Available online at www.sciencedirect.com
SCIENCE DIRECT®
Children and Youth Services Review
27 (2005) 409–427
Children and Youth Services Review
www.elsevier.com/locate/childyouth

The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management

Aron Shlonsky[a,*], Dennis Wagner[b]

[a]*School of Social Work, Columbia University, 1255 Amsterdam Avenue, 10027 New York, NY, United States*
[b]*Children's Research Center, United States*

Available online 22 December 2004

Abstract

Assessment in child welfare involves at least two distinct processes: an assessment of risk (prediction of future harm) and a contextual assessment of child and family functioning used to develop case plans. Both types of assessment are critical decision aids, yet there has been confusion in the field about their respective uses. Actuarial risk assessment instruments clearly have the greatest potential to reliably and accurately estimate the recurrence of child maltreatment. This type of risk assessment, however, does not indicate which clinical factors are most important to address and certainly does not indicate which services are most likely to be effective. The structured decision making (SDM) approach is an example of an effort to integrate predictive and contextual assessment strategies into child welfare practice. Clinical decision makers complete both an actuarial risk assessment and an objective assessment of family strengths and needs. Both assessments incorporate clinical input in their design and completion. Yet this is still not enough. Clinicians must translate information from both forms of assessment into the choice of a set of effective service interventions. The *process* of evidence-based practice and the establishment of structural supports for this practice model may be the next step in the evolution of child protective services.

Keywords: Risk assessment; Structured decision making; Child welfare

* Corresponding author. Tel.: +1 2128512247.
E-mail address: as2156@columbia.edu (A. Shlonsky).

0190-7409/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.childyouth.2004.11.007

1. The problem of risk assessment in child protective services

Child protective services workers are charged with a series of difficult tasks, not the least of which is to assess the level of risk posed by parents to their children (Gambrill & Shlonsky, 2001) and to make a series of casework decisions based upon this appraisal (Rycus & Hughes, 2003). The former requires an accurate assessment of the risk of future harm and the latter requires a detailed assessment of individual and family functioning and related factors. Both types of assessment are critical decision aids, yet there has been confusion in the field about their respective uses (Rycus & Hughes, 2003). Mistakes at any juncture can lead to serious harm to children, either through subsequent maltreatment or unwarranted separation from their parents. Yet this combination of severe consequences, the inherent difficulty of conducting accurate assessments, and the variation in skills among CPS workers is a set up for unreliable case decision making.

Clinical decisions involving the assignment of risk are marked by cognitive biases and thinking errors, resulting in decisions that tend to have limited predictive validity (Dawes, 1994; Grove & Meehl, 1996). Clinicians must observe, select, and organize "risk factors" from a vast array of case specific information presented to them in initial contacts with families. Assessment of risk must be made quickly due to the potential danger faced by vulnerable children and the demands of the administrative setting in which the decision is made. Yet how do caseworkers select the most important factors? Are they able to weigh each observed risk factor and formulate an accurate prediction of future harm? There is evidence to suggest that unassisted clinicians, even those with experience and a great deal of case information at their disposal, are not very good at predicting future client behavior (Dawes, 1994; Dawes, Faust, & Meehl, 1989). The reasons for this are manifold, including flawed heuristic strategies (see, for example, Nisbett & Ross, 1980), but basically boil down to an inability to accurately weigh and combine large amounts of disparate and often conflicting information. In essence, too much data may prompt the clinician to select factors for the decision that have no relationship to the behavioral outcome being forecast (Faust, 1984).

Whereas clinical judgment (in this case, prediction of future behavior) is based on case study, intuitive judgment, or the professional experience of the worker, formal risk assessment measures have been developed in child welfare in order to combat the shortfalls of unassisted clinical judgment (Rycus & Hughes, 2003). These generally fall into two categories: consensus-based and actuarial. Both are designed to improve clinical judgment by identifying specific characteristics for examination during the process of risk assessment. Consensus risk assessment models are compiled by "experts" who may draw upon previous research findings, clinical experience, or a combination of both, but an empirical study is usually not conducted to validate the assessment. Actuarial assessments, on the other hand, are developed by observing the behavioral outcomes of interest to the predictive enterprise. Empirical research procedures are employed to identify a set of risk factors with a strong statistical relationship to the behavioral outcome. These are then weighted and combined to form an assessment tool that optimally classifies families or individuals according to the "risk" that they will exhibit the behavior. The use of actuarial versus clinical prediction has had a fairly long history of controversy in psychology (Table 1). However, this is one of the few controversies that appear to have been settled.

Table 1
History of the actuarial vs. clinical prediction controversy in psychology

1943	Sarbin's (1943) study of university admissions finds that a simple formula estimates academic success better than university admission counselors.
1954	Blenkner finds that clinical social workers could observe and score specific case characteristics, which could be combined in a simple, but effective, actuarial tool. She also finds that the actuarial tool provides a better estimate of future client success in a family casework agency than the prognosis of clinical social workers.
1954	Meehl reviews 20 research studies in which simple actuarial instruments proved either superior or equivalent to clinical prediction in estimating future behavior.
1966	Sawyer (1966) reviews 45 research studies, which compare actuarial and clinical prediction. He concludes that actuarial procedures proved better than or equal to clinical decisions in every study.
1989	Dawes, Faust, and Meehl restate these findings.
1996	Grove and Meehl update Meehl's (1954) study with same result, this time using 136 studies over a wide range of outcomes.
The controversy in child welfare	
1984	Johnson and L'Esperance conduct the first study comparing a statistical model to clinical prediction of subsequent maltreatment in a child welfare sample.
1990	Wald and Woolverton write a seminal article pointing out the difficulties of risk assessment and, while advocating the use of actuarial instruments, highlighting some of the challenges of constructing and using such tools.
1996	Rossi, Schuerman, and Budde (1996) conduct an experimental study using vignettes, finding that clinical decisions to place children are fairly unreliable and are influenced by factors unrelated to the family.
1999	Baird, Wagner, Healy, and Johnson compare the reliability of three risk assessment instruments, one actuarial and two consensus-based, finding that the actuarial model had better reliability.
2000	Baird and Wagner compare the validity of three risk assessment instruments, one actuarial and two consensus-based, finding that the actuarial model more accurately categorized children into varying degrees of risk of maltreatment.

Put simply, carefully validated actuarial models outperform clinical judgment at estimating future behavior.

