

Risk and Safety Assessment in Child Welfare: Instrument Comparisons

Reader Note: The goal of a literature review is to identify the most recent, relevant, and rigorous research in order to categorize what is known, to date, about a topic of interest. It focuses on the past in order to inform the future. It does not address the present where emerging and promising practices are being carried out by practitioners. These current practices often do not have the benefit of either formal evaluations or the availability of published reports on the nature of the activities, lessons learned, or research findings. As a result, a literature review is designed to analyze what is known in order to foster critical thinking about the nature of current practices, including new ideas and directions as well as validation of current approaches. A literature review is not designed to disrupt current efforts to experiment with new, untested approaches to practice.

In order to locate and evaluate evidence for use in practice, a special kind of search strategy was used to identify published and unpublished studies using specific search terms and databases. A detailed description of the search process is located in the full report of this structured literature review available at www.bassc.net.

Introduction

Before child welfare agencies intervene with families, they are generally required to identify maltreatment or the risk of maltreatment.¹ Therefore, the assessment of risk is a critical part of child welfare agency work.² Most states in the US formalize the process of assessing risk by using some type of structured decision-making process or tool. Risk assessment tools generally include broad categories related to abuse and neglect, behavioral descriptions, procedures to determine levels of risk, and standardized forms for recording this information.³

This report, commissioned by the Bay Area Social Services Consortium, reviews the research evidence related to risk and safety assessment instruments. The first section of the report describes current approaches to risk and safety assessment. The second section describes a number of instruments used to assess risk, and summarizes research findings regarding these instruments' validity, reliability, outcomes, and use with children and families of color. The third section raises the following issues related to implementation of risk assessment models in child welfare: supports for implementation, use across the life of a case, and liability concerns. It also describes California's efforts to institute a standardized process of risk and safety assessment across the state. The report concludes with implications for practice.

Different Approaches to Risk Assessment

A variety of researchers from various academic fields have studied human decision-making and have identified a number of common errors that people make in their predictions and decisions. For example, researchers have found that people tend to: 1) ignore the probability of an event in making predictions about the likelihood of its occurrence;⁴ 2) be overconfident of

their ability to predict an event;⁵ and 3) have difficulty weighing factors related to a decision.⁶ Studies in social work suggest that child welfare workers are prone to the same difficulties in decision-making. Social workers have been found to make different decisions from one another in regards to the same situation⁷ and to be reluctant to revise their judgments of families once made.⁸ They also display skepticism of new information when it conflicts with their initial view of a family as well as a lack of skepticism regarding new information when it supports their initial view.⁹ Such studies have contributed to the notion that risk assessment processes are needed to aid social workers in decision-making.

Currently, there are two major approaches to risk assessment in child welfare decision-making: a *consensus-based* model and an *actuarial* model. Both involve a list of family or case characteristics believed to be associated with risk of maltreatment (a risk assessment instrument). However, the two approaches differ in the processes used to identify characteristics for inclusion on the instrument, and how the instruments are utilized in practice. In the field of child welfare, there is considerable debate regarding the best approach to assessing risk (Note #1).

Consensus-based instruments emphasize a comprehensive assessment of risk.¹⁰ The items are based upon various theories of child maltreatment, research literature on maltreatment, and/or the opinions of expert practitioners (hence the name “consensus-based”). Items on one instrument are often combined with items from another instrument, creating hybrid tools that vary according to the needs or beliefs of the user. Sometimes characteristics are assessed numerically (categorizing families by their total score) while other instruments simply describe areas that are to be assessed by the worker (items are coded high, moderate or low risk based upon worker judgment). A consensus-based model tends to use a single instrument to predict all forms of child maltreatment.¹¹

Consensus-based instruments can help workers structure their information gathering for clinical assessments of risk¹² as well provide documentation of the reasoning underlying their decision-making.¹³ Some argue that the more comprehensive approach of consensus-based instruments provides better information for casework decision-making.¹⁴ However, consensus-based models are criticized in social work research literature for the following reasons: 1) poor conceptualization (“measures are often poorly defined, nebulous and ambiguous, overly global,

illogical, and very subjective”);¹⁵ 2) inconsistency in the type and number of variables included;¹⁶ 3) use of the same variables to predict physical abuse, neglect, and sexual abuse, even though contributors and dynamics are often different for these types of maltreatment;¹⁷ and 4) reliance upon characteristics associated with maltreatment, rather than recurrence of maltreatment,¹⁸ or upon characteristics for which there is no research support.¹⁹

Actuarial instruments use statistical procedures to identify and weigh factors that predict future maltreatment.²⁰ Often the statistical analysis is done in the state or county in which the instrument will be applied. Actuarial instruments tend to use fewer factors than do consensus-based models and generally use different factors to predict the likelihood of physical abuse and of neglect. Each factor is scored and scores are summed into overall risk scores. Families are categorized into low, moderate or high risk groups, and receive correspondingly different service responses. Some researchers assert that actuarial models are more reliable and valid than consensus-based models since they help practitioners focus their risk assessments on a small set of case factors that have demonstrated a strong statistical relationship to future maltreatment.²¹

However, the actuarial instruments have detractors as well. Some feel actuarial tools do not incorporate or facilitate the clinical judgment of skilled practitioners. Additionally, since the basis for including a factor in an actuarial instrument is its statistical association with recurrence of maltreatment, factors may not appear to be causally related to the outcome (although in fact they may be). This perceived lack of a logical, theoretical ‘glue’ may cause social workers to discount the value of an actuarial instrument and object to its use.²²

Research on Risk Assessment

This structured review of the risk assessment research literature examined the following five risk and safety assessment instruments: 1) the Washington Risk Assessment Matrix (WRAM); 2) the California Family Assessment Factor Analysis (CFAFA, or the “Fresno” model); 3) the Child At Risk Field System (CARF); 4) the Child Emergency Response Assessment Protocol (CERAP); and 5) the actuarial Risk Assessment instruments developed by the Children’s Research Center (CRC) (Note #2).

