

Quality and Quality Improvement in Forensic Mental Health Evaluations

Robert M. Wettstein, MD

Despite the growing attention to quality and quality improvement in health care in the United States, forensic psychiatry has yet to incorporate relevant developments and information and make quality an important item on the agenda. This article reviews the empirical research regarding the perceived quality of forensic evaluations, which has primarily examined criminal rather than civil forensic evaluations. Beyond the available research, many important policy and empirical questions must be addressed, including the definition of a quality forensic evaluation, the process used to assess quality, the indicators and measures used, the methods that provide incentives for performing quality evaluations, the role of forensic psychiatry training programs, and the role of the American Academy of Psychiatry and the Law (AAPL) or other professional organizations in the quality improvement enterprise.

J Am Acad Psychiatry Law 33:158–75, 2005

In 1996, Dietz¹ encouraged forensic psychiatrists to aspire to excellence in their chosen field. He emphasized the need to search for the truth, to be honest, to disclose biases and credentials to potential clients, and to accept cases with caution. He asked us to disclose completely “all information, all reasoning, and all opinions” (Ref. 1, p 162). He recommended that we volunteer the weaknesses of the case, despite the client’s resistance. He cautioned us not to serve as advocates for a political or social cause, much as a lobbyist might do, or as a frustrated lawyer advocating for a party to the litigation. Rather, he advised that we adopt the role of the forensic scientist, and present our findings and opinions with “scrupulous fairness” (Ref. 1, p 161). He noted that the “ultimate arbiters of excellence in forensic practice” are our peers rather than our clients, who are the attorneys and courts (Ref. 1, p 154).

Though useful, these ideas are preliminary to a more far-reaching goal of addressing quality and quality improvement in forensic evaluations.² In this

article, some important concepts and developments regarding quality improvement in clinical medicine and mental health care are reviewed. Knowledge of these developments should further efforts to improve forensic evaluations. Later, a review of empirical studies of the quality of forensic evaluations is provided. Finally, some of the quality and quality improvement in forensic evaluations are discussed.

Quality and Quality Improvement in Clinical Medicine

Two significant themes in the culture of medical practice in the United States are now the focus of substantial attention: patient safety and quality of medical care.^{3,4} Patient safety is just one component of quality. Quality care is certainly safe, but it is also timely, patient centered, efficient, effective, and equitable (i.e., quality does not vary because of gender, ethnicity, location, and socioeconomic status).⁵ We might, therefore, refer to the quality of health care in the plural rather than the singular. There are health care “qualities” rather than one quality, and each should be delineated and assessed.^{6,7} The different stakeholders in the health care system may each have diverging goals for a quality system, with patients focused on effectiveness and safety, but insurers focused on efficiency. The concern about both health

Dr. Wettstein is Clinical Professor of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA. This article is a modified version of the author’s President’s Address at the Annual Meeting of the American Academy of Psychiatry and the Law, Scottsdale, AZ, October 21, 2004. Address correspondence to: Robert M. Wettstein, MD, University of Pittsburgh School of Medicine, 401 Shady Avenue, Suite B103, Pittsburgh, PA 15206. E-mail: wettsteins@aol.com

care safety and quality reflects increased scrutiny and demand for accountability by the greater, non-health care community.⁸

Superimposed on fiscal concerns is the more recent concern about patient safety, medical error, and the cost of professional liability insurance.³ The Institute of Medicine's 1999 landmark report on medical error, *To Err is Human: Building a Safer Health System*,⁹ contended that there are as many as 100,000 fatalities annually in hospitals in the United States due to medical errors or adverse events. This report has prompted substantial interest in examining systems of medical care rather than individual physicians as the ultimate cause of medical error. It is revolutionary for medicine to examine health care systems as the source of adverse events, rather than to blame individual practitioners.^{10,11}

The Institute of Medicine has drawn further attention to health care quality through its publication, *Crossing the Quality Chasm*,⁵ which recommended a framework and strategies to improve health care quality. Toward this end, the Institute of Medicine has adopted the following definition of health care quality: Quality is the "degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" (Ref. 5, p 232).

Quality can be evaluated based on the structure of health care (i.e., physical facilities, staffing, organization, financing, provider type and duties, utilization controls and incentives), its process (i.e., the delivery of care, such as actual interventions of assessment and treatment), or its outcome (i.e., the results of care, such as morbidity, mortality, and patient function and satisfaction).⁶ Clinicians may be more comfortable measuring the process of care because of the multidetermined nature of clinical outcomes, and purchasers are more interested in measuring patient satisfaction and utilization and the outcome of health care.⁷ Cost control efforts focus on process and outcome, rather than the structure of care.⁶ Generally, health care quality is a property of the health care system. But, to improve the daily practice of medicine, we should monitor its performance and assess change in health care practice, once quality improvement efforts have been implemented.⁷

National efforts to assess and improve health care quality have been ongoing for several years through numerous organizations and agencies.¹² The Centers

for Medicare and Medicare Services (CMS) has published hospital morbidity and mortality data. Another federal agency, the Agency for Health Care Research and Quality (AHRQ), has developed a computerized quality-measurement system, with a database of specific measures, and sets of clinical performance measures. The National Committee on Quality Assurance (NCQA), composed of managed health care organizations, has produced Health Plan Employer Data and Information Set (HEDIS), which provides standardized evaluation criteria or performance measures to help employers select health plans; these criteria are predominately medical rather than psychiatric. The Joint Commission on the Accreditation of Health Care Organizations (JCAHO) has developed its own system for accreditation and performance measures. Some states publish data on various providers' outcomes for selected diseases or procedures. With the diverging priorities of these organizations, it has not been possible to obtain a consensus on performance measures.⁴

Measures of health care quality are essential for health care purchasing, monitoring, accreditation, and advancement of performance.⁷ Outcome measures of care in medicine are increasingly evaluated and include appropriate use of medication for chronic disease (diabetes, heart disease, post-myocardial infarction, asthma, and hypertension), complication rates after surgery, prolonged hospital lengths of stay, and rehospitalization rates.^{4,12} Quality of medical care indicators for screening include screening for breast cancer, cervical cancer, colorectal cancer, childhood immunization, and cholesterol levels.⁴ Ratings of physician performance include patient satisfaction, clinical outcome, and compliance with practice parameters and professional guidelines.^{13,14} Patient satisfaction ratings include timely access to care, provider-patient communication, rating of medical care, office organization, and health care quality of life. Patient and family satisfaction ratings are often used but have been criticized as failing to measure quality of health care.

Empirical study of the quality of medical care actually delivered in the United States reveals disappointing results. McGlynn and colleagues¹⁵ determined that only one-half of the recommended preventive, acute, or chronic care was actually provided to patients. More surprising was that quality of delivered care varied substantially by medical condition. These data bring a sense of urgency to the fail-

ure of our health care system as a whole. Given the failure of traditional medical care, integrated disease management programs that educate consumers and support practitioners are being developed and implemented for some chronic diseases.^{12,16}

There are marked disparities in health care access, use, quality, and outcome by race, ethnicity, income, education, place of residence, and sex.^{17,18} Such disparities persist, even when availability of health insurance is eliminated as a confounding variable. Clinician bias is one explanatory factor in these disparities. Efforts to improve the overall quality of medical care have been demonstrated to reduce, but not eliminate, these disparities.¹⁹

Some insurers and managed care organizations offer a performance or quality financial bonus for excellent health care.^{20,21} Employers, distressed over large increases in health insurance premiums, could require quality assessment and improvement efforts and reward providers of high quality, efficient health care who obtain better patient outcomes. Family practitioners in the United Kingdom can contract with the National Health Service to link physicians' pay to the quality of health care provided.²² Such a quality bonus initiative requires that the medical practice regularly collect data and hire additional staff. Investment in health care technology, such as electronic medical records and prescriptions, is thus encouraged to improve quality and efficiency and reduce medical error.

Current efforts to improve patient safety and health care quality use several methods. Some insurers and managed care organizations have published health care quality results, so-called report cards, for hospitals and outpatient practices.²³⁻²⁵ Morbidity and mortality conferences parade poor outcomes before the hospital medical staff or department faculty. Incident reports to the hospital, state department of health, or other state agency, are increasingly required. Hospitals and other facilities are accredited, based on a variety of quality-of-care indicators and measures. Confidential peer review is undertaken by hospital staff to an unknown extent.²⁶ Professional liability suits are brought, in part, to deter future adverse outcomes and improve care. Most of these health care quality improvement mechanisms have not enjoyed significant implementation with regard to improving forensic mental health evaluations.

There are several barriers to repeated measuring of health care quality, including motivational, finan-

cial, organizational, and technical.^{4,5,7,15,22,25} Measurement systems should be developed, implemented, and financed, with many methodological challenges. Specific goals are needed to implement practice changes. Attitudinal and motivational problems apply to the individual clinicians and the hospital or health care system seeking quality improvement. Resources devoted to quality improvement could compromise attention to other health care needs, such as providing for the care of the uninsured and underinsured.

Many of these matters regarding the quality of health care can apply to the quality of forensic evaluations and will be elaborated herein. Regard for the numerous dimensions of health care quality, absence of consensus on performance measures, challenges in evaluating health care quality, importance of the system of care in understanding quality, attention to process and outcome of care, presence of several stakeholders in the endeavor, usefulness of technology in the improvement of quality, disappointing actual quality of health care, impact of financial considerations on health care quality improvement, and future use of quality report cards are each potential problems for the quality of forensic evaluations, as well.