Reflecting on the overwhelming set of research findings that examined this issue, Dawes (1993) comments:

> . . .objections [to using statistical models] ignore the data from well over 100 studies. . . The objections to using statistics also ignore the ethical mandate that, for important social purposes such as protecting children, decisions should be made in the best way possible. If relevant statistical information exists, use it. If it doesn't exist, collect it. (p. 39)

The evidence that actuarial estimates perform better than clinical judgment at forecasting client behavior is clear. However, Meehl (1954), among other advocates of actuarial prediction in psychology, recognized long ago that clinical judgment remained critical to the effective treatment of client conditions and noted that clinicians identified the risk factors employed in actuarial tools. In psychology, it was understood that actuarial instruments could not assess family dynamics or develop treatment plans. As one psychologist noted, the optimal combination of actuarial method and clinical judgment must be sought for the task at hand (Holt, 1958). Child welfare also needs to integrate

actuarial information into the clinical decision making process required to address client problems. This paper discusses just such an approach.

2. A brief history of actuarial risk assessment in CPS

Actuarial studies of risk assessment in child welfare began in the early 1980s with a seminal study by Johnson and L'Esperance (1984), who developed and tested a predictive model for maltreatment recurrence. In 1990, Wald and Woolverton (1990) reviewed proceedings of the APWA/AHA risk assessment roundtable (American Public Welfare Association, 1988) and wrote a key article identifying many concerns about the use of risk assessment in child protective agencies. They noted that most risk assessments used by child welfare agencies had not been empirically validated. At that time, only two actuarial risk assessments had been developed, yet they maintained that actuarial processes had the greatest potential for creating reliable and valid risk assessment instruments. They cautioned, however, that even validated risk assessments were far from perfect and that development of accurate actuarial models would take time. These included the dangers of misclassification, concerns about a lack of clear definitions of critical concepts (e.g., child neglect), implementation errors, and various statistical pitfalls. McDonald and Marks (1991) reached similar conclusions in a contemporary review.

Building on empirical research conducted in child welfare agencies in Alameda County, California Johnson and L'Esperance (1984), Alaska and Michigan as well as studies from the adult and juvenile justice systems, Children's Research Center (CRC) conducted an Office of Child Abuse and Neglect (OCAN) funded study in three sites evaluating the reliability and predictive validity of one actuarial and two consensus-based risk assessments under actual field conditions (Baird & Wagner, 2000; Baird, Wagner, Healy, & Johnson, 1999). The actuarial instrument demonstrated higher inter-rater reliability and superior estimates of future maltreatment than the consensus instruments. Actuarial instruments are now used for estimating the future risk of maltreatment at the close of an investigation by several state or county child welfare agencies.

Recent CPS agency interest in actuarial risk assessment comes, in part, from increased demands placed on child protective services in an era of declining staff resources, high staff turnover, and increased public accountability (United States General Accounting Office, 1997). The federal Child and Family Service Review (CFSR) criteria now include new substantiations after the close of an investigation as a measure for evaluating agency performance, further generating interest in simple, risk assessment procedures with known predictive attributes. There is also an increasing recognition in social work that assessment procedures and service interventions need to be empirically valid. While there is considerable debate about the form and function of evidence-based practice (Gambrill, 2003b), there appears to be a growing trend in the use of the term "evidence-based" as well as an increase in the reporting of clinical trials as a proportion of literature published in several human services fields (Shlonsky & Gibbs, 2004).

The primary mission of every CPS agency is to manage risk to children (i.e., the risk of maltreatment), essentially meaning that cases need to be managed in a way that reduces harm to children. Effective agency case management, therefore, requires a reliable and valid estimate of future child maltreatment in order to inform key decisions that involve risk to children (Rycus & Hughes, 2003). Actuarial risk assessments do not yield infallible estimates of future harm but there is evidence that they provide the best available mechanism for estimating the probability of future maltreatment at a critical point—the close of the protective service investigation (Baird & Wagner, 2000; Baird et al., 1999). This critical juncture marks the point at which a case is either opened for agency child welfare services (e.g., voluntary or court ordered in-home service or placement) or closed (i.e., with a community service referral, or no service).

3. Developing an actuarial risk assessment instrument

A better understanding of the development of actuarial assessments may help dispel resistance to the use of actuarial risk assessment among those who believe that it undermines clinical assessment. An actuarial risk assessment instrument is based on direct observation of case outcomes and related predictors in a given jurisdiction. It also employs a great deal of input from agency social workers. In a recent California risk assessment study (Wagner, Johnson, & Johnson, 1998), the researchers convened a group of experienced workers and supervisors from several counties to guide the research effort and review the findings. Group members reviewed relevant research and drew upon their experience in the field. They were asked to identify and carefully define family risk factors that in their opinion: (1) had a relationship to future maltreatment and (2) could be reliably observed and scored by a worker at the close of a CPS investigation. This information was used to develop the data collection instrument.

This approach was pursued to improve the practical utility of the assessment (after all they know what is in their case files) and build face validity among workers who must use it. It also recognizes the critical role clinical staff can play in identifying risk factors for test in an actuarial study. This role was acknowledged by Meehl (1954) after his review of a study by Blenkner (1954). Blenkner's study found that clinical social workers could identify risk factors that could be effectively employed in a simple actuarial instrument. The California risk assessment study serves as an example. Workers identified the great majority of factors that appear on the actuarial risk assessment (shown in Fig. 1) and it is not surprising that many of the factors require clinical judgment to score. The workgroup also defined each of the concepts and related attributes contained in the measure (i.e., related evidentiary standard of abuse) and collaborated with the researchers to develop a set of instructions for scoring items. These were then translated into user-friendly paper manuals and pop-up screens for electronic versions, and was integrated into mandatory training sessions.

The field research used to develop the data collection tool examined a random sample of 2511 families investigated for abuse or neglect. Potential family and child risk factors were drawn only from information available in the case files at investigation closure. Cases were followed for 24 months after investigation closure and outcomes measured