Figures 1-5 in this section describe each instrument and summarize its performance in achieving certain

measurement qualities and outcomes (Note #3). In addition, to illustrate how each instrument structures a worker's assessment of risk, a sample question related to substance abuse is provided. Each instrument was assessed on the following criteria:

Predictive Validity: The ability of some measure to predict a particular outcome. Often the ability of an instrument to accurately classify cases into low, medium or high risk groups is assessed.

Convergent Validity: The degree to which a measure of a factor corresponds to other measures of the same or similar factors (providing evidence that the measure is assessing what it is intended to assess). While convergent validity does not describe the ability of an instrument to predict maltreatment, others argue that convergent validity is relevant to risk and safety assessment.²³

Inter-rater Reliability: The degree to which a tool results in similar decisions on similar cases, when those cases are assessed by different workers.²⁴

Outcomes: Use of a risk assessment instrument may affect certain outcomes. For example, if a newly implemented risk assessment instrument improved the ability of workers to assess risk, and workers then removed children who were at highest risk, one might expect to see fewer recurrences of maltreatment.

Racial/ethnic Group Differences: Given the well-known discrepancies in child welfare outcomes for children and families of color, it is important to consider how the use of an instrument affects different racial/ethnic groups. A risk and safety assessment instrument should be equally valid with different racial/ethnic groups.

Figure 1: The Washington Risk Assessment Matrix (WRAM)

I. INSTRUMENT DESCRIPTION

Type of Instrument: Consensus-based.

Development: The WRAM was developed by Washington State social service agency in 1986. Its composition is constantly evolving based on new research evidence.

Decision Point: The instrument is used at the initial investigation.

Domains and Items: Currently the tool has 37 items that fit within seven theoretical domains: 1) child characteristics, 2) severity of abuse/neglect, 3) chronicity of abuse/neglect, 4) caretaker characteristics, 5) caretaker/child relationship, 6) socio-economic factors, and 7) perpetrator access.

Use and Scoring: To use the instrument, social workers assess and rate the level of risk that they perceive for each item on a scale from 0-6. An overall risk score is tabulated by adding all the ratings. Based on these ratings, families are categorized into six risk levels. The instrument assesses risk of maltreatment in general, rather than considering risk for different kinds of abuse separately (e.g. neglect, abuse, sexual abuse, etc). Appendix A of the full report provides a copy of the instrument.

SAMPLE QUESTION: SUBSTANCE ABUSE

The WRAM provides the following behavioral descriptors to aid workers in assessing the risk factor of "Substance Abuse":

- **No risk (0)** = "No past or present substance abuse"
- **Low risk (1)** = "History of substance abuse but no current problem"
- **Moderate risk (3)** = "Reduced effectiveness due to substance abuse or addiction"
- **High risk (5)** = "Substantial incapacity due to substance abuse or addition"

II. INSTRUMENT PERFORMANCE

Predictive Validity: In tests of predictive validity, the WRAM did not perform very well. In one study, while rates of subsequent investigation were higher for moderate or high risk families than for low risk families, rates of *substantiated* maltreatment for families in low, moderate or high risk groups were not significantly different.²⁵ Other studies using different analytic techniques and outcomes have produced similar findings.²⁶ The authors of one of these studies concluded that the performance of the instrument "...might be characterized as generally poor."²⁷

Convergent Validity: Only a few of the items on the WRAM have been tested for convergent validity and results are mixed. In one study, nine of the 37 items on the WRAM were tested against other measures of the same constructs or ideas and four positive associations were found.²⁸

Inter-rater Reliability: In one reliability study, the WRAM performed poorly. In this study, four raters were asked to assess risk on the same 80 cases using the WRAM instrument. Less than 14% of the time, all four workers classified families in the same way; just over half the time, three out of four workers did so. Using a special procedure that corrects for agreements due to chance, a "kappa" score for the WRAM was calculated. Kappa varies from -1 to +1; a kappa of 0 would mean the performance of the tool was no better than chance. According to the authors of the study, kappa's in the range of .50-.60 are generally considered acceptable; the WRAM kappa was 0.18.²⁹

Outcomes: No studies were found that assessed the effects of WRAM on case outcomes.

Racial/Ethnic Group Differences: Findings from several studies assessing the utilization of the WRAM with different racial/ethnic groups are mixed. One study found that African American and Native American families were more likely to be assigned to the highest risk level with WRAM than their numbers in the population would suggest, while Asian American families on the other hand were under-assigned to the highest risk level.³⁰ However, another study found that approximately equal percentages for African American and white families were classified into each risk level by social workers using the WRAM.³¹

Figure 2: The California Family Assessment and Factor Analysis (CFAFA, or the “Fresno Model”)

I. INSTRUMENT DESCRIPTION

Type of Instrument: Consensus-based.

Development: The CFAFA, or the “Fresno Model,” is derived from an instrument originally developed by the state of Illinois (the Child Abuse and Neglect Tracking System or CANTS 17B). The instrument is no longer used in Illinois, but has been used most recently in California.

Decision Point : The instrument is to be used throughout the life of a case.

Domains and Items: The instrument has 23 items that fit within five theoretical domains: 1) precipitating incident, 2) child assessment, 3) caregiver assessment, 4) family assessment, and 5) family-agency interaction. All types of maltreatment are considered together.

Use and Scoring: A social worker rates each item as low, moderate or high risk; sums the number of items coded at each risk level and decides the overall level of risk. Appendix B of the full report provides a copy of the instrument.

II. INSTRUMENT PERFORMANCE

Predictive Validity: In tests of predictive validity, the CFAFA did not perform well. In one study, while rates of subsequent investigation were higher for moderate or high risk families than for low risk families, rates of *substantiated* maltreatment for families in low, moderate or high risk groups were not significantly different.³² Another study of the CANTS 17B using different analytic techniques and outcomes produced similar findings. The authors of that study concluded that the performance of the instrument “...might be characterized as generally poor.”³³

Convergent Validity: No studies assessing the convergent validity of the CFAFA were found.

Inter-rater Reliability: Like the WRAM, the CFAFA performed poorly in a reliability study. Four raters asked to rate the same 80 cases using the CFAFA instrument classified families into the same risk groups just over 16% of the time; about 45% of the time, three out of four workers did so. The kappa for the CFAFA was 0.184, indicating poor reliability.³⁴

Outcomes: No studies were found that assessed the effects of CFAFA on case outcomes.