Quality and Quality Improvement in Psychiatry

We know relatively little about the quality of mental health treatment programs, how well they work, and how to measure their effectiveness.²⁷ Quality improvement research activity has focused on the detection and treatment of serious depression, schizophrenia, and suicide.

Quality indicators specific to psychiatry include time until treatment, detection of mental disorder, and adequacy of treatment. Specific quality measures include rehospitalization rates, time interval until an outpatient appointment after hospitalization, adequacy and duration of acute and maintenance antidepressant medication treatment, frequency and duration of use of hospital seclusion or restraint, and use of appropriate laboratory testing. Outcome measures include symptomatic improvement or remission, functional status, return to employment, or complications such as criminal arrest or suicidal behavior.^{28,29}

Some researchers have investigated the adequacy of treatment of mental disorders by primary care

physicians and psychiatrists.^{30,31} Results generally indicate that primary care physicians fail to detect depression and anxiety or to treat it properly.³² In one study, only 19 percent of those with a probable depressive or anxiety disorder who were seen by a primary care physician received appropriate care (i.e., two months of psychotropic medication in the past year, or at least four counseling visits).³³ Appropriate care was less likely to be provided to those patients who were black, less educated, male, younger than 30 or older than 59 years. When seen by primary care physicians, depressed patients are undermedicated, not treated at all, or not referred for counseling. Antidepressant medication prescriptions are written without follow-up, leading to the patient's noncompliance and discontinuation of medication and symptom chronicity.³⁰⁻³³

Quality improvement research has demonstrated improved outcomes for depression and schizophrenia, using models for the management of chronic illness. These models include training clinicians in appropriate care, repeated and systematic follow-up of patients, monitoring, outreach programs, telephone support of patients, and collaboration between the treating mental health professional and the primary care physician.^{31,34-36}

Even in private practice, clinicians are increasingly using outcome measures. Hatfield and Ogles³⁷ surveyed a national sample of psychologists, almost all of whom had doctoral-level education and had been licensed for a mean of 18 years. Of the sample ($n = 874$), 37 percent reported that they used some form of outcome assessment, including standardized and individualized measures. They reported both internal and external motivations for using outcome measures.

Consumer ratings of behavioral health services in managed health plans are now commonplace and are required by the National Committee for Quality Assurance (NCQA) accreditation process.^{38,39} Surveys ask consumers about their experience in locating a clinician, seeking approval from the health plan, selecting a clinician with whom they were satisfied, and the time taken to obtain treatment.

As noted earlier in the discussion of quality of general medical care, the problems and challenges with respect to quality improvement of mental health care are relevant to quality improvement efforts regarding forensic evaluations. The absence of consensus quality or performance measures for men-

tal health treatment, the need to examine systems of mental health care, the attention to process and outcome dimensions of mental health care, the frequently low quality of measured actual mental health care, and the usefulness of consumer ratings of quality are relevant to the quality improvement process regarding forensic evaluations.

Empirical Data Regarding Quality of Forensic Evaluations

Relatively little conceptual analysis has been undertaken with regard to the quality of forensic evaluations. The published literature, with few exceptions, has not benefited from current knowledge and research findings obtained from the work reviewed in the prior section regarding the general quality of health care. The published literature on the quality of forensic evaluations has largely been empirical and has concentrated on the adequacy of evaluation reports.

A variety of empirical studies have been conducted relevant to standards and quality of forensic evaluations.⁴⁰ Topics investigated include contents of actual forensic reports, desired contents of forensic reports, perceived deficiencies of reports and evaluations, and prevalence of the use of diagnostic tests. Some studies report only descriptive data regarding expert report contents, while others include quality ratings. The studies have included the perspectives of attorneys, judges, and forensic mental health experts. Relatively few studies have been conducted regarding the quality of forensic evaluations, except in the case of child custody evaluations. Many of the published studies consist of small sample sizes in a restricted geographical area. A representative sample of these will now be reviewed, in chronological order, dealing with general forensic practice, psychological test usage, child custody, and surveys of judges and attorneys.

General Forensic Practice

Petrella and Poythress⁴¹ examined reports of competency-to-stand-trial and criminal responsibility evaluations conducted at Michigan's Center for Forensic Psychiatry using several quality measures. Their principal concern was the respective quality of evaluations conducted by psychiatrists, psychologists, and social workers at the Center, all of whom had at least some forensic evaluation training and used a standardized evaluation format. They used

two measures of an evaluation's thoroughness: first, the use of data sources other than the interview; and, second, the quantity and comprehensiveness of the evaluator's notes recording the interview. In addition, they had three outside raters (a practicing attorney, a trial judge, and a law professor) rate the quality of a sample of forensic reports ($n = 30$), using a nine-point Likert scale, on a set of quality measures (e.g., for criminal responsibility reports: use of proper legal criteria, a clearly stated opinion, adequate basis for opinion, use of psychiatric jargon, clinical characteristics of the defendant, and overall report quality). The obtained results were that reports by psychologists were more thorough than those by psychiatrists, based on the use of more collateral data sources. In trial competency reports, the frequency of contacting attorneys was 43 percent by psychiatrists, 45 percent by psychologists, and 53 percent by social workers. Review of the defendant's previous medical or psychiatric records was even less frequent in trial competency evaluation reports, with 10 percent by psychiatrists, 25 percent by psychologists, and 35 percent by social workers. Psychologists' reports were often considered to be of higher quality when rated blindly by the outside raters.

Heilbrun and Collins⁴² rated trial competency and criminal responsibility reports of evaluations conducted in the hospital and in the community in Florida. Raters examined preadjudication (i.e., those recommended as incompetent to stand trial) and postadjudication trial competency reports ($n = 277$) prepared by psychologists and psychiatrists. Only 19 percent of the latter were certified by the American Board of Psychiatry and Neurology. Some evaluations included both trial competency and criminal responsibility. The reports were relatively brief, with a mean length of 3.9 pages for both hospital and community evaluations. Only 30 percent of community-based evaluations included a notice of nonconfidentiality to the evaluatee. Psychological testing (i.e., the MMPI, and the Revised Wechsler Adult Intelligence Scale; WAIS-R) was used in 13 percent of the hospital-based evaluations and 41 percent of the community evaluations. Hospital-based evaluators reviewed previous mental health evaluations in 81 percent of cases, but community-based evaluators did so in only 30 percent of cases. Arrest reports were reviewed in 95 percent of hospital evaluations and 48 percent of community evaluations. Arrest reports were reviewed in 42 percent of competency evalua-

tions, but in 67 percent of criminal responsibility evaluations. Jail staff were interviewed in 1 percent of hospital evaluations, but 17 percent of community evaluations. Evaluators did not contact attorneys when conducting trial competency evaluations. The reports addressed the relevant state legal competency criteria in 95 percent of hospital reports and 61 percent of community reports. An ultimate issue opinion was offered in 95 percent of hospital reports, 99 percent of community reports, and 47 percent of criminal responsibility reports. No quality ratings were conducted.

Borum and Grisso⁴³ surveyed a national sample of forensic psychologists ($n = 53$) and psychiatrists ($n = 43$) concerning their beliefs about the appropriate content for reports (i.e., not evaluations) on competency to stand trial and criminal responsibility. This novel approach did not examine evaluators' reported or actual behavior as experts. Of the sample, 81 to 83 percent had forensic board certification (American Board of Forensic Psychiatry, American Board of Forensic Psychology), with an average of 17 years of experience in conducting forensic evaluations, and 52 to 54 percent of their practice consisted of forensic evaluations. Each had conducted an average of 48 trial competency evaluations and 35 criminal responsibility evaluations in the prior year. On the written survey, each subject rated the importance of potential content items for competency and responsibility reports on an importance scale of essential, recommended, optional, or contraindicated. The investigators defined a "consensus" on these ratings when a content item was endorsed by 70 percent or more of the respondents. Content and quality ratings of the experts' actual reports were not evaluated, but the investigators considered only opinions about the appropriate content of criminal forensic reports. Given the large number of content items in both types of reports, there was considerable variability regarding what experts thought was essential or recommended. For instance, there was much disagreement about whether the defendant's description of the alleged offense should be included in trial competency reports. One-third of the psychiatrists thought that it was contraindicated in reports, but one-half of the psychiatrists believed that it was important to include. Similarly, one-third of psychiatrists and psychologists thought that it was important to include the police version of the alleged offense in a trial competency report, but 27 percent of psychi-

atrists thought that inclusion was contraindicated. Only one-fourth of forensic psychologists thought that psychological testing was essential for competency and responsibility evaluations. On the one hand, two-thirds (67%) of psychiatrists indicated that ultimate issue opinions were essential for trial competency evaluations; on the other hand, 13 percent of psychiatrists indicated that it was contraindicated to include ultimate issue opinions in competency reports. Regarding criminal responsibility reports, 59 percent of psychiatrists thought that ultimate issue opinions were essential, but 20 percent thought that they were contraindicated.

Robbins and colleagues⁴⁴ evaluated 66 systematically selected competency-to-stand-trial reports from New Jersey (inpatient evaluations) and Oklahoma (community-based evaluations).⁴⁴ Two raters independently coded the reports to determine whether the reports provided relevant functional psycholegal ability data. In all but four reports, examiners expressed an ultimate issue opinion about the defendant's trial competency. More than 25 percent of the defendants were deemed "incompetent" in the reports by the examiners, but examiners sometimes failed to provide treatment recommendations to restore competency. In all but four reports, examiners recorded a psychiatric diagnosis, but only 27 percent of reports stated how the diagnosis affected the defendant's functional ability. Only 11 reports included psychological testing or review of the treatment record, but most of these failed to describe how these data related to functional deficits; 26 of the reports were based only on an interview with the defendant, without testing or third-party data. Examiners included extraneous information, such as treatment for conditions unrelated to trial competency deficits, treatment of individuals found to be competent, or opinions about risk of violence and criminal responsibility. The authors expressed disappointment in the quality of sampled competency reports. They recommended a national sample of trial competency assessment reports to determine practice in this area, and a standardized format for conducting and recording competency-to-stand-trial evaluations.

