

1993

Improving clinical judgment and decision making in forensic evaluation

Randy Borum, *University of South Florida*

Randy Otto, *University of South Florida*

Steve Golding, *University of Utah*

Psychiatry & Law

Improving clinical judgment and decision making in forensic evaluation

BY RANDY BORUM, PSY.D., RANDY OTTO, PH.D.,
AND STEPHEN GOLDING, PH.D.

Improving clinical judgment and decision making in forensic evaluation

BY RANDY BORUM, PSY.D., RANDY OTTO, PH.D.,
AND STEPHEN GOLDING, PH.D.

Psychologists, psychiatrists and other mental health professionals are frequently involved as expert witnesses in legal proceedings. However, clinical judgment and decision making, which play a role in almost all clinical evaluations, have problems and limitations. Mental health professionals who conduct forensic examinations should be aware of these problems and take steps to address them. This article details the limitations of clinical judgment and decision making, and suggests ways to minimize associated problems, thereby improving the validity and utility of forensic evaluations.

AUTHORS' NOTE: We gratefully acknowledge the valuable assistance of Dr. Hal Arkes and Dr. Thomas Grisso in the preparation of this manuscript.

and mental health professionals have criticized, on a variety of grounds, the scientific bases of expert testimony that these professionals offer.²

Among the most vocal critics are Faust and Ziskin,³ who offer numerous criticisms of expert testimony as it is currently delivered in the courts. Among them, they cite the limitations of clinical judgment as particularly problematic when considering introduction of expert mental health testimony into evidence.

An extensive experimental literature has developed detailing the problems that may occur when individuals, whether professionally trained or not, are faced with decision-making tasks.⁴

However, there is reason to believe that clinical decision making is not as problematic as some writers suggest. Much of the judgment research has been criticized because of limits on its external validity.⁵ The majority of studies that show poor performance on decision-making tasks have been conducted in laboratory situations and may not generalize well to clinical evaluation and practice. Even when clinicians have been used as subjects the experimental tasks are not analogous to those required in clinical practice. Critics of this research point out that individuals seem to perform well in natural settings, so that the "real life" consequences of any cognitive errors may be insignificant. Still other researchers have criticized the current studies and challenged the findings based on theoretical objections.⁶ Their own studies show how various "cognitive errors" disappear when framed and/or interpreted differently.

Christensen-Szalanski and Beach⁷ also argue that this literature is not as consistent as some reviewers suggest, and that there is a "citation bias" operating. They demonstrated that studies with positive findings about clinical judgment receive

less attention, and are less likely to be cited, than those with negative findings.

While some continue to debate the applicability of the experimental literature to applied clinical decision making, the extant literature nonetheless suggests that a variety of factors can operate to limit the accuracy of clinical judgment, regardless of its context. Accordingly, it is incumbent upon mental health professionals to be aware of such factors and to take steps to address them. This is particularly important when conducting forensic evaluations, given the nature of the questions and the legal and professional requirements that apply.

In this article we first describe and explain the biases and judgment errors that may occur in the context of forensic evaluations as well as in other clinical decision-making tasks more generally. Then we review specific techniques that forensic examiners can employ to minimize such biases, providing clinical examples to demonstrate how such biases may operate and how such corrective strategies can be applied.

Issue: Inaccuracy from overreliance on memory

Clinical example

You are a clinician hired by an attorney representing a father who wants custody of his eight-year-old daughter Stacy. Stacy's parents were divorced three years ago, and the mother was awarded sole custody. Both parents have subsequently remarried and have stepchildren living in their homes. The parents are now accusing each other of abuse and neglect. The attorney is convinced that this case will go to trial and, having a fair amount of experience with custody cases, requests that you complete a comprehensive and detailed custody evaluation, including a complete review of all relevant records and interviews with all relevant parties (e.g., parents, step-parents, step-siblings, other caretakers, etc.).

Background

Contrary to some common analogies, the brain does not function as a video recorder. It does not record and store every detail of events that we experience. We store event memories in a constructive and often nonsequential manner. In addition, interactive retrieval mechanisms preclude effortless recall of events as originally encoded.⁸ Therefore, an overreliance on memory in forensic evaluation can result not only in a decrease in information with which to make a judgment, but also in an increase in judgment bias.⁹ When asked to recall details of an observed clinical interaction sequence, it is much easier to recall information that is cued (perhaps by the questions of an attorney) or that confirms the hypothesis or conclusion being considered.¹⁰ For example, in a study by Arkes and Harkness¹¹ (1980), hearing and speech students were asked to recall symptoms of a patient with a particular diagnosis. The subjects were more likely to remember those features that were consistent with the diagnosis and found it more difficult to remember those features that were inconsistent with the diagnosis. In some cases the students recalled symptoms that were not even presented by the patients, apparently because they were "typical" features of the particular disorder.

Information that is more vivid or salient will also be remembered more readily than information that is bland or ordinary. This phenomenon is another memory bias that can affect clinical judgment and decision making.¹² Thus unusual features, striking events, or extraordinary circumstances may stand out in our memory and make more complete and accurate recall of more mundane but potentially relevant details less likely.

Overreliance on memory may also increase the effect of other judgment biases. When one must depend solely on memory, estimates of covariance¹³ are likely to be less accurate than when the actual information is available for review.¹⁴ The bias of illusory correlation¹⁵ may also become more pronounced when there is a greater amount of information to be processed.¹⁶

Corrective measure

Provide the best documentation possible

Given the potential for such memory biases, clinicians should take steps to document as accurately as possible all information that they gather, whether based upon interviews, test data, or collateral data from third parties. Relying on careful documentation rather than distant memories allows for more accurate recall of events in the examination and reduces the likelihood of memory-based errors.¹⁷ Clinicians should develop a system of note-keeping in which they can distinguish between observations and inferences. For example, clinicians should document differences between claims made by interviewees (e.g., "I am depressed") and interpretations or opinions about interviewees ("She appeared depressed"). Similarly, clinicians should distinguish among observations (e.g., "she cried throughout the interview"), intermediate interpretations ("she sobbed throughout the interview"), and interpretations involving even greater inference ("she was depressed").

Clinicians conducting forensic evaluations should consider the advantages of documenting their interviews by means of audio- or videotaping in addition to copious note-taking. The use of such measures in the forensic context is further supported by "Specialty Guidelines for Forensic Psychologists,"¹⁸ promulgated by the American Psychological Association (Division 41 of the American Psychological Association) and the American Academy of Forensic Psychology. The guidelines direct that when conducting a forensic examination, the psychologist incurs "a special responsibility to provide the best documentation possible under the circumstances."¹⁹ This obligation applies "from the moment they know or have a reasonable basis for knowing that their data and evidence derived from it are likely to enter into legally relevant decisions."²⁰ A similar prescription is included in the recently revised "Ethical Principles of Psychologists and Code of Conduct":²¹ "When psychologists have reason to believe that records of their professional services will be used in legal proceedings . . . they have a responsibility to create and maintain documentation in the kind of detail and quality

that would be consistent with reasonable scrutiny in an adjudicative forum."²²

Some suggest that the "best documentation" is a videotaped or audiotaped recording of the evaluation, while others contend that careful notes are sufficient to meet this standard. Regardless of the type of documentation selected, the standard is clearly higher than that required for general clinical practice.

In the case example, the evaluator may be overwhelmed by the amount of information that he or she is required to gather and interpret. The evaluation requires contact with a large number of people and involves numerous but differing accounts of the same or similar events and situations. Such a large amount of information increases the potential that the examiner may forget or confuse issues or may selectively remember data as a function of saliency, the point in time at which they were delivered, etc. Accordingly, careful documentation of the various contacts and evaluations would minimize the biases of memory in this case. Videotaped or audiotaped interviews, separate files for different interviewees, and careful and precise written documentation will minimize the potential for such memory biases as well as preserve evidence for adequate cross-examination.²³

Issue: potential limitations in complex configurational analysis and the utility of additional information

Clinical example

You are a clinician retained by the attorney for the plaintiff in a personal injury case. The plaintiff, an adult who was sexually abused by a priest at the age of seven, alleges that he suffers a variety of psychological impairments as the result of this abuse. The plaintiff has been in outpatient therapy for a number of years and has been hospitalized for depression and alcohol abuse on three occasions. In addition to the psycho-

Background

logical testing and interview material that you have collected, the plaintiff has undergone three extensive psychological evaluations in the past and has also been evaluated by the defense expert. Also available for review are all prior records of inpatient and outpatient treatment. The plaintiff's attorney asks that you conduct your own evaluation and consider results of all previous evaluations in doing so.