Racial/Ethnic Group Differences: Findings from a single study of the use of the CFAFA with different racial/ethnic groups indicate that the CFAFA classified African American and White families equitably into each risk level.³⁵

SAMPLE QUESTION: SUBSTANCE ABUSE

The CFAFA provides the following behavioral descriptors to aid workers in assessing the risk factor of “Substance Abuse”:

- **Low risk (1)** = “Alcohol is consumed only in moderation and caretaker is in control of his/her actions”
- **Moderate risk (3)** = “Caretaker is currently experimenting with or using several substances; use tends to be episodic with no serious consequences or significantly reduced ability to parent; drug/alcohol abuse is not physically/psychologically addictive at this time, but pattern of misuse may be escalating”
- **High risk (5)** = “Caretaker’s life revolves around the use or attainment of drugs or alcohol, endangering the child; substance misuse poses risk to family’s financial resources and negatively affects caretaker’s ability to meet basic needs of child”

Figure 3: Child at Risk Field System (CARF)

I. INSTRUMENT DESCRIPTION:

Type of Instrument: Consensus-based.

Development: CARF is one of the first risk assessment instruments to focus on safety as distinct from risk. It was developed by ACTION for Child Protection.

Decision Point: The instrument is to be used throughout the life of a case.

Domains and Items: Fourteen items within the following 5 domains: 1) child; 2) parent; 3) family; 4) maltreatment; and 5) intervention. Four “qualifiers” are also to be considered: 1) duration of a negative influence; 2) pervasiveness of a negative influence; 3) acknowledgement by parents of a negative influence; and 4) control of the negative influence. All types of maltreatment are considered together.

Use and Scoring: Each item or qualifier is rated on a four point scale; the average of the 14 items plus the average of the four qualifiers is summed and divided by 2 to arrive at a final risk score. The family is then categorized into no risk, low risk, moderate risk, significant risk, or high risk groups. Appendix C of the full report provides a copy of the instrument.

II. INSTRUMENT PERFORMANCE

Predictive Validity: The performance of CARF on tests of predictive validity was mixed. Families assigned to the highest risk group were more likely to have a subsequent referral than families assigned the lowest risk group, though the relationship only “approached” statistical significance. Particular items were not found to be associated with subsequent maltreatment.³⁶

Convergent Validity: CARF did not perform well in a test of convergent validity. Two domains tested were not found to be associated with any of eight clinical measures of related constructs. Another domain was found to be positively associated with only one out of seven measures of a similar construct.³⁷

Inter-rater Reliability: No studies were found on the reliability of CARF.

Outcomes: CARF had no discernable affect on outcomes. In a study comparing substantiation rates before and after the implementation of CARF, no differences were found. No difference in “before and after” substantiation rates were found for most types of abuse although physical neglect was found to be somewhat more likely to be substantiated subsequent to implementation of CARF.³⁸

Racial/Ethnic Group Differences: No studies were found on the use of CARF with different racial/ethnic groups.

SAMPLE QUESTION: SUBSTANCE ABUSE

No questions on the instrument relate directly to substance abuse, but several questions relate to family functioning:

- What are the pervasive behaviors, feelings, and levels of adaptation apparent in the parent (s)?
- What is the history of the adults (parents, caretakers) in the family (recent/past)?

Figure 4: The Child Emergency Response Assessment Protocol (CERAP)

I. INSTRUMENT DESCRIPTION

Type of Instrument: Consensus-based.

Development: The CERAP was developed by Illinois Department of Child and Family Services, the American Humane Association, the University of Illinois, and field experts as a “safety assessment.”

Decision Point : The instrument is to be used throughout the life of the case.

Domains and Items: The CERAP is a single list of 14 items. All types of maltreatment are considered together.

Use and Scoring: The social worker notes the presence or absence of each item; if any of the items are present, the social worker decides whether the child is “safe” or “unsafe.” If the worker decides the child is unsafe, a safety plan is developed. Training includes a rigorous testing and certification process. Appendix D of the full report provides a copy of the instrument.

II. INSTRUMENT PERFORMANCE

Predictive Validity: Findings regarding the predictive validity of the CERAP are mixed. One study found that at initial investigation neither the overall safety assessment nor the number of safety factors identified was associated with subsequent referral within 60 days. However, when assessed within five days of case opening, both the overall safety assessment and the number of safety factors were associated with subsequent referral within 60 days in bivariate statistical tests, although these associations did not persist once other variables (prior reports, caregiver problems, and service receipt) had been controlled for.³⁹

Convergent Validity: No studies were found that assessed the convergent validity of the CERAP.

Inter-rater Reliability: No studies were found that assessed the reliability of the CERAP.

Outcomes: Several studies have found improved 60 day maltreatment recurrence rates after implementation of CERAP.⁴⁰ Reductions have been maintained for six years following implementation.⁴¹ Several alternative explanations for the reductions (e.g., increased use of out of home placement, other agency policies, and the possibility of a nationwide trend) were ruled out in a follow-up study.⁴²

Racial/Ethnic Group Differences: No studies were found that assessed the use of the CERAP with different racial/ethnic groups.

SAMPLE QUESTION: SUBSTANCE ABUSE

The CERAP asks the worker to note the presence or absence of the following condition in assessing the risk factor of “Substance Abuse”:

“Any member of household’s alleged or observed drug or alcohol abuse may seriously affect his/her ability to supervise, protect, or care for the child.”

- **Risk Factor Present** = Child unsafe, safety plan needed
- **Risk Factor Not Present** = Child not unsafe due to this risk factor

Figure 5: CRC Actuarial Models for Risk Assessment

I. INSTRUMENT DESCRIPTION

Type of Instrument: Actuarial.

Development: A number of different versions of a risk assessment instrument have been developed for various jurisdictions by the Children’s Research Center (CRC). These instruments are based upon the statistical association of variables with substantiated maltreatment injury, foster care placement, and reinvestigation within two years in each location.

Decision Point: The instrument is used at the initial investigation.

Domains and Items: The risk assessment instrument includes two subscales of ten items each; one subscale assesses risk of neglect and the other risk of physical or sexual abuse.