CALIFORNIA
FAMILY RISK ASSESSMENT c: 06/02

Referral Name: _____ **Referral #:** _____ **County:** _____

County Name: _____ **Worker Name:** _____ **Worker ID#:** _____

NEGLECT		Score	ABUSE		Score
N1.	Current Complaint is for Neglect a. No ... 0 b. Yes ... 1	_____	A1.	Current Complaint is for Abuse a. No ... 0 b. Yes ... 1	_____
N2.	Prior Investigations (assign highest score that applies) a. None ... 0 b. One or more, abuse only ... 1 c. One or two for neglect ... 2 d. Three or more for neglect ... 3	_____	A2.	Number of Prior Abuse Investigations/Assessments a. None ... 0 b. One ... 1 c. Two or more ... 2 (actual number: _____)	_____
N3.	Household has Previously Received CPS (voluntary/court-order) a. No ... 0 b. Yes ... 1	_____	A3.	Household has Previously Received CPS (voluntary/court-ordered) a. No ... 0 b. Yes ... 1	_____
N4.	Number of Children Involved in the CA/N Incident a. One, two, or three ... 0 b. Four or more ... 1	_____	A4.	Prior Injury to a Child Resulting from CA/N a. No ... 0 b. Yes ... 1	_____
N5.	Age of Youngest Child in the Home a. Two or older ... 0 b. Under two ... 1	_____	A5.	Primary Caretaker's Assessment of Incident (check applicable items and add for score) a. Not applicable ... 0 b. ____ Blames child ... 1 c. ____ Justifies maltreatment of a child ... 2	_____
N6.	Primary Caretaker Provides Physical Care Inconsistent with Child Needs a. No ... 0 b. Yes ... 1	_____	A6.	Domestic Violence in the Household in the Past Year a. No ... 0 b. Yes ... 2	_____
N7.	Primary Caretaker has a Past or Current Mental Health Problem a. No ... 0 b. Yes ... 1	_____	A7.	Primary Caretaker Characteristics (check applicable items and add for score) a. Not applicable ... 0 b. ____ Provides insufficient emotional/psychological support 1 c. ____ Employs excessive/inappropriate discipline ... 1 d. ____ Domineering caretaker(s) ... 1	_____
N8.	Primary Caretaker has Historic or Current Alcohol or Drug Problem. (Check applicable items and add for score) a. Not applicable ... 0 b. ____ Alcohol (current or historic) ... 1 c. ____ Drug (current or historic) ... 1	_____	A8.	Primary Caretaker has a History of Abuse or Neglect as a Child a. No ... 0 b. Yes ... 1	_____
N9.	Characteristics of Children in Household (Check applicable items and add for score) a. Not applicable ... 0 b. ____ Medically fragile/failure to thrive ... 1 c. ____ Developmental or physical disability ... 1 d. ____ Positive toxicology screen at birth ... 1	_____	A9.	Secondary Caretaker has Historic or Current Alcohol or Drug Problem a. No ... 0 b. Yes, alcohol and/or drug (check all applicable) ... 1 _____ Alcohol ___ Drug	_____
N10.	Housing (check applicable items and add for score) a. Not applicable ... 0 b. ____ Current housing is physically unsafe ... 1 c. ____ Homeless at time of investigation ... 2	_____	A10.	Characteristics of Children in Household (check appropriate items and add for score) a. Not applicable ... 0 b. ____ Delinquency history ... 1 c. ____ Developmental disability ... 1 d. ____ Mental health/behavioral problem ... 1	_____
	TOTAL NEGLECT RISK SCORE	_____		**TOTAL ABUSE RISK SCORE**	_____

SCORED RISK LEVEL. Assign the family's scored risk level based on the highest score on either the neglect or abuse instrument, using the following chart:

Neglect Score	Abuse Score	Scored Risk Level
_____ 0 – 1	_____ 0 – 1	_____ Low
_____ 2 - 4	_____ 2 – 4	_____ Moderate
_____ 5 - 8	_____ 5 – 7	_____ High
_____ 9 +	_____ 8 +	_____ Very High

POLICY OVERRIDES. Circle yes if a condition shown below is applicable in this case. If any condition is applicable, override final risk level to very high.

Yes No 1. Sexual abuse case AND the perpetrator is likely to have access to the child victim.
Yes No 2. Non-accidental injury to a child under age two.
Yes No 3. Severe non-accidental injury.
Yes No 4. Caretaker(s) action or inaction resulted in death of a child due to abuse or neglect (previous or current).

DISCRETIONARY OVERRIDE. If a discretionary override is made, circle yes, circle override risk level, and indicate reason. Risk level may be overridden one level higher.

Yes No 5. If yes, override risk level (circle one): Low Moderate High Very High

Discretionary override reason: _____

Supervisors Review/Approval of Discretionary Override: _____ Date: ___/___/___

FINAL RISK LEVEL (circle final level assigned): Low Moderate High Very High

O:\555-1\Training Materials 6-03\Worker Training\P&P Manual.doc

Fig. 1.

included: (1) subsequent investigations, (2) subsequent substantiations of maltreatment, (3) severity of subsequent maltreatment (injury, hospitalization, medical exam, or death), and (4) subsequent child placement in foster care. Using case-level data, separate

analyses were conducted to develop risk assessments for both abuse and neglect outcomes (e.g., abuse and neglect investigation, substantiation, and child injury outcomes). In each case, risk factors were identified that had a strong statistical relationship to case outcome measures in a construction sample (i.e., randomly selected portion of the entire sample). These were then placed in separate multivariate models (statistical models that measure the effect of each predictor in the presence of other predictors) in order to control for potential overlap between risk factors and to create a parsimonious model.[1] Those items drawn from multivariate analyses were then weighted (given more or less importance based upon the strength of their relationship with the case outcomes relative to other variables) and arranged on two assessments—one for neglect and the other for abuse (see Fig. 1). The assessment was then validated (a process of testing its ability to estimate maltreatment) in a second sample of cases that were not used in the construction sample (see, for example, Altman & Royston, 2000; Gottfredson & Gottfredson, 1979; Silver & Chow-Martin, 2002). The neglect and abuse assessments both yield a four-level classification. The higher of the two is employed as the scored risk level or final risk classification for maltreatment.

4. Two types of actuarial risk assessment estimates

While actuarial risk assessment may have the greatest potential to predict recurrence of maltreatment, it still faces serious methodological challenges common to most empirical measurements that employ field data. In particular, child welfare outcomes can be somewhat "noisy" or inexact. For example, substantiation may be influenced by structural or institutional factors that have nothing to do with child or parent characteristics. In addition, the case outcome measure examined (e.g., maltreatment substantiation documented by CPS agency staff) may not capture every incident of maltreatment (English, Marshall, & Orme, 1999). There may also be critical pieces of predictive information about the case that remain undiscovered or cannot be accurately ascertained during initial contacts with the family (e.g., parental history of abuse in childhood).