Skeem and colleagues⁴⁵ analyzed community examiners' reports of trial competency in Utah conducted between 1991 and 1994. Two clinicians examined each of the fifty defendants. There were 18 examiners, 80 percent of whom were psychologists.

Only two examiners had forensic diplomate status, so that most were general clinicians rather than trained forensic experts. With regard to the use of collateral data in the assessments, 65 percent of the examiners reviewed police reports, 37 percent reviewed mental health records, and 9 percent contacted defense counsel. Most examiners failed to consider the defendant's legal decision-making capacity, but they did address related abilities, such as appreciation of charges, knowledge of courtroom personnel, and ability to disclose information to counsel. Examiners typically failed to relate the defendant's psychopathology to impaired competency to stand trial. Even when examiners opined that the defendant was competent, they only infrequently provided details regarding the defendant's psycholegal abilities. The examiner pairs agreed as to clinical diagnosis (79%) and trial competence (82%), but not as to the specific trial competence domains.

Hecker and Steinberg⁴⁶ examined 172 juvenile predisposition reports from a Philadelphia area juvenile court jurisdiction between 1992 and 1996. The purpose of the reports was to provide relevant information to the juvenile court judge to guide court disposition. The reports had been prepared by four licensed psychologists. Report contents were quantitatively rated on a three-point scale for each content area in question. The investigators also examined the concordance between the psychologists' recommendations and the courts' disposition of the case. Perhaps, the most significant finding of the study was that only seven percent of the reports were deemed to have sufficient or better ratings of the evaluators' explanation of disposition recommendations. Judges were more likely to implement the evaluators' recommendations when there was adequate mental health information presented in the report and the evaluator adequately explained the recommendations.

Ryba and colleagues⁴⁷ surveyed psychologists regarding appropriate practice of competence-to-stand-trial evaluations of juveniles. Respondents ($n = 82$) were doctoral level psychologists throughout the United States with considerable experience in conducting adult and juvenile trial competency evaluations. The mailed, written survey instrument was modeled after that of Borum and Grisso,⁴³ which solicited information about the importance of elements of the competency evaluation instead of examining evaluators' reports. Similar results were ob-

tained in this study, despite the different subject pool and juvenile evaluation population, though more subjects in this study indicated that particular items were essential or required than in the earlier study. Again, 68 percent of the sample indicated that ultimate issue trial competency opinions were essential, but 10 percent thought that they were contraindicated. Differences in opinion on this matter may reflect divergent jurisdictional requirements. Ten percent thought that proffered opinions on factors other than the referral issue of trial competency were essential, but 33 percent thought that they were contraindicated. Regarding psychological testing, 44 percent indicated that testing was essential, and 35 percent recommended it, predominately intelligence testing. Forensic assessment instruments were thought to be essential in 30 percent of juvenile competence evaluation reports and recommended in 40 percent, predominately the Competence Assessment for Standing Trial for Defendants with Mental Retardation (CAST-MR). Subjects were also queried about the frequency of the use of tests and instruments, revealing somewhat lower frequencies. The investigators did not determine the accuracy of these self-reported test frequencies to ascertain whether evaluators actually do what they report.

Ryba and colleagues⁴⁸ separately reported on the assessment of juvenile maturity by psychologists in competency-to-stand-trial evaluations. A wide variety of techniques and as many as 34 tests were reportedly used in this regard, suggesting the absence of a definitive standard of practice in this area.

Zapf and colleagues⁴⁹ examined competency-to-stand-trial reports in Alabama that had been completed between 1994 and 1997. Only those reports in which the defendant was assessed as incompetent to stand trial were reviewed ($n = 53$). Similar to findings of other studies in this area, these reports omitted discussion of several psycholegal functional areas required by law. For instance, 22 percent of the reports did not address the defendant's understanding of the nature of the proceedings; 15 percent did not address the defendant's ability to participate effectively in the trial process; 93 percent failed to discuss the defendant's ability to appreciate his or her role in the legal proceedings; and 27 percent did not discuss the defendant's restorability to competence.

Christy and colleagues⁵⁰ evaluated juvenile competency-to-stand-trial evaluations conducted in Florida between 1997 and 2001 by private practitio-

ners, 95 percent of whom had doctoral or medical degrees. Only those evaluations of juveniles adjudicated incompetent to proceed and referred to the state's juvenile restoration program were reviewed (1,357 reports for 674 juveniles). The juveniles had a median of two reports each, and reports were a median of four pages in length (range, 0.5–15 pages). Nearly one-half (48%) of the reports failed to state where the evaluation was conducted. Arrest reports were reviewed in 38 percent of cases, school records in 12 percent, and interviews with defense attorneys in 2 percent. Sixty percent of reports indicated at least one third-party interview. Intelligence testing was used in 44 percent of cases, reflecting that mental retardation was the basis for a finding of incompetence in half the cases. Forensic assessment instruments were used in 29 percent of evaluations. Important cognitive mental status information was omitted in one-half the reports. However, evaluators' descriptions of the defendant's psycholegal functioning was better than that of clinical functioning, but often failed to provide examples or supporting data regarding psycholegal capacities. Ultimate-issue opinions were provided in 96 percent of cases, as required by Florida law. In only 62 percent of cases was the predicate condition for the incompetency finding stated in the report, contrary to legal requirements.

Warren and colleagues⁵¹ reviewed 5,175 criminal responsibility evaluations, over a 10-year period, conducted by 222 evaluators trained at the University of Virginia. Training was a five- to seven-day program that included a sample report segment and completion of a written examination. Results revealed that these evaluators spent three hours interviewing the defendant, and eight hours total on the case, including report preparation. A surprising finding was that evaluators often did not have available the defendant's criminal history, psychiatric or medical records, defendant's statements about the charges, or witness accounts. Psychologist evaluators used testing in 22 percent of cases, and psychiatrists did so in 6 percent. Of interest was the finding that less experienced evaluators were more likely to opine that the defendant met the standard for legal insanity. With regard to race, 8.5 percent of minority defendants were judged legally insane, compared with 11.4 percent of white defendants, raising a question about racial disparity in the conduct of forensic evaluations.

Psychological Test Usage

Psychological tests are an important and sometimes essential component of forensic evaluations. Testing, in theory, can improve the quality of forensic assessments, although data to establish this hypothesis are lacking. Yet, there are many relevant questions about test selection, administration, interpretation, qualification of the examiner, and admissibility in court required under a general acceptability or peer-based admissibility criterion.^{52,53} Data about actual test usage can provide answers to these questions.

Borum and Grisso^{43,54} studied the frequency of reported psychological test usage in criminal forensic evaluations, examining the same national sample reported earlier. Evaluators were asked their opinions about the importance of psychological testing in trial competence and criminal responsibility evaluations, as well as their reported use of tests and forensic instruments. They provided data on the types of tests and the specific tests used for these evaluations. Actual test usage was not determined, and the data were self-report only. Results indicated that psychiatrists and psychologists considered testing to be equally important (i.e., ratings of “essential” or “recommended”), but psychologists reported using or ordering tests more frequently than psychiatrists. Both groups considered psychological testing to be more important for conducting criminal responsibility than trial competence evaluations. Objective personality testing, especially with the MMPI-2, was the most commonly used psychological test for either evaluation.

Lees-Haley and colleagues⁵⁵ assessed the frequency of use of various neuropsychological tests in adult personal injury evaluations by a national sample of 100 expert neuropsychologists. Only reports of evaluations of traumatic brain injuries and toxic exposures were considered. The study did not rely on the self-report and recall of evaluators. The study observed that no single neuropsychologist used exactly the same test battery as any other and that the number of tests administered ranged from 1 to 32, with a mean of 11.7. The WAIS/WAIS-R, MMPI/MMPI-2, and the Wechsler Memory Scale-Revised WMS/WMS-R were the most frequent instruments used by examiners, and that finding represented a significant change from results of earlier studies of the use of neuropsychologist testing.

Boccaccini and Brodsky⁵⁶ examined the reported frequency of psychological testing by a national sample of forensic psychologists in emotional injury cases. The sample ($n = 80$) reported the percentage of evaluators who used a particular instrument, the percentage of cases in which a particular instrument was used, and the reasons for selecting a particular instrument. Results indicated that there was substantial variation of test selection across evaluators and that there was no standard assessment procedure, even in emotional injury cases. The MMPI was the only instrument used by more than one-half of the evaluators. The evaluator’s clinical experience with a specific test was as important as the availability of test norms in selecting tests. The Rorschach was the fourth most popular test used. Evaluators typically used four or five instruments in each evaluation. Actual frequency of test usage was not studied or reported.

Lally⁵⁷ surveyed a national sample of board-certified forensic psychologists regarding the frequency of use and opinions about acceptability of psychological test instruments in six criminal forensic practice areas. The evaluators ($n = 64$) were asked to rate specific instruments as to whether they were recommended, acceptable, or unacceptable and to report the frequency of test use. Evaluators used four to six tests, depending on the type of forensic evaluation, for each evaluation, with at least minimal frequency. A test was categorized as acceptable when at least half of the sample rated it acceptable. Projective tests generally were viewed as unacceptable across forensic evaluations, although the Rorschach was viewed less unfavorably than the others. Actual test usage by the forensic psychologists was not calculated.