Use and Scoring: Each item is scored with a 0, 1, or 2 as indicated on the instrument and each subscale is summed. Based on the highest subscale score, a family is classified into a low, moderate, high, or very high risk category. In most jurisdictions, workers can override the risk classification and increase the risk rating by one level.

II. INSTRUMENT PERFORMANCE

Predictive Validity: The CRC risk assessment instrument has performed well in tests of predictive validity. The rate of subsequent maltreatment was higher for families in the high risk group than for families in the moderate risk group, and higher for the moderate risk group than for families in the low risk group. This was found to be true for subsequent maltreatment within 6 months,⁴³ 18 months,⁴⁴ and 24 months.⁴⁵

Convergent Validity: No studies were found that assessed the convergent validity of the CRC risk assessment instrument.

Inter-rater Reliability: The CRC risk assessment instrument performed fairly well in reliability studies. Four raters considering the same 80 cases assigned families to similar risk groups over half of the time; 85% of the time, three out of four workers did so. The kappa score for the CRC Risk Assessment instrument was 0.562, indicating moderate reliability.⁴⁶ In a second study, most workers scored families within a “...small and not excessively skewed” range of 4 points; somewhat lower consistency was realized when scores were combined for an overall risk score.⁴⁷

Outcomes: One study found improvement in outcomes following the implementation of an array of CRC instruments that included the risk assessment. Compared to a demographically matched set of counties that did not implement the CRC instruments, counties implementing the array of instruments had lower referral rates, substantiation rates, removal rates, and fewer injuries than comparison counties.⁴⁸

Racial/Ethnic Group Differences: Findings from studies that assess the use of the CRC instruments with different racial/ethnic groups are mixed. Equitable results were found for important areas: studies found that the instruments classify approximately equal percentages of all ethnic groups into each risk level,⁴⁹ that rates of recurrence for different risk categories are consistent across ethnic groups,⁵⁰ and that the association of scores with subsequent maltreatment does not differ by ethnic group.⁵¹ However, some differences were seen: studies have found white families were somewhat more likely to be coded higher risk on some items than families of color,⁵² that risk ratings of ethnic groups varied,⁵³ and distinctions between maltreatment rates by risk category were somewhat smaller for Native American families.⁵⁴

SAMPLE QUESTION

The CRC Risk Assessment asks the worker to note the presence or absence of the following conditions in assessing the risk factor of “Substance Abuse”:

“Primary caretaker has historic or current alcohol or drug problem”

- **Not applicable** = 0 points
- **Alcohol (current or historic)** = 1 point
- **Drug (current or historic)** = 1 point
(points for factor are summed)

Risk Assessment Implementation Issues

This section describes three implementation issues: 1) supports for implementation, 2) use at decision points across the life of a case, and 3) liability issues. The effort to institute a standardized risk assessment system across the state of California is also described.

Support for Implementation

The implementation of a risk assessment system can be problematic. One study of the CARF system reported that less experienced workers found the instrument more useful than did more experienced workers, who felt it limited their ability to utilize their expertise. In addition, "...workers manipulated the quantitative indices in order to be consistent with their clinical judgments," rather than using the risk assessment instrument to guide their decisions.⁵⁵ Similarly, in another study examining the Illinois CANTS-17B, workers admitted that they inflated risk levels at intake in order to guarantee that a family would receive services.⁵⁶ And several studies of risk assessment implementation found that risk assessment forms were often missing from case files or incomplete.⁵⁷

A number of risk assessment researchers suggest that the success or failure of any instrument is based less upon whether it is actuarial or consensus-based and more upon the quality of implementation.⁵⁸ There is some consensus that the successful implementation of risk assessment instruments requires: 1) high quality and comprehensive training,⁵⁹ 2) supervisory and management support,⁶⁰ and 3) the involvement of supervisors and direct line staff in the planning and implementation process.⁶¹

Use Across the Life of a Case

Since a family's risk may change over time, it is important that risk be periodically reassessed.⁶² Available research suggests that decisions in child welfare at these various points in the life of a case are as likely to be vulnerable to error as are decisions regarding the assessment of risk at intake.⁶³ In addition, caseworkers have difficulty revising their assessments of families once they have been made.⁶⁴ Therefore, a structured instrument could help workers attend to critical factors that would indicate changes in risk. However, there are those who assert that using the *same* instrument to assess risk at different points of time in a case is unwise.⁶⁵ The factors that predict maltreatment at one point (e.g. at investigation and prior to services) may not be the same as those that predict subsequent maltreatment at some other point in time. Unfortunately, there is very little research regarding the reliability or

validity of instruments used at other points in the life of a case.⁶⁶

Liability Issues

While legal liability in CPS work has long been a concern,⁶⁷ little has been written specifically about liability as it relates to the use of risk assessment instruments. As more states and counties mandate the use of a risk assessment model, concerns have been raised about the potential increases in the risk of legal liability for workers and agencies.

Opinions about the issue differ. Some analysts think mandating the use of risk assessment instruments increases an agency's potential liability. While section 1983 of the Civil Rights Act allows for individuals to bring suit against government employees, "...the duties of most government child welfare workers are discretionary in nature and, as such, provide them with either absolute, or ...qualified immunity for their actions."⁶⁸ By making the use of a risk assessment instrument a part of agency policy (rather than leaving its use to the worker's discretion) a worker's right to claim immunity may be reduced.⁶⁹ Secondly, some argue that liability could depend on the extent to which a risk assessment instrument is utilized. If a county or state has required the use of a risk assessment instrument and a worker fails to use the instrument correctly or does not use the instrument at all, these actions could be interpreted as negligence or failure to follow agency policies.⁷⁰ And third, liability may depend upon the validity of the tool. Use of un-tested instruments and/or inconsistent worker training on risk assessment may increase an agency's vulnerability.⁷¹ A contrasting opinion is that worker discretion and the good-faith conduct of job duties will continue to protect workers from potential lawsuits. In this case, the use of risk assessment instruments would neither improve nor worsen a worker's legal liability.⁷²

Risk and Safety Assessment in California

California's 58 counties have adopted a variety of strategies for conducting risk assessments. In the past, most counties used either the consensus-based CFAFA (the "Fresno" model), which was adopted by the state in 1991, or the CFRA (California Family Risk Assessment), the actuarial instrument developed by the Children's Research Center (CRC). Some counties do not require the use of any instrument. A major three-year systems reform effort in California (2001-2004) resulted in a recommendation that a standardized risk assessment procedure be implemented throughout the state.⁷³ A workgroup comprised of county representatives and state

personnel was convened to outline the risk and safety assessment procedure to be implemented by the state. Rather than identifying a particular instrument for state-wide implementation, the workgroup developed a framework of the various domains and items that should be assessed for risk and safety across the life of a case (see Figure 6). The decision points throughout the life of a case at which risk and safety should be assessed include: 1) determining a response to a referral, 2) making the initial safety decision, 3) placement, 4) the referral disposition, 5) ongoing case planning decisions, 6) reunification, and 7) case closure. A county can use either an actuarial or a consensus-based instrument as long as the areas relevant for each decision point are assessed.

