 cases (57%) for “poor” level of agreement ($k = .16$, $p = .11$). Judges and DOH agreed in 30 cases (61%), for “poor” level of agreement ($k = .30$, $p < .01$). Overall agreement for CR among all mental examiners and judicial determination was very “poor,” $ICC = .11$, $p < .05$; 95% CI: $.00$ to $.26$.

4. Discussion

This study found that quality of CR reports submitted to the Hawaii judiciary is poor. A great deal of variability was found in examiner methods, report formats, and findings. This appears to be reflected in the virtual chaos found in concordance statistics among examiners and judges. These levels of agreement are lower than levels found in Hawaii CST examinations. Robinson and Acklin (2010) found unanimous inter-examiner agreement rates of 70% on CST recommendations. Gowensmith et al. (in press) found unanimous rates of CST recommendations of 71% for initial CST examinations and 61% for follow-up CST examinations. Additionally, agreement rates between majority examiner opinions and judicial determinations on CST cases were reported at 90% (Robinson & Acklin, 2010) and 89% (Gowensmith et al., in press). Finally, the rates found in the current study are comparable to rates found by McNichols, Jul, and Gowensmith (in press) on a similar sample of CR evaluations (54% for inter-examiner unanimous agreement, and 87% for agreement between the majority examiner recommendation and judicial determination). It is not surprising that quality of CR reports, including level of agreements between evaluators and the court, differs from CST reports. In many respects, CR evaluations are a more complex task than CST evaluations. Competency to stand trial evaluations involve “contemporaneous” assessments, a solid research base, and well established forensic assessment instruments. Conditional release evaluations involve higher stakes, rely on the vagaries of risk

assessment, and methods of evaluation—whether clinical, actuarial, or structured professional judgment—are still in flux.

A number of recent papers address the issue of report improvement (e.g., Grisso, 2010; Wettstein, 2005). Generic suggestions for improving report quality across forensic referral questions include formal training in the conduct of forensic evaluations and report preparation. A significant problem in the current study was lack of uniformity in examiner methods. Report quality may be specifically improved by standardizing methods and report formats, as well as clearly describing rationales that integrate clinical findings, functional legal impairments, and forensic opinions (Grisso, 2003; Robinson & Acklin, 2010; Wettstein, 2005). The use of FAIs as a standard of practice is a nationally emerging trend, but the current research shows that this trend has not reached Hawaii (Gowensmith et al., in press; Robinson & Acklin, 2010). Since this appears to be the first study of CR report quality to appear in the literature, we are unable to assess the degree to which the use of FAIs, either actuarial or structured professional judgment, has penetrated practice nationally. Specific to CR evaluation reports, implementing formal risk assessment methodologies—whether purely actuarial or structured professional judgment including dynamic risk factors—is likely to be a powerful means to improve forensic report quality. A second problem with report quality was examiner failure to communicate meaningful linkage between clinical, legal, and forensic opinion. Pertinent to information conveyed in forensic reports, Wettstein (2005) writes, “Perhaps the most significant problem area identified by the empirical studies of forensic reports is the evaluator's failure to determine the reasons by which the opinions are reached. Evaluators have repeatedly failed to link psychopathology, psychological test results, and legal functional impairments. The omission is characteristic of both forensically and untrained evaluators” (pp. 170–171). Including a specific section of the report outlining forensic reasoning (Grisso, 2003) is likely to be an effective remedy. The use of checklists has also been shown to be helpful in improving report quality (Witt, 2010).

CST evaluations have a longer and more developed history than CR evaluations and, consequently, more FAIs have been developed and empirically validated in that arena (Acklin, 2010). The earliest FAI for CST, the Checklist for Criteria for CST, dates back to 1965 (Heilbrun, Rogers, & Otto, 2002). Several FAIs related to assessing CST were developed in the 1970s but were not widely available, used, or well grounded in psychometrics (Heilbrun et al., 2002). The following decades saw an emergence of tools developed for CST including the MacArthur Competence Adjudication Tool-Criminal Adjudication, Evaluation of Competency to Stand Trial-Revised, Competency Screening Test, Georgia Court Competency Test, and Fitness Interview Test-Revised, many of which are now in the second generation of development. These measures include a healthy degree of empirical validation and development based on a theoretical framework and Dusky legal standard relating to competency to stand trial (Zapf, Hubbard, Cooper, Wheelles, & Ronan, 2004).