When clinicians who testify in court are asked how they came to a particular opinion, they frequently offer "My opinion is based on an integration of all of the data I collected" or some equivalent response. One aspect of data integration, referred to as "configural analysis," involves an examination of inter-actional relationships among various scores and data points. This interpretive strategy has been advocated by prominent scholars and practitioners in psychological and neuropsychological assessment.²⁴ Some, however, have questioned whether the decision-making processes employed by clinicians in those situations where test scores are quantified are configural or are better understood as less complex, linear decision making.²⁵

A number of investigators have examined the ability of human judges to perform configural analysis.²⁶ Investigators have been able to replicate judgments made by clinicians using simple linear equations involving only a few variables,²⁷ suggesting that data integration and corresponding judgments may not be as complicated as proponents of configural analysis claim. However, because "linear models can sometimes duplicate judgments performed via configural strategies,"²⁸ these findings do not conclusively prove that clinicians employ more basic linear methods in their decision making. Some have suggested that use of a variable-weighting system results in little if any improvement over any equal-weighting system.²⁹

In general, research has demonstrated the ability of judges to perform basic configurational analyses,³⁰ but the ability to perform the type of integrative analysis typically described by clinicians has not been systematically investigated. Most of the research that does exist in this area is highly artificial and does not generalize well to the clinical situation.³¹ Therefore, the capacity of individuals to perform complex configurational analysis has not yet been empirically demonstrated, and it is even possible that such analysis adds very little to more simple linear approaches.³²

A related issue is the ability of clinicians to manage and integrate additional information providing incremental validity. There may be limitations on the amount of information that can be managed and utilized by clinicians in their clinical decision making. Some early studies indicated that additional information does not significantly improve judgment accuracy once an individual has several pieces of valid information.³³

Currently, there is a body of literature indicating that accuracy of judgment does improve somewhat when additional information of proven validity is provided.³⁴ However, including data of questionable validity or data that are less valid than those already possessed may not increase, and can sometimes decrease, accuracy of judgments.³⁵

Corrective measures
Consider the issue of incremental validity in data gathering

The literature reviewed above suggests that merely adding extra information, especially if it is of questionable validity, does not increase, and may decrease, judgment accuracy. Thus, more information or data is not necessarily better. The clinician should consider the issue of incremental validity when deciding whether or not to collect any particular type of data (i.e., "Will this information assist me in developing a more valid opinion?"). Prior to gathering data, the clinician should take steps to ensure that he/she has a good understand-

ing of the relevant psycholegal questions via consultation with the referral source. Once the questions are well defined, the evaluator should determine the techniques, tests, or approaches that are most valid in providing information relevant to the question(s) at hand. Such an approach should result in the evaluator's identifying a valid and efficient means of conducting the forensic evaluation.

In the case above, there is an extensive body of available information. Prior to meeting with the examinee, the clinician should contact the attorney and identify specific questions that the attorney wants addressed (e.g., "What is the examinee's current level of psychological distress, and what are the possible causes or sources of this distress? How might the examinee's abilities with respect to forming relationships with others be affected as the result of his abuse? How likely is it that the examinee is either minimizing or exaggerating his difficulties? What are his abilities with respect to obtaining and maintaining employment? What kind of treatment is recommended at the current time, and what is the prognosis for better adjustment through treatment?").

Having the specific questions in hand, the clinician should be able to identify those techniques that are most helpful. A standardized measure of adjustment/psychopathology is indicated (e.g., MMPI-2, MCMI-II), and, given the constant concern about accuracy and veracity of self-report in forensic and other clinical situations,³⁶ using instruments with validity measures or techniques specifically designed to assess malingering and deception should also be considered. Numerous third-party contacts are also indicated to best develop a picture of the examinee's adjustment across different situations and further address the issue of dissimulation. On the other hand, assessment of intellectual functioning does not seem indicated, given the referral questions, and would appear to add little to the relevant clinical picture.

Identify and weight the most valid sources of data

When developing clinical hypotheses, it is helpful to list separately the most valid sources of data, along with the inferences derived from them, and then to determine what conclusion might be reached just on the basis of these data. In the above example, the evaluator would generate all hypotheses that are supported by the MMPI-2 profile, all hypotheses supported by interviews with various third parties, all hypotheses supported by results of prior testing and assessments, etc. Then all hypotheses generated should be evaluated for the strength of support provided by the various sources. A conclusion can then be made on the basis of those hypotheses that receive the strongest support. The most valid information should ultimately be most heavily weighed in the decision, so further consideration or integration of the rest of the data typically should not substantially alter this opinion.

Issue: underutilization of base rates

Clinical example

In your work at a rural community mental health center, you are asked to evaluate the suicide potential of a 34-year-old separated black female, Diana, who came to the outpatient center in crisis; her husband moved in with another woman three days ago and announced his intention to divorce. Diana reports getting little sleep over the past three days, crying frequently, isolating herself, and eating little. She denies using drugs or alcohol. Diana reports thinking about killing herself, but she does not have a plan; she identifies her strong beliefs in the Roman Catholic faith as an obstacle to hurting herself. Diana requests medication and outpatient therapy, but she is unwilling to admit herself to the hospital because of "what others might think."

Background

"Base rate" refers to the prevalence of a given characteristic in a population.³⁷ In the prediction literature, it is sometimes referred to as prior probability or antecedent probability. When making predictions or classifications, the applied value

of a diagnostic sign or psychometric score is a function of the overall correct classification rate of the sign/instrument (including separate analyses of false positive and false negative rates) and the base rate (i.e., prevalence) of the characteristic of interest in the population to which the sign/instrument is being applied.³⁸ Infrequently occurring conditions are by their very nature difficult to predict. However, low-frequency events are often the very issues that forensic clinicians are called upon to address (e.g., violence risk, suicide potential).

Although prior probability does affect predictive accuracy,³⁹ research suggests that individuals often do not sufficiently attend to this issue when making judgments.⁴⁰ Base rate information is more likely to be used under certain conditions, such as when a causal relationship to the specific case is provided,⁴¹ when predictions are dichotomous rather than probabilistic,⁴² or when the case is considered to be randomly drawn from the population and no individualized information is made available.⁴³ Failure to consider base rates when making judgments and predictions about low base rate behaviors (even when very accurate prediction techniques are used) may result in overprediction of the behavior of interest (i.e., a high false positive rate).

Corrective measures

Identify relevant base rates

When making judgments or predictions about an individual case, the first step is to identify the population to which the predictee belongs and then attempt to identify the base rates of the behavior of interest within that population. Whenever possible, the forensic clinician should use those norms that are most appropriate for the individual case. Variables such as age, gender, educational level, or social factors may affect applicability of base rates, depending on the behavior of interest. For example, the prevalence of suicide attempts among young, white, male jail detainees is different from that of middle-aged African-American females in the community.

Although base rate data are often unavailable or difficult to find, psychologists and other mental health professionals have begun to compile such information. For example, Otto⁴⁴ and Swanson and his colleagues⁴⁵ provide data relevant to the base rates of violence against others in a variety of relevant populations. Maris et al.⁴⁶ provide information regarding the base rates of suicide among various populations of interest. Heaton, Grant, and Matthews⁴⁷ have provided some updated and comprehensive normative data for the Halstead-Reitan and several supplemental tests used in neuropsychological assessment. In those cases where normative data are not as strong (e.g., sample size is small, sample is not representative of population, data are not provided for specific, presumably relevant factors such as gender, age or education), one should be cautious about the interpretation and consider the effects of these variables as potential alternative explanations for the findings. Absent such specific base rate information, the clinician should consider base rates of similar behaviors or base rates of the behavior of interest in other, similar populations.

The evaluator should use the base rate to set the starting point for subsequent evaluation of probability. That is, higher frequency events will be seen as more likely, and low frequency events will be seen as less likely. With the base rates as a starting point, the examiner can then consider assessment data and (cautiously) modify these rates accordingly to make a judgment about the individual case.

In the case example, it should first be considered that the base rate of suicide among black middle-aged females is quite low.⁴⁸ From this primary assumption of low likelihood, however, the evaluator should consider modifying estimates based on particular facts of the case that are empirically demonstrated to increase or decrease suicide risk. Factors that might increase Diana's suicide risk include her depression and suicidal ideation. There are a number of factors, however, that decrease her risk, including the lack of such behavior on her part in the past, the absence of drug or alcohol

abuse, and her strong religious beliefs proscribing suicide. In this case, while there are no base rates available for suicide among middle-aged, non-drug-abusing, devoutly Catholic, recently separated African-American females living in rural areas, basic base rates can be used as a starting point from which to tailor predictions.⁴⁹

Issue: confirmatory bias

Clinical example

You have been retained by the state attorney to evaluate a 29-year-old male recently convicted of the capital murder of a police officer. The state is seeking the death penalty, and the defense has indicated that it plans to introduce the testimony of a psychiatrist regarding mitigating factors. The defendant had two prior convictions for drug offenses and two prior convictions for unlawful possession of firearms. None of the convictions involved acts of violence. The state requests that you evaluate the defendant and offer an opinion regarding the probability that the defendant will commit criminal acts of violence and continue to pose a threat to society.

