

## Risk Assessment: The Emperor's New Clothes?



MICHAEL S. WALD  
MARIA WOOLVERTON

*This article reviews the risk assessment procedures and instruments that are being used by child protection agencies. Although supportive of the concept of risk assessment, the authors argue that all current instruments have major methodological deficiencies that limit the utility of such instruments as a means of predicting future abusive or neglectful behavior. The authors suggest how risk assessment decisions should be made at each stage of a child protection investigation.*

Over the last few years, child welfare agencies throughout the United States have been adopting risk-assessment methodologies in an effort to improve

---

*Michael S. Wald, J.D., is Professor of Law, Stanford University, Stanford, CA. Maria Woolverton is Graduate Student in Psychology, Stanford University, Stanford, CA. This article is a revised version of a speech given by Michael Wald at the Third National Roundtable on Risk Assessment, August, 1989. Many persons have made helpful suggestions. The authors would especially like to thank Chris Baird, Diana English, Will Johnson, Jonathan Koehler, Peter Pecora, and Dee Wilson. Kimberly Gordon, a Stanford graduate student, provided valuable research assistance. This research was supported in part by the Stanford Legal Research Fund, made possible by a bequest from Ira S. Lillick and by gifts from other friends of the Stanford Law School, and by funds from the Stanford Center for the Study of Children, Youth, and Families.*

their ability to protect abused and neglected children. As part of this process, a number of risk-assessment instruments have been developed for use by workers in the child protective system. The effort to develop such instruments is, in many respects, a positive occurrence. Risk-assessment procedures potentially can improve decision-making, facilitate internal supervision, and lead to more efficient resource allocation. Moreover, to properly develop a risk-assessment system, child welfare professionals would need to face directly the value judgments—often unstated—that underlie the child protection system.

Despite the promise, we believe that risk-assessment instruments have only limited utility at present. Many agencies have acted prematurely, implementing risk-assessment instruments that have not been adequately designed or researched.<sup>1</sup> It is not possible to make highly accurate predictions of risk with existing instruments. Unfortunately, some child protection services (CPS) agencies appear to be using risk-assessment instruments in an unjustifiable manner, given the limited knowledge base regarding the validity of these instruments. Moreover, we are concerned that many agencies are adopting risk-assessment instruments in lieu of addressing fundamental problems in existing child protection systems, such as the excessive number of inexperienced or incompetent workers and the lack of adequate resources. In fact, use of inadequately designed or researched risk-assessment instruments may result in poorer decisions, because workers will rely on mechanical rules and procedures instead of trying to develop greater clinical expertise.

In this article, we examine some of the deficiencies in the current approach to risk assessment, in the hope that agencies will reexamine how and when they should adopt a risk-assessment instrument. Our goal is to clarify the potential strengths and limitations of various risk-assessment procedures. We also discuss how risk-assessment procedures should be incorporated into child protection decision-making.

### The Logic of Risk Assessment

The movement toward risk assessment grows logically out of changes in the direction of child protection policies over the past 15 years. Before the 1970s, intervention by child welfare agencies and juvenile courts often was justified as a way of helping children who were exposed to "inadequate care." No specific harm was required, and a likelihood of future harm was not a prerequisite for court intervention. Child protection agencies were therefore not generally concerned with carefully assessing risk. Moreover, since the laws

defining juvenile court jurisdiction over abused and neglected children permitted intervention if the child was in an "unfit home," the parents were "unsuitable or neglectful," or other such vague and undefined terms, anyone interested in risk assessment could not know the behavior that needed to be predicted.<sup>2</sup>

More recently, many legislatures have made it clear that CPS intervention is justified only when a child has suffered (or is highly likely to suffer due to current parent behavior) specific types of maltreatment.<sup>3</sup> The purpose of intervention is to prevent further maltreatment, not just to provide general services to the family. Legislatures have begun to define abuse and neglect more carefully, in terms of particular physical harms or injuries, sexual conduct involving a child, or severe emotional damage.<sup>4</sup> When laws provide more specific definitions of physical or emotional harm and sexual abuse, it becomes possible to ask, "How likely is it that a parent who has once harmed a child *in a specific way* will harm that child again in a specific way?"<sup>5</sup>

Changes also have taken place in the legal standards governing removal from their homes of abused or neglected children, and in the standards for reunification of children with their biological parents. Under older laws, still in place in many jurisdictions, the standard for removal and for return was whether this action would be in the best interests of the child. The term best interests was undefined; thus there was no standard for future parental behavior that might be the focus of a risk-assessment procedure. Today, many states permit removal only if the child cannot be protected from future abuse if left at home (throughout this article the term "reabuse" will also mean continued neglect); reunification is mandated unless continued foster placement is necessary to protect the child from reabuse.<sup>6</sup> Since the legal question is whether the parents are likely to behave in a certain way (e.g., in ways that will cause physical injury to the child), it is at least theoretically possible to develop instruments that will help predict the likelihood of this behavior.

We support these legal changes and believe that child protection agencies properly view their task as one of risk assessment. Risk-assessment instruments can help workers make these decisions. In fact, in some areas of legal decision-making that involve predicting risk, such as decisions whether to release people from mental hospitals or correctional facilities, there is reason to believe that the risk predictions are best made by looking only at a set of specific factors, applied in an actuarial manner. This is because there are specific factors that, when used in a statistical or actuarial manner, are more likely to result in accurate predictions than are well-trained clinicians using clinical judgment.<sup>7</sup>

It is possible that if child protection agencies engage in a substantial amount

of research we will learn which factors are most predictive of future abuse; the time may even come when agencies will be justified in using an actuarial approach in making child protection decisions.

That time, however, is far away. Nevertheless, agencies throughout the country have put into operation risk-assessment systems before conducting the research necessary to validate the risk-assessment instrument. It is often unclear what purposes those instruments are meant to serve. Existing instruments have both theoretical and methodological flaws that raise questions about the accuracy of any predictions made by using these instruments. Until the conceptual issues are resolved, and the necessary research completed, risk-assessment instruments should be used very cautiously, primarily as a means of training workers, not as a mechanical basis for making decisions.

In the following sections we discuss the conceptual and methodological deficiencies in many risk-assessment systems. We then examine each of the major decision points in the child protection system and propose how risk-assessment instruments should be used in making these decisions. Finally, we discuss the relationship between risk assessment and the use of resources by child protection systems.

### **Assessing Risk-Assessment Procedures: Some Conceptual Issues**

It is difficult to do a general critique of risk assessment that applies to each and every current system. Throughout the country many different systems are being used in various ways. The utility of a given approach will depend upon the purpose for adopting risk assessment, the use being made of the information, as well as the quality of the risk-assessment tool. We will not attempt to evaluate every current system. Instead, we raise some issues that are applicable to most systems, recognizing that, with regard to some of our criticisms, there are exceptions to our descriptions. We begin by identifying three conceptual issues that must be resolved by anyone trying to develop a risk-assessment system.<sup>8</sup>

#### *Definitional Confusions*

The first conceptual problem is a lack of clarity about the meaning of risk assessment. The term is used inconsistently; it has been applied to at least three quite different procedures. The first, and in our mind proper, meaning is *a process for assessing the likelihood that a given person (usually a parent) will harm a child in the future*. In essence, risk assessment involves making a probability statement regarding a future event. Thus, someone evaluating

a risk-assessment instrument should look at the theoretical justifications and the empirical evidence supporting the validity of the predictive factors used in the instrument.