Violence risk assessments, on the other hand, are a relatively newer type of forensic assessment, first emerging in the 1990s (Heilbrun et al., 2002). Forensic assessment instruments represent a significant conceptual and methodological advance in the field of dangerousness assessment. The introduction of actuarial risk measures and, later, structured professional judgment approaches, has advanced the behavioral science of violence risk assessment and management. According to Grisso (2003), “FAIs may improve our ability to conceptualize the relations between legal definitions of abilities and psychological constructs associated with human capacities.” Actuarial risk methods with structured professional judgment should be the standard used to inform the decision making process by examiners and, if possible, mandated by the courts for inclusion. The use of empirically validated assessment measures has higher reliability than unstructured clinical judgment and should improve inter-rater agreement (Otto & Heilbrun, 2002). Current

tools include the HCR-20, PCL-R, Static 99 and VRAG and the newly introduced Violence Risk Scale shows promise (Wong & Gordon, 2006; Yang, Wong, & Coid, 2010).

Despite evidence that formal risk assessment methods have better predictive validity than clinical judgment, utilization of risk assessment and methodologies was commonly omitted from the CR evaluations examined in this study. Despite a rapidly evolving literature and constant stream of new instruments and models, evidence-based methods of violence risk assessment (Gowensmith, 2009) was rarely conducted and reported by examiners. Examiners often failed to consider violence risk from a comprehensive viewpoint that included probability, imminence, frequency, severity, and nature of the violence. Additionally, examiners failed to consider base rates in their predictions of violence or if they did, it was not reflected in their rationales. There appears to be a disconnection in our sample between what professionals consider as “recommended” (PCL-R) or “acceptable” (PCL-SV and VRAG) methods to assess violence risk (Lally, 2003).

Conroy and Murrie (2007) have suggested a model for assessing risk that includes consideration of base rates, risk factors, and nomothetic and idiographic data in the assessment of violence. They suggest that before beginning the assessment process that referral questions be clarified so that questions have specificity. Indeed, ambiguity in statutory language may play a role in the poor levels of agreement observed here. Court orders for CR evaluations, which are based on Hawaii Revised Statutes (HRS chapter 704), typically require examiners to provide:

An opinion as to the risk of danger, if any, which the Defendant currently present to himself/herself or the person or property of others as a result of such condition, and if there is such a danger, whether the Defendant may still be released on conditions under which he/she can be controlled adequately and given property care, supervision and treatment or whether the Defendant may be discharged because he/she is no longer affected by a physical or mental disease, disorder or defect, or if so affected, he/she no longer presents a danger to himself or others and is not in need of care supervision or treatment. If your answer is that Defendant can be released on conditions, state the conditions.

Focusing on assessment, prediction, and management, this statutory language leaves a wide range of interpretation between examiners and the court. CR examiners are left to their own devices to interpret what is meant by the statutory criteria and each examiner may in fact operationalize this question very differently. In the presence of ambiguity, examiners may resort to heuristics which impair the accuracy and reliability of clinical decision-making (Garb, 1998). “Specification and clarity of statutory language [and] operationalization of legal definitions” are likely to increase agreement rates (Acklin & Robinson, 2007). Feedback to examiners has also been found to increase level of agreement (Haynes, 1978). Hawaii courts do not ordinarily provide timely feedback to examiners. Feedback allows examiners to engage in greater reflexivity and self-monitoring of their performance and should increase the accuracy and consistency of the forensic product. Nevertheless, the absence of controlled “gold standard” research on the accuracy of judgments is formidable (Mossman et al., 2010).