Yet these problems impact actuarial, consensus, and clinical assessment methods alike. There is no free ride. To argue that consensus-based and clinical models of decision-making are not subject to the need for empirical validation is clearly a disservice to children and families. Validation can, and should, occur with any of these measures. The difference is that actuarial tools are built upon the very outcome measures being employed. If a better outcome measure can be found or constructed, it, too, can be used to create an empirically valid tool. In the absence of more accurate markers of outcome, a more useful strategy might be to adjust the tool to better reflect field contingencies. In the noisy and fragmented information environment of CPS, the

[1] Separate models for abuse and neglect outcomes are estimated because the dynamics of abuse and neglect may differ as do risk factors with a significant relationship to abuse or neglect case outcomes. Note that the case outcomes are not independent of one another (i.e., a family must be investigated to be substantiated and child injury or placement largely occurs in the presence of a substantiated maltreatment incident).

classical binary prediction schema where a specific behavioral outcome is declared in advance (e.g., the family will maltreat their child) may not hold much promise. Rather, a classification scheme may be necessary whereby families are placed in at least three graduated risk groups based on their estimated probability of future maltreatment (for an example in an allied field, see Monahan & Steadman, 1996).

4.1. Traditional binary

Binary (yes/no) prediction measures in child welfare, while far better than non-data driven approaches, tend to have inadequate predictive validity (Lyons, Doueck, & Wodarski, 1996; Munro, 2004; Ruscio, 1998). In general, this type of schema allows for four possible predictive outcomes: (1) true positives—children who are predicted to be subsequently maltreated who actually are maltreated; (2) false positives—children predicted to be subsequently maltreated but are not; (3) false negatives—children predicted not to be maltreated who actually are maltreated; and (4) true negatives—children predicted not to be maltreated who are not (Table 2). There are essentially four constructs found on the margins of this table that determine the predictive validity of an instrument: (1) positive predictive value or PPV—the probability that a case with a positive test (predicted to maltreat) will actually maltreat; (2) negative predictive value or NPV—the probability that a case with a negative test (predicted not to maltreat) does not maltreat; (3) sensitivity—the proportion of cases with a positive test (will maltreat) who are accurately identified by the measure; (4) specificity—the proportion of cases with a negative test (will not maltreat) who are accurately identified by the measure.

Although not optimized for a binary prediction schema (i.e., ROC analysis), data from the California validation study were used to collapse the four-level classification system into two categories: 'predicted *not* to be maltreated' (combination of low and moderate risk) and 'predicted to be maltreated' (combination of high and very high risk) (Table 3). While actual child maltreatment was more prevalent among children who were predicted to be maltreated (PPV=0.35) than among children who were predicted not to be maltreated (NPV=0.88), 65% of the maltreated children were incorrectly labeled. That is, more children were *incorrectly* predicted to be maltreated than were correctly predicted to be maltreated. The sensitivity (degree to which true cases are correctly identified) of this measure is 0.69, meaning that 31% of the true high risk cases were missed. The specificity (degree to which non-abusing families are correctly

Table 2

	Abuse found	Abuse not found	
Abuse predicted	True positives (TP)	False positives (FP)	Positive predictive value TP/(TP+FP)
Abuse not predicted	False negatives (FN)	True negatives (TN)	Negative predictive value TN/(TN+FN)
	Sensitivity TP/(TP+FN)	Specificity TN/(FP+TN)	Prevalence rate (TP+FN)/(TP+FP+FN+TN)

Adapted from Gibbs (2003).

Table 3
Reclassified California actuarial data from validation study

Follow-up substantiation		
Binary risk prediction	Sub	No sub
High or very high	380	712
Low or moderate	174	1245
Total	554	1957
Predictive validity		
Overall hit rate	Sub	No sub
Low or moderate	0.12	0.88
High or very high	0.35	0.65
True positive	380	
False positive	712	
False negative	174	
True negative	1245	
Positive predictive value	0.35	
Negative predictive value	0.88	
Sensitivity	0.69	
Specificity	0.64	
Prevalence rate	0.22	

identified) stands at 0.64, meaning that 36% of the truly low risk cases were missed. While better than chance alone, the predictive properties of this contrived tool are not of sufficient quality to warrant its use as the sole predictive measure of whether or not children will be harmed in the future.[2]

4.2. Classification scheme

A more realistic and clinically informative schema is to classify cases into varying levels of risk associated with an outcome of interest, moving away from the binary prediction framework. Rather than predicting what will occur, classification of greater or lesser degree of risk simply informs clinicians and agencies about which cases are more likely than others to be high risk. While not stating with certainty that a child will be maltreated, a risk classification indicates that the likelihood of maltreatment is greater for one case than it is for others in the near future (typically 6, 12, or 24 months). Human behavior is extremely difficult to predict, and classification recognizes this inherent limitation. Rather than usurping a clinical decision, a risk classification summarizes key case information observed during an investigation into what is currently the most reliable and valid estimate of the risk of future harm. Once the risk assessment is scored, clinical judgment must still be employed to make the final decision whether to offer services or

[2] For another explanation of some of these key psychometric properties, particularly with respect to base rate of occurrence, please see Munro (2004).

close the case. At this pivotal transition, actuarial estimation is integrated with, and supplemented by, clinical judgment. For example, in the California model, a caseworker and supervisor can make the decision to exercise a discretionary override to the scored risk classification.

The logic of the transition from actuarial scoring to clinical judgment was originally described by Meehl (1954). He noted that actuarial assessments had a limited purpose and relied on limited client information. Only the clinician could interview the client directly and make certain qualitative judgments. Meehl referred to this blind spot in the actuarial method as the "broken leg" problem. An example is a situation where an actuarial formula derived from an analysis of past race performance predicts that a runner will finish a 10-km race in approximately 45 min. If the runner has a broken leg, the actuarial prediction will be wrong. This kind of contingency cannot be accounted for in an actuarial assessment, but a clinician, given access to the subject, can often observe such information. Actuarial assessment in child welfare should not be used to dictate type or scope of service. Rather, it should be used to improve clinical decisions (particularly those made at the close of the investigation), thereby improving the agency's ability to manage risk to children.