For purposes of child protection decisions, both the likelihood of any future abuse and the likely severity of any harm that would result from such behavior are important. Decision-makers should consider both factors in deciding whether and how to intervene. A risk-assessment instrument might help with both predictions. It must be recognized, however, that assessing the probability that a parent will engage in abusive behavior and assessing the likelihood of a child being seriously harmed requires asking separate questions (and, of course, defining "serious"). The probability that a child will suffer any future harm may be quite different from the probability that he or she will suffer serious future harm. As we discuss later, the predictive factors in current instruments relate to the likelihood of some harm occurring; they have not been shown to predict seriousness of future harm.

This use of the term risk assessment, that is, the prediction of future behavior, is central to most current risk-assessment procedures. In many states, however, the term risk assessment is used to describe procedures that do not involve making probability statements. Two quite different procedures are sometimes also called risk assessment. First, in some states, part of the risk-assessment process involves a procedure whereby the agency categorizes the seriousness of an existing injury to a child. For example, a risk-assessment instrument may instruct workers to categorize as high risk those cases that involve injuries to the face or injuries that entail the risk of death or physical impairment.

There are two possible reasons why an agency may require workers to rate the seriousness of the present harm; one reason is consistent with our definition of risk assessment, the other is not. Some agencies may want to rate the severity of the current injuries because they believe that the severity of the present harm is a valid predictor of the likelihood of any future harm or of serious future harm. If so, there is no conceptual problem; the only issue is whether seriousness of present harm is a good predictor. We believe that the little available research does not provide much support for using the severity of the present harm as one of the factors predictive of future harm.<sup>9</sup> This is, however, an empirical issue that can be studied.

In other agencies, however, the severity of the present harm, in and of itself, is considered relevant to decision-making, regardless of whether it is a good predictor of future harm. Child welfare officials might decide, for example, that intervention is necessary in virtually every case involving certain serious injuries, even if the apparent risk of recurrence is low. Such a policy

might be justified on the grounds that once a parent has injured a child seriously, the agency does not want to take any chance of reinjury. To guide these decisions, the agency therefore must categorize those injuries it considers serious. As we discuss below, we would support basing some decisions largely on the seriousness of the current injury. Thus, it is reasonable for agencies to create taxonomies, classifying the seriousness of injuries. But it must be recognized that decisions based on such taxonomies do not turn on the probability of future behavior. Therefore, the process should not be called risk assessment.

Finally, in some states the term risk assessment is sometimes used interchangeably with the term needs assessment. Needs-assessment instruments require workers to focus on the aspects of a family's environment or behavior that are to be the focus of service intervention, such as poor family communication or high parental conflict. These factors are not, however, necessarily predictive of future abuse, although they may be related to the future well-being of the adults and/or children.<sup>10</sup> Moreover, there are some variables that may be predictive of reabuse but are not relevant to a needs assessment. A history of criminality and the number of previous referrals, for example, have each been found to be good predictors, that is, they are correlated with reabuse. But these factors do not cause the future conduct and they are not subject to change. They are predictive because they are, in effect, proxies for some unknown variables that are actually causal.<sup>11</sup>

Thus a needs-assessment instrument should include different factors than a risk-assessment instrument. Although there would be overlap, a needs-assessment instrument should focus on factors that cause the undesirable behavior and are amenable to agency services. Needs assessments can be useful in setting up a casework plan, if intervention is otherwise justified, because the child has been abused or neglected. These instruments, however, cannot be used to justify decisions to intervene, to remove a child, or to reunify a family.

The confusion of terms just described is disturbing. Very different procedures are necessary to develop instruments for predicting future parent behavior than to develop taxonomies of seriousness, or needs-assessment tools. As just indicated, the appropriate uses of each type of instrument also vary. Because it is often difficult to determine what an agency means when it says that it has a risk-assessment system, we have been unable to determine whether the instruments being used are likely to be valid means of accomplishing each agency's goals. It is critical that administrators, workers, and judges understand, at a conceptual level, the appropriate uses of each process—that is, risk assessment, rating levels of seriousness, and needs as-

essment. Adopting a common and precise meaning for each term is an essential first step.

#### *Application of Same Predictive Factors to Varied Decisions*

The types of decisions that are thought to be enhanced through the use of risk-assessment instruments vary enormously. Agencies are using risk-assessment instruments in making some or all of the following decisions: (1) whether, and how soon, to investigate a report of abuse or neglect; (2) whether to substantiate such a report after investigation; (3) whether, in substantiated cases, to divert a case to an alternative system, to file a court petition, or even to close the case entirely; (4) whether to remove a child temporarily during the course of an investigation; (5) how much service to give a family that has been brought within the protective services system; (6) whether to remove a child into long-term foster care following an adjudication that the child has been abused and neglected; (7) whether to return to a parent a child who has been in foster care, or in the alternative, whether to seek termination of parental rights; and (8) whether to close a case.

In evaluating the utility of a given risk-assessment procedure for making a specific decision, two questions must be asked. First, is the particular type of decision amenable to risk assessment? Second, is the instrument being used appropriate for making that type of decision? With regard to the first question, we believe, for reasons discussed later, that some decisions—for example, whether to substantiate a case—can never be made on the basis of a risk-assessment instrument. The second question raises different issues. Many agencies are using the same instruments to make many different decisions. Yet, as we discuss below, there is no theoretical or empirical basis for applying the same instrument for each of the varied predictions described above. It seems almost certain that different instruments will be necessary to make different decisions.

#### *What Function Is the Instrument Serving?*

From the literature, it appears that agencies are using risk-assessment instruments in several ways. Some agencies are using them purely as training devices, to get workers to apply more consistently whatever clinical criteria they use in making individual decisions, and/or to get all workers within an agency to focus on those factors that are believed, on the basis of research or expert judgment, to be most relevant to the decisions that are being made. Although the workers are taught to focus on particular factors, they are not bound to make decisions based on an individual's risk score, nor are they

limited to considering only those factors in the instrument. In these agencies, risk-assessment instruments function basically as a means of improving clinical judgment.

In other agencies, these instruments apparently are being applied, on a case-by-case basis, to make decisions in a quasi-actuarial or statistical fashion. In some states, workers add up the number of risk elements and base their decision about the amount of risk on the number of negative factors.<sup>12</sup> This may, in fact, occur in most agencies, even when it is agency policy to treat the instrument only as a guide for clinical judgment, since it is likely that workers, especially new or less competent workers, will apply the instruments in a mechanical fashion in order to avoid criticism and possible legal liability, or because they recognize that they lack the skills for making clinical judgments.