Education and training of examiners should also improve report quality. Robinson and Acklin (2010) showed that CST report quality significantly improved after a single training. Training should focus on both the methodology of the assessment process and the written product submitted to the court. Methodological training should include use of actuarial risk assessments and structured professional judgment (Douglas, 2009) and how to integrate the results from the testing with the unstructured clinical judgment of the examiners while utilizing base rates. Training should also include fundamentals

of report writing and education about how to report findings as well as convey the examiner's thought processes and psycholegal rationale underlying their opinions (Grisso, 2003; Wettstein, 2005). These findings on the quality of Hawaii CR reports contribute to two emerging literatures: quality of forensic reports (Acklin et al., 2005; Borum, 1994; Christy et al., 2004; Hecker & Steinberg, 2002; Heilbrun & Collins, 1995; Nicholson & Norwood, 2000; Robinson & Acklin, 2010; Skeem & Golding, 1998; Wettstein, 2005), and inter-rater agreement in forensic mental health assessment (Acklin et al., 2005; Goldstein & Stone, 1977; Poythress & Stock, 1980; Robinson & Acklin, 2010; Roesch & Golding, 1980; Skeem & Golding, 1998; Skeem et al., 1998), and moves the field toward emerging practice guidelines and standards of care in forensic mental health evaluation (Heilbrun, DeMatteo, & Marczyk, 2008; Wettstein, 2005). It may be the first study examining quality of CR reports to appear in the literature.

Hawaii's three panel system serves as a natural laboratory for examining the effects on inter-rater agreement (Acklin, 2010). In comparison to the quality of CR reports nationally, our results may be not atypical. For a variety of reasons, including inadequate or inconsistent training, non-standardization of procedures and methods, non-utilization of FAls, weak understanding of violence risk assessment methodologies, including instruments and the availability and application of base rates, and deeper challenges and limitations of violence risk prediction and assessment in general (Heilbrun, 1991), CR reports may be particularly vulnerable to poor or inconsistent quality. Recent commentators have highlighted the challenges of risk prediction, including inherent limitations in the predictive efficacy of violence risk instruments “the ceiling of predictive efficacy may have been reached with the available technology” (Yang et al., 2010, p. 759), and questionable use of group statistics in making individual predictions “we conclude that predictions of future offending cannot be achieved in the individual case with any degree of confidence” (Cooke & Michie, 2010, p. 259). Hanson and Howard (2010) make a compelling argument that “the confidence intervals for the recidivism outcome predicted for a single case will range from zero to one (i.e., be uninformative) when the outcome is dichotomous and the predicted probability is between .05 and .95” (p. 275), though this assertion has been vigorously disputed (Harris et al., 2008). Testing the degree to which these findings are generalizable to other jurisdictions will be a task for future research. Our findings appear to provide clear guidelines for identifying report weakness and improving report quality.

References

- Acklin, M.W. (2010). Quantifying consensus: Methodological issues in competency to stand trial evaluations (manuscript in submission).
- Acklin, M., Kennedy, R., Robinson, R., Dunkin, B., Dwire, J., & Lees, B. (2005). *Quality of forensic reports: An empirical investigation of three panel reports*. Honolulu, HI: Paper presented at the Annual State of Hawaii Forensic Mental Health Examiner Training Conference.
- Acklin, M. & Robinson, R. (2007). *Quantifying Consensus: Methodological Issues in Competency to Stand Trial Evaluations*. (Unpublished manuscript).
- Bertman-Pate, L. J., Burnett, D. M. R., Thompson, J. W., Bertman-Pate, L. J., Burnett, D. M. R., Thompson, J. W., Calhoun, C. J., Jr., Deland, S., Fryou, R. M., Bertman-Pate, L. J., Burnett, D. M. R., Thompson, J. W., Calhoun, C. J., Jr., Deland, S., & Fryou, R. M. (2004). The New Orleans forensic aftercare clinic: A seven year review of hospital discharged and jail diverted clients. *Behavioral Sciences & the Law*, 22(1); 1, 159–169.
- Bloom, J. D., Williams, M. H., & Bigelow, D. A. (1991). Monitored conditional release of persons found not guilty by reason of insanity. *The American Journal of Psychiatry*, 148(4), 444.
- Borum, R. (1994). Standards and practices in forensic evaluation. *American Psychology Law Society News*, 14(2), 2–4.
- Cavanaugh, J., & Wasyliv, O. (1985). Adjustment of the not guilty by reason of insanity (NGRI) outpatient: An initial report. *Journal of Forensic Sciences*, 30, 24–30.
- Christy, A., Douglas, K. S., Otto, R. K., & Petrila, J. (2004). Juveniles evaluated incompetent to proceed: Characteristics and quality of mental health professionals' evaluations. *Professional Psychology: Research and Practice*, 35(4), 380–388.
- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.