Background

Confirmatory bias refers to a tendency to look for evidence that supports one's hypothesis (what one is expecting or hoping to find) and to ignore, or fail to seek, information that is not consistent with that hypothesis.⁵⁰ This bias could also extend into the interpretation phase, where one may interpret the same piece of data in the way that supports one's perceptions/preconceptions when either of two interpretations is equally possible. A related bias, the "anchoring" or "primacy" effect, occurs when information gathered earlier in the evaluation process is weighted heavily and there is insufficient modification of these initial impressions even upon presentation of contradictory information.⁵¹ It has been suggested that simply having to consider a specific cause increases the perceived probability of that cause.⁵² Thus it seems easier to create a first impression than to change one.⁵³

Such biases can affect forensic evaluations because the initial impression of a defendant often comes from the defendant's attorney or others who may not be objective. For example, with this sentencing case, the prosecutor may have called you and said something like "Doctor, I've got a pretty clear-cut case here. This is a defendant who viciously murdered a cop. He's the kind of guy who would probably do it again given half a chance—so we're going for the death penalty. We'd like you to evaluate him and give testimony about your opinion of his future risk for violence. Let me know what you think."

Corrective measures
Search for and list disconfirmatory information

One way to counter the confirmatory bias during the data-gathering process is to specifically search for disconfirmatory information for the initial hypothesis. Faust⁵⁴ suggests looking specifically for exclusionary criteria in addition to exploring the more common features/signs that may be associated with several disorders or conditions. Disconfirming information should be used to modify the initial impressions as appropriate. During the interpretation stage and when forming conclusions, the disconfirmatory information should be listed and considered with the other relevant data. The identification of competing hypotheses may also help by broadening the scope of inquiry and evaluation.

In the case above, the evaluator must be aware of and fight the tendency to identify and overemphasize factors in the defendant's past that confirm his or her tendency to believe the defendant is "dangerous" and to ignore or minimize factors suggesting that the defendant presents less of a risk.⁵⁵ Perhaps the evaluator will focus on the defendant's two prior convictions and his relatively young age and cite these factors as confirming the potential for violence, while ignoring other relevant factors such as the apparent lack of violent behavior prior to the offense for which he was convicted (the two prior convictions were for nonviolent offenses).⁵⁶

Consider the tendency to underrevise initial hypotheses

As a method of combating such biases, the evaluator should generate and list possible hypotheses.⁵⁷ Disconfirming or contradicting evidence for all hypotheses should be specifically listed and considered, especially for those hypotheses to which the evaluator is "drawn."⁵⁸

As the forensic clinician encounters information that does not support his or her initial impressions, it is necessary to consider whether such data warrant a modification. In making this decision, one should keep in mind that there is a tendency to underrevise initial hypotheses. More specifically, there is a tendency to not revise an opinion based on disconfirmatory evidence, and, when a revision is made, it is often not revised substantially enough, given the subsequent information.⁵⁹ As applied to the case example, in reevaluating clinical hypotheses the evaluator needs to be aware of the general tendency to be drawn to initially established hypotheses. Does the evaluator find himself or herself explaining away data that do not support the initial hypothesis that the client poses some kind of increased risk for violent behavior (e.g., perhaps assuming that although the defendant was only convicted of two prior offenses, he probably committed a number of violent offenses for which he was not caught; or noting that although the two prior convictions were for nonviolent offenses, the defendant presumably was willing to engage in violent behavior in the course of violating the law)?

Issue: misestimation of covariation

Clinical example

You have been retained by the state in a hearing for termination of parental rights. Nancy is a three-year-old girl recently taken from the custody of her mother and stepfather by child protective workers. Nancy's father alleged that Nancy was being sexually abused by her stepfather and that her mother knew of the abuse and failed to protect her. The child protective investigators suspect that sexual abuse has occurred, given the number and variety of behaviors and symptoms

they have documented: sleep difficulties, nightmares, decreased appetite, bed-wetting, shyness, inexplicable crying, and apparent masturbation. The attorney representing the state intends to remove Nancy from her mother's home for an extended period of time, and he requests that you evaluate Nancy and offer an opinion with respect to whether Nancy may have been sexually abused.

Background

There are two major issues related to miscalculation of covariation. Most important to note is the difficulty establishing whether or not there is a *true* relationship between two or more factors or variables, such as the relationship between a symptom and a disorder. The following figure shows the four conditions relevant to assessing covariation. Cases in Cell A are those in which both the symptom and the disorder are present (cases that would be correctly diagnosed as having the disorder based on the presence of the symptom). Cases in Cell B are those in which the symptom is present but the disorder is not (cases that would be incorrectly "diagnosed" if concluded to have the disorder based on the presence of the symptom). Cases in Cell C are those in which the symptom is not present but the disorder is (cases that would be incorrectly diagnosed as not having the disorder based on the absence of the symptom). Cases in Cell D are those in which neither the symptom nor the disorder is present (cases that would correctly be diagnosed as not having the disorder based on the absence of the symptom).

The distribution of the population of cases across all four cells is important when drawing conclusions regarding covariation (and the predictive utility of a particular symptom or sign). However, research indicates that individuals typically draw such conclusions based mainly on selective recall of case representations in Cell A (when both symptom and disorder are present) and pay less attention to Cells B, C, and D.⁵⁰

Relationship between symptom/sign and disorder

	Disorder/ Condition Present	Disorder/ Condition Absent
Sign/Symptom Present	A	B
Sign/Symptom Absent	C	D

A related issue in estimating covariation is the phenomenon of "illusory correlation." Chapman defined illusory correlation as "the report by observers of a correlation between two classes of events which, in reality, (a) are not correlated, (b) are correlated to a lesser extent than reported, or (c) are correlated in the opposite direction from that which is reported."⁵¹

These erroneous associations may be based on verbal associative connections (e.g., the belief that large eyes in human figure drawings indicate suspiciousness),⁵² or they may result from reliance on the "degree of representativeness (that is, on the quality of the match between the selected outcome and the input) with little or no regard for the factors that limit predictive accuracy."⁵³ An illusory correlation may also result from other errors in the assessment of covariation mentioned above, such as noticing only when an association is present (Cell A) and not noticing when it is absent (Cells B & C).⁵⁴ Regardless of their origin, these associations may be passed along and perpetuated across generations of clinicians as a part of "clinical lore."⁵⁵

Corrective measures
Assess all aspects of covariation

Many errors in estimating covariation come from an incomplete assessment of the association between the sign (score) and the outcome. Investigators sometimes fail to report the information from all four cells in research publications. In order to accurately assess the predictive utility of a particular sign or symptom, information from all four cells is necessary. The clinician should be cautious in making inferences about the predictive utility of a particular sign or symptom when information from all four cells is unknown.

In order to draw conclusions about the diagnostic utility of the symptoms presented above, the clinician must have a handle not only on the co-occurrence of sexual abuse and bedwetting, nightmares, and academic difficulties, but also on the occurrence of such symptoms in the absence of sexual abuse. Only when the occurrence of these symptoms is greater in the presence of sexual abuse than in its absence do they have any diagnostic potential.⁶⁶ Clearly, clinicians cannot gather these data in routine clinical practice. They should, however, read the existing empirical literature with an eye toward the "missing cells."

Rely on empirically established relationships

Illusory correlations seem to arise from a reliance on verbal associations or degree of representativeness. In essence, this type of error occurs when individuals form intuitive connections for events that "seem to go together" or relationships that "seem to make sense." Thus, the appropriate measure to correct for this difficulty is to rely on empirically established relationships and to treat relationships that have been "established" in any other way (e.g., "in my extensive clinical experience") very skeptically. Reliance solely on clinically perceived relationships increases the opportunity for these types of errors to occur.

In this case the evaluator must rely on empirically established indicators of sexual abuse to the extent that they exist. Familiarity with the most recent and comprehensive literature regarding sexual abuse profiles would lead most clinicians to

Issue: hindsight bias

Clinical example

proceed very carefully in this case and limit their conclusions accordingly.⁶⁷ The mental health professional must be particularly careful not to rely on and fall victim to illusory correlations that he or she may have developed from clinical lore or as a function of exposure to a select sample of sexually abused children seen in clinical practice.