Much additional research is needed on actual and desired use of psychological testing and the frequency and rationale for test administration, each across a variety of forensic settings and content areas. Investigation regarding the incremental value (i.e., additional information) that is produced by testing in many forensic contexts must be conducted. In many forensic evaluation contexts, it remains uncertain what tests, if any, and administered by whom, are essential to a quality forensic evaluation.

Several studies regarding psychological test selection are discussed in the next section with regard to child custody evaluations.

Child Custody Evaluation Practices

Keilin and Bloom⁵⁸ sent a written questionnaire to psychologists, psychiatrists, and masters-level practitioners who were child custody experts in private practice to assess their “activities, beliefs, and experiences relating to child custody” (Ref. 58, p 339). The evaluators ($n = 82$) provided data on the frequency of use of various custody evaluation procedures, such as individual interviews, observations of interactions, and school or home visits. They reported the use and frequency of use of various psychological tests for the parents and children, but did not report the overall frequency of testing. Overall, a mean of 19 hours was reportedly spent on each custody evaluation. The evaluators were asked to rate the importance of various decision-making criteria for custody in hypothetical cases based on a nine-point Likert scale (e.g., child’s preferences, parental alienation, quality of bonding, parental stability). They reported their actual and preferred child custody recommendations.

Ackerman and Ackerman⁵⁹ surveyed 201 doctoral-level psychologists from 39 states about 112 aspects of child custody evaluations. Almost all of the sample (88%) worked in private practice. The survey paralleled that of Keilin and Bloom,⁵⁸ but excluded psychiatrists and social workers. Eight percent of respondents stated that they did not test children, and two percent did not test adults. Respondents reportedly spent a mean of 26 hours on each custody evaluation, excluding consulting with attorneys and court testimony.

LaFortune and Carpenter⁶⁰ surveyed by mail mental health professionals from five states who were experienced child custody evaluators. No psychiatrists were among the sample ($n = 165$). The investigators did not review child custody reports, but, using a five-point scale, queried the sample about their training and experience, interaction with the legal profession, desirable characteristics of evaluators, preferences for retention, and the role of experts. The respondents reported spending 21 hours completing a custody evaluation, including report writing. An “advocacy index” was presented based on the sample’s response to questions about involvement with the retaining attorney.

Bow and Quinnell⁶¹ evaluated psychologists’ reported practices in child custody evaluations. Respondents ($n = 198$) were a national sample of clinical and forensic psychologists with master’s or

doctoral degrees, with 92 percent in private practice. Evaluators were asked to rank child custody evaluation procedures according to their importance, and state their preferred procedures and average time spent thereon. They rated the importance of statutory child custody criteria. Nearly the entire sample (94%) reported that they made an explicit custody/visitation recommendation in their reports. Custody reports averaged 21 pages, with a range of 4 to 80 pages. They also reported use of psychological tests and parent-rating scales. Psychological testing was considered as adjunctive to clinical interviews with the parents and children, rather than the primary procedure, though the parent was reportedly administered an MMPI in 88 percent of cases.⁶²

Bow and Quinnell⁶³ examined 52 child custody reports prepared by a national sample of doctoral-level psychologists. The reports ranged from 5 to 63 pages, with a mean of 24 pages. The report format and content varied widely. Specific problems noted by the investigators included failure to identify the specific evaluation procedures, the reason for the referral, or the specific testing used. Some reports did not use collateral contacts, demonstrated over-reliance on testing, were adult rather than child focused, or were conducted by an evaluator who had served as a therapist.

Surveys of Judges and Attorneys

Owens and colleagues^{64,65} surveyed and interviewed two small samples of trial judges in New York about their reasons for ordering trial competency evaluations and their satisfaction with them. Generally, they indicated that the evaluations were useful to them and had few complaints about them. However, some judges were interested in defendants’ backgrounds and clinical data and often transformed the competency evaluation into an evaluation of violence risk, need for treatment, or sentencing.

LaFortune and Nicholson⁶⁶ conducted via mail a semistructured survey of Oklahoma judges and attorneys about the adequacy of already submitted competency-to-stand-trial evaluations in two metropolitan areas. The subjects ($n = 110$) were asked to rate retrospectively six characteristics of expert reports provided by mental health professionals, using a five-point Likert scale: (1) timeliness of the reports; (2) familiarity of the examiners with legal criteria; (3) use of understandable language by the examiners; (4) presence of a factual basis for examiners’ conclusions;

(5) usefulness of reports in decision-making; and (6) overall quality of reports. Outpatient competency reports were judged to be of higher quality than inpatient reports. Subjects also reported their perceptions of the actual and optimal frequency of describing defendant characteristics in competency reports, typically indicating that reports tended to summarize rather than detail the relevant functional capacities. Generally, attorneys requested more specificity in reports, such as “more detailed descriptions of the factual bases for legal conclusions” (Ref. 66, p 248). Attorneys also stated their request for criminal responsibility information and opinions, even though that is an extraneous issue.

Redding and colleagues⁶⁷ surveyed by mail experienced Virginia judges, prosecutors, and criminal defense attorneys about the testimony of mental health experts, by using a hypothetical insanity defense case. Questionnaires including vignettes were sent to the subjects, who were asked to rate the importance of eight types of mental health testimony evidence on a nine-point Likert scale. The subjects ($n = 72$) responded that they were primarily interested in the defendant’s clinical diagnosis, whether the clinical condition met the relevant criminal responsibility standard, and the expert’s ultimate opinion on the legal issue. The subjects were less interested in research or actuarial evidence from experts, such as diagnostic reliability and statistical crime data related to the diagnosis. Ratings of judges and prosecutors correlated strongly, but defense attorneys ranked clinical diagnosis, statistical data on diagnostic reliability, and explanations of criminal behavior to be more probative than did the other two groups. This difference probably reflects the divergent roles and needs of the subjects for mental health testimony. These responses from members of the bar highlight the hazards of over-reliance by judges and attorneys on conclusory legal testimony by mental health experts, as well as the need to educate the judiciary about the value of social science.

Dahir and colleagues⁶⁸ conducted a telephone and mail survey of a national sample of 325 state trial judges in 1998 and 1999. While judges supported their gatekeeping role as defined by *Daubert*,⁵³ they lacked the scientific literacy required by *Daubert*. Few judges understood concepts such as falsifiability or error rate.⁶⁹ They therefore ruled, improperly, on the admissibility of psychological syndrome evidence in cases at bar, based primarily on the qual-

ifications of the expert and general acceptance considerations.⁶⁸ Additional training to improve judges’ scientific understanding was recommended.

Analysis of Empirical Forensic Practice Studies

There are many limitations to the research that has been conducted regarding actual or reported forensic practice, though they provide some useful information. These empirical studies almost exclusively rely on expert reports, which, as noted, represent just one component of the evaluation. Characteristic of any documentation, the forensic report is only a window into the evaluation itself and is therefore limited. The report can only be as good as the evaluation that precedes it. We can work toward improving report content, writing, exposition, critical thinking, and decision making, but that effort goes only so far. In addition, the studies focus on criminal forensic evaluations in state court and child custody evaluations, with much less empirical literature on other forensic evaluations. Studies of trial competence evaluations predominate over those of criminal responsibility. Generalization of the study data are an important concern, given the limited sample sizes, sampling procedure, geography, legal jurisdiction, statutory requirements, characteristics of defendants and examiners, and location of evaluation in the hospital or community. Subject selection and recruitment are generally not random, and it is often unclear what would constitute a random sample of evaluators.

As yet, there are no published studies of observed forensic evaluations, or of observed evaluations in conjunction with written reports. The studies reveal problems in expert reports, such as the absence of stated reasoning to support the expert opinion (i.e., on causation, psycholegal function), the use of ultimate issue conclusions, the offering of opinions on extraneous areas, and the failure to acknowledge the limitations of the report or evaluation.^{45,70} Only some of the studies included designated quality measures, which are necessary to evaluate the adequacy of forensic reasoning and decision making in evaluation reports. Some studies inquire about the beliefs of evaluators, or their report of the frequency with which they use a test or instrument, but there is no accompanying evidence of the accuracy of these self-reports. Thresholds of evaluation quality probably differ from study to study.

Nevertheless, the studies raise substantial questions about the quality and thoroughness of criminal

forensic evaluations, given the variability of the use of collateral information, contact with an attorney, psychological testing, and use of forensic assessment instruments.^{40,45,70} One of the most common report weaknesses is the failure to substantiate expert opinions and the related failure to relate psychopathology to expert opinions regarding psycholegal abilities.^{40,45,70} The relationships among symptom, diagnosis, psychological test-identified deficits, and psycholegal functional impairment is too often neglected.^{45,70} Failure to explicate these links comprehensively, including the evaluator's reasoning in reaching opinions, is likely to render the evaluation report less useful and cogent to the attorney or court that requested it. While there is sometimes consensus about evaluators' views regarding the desired content of reports, there is often lack of such consensus, depending on the specific content.

The empirical studies of psychological test usage reveal different types of information: reported frequency of test usage, importance of the test, acceptability of test usage, and reasons for test selection. Most of the studies are limited by reliance on the evaluators' retrospective memories for test selection and use, which are likely to be highly fallible. Test selection is reported without regard for the particular referral or clinical situation.⁶² In the only study to examine actual test usage, Bow and Quinnell⁶³ examined actual reports submitted to them by evaluators. Some tests were judged acceptable for some forensic evaluations but not for others. Examining reported test use is not equivalent to judging whether a test satisfies general acceptance criteria under *Frye*⁵² or *Daubert*.^{53,57,71} Even if a test is judged as recommended by evaluators, such an opinion does not indicate that the standard of quality care requires such test use. It is noteworthy that the studies often reveal great divergence in test selection for forensic evaluations.