- Cohen, J. W. (1988). *Statistical power and analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conroy, M. A., & Murrie, D. C. (2007). *Assessment of violence risk: A guide for risk assessment and risk management*. New York: John Wiley & Sons.
- Cooke, D., & Michie, C. (2010). Limitations of diagnostic precision and predictive utility in the individual case: A challenge for forensic practice. *Law and Human Behavior, 34*, 259–274.
- Douglas, K. (2009). *Assessing risk of violence using Structured Professional Judgment*. Winter: AP-LS NEWS p. 12.
- Elbogen, E. B., & Tomkins, A. (2000). From the psychiatric hospital to the community: Integrating conditional release and contingency management. *Behavioral Sciences and the Law, 18*, 427–444.
- Evans, D. (1999). One hundred years of conditional release in Canada. *Corrections Today, 61*(5), 132.
- Foucha v. Louisiana, 504 U.S. 71 (1992).
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Goldstein, A. M. (2003). Overview of forensic psychology. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology*. (pp. 11. (pp. 3–20) Hoboken, NJ, US: John Wiley & Sons Inc.
- Goldstein, R. L., & Stone, M. (1977). When doctors disagree: Differing views on competency. *Bulletin of the American Academy of Psychiatry and the Law, 5*, 90–97.
- Gowensmith, W. N. (2008). *The conditional release program in Hawaii: Successes and challenges*. Paper presented at the National GAINS Center Conference. D.C.: Washington.
- Gowensmith, W. N. (2009). *Evidence-based assessment of dangerousness*. Paper presented at Best Practices in Forensic Mental Health: Responsibility and Recovery in the Legal System Conference. HI: Honolulu.
- Gowensmith, W. N., Murrie, D. M., Boccaccini, M. T. (in press). Field reliability of competency to stand trial evaluations: How often do evaluators agree, and what do judges decide when evaluators disagree? *Law and Human Behavior*.
- Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology, 2* Available at: http://works.bepress.com/thomas_grisso/4
- Grisso, T. (2003). *Evaluating competencies: Forensic assessment instruments* (2nd ed.). New York: Plenum.
- Grounds, A., Snowden, P., & Taylor, P. J. (1993). Forensic psychiatry in the National Health Service in England and Wales, with commentary on Scotland and other countries. In J. Gunn, & P. J. Taylor (Eds.), *Forensic Psychiatry: Clinical, legal, and ethical issues* (pp. 691–731). Oxford Butterworth-Heinemann.
- Grove, W. M., Andreasen, N., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability studies of psychiatric diagnosis: Theory and practice. *Archives of General Psychiatry, 38*, 408–413.
- Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *American Psychological Association, 2*(2), 293–323.
- Hafemeister, T. L. (1998). Legal aspects of the treatment of offenders with mental disorders. In R. Wettstein (Ed.), *treatment of offenders with mental disorders* (pp. 44–125). New York, NY, US: The Guilford Press.
- Hanson, R., & Howard, P. (2010). Individual confidence intervals do not inform decision-makers about the accuracy of risk assessment evaluations. *Law and Human Behavior, 34*, 275–281.
- Harris, G., Rice, M., & Quinsey, V. (2008). Shall evidence-based risk assessment be abandoned? *British Journal of Psychiatry, 192*, 154.
- Haynes, S. N. (1978). *Principles of behavioral assessment*. Oxford, England: Gardner.
- Hecker, T., & Steinberg, L. (2002). Psychological evaluation at juvenile court disposition. *Professional Psychology: Research and Practice, 33*(3), 300–306.
- Heilbrun, K. (1991). Prediction versus management models relevant to risk assessment: The importance of legal decision-making context. *Law and Human Behavior, 21*(4), 347–359.
- Heilbrun, K. (2010). Developments in risk assessment. Violence risk and sexual violence risk. *Presentation to the Courts & Corrections Branch, Adult Mental Health Division, Annual Forensic Training Conference, September '14, 2010*.