The use of instruments as a statistical tool raises different problems than their use as a way of focusing clinical judgment. Both the way workers think about the instruments and portray them in court proceedings will be altered. It is, therefore, very important in evaluating a risk-assessment instrument to focus on how it is being used.

In sum, because of the substantial variations in the nature of the instruments being used, the purposes for which they are being used, the decisions for which they are being used, and the research base supporting any given instrument, our discussion of the strengths and weaknesses of risk assessment must be general. Nevertheless, we believe that there are some important generalizations that can be made regarding all current instruments and procedures.

### **Limits on the Utility of Current Risk-Assessment Systems**

#### *What a Good Risk-Assessment System Must Do*

To understand the limitations of existing instruments, it is first necessary to understand what it means to engage in risk assessment. In essence, risk assessment entails using known information about a given person or situation to estimate the likelihood that the person will engage in a particular behavior or behaviors in the future. For such a procedure to have validity, there must be adequate evidence that the identified factors are related to the chance that the person will engage in the conduct. Second, it is essential that the factors be measured accurately; for example, it is no good to know that current drug use is predictive of future behavior if we cannot tell whether a person is using drugs.

There is a third element that is critical in evaluating the utility of a risk-assessment system designed for improving child protection decisions, an element that is frequently ignored. Under the laws of most states, and the mandate of Public Law 96-272, the goal of protective intervention is not simply to protect presently abused or neglected children from future reabuse. If that were the case, the goal could be met by simply placing all abused and neglected children in foster care.<sup>13</sup> Public policy dictates, however, that the goal of intervention is to try to prevent further maltreatment while keeping families together, to the degree possible. When children are removed, the goal generally is to work for reunification. In some states, agencies are required to intervene in the least intrusive way possible.

Therefore, in the context of child protection services, a fully adequate risk-assessment procedure must address the following question: "How high is the risk of reabuse if the parent receives a particular treatment or set of services?" In fact, in light of current laws, the appropriate questions might be: "Under what conditions is the parent *unlikely* to reabuse a child? Is supervision of the family (plus the provision of services) necessary to make such behavior unlikely? Is it sufficient?" If the behavior is unlikely even without services, then any intervention may be inappropriate. If it is likely even with supervision and services, then removal usually may be necessary, depending on the seriousness of the potential harm and the availability of adequate placements. Similar questions must be asked regarding reunification, closing of open cases, and termination of parental rights. Since the availability of services, treatment, or monitoring will alter the risk posed by a given individual, a risk-assessment instrument is truly useful only if it identifies the likelihood of reabuse *given specific interventions*.

Development of an adequate risk-assessment system is not merely a methodological task, however. Ultimately, risk assessment requires making value judgments. To begin with, it is necessary to specify the types of harm that are to be predicted. Abuse and neglect can be defined in any way the users of the instruments wish, but they must be defined; there can be no risk assessment unless one knows the risk that is being assessed. Obviously, the validity of the instrument or predictive factors must be measured in terms of the ability to predict whether the parents will abuse or neglect their children in the ways indicated in the definition.

In addition, an acceptable risk-assessment system must specify the criteria for determining the relationship between a given level of risk and the decisions that must be made by the child protection system—for example, whether to intervene, whether to remove a child, and so on. All risk assessment involves estimating probabilities. Child protection workers must assess (or guess) the

probability that a given behavior (e.g., the parent will hit the child) or given outcome (e.g., the child will sustain a life-threatening injury) will occur in the future. Risk-assessment instruments may help maximize the accuracy of these predictions, but even the best prediction instruments will never be 100% accurate. There will always be false positives—people who are labeled as likely to engage in the behavior who will not engage in it, and false negatives—people labeled as unlikely to commit certain acts who will commit those acts. In fact, there is a trade-off in the application of any instrument in the ratio between false positives and false negatives; in general, limiting the number of false positives increases the proportion of false negatives.<sup>14</sup>

The level of risk that one is willing to accept always involves making value judgments. Both types of errors involve serious consequences: children may be placed in foster care who would have been safe at home or children may be reinjured because the family was mistakenly considered low risk. How many false positives and false negatives are acceptable? It is unclear who, if anybody, is making these value judgments or how they are being made.<sup>15</sup> It is critical that judges and agency administrators openly acknowledge that the ultimate decisions involve value judgments and that a risk-assessment instrument merely tries to quantify the risk.

When decision-makers face these value judgments, they should be aware of a major limitation of all current risk-assessment systems. To make decisions sensibly, it is necessary to know the general parameters of the trade-offs. This requires that the risk decision be quantified. Unfortunately, no present risk-assessment system specifies what is meant, in quantitative terms, by the terms high, medium, or low risk. Thus, we do not know if high risk means that there is a 10%, 30%, or an 80% probability that this family will, in fact, reabuse its children.<sup>16</sup> Nor do we know what it means with regard to the probability of severe injury. The same is true of the term low risk.

In fact, for reasons discussed below, no risk-assessment system is capable of providing such information, since the research necessary to make quantitative statements has not been done. The best that can be said for existing instruments is that they are able to rank cases, more or less accurately, along a risk continuum, without telling us how close the case is to either end of the continuum, or how much distance there is between cases with different rankings. Obviously, this means that these instruments provide quite limited information to help inform the value-decision process.

#### *Methodological Deficiencies in the Development of Present Instruments*

If present instruments cannot provide quantitative estimates of the probability of reabuse, what can they do? At a minimum, they should help decision-

makers focus on those factors that are most likely to be related to recurrence of maltreatment and they should enable cases to be ranked, however roughly, along a risk continuum. How good are they for these purposes?

As Dee Wilson, Coordinator of the Washington State Risk Assessment Project, wrote in 1988, "[To] date, risk-assessment models have been developed in an atheoretical or even antitheoretical professional milieu."<sup>17</sup> From a methodological perspective, no current instrument was developed in a scientifically adequate fashion, although some agencies are in the process of trying to validate instruments properly.<sup>18</sup> We will not detail all the methodological shortcomings of existing instruments; Pecora has recently written an excellent review.<sup>19</sup> We do want to identify some prevalent misconceptions about existing instruments.

To begin with, the factors relied on to predict future behavior in virtually all states were not derived from research regarding reabuse by parents. Ideally, an instrument would be used only after the agency had identified a number of potentially predictive factors of reabuse and then tested their predictiveness through prospective, longitudinal research. This would require identifying a group of parents who had abused or neglected their children and monitoring the family for a period of time to see which parents reabused their children and which factors predicted this behavior. There have been a few such studies in recent years, but these studies followed, rather than preceded, adoption of a risk-assessment instrument by the agencies that participated in the research.<sup>20</sup> Moreover, even these few completed studies are quite limited in scope. It is unknown whether the findings are generally applicable beyond the particular sample studied.<sup>21</sup> In addition, none of these studies examined the impact of services in reducing risk, which, as we have argued, is a critical element of an adequate system.