Practice Parameters and Guidelines for Forensic Evaluations

Examination of practice guidelines with regard to conducting forensic evaluations is relevant to determining evaluation quality and to quality improvement. Practice guidelines recommend specific professional conduct but usually differ from standards that are mandatory and prescriptive.⁷² They are designed to assist and educate clinicians and patients, rather than to legislate practice in a particular area. In

contrast, medical review criteria are designed to evaluate health care outcomes and decisions and can differ from guidelines.⁷³ Practice guidelines are typically prepared by many of the most knowledgeable experts in a particular area and are sponsored by professional medical societies, health plans, or government agencies. Ideally, they reflect professional consensus in a field. Practice guidelines are much more prevalent for medical treatment rather than for evaluation, and some organizations differentiate between treatment guidelines and practice guidelines.⁷⁴ There are thousands of published practice guidelines in medicine, but relatively few for forensic evaluations.⁷⁵

Existing practice guidelines pertinent to forensic evaluations include those for child custody,⁷⁶⁻⁷⁹ child and adolescent abuse victims,⁸⁰ child protection,⁸¹ juvenile sex offenders,⁸² conduct disorder,⁸³ post-traumatic stress disorder,^{84,85} criminal responsibility,⁸⁶ and physicians' fitness for duty.⁸⁷ The practice guideline regarding suicide, published by the American Psychiatric Association, also discusses suicide assessment.⁸⁸ These guidelines are applicable to both generalist and specialist practitioners in the field.

Although practice parameters in forensic psychiatry can be useful in determining what constitutes a quality forensic evaluation, practice parameters are problematic in this regard. Most guidelines are written generally rather than in a detailed, specific, cookbook-like manner and cannot be readily translated or put into practice or even serve as practice review criteria.⁸⁹ Many are presented as guidelines that are to be aspired to rather than adhered to. Practice guidelines generally are slow and expensive to develop and are soon out of date.⁹⁰ Many guidelines are not evidence based, but only state the beliefs and practices of the authors. Many do not properly synthesize the available scientific evidence.⁹¹ Other barriers to successful implementation include that physicians are unaware of their existence, lack knowledge of their content, disagree with their parameters, have negative expectations regarding their likely outcome, and have difficulty changing practice habits.⁹² External barriers to their use include lack of time and resources, organizational constraints, lack of reimbursement, inability to generalize their use to every situation, fear of liability, and difficulty of use.⁹² Nevertheless, quality improvement activities in forensic evaluations must consider the extant forensic

Table 1 Sample Quality Measures for Forensic Evaluations

| |
|--|
| Likert scale ratings of the following: |
| Timeliness of evaluation and report |
| Comprehensiveness of evaluation |
| Evaluator's demeanor with evaluatee |
| Use of specific collateral sources |
| Use of specific tests and forensic instruments |
| Factual support for clinical diagnosis in report |
| Organization of report and use of jargon |
| Knowledge of legal criteria |
| Stated psycholegal functional ability, with specific support for opinion |
| Use, or avoidance, of ultimate issue opinions |
| Scientific accuracy of opinions (i.e., course of illness) |
| Bias, objectivity, honesty, fairness, partisanship, humility |
| Stated limitations to evaluation and conclusions |
| Absence of irrelevant opinions |
| Overall quality |

practice parameters. Much work remains to be accomplished in the development of forensic practice guidelines that reconcile the professional differences across all types of forensic practitioners regarding optimal practice.

Quality Concerns in Forensic Evaluations

The factors in quality of health care that were discussed earlier are relevant to our approach to quality assessment and improvement of forensic evaluations. As described earlier, there are several dimensions or domains of forensic evaluation quality. An evaluation, therefore, can be evaluated and improved in different respects. In this regard, the qualities of a forensic evaluation include timeliness, safety to evaluatee and evaluator, effectiveness, efficacy, efficiency, evaluatee-centeredness, and absence of disparity. Quality can be assessed by examination of the evaluation process, structure, and outcome. In these different respects, quality ratings can be individually or simultaneously conducted by the evaluator, evaluator's employer, client/payer, evaluatee, professional peer, and others.

Quality indicators for forensic evaluations (i.e., general areas of interest but not specific measures) remain to be developed, after additional research. Typical indicators might include expertise (training and experience), the expert's role, data that form the basis of the evaluation, demeanor during interview, and expert decision-making. Specific measures regarding each indicator would then be developed. For the indicator of demeanor or communication with

the evaluatee, for instance, specific items to be measured could include whether the evaluator is disrespectful, argumentative, cynical, rejecting, angry, or biased. Table 1 lists several other sets of measures for evaluations. As in the general health care setting, achieving consensus regarding performance measures is a critical task for the forensic mental health field, and has yet to be achieved.

As in general medical care, there are many stakeholders with regard to forensic evaluations, and their repeated input is essential. Arguably, our peers are the best judges of the quality of our work—better than the retaining attorneys. Attorneys excessively rely on cross-examination of court testimony to expose the weaknesses and failures of forensic evaluations, with various degrees of success.^{93–95} Cross-examination is not, however, a quality improvement mechanism, but has different purposes. It is unsurprising, therefore, that attorneys, who are the ones who typically retain evaluators, are complacent about the deficient quality of forensic evaluations and reports. Nevertheless, soliciting client feedback through the use of satisfaction ratings could be useful. These ratings would be completed by the client (i.e., attorney, court, agency, evaluatee, payer). Ratings could occur on a prospective or retrospective basis, after every evaluation, or a subsample of them. Table 2 lists several possible areas of inquiry. The primary purpose of the client satisfaction instrument should be for quality improvement rather than marketing.

A major lesson for those who work in the forensic mental health field, derived from the work on quality in health care, is that, rather than examine only the work of the individual clinician, we must consider the entire system of care. Third-party and other collateral data are essential for conducting forensic mental health evaluations, and barriers to obtaining such

Table 2 Client Satisfaction Form

| |
|---|
| Likert scale ratings of the following: |
| Were the data in the report accurate? |
| Was the evaluation comprehensive? |
| Was the evaluation adequate? |
| Was the clinical content of the report informative? |
| Was the forensic content of the report informative? |
| Was the report well written and understandable? |
| Were the conclusions and recommendations of the report useful? |
| Was there adequate opportunity to communicate with the evaluator? |
| Was the evaluation and report conducted in a timely manner? |
| What is the overall impression of the evaluation and report? |
| How could the evaluation and report be made more useful? |

Table 3 Deficiencies in Forensic Evaluations and Reports

| |
|---|
| Lack of training and expertise in the psycholegal content area |
| Advocacy, impartiality, lack of objectivity (e.g., therapist evaluator) |
| Other forensic boundary crossing or violation ⁹⁸ |
| Inadequate database of documents |
| Over-reliance on litigant self-report |
| Lack of assessment of litigant response bias |
| Inadequate collateral information ^{96,97,99} |
| Failure to use validated forensic assessment instruments appropriately ^{100,101} |
| Failure to support clinical diagnoses |
| Psychopathology not linked to expert opinions |
| Psychological test results not linked to expert opinions |
| Inadequate support and explanation for expert opinions |
| Failure to state limitations of methods and opinions |

information must be eliminated.^{96,97} Process and outcome must be addressed as well. The costs of quality assessment and improvement must be considered, perhaps tempered by the ability in the future to adopt technology, such as software applications, to assist with review of expert reports in critical areas.

Raising the Quality of Forensic Evaluations

The general principles of quality improvement are useful in the context of advancing the quality of forensic evaluation. Initially, a quality improvement agenda must be established, with advance planning, motivation, and incentives. There must be financial support for quality improvement efforts. Quality indicators and measures must be developed and psychometrically evaluated. Other principles include the use of specific quality improvement intervention, regular audit and feedback, empirical assessment of change, and assessment of the barriers to change in individuals and the system.

As discussed, the empirical literature regarding forensic evaluations documents the need to raise the quality of evaluations and reports (Table 3). This is especially a concern regarding evaluations conducted by the generalist clinician who has no formal training or supervised experience in conducting forensic evaluations. Tolman and Mullendore¹⁰² reported that risk assessments performed by forensic specialists were of higher quality than those conducted by generalists with regard to the use of psychological assessment instruments; further, specialists were more familiar with the risk assessment literature than were generalists. In general medicine, research has been conducted on the relationship between board certification and clinical patient outcomes, with mixed

results, though many studies reveal a positive correlation in this regard.¹⁰³ As to forensic mental health evaluations, Otto and Heilbrun¹⁰⁴ delineated several categories of forensic experts: accidental experts brought into court by their patient’s attorney; legally informed clinicians without specialized forensic experience or training; proficient clinicians with supervised forensic experience; and specialist clinicians with the highest level of forensic experience, formal training, and board certification.¹⁰⁴ Further research regarding the respective quality of evaluations conducted by these categories of experts is needed. Quality improvement approaches probably should be individualized for the different levels of practitioners, depending on the identified deficiencies.

Review of quality improvement activities in general medicine leads to the conclusion that no single, specific intervention is likely to be successful in improving the quality of forensic evaluation. Rather, a variety of interventions is appropriate, depending on the forensic service organization, model of service delivery, density of local forensically trained specialists and educators, and other factors. Table 4 lists some of these interventions. Successful interventions are not likely to be single events, but will require repetition and modification over time.