- Heilbrun, K., & Collins, S. (1995). Evaluations of trial competency and mental state at time of offense: Report characteristics. *Professional Psychology: Research and Practice, 26*(1), 61–67.
- Heilbrun, K., DeMatteo, D., & Marczyk, G. (2008). Standards of practice and care in forensic mental health assessment: Legal, professional, and principles-based considerations. *Psychology: Public Policy and Law, 14*, 1–26.
- Heilbrun, K., & Griffin, P. (1998). Community-based forensic treatment. In R. Wettstein (Ed.), *Treatment of offenders with mental disorders* (pp. 1–43). New York, NY, US: The Guilford Press.
- Heilbrun, K., Rogers, R., & Otto, R. (2002). Forensic assessment: current status and future directions. In J. R. P. Ogloff (Ed.), *Taking psychology and law into the twenty-first century* (pp. 119–146). New York, NY: Kluwer Academic/Plenum Publishers.
- Jackson, J. (1997). A conceptual model for the study of violence. In C. D. Webster, & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment and treatment* (pp. 223–247). New York: Guilford.
- Kozol, H., Boucher, R., & Garofolo, R. (1972). The diagnosis and treatment of dangerousness. *Crime and Delinquency, 18*, 371–392.
- Lally, S. (2003). What tests are acceptable for use in forensic evaluations: A survey of experts. *Professional Psychology: Research and Practice, 34*(5), 491–498.
- Lamb, H. R., Weinberger, L. E., & Gross, B. H. (1988). Court-mandated community outpatient treatment for persons found Not Guilty by Reason of Insanity: A five-year follow-up. *American Journal of Psychiatry, 145*, 450–456.
- Landis, J., & Koch, G. G. (1977). The measurement of observed agreement for categorical data. *Biometrics, 33*, 159–174.
- Lessard v. Schmidt, 414 U.S. 473 (1974).
- Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. *Journal of American Medical Association, 269*, 1007–1011.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.
- McGreevy, M. A., Steadman, H. J., Dvoskin, J. A., & Dollard, N. (1991). New York State's system of managing insanity acquitees in the community. *Hospital and Community Psychiatry, 42*, 512–517.
- McNichols, B. J., Jul, E., & Gowensmith, W. N. (2010). *Forensic evaluators, conditional release evaluations, and community success: Is evaluator agreement of CR readiness related to longer community tenure? Paper accepted for presentation at 2011 American Psychology – Law Society Annual Conference, Miami*.
- Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4*(4), 268–273.
- Monahan, J. (1981). *Predicting violent behavior: An assessment of clinical techniques*. Beverly Hills, CA: Sage Publications.
- Nicholson, R., & Norwood, S. (2000). *The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gaps between promise and practice*. : Law and Human.
- Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting & Clinical Psychology, 62*, 783–792.
- Mossman, D., Bowen, M., Vanness, D., Bienenfeld, D., Correll, T., Kay, J., Klyko, W., & Lehrer, D. (2010). Quantifying the accuracy of forensic examiners in the absence of a "gold standard". *Law and Human Behavior, 34*, 402–417 Behavior, 24, 9–44).
- O'Connor v. Donaldson, 422 U.S. 563 (1975).
- Otto, R. K., & Heilbrun, K. (2002). The practice of forensic psychology: A look to the future in light of the past. *American Psychologist, 57*(1), 5.
- Patten, M. L. (2002). *Understanding research methods*. Los Angeles: Pyrczak Publishing.
- Petrella, R. C., & Poythress, N. G. (1983). The quality of forensic evaluations: An interdisciplinary study. *Journal of Consulting and Clinical Psychology, 51*(1), 76–85.
- Pettila, John (2007). Fitness to proceed. *Presentation at AMHD 5th Annual Forensic Mental Health Examiner Training, Kaneohe, Hawaii*.
- Poythress, N. G., Bonnie, R. J., Monahan, J., Otto, R., & Hoge, S. K. (2002). *Adjudicative competence: The MacArthur studies*. New York, NY: Kluwer Academic/Plenum Publishers.