Thus, existing instruments basically reflect the best available guesses about which factors are most likely to be related to reabuse. In general, the factors were derived from two sources: the judgments of experienced workers as to which factors are most likely to be associated with reabuse, and literature reviews that examined those factors generally found to be associated with people who *initially* abuse or neglect their children. Both of the sources are suspect, and thus there is good reason to question the validity of a number of the factors contained in many current instruments.

It is not inappropriate to begin with factors suggested by experienced workers. It must be remembered, however, that until tested empirically, these factors cannot be considered to be scientifically validated. The suggested factors undoubtedly reflect each worker's sense of the factors that were present in cases where reabuse occurred. For example, a worker may believe that in

her or his experience reabuse occurs more frequently if parents are substance abusers, isolated from other family members, or very young. Undoubtedly, many of these factors are predictive of reabuse. Because clinical judgment is not perfect, however, it seems highly likely that some false predictors have been included in current instruments.<sup>22</sup> In addition, workers will identify some factors that are spuriously correlated with the likelihood of future maltreatment. For example, a worker may notice that younger parents seem to have higher rates of future abuse. It may be, however, that the critical factor is really economic status, or family disorganization, and that younger parents are generally poorer or more disorganized. A better-off or better organized younger parent may be no more likely to reabuse than an older parent.

Relying on factors derived from literature reviews of factors associated with initial abuse and neglect also is problematic.<sup>23</sup> First, the evidence about the causes of initial abuse and neglect is far from definitive. Many of the studies were methodologically weak. Moreover, few studies examined neglect; even fewer pertained to sexual abuse. Thus, the research gives us little help in developing instruments predictive of continued neglect or the recurrence of sexual abuse.<sup>24</sup>

Most importantly, the evidence derived from literature reviews is of only limited relevance to the risk-assessment decision. These reviews generally examined studies that compared people who had abused their children with nonabusers. These studies help identify factors that predict who, among the general population, will abuse their children. The prediction confronting CPS workers, however, is not which parents will abuse or neglect their children initially; it is which parents will reabuse their children. The factors that are predictive of initial abuse or neglect are not necessarily the same factors as those correlated with future maltreatment. In fact, since all of the parents being assessed have already abused or neglected their children, most cases should have many of the factors associated with initial abuse, otherwise the research is wrong. Risk assessment involves predicting which subset of this population will continue to harm their children. The deficiency of current instruments is even greater with regard to predictions of the likely severity of any future harm. There is virtually no clinical or empirical basis for identifying those factors that are likely to be associated with severity. To our knowledge, no instruments are designed to assess the probability that a child will suffer serious injury.<sup>25</sup>

Since the factors most predictive of recurrence or of severity are unknown, many workers may be inclined just to add up the number of risk factors and assume that people with more such factors pose a greater risk of reabuse. Many instruments, in fact, instruct the workers to calculate a total score and

to label cases as high, medium, or low risk based on the number of risk factors present. Even if one assumes that the factors contained in the instrument generally had predictive validity, this procedure is unjustified.

To begin with, without research we have no idea where cut points should be—that is, how many factors must be present before the risk is increased?<sup>26</sup> Moreover, an additive model is unlikely to accurately predict risk in this context. First, many of the factors contained in current instruments probably are intercorrelated—that is, when one of the factors (for example, a young parent) is present, certain other factors (for example, low income) also tend to be present. To the extent that this is true, it will produce systematic bias in judgments about risk. Adding factors that are correlated with each other may result in predictions of higher levels of risk than are actually warranted.<sup>27</sup> In addition, some factors may be much more important than others; some may come into play only in the presence of others; and the effects of some factors may be multiplicative, not additive. At present, such factors cannot be identified.

Most advocates of risk assessment recognize that adding up the number of risk factors is not appropriate; they assert that it is the interaction of factors that matters. Yet nobody is able to point out which interactions of factors make a difference, making it impossible to draw any useful practice guidelines from this constant pointing out that it is the interaction of factors that matters. Although there are some promising starts in developing models that specify interaction effects, it is extremely difficult to conduct research that has a sample size large enough to permit the necessary statistical analysis.

Any predictions based on these instruments are further suspect if the same instrument is used in cases of physical abuse, sexual abuse, and neglect. Both the theoretical literature and the sparse empirical literature indicate that different factors may be predictive depending on the type of behavior being predicted. In addition, a large number of cases involve multiple types of maltreatment. The factors that contribute to continued maltreatment in such situations may still be different.

Another type of problem also undermines the utility of many present instruments: the reliability of the instruments. As several researchers have noted,<sup>28</sup> in many states the items that are included in the instrument are not rated consistently by various workers; in fact, interrater reliability is below acceptable levels for most instruments. This is not surprising given the vagueness of many items. For example, one risk-assessment instrument asks workers to decide whether there are "mild discrepancies or inconsistencies in the parent-child relationship," or the "child evidences an anxious or disturbed attachment to the parent." What do these terms mean? How are they measured?

Are they relevant beyond infancy? These questions plague developmental psychologists. How can a worker in the field reliably assess these factors? These judgments would be difficult for experienced, well-trained professionals. They are impossible for inexperienced, untrained workers. Until there is satisfactory interrater reliability for a given instrument, it remains an unacceptable means for making critical decisions.

Fortunately, many agencies are aware of these problems and are trying to improve their instruments. However, we have one concern that is not, to our knowledge, being addressed. As noted, the critical question in child protective cases is how will a given parent respond to treatment, not how many risk elements can be identified. In deciding whether to divert a case, petition a court, or remove or return a child, a worker must estimate the probability that the parent will respond to treatment. Yet, virtually no research examines the impact of different intervention strategies on people with different risk profiles. Research that identifies factors associated with initial abuse or neglect tells us nothing about this question. Neither does the limited research regarding reabuse, since these studies have not controlled for treatment.<sup>29</sup>

Finally, it must be remembered that no present instrument is able to make statistical predictions regarding the probability that a given individual will reabuse a child. Therefore these instruments should not be considered as providing a "scientific" alternative to clinical judgment, however unscientific such judgments may be. Developing numerical probabilities requires substantial research. Unfortunately, it seems unlikely that such research will be available in the foreseeable future.

#### *Using the Same Instrument for Multiple Purposes*

As previously indicated, many agencies are using the same set of risk factors, actually the same instrument, to make a number of different types of decisions: whether to divert a case following substantiation; whether to remove a child during the investigation stage; whether to place a child in foster care in an adjudicated case; whether to reunify a child with her or his family.<sup>30</sup> This is inappropriate. Theoretically, it seems unlikely that the same set of factors is equally predictive of future behavior under quite different conditions; neither is there research evidence that would justify this procedure. Although some factors may be predictive at all points in time, other factors will be useful only for particular decisions.