Discussion

Overall, the available research suggests that the actuarial risk assessment models appear to have greater predictive validity and inter-rater reliability than the consensus-based models. The better predictive validity of the actuarial instruments may be related to the statistical identification of the strongest predictors of a particular outcome in that state or county. Unless the sample used to develop that model was different from the typical population referred to child welfare agencies in that jurisdiction (or there were major changes in the local context or population demographics), it would be reasonable to assume that those variables in the model would continue to be predictive on later cohorts. The processes for identifying factors for consensus-based instruments, as outlined earlier, may simply be less accurate at identifying the strongest predictors of maltreatment.

Convergent validity was not assessed for the actuarial instruments and the performance of consensus-based instruments in this area was generally poor. These instruments may be unreliable, they may not be good measures of underlying concepts, or both problems may exist.

Actuarial instruments may show better reliability because factors in these instruments are more often objective; the factors in consensus-based instruments tend to be more subjective and less precise. For example, in a question related to prior CPS history using the consensus-based WRAM, the social worker is required to determine whether past incidents were ‘isolated’ or ‘intermittent,’ and whether there is evidence of ‘minor’ abuse and neglect or ‘moderate’ abuse and neglect. In comparison, the CRC instrument asks ‘whether or not’ there was a prior injury to a child from abuse or neglect,

or ‘whether or not’ there was a prior investigation. Generally, more precise, well-defined, objective, and clearly articulated measures are more likely to be reliable because differences of opinion about the meaning or coding of factors are minimized.⁷⁴

In terms of outcomes, studies of the CERAP and the CRC instruments suggested that implementation resulted in improved outcomes. These instruments may be improving the accuracy of worker assessments of risk, resulting in fewer high risk children being left at home to be re-abused. The findings regarding the use of the instruments with different racial/ethnic groups are mixed.

It is important to note that the available research is limited. For any particular instrument, there were only a few studies available. Therefore, conclusions about the risk assessment instruments should be considered preliminary and in need of further study.

Implications for Practice

Some of the debate regarding the best approach to assessing risk and safety may be due to a lack of clarity regarding the purposes of “risk assessment.” Distinctions between *risk assessment* and *family assessment* can be somewhat unclear⁷⁵ and a number of researchers have argued that they have often been confused.⁷⁶

If the goal of an assessment is to *predict the likelihood of the recurrence of maltreatment* in order to provide services to the families at greatest risk, this is clearly a risk assessment. The research evidence strongly suggests that the actuarial instrument will produce a more accurate and reliable prediction than the consensus-based instruments. On the other hand, if the goal of an assessment activity is to *gain a comprehensive understanding of the service needs of a family or individual*, a family needs assessment instrument may be more appropriate. In this situation, a consensus-based instrument may do a better job because it incorporates more items and thus provides more information. However, consensus-based instruments did not have high convergent validity, suggesting they may not accurately measure the relevant characteristics and thus would not necessarily be helpful in family assessments.

Finally, most research on risk assessment acknowledges that the use of any kind of risk assessment instrument, actuarial or consensus-based, requires good clinical skills.⁷⁷ Even the CRC actuarial instrument contains numerous items that require clinical judgment to score,