Presented here are the results from the validation study of the California Family Risk Assessment. This assessment uses four risk classifications ranging from low to very high. The percent of children experiencing a subsequent substantiated maltreatment reoccurrence, subsequent placement, and child injury are displayed over each risk level (Fig. 2). As level of risk increases, so, too, does the percent of children experiencing these outcomes. Children falling into the highest risk categories have a higher likelihood of experiencing these events, while children at the lower risk levels have a lower likelihood. In general, the base rate of occurrence (the rate of occurrence within the population of interest) should fall below the highest risk classifications and below the lowest

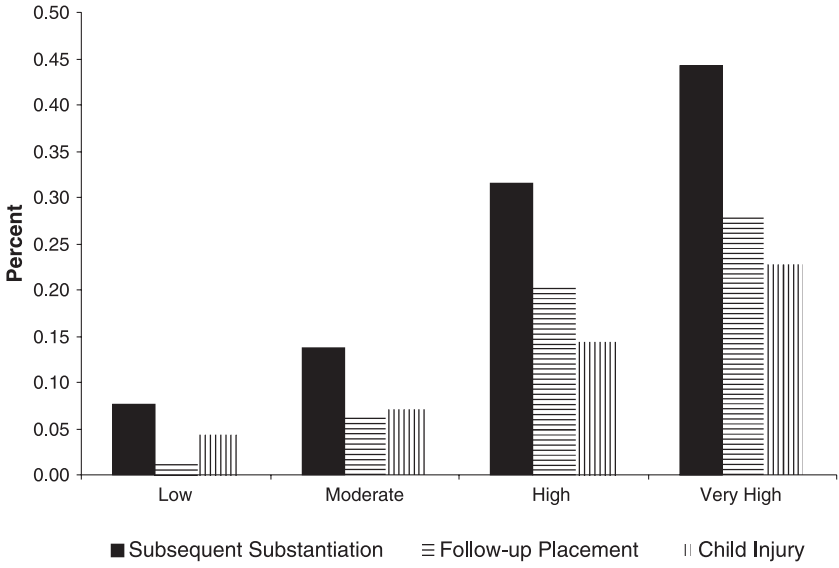

Fig. 2. California final family risk classification by follow-up substantiation, placement, and injury.

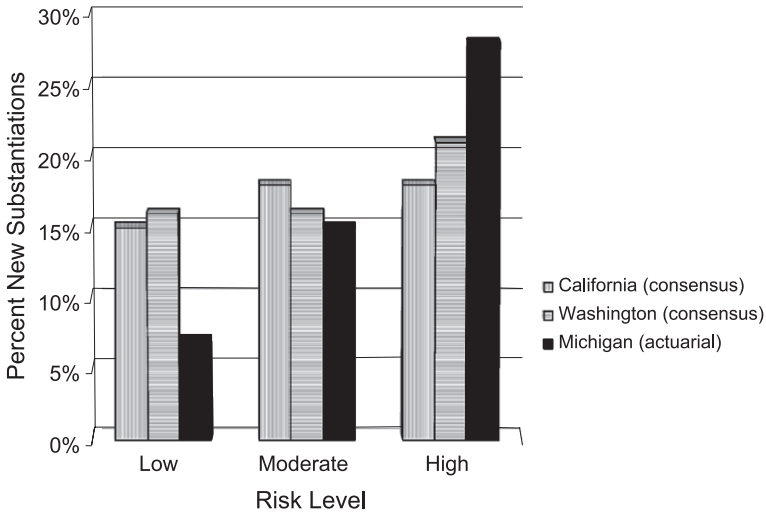

Fig. 3. Actuarial vs. consensus-based classification of subsequent substantiations.

classifications.[3] Thus, in relative terms, the model distinguishes between risk levels fairly well. Although there are still a high number of misclassifications, the assessment does not purport to predict which cases will have, for example, a subsequent substantiated maltreatment recurrence. The risk assessment tool simply assigns a risk classification that estimates maltreatment relative to other cases. The clear presentation of the expected percentage within each category, as well as training pertaining to the base rate of event occurrence in the population, provides workers and supervisors with a realistic expectation of these future events rather than relying on speculation. This method has far more transparency than consensus-based instruments or a clinical decisions that do not present such figures.

When actuarial models are compared to consensus-based tools, the actuarial measure clearly surpasses expert-driven tools in its ability to discriminate levels of risk. Using the same outcome measures (follow-up investigation, substantiation, and placement), Baird and Wagner (2000) compared Michigan's actuarial assessment to the Washington Assessment of Risk Matrix (WARM) and the California (Fresno) risk assessment tool (both consensus-based models). The Michigan actuarial model clearly discriminated risk levels in the desired direction (higher proportion of substantiations as risk level increased), while the WARM had difficulty discriminating between low and moderate risk and the Fresno model was virtually unable to distinguish between moderate and high levels of risk (Fig. 2). Although the WARM was able to perform fairly well on some outcome measures (e.g., subsequent placements), it never performed as well as the Michigan actuarial model (Fig. 3).

[3] For a good discussion of baserates and their implications in child welfare sample, please see Munro (2004) and Wood and Writght (1995).

There appear to be several advantages to the use of actuarial risk assessment strategies in child welfare. First, the models have a demonstrated capacity to classify families into groups with significantly different probabilities of future maltreatment. In addition to construct validity, workers tend to classify families more consistently (inter rater reliability) (Baird et al., 1999). That is, different workers tend to estimate the same level of risk given the same child and family circumstances. This is likely due to the fact that actuarial instruments tend to employ risk factors that have more objective scoring criteria and rely less upon subjective judgments (e.g., parental cooperation). If actuarial risk assessment studies actively seek worker input during their development, the risk factors employed are more likely to be easily observed by workers during an investigation and the assessment will have considerable face validity. In addition, a carefully designed actuarial risk assessment also increases the transparency in the case decision making process that leads to case opening and it provides the investigating worker with a better estimate of future maltreatment than other approaches. The primary mission of child welfare agencies is to effectively manage the risk of child maltreatment and actuarial risk assessment can help agencies target scarce service interventions to children at the greatest risk of harm. Despite these attributes, there are clear limitations.

A risk classification, whether obtained from a consensus or actuarial assessment, estimates the probability of child maltreatment among families with similar characteristics. It is not an infallible prediction, nor is it a substitute for the exercise of sound professional judgment by the investigating worker who completes it. An actuarial risk assessment finding should not be the sole basis for a case decision. Appropriate use in the field requires that workers understand how actuarial risk assessments work, know the limitations of the estimates they make, and receive the training and policy guidance necessary to employ them effectively in the field. For example, agency policy should include clear guidelines as to how to proceed with a family, given a certain risk level, and under which circumstances overrides based on agency policy (for sex abuse, serious injury to a child, etc.) or the discretion of the worker (clinical judgment) should be made.

5. The role of clinical judgment in actuarial risk assessment

> A risk estimate is like a weather forecast. The forecast may indicate a 60% or higher probability of rain the next day. That forecast may prove to be wrong but since it rained 60% of the time on similar days in the past, it may be a good idea to carry an umbrella (Gottfredson & Gottfredson, 1985).