You have been retained by the plaintiff's attorney in a negligent release case. Dr. Wilson, a psychiatrist, admitted Mr. Hannigan to a hospital for severe depression and suicidal ideation in response to a suicide attempt. After a 13-day stay, Dr. Wilson discharged Mr. Hannigan despite Mrs. Hannigan's warnings that her husband might try to kill himself again, just as he did immediately prior to his hospitalization. Mr. Hannigan subsequently shot and killed himself the day after being discharged. Mrs. Hannigan is suing Dr. Wilson alleging malpractice/negligent release. Her attorney requests that you review the record and comment upon whether Dr. Wilson's treatment and discharge planning met the community's requisite standard(s) of care.

Background

Forensic clinicians are sometimes asked (typically in malpractice cases) to assess the appropriateness of professional practice in light of subsequent (typically negative) outcomes or to assess the foreseeability (at some prior point in time) of an event that has already occurred. In these instances there is a potential for hindsight bias to affect the clinical judgments that are offered. Hindsight bias "refers to the fact that after an event has taken place, its occurrence seems so inevitable that one believes that it could have easily been predicted in advance."⁶⁸ Judgments regarding the appropriateness of particular behaviors that may or may not have taken place can be affected as a consequence of knowledge of the outcome. The phenomenon of hindsight bias, sometimes referred to as the

"knew-it-all-along effect"⁶⁶ or "Monday morning quarterback," has been documented in numerous studies.⁶⁶

Corrective measures
List supporting evidence for alternative conclusions

Arkes et al.⁷¹ proposed one technique for reducing hindsight bias the has met with some success. In an experimental analogue, Arkes and his colleagues provided neuropsychologists with a case history and three possible diagnoses. They created a foresight condition (the case history did not include a primary diagnosis) and a hindsight condition (the case history did provide a primary diagnosis). After reading the case history, half of the subjects in both conditions rated the probability that each of the diagnoses was the *primary* one. The remaining subjects were provided with the following instructions before rating the probability of the diagnoses: "After *each* of the following diagnoses, please jot down one piece of evidence from the case history that would support that particular diagnosis as the primary one."⁷² Arkes and his colleagues found that requiring clinicians to generate and list reasons for alternative diagnoses significantly reduced the impact of hindsight bias—that is, it reduced the subjects' overconfidence in their postdicted choices.

Such an approach is easily applied to clinical practice. When assessing the probability of an event that has already occurred, generate potential alternative outcomes and then interpret and review the data with an eye toward each of these outcomes, listing supportive evidence and reasons for each.

In the case example, the clinician should list alternative outcomes or scenarios that could have occurred in response to the treating psychiatrist's actions (identifying and listing supporting evidence for each of them), and then identify alternative courses of action and possible outcomes.

Evaluate probability in absence of outcome information

Since by definition the hindsight bias involves a postdictive estimation of probability when the actual outcome is already known, one possible solution is to make an estimation before being exposed to the outcome. Poythress⁷³ has formally proposed this strategy for evaluating negligent release claims such as that in the case example. This strategy requires some intricate preplanning and precautions with respect to contacts with the referral source to prevent inadvertent exposure of outcome information prior to rendering an opinion about the appropriateness of the practice that is alleged to constitute malpractice.

Adopting this approach in clinical practice, and the case above, is difficult and unwieldy, although not impossible. If nothing else, it would be interesting to present the case facts, along with various possible outcomes, to colleagues ignorant about the actual case and outcome and then ask their opinions about the likelihood of those outcomes. This provides a way of permitting evaluation that is not tainted by hindsight bias.⁷⁴

Issue: overconfidence

Clinical example

A public defender asks you to evaluate a 30-year-old homeless male for criminal responsibility. He is charged with aggravated battery in an incident that occurred 18 months ago. At that time, the defendant allegedly cut a shopping plaza security guard who woke him by "poking" him with a nightstick. On examination, he is disheveled and paranoid. He is able to give a lucid account of the events of the assault, although he is a rather poor historian for other background information. He does admit to hearing voices in the past and currently has persecutory delusions. He reports being hospitalized once but does not remember many details of the hospitalization. The victim is unavailable, the police report is sparse, and there are no medical records or third-party informants who had contact with the defendant on or around the date of the incident.

Background

A concern has been raised in the psychological literature that individuals express more confidence in their judgments than is actually warranted.⁷⁵ Although much of this research comes from the social psychology literature, there has also been concern about overconfidence on the part of clinicians.⁷⁶ Several studies have investigated this issue using clinical experts.⁷⁷ Generally, these results also suggest a trend toward overconfidence.

However, Garb,⁷⁸ in a comprehensive review of the literature, identified several more specific patterns. These include a trend in which use of more valid information produces more appropriate confidence ratings, and a tendency for more experienced clinicians to make more appropriate confidence ratings. Garb concluded that "clinicians may be overconfident for only a few sources of information or for only a few types of judgment tasks"⁷⁹ (p. 194). In addition, Lindsay⁸⁰ reported that the correlations found between confidence and accuracy in empirical studies are significantly influenced by the design of the research.

Nonetheless, overconfidence is an important issue in forensic assessment for two reasons. First, an expert witness's level of confidence can affect the fact finder's reliance on the testimony he or she offers. Second, overconfidence may affect clinical judgment by causing the clinician to disregard, or at least not to search for, disconfirmatory information.⁸¹

Corrective measures

Two of the most effective techniques for limiting overconfidence have already been discussed in detail under previous headings. Clinicians should (1) identify the competing/alternative explanations (and list supporting evidence for each) and (2) search for and list information that would disconfirm the hypothesis being considered.⁸² Two other measures are offered below.

Vary level of confidence according to validity of data sources

One cause of overconfidence in judgment is an erroneous belief in the accuracy of unreliable or invalid sources of data. However, the validity of an opinion is only as strong as its foundation. Therefore, having access to more valid information may produce a more valid opinion, and the clinician's level of confidence may be higher. For an opinion that is based on sparse or unreliable information (as might occur when more valid data are unavailable), the clinician should adjust the confidence rating downward (assigning less confidence).

In the case example, almost all information about the incident comes from the defendant, who has the potential for secondary gain from distorting information and who may actually have difficulty providing an accurate account if his current or past thought processes interfere with his processing or recall of information. Simply stated, the defendant's account alone may not provide a very strong or valid basis for a conclusion about his mental state at the time of the offense. Most of the information that would increase confidence in this foundation is unavailable. Consequently, the evaluator should be more cautious about his/her level of confidence in any opinion based solely on the account of the defendant, while examiners can be more confident of opinions based on multiple sources of information.

Gather follow-up information on accuracy

To improve calibration of confidence levels, it is helpful to gather some follow-up information on judgments and predictions. This can be problematic, however, since concrete feedback or appropriate criteria may be difficult to obtain. For example, if a defendant is found incompetent to proceed, then it is not possible to observe his or her actual behavior during the trial that would have occurred. Still, it is possible to get some idea of how one's judgments correspond to various criteria and to a confidence level—for example, one can obtain feedback about the utility of evaluations from the referral source (although this is helpful, it is not always objective and does not necessarily reflect accuracy) and acquire informa-

tion about a patient/defendant's behavior subsequent to an evaluation. Knowledge of accuracy (validity) and systematic feedback should help to calibrate confidence.

Issue: overreliance on unique data

Clinical example

The public defender has retained you to evaluate a 43-year-old white male charged with the first degree murder of his uncle. The public defender plans to raise the insanity defense, alleging that the defendant suffered from post-traumatic stress disorder as a result of his extensive Vietnam combat experience. The defendant claims that he has a history of explosive rage episodes since his return from Vietnam. The defendant reports that he had been drinking when he went to his uncle's house to talk to his wife, who had separated from him two weeks previously. His uncle met him at the door and refused him entrance. The defendant has a memory only for a fistfight starting—the rest “is a blur.” His wife reports that he took a knife from his jacket and stabbed his uncle 43 times.

Background

Working in the forensic arena provides an opportunity to evaluate some interesting cases with very unusual features. However, there is a tendency to give excessive attention to specific symptoms or features that are exotic, interesting, or highly unusual. This tendency can produce two problems in clinical judgment. First, it may lead one to believe that the case itself is unique and that all information based on common features therefore does not apply. Second, it may lead one to assume that such unusual signs have greater predictive validity than familiar ones.³³

Corrective measures

Search for features that are common to other cases

When a case seems to be highly unusual or “unique,” it is helpful to search for those features of the case that are familiar or similar to other cases. The salience of these exotic elements should be modified and judgments balanced against the weight of all the common features.³⁴ A listing of these common features is likely to show that the case is more similar than different; more common than unique. Typically, the decision should be based primarily on the common, routine factors in the case, with only careful modification for those elements that are seen as unique.