Figure 6: Recommended Statewide Safety Assessment System for California

	DETERMINE RESPONSE	INITIAL SAFETY DECISION	PLACEMENT	REFERRAL DISPOSITION	INITIAL/CHANGES IN CASE PLANNING	REUNIFICATION	CASE CLOSURE
Purpose	Determine the appropriateness of the Child Abuse Report for a response. Determines the urgency of the response needed & the Response Team (who will respond).	Ensuring the safety of the child. Begin to identify risk factors of concern in the family.	Ensure the safety of the child by providing the least-restrictive alternative to their home environment.	Determine if a family needs continued public child welfare services, community services or has no current service needs. Examine why problems may be present, and what may be causing the symptoms and behaviors that may result in maltreatment and begins the process of helping the family.	To identify the best possible strategies for changing the conditions/behaviors that lead to harm or the risk of harm to the child.	To determine if it is safe to return a child to his or her home.	To safely terminate public CWS involvement.
Standard Areas For Review	<ol style="list-style-type: none"> 1. Current and Prior Maltreatment. 2. Current and Prior CWS history. 3. Child’s strengths and vulnerability. 4. Cultural and language considerations. 5. Perpetrator access to child. 6. Violence Propensity. 7. Social Environment. 8. Caregiver Protective Capacity. 9. Home Environment. 10. Ability to meet child’s needs. 11. Caregiver / child interaction. 12. Ability to locate. 	<ol style="list-style-type: none"> 1. Current and prior maltreatment. 2. Child’s strengths and vulnerability. 3. Cultural and language considerations. 4. Perpetrator access to child. 5. Violence propensity. 6. Social environment. 7. Caregiver protective capacity. 8. Home environment. 9. Ability to meet child’s needs. 10. Caregiver / child interaction. 11. Safety interventions. 12. Pre-placement preventive services. 	<ol style="list-style-type: none"> 1. Current and prior CWS history. 2. Child’s strengths and vulnerability. 3. Cultural and language considerations. 4. Child’s immediate and ongoing needs. 5. Level of care to meet child’s needs. 6. Substitute Care provider’s willingness and ability to provide care and ensure safety. 7. Substitute care provider’s strengths and willingness to support child’s case plan. 8. Sibling placement considerations. 9. Child’s permanency needs. 10. Visitation. 	<ol style="list-style-type: none"> 1. Current and prior maltreatment. 2. Current and prior CWS history. 3. Child’s strengths and vulnerability. 4. Cultural and language considerations. 5. Perpetrator access to child. 6. Violence Propensity. 7. Social Environment. 8. Caregiver Protective Capacity. 9. Home Environment. 10. Ability to meet child’s needs. 11. Caregiver / child interaction. 12. Ability to locate. 13. Safety interventions. 14. Caregiver willingness to change. 	<ol style="list-style-type: none"> 1. Current and prior maltreatment. 2. Child’s strengths and vulnerability. 3. Cultural and language considerations. 4. Social Environment. 5. Child’s permanency needs. 6. Visitation. 7. Contributing factors requiring intervention. 8. Current and previous social services. 9. History of criminal behavior. 10. Basic needs. 11. Medical / Dental Care. 12. Mental Health / Coping Skills. 13. Child Development. 14. Education needs. 15. Parenting Skills and practices. 16. Child’s relationships with peers and adults. 17. Substance abuse. 18. Domestic violence. 19. Delinquent behavior. 	<ol style="list-style-type: none"> 1. Current and prior maltreatment. 2. Current and prior CWS history. 3. Child strengths and vulnerability. 4. Cultural and language considerations. 5. Perpetrator access to child. 6. Violence propensity. 7. Social Environment. 8. Caregiver protective capacity. 9. Home Environment 10. Ability to meet child’s needs. 11. Caregiver / child interaction. 12. Child’s permanency needs. 13. Visitation. 14. Subsequent referrals. 15. Caregiver’s compliance / progress toward objectives within case plan. 	<ol style="list-style-type: none"> 1. Current and prior maltreatment. 3. Child’s strengths and vulnerability. 4. Cultural and language considerations. 5. Perpetrator access to child. 6. Violence Propensity. 7. Social Environment. 8. Caregiver protective capacity. 9. Home Environment. 10. Ability to meet child’s needs. 11. Caregiver / child interaction. 12. Ability to locate. 13. Child’s permanency needs. 14. Subsequent referrals. 15. Caregiver’s compliance / progress toward objectives within case plan.

and allow for a clinical override based on family characteristics or dynamics that are likely to affect risk, but are not included on the actuarial instrument. As Ereth et al. (2003) have noted, "...A caseworker can sense things that an actuarial instrument would ignore or could not employ... Many characteristics of human subjects simply cannot be quantified empirically and actuarial models cannot easily account for rare events."⁷⁸ Therefore, clinical judgment can never be eliminated from any risk assessment process. In fact, many researchers in child welfare stress that instruments for risk and safety assessment should be understood as decision aids to enhance or expand upon clinical judgment, rather than as a competing approach.⁷⁹ As Munro (1999) observed, "...Errors can be reduced if people are aware of them and strive consciously to avoid them. The challenge is to devise aids to reasoning that recognize the central role of intuition and do not seek to ignore or parallel it but, by using our understanding of its known weakness, offer ways of testing and augmenting it."⁸⁰

Conclusion

This structured review of the available research literature on instruments of risk and safety assessment in child welfare suggests that actuarial models have stronger predictive validity and inter-rater reliability than do consensus-based models. This review was limited by: 1) the lack of studies on decision points other than initial investigation, 2) the variability in definitions and measures across studies, and 3) the relatively small number of studies examining risk assessment instruments. Nonetheless, the findings of this structured literature review should be useful to agencies and practitioners when considering the various approaches to risk and safety assessment in child welfare.

Notes

1. It is not clear which approach is more commonly used, as there is no national database regarding risk assessment approaches used by states. As of 1996, one report found 15 states used the Illinois CANTS 17B model or some derivation, 4 used the CARF system, and 4 used WARM or some derivation, all consensus-based models.⁸¹ However, an increasing number of states are using the actuarial model developed by the Children's Research Center.
2. Studies were excluded if they did not assess a particular instrument. In addition, only instruments for which there were at least two studies conducted are included in this Executive Summary.
3. Comparisons between instruments are challenging: not all instruments have been assessed for their validity, reliability, outcomes, or use with various

racial/ethnic subpopulations; and existing studies use different research strategies and different definitions of "maltreatment recurrence."

Endnotes

- ¹ Wald & Woolverton (1990)
- ² Wald & Woolverton (1990)
- ³ Rycus & Hughes (2003)
- ⁴ Kahneman & Tversky (1982)
- ⁵ Kahneman & Tversky (1982)
- ⁶ Grove & Meehl (1996)
- ⁷ Scheurman, Rossi & Budde (1999)
- ⁸ Munro (1999)
- ⁹ Munro (1999)
- ¹⁰ English (1999)
- ¹¹ English (1999)
- ¹² English (1999)
- ¹³ Doueck, English, DePanfalis & Moote (1993); English (1999)
- ¹⁴ Nasuti & Pecora (1993)
- ¹⁵ Rycus & Hughes (2003) p.13
- ¹⁶ Rycus & Hughes (2003)
- ¹⁷ Rycus & Hughes (2003)
- ¹⁸ Wald & Woolverton (1990)
- ¹⁹ McDonald & Marks (1991)
- ²⁰ Rycus & Hughes (2003)
- ²¹ Ereth, Johnson & Wagner (2003)
- ²² Grove & Meehl (1996); Schwalbe (2004)
- ²³ English & Graham (2000)
- ²⁴ Rycus & Hughes (2003); Schuerman et al. (1999)
- ²⁵ Baird & Wagner (2000)
- ²⁶ Camasso & Jagannathan (2000)
- ²⁷ Camasso & Jagannathan (1995) p.182