To give one illustration, the degree to which a parent cooperated with a caseworker while a child was in long-term foster placement may be quite predictive of the risk of reabuse following reunification. In contrast, the fact that a parent is uncooperative at the point of initial investigation may not be

at all predictive of the likelihood of reabuse prior to adjudication; uncooperativeness may merely reflect the parent's hostility over the investigation process or the way the individual worker approached the parent. (In fact, one benefit that may come from the development of risk-assessment instruments is knowledge that some factors commonly used in clinical judgment are not really predictive of future behavior—parental attitudes may be just such a factor.)

Therefore, separate instruments should be used for each decision. To develop separate instruments relevant to each type of decision, studies are needed that identify those factors that are predictive of reabuse pending adjudication, reabuse of children left at home following court adjudication, reabuse following reunification, and so on. Little such research is currently in progress.

#### *Use of Risk Assessment To Substantiate Cases*

The previous discussion concerned methodological issues that generally limit the utility of risk-assessment instruments, even when the instruments are being applied in otherwise appropriate situations. One decision, however, is not amenable to risk assessment for reasons totally independent of deficiencies in the instruments—the decision to substantiate an allegation that a child has been abused or neglected.

As we have noted, risk assessment is a tool for predicting future behavior; it is not a method for determining whether past behavior has occurred. Yet, to substantiate an allegation of abuse or neglect, a CPS worker must find evidence that a parent already has harmed the child or has created conditions that have a high likelihood of harming the child (e.g., leaving a young child unsupervised, having a home with exposed wires, broken glass, etc).<sup>31</sup> Under virtually all states' laws, statistical evidence that a person has characteristics associated with parents who might abuse or neglect their children is not a basis for legal intervention. The parent must have engaged in actual behavior that meets the legal definition of abuse or neglect.

The issue is not just one of legal definitions. The fact that a parent has already engaged in abusive or neglectful behavior is critical to the acceptability of the risk-assessment process. The accuracy of predictions is greatly increased because we are dealing with a group of known abusers, since the rate of reabuse is greater than the rate of initial abuse in the general population. Using the presence of risk factors as evidence that abuse or neglect has occurred, when this cannot be proven independently, would certainly result in a substantial number of errors. If an allegation of abuse or neglect cannot be substantiated but the family appears to be at risk of abusive or neglectful behavior, an offer of voluntary services may be appropriate. But coercive intervention is never justified solely on the basis of risk factors.<sup>32</sup>

## Implications

What are the implications of the deficiencies in existing instruments, and the limited research support for them? We do not believe that these instruments should be abandoned until the appropriate research is done. They can play a useful role in improving decision-making. Agency administrators must recognize, however, the limitations of these instruments and use them in a way that reflects these limitations. We recommend several courses of action.

### *Instruments Should Be Used as a Means of Improving Clinical Judgment, Not as an Actuarial Device*

Because risk-assessment instruments currently are embodiments of the best available clinical judgment, they should be treated as such. Even well-conceived instruments cannot be used in a statistical or actuarial manner. Instead, they should be used to teach workers to focus on particular factors, and not to consider other factors that might be arbitrary or irrelevant.<sup>33</sup> Workers should be taught that the factors in the instruments constitute a general set of guidelines to focus their thinking. Training should concentrate on improving clinical judgment, so that workers can make better educated guesses about whether the particular combination of factors seen in any given case is predictive of future harm. Agencies should be especially diligent in preventing inexperienced, less competent workers from using these instruments in a mechanical fashion, which they might do in order to avoid liability or criticism, or because they do not have the skills to use the information in a more sophisticated manner.

To prevent misuse of the instrument, workers should be instructed *not* to add up the number of risk factors. Instead, they should be required to describe, in writing, why they believe a given situation entails high or low risk. Workers also must be taught to ask the critical question: "What is necessary to prevent recurrence of the behavior?" In essence, they should come to see the risk-assessment protocol as a screening tool, helping them to perceive the kinds of resources or services that would be needed to deal with each of the risk factors, rather than just adding up risk factors to make a decision about removal or nonremoval, reunification or nonreunification. They must become sophisticated at judging intervention strategies; this is why agencies cannot settle for high worker turnover.

Third, agencies must ensure that workers receive good supervision. Risk-assessment procedures are, in fact, a form of supervision. They are important additions to, but not substitutes for, ongoing supervisory review.

Finally, we must continue to recognize the value of providing workers with the opportunity to exercise judgment. As one author has written:

Competence and sensitivity are themselves not static phenomena. Conscientiousness, dedication, idealism, openness, enthusiasm, willingness to listen and learn . . . are at least to some extent evoked by a sense of responsibility, by the realization that one has been entrusted with a great and important task. I can think of nothing more subversive to the . . . inner sense of responsibility than the notion that, to the greatest possible extent, all the important shots will be called by someone else because we don't believe in his or her competence and sensitivity.<sup>34</sup>

Wald, among others, has written about the dangers of giving workers too much discretion;<sup>35</sup> however, the dangers of giving too little discretion must also be recognized.

#### *Application to Specific Decisions*

*Whether, or how soon, to investigate.* Current risk-assessment instruments are not appropriate for use in connection with these decisions. First, all allegations that might constitute abuse or neglect, as defined by law in a given state, must be investigated. The likelihood of future injury cannot be known in advance of investigation; it is legally irrelevant in any case.

It has been argued to us that, since agencies faced with limited resources do not investigate all cases, it would be better for them to investigate cases that are more likely to be confirmed, since this group presumably is at greater risk of future maltreatment. Some researchers have developed instruments that predict confirmation rates.<sup>36</sup> Although confirmed cases probably are at greater risk, we believe that using risk instruments in this way is inappropriate. If legislators do not wish to provide adequate funds to allow for the investigation of all allegations of maltreatment, then the legislature should face up to the value issue and decide which types of cases do not warrant investigation. A legislature might choose to narrow the definition of maltreatment or to instruct agencies to focus on the seriousness of the alleged harm. It also might direct that priority be given to cases most likely to be confirmed, those most likely to respond to intervention (assuming either of these categories can be identified), or on some other basis. The critical point is that the issue is not purely a research problem; at core it is a value decision, which should be made explicit.

The issue of priority is more complex. The first question is again one of value—which cases deserve priority. A priority system could, for example, focus on the likelihood of confirmation, the probable seriousness of any maltreatment that might occur if investigation is delayed, or on cases that pose the greatest threat of imminent reabuse, regardless of seriousness.

We believe that the primary considerations should be the likelihood that the child will be maltreated in the near future and the potential seriousness of any injury. There are no studies that address these issues; thus current risk-assessment instruments are not useful in making priority decisions, if these are the criteria.<sup>37</sup> Priority must therefore be based on value judgments regarding the seriousness of the potential harm and commonsense decisions regarding the factors that relate to whether reabuse is imminent, such as evidence of frequent previous maltreatment, any known serious parental substance abuse or mental illness, and whether the alleged perpetrator still has access to the child. Since there is no research indicating which factors predict severity of future harm, it is reasonable to look at severity of the present injury and assume that past behavior is the best predictor of future behavior. The class of injuries or harms that are considered serious should be carefully defined; this is the appropriate use of taxonomies.