In the case example, the clinician should avoid jumping to conclusions regarding the defendant's criminal responsibility, based on his claims about combat experience and PTSD. First, the question always remains about the examinee's veracity in his accounts of his combat experience.³⁵ Assuming that the combat claim is true, the examining clinician should be careful not to assume that the combat experience necessarily resulted in PTSD. The majority of persons experiencing combat do not suffer from post-traumatic stress disorder;³⁶ thus combat experience itself (a highly salient piece of information) cannot be taken to indicate the presence of PTSD. Finally, even if the diagnosis of PTSD is established (a highly salient factor), its presence alone cannot be interpreted as necessarily affecting the behavior of the defendant in such a way that would exculpate him (ultimately to be decided by the judge or jury, in any case). Rather, after the combat experience and the presence of post-traumatic stress disorder have been established by the clinician, he or she must continue to investigate and probe to determine the specific effects that the disorder has on the defendant's adjustment and behavior.

In addition, the careful reader may note that the defendant stabbed his uncle a number of times exactly equal to his own age. One could base an opinion on this single piece of data and offer the following testimony: “In my opinion, the man did not act in an uncontrolled explosive rage. The man, 43 years of age, stabbed his victim 43 times. The odds that, by

chance, he would stab the victim the same number of times as his age are infinitesimal. The only other plausible explanation is that he was counting and stopped when he reached 43. If he could count, and stop at 43, he could have stopped at zero or one.⁷ No matter how many times the defendant stabbed the victim, one could probably find some significance of that number in the defendant's life.⁸ In searching for common features of this case, however, one would inevitably find that they far exceed the exotic ones.

*Be cautious
about
overriding
established
decision rules*

Again, the dangers of overfocusing on unusual or unique features of a case are that those features may be seen as better predictors and may create the impression that no other common information would apply. As a result, one may be tempted to assume that actuarial formulae or useful decision rules do not apply in such a case. In those cases where useful decision rules are available, the research generally suggests that they can perform better than clinical judgments.⁹ Arkes, Daves, and Christensen¹⁰ found that as individuals were given more motivation to perform better on judgment tasks, they increasingly disregarded the actuarial information and made an alternate decision. Consequently, their performance was also worse. Therefore, one should be cautious about overriding useful decision rules unless there is compelling evidence to suggest that the rule does not apply.

Issue: confusing fact and statistical artifact

**Clinical
example**

Katie, a 19-year-old female, was involved in a motor vehicle accident in which she sustained a closed head injury. You have been asked by her attorney to evaluate her for her personal injury claim. In reviewing her past records, you find that she was tested in the school system a year prior to the accident and obtained a Full Scale WAIS-R IQ of 135. On your evaluation she obtains a WAIS-R IQ of 124. The attorney wants to use these data as evidence of significant cognitive (neuropsychological) damage.

Background

Although many forensic clinicians have been well trained in statistical methodology and concepts, they often fail to apply them or to recognize their effects in clinical practice. The concepts of regression to the mean and sampling bias are two such frequently neglected concepts.¹¹ Regression to the mean predicts that extreme events will be followed by less extreme events. Accordingly, we would expect an extreme behavior or test score, when sampled again, to be somewhat less extreme. This possibility should be considered when developing causal explanations for extreme events, especially when evaluating changes across subsequent evaluations. At least two types of sampling problems are also overlooked in day-to-day practice: insensitivity to sample bias, and insensitivity to small sample size. As a result, clinicians may make inappropriate generalizations about a class of individuals based on their experience with some subgroup of that population (e.g., defendants who are detained in an inpatient forensic facility). They may also fail to recognize that small samples are not as representative as large samples and are more prone to chance variation. This may result in such errors as prematurely abandoning some assessment technique because it did not seem to work well the one or two times it was used, or overgeneralizing about an individual from a small sample of interview behavior.¹²

**Corrective
measures
Consider
regression
effects**

Regression effects occur in a variety of natural phenomena and clinical situations. However, clinicians often fail to consider the effects of regression, or they develop "spurious" substantive explanations for such effects when they are observed.¹³ The issue may arise (1) when one is making a prediction of a future event following an extreme event or (2) when assessing the cause of a present state that has followed an extreme event. The clinical case provides an example of the second scenario.

In this scenario, Katie's pre-accident WAIS-R score is 135 (99th percentile) and her post-accident WAIS-R score is 124 (95th percentile). Assuming that her obtained score actually

represents the sum of a "true score" and a component of random error, and given that "the observed score (on her first testing) is considerably higher than the population mean, it is more likely that the error component is positive and that (she) will obtain a somewhat lower score on subsequent tests."⁹³ Therefore, to appropriately interpret the change between Katie's two scores, one would need to (1) consider that the obtained score is likely to contain a component of random error, (2) consider that a somewhat lower score on retesting would have been predicted based solely on effects of regression, and (3) be careful not to develop spurious explanations for regressive phenomena. As Wedding and Faust point out, "Regression towards the mean is more likely to be recognized if a clinician habitually asks the question: 'Is the change I have observed partially or wholly predictable on the basis of regression towards the mean?' This simple question should lead one to undertake the calculations needed to provide the answer."⁹⁴

*Consider
representativeness
of small (or
biased) samples*

Again, the corrective strategies are relatively straightforward. Before making any generalizations or drawing inferences based on collected or experiential data (i.e., "norms" developed as a function of clinical experience), the clinician should consider at least two factors: (1) whether the sample is large enough to represent the parameters of the population and consequently to support reliable and valid inferences; and (2) whether the sample is similar to or representative of the case under consideration. Because of this latter principle, clinicians should seek and use the most appropriate normative data possible. It also necessitates that clinicians exercise extreme caution in the use of "experiential norms" based on nonsystematic clinical observation and experience. Such "data" have a high probability of being biased and not representative of population parameters. For example, a clinician who practices in a forensic setting could erroneously conclude that the population of mentally ill are quite violent simply because many of the mentally ill people seen in his or her daily practice seem to have a history of violence. This type of

*Be aware of
problems
associated with
use of highly
correlated
measures*

conclusion ignores the bias that the violence may have been what selected that subgroup of mentally ill individuals to be admitted to the facility.

Another issue related to separation of fact and statistical artifact is the use of measures that are highly correlated with one another. There are two primary situations in which this may arise: (1) when test results are unclear and additional instruments are administered for clarification; and (2) when multiple instruments are administered and one is assessing consistency with regard to the criterion.

When results of initial assessment or testing are unclear, clinicians sometimes select a very similar (and correlated) measure to assist in interpretation. The potential problem with this practice is that to the extent that the instruments are correlated, the similarity in results may represent redundancy in measurement rather than incremental information about the examinee's ability or condition. Ideally, the second measure would be positively correlated with the criterion but negatively correlated with the first instrument.⁹⁵

When highly correlated multiple measures are administered, one must again consider the extent to which the consistency of test results reflects the correlation between the measures rather than convergence upon the criterion. Information about interest correlations can typically be found in the test manuals, since concurrent validity data often include comparisons with other commonly used instruments. For example, suppose a clinician was assessing for mental damages in a personal injury claim and used the Beck Depression Inventory and the Hamilton Rating Scale for Depression. Whether used in combination for convergence or consecutively for clarification, the clinician should account for the high correlation between these instruments when making inferences about their combined value.

To address this issue, the following suggestions are provided: (1) be cautious about relying too heavily on highly correlated measures for purposes of clarifying equivocal results; (2) when considering the consistency of test results, consider the possibility that measurement redundancy may explain some of the convergence.

Summary and conclusions

Given the nature of legal proceedings and the stakes involved, mental health professionals who conduct forensic evaluations should rely on the most valid information in reaching an opinion. Clinical judgment, which is an integral part of both traditional and forensic evaluations, has problems and limitations. Because these errors and biases can occur at all levels of clinical practice, the type of information outlined above should be integrated into the general curriculum of graduate clinical training in psychology/psychiatry and/or included in continuing education programs. Most clinical training currently focuses on the content of assessment (e.g., diagnostic criteria or meanings of scores on psychological tests) rather than on the process of clinical judgment and decision making. However, the literature suggests that knowledge or awareness of these limitations alone is insufficient. Clinicians must know the research on these limitations, how they are manifested in clinical practice, and how to avoid, or at least minimize, their impact.

Specifically, evidentiary requirements and professional ethics mandate that those who conduct mental health evaluations for the courts be familiar with problems and limitations in their assessments and take steps to minimize the impact of such weaknesses and biases in their evaluation process. This article is our attempt to provide this information to practicing clinicians in order to improve clinical judgment and decision making in forensic evaluation.