The difference between prediction and causality is monumental. It is foolish to think that just because an item is on an actuarial risk assessment instrument and has an established association with maltreatment that it is anything more than a visible link in the causal chain. An actuarial risk classification is not intended to be the only information employed in the case assessment and decision making process. As noted above, Paul Meehl (1954), among others, recognized that clinical judgment was not suspended by a risk assessment estimate because there are many case circumstances

that risk assessments do not capture (e.g., the "broken leg"). That being said, there is every reason to design actuarial assessments in a way that practitioners can understand and appreciate the various items they contain. If the instrument or the items it contains can convey a story that meshes with case worker experience, they may be more comfortable and vigilant in its use (Schwalbe, 2004). Even the most predictive, psychometrically sound instrument, if not used, will be ineffective in the field. There is some evidence that a well-constructed, easily scored actuarial instrument can be effectively used in the field. Johnson (2004) conducted a prospective revalidation study of 7635 cases from a variety of California counties in which the California Risk Assessment Instrument was completed by trained workers in the field during the course of their investigation. He found that the instrument maintained its psychometric properties indicating that, with proper training, the scale can be relied upon even after field implementation.

In child welfare, there are also several unique methodological challenges to maximizing predictive capacity (Gambrill & Shlonsky, 2000). For instance, actuarial tools are necessarily built upon a review of cases where children were not immediately and permanently removed. Thus, it may be that a good portion of the worst cases are removed from the risk set at the outset.[4] As mentioned previously, the rate of misclassification is still fairly high. Actuarial tools only estimate specific types of risk, but there may be other forms of risk that are of import (e.g., risk of poor developmental outcomes). Despite these limitations (and others), the proper use of empirically derived information offers the greatest opportunity for improving casework decisions.

Actuarial risk assessment does not assist in case specific clinical decisions, nor does it engage the family in cooperative case planning, assess their functioning, establish case plan goals, or choose treatment interventions, nor would the predictive elements of consensus or clinical assessment. Risk assessment simply informs the worker about the likelihood of future maltreatment at a given point in time. At such a transition point, the actuarial assessment is completed and the critical, clinical judgments about service interventions that may reduce the likelihood of future maltreatment must be made (Holt, 1958; Meehl, 1954). A comprehensive, contextualized family assessment is required to identify and clarify relevant problems at the individual, family, community, and societal level (Gambrill, 1997). In particular, specific problem behaviors and the context in which they occur must be clearly identified and described. Obtaining information about how the clients view the problem is also key. Client perception may be more accurate and informs case planning in terms of setting and attaining agreed-upon goals. Likewise, individual and family strengths (e.g., social support, employment, strong emotional ties) must be found, acknowledged, and integrated into case plans.

5.1. Integrating risk and context: A standardized approach

The structured decision making (SDM) process offers one example of an assessment framework in which actuarial models function in concert with the contextual clinical

[4] Though some might argue that the reliability of casework decisions is high (see, for example, Lindsey, 1991, 1992), minimizing this concern.

assessment necessary to plan service interventions.[5] The actuarial risk assessment tool is used to help establish the intensity of the CPS agency response (i.e., set case opening/service priority), but the case planning process proceeds with a separate contextual component, the Family Strengths and Needs Assessment (FSNA),[6] that is completed after the risk assessment. The combined use of actuarial risk assessment and an objective, clearly articulated structure for documenting clinical assessment findings is intended to help agencies reduce subsequent maltreatment through improved targeting of limited resources for service intervention. For example, high risk families can be given priority for intensive case intervention services used to maintain families or reunify them, as well as worker case management activity and contact. However, a deep understanding of individual and family functioning may be required in order to select those services most likely to address key areas of need. The FSNA provides a first step for engaging families in that process.

The California FSNA assessment, which is used to inform case planning, was also developed in a work group setting by experienced social workers from the agency that subsequently employed it. The group is asked to identify major areas of family functioning that may pose problems requiring service intervention or serve as strengths in the casework process. The result is a consensus-based assessment instrument with carefully defined items designed to objectively score child/family functioning in several areas including: substance abuse, mental health, physical health, family relationships, housing, and social support. In some California counties, the FSNA is completed as part of a case or family group decision-making conference, allowing families the opportunity to more fully participate in the assessment and case planning process.

The FSNA findings simply provide the initial framework for a comprehensive clinical assessment. Workers use it in an effort to help them: (1) conduct and document a reliable family assessment; (2) screen families for more specialized assessment procedures (e.g., substance abuse or mental health evaluations); (3) choose service interventions (including referrals to outside agencies); (4) develop a case plan which identifies specific goals; and (4) monitor service interventions. While relatively simple in form, this assessment provides a concise evaluation of family functioning for review by other workers, the clients, supervisors, or the court. In addition, assessment findings are easily entered into a data base where, in aggregate form, they can support agency service delivery planning (e.g., if a large number of families experience a certain type of problem, services can be developed or sought that effectively address this issue). Over time, the empirical data secured from actuarial tools, clinical assessments, service participation, and case outcomes can be used to evaluate which service interventions

[5] Although risk and family assessments are highlighted in this paper, SDM employs separate assessments at other critical decision points in child welfare cases. Specifically, instruments are used at: (1) CPS intake/referral screening (i.e., the "hot line"); (2) investigation response priority assignment (i.e., immediate, 72 h); (3) standardized family safety assessment (during the investigation); (4) actuarial risk assessment for estimating subsequent abuse/neglect (at close of investigation); (5) standardized family strengths and needs assessment for case planning prior to service intervention; (6) standardized risk/needs reassessments for open family cases and reunification assessments for children in foster care.

[6] The Family Strengths and Needs Assessment is used for this purpose in the structured decision making model, though there are clearly other comprehensive assessments that could be used.

worked for whom. This clinical assessment portion of SDM is designed to help workers carefully evaluate a potentially critical set of child and family strengths and needs before developing a case service plan. Standardization makes worker assessments more reliable, furnishes a brief format for documenting case notes, supplies additional criteria for classifying cases based on prioritized service or treatment needs, and provides useful information for constructing fundamental progress indicators. While fairly comprehensive, the FSNA is fairly unidimensional. That is, it does not provide a clinically sensitive assessment of the dynamics of the client's problems (e.g., does the client drink because of depression, job loss, social isolation, or some other factor?). What it does offer is a framework for exercising clinical judgment about the case dynamics and the services that may be required given the issues at hand.