Many of these factors will not be known when a report is received. Therefore, first priority should be given to allegations of serious injury. We do not have suggestions regarding less serious injuries. We suggest that agencies conduct research aimed at identifying factors that predict which situations put the child most at risk of immediate harm; only if such factors cannot be identified might it be useful to try to identify those cases most likely to be confirmed.

*Whether to intervene, close, or divert.* When allegations of abuse or neglect are substantiated, a CPS worker must decide whether the risk of recurrence is substantial enough, or the harm serious enough, to justify agency intervention. As an alternative to intervention, a worker might close the case entirely or divert the case to an alternative treatment system. Risk-assessment procedures are, for the most part, appropriately used in making these decisions. The exception relates to seriousness; an agency might decide that some involvement is always needed in cases of more serious injuries, regardless of a parent's risk profile.<sup>38</sup> These judgments should be spelled out, and the value premises specified.

Based on present research, it may be especially appropriate to use risk assessment to screen out cases that involve less serious actual or potential harm and where the risk of reabuse is apparently low. Several studies, using risk-assessment instruments and some degree of random assignment to treatment and nontreatment groups, have reported very low recidivism rates for families classified as low risk that were diverted to alternative systems or whose cases were closed after only a brief intervention.<sup>39</sup> These studies must be viewed cautiously, since some of these studies followed the cases for only six months. Still, some agencies may unnecessarily tax their already inade-

quate resources by opening cases that probably can be diverted without jeopardizing the child's well-being.

We stress the need for accompanying research, however. The research base is not adequate to allow agencies to determine with confidence which families are truly low risk. Moreover, for diversion to work, good programs must be available outside the CPS system. The quality of the service provided by those alternatives must be regularly checked. Research should examine which alternatives work and which families do not respond. Without this research, the validity of decisions based on risk-assessment instruments will remain suspect.

Agencies might also try to get relatively low-risk parents to accept agency supervision voluntarily, as an alternative to invoking court jurisdiction. Although the level of risk might be used to decide not to seek court involvement, research has not so far indicated which low-risk cases should be closed, which diverted, and which encouraged to use voluntary services.

*Pretrial removal.* One of the most critical decisions facing CPS workers is whether to remove a child during the investigation process. The pertinent questions at this stage are whether the parent is likely to reinjure the child prior to trial, the potential seriousness of any reinjury, and whether the parents are likely to flee the jurisdiction. The needs of the child also may be relevant; for example, a child may need special medical care.

Current research tells us little about any of these issues. Thus, we would not recommend using current instruments to make this decision. Instead, we suggest that workers focus on the seriousness of the present injury and any clear-cut parental conditions, such as serious substance abuse or mental illness, that might influence their conduct in the short run, plus any previous pattern of abuse or neglect by the parent. In general, children who have been seriously injured—for example, broken bones, burns—should be removed because we do not want to risk reinjury of this magnitude even in one case; thus the probability of reinjury is less relevant (an exception may be cases where the perpetrator is no longer present). Most other factors found in present instruments are probably irrelevant to short-run predictions; some may even lead to poor decisions—for example, too much weight may be paid to parental attitude—which may be misleading or improperly assessed at this stage of the process.

In making this decision, the reasonable efforts mandate must also be considered. The agency is obligated to explore interventions that might alleviate the risk and the need for removal. These services often can eliminate the risk in neglect cases—at least on a temporary basis.

*Home supervision versus foster placement.* This is perhaps the most critical,

and most difficult, decision for the protective services system. It is also the decision most amenable to risk assessment. The little available research is most germane to this issue, since several studies have examined factors associated with reabuse of children left at home. In addition, this is the point in the process where the data about the family are most likely to be accurate and complete, since a full investigation will have been completed.

Risk-assessment instruments can be used to help workers focus on the factors that are appropriately considered. Yet, even in this area, there is not nearly enough evidence indicating which factors are predictive of reabuse to allow for mechanical applications of assessment tools. Clinical judgment is still critical to determining the likely impact of services, the parents' responsiveness to services, and the likely impact of various interventions on the child. These decisions are complicated—the risk to the child resulting from removal, as well as from nonremoval, must be considered. Because of the complexity of the judgment, agencies ought to consider establishing internal review committees to make these decisions in the more serious cases—perhaps in all cases in which removal is contemplated.

*Reunification.* The factors in most instruments probably are not particularly good at predicting successful reunification. Some of the factors frequently found in present instruments are irrelevant or hard to measure at the time of the reunification decision (for example, most factors dealing with the parent-child relationship do not make sense following placement, although it may be possible to assess interaction during visits). There is no research indicating what factors are most predictive; in fact, there is little information indicating the percentage of children who are reabused following reunification—information that is needed to determine the likely error rates associated with any prediction instrument. Reunification decisions require substantial clinical judgment.

#### *Revision of Instruments*

Even if restricted in the manner just suggested, current instruments should be redesigned to separate those factors that are basically value judgments about seriousness of harm from those factors that are thought to be predictive of reabuse. A separate set of rules ought to be adopted by the agency regarding how the potential seriousness of any future injury should be weighed, recognizing that this question is independent of the likelihood of its occurring.

Also, current instruments should be reevaluated to determine whether there really is a sound theoretical or experiential basis for each of the items now included. Those items that are questionable should be eliminated, and those that cannot be measured reliably should be revised or eliminated.

### *Conducting Research*

Obviously, from all we have said previously, we believe that if decision-making is really to be improved and risk assessment is to become a more scientific procedure, agencies must support, either within the agency or through connections with outside researchers, the type of research that is needed to make any risk-assessment procedure more useful. Over the long run, the development of more statistical bases for making judgments is certainly desirable. In the short run, faced with enormous budget problems and limited personnel, it is understandable that agencies would not choose to use limited resources for research.

These are not necessarily competing uses of resources, however. Money may be available for research that is not available for direct services. Carefully conceived research, such as has been done by the National Council on Crime and Delinquency in Alaska and by other organizations in other states,<sup>40</sup> will certainly improve decision-making and may ultimately lead to a point where risk-assessment instruments can be used in a more sophisticated fashion. Moreover, much of the data needed for prediction studies could be obtained by agencies as part of a comprehensive management information system. If agencies begin to collect information systematically about children and families throughout the period of agency involvement, we can expect much more and better research.

### **Risk Assessment and the Politics of Child Protection**

One of the most startling aspects of the risk-assessment movement is the rapidity of its adoption throughout the country—which might be contrasted with the considerably slower pace in the implementation of reasonable efforts procedures, home-based services, and permanency planning.

Why have agency administrators throughout the country latched on to risk assessment so rapidly? Undoubtedly, part of the reason is that risk assessment is being sold as critical to good practice. And, as we have noted, eventually it may be. At that time, it will deserve widespread adoption.