- Quinsey, V., Rice, M., Harris, G., & Cormier, C. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.
- Rietveld, T., & van Hout, R. (1993). *Statistical techniques for the study of language and language behavior*. New York: Mouton de Gruyter.
- Robinson, R. (2007). *Quality of forensic psychological reports submitted to the first circuit court of the state of Hawaii*. Unpublished doctoral dissertation, Argosy University, Hawaii.
- Robinson, R., & Acklin, M. W. (2010). Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary. *International Journal of Law & Psychiatry, 33*, 131–137.
- Roesch, R., & Golding, S. L. (1980). *Competency to stand trial*. Urbana/Champaign: University of Illinois Press.
- Rogers, R., & Wettstein, R. (1984). Relapse of NGRi outpatients: An empirical study. *International Journal of Offender Therapy and Comparative Criminology, 28*, 227–235.
- Rogers, J., Bloom, J. D., & Manson, S. (1986). Oregon's new insanity defense system: A review of the first five years, 1978 to 1982. *Bulletin of the American Academy of Psychiatry and the Law, 12*, 383–402.
- Sanschagrin, K. A. (2006). The quality of forensic mental health assessments of juvenile offenders: An empirical investigation. (ProQuest Information & Learning). *Dissertation Abstracts International: Section B: The Sciences and Engineering, 66* (11), 6292.
- Segal, S. P., & Burgess, P. M. (2006). Conditional release: A less restrictive alternative to hospitalization. *Psychiatric Services, 57*(11), 1600–1606.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420–428.
- Silver, S. B. (1983). *Treatment and aftercare of insanity acquitees in Maryland*. Testimony before the Subcommittee on Criminal Law of the Committee on the Judiciary, United States Senate. Washington DC: U.S. Government Printing Office Serial No. J-97-122.
- Skeem, J. L., & Golding, S. L. (1998). Community examiner's evaluations of competence to stand trial: Common problems and suggestions for improvement. *Professional Psychology: Research and Practice, 29*, 357–367.
- Skeem, J. L., Golding, S. L., Berge, G., & Cohn, N. B. (1998). Logic and reliability of evaluations of competence to stand trial. *Law and Human Behavior, 22*(5), 519–547.
- Stafford, K. P., & Karpawich, J. J. (1997). Conditional release: Court-ordered outpatient treatment for insanity acquitees. *New Directions for Mental Health Services, 75*, 61–72.
- Thornberry, T., & Jacoby, J. (1979). *The criminally insane: A community follow-up of mentally ill offenders*. Chicago: University of Chicago Press.
- Warner, R. M. (2008). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks: Sage Publications.
- Webster, C. D., & Hucker, S. J. (2003). *Release decision making*. Ontario: St. Joseph's Healthcare.
- Wettstein, R. M. (2005). Quality and quality improvement in forensic mental health evaluations. *The Journal of the American Academy of Psychiatry and the Law, 33*, 158–175.
- Witt, P. H. (2010). Forensic Report Checklist. <http://www.forensicpsychologyunbound.ws/2>: 233–240.
- Wiederanders, M. R. (1992). Recidivism of disordered offenders who were conditionally vs. unconditionally released. *Behavioral Sciences and the Law, 10*, 141–148.

- Wiederanders, M. R., Bromley, D. L., & Choate, P. A. (1997). Forensic conditional release programs and outcomes in three states. *International Journal of Law and Psychiatry*, 20(2), 249–257.
- Wong, S., & Gordon, A. (2006). The validity and reliability of the Violence Risk Scale: A treatment-friendly violence risk assessment tool. *Psychology, Public Policy and Law*, 3, 279–309.
- Yang, M., Wong, S., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136(5), 740–767.
- Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheelles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice*, 4, 27–44.