-
- ²⁸ English & Graham (2000)
- ²⁹ Baird, Wagner, Healy & Johnson (1999)
- ³⁰ English, Marshall, Brummel & Orme (1995)
- ³¹ Baird & Wagner (2000)
- ³² Baird & Wagner (2000)
- ³³ Camasso & Jagannathan (1995) p.182
- ³⁴ Baird et al. (1999)
- ³⁵ Baird & Wagner (2000)
- ³⁶ Doueck et al. (1993)
- ³⁷ Kolko (1998)
- ³⁸ Doueck Levine & Bronson (1993)
- ³⁹ Fuller, Wells & Cotton (2001)
- ⁴⁰ Fluke, Edwards, Bussey, Wells & Johnson (2001)
- ⁴¹ Garnier & Nieto (2002); Nieto & Garnier (2001)
- ⁴² Fluke et al. (2001)
- ⁴³ Johnson (2004)
- ⁴⁴ Baird & Wagner (2000)
- ⁴⁵ Johnson (2004)
- ⁴⁶ Baird et al. (1999)
- ⁴⁷ Loman & Siegel (2004)
- ⁴⁸ Wagner, Hull & Luttrell (1995)
- ⁴⁹ Baird & Wagner (2000)
- ⁵⁰ Baird, Ereth & Wagner (1999)
- ⁵¹ Johnson (2005)
- ⁵² Johnson (2005)
- ⁵³ Loman & Siegel (2004)
- ⁵⁴ Loman & Siegel (2004)
- ⁵⁵ Doueck et al. (1993)
- ⁵⁶ Lyle & Graham (2000)
- ⁵⁷ Kolko (1998); Lyle & Graham (2000)
- ⁵⁸ Hollinshead & Fluke (2000)
- ⁵⁹ Costello (1995); Depanfilis (1996)
- ⁶⁰ Albers & Roditti (2004); Depanfilis (1996); Rycus & Hughes (2003); Sheets (1992)
- ⁶¹ Sheets (1992)
- ⁶² Munro (2004)
- ⁶³ Tumlin & Geen (2000)
- ⁶⁴ Munro (1999)
- ⁶⁵ Wald & Woolverton (1990)
- ⁶⁶ Rycus & Hughes (2003); Zuravin, Orme & Heger (1995)
- ⁶⁷ American Federation of State County and Municipal Employees (1999); Kanani, Regehr & Bernstein (2002)
- ⁶⁸ Curran (1995) p.17
- ⁶⁹ Curran (1995); Davidson (1991)
- ⁷⁰ Curran (1995); Davidson (1991)
- ⁷¹ Curran (1995); Davidson (1991); Rycus & Hughes (2003)
- ⁷² Nohejl, Doueck & Levine (1992)
- ⁷³ Child Welfare Services Stakeholders Group (2003)
- ⁷⁴ Rycus & Hughes (2003)
- ⁷⁵ Lyons, Doueck and Wodarski (1996); Lyons et al. (1996)
- ⁷⁶ Rycus & Hughes (2003); Wald & Woolverton (1990)
- ⁷⁷ Doueck et al. (1993)
- ⁷⁸ Ereth et al. (2003) p.3
- ⁷⁹ Ereth et al. (2003); Fuller et al. (2001); Munro (1999)
- ⁸⁰ Munro (1999) p.756
- ⁸¹ Lyons et al. (1996)

References

- Albers M. & Roditti M. (2004). *Management values and decision-making tools: Nuts and bolts of risk assessment in California*. Bay Area Academy, Berkeley CA.
- American Federation of State County and Municipal Employees (1999). *Liability and child welfare workers*. Available online at: <http://www.afsane.org/publications/child/cww99205.htm>.
- 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.
- Camasso, M. J. & Jagannathan, R. (1995). Prediction accuracy of the Washington and Illinois risk assessment instruments: An application of receiver operating characteristics curve analysis. *Social Work Research*, 19(3), 174-183.
- Child Welfare Services Stakeholders Group (2003). *CWS Redesign: The future of California's child welfare services - final report*. State of California Health and Human Services Agency, Department of Social Services.
- Cicchinelli, L. (1990). Risk assessment: Expectations, benefits and realities. *Fourth National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Collins, B., Shearer, J. M. & Annin, J. B. (1987). Risk assessment development and implementation in a rural State. *National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Costello, T. (1995). Why is it so hard to implement risk assessment? *Ninth National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Curran, T. F. (1995). Legal issues in the use of CPS risk assessment instruments. *The APSAC Advisor*, 8(4), 1-4.
- Davidson, H. (1991). Risk assessment and the law: Unsettled issues. *Fifth National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Dawes, R.M. (1994). *House of cards: Psychology and psychotherapy built on myth*. New York NY: The Free Press.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision-making. *American Psychologist*, 34, 571-582.
- DePanfilis, D. (1996). Implementing child mistreatment risk assessment systems: Lessons from theory. *Administration in Social Work*, 20(2), 41-59.
- Doueck, H.J., English, D.J., DePanfalis, D. & Moote, G.T. (1993). Decision-making in child protective services: A comparison of selected risk assessment systems. *Child Welfare*, 72(5), 441-452.
- Doueck, H.J., Levine, M., & Bronson, D.E. (1993). Risk assessment in child protective services: An evaluation of the Child at Risk Field System. *Journal of Interpersonal Violence*, 8(4), 446-467.
- Downing J. D. Wells S. J. & Fluke J. (1990). Gatekeeping in child protective services: A survey of screening policies. *Child Welfare*, 69(4), 357-369.
- English, D. J. (1999). Evaluation and risk assessment of child neglect in public child protection services. In H.Dubowitz (Ed.) *Neglected children: Research practice and policy* (pp. 191-210). Thousand Oaks CA: Sage Publications Inc.
- English, D. J. & Graham, J. C. (2000). An examination of relationships between children's protective services social worker assessment of risk and independent LONGSCAN measures of risk constructs. *Children and Youth Services Review*, 22(11/12), 897-933.
- Ereth, J., Johnson, K. & Wagner, D. (2003). *New Mexico Children Youth and Families Department: Foster Provider Risk Assessment Study*. Children's Research Center.
- Fluke J. Edwards M. et al. (2001). Reducing recurrence in child protective services: Impact of a targeted safety protocol. *Child Maltreatment*, 6(3), 207-218.
- Fuller T. L. Wells S. J. & Cotton E. E. (2001). Predictors of maltreatment recurrence at two milestones in the life of a case. *Children and Youth Services Review*, 23(1), 49-78.
- Garnier, P.C. & Nieto, M. (2002). *Illinois Child Endangerment Risk Assessment Protocol evaluation: Impact on short-term recurrence rates – year six*. Children and Family Research Center, School of Social Work, University of Illinois at Urbana-Champaign.
- Gleeson J. P. (1987). Implementing structured decision-making procedures at child welfare intake. *Child Welfare*, 66(2), 101-112.
- Grove W. M. & Meehl P. E. (1996). Comparative efficiency of informal (subjective impressionistic) and formal (mechanical algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology Public Policy and Law*, 2(2), 293-323.
- Hollinshead D. & Fluke J. (2000). What works in safety and risk assessment for child protective services. In M. Kluger G. Alexander and P. Curtis (Eds.), *What Works in Child Welfare* (pp. 67-74). Washington D.C.: CWLA Press.
- 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*. Alameda County Social Services Agency. Unpublished report.
- Johnson, W. (2005). Effects of a research-based risk assessment on racial/ethnic disproportionality in service provision decisions. In D.M. Derezotes, J. Poertner, & M.T. Testa (Eds.), *Race Matters in Child Welfare* (pp. 147-162). Washington D.C.: Child Welfare League of America.