We are concerned that the rush to implement risk assessment is being propelled by less justifiable motives, as well. As the literature makes clear, risk assessment also is viewed as a way of coping with two problems that plague administrators—inadequate resources and incompetent personnel.<sup>41</sup> This is unacceptable. It is also seen as a way of avoiding agency liability for “incorrect” decisions. This is unlikely.

We recognize that it is often difficult to recruit and retain highly competent

workers. We also are sympathetic with concern about agency liability for "wrong" decisions. It is not surprising that administrators look for ways to minimize the negative effects of staff deficiencies or the lack of treatment resources. The use of risk-assessment procedures, however, cannot adequately compensate for inadequate staff or other resources.

There are a number of reasons why risk-assessment procedures cannot be used as a substitute for competent workers. Since no current risk-assessment instrument has been shown to make accurate predictions statistically, the judgment of workers remains an important element in most child protection decision-making. Moreover, even if more scientific instruments are developed, many of the factors or measures probably will require worker judgment (for example, parental cooperativeness and/or levels of functioning). Unless workers are competent and well trained, they will not even gather the basic information reliably, thereby reducing the validity of predictions.

Finally, and perhaps most importantly, risk assessment always should be tied to case planning.<sup>42</sup> The risk of a child being injured by a parent who has already injured the child (or who has engaged in dangerous behavior) depends heavily upon the kind of intervention that takes place following the determination that the child is at risk. Appropriate interventions will lessen the risk in some cases; inappropriate interventions will increase the risk. Only competent, trained workers will be able to make these decisions.

Risk-assessment instruments also should not be used to compensate for inadequate resources. The use of risk-assessment tools can lead to more efficient utilization of resources, but they will not enable agencies to provide adequate intervention if caseloads are excessive or treatment programs limited. We believe that it is ethically wrong for administrators to adopt risk-assessment procedures in lieu of pushing the political system for more resources. Politicians must be made to bear the consequences of their unwillingness to adequately fund the child protection system.

All children who come within the definition of abuse and neglect are legally and morally entitled to that level of services necessary to protect them from future harm.<sup>43</sup> Of course, when resources are inadequate to respond to each case or to provide appropriate services, allocation decisions must be made. We would focus on cases involving the most serious actual harm to children as a first priority. After that we would want to balance the seriousness of potential harm with the likelihood of its occurrence. Risk-assessment instruments can help inform that value judgment.<sup>44</sup>

We would not, however, imply in any way that this was a satisfactory outcome. The public and its legislators ought to be reminded regularly that as a result of inadequate resources children who require protection are not

getting it. Legislators ought to be forced to confront the value judgment about what level of risk is deemed acceptable, or not worthy of funds for intervention. In our view, agency administrators are obligated to carry on the advocacy necessary to obtain adequate resources and personnel so that they can meet the child-protection goals of the system.

Finally, we doubt that these instruments will protect agencies from liability. How will the agency justify the use of an improperly developed instrument? How will untrained workers be able to explain why they relied on the factors identified in the instrument? In addition, as we have previously noted, in the hands of unskilled workers, these instruments may actually produce worse decisions. Agencies can protect themselves from liability. Clear-cut guidelines, based on the best available knowledge, will help. But the best defense is a well-trained worker, who is able to explain the basis for her or his decisions, and whose work is regularly subjected to good internal review.<sup>45</sup>

### Conclusion

Over the past five years, agencies throughout the country have become increasingly sophisticated in developing and employing risk-assessment systems. Project coordinators in several states have identified many of the concerns we discuss in this article and have warned against viewing risk assessment as a magical cure-all.<sup>46</sup> It is not clear, however, whether their voices are being heard.

Implementation of a risk-assessment system can be a stimulus for major changes in a child protection system. If implementation is accompanied by clarification of goals, analysis of resource needs, increasing training of workers, good supervision, and high-quality services, it will certainly result in a vastly improved system. But without the additional changes, risk assessment will be the emperor's new clothes. ♦

### Notes

1. As Dee Wilson, Project Coordinator of the Washington State Risk Assessment Model, has noted, when risk-assessment instruments first became popular, many child welfare professionals made exaggerated claims about their utility and often seemed to misunderstand what was necessary to develop these instruments. See Wilson, D., "CPS: A Working Agenda," in *Research Issues in Risk Assessment For Child Protection*, edited by P. Schene and K. Bond. Denver, CO: American Association for Protecting Children, 1988. (Hereinafter cited as *Research Issues*). Although discussion is generally more sophisticated today, conceptual confusion and inflated claims still abound.

2. Of course, the developers of the instrument could define the harms, but they could not be sure whether these were the harms of concern to the legislators.
3. See, for example, California Welfare and Institutions Code Section 300 (1989). This view is shared by many child welfare professionals. See National Association of Public Child Welfare Administrators, *Guidelines for a Model System of Protective Services for Abused and Neglected Children and their Families*. Washington, DC: APWA, 1988. Over the past 15 years, much of the work of the first author has been devoted to developing legal standards that specify guidelines for the difficult decisions that must be made in the child protection system. I have tried to develop both a theoretical and value framework for the system that spells out justifications for state intervention to protect children and the ways in which intervention should take place. The framework defines the harms from which children should be protected. I have argued that state intervention is justified only when a child has sustained, or there is a high likelihood based on current parental behavior that the child will sustain, a particular harm. I have also argued that intervention should be limited to preventing that harm from occurring or recurring. Thus, I believe that the legal framework put in place over the past 15 years had risk assessment as the ultimate goal. See Wald, M., "State Intervention on Behalf of 'Neglected' Children: Standards for Removal of Children from their Homes, Monitoring the Status of Children in Foster Care, and Termination of Parental Rights," *Stanford Law Review* 28 (1976): 623.
4. Although the focus is on specific harms to the child, intervention can be based on parental behavior that has a high likelihood of causing those harms, if it is demonstrated both that the behavior has occurred, and that there is a close relationship between the behavior and the potential harm. Thus, a parent who hits a child with an extension cord creates a substantial danger of injury, even if no injury occurred in a particular instance. The same is true if a parent leaves a young child unattended for lengthy periods—the risk of injury is great. But parental conditions alone, such as drug use, would not justify intervention, just because drug use is a risk factor.
5. The future harm need not be of the same level of severity as the harm that resulted in initial intervention, but it should meet the state's definition of abuse or neglect. It might also be possible to develop a risk assessment instrument to predict which caregivers are likely to abuse their children, regardless of whether the caregivers have engaged in a particular behavior. We believe that such predictions should not be a basis for court intervention in cases where no harm has occurred, unless the caregiver has engaged in behavior that is highly likely to cause harm, as for example leaving a toddler unattended. Risk factors alone are not sufficiently predictive to justify coercive intervention. See *Early Prediction and Prevention of Child Abuse*, edited by K. Browne, D. Davies, P. Stratton. New York: John Wiley and Sons, 1988.
6. See, for example, California Welfare and Institutions Code Sections 358-366 (1989).
7. See Dawes, R.M.; Faust, D.; and Meehl, P.E. "Clinical Versus Actuarial Judgment," *Science* 243 (March 31, 1989): 668.
8. We have not reviewed all risk-assessment instruments and systems in current use. We have focused, in particular, on the systems and instruments described in *Second National Round-table on CPS Risk Assessment—Summary of Highlights*. Washington, DC: APWA, 1988; and in *Validation Research in CPS Risk Assessment: Three Recent Studies*. Washington, DC: APWA, 1988. (Hereinafter cited as *Validation Research*). For descriptions of other systems see Holder W., and Corey, M., *Child Protective Services Risk Management: A Decisionmaking Handbook*. Charlotte, NC: ACTION for Child Protection, 1986; Magura,

S.; Moses, B.; and Jones M., *Assessing Risk and Measuring Change in Families: The Family Risk Scales*. Washington, DC: Child Welfare League of America, 1987. We recognize that many states are regularly redefining their systems, in an attempt to overcome problems such as those discussed in this article.