6. The next step—evidence-based practice in child welfare[7]

The approach described represents the very beginning of what is necessary in child protective services. Correct information from actuarial and clinical forms of assessment must somehow translate to a set of effective services. The *process* of evidence-based practice and the structural elements to support its full-scale implementation may be the next step in the evolution of child protective services. The integration of current best evidence, clinical expertise, and client state/preferences (Sackett, Richardson, Rosenberg, & Haynes, 1997) is a conceptual model used in evidence-based medicine that incorporates elements of assessment into practice. The conceptual model is applied through a process of posing an answerable question, querying a database in order to find current best evidence, and applying it to client and clinical context. Specifically, the five step model proposed by Sackett, Straus, Richardson, Rosenberg, and Haynes (2000) includes:

Step 1 Converting the need for information (about prevention, diagnosis, prognosis, therapy, causation, etc.) into an answerable question
Step 2 Tracking down the best evidence with which to answer the question
Step 3 Critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect), and applicability (usefulness in our clinical practice)
Step 4 Integrating the critical appraisal with our clinical expertise and with our patient's unique biology, values, and circumstances
Step 5 Evaluating our effectiveness and efficiency in executing steps 1–4 and seeking ways to improve them both for next time (pp. 3–4).

Advantages of EBP include: (1) enabling staff and clients to make decisions based on the best available evidence (e.g., helping practitioners to keep up-to-date with current research findings related to important decisions that affect children and their families); (2) encouraging participation in evidence-based continuing education programs that

[7] Portions of this section were adapted from Shlonsky and Gambrill.

contribute to high quality practice; (3) honoring ethical obligations to clients (e.g., to offer competent services and to fully inform them); (4) clearly describing outcomes sought and progress indicators, and tracking these on an ongoing basis. Hallmarks of EBP are transparency of what is done to what effect and a consideration of the values and expectations of clients, including involving clients as informed (rather than misinformed or uninformed) participants in decisions made (Gambrill, 2003a). Transparency of what is done to what effect should encourage risk management programs that minimize errors, mistakes, and harm and maximize use of services found to help clients achieve valued outcomes (Gambrill & Shlonsky, 2001). Increased attention to errors, mistakes, and harm in the helping professions and recognition of limited resources should also encourage risk management programs. These trends will increase accessibility of information related to important decisions to both clients and professionals.

The relationship of EBP to risk assessment is suggested in Fig. 4 following a recent conception of the EBP model by Haynes, Devereaux, and Guyatt (2002). Using current best evidence as an entry point, an actuarial assessment of risk can target scarce resources to clients at highest risk. Relevant data sources on current best evidence include the Cochrane and Campbell Collaboration databases, extending to databases such as Medline, Psycinfo, CINAHL, Social Services Abstracts, and others. The contextual assessment uses clinical expertise to elicit key strengths and needs as well as client preferences as movement is made toward service provision. At this point, current best evidence is again sought regarding other assessment tools (e.g., depression inventories, child behavioral indicators) and the effectiveness of various service options (e.g., MST, CBT for depression, parenting classes). What is found is integrated with client circumstances and characteristics, including their preferences, again drawing on clinical expertise to integrate data from disparate sources.

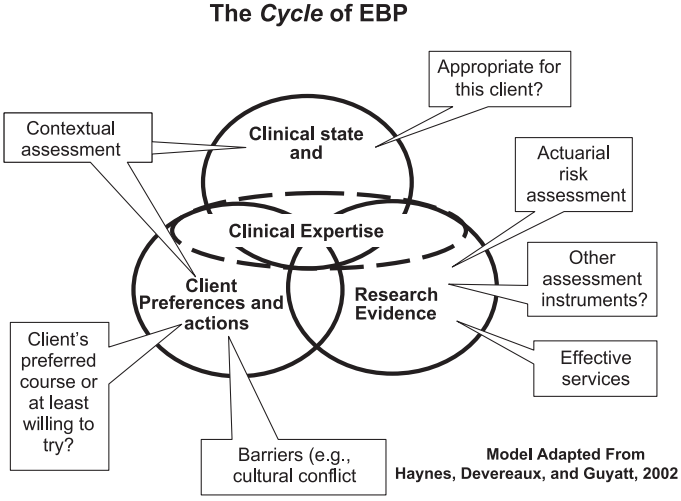

Fig. 4. MS document object.

Risk assessment and contextual assessment can be seen as decision aids employed in the service of evidence-based practice, but these aids must be supported by child welfare agencies in order for them to be viable. This will require an agency in which the values of EBP are incorporated including transparency of what is done to what effect, integrating research and practice, involving clients as informed participants, providing infrastructure (e.g., good computers, access to databases) and technical support (e.g., access to an information specialist), and addressing application barriers. Although evaluations of whether social workers can be influenced to use evidence in practice are virtually non-existent, early findings from medicine and allied fields indicate that certain types of hands-on educational outreach, audit and feedback, continuing education, and increased use of computers as decision aids have some measurable impact on provider behavior (Gira, Kessler, & Poertner, 2004).

In order to facilitate such a large-scale innovation, a culture of inquiry must be developed and maintained, and discovery translated into services to clients. Specifically, effective services must be identified and made available for locally prevalent problems. For instance, using data from actuarial and contextual assessment tools, as well as administrative data, each agency must identify a core set of commonly needed services (e.g., parenting education, substance abuse treatment), find those interventions with a proven track record with high risk families, and invest the time and resources to ensure that such services are available within the community. Psychometrically sound rapid assessment instruments (RAIs) for frequently encountered functional impairments (e.g., depression) should also be made available to case workers to further enhance assessments and case plans. Less conventional services would be located and engaged on an individual basis (e.g., interventions for children with autism) using the same EBP process. In all likelihood, this will require the services of a competent information specialist (for larger agencies, perhaps an information unit) in order to continuously identify effective core services, relieve some of the burden of searching from caseworkers, and train staff to become expert and efficient searchers for current best evidence (Gray, 1998).

This framework also requires an active effort to evaluate the impact of services. The use of high-quality experimental and quasi-experimental designs that incorporate intent to treat analyses, where appropriate, can help explore the effectiveness of interventions for clients at various risk levels. Administrative data can be used to monitor broad trends and suggest future directions for services. A continuous feedback loop utilizing items contained in administrative data and assessment tools could be used to ensure that the correct core services continue to be available, and support for such services could be expanded or contracted based on overall service need. In addition, we should take advantage of helpful guidelines for communicating and understanding risk (e.g., Gigerenzer & Edwards, 2003; Schwalbe, 2004). For instance, Monahan and Steadman (1996) propose that a combination of ordinal classification and probability might provide users with a more nuanced understanding of the actual risk posed. Additionally, we should attend carefully to involved parties' interests and background beliefs about risk so that appropriate risk communicative tools are created (Morgan, Fischoff, Bostrom, & Atman, 2002). Without such an approach, even the very best client assessment tools will do little to improve the lives of children and families coming into contact with the child protective services system.

References

Altman, D. G., & Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine*, *19*(4), 453–473.