Perhaps the most significant problem area identified by the empirical studies of forensic reports is the evaluator’s failure to determine the reasoning by which the opinions are reached.^{40,45} Evaluators have repeatedly failed to link psychopathology, psychological test results, psychiatric diagnosis, and legal functional impairments. This omission is character-

Table 4 Raising the Quality of Forensic Evaluations

| |
|---|
| Forensic service credentialing and certification by law or policy |
| Incentivize quality and quality improvement, with quality bonus |
| Establish quality guidelines and standards ⁹⁹ |
| Define, test, and operationalize quality performance measures and tools ⁹⁹ |
| Collaboration of generalists with forensic specialists |
| Audiotape or videotape forensic interviews ¹⁰⁵ |
| Use of specific checklists and contracts by referral sources |
| Education of referral sources about evaluation guidelines, and sharpening of referral questions |
| Model excellence of forensic evaluations to attorneys and courts |
| Peer review of evaluations, reports, and testimony |
| Encourage cross-examination regarding participation in QA activities |
| Mandatory forensic continuing medical education on quality of evaluations |
| Maintenance of forensic board certification predicated on quality improvement activity |

istic of both forensically trained and untrained evaluators. For maximum effect, quality improvement activities must focus on this deficit, and the other frequently identified evaluation weaknesses, such as the use of collateral informants and data.^{96,97} Such an improvement will strengthen the evaluation's credibility and utility to attorneys and courts.

Several states have, by statute or general practice, developed credentialing procedures for forensic evaluators, usually in criminal or child custody cases.¹⁰⁶ Those evaluators have attended mandatory training and may have completed a local examination. Recertification is, or could be, accomplished by incorporating quality audits of the evaluator's work product. The Commonwealth of Massachusetts, a leader in this respect, has had for many years a routine quality review and improvement component of expert reports for Designated Forensic Professionals who conduct criminal pretrial evaluations.^{107,108} Similarly, criminal court clinics are ideally situated to institute quality review activities of their employee-evaluators, as a condition of employment and promotion. Forensic credentialing for faculty can also be accomplished by university departments of psychiatry and hospital medical staff committees. The forensic mental health field could encourage legislation or policy that further promotes forensic credentialing, accompanied by mandatory quality audits and improvement activity. Volume credentialing requirements could also be set as they are for radiologists (i.e., mammography) and cardiothoracic surgeons, based on the reported association between volume and evaluator performance demonstrated even in mental health treatment.¹⁰⁹⁻¹¹² Specific credentialing for types of forensic evaluations could be implemented. For instance, mandatory forensic credentialing for participation in pretrial or postconviction death penalty cases is desirable, given the importance of such evaluations.¹¹³

The Maintenance of Certification program of the American Board of Medical Specialties is intended to increase the "accountability, service and quality of medical care" (Ref. 114, p 4). Components of this program include self-assessment, life-long learning, and performance in practice.¹⁰³ Though problems with physician clinical performance assessment have been identified, the assessment of the competence and performance of forensic evaluators is certainly a feasible enterprise once consensus on performance measures has been obtained.^{115,116}

Credentialing and recertification should require some review by peers of one's actual, not reported, work. Peer review of forensic testimony has been offered through the national and local efforts of members of the American Academy of Psychiatry and the Law (JAAPL), although few expert witnesses have participated.^{117,118} No attempt has been made for peer review of actual evaluations, which consumes more resources but is likely to be more valuable.

Role of AAPL

AAPL has many opportunities to initiate quality improvement activities in forensic psychiatry. Educational, consultative, and research activities are within the purview of AAPL. The organization, through its committees or new task forces, could identify quality indicators, measures, and barriers to change. AAPL could prepare resource documents, position statements, and practice guidelines regarding evaluation of quality and quality improvement. The Awards Committee could acknowledge forensic service programs that have instituted quality improvement activities, much as the American Psychiatric Association gives awards for exemplary clinical service programs. The newly created AAPL Institute for Education and Research could play a substantial role in funding researchers to develop quality indicators and measures or to train forensic evaluators to improve the quality of their work. Beyond AAPL's traditional involvement in education and research, an expanded role could include performing quality audits of forensic evaluations for organizations on request.

Liaison among forensic mental health organizations and groups is essential in the quality improvement enterprise, and AAPL should be a leader in this field. Forensic psychiatrists often work on individual cases in conjunction with clinical or neuropsychologists. Almost all of the empirical work on forensic reports has been conducted by psychologists, and the subjects of those investigations have predominately been psychologists. The studies reveal significant differences between psychiatrists and psychologists in the frequency of use of psychological testing and other evaluation modalities. Even among psychologists, selection of tests and forensic assessment instruments varies widely in forensic evaluations. The role and appropriateness of the psychological tests administered by psychiatrists has not been explicated.

Consensus among differing forensic mental health practitioners in these areas is important.

Conclusions

Much work remains to be accomplished in improving the quality of forensic evaluations of evaluators with or without specific forensic training and experience.¹¹⁹ There are numerous barriers to quality improvement, including motivational or psychological factors. Of great concern is that practitioners are likely to be satisfied with the quality of their evaluations, believe that little improvement is needed, and fail to undertake self-assessment or quality improvement unless externally mandated.¹⁰² Imposing an external quality agenda or standard may be considered a threat to the evaluator's autonomy and narcissism.¹²⁰ Within organizations, barriers include financial and bureaucratic resistance to change. Scientific barriers include the absence to date of established quality indicators and performance measures. There is resistance to change within the legal system, including infrequent use of court-appointed evaluators, except in child custody cases. Despite the adoption of gatekeeping efforts incorporated in the rules of evidence, our adversarial legal system has excessively relied on cross-examination as its primary filter for excluding pseudoscience and inadequate expertise.⁵³ Legal issues such as confidentiality and privilege of evaluations and reports must be addressed before they are shared with quality improvement staff or peers.⁹⁹ Further, due to extant financial incentives, there is, regrettably, a market for mediocre quality forensic evaluations conducted with questionable or antiquated methods.^{121,122}

In the long-term future, we expect that quality improvement at a more sophisticated level will transcend anything discussed heretofore. We hope to move beyond the exclusive reliance on unguided clinical judgment as the basis for expert forensic opinion, given its well-documented fallibility.^{123,124} Improvements in clinical and forensic decision-making are essential developments for the field.^{125,126} Just as technological developments will no doubt be used to improve the quality of general health care, we hope to see computer-assisted decision-making for risk assessments and other forensic evaluations,¹²⁷⁻¹³⁰ use of forensic cognitive neuroscience tools and specific forensic assessment instruments,⁹⁷ and objective psychophysiological testing. We need a more complete and scientific understanding of self-

report,¹³¹ clinical judgment,¹²⁴⁻¹²⁶ heuristics,^{126,132} and memory,^{133,134} given their central roles in forensic evaluation content and expert decision making. We also need improved methods of detecting, quantifying, and correcting for bias and lack of objectivity in experts.¹³⁵⁻¹³⁸

References

1. Dietz PE: The quest for excellence in forensic psychiatry. *Bull Am Acad Psychiatry Law* 24:153-63, 1996
2. Barnum R: An agenda for quality improvement in forensic mental health consultation. *Bull Am Acad Psychiatry Law* 21:5-21, 1993
3. Leape LL, Berwick DM, Bates DW: What practices will most improve safety? *JAMA* 288:501-7, 2002
4. Agency for Healthcare Research and Quality: 2004 National Healthcare Quality Report. Rockville, MD: Agency for Healthcare Research and Quality, December 2004; AHRQ Publication No. 05-0013
5. Institute of Medicine: Crossing the Quality Chasm. Washington, DC: National Academy Press, 2001
6. Morreim H: Assessing quality of care: new twists from managed care. *J Clinical Ethics* 10:88-99, 1999
7. Pronovost P, Nolan T, Zeger S, *et al*: How can clinicians measure safety and quality in acute care? *Lancet* 363:1061-7, 2004
8. Cassel C: Quality of care and quality of training: a shared vision for internal medicine? *Ann Intern Med* 140:927-8, 2004
9. Institute of Medicine: To Err is Human: Building a Safer Health System. Washington, DC: National Academy Press, 1999
10. Landrigan CP, Rothschild JM, Cronin JW, *et al*: Effect of reducing interns' work hours on serious medical errors in intensive care units. *N Engl J Med* 351:1838-8, 2004
11. Koppel R, Metlay JP, Cohen A: Role of computerized physician order entry systems in facilitating medication errors. *JAMA* 293:1197-203, 2005
12. Lighter DE, Lighter SA: Legal and regulatory issues in quality improvement, in *Quality Management in Health Care* (ed 2). Edited by Lighter DE, Fair DC. Sudbury, MA: Jones and Bartlett Publishers, 2004, pp 429-540
13. Saag KG, Olivieri J, Patino F, *et al*: Measuring quality in arthritis care: The Arthritis Foundation's quality indicator set for analgesics. *Arthritis Rheum* 51:337-49, 2004
14. Romano PS, Mutter R: The evolving science of quality measurement for hospitals. *Int J Health Care Finance Econ* 4:131-57, 2004
15. McGlynn EA, Asch SM, Adams J, *et al*: The quality of health care delivered to adults in the United States. *N Engl J Med* 348:2635-45, 2003
16. Landon BE, Wilson IB, McInnes K, *et al*: Effects of a quality improvement collaborative on the outcome of care of patients with HIV infection. *Ann Intern Med* 10:887-96, 2004
17. Institute of Medicine: Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare. Washington, DC: National Academy Press, 2003
18. Agency for Healthcare Research and Quality: 2004 National Healthcare Disparities Report. Rockville, MD: Agency for Healthcare Research and Quality, December 2004; AHRQ Publication No. 05-0014
19. Sehgal AR: Impact of quality improvement efforts on race and sex disparities in hemodialysis. *JAMA* 289:996-1000, 2003
20. Rosenthal MB, Fernandopulle R, Song HSR, *et al*: Paying for quality: providers' incentives for quality improvement. *Health Affairs* 23:127-41, 2004