Notes

1. Curran W, McGarry A, Shah S (eds): *Forensic Psychology and Psychiatry*. Philadelphia: Davis, 1986; Grisso T: *Evaluating Competencies*. New York: Plenum Press, 1986; Melton G, Pettila J, Poythress N, Slobogin C: *Psychological Evaluations for the Courts: A Handbook for Mental Health Professionals and Lawyers*. New York: Guilford, 1987.
2. Bazelon D: Psychiatrists and the adversary process. *Scientific American* 230:18-23, 1974; Bazelon D: Can psychiatry humanize the law? *Psychiatric Annals* 7(5):29-39, 1977; Melton, Pettila, Poythress, Slobogin, *supra* note 1; Morse S: Crazy behavior, morals, and science: An analysis of mental health law. *Southern California Law Review* 51:527-654, 1978a; Morse S: Law and mental health professionals: The limits of expertise. *Professional Psychology: Research and Practice* 9:389-399, 1978; Poythress N: Mental health expert testimony: Current problems. *Journal of Psychiatry and Law* 5:207-227, 1977; Poythress N: Concerning reform in expert testimony: An open letter from a practicing psychologist. *Law and Human Behavior* 6:39-43, 1982; Shapiro D: *Psychological Evaluation and Expert Testimony*. New York: Van Nostrand Reinhold, 1984.
3. Faust D, Ziskin J: The expert witness in psychology and psychiatry. *Science* 241:31-35, 1988; Ziskin J, Faust D: *Coping with Psychiatric and Psychological Testimony* (3rd ed.). Venice, CA: Law & Psychology Press, 1988.
4. An extensive review of the experimental decision-making literature is beyond the scope of this paper. A number of such reviews exist; see *supra* notes 1-3; Arkes HR: Impediments to accurate clinical judgment and possible ways to minimize their impact. *Journal of Consulting and Clinical Psychology* 49:323-330, 1981; Faust D: *The Limits of Scientific Judgment*. Minneapolis: University of Minnesota Press, 1974; Kahneman D, Slovic P, Tversky A (eds): *Judgment Under Uncertainty*. New York: Cambridge University Press, 1982; Tversky A, Kahneman D: Judgment under uncertainty: Heuristics and biases. *Science* 185:1124-1131, 1974.
5. Berkeley D, Humphreys P: Structuring decision problems and the bias heuristic. *Acta Psychologica* 50:201-252, 1982; Cohen LJ: Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences* 4:317-370, 1981; Edwards W: Human cognitive capabilities, representativeness, and ground rules for research. In Humphreys P, Swenson O, Vari A (eds): *Analyzing and Aiding Decision Processes* (pp. 507-513). New York: North-Holland, 1983; Funder DC: Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin* 101:75-90, 1987; Hogarth RM: Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin* 90:197-217, 1981;

- Phillips LD: A theoretical perspective on heuristics and biases in probabilistic thinking. In Humphreys P, Svenson O, Vári A (eds): *Analyzing and Aiding Decision Processes* (pp. 525-543). New York: North-Holland, 1983.
- For a review see Gigerenzer G: How to make cognitive illusions disappear: Beyond "heuristics and biases." In Stroebel W, Hewstone M (eds): *European Review of Social Psychology* 2:83-115. London: Wiley, 1991.
- Christensen-Szalanski JJ, Beach LR: The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist* 39:75-78, 1984.
- Ceci SJ, Bruck M: (in press). The suggestibility of the child witness: An historical review and synthesis.
- Supra* note 4.
- Snyder M: Seek and ye shall find: Testing hypotheses about other people. In Higgins ET, Zanna MP (eds): *Social Cognition: The Ontario Symposium on Personality and Social Psychology*. Hillsdale, NJ: Lawrence Erlbaum, 1981.
- Arkes HR, Harkness AR: Effect of making a diagnosis on subsequent recognition of symptoms. *Journal of Experimental Psychology: Human Learning and Memory* 6:568-575, 1980.
- Reyes RM, Thompson WC, Bower GH: Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social Psychology* 39:2-12, 1980.
- See text pp. 49-53 *infra*.
- Arkes HR, Rothbart M: Memory, retrieval, and contingency judgments. *Journal of Personality and Social Psychology* 49:598-606, 1985; Ward WC, Jenkins HM: The display of information and the judgment of contingency. *Canadian Journal of Psychology* 19:231-241, 1965.
- Chapman LJ, Chapman JP: Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74:271-280, 1969.
- Leuger RJ, Petzel TP: Illusory correlation in clinical judgment: Effects of amount of information to be processed. *Journal of Consulting and Clinical Psychology* 47:1120-1121, 1979.
- Arkes, *supra* note 4.
- Committee on Ethical Guidelines for Forensic Psychologists: Specialty guidelines for forensic psychologists. *Law and Human Behavior* 15:655-665, 1991.
- Id.* at 661.
- Id.* at 661.
- American Psychological Association: Ethical principles of psychologists and code of conduct. *American Psychologist* 47:1597-1628, 1992.
- Id.* at 1602.
- Golding SG: Mental health professionals and the courts: The ethics of expertise. *International Journal of Law and Psychiatry* 13:281-307, 1990.
- Exner J: *The Rorschach: A Comprehensive System*. New York: Wiley & Sons, 1974; Lenzak M: *Neuropsychological Assessment* (2nd ed). New York: Oxford University Press, 1983; Kleinmuntz B: Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin* 107:296-310, 1990.
- Dawes RM: The robust beauty of improper linear models in decision making. *American Psychologist* 34:571-582, 1979; Dawes RM, Corrigan B: Linear models in decision making. *Psychological Bulletin* 81:95-106, 1974; Faust, *supra* note 4; Faust D: Research on human judgment and its application to clinical practice. *Professional Psychology: Research and Practice* 17:420-430, 1986; Faust D: Data integration in legal evaluations: Can clinicians deliver on their promises? *Behavioral Sciences & the Law* 7:469-483, 1989; Nisbett RE, Borgida E, Crandall R, Reed H: Popular induction: Information is not necessarily informative. In Carroll J, Payne J (eds). *Cognition and Social Behavior* (pp. 227-236). Hillsdale, NJ: Lawrence Erlbaum, 1976; Goldberg LR: Simple models or simple processes? Some research on clinical judgments. *American Psychologist* 23:483-496, 1968.
- Hoffman PJ, Slovic P, Rorer LG: An analysis-of-variance model for the assessment of configurational cue utilization in human judgment. *Psychological Bulletin* 69:338-349, 1968; Newell A, Simon HA: *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall, 1972; Slovic P, McPhiliamy SC: Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance* 11:172-194, 1974; Slovic P: Cue-consistency and cue-utilization in judgment. *American Journal of Psychology* 79:427-434, 1966.