American Public Welfare Association. (1988). *Summary of highlights.* Paper presented at the Second Annual Roundtable on CPS Risk Assessment, Washington, DC.

Baird, C., & Wagner, D. (2000). The relative validity of actuarial and consensus-based risk assessment systems. *Children and Youth Services Review*, *22*(11/12), 839–871.

Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare*, *78*(6), 723–748.

Blenkner, M. (1954). Predictive factors in the initial interview in family casework. *Social Service Review*, *28*, 65–73.

Dawes, R. M. (1993). Finding guidelines for tough decisions. *Chronicle of Higher*, 39.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York: Free Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674.

English, D. J., Marshall, D. B., & Orme, M. (1999). Characteristics of repeated referrals to child protective services in Washington State. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, *4*(4), 297–307.

Faust, D. (1984). *The limits of scientific reasoning*. Minneapolis: University of Minnesota Press.

Gambrill, E. (1997). *Social work practice: A critical thinkers guide*. New York, NY: Oxford University Press.

Gambrill, E. (2003a). Evidence-based practice: Implications for knowledge development and use in social work. In A. Rosen, & E. K. Proctor (Eds.), *Developing practice guidelines for social work intervention: Issues, methods, and research agenda* (pp. 37–58). New York: Columbia University Press.

Gambrill, E. (2003b). Evidence-based practice: Sea change or the emperor's new clothes. *Journal of Social Work Education*, *39*(1), 3–23.

Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review*, *22*(5–6), 813–837.

Gambrill, E., & Shlonsky, A. (2001). The need for a comprehensive risk management system in child welfare. *Children and Youth Services Review*, *23*(1), 79–107.

Gibbs, L. (2003). *Evidence-based practice for the helping professions: A practical guide with integrated multimedia*. Pacific Grove, CA: Brooks/Cole-Thomson Learning.

Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *BMJ*, *327*(7417), 741–744.

Gira, E. C., Kessler, M. L., & Poertner, J. (2004). Influencing social workers to use research evidence in practice: Lessons from medicine and the allied health professions. *Research on Social Work Practice*, *14*(2), 68.

Gottfredson, D., & Gottfredson, S. (1985). Screening for risk among parolees: Policy, practice, and methods. In D. Farrington, & R. Tarlington (Eds.), *Prediction and criminology* (pp. 54–77). Albany, NY: State University of New York Press.

Gottfredson, S., & Gottfredson, D. (1979). *Screening for risk: A comparison of methods*. Washington, DC: National Institute of Corrections.

Gray, J. A. (1998). Where's the chief knowledge officer? To manage the most precious resource of all [see comment]. *British Medical Journal*, *317*(7162), 832.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures. *Psychology, Public Policy, and Law*, *2*(2), 293–323.

Haynes, R. B., Devereaux, P. J., & Guyatt, G. H. (2002). Clinical expertise in the era of evidence-based medicine and patient choice. *ACP Journal Club*.

Holt, R. R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology*, *56*, 1–12.

Johnson, W. (2004). *Effectiveness of California's child welfare structured decision-making (SDM) model: A prospective study of the validity of the California Family Risk Assessment*. Oakland, CA: Alameda County Social Services Agency.

Johnson, W., & L'Esperance, J. (1984). Predicting the recurrence of child abuse. *Social Work Research and Abstracts*, *20*(2), 21–26.

Lindsey, D. (1991). Factors affecting the foster care placement decision: An analysis of national survey data. *American Journal of Orthopsychiatry*, *61*, 272–281.

Lindsey, D. (1992). Reliability of the foster care placement decision: A review. *Research on Social Work Practice*, *2*(1), 65–80.

Lyons, P., Doueck, H. J., & Wodarski, J. S. (1996). Risk assessment for child protective services: A review of the empirical literature on instrument performance. *Social Work Research*, *20*(3), 143–155.

McDonald, T., & Marks, J. (1991). A review of risk factors assessed in child protective services. *Social Service Review*, *65*, 112–132.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.

Monahan, J., & Steadman, H. J. (1996). Violent storms and violent people: How meteorology can inform risk communication in mental health law. *American Psychologist*, *51*(9), 931–938.

Morgan, M. G., Fischoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk communication*. New York: Cambridge University Press.

Munro, E. (2004). A simpler way to understand the results of risk assessment instruments. *Children and Youth Services Review*, *26*(9), 873–883.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Edgewood Cliffs, NJ: Prentice-Hall.

Rossi, P. H., Schuerman, J. R., & Budde, S. (1996). *Understanding child maltreatment decisions and those who make them*. Chapin Hall Center for Children, Chicago: University of Chicago.

Ruscio, J. (1998). Information integration in child welfare cases: An introduction to statistical decision making. *Child Maltreatment: Journal of the American Professional Society on the Abuse of Children*, *3*(2), 143–156.

Rycus, J. S., & Hughes, R. C. (2003). *Issues in risk assessment in child protective services: A policy white paper*. Columbus, Ohio: North American Resource Center for Child Welfare Center for Child Welfare Policy.

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1997). *Evidence-based medicine: How to practice and teach EBM*. New York: Churchill Livingstone.

Sackett, D. L., Straus, S. E., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.). New York: Churchill Livingstone.

Sarbin, T. R. (1943). A contribution to the study of actuarial and individual methods of prediction. *American Journal of Sociology*, *48*, 593–602.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin*, *66*(3), 178–200.

Schwalbe, C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review*, *26*(6), 561–576.

Shlonsky, A., & Gibbs, L. (2004). Will the real evidence-based practice please step forward: Teaching the process of EBP to the helping professions. *Journal of Brief Therapy and Crisis Intervention*, *4*(2), 137–153.

Silver, E., & Chow-Martin, L. (2002). A multiple models approach to assessing recidivism risk—Implications for judicial decision making. *Criminal Justice and Behavior*, *29*(5), 538–568.

United States General Accounting Office. (1997). *Child welfare: Complex needs strain capacity to provide services (No. GAO/HEHS-97-115).* Washingtion DC.

Wagner, D., Johnson, K., Johnson, W. (1998). *Using actuarial risk assessment to target service interventions in pilot California counties.* Paper presented at the 13th National Roundtable on CPS Risk Assessment, San Francisco, CA.

Wald, M. S., & Woolverton, M. (1990). Risk assessment: The emperor's new clothes? *Child Welfare*, *69*(6), 483–511.

Wood, J. M., & Wright, L. (1995). Evaluation of children's sexual behaviors and incorporation of base rates in judgments of sexual abuse. *Child Abuse and Neglect*, *19*(1263–1273).