21. Epstein AM, Lee TH, Hamel MB: Paying physicians for high-quality care. *N Engl J Med* 350:406–10, 2004
22. Roland M: Linking physicians' pay to the quality of care: a major experiment in the United Kingdom. *N Engl J Med* 351:1448–54, 2004
23. Krumholz HM, Rathore SS, Chen J, *et al*: Evaluation of a consumer-oriented internet health care report card. *JAMA* 287:1277–87, 2002
24. McCormick D, Himmelstein DU, Woolhandler S, *et al*: Relationship between low quality-of-care scores and HMOs' subsequent public disclosure of quality-of-care scores. *JAMA* 288:1484–90, 2002
25. Werner RM, Asch DA: The unintended consequences of publicly reporting quality information. *JAMA* 293:1239–44, 2005
26. Grol R, Lawrence M: *Quality Improvement by Peer Review*. New York: Oxford University Press, 1995
27. Edmunds M, Frank R, Hogan M, *et al*: *Managing Managed Care: Quality Improvement in Behavioral Health*. Washington, DC: National Academy Press, 1997
28. American Psychiatric Association Task Force on Quality Indicators: *Quality Indicators: Defining and Measuring Quality in Psychiatric Care for Adults and Children*. Washington, DC: American Psychiatric Publishing, 2002
29. Ishak WW, Burt T, Sederer LI (editors): *Outcome Measurement in Psychiatry*. Washington, DC: American Psychiatric Publishing, 2002
30. Schwenk TL, Evans DL, Laden SK, *et al*: Treatment outcome and physician-patient communication in primary care patients with chronic, recurrent depression. *Am J Psychiatry* 161:1892–901, 2004
31. Wells KB, Sherbourne C, Schoenbaum M, *et al*: Impact of disseminating quality improvement programs for depression in managed primary care. *JAMA* 283:212–20, 2000
32. Weilburg JB, O'Leary KM, Meigs JB, *et al*: Evaluation of the adequacy of outpatient antidepressant treatment. *Psychiatr Serv* 54:1233–9, 2003
33. Young AS, Klap R, Sherbourne CD, *et al*: The quality of care for depressive and anxiety disorders in the United States. *Arch Gen Psychiatry* 58:55–61, 2001
34. Dietrich AJ, Oxman TE, Williams JW: Re-engineering systems for the treatment of depression in primary care: cluster randomized controlled trial. *BMJ* 329:602–5, 2004
35. Gilbody S, Whitty P, Grimshaw J, *et al*: Educational and organizational interventions to improve the management of depression in primary care. *JAMA* 289:3145–51, 2003
36. Asarnow JR, Jaycox LH, Duan N, *et al*: Effectiveness of a quality improvement intervention for adolescent depression in primary care clinics. *JAMA* 293:311–19, 2005
37. Hatfield DR, Ogles BM: The use of outcome measures by psychologists in private practice. *Prof Psychol* 35:485–91, 2004
38. National Committee for Quality Assurance, available at <http://www.ncqa.org> (accessed March 28, 2005)
39. Garland AF, Aarons GA, Hawley KM, *et al*: Relationship of youth satisfaction with mental health services and changes in symptoms and functioning. *Psychiatr Serv* 54:1544–6, 2003
40. Nicholson RA, Norwood S: The quality of forensic psychological assessments, reports and testimony: acknowledging the gap between promise and practice. *Law Hum Behav* 24:9–44, 2000
41. Petrella RC, Poythress NG: The quality of forensic evaluations: an interdisciplinary study. *J Consult Clin Psychol* 51:76–85, 1983
42. Heilbrun K, Collins S: Evaluations of trial competency and mental state at time of offense: report characteristics. *Prof Psychol* 26:61–7, 1995
43. Borum R, Grisso T: Establishing standards for criminal forensic reports: an empirical analysis. *Bull Am Acad Psychiatry Law* 24:297–317, 1996
44. Robbins E, Waters J, Herbert P: Competency to stand trial evaluations: a study of actual practice in two states. *Bull Am Acad Psychiatry Law* 25:469–83, 1997
45. Skeem JL, Golding SL, Cohn NB, *et al*: Logic and reliability of evaluations of competence to stand trial. *Law Hum Behav* 22:519–47, 1998
46. Hecker T, Steinberg L: Psychological evaluation at juvenile court disposition. *Prof Psychol* 33:300–6, 2002
47. Ryba NL, Cooper VG, Zapf PA: Juvenile competence to stand trial evaluations: a survey of current practices and test usage among psychologists. *Prof Psychol* 34:499–507, 2003
48. Ryba NL, Cooper VG, Zapf PA: Assessment of maturity in juvenile competency to stand trial evaluations: a survey of practitioners. *J Forensic Psychol Pract* 3:23–45, 2003
49. 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* 4:27–44, 2004
50. Christy A, Douglas KS, Otto RK, *et al*: Juveniles evaluated incompetent to proceed: characteristics and quality of mental health professionals' evaluations. *Prof Psychol* 35:380–8, 2004
51. Warren JI, Murrie DC, Chauhan P, *et al*: Opinion formation in evaluating sanity at the time of the offense: an examination of 5175 pre-trial evaluations. *Behav Sci Law* 22:171–86, 2004
52. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923)
53. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993)
54. Borum R, Grisso T: Psychological test use in criminal forensic evaluations. *Prof Psychol* 26:465–73, 1995
55. Lees-Haley PR, Smith HH, Williams CW, *et al*: Forensic neuropsychological test usage: an empirical survey. *Arch Clin Neuropsychol* 11:45–51, 1996
56. Boccaccini MT, Brodsky SL: Diagnostic test usage by forensic psychologists in emotional injury cases. *Prof Psychol* 30:253–9, 1999
57. Lally SJ: What tests are acceptable for use in forensic evaluations?—a survey of experts. *Prof Psychol* 34:491–8, 2003
58. Keilin WG, Bloom LJ: Child custody evaluation practices: a survey of experienced professionals. *Prof Psychol* 17:338–46, 1986
59. Ackerman MJ, Ackerman MC: Custody evaluation practices: a survey of experienced professionals (revisited). *Prof Psychol* 28:137–45, 1997
60. LaFortune KA, Carpenter BN: Custody evaluations: a survey of mental health professionals. *Behav Sci Law* 16:207–24, 1998
61. Bow JN, Quinnell FA: Psychologists' current practices and procedures in child custody evaluations: five years after American Psychological Association guidelines. *Prof Psychol* 32:261–8, 2001
62. Quinnell FA, Bow JN: Psychological test used in child custody evaluations. *Behav Sci Law* 19:491–501, 2001
63. Bow JN, Quinnell FA: A critical review of child custody evaluation reports. *Family Court Rev* 40:164–76, 2002
64. Owens H, Rosner R, Harmon RB: The judge's view of competency evaluations. *Bull Am Acad Psychiatry Law* 13:389–97, 1985
65. Owens H, Rosner R, Harmon RB: The judge's view of competency evaluations II. *Bull Am Acad Psychiatry Law* 15:381–9, 1987
66. LaFortune KA, Nicholson RA: How adequate are Oklahoma's mental health evaluations for determining competency in criminal proceedings?—the bench and the bar response. *J Psychiatry Law* 23:231–62, 1995
67. Redding RE, Floyd MY, Hawk GL: What judges and lawyers