27. Goldberg, *supra* note 25; Rorer LG, Hoffman HD, Dickman HD, Slovic P: Configural judgments revealed (summary). Proceedings of the 75th Annual Convention of the American Psychological Association 2:195-196, 1967; Wiggins N, Hoffman PI: Three models of clinical judgment. *Journal of Abnormal Psychology* 73:70-77, 1968.
28. Faust, 1986, *supra* note 25 at 225.
29. Dawes, *supra* note 25.
30. Einhorn HJ: Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance* 6:1-17, 1971; Slovic, *supra* note 26; Slovic & McPhillamy, *supra* note 26.
31. Wiggins JS: *Personality and Prediction: Principles of Personality Assessment*. Reading, MA: Addison-Wesley, 1973.
32. Slovic P, Lichtenstein S: Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* 6:649-744, 1971.
33. Oskamp S: Overconfidence in case-study judgments. *Journal of Consulting Psychology* 29:261-265, 1965; Sines LK: The relative assessment. *Journal of Consulting Psychology* 23:483-492, 1959.
34. Duker J: The utility of the MMPI atlas in the derivation of personality descriptions. *Dissertation Abstracts* 19:3021. (University Microfilms No. 59-1259), 1959; Garb HN: The incremental validity of information used in personality assessment. *Clinical Psychology Review* 4:641-655, 1984; Golden M: Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology* 28:440-446, 1964; Goldstein SG, Deyssach RE, Kleinknecht RA: Effect of experience and amount of information on identification of cerebral impairment. *Journal of Consulting and Clinical Psychology* 41:33-34, 1973; Graham JR: A O-sort study of the accuracy of clinical descriptions based on the MMPI. *Journal of Psychiatric Research* 5:297-305, 1967; Kostlan A: A method for the empirical study of psychodiagnosis. *Journal of Consulting Psychology* 18:83-88, 1954; Marks PA: An assessment of the diagnostic process in a child guidance setting. *Psychological Monographs* 75 (3, Whole No. 507) 1961.
35. Barendregt JT: *Research in Psychodiagnostics*. The Hague: Mouton (1961); Cochrane CT: Effects of diagnostic information on empathic understanding by the therapist in a psychotherapy analogue. *Journal of Consulting and Clinical Psychology* 38:359-365, 1972; Einhorn, *supra* note 30; Gadol I: The incremental and predictive validity of the Rorschach test in personality assessments of normal, neurotic, and psychotic subjects. *Dissertation Abstracts* 29:3482-B. (University Microfilms No. 69-4469) 1969; Leung & Petzel, *supra* note 16; Perez FI: Behavioral analysis of clinical judgment. *Perceptual and Motor Skills* 43:711-718, 1976; Pokay CR, Ward EF: Clinical judgment under varied informational conditions: Rorschach, personal data, and best guess. *Journal of Consulting and Clinical Psychology* 39:513, 1972; Sines, *supra* note 33; Soskin WF: Bias in postdiction from projective tests. *Journal of Abnormal and Social Psychology* 49:69-74, 1954; Soskin WF: Influence of four types of data on diagnostic conceptualization in psychological testing. *Journal of Abnormal and Social Psychology* 58:69-78, 1959; Wildman RW, Wildman RW II: An investigation into the comparative validity of several diagnostic tests and test batteries. *Journal of Clinical Psychology* 31:455-458, 1975; Winch RF, More DM: Does TAT add information to interview? Statistical analysis of the increment. *Journal of Clinical Psychology* 12:316-321, 1956.
36. Heilbrun K: The role of psychological testing in forensic assessment. *Law and Human Behavior* 16:257-272, 1992; Rogers R: *Clinical Assessment of Malingering and Deception*. New York: Guilford, 1988.
37. Arkes HR: Principles in judgment/decision making research pertinent to legal proceedings. *Behavioral Sciences & the Law* 7:429-456, 1989.
38. Other factors are also relevant to determining the applied utility, such as the costs associated with various outcomes. See Wiggins, *supra* note 31, for a discussion.
39. Manis M, Dovatina I, Avis NE, Cardoze S: Base rates can affect individual predictions. *Journal of Personality and Social Psychology* 38:231-248, 1980.
40. Arkes, *supra* note 4; Kahneman D, Tversky A: On the psychology of prediction. *Psychological Review* 80:231-251, 1973; Meehl PE, Rosen A: Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin* 52:194-216, 1955; Nisbett, Borgida, Crandall, Reed, *supra* note 25; Tversky A, Kahneman D: Causal schemata in judgments under uncertainty. In Fishbein M (ed), *Progress in Social Psychology* (pp. 49-72). Hillsdale, NJ: Lawrence Erlbaum, 1978; Wiggins, *supra* note 31.
41. Aizen I: Intuitive theories of events and effects of base-rate information on prediction. *Journal of Personality and Social Psychology* 35:303-314, 1977; Arkes, *supra* note 37; Bar-Hillel M: The base-rate fallacy in probability judgments. *Acta Psychologica* 44:211-233, 1980.

42. Maris, Dovalina, Avis, Cardoze, *supra* note 39.
43. Hansen RD, Donoghue JM: The power of consensus: Information derived from one's own and others' behavior. *Journal of Personality and Social Psychology* 35:294-302, 1977; Wells GL, Harvey JH: Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology* 35:279-293, 1977.
44. Otto RK: Prediction of dangerous behavior: A review and analysis of "second generation" research. *Forensic Reports* 5:103-133, 1992.
45. Swanson JW, Holzer CE, Ganzu VK, Jono RT: Violence and psychiatric disorder in the community: Evidence from the epidemiological catchment area surveys. *Hospital and Community Psychiatry* 41:761-770, 1990.
46. Maris RW, Berman AL, Malsberger JT, Yulif RI (eds): *Assessment and Prediction of Suicide*. New York: Guilford, 1992.
47. Heaton RK, Grant I, Matthews CG: *Comprehensive Norms for an Expanded Halstead-Reitan Battery*. Odessa, FL: Psychological Assessment Resources, 1991.
48. Maris, Berman, Malsberger, Yulif, *supra* note 46.
49. We realize that this approach may be somewhat frustrating to professionals insofar as we do not identify any kind of formula that can be used. Should we weigh the absence of alcohol or drug abuse equally to the current report of depression? The current state of the art does not provide us with the information necessary to develop such formulas, and the clinician must choose between simply following the base rate or choosing the base rate and then altering predictions based on other salient factors, weighing the information to the best of his or her ability.
50. Greenwald AG, Pratkanis AR, Leippe MR, Baumgardner MH: Under what conditions does theory obstruct research progress? *Psychological Review* 93:216-229, 1986; Koehler DJ: Explanation, imagination, and confidence in judgment. *Psychological Bulletin* 110:499-519, 1991; Koriat A, Lichtenstein S, Fischhoff B: Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory* 6:107-118, 1980; Mahoney MJ: Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research* 1:161-175, 1977; Mitruff I: *The Subjective Side of Science*. Amsterdam: Elsevier, 1974; Nisbett RE, Ross L: *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall, 1980; Rorer LG: The proper domain of prediction. Paper presented at the meeting of the Western Psychological Association, Honolulu (June 1965); Snyder, *supra* note 10; Tweney RD, Doherty ME, Mynatt CR (eds): *On Scientific Thinking*. New York: Columbia University Press, 1981; Wedding D: Clinical judgment in forensic neuropsychology: A comment on the risks of claiming more than can be delivered. *Neuropsychology Review* 2:233-239, 1991; Smedsund J: Note on learning contingency and clinical experience. *Scandinavian Journal of Psychology* 7:265-266, 1966.
51. Anderson CA, Lepper MR, Ross L: Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology* 39:1037-1049, 1980; Bruner JS, Potter MC: Interference in visual recognition. *Science* 144:424-425, 1964; Dailey CA: The effects of premature conclusions upon the acquisition of understanding of a person. *Journal of Psychology* 33:133-152, 1952; Friedlander ML, Stockman S: Anchoring and publicity effects in clinical judgment. *Journal of Clinical Psychology* 39:637-643, 1983; Tversky & Kahneman, *supra* note 4.
52. Fischhoff B, Slovic P, Lichtenstein S: Fault trees: Sensitivity of estimated failure probabilities of problem representation. *Journal of Experimental Psychology: Human Perception and Performance* 4:330-344, 1978; Koehler, *supra* note 50; Ross L, Lepper MR, Strack F, Steinmetz JL: Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology* 35:817-829, 1977; Sherman SJ, Skov RB, Hervitz EF, Stock CB: The effects of explaining hypothetical future events: From possibility to probability to actuality and beyond. *Journal of Experimental Social Psychology* 17:142-158, 1981.
53. Bruner & Potter, *supra* note 51.
54. Faust 1986, *supra* note 25.
55. Of course, expert witnesses should strive for objectivity and not accept or conduct an evaluation with the intent of advocating for one side or the other. After reaching an opinion, however, the examiner can conceivably be an advocate *for that opinion*, and not for a particular side. It is important that the evaluator not identify with one side *before* the evaluation has been independently performed. Research findings suggest that expert witnesses may identify with the side by which they are retained (by a hypothesized phenomenon identified as "forensic identification"), so that bias (both conscious and unconscious) can occur. See, for example, Homant RJ, Kennedy DB: Subjective factors in clinicians' judgments of insanity: A comparison of a hypothetical and an actual case. *Professional Psychology: Research and Practice* 18:439-446, 1987; Homant RJ, Kennedy DB: Subjective factors in the judgment of insanity. *Criminal Justice and Behavior* 14:38-61, 1987; Otto RK: Bias and expert testimony of mental health professionals in adversarial