- Kahneman, D. & Tversky, A. (1982). On the psychology of prediction. in D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 50-98). New York, NY: Cambridge University Press.
- Kanani, K., Regehr, C. & Bernstein, M.M. (2002). Liability considerations in child welfare: Lessons from Canada. *Child Abuse and Neglect*, 26, 1029-1043.
- Loman L. A. & Siegel G. L. (2004). *An evaluation of the Minnesota SDM Family Risk Assessment*. Available from: <http://www.iarstl.org>.
- Lyle C. G. & Graham E. (2000). Looks can be deceiving: Using a risk assessment instrument to evaluate the outcomes of child protection services. *Children and Youth Services Review*, 22(11/12), 935-949.
- 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.P. & Marks, J. (1991). A review of risk factors assessed in child protective services. *Social Services Review*, 65, 112-132.
- Munro E. (2004). A simpler way to understand the results of risk assessment instruments. *Children and Youth Services Review*, 26(9), 873-883.
- Munro E. (1999). Common errors of reasoning in child protection work. *Child Abuse and Neglect*, 23(8), 745-758.
- Nasuti J.P & Pecora P.J. (1993). Risk assessment scales in child protection: A test of the internal consistency and inter-rater reliability of one stateside system. *Social Work Research and Abstracts*, 29(2), 28-33.
- Nieto M. & Garner P. C. (2001). *Illinois Child Endangerment Risk Assessment Protocol evaluation: Impact on short-term recurrence rates- Year five*. Urbana, IL: Children and Family Research Center.
- Nohejl C. Doueck H. J. & Levine M. (1992). Risk assessment implementation and legal liability in CPS practice. *Law & Policy*, 14(2 & 3), 185-203.
- Repetosky C. & Bailey D. (1988). The risk "fit": Integrating risk assessment with case management and practice decisions in child protection and child welfare services. *Second National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Rycus J.S. & Hughes R.C. (2003). *Issues in risk assessment in child protective services: Policy white paper*. Columbus, OH: North American Resource Center for Child Welfare, Center for Child Welfare Policy.
- Schuerman, J., Rossi, P.H. & Budde, S. (1999). Decisions on placement and family preservation: Agreement and targeting. *Evaluation Review*, 23(6), 599-618.
- Schwalbe C. (2004). Re-visioning risk assessment for human service decision making. *Children and Youth Services Review*, 26(6), 561-576.
- Sheets D. (1992). How Texas Got S.M.A.R.T.: A description of the rapid application design and development process used by Texas to design and implement a statewide risk assessment system. *Sixth National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Thompson L. Brown J. & Pecora P. J. (1989). Implementation of Risk Assessment Systems: Training and Quality Assurance Issues. *Third National Roundtable on CPS Risk Assessment*. San Francisco, CA: American Public Welfare Association.
- Tumlin K.C. & Geen R. (2000). The decision to investigate: understanding state child welfare screening policies and practices. *New Federalism Issues and Options for States, Series A No.38*. Washington D.C.: The Urban Institute.
- Tversky A. & Kahneman D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Wald M.S. & Woolverton M. (1990). Risk assessment: The emperor's new clothes? *Child Welfare*, 69(6), 483-511.
- Zuravin S. J. Orme J. G. & Hegar R. L. (1995). Disposition of child physical abuse reports: Review of the literature and test of a predictive model. *Children and Youth Services Review*, 17(4), 547-566.

This structured review of the literature is sponsored by the eleven county social service agencies who are members of the Bay Area Social Services Consortium (BASSC) located in the Northern California region surrounding the San Francisco Bay. It is partially funded by the Zellerbach Family Foundation and the VanLobenSels/RembeRock Foundation and is part of a BASSC Series on Evidence For Practice developed by the following members of the BASSC Research Response Team:

Michael J. Austin, PhD, BASSC Staff Director and Professor of Management and Planning, School of Social Welfare, University of California, Berkeley

Amy D'Andrade, PhD, BASSC Research Director

Kathy Lemon, MSW, Doctoral Research Assistant

Amy Benton, MSW, Doctoral Research Assistant

Belinda Chow, Masters Research Assistant

Carolina Reyes, Undergraduate Research Assistant

The Center for Social Services Research
School of Social Welfare
University of California, Berkeley

BASSC Evidence For Practice Series www.bassc.net

1. Understanding and Addressing Racial/Ethnic Disproportionality, July 2005
2. Risk and Safety Assessment in Child Welfare: Instrument Comparisons, July 2005
3. Understanding and Measuring Child Welfare Outcomes, July 2005
4. Assessing the Outcomes of Substance Abuse Treatment Programs for Child Welfare Populations, Spring 2006
5. Assessing the Outcomes of Parent Education Programs for Child Welfare Populations, Spring 2006
6. Family Assessment in Child Welfare: Instrument Comparisons, Spring 2006
7. Impact of Organizational Change on Organizational Culture: The Prospects for Evidence-based Practice, Spring 2006
8. Dissemination and Utilization of Evidence for Promoting Evidence-based Practice, Spring 2006