Quality Improvement in Mental Health Evaluations

- think about the testimony of mental health experts: a survey of the courts and bar. *Behav Sci Law* 19:583–94, 2001
68. Dahir VB, Richardson JT, Ginsburg GP, *et al*: Judicial application of Daubert to psychological syndrome and profile evidence. *Psychol Public Policy Law* 11:62–82, 2005
 69. Gatowski SI, Dobbin SA, Richardson JT: Asking the gatekeepers: a national survey of judges on judging expert evidence in a post-Daubert world. *Law Hum Behav* 25:433–58, 2001
 70. Skeem JL, Golding SL: Community examiners' evaluations of competence to stand trial: common problems and suggestions for improvement. *Prof Psychol* 29:357–67, 1998
 71. Weiner IB, Exner JE, Sciarra A: Is the Rorschach welcome in the courtroom? *J Pers Assess* 67:422–4, 1996
 72. Field MJ, Lohr KN (editors): *Clinical Practice Guidelines*. Washington, DC: National Academy Press, 1990
 73. Field MJ, Lohr KN (editors): *Guidelines for Clinical Practice*. Washington, DC: National Academy Press, 1992
 74. American Psychological Association: Criteria for practice guideline development and education. *Am Psychol* 57:1048–51, 2002
 75. National Guideline Clearinghouse, <http://www.guideline.gov> (accessed March 28, 2005)
 76. American Academy of Child and Adolescent Psychiatry: Practice parameters for child custody evaluation. *J Am Acad Child Adolesc Psychiatry* 36:57S–68S, 1997
 77. American Psychological Association: Guidelines for child custody evaluations in divorce proceedings. *Am Psychol* 49:677–80, 1994
 78. Association of Family and Conciliation Courts: Model standards of practice for child custody evaluations. *Fam Concil Courts Review* 32:504–13, 1994
 79. American Psychiatric Association: Child custody consultation: a report of the Task Force on Clinical Assessment in Child Custody. Washington, DC: American Psychiatric Association, 1981, revised 1988
 80. American Academy of Child and Adolescent Psychiatry: Practice parameters for the forensic evaluation of children and adolescents who may have been physically or sexually abused. *J Am Acad Child Adolesc Psychiatry* 36:37S–56S, 1997
 81. American Psychological Association: Guidelines for psychological evaluations in child protection matters. *Am Psychol* 54:586–93, 1999
 82. American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of children and adolescents who are sexually abusive of others. *J Am Acad Child Adolesc Psychiatry* 38:55S–76S, 1999
 83. American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of children and adolescents with conduct disorder. *J Am Acad Child Adolesc Psychiatry* 36:122S–139S, 1997
 84. American Psychiatric Association: Practice guideline for the treatment of patients with acute stress disorder and posttraumatic stress disorder. *Am J Psychiatry* 161:1S–31S, 2004
 85. American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *J Am Acad Child Adolesc Psychiatry* 37:997–1001, 1998
 86. American Academy of Psychiatry and the Law: Practice guideline for forensic psychiatric evaluation of defendants raising the insanity defense. *J Am Acad Psychiatry Law* 30:S1–S40, 2002
 87. Anfang SA, Faulkner L, Fromson J, *et al*: American Psychiatric Association's Resource Document on Guidelines for Psychiatric Fitness-for-Duty Evaluations of Physicians. *J Am Acad Psychiatry Law* 33:85–8, 2005
 88. American Psychiatric Association: Practice guideline for the assessment and treatment of patients with suicidal behaviors. *Am J Psychiatry* 160:1S–60S, 2003
 89. Walter LC, Davidowitz NP, Heineken PA, *et al*: Pitfalls of converting practice guidelines into quality measures. *JAMA* 291:2466–70, 2004
 90. Shekelle PG, Ortiz E, Rhodes S, *et al*: Validity of the Agency for Healthcare Research and Quality Clinical Practice Guidelines: how quickly do guidelines become outdated? *JAMA* 286:1461–7, 2001
 91. Saneyfelt TM, Mayo-Smith MF, Rothwangl J: Are guidelines following guidelines? *JAMA* 281:1900–5, 1999
 92. Cabana MD, Rand CS, Powe NR: Why don't physicians follow clinical practice guidelines? *JAMA* 282:1458–65, 1999
 93. Aron R, Duffy KT, Rosner JL: *Cross-Examination of Witnesses: the Litigator's Puzzle*. New York: McGraw Hill, 1989 (1993 suppl)
 94. Brodsky SL: *Coping with Cross-Examination*. Washington, DC: American Psychological Association, 2004
 95. Krauss DA, Sales BD: The effects of clinical and scientific expert testimony on juror decision making in capital sentencing. *Psychol Public Policy Law* 7:267–310, 2001
 96. Heilbrun K, Warren J, Picarello K: Third party information in forensic assessment, in *Handbook of Psychology* (vol 11), *Forensic Psychology*. Edited by Goldstein AM. New York: John Wiley & Sons, 2003, pp 69–86
 97. Heilbrun K: *Principles of Forensic Mental Health Assessment*. New York: Kluwer Academic, 2001
 98. Simon RI, Wettstein RM: Toward the development of guidelines for the conduct of forensic psychiatric examinations. *J Am Acad Psychiatry Law* 25:17–30, 1997
 99. Heilbrun K, DeMatteo D, Marczyk G: Pragmatic psychology, forensic mental health assessment, and the case of Thomas Johnson. *Psychol Public Policy Law* 10:31–70, 2004
 100. Heilbrun K, Rogers R, Otto RK: Forensic assessment, in *Taking Psychology and Law into the Twenty-First Century*. Edited by Ogloff JRP. New York: Kluwer Academic Press, 2002, pp 119–46
 101. Grisso T: *Evaluating Competencies: Forensic Assessments and Instruments* (ed 2). New York: Kluwer Academic Publishers, 2003
 102. Tolman AO, Mullendore KB: Risk evaluations for the courts: is service quality a function of specialization? *Prof Psychol* 34:225–32, 2003
 103. Brennan TA, Horwitz RI, Duffy FD, *et al*: The role of physician specialty board certification status in the quality movement. *JAMA* 292:1038–43, 2004
 104. Otto RK, Heilbrun K: The practice of forensic psychology. *Am Psychol* 57:5–18, 2002
 105. Pitt SE, Spiers EM, Dietz PE, *et al*: Preserving the integrity of the interview: the value of videotape. *J Forensic Sci* 44:1287–91, 1999
 106. Farkas GM, DeLeon PH, Newman R: Sanity examiner certification: an evolving national agenda. *Prof Psychol* 28:73–6, 1997
 107. Fein RA, Appelbaum KL, Barnum R, *et al*: The designated forensic professional program: a state government-university partnership to improve forensic mental health services. *J Ment Health Admin* 18:223–30, 1991
 108. Massachusetts Department of Mental Health, Division of Forensic Mental Health: Massachusetts General Laws, Chapter 123 Section 15. Report Guidelines, and Quality Assurance Review Form for Inpatient Forensic Evaluations, 1996
 109. Hodgson DC, Zhang W, Zaslavsky AM: Relation of hospital volume to colostomy rates and survival for patients with rectal cancer. *J Natl Cancer Inst* 95:708–16, 2003

110. Peterson ED, Coombs LP, DeLong ER: Procedural volume as a marker for quality for CABG surgery. *JAMA* 291:195–201, 2004
111. Druss BG, Miller CL, Pincus HA, *et al*: The volume-quality relationship of mental health care: does practice make perfect? *Am J Psychiatry* 161:2282–6, 2004
112. Esserman L, Cowley H, Eberle C, *et al*: Improving the accuracy of mammography: volume and outcome relationships. *J Natl Cancer Inst* 94:369–75, 2002
113. Cunningham MD, Reidy TJ: A matter of life or death: special considerations and heightened practice standards in capital sentencing evaluations. *Behav Sci Law* 19:473–90, 2001
114. American Board of Psychiatry and Neurology: ABPN Update 11:1-7, winter 2005
115. Epstein RM, Hundert EM: Defining and assessing professional competence. *JAMA* 287:226–35, 2002
116. Landon BE, Normand S, Blumenthal D, *et al*: Physician clinical performance assessment. *JAMA* 290:1183–9, 2003
117. American Psychiatric Association Council on Psychiatry and Law: Peer review of psychiatric expert testimony. *Bull Am Acad Psychiatry Law* 20:343–52, 1992
118. American Psychiatric Association Task Force on Peer Review of Expert Testimony: American Psychiatric Association resource document on peer review of expert testimony. *J Am Acad Psychiatry Law* 25:359–73, 1997
119. Budd KS, Poindexter LM, Felix ED, *et al*: Clinical assessment of parents in child protection cases: an empirical analysis. *Law Hum Behav* 25:93–108, 2001
120. Gutheil TG, Simon RI: Narcissistic dimensions of expert witness practice. *J Am Acad Psychiatry Law* 33:55–8, 2005
121. Grisso T: The economic and scientific future of forensic psychological assessment. *Am Psychol* 42:831–9, 1987
122. Haas L: Competence and quality in the performance of forensic psychologists. *Ethics Behav* 3:251–66, 1993
123. Hanson RK, Morton KE, Harris AJ: Sexual offender recidivism risk. *Ann NY Acad Sci* 989:154–66, 2003
124. Garb HN: *Studying the Clinician: Judgment Research and Psychological Assessment*. Washington, DC: American Psychological Association, 1998
125. Garb HN, Schramke CJ: Judgment research and neuropsychological assessment: a narrative review and meta-analysis. *Psychol Bull* 120:140–53, 1996
126. Borum R, Otto R, Golding S: Improving clinical judgment and decision making in forensic evaluation. *J Psychiatry Law* 21:35–76, 1993
127. Garg AX, Adhikari N, McDonald H, *et al*: Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA* 293:1223–38, 2005
128. Garb HN: Computers will become increasingly important for psychological assessment: not that there's anything wrong with that. *Psychol Assess* 12:31–9, 2000
129. Grove WM, Zald DH, Lebow BS, *et al*: Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess* 12:19–30, 2000
130. Snyder DK: Computer-assisted judgment: defining strengths and liabilities. *Psychol Assess* 12:52–60, 2000
131. Stone AA, Turkkan JS, Bachrach CA, *et al* (editors): *The Science of Self-Report*. Mahway, NJ: Lawrence Erlbaum Associates, 2000
132. Borrell-Carrio F, Epstein RM: Preventing errors in clinical practice: a call for self-awareness. *Ann Fam Med* 2:310–16, 2004
133. Haber RN, Haber L: Experiencing, remembering and reporting events. *Psychol Public Policy Law* 6:1057–97, 2000
134. Mannuzza S, Klein RG, Klein DF: Accuracy of adult recall of childhood attention deficit hyperactivity disorder. *Am J Psychiatry* 159:1882–8, 2002
135. Gutheil TG, Simon RI: Avoiding bias in expert testimony. *Psychiatr Ann* 34:260–70, 2004
136. Gutheil TG, Simon RI: Attorneys' pressures on the expert witness: early warning signs of endangered honesty, objectivity, and fair compensation. *J Am Acad Psychiatry Law* 27:546–62, 1999
137. Commons ML, Miller PM, Gutheil TG: Expert witness perceptions of bias in experts. *J Am Acad Psychiatry Law* 32:70–5, 2004
138. Malsch M, Freckelton I: Expert bias and partisanship. *Psychol Public Policy Law* 11:42–61, 2005