- proceedings: A preliminary investigation. *Behavioral Sciences and the Law* 7:267-273, 1989; Simon J, Zusman J: The effect of contextual factors on psychiatrists' perceptions of illness: A case study. *Journal of Health and Social Behavior* 24:186-198, 1983; Zusman J, Simon J: Differences in repeated psychiatric examinations of litigants to a lawsuit. *American Journal of Psychiatry* 140:1300-1304, 1983; but such an effect has not always been observed—see Beckham JC, Annis LV, Gustafson DL: Decision making and examiner bias in forensic expert recommendations for not guilty by reason of insanity. *Law and Human Behavior* 13:79-87, 1989.
56. The facts of this clinical example are taken from *Estelle v. Smith*, 451 U.S. 454 (1981), in which a well-known psychiatrist, Dr. James Grigson, after interviewing the defendant for 90 minutes, concluded that the defendant was "a very severe sociopath" who would continue to "commit other similar or same [sic] criminal acts if given the opportunity to do so" and whose condition would "only get worse" (pp. 459, 460). In another Supreme Court case (*Barclay v. Estelle*, 463 U.S. 880, 1983), Dr. Grigson described a defendant as an "above ten" sociopath on a scale of one to ten, and characterized the probability of the defendant's committing future acts of criminal violence as "one hundred percent and absolute" (p. 919). Such opinions have led some to refer to Dr. Grigson as "Dr. Death."
57. See text p. 54 *infra*.
58. It should be noted, however, that some investigators have found such procedures to be of little value. Golding SL, Rorer LG: Illusory correlation and subjective judgment. *Journal of Abnormal Psychology* 80:249-260, 1972.
59. Friedlander ML, Phillips SD: Preventing anchoring effects in clinical judgment. *Journal of Consulting and Clinical Psychology* 52:366-371, 1984.
60. Arkes HR, Harkness AR: Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General* 112:117-135, 1983; Nisbett & Ross, *supra* note 50; Shadlee H, Wasseman EA: Judging interevent contingencies: Being right for the wrong reasons. *Bulletin of the Psychonomic Society* 24(2):91-94, 1986; Shustak MW, Sternberg RJ: Evaluation of evidence in causal inference. *Journal of Experimental Psychology: General* 110:101-120; Smedlund, *supra* note 50; Ward & Jenkins, *supra* note 14; Wasseman EA, Donner WW, Kao SF: Contributions of specific cell information to judgments of inter-event contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:509-521, 1990.
61. Chapman LJ: Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior* 6:151-155, 1967; see p. 151.
62. Chapman LJ, Chapman JP: Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology* 72:193-204, 1967; Chapman & Chapman, *supra* note 15; Golding & Rorer, *supra* note 58; Kurtz RM, Garfield SL: Illusory correlation: A further exploration of Chapman's paradigm. *Journal of Consulting and Clinical Psychology* 46:1009-1015, 1978.
63. Tversky & Kahneman, *supra* note 4.
64. Rorer, *supra* note 50.
65. Wedding D, Faust D: Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology* 4:233-265, 1989.
66. And applied diagnostic or predictive utility is not guaranteed even when this requirement is met, given the problems associated with the occurrence of low base rate behaviors; see text pp. 44-45, *supra*.
67. See, for example, Ceci & Bruck, *supra* note 8; Utah v. Rimmach, 775 P.2d 388 (S. Ct. Utah, 1989); Youniss D: Constitutional perspective: Note-evaluating and admitting expert opinion testimony in child sexual abuse prosecutions. *Duke Law Journal*, 42:691-725, 1991.
68. Arkes, *supra* note 4.
69. Wood G: The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance* 4:345-353, 1978.
70. Arkes HR, Wortman RL, Saville P, Harkness AR: The hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology* 66:252-254, 1981; Arkes HR, Faust D, Guimette TJ, Hart K: Eliminating the hindsight bias. *Journal of Applied Psychology* 73:305-307, 1988; Dawson NV, Arkes HR, Blinkhorn R, Lakshmanan M, Petrelli M: Hindsight bias: An impediment to accurate probability estimation in clinicopathologic conferences. Paper presented at the meeting of the Society for Medical Decision Making, Chicago (October 1986); Fischhoff B: Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance* 1:288-299, 1975; Hawkins SA, Hastie R: Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin* 107:311-327, 1990; Leary MR: Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin* 8:257-263, 1982; Mitchell TR, Kalb LS: Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology* 66:605-612, 1981.

71. Arkes, Faust, Guilmette, Hart, *supra* note 70.
72. *Id.* at 306.
73. Poythress NG: Negligent release litigation: A proposal for procedural reform. *Journal of Psychiatry and Law* 17:595-605, 1989.
74. Of course, there are some problems with this approach. It is reasonable to expect that the colleagues who are enlisted would realize that "something wrong" must have happened or there would be no inquiry. Therefore, hindsight bias could not be truly eliminated, since they would suspect that some kind of negative outcome occurred. However, this potential problem can be minimized by providing the reviewers with three or four additional (actual) cases to comment upon, none of which has a negative outcome. Using this approach, the reviewers would not know in which case there was a bad outcome. This approach, while an interesting one, is not likely to prove useful, given its complexity and evidentiary requirements. For a proposal and discussion of other debiasing techniques when commenting on and establishing standards of care with respect to malpractice, see Weiner R: A psycholegal and empirical approach to the medical standard of care. *Nebraska Law Review* 69:112-154, 1990.
75. Fischhoff B: Debiasing. In Kahneman D, Slovic P, Tversky A (eds), *Judgment Under Uncertainty* (pp. 422-444). New York: Cambridge University Press, 1982; Fischhoff, Slovic, Lichtenstein, *supra* note 52; Kahneman & Tversky, *supra* note 40; Nisbett & Ross, *supra* note 50; Slovic P, Fischhoff B, Lichtenstein S: Behavioral decision theory. *Annual Review of Psychology* 28:1-39, 1977; see note 56 for an example of such overconfidence.
76. Arkes, *supra* note 4; Faust D: Learning and maintaining rules for decreasing judgment accuracy. *Journal of Personality Assessment* 50:585-600, 1986; Faust 1986, *supra* note 25; Goldberg, *supra* note 25; Lanyon RI, Goodstein LD: *Personality Assessment* (2nd ed). New York: Wiley, 1982; Mischel W: *Personality and Assessment*. New York: Wiley, 1968.
77. Dawson NV, Connors AF, Hsiue I, Shaw P: Hemodynamic assessment in the critically ill: Is physician confidence warranted? *Clinical Research* 35:737A, 1987; Faust D, Hart K, Guilmette TJ, Arkes HR: Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice* 19:508-515, 1988; Holsopple J, Phelan J: The skills of clinicians in analysis of projective tests. *Journal of Clinical Psychology* 10:307-320, 1954; Moxley A: Clinical judgment: The effects of statistical information. *Journal of Personality Assessment* 37:86-91, 1973; Poses RM: Physicians' judgments in the real world: Cause for concern. Presented at the annual meeting of the Society for Medical Decision Making, Richmond, VA (October 1988); Wedding D: Clinical and statistical prediction in neuropsychology. *Clinical Neuropsychology* 5:49-55, 1983.
78. Garb HN: The appropriateness of confidence ratings in clinical judgment. *Journal of Clinical Psychology* 42:190-197, 1984.
79. *Id.* at 194.
80. Lindsay RCL: Confidence and accuracy of eyewitness identification from lineups. *Law and Human Behavior* 10:229-239, 1986.
81. Einhorn HJ: Overconfidence in judgment. In Shweder R, Fiske D (eds), *New Directions for Methodology of Behavioral Research: Fallible Judgment in Behavioral Research*. San Francisco: Jossey-Bass, 1980; Einhorn HJ, Hogarth RM: Confidence in judgment: Persistence in the illusion of validity. *Psychological Review* 85:395-416, 1978; Friedlander & Phillips, *supra* note 59; Koriat, Lichtenstein, Fischhoff, *supra* note 50.
82. Koriat, Lichtenstein, Fischhoff, *supra* note 50.
83. Faust D, Nurcombe B: Improving the accuracy of clinical judgment. *Psychiatry* 52:197-208, 1989.
84. Kahneman, Slovic, Tversky, *supra* note 4.
85. For a discussion of this issue specific to the claim of PTSD related to Vietnam combat experiences, see the special PTSD issue of *Behavioral Sciences and the Law*, Vol. 1 No. 3, 1983.
86. Egendorf A, Kadushin C, Laufer RS, Rothbart G, Sloan L: *Legacies of Vietnam: Comparative Adjustment of Veterans and Their Peers*. New York: Center for Policy Research, 1981.
87. Credit is given to Thomas Grisso for recognizing and developing this example.
88. Dawes R, Faust D, Meehl P: Clinical versus actuarial judgment. *Science* 243:1668-1674, 1989.
89. Arkes HR, Dawes RM, Christensen C: Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes* 37:93-110, 1986.
90. Tversky & Kahneman, *supra* note 4.
91. Faust, *supra* note 76; Faust & Nurcombe, *supra* note 83; Tversky & Kahneman, *supra* note 4.
92. Kahneman & Tversky, *supra* note 40.

93. Kahneman & Tversky, *supra* note 40 at p. 250. Of course, in this type of evaluation there are a number of other factors to consider, such as practice effects, the different contexts of the evaluations, influence of emotional distress, full-scale IQ reliability, etc. This case is used simply to illustrate a situation in which regression effects should be considered.
94. Wedding & Faust, *supra* note 65 at 256.
95. Goldberg LR: Human mind versus regression equation: Five contrasts. In Cicchetti D, Grove W (eds), *Thinking Clearly About Psychology. Volume 1: Matters of Public Interest* (pp. 173-184). Minneapolis: University of Minnesota Press, 1991; Vanderploeg R: The evaluation process: Interview and testing. In Vanderploeg R (ed), *Neuropsychological Assessment Practice*. Hillsdale, NJ: Lawrence Erlbaum, in press; Wedding & Faust, *supra* note 65 at 256.