9. It is unclear whether the seriousness of current abuse or neglect is related to the likelihood and/or the seriousness of any future abuse or neglect. Several researchers have argued to us that seriousness has, in fact, been shown to be a reasonable predictor. They have cited several studies that stated that severity was predictive.

We have reviewed all these studies and concluded that they do not justify including seriousness as a factor. Unfortunately, in these studies the term severity was used quite differently than it is used in most risk-assessment instruments. In all of the studies, severity was defined to include both the degree of injury and the frequency of any abuse or neglect. It is impossible to tell which of these factors—degree or frequency—is most important; one of the two factors may actually be irrelevant. Moreover, these studies never made clear what types of injuries or conduct were considered more serious. The studies, therefore, cannot be used to produce a seriousness scale.

Most risk-assessment instruments include extent of injury as a separate predictive risk factor. We can think of reasons, on the one hand, why the extent of an injury may turn out to be a good predictor—for example, a more extensive injury may reflect parental hostility toward the child or a total inability to control behavior. On the other hand, the extent of injury can be fortuitous—a child turns at the wrong time and thereby receives a more severe injury. Until there is far more evidence indicating whether severity predicts reabuse, we would not include severity as a predictor. Unfortunately, it may be difficult to obtain adequate evidence, since severe injuries often result in removal of the child, thereby reducing the likelihood of reabuse.

10. Children may be harmed in ways that do not constitute abuse or neglect. For example, poor parent-child attachment may result in the child's having poor peer relations. Services might therefore be directed at improving attachment even though this will not necessarily prevent reabuse. Rhode Island has adopted a family assessment system that seems to contemplate use primarily as a service guide. See presentation by Friedman, C., *Third National Roundtable on Risk Assessment*. Washington, DC: APWA, 1990.
11. It is beyond the scope of this article to discuss the policy issues raised by restng critical decisions on the basis of factors that parents cannot alter, such as past criminality or race. If our only concern is prediction, use of such factors may be appropriate. But, we believe, decision-makers have to focus on more than just prediction *per se*. We hope others will address this issue, since such variables may prove to be quite good predictors. See Baird, C., "Development of Risk Assessment Indices for the Alaska Department of Health and Social Services," in *Validation Research*, *supra* note 8. (Hereinafter cited as Baird).
12. Workers may be given authority to override the numbers.
13. We recognize that this is not a perfect solution, even in terms of preventing further abuse, since children are abused by foster parents. Abuse in foster care is much less likely than reabuse by parents, however.
14. The proportion of false positives and false negatives will be affected by various factors, including the base rate of the behavior being predicted (that is, the total proportion of all abusing and neglecting parents who reabuse or again neglect their children) and the cut points used for labeling a case as likely to reabuse. In defending the accuracy of risk-

assessment instruments, most commentators fail to examine a number of important statistical issues related to instrument development.

15. Discussions of this kind are going on in Washington State, according to Diana English, research director of the Washington State Risk Assessment Project. Private conversation July 1989. We hope these issues also are being discussed in other states.
16. The little research done on current instruments indicates that we can expect a significant number of false positives using these instruments. For example, in a study in Vermont, 12% of all families were labeled high-risk. Of these, 61% actually reabused or again neglected their children; 39% did not. Only 36% of those labeled medium-risk abused or neglected a child again during the study period. In contrast, 25% of those families labeled low-risk reabused or neglected their children. See Weadon, J.; Torti, T.; and Zunder, P., "Vermont Division of Social Services Family Risk Assessment Matrix Research and Evaluation," in *Validation Research*, *supra* note 8. (Hereinafter cited as Weadon). The instrument developed by NCCD and tested in Alaska produced a smaller proportion of false positives, at least when very high cut points were used. See Baird, *supra* note 13. The error rate varied by type of abuse.

A number of studies indicate that, in general, the base rate for reabuse is quite high—most studies find that between 30% and 50% of all families reabuse, even when they are involved in a treatment program. Thus we would not expect risk-assessment instruments to produce extremely high percentages of false positives. However, the few instruments that have been researched seem much better at correctly identifying low-risk cases than at identifying high-risk cases.

Although some data indicate base rates of reabuse during treatment, we lack comparable data pertaining to other decisions—for example, reabuse during the investigation process when children are not temporarily removed.

17. Wilson, *supra* note 1 at p. 101. Wilson's thoughtful piece raises several of the issues we address.
18. Some of these efforts are described in *Validation Research*, *supra* note 8.
19. See Pecora, P., "Evaluating Risk Assessment Systems—Methodological Issues and Selected Research Findings," in *Research Issues*, *supra* note 1 at 47. We support his recommendations.
20. One exception is the work done by the National Council on Crime and Delinquency in Alaska. See Baird, *supra* note 13. NCCD is repeating this procedure in Michigan. There are other ways to validate instruments; it is beyond the scope of this article to discuss the strengths and weaknesses of different validation procedures.
21. Some of the best research from a methodological perspective has been done with small or nonrepresentative samples. See the three studies reported in *Validation Research*, *supra* note 8. See also Johnson, W., and L'Esperance, J., "Predicting the Recurrence of Child Abuse," *Social Work Research and Abstracts* 29 (1984): 21.
22. The inability to separate out such factors is not uncommon in clinical judgment. See Dawes et al., *supra* note 7.
23. For a discussion of how some researchers have tried to use previous research in selecting risk items, see Miller, J.; Williams, K.; English, D.; and Olmstead, J., *Risk Assessment in Child Protection: A Review of the Literature*. Washington, DC: APWA, 1988. These authors are more optimistic than we are about the usefulness of these research findings.
24. Peter Pecora and Marjam Martin recently reviewed factors that may be predictive of sexual

abuse. See Pecora, P., and Martin, M., "Risk Factors Associated with Child Sexual Abuse: A Selected Summary of Empirical Research," in *Research Issues*, *supra* note 1 at 60. Like reviews of physical abuse, the reported studies examined factors associated with initial sexual abuse, not recidivism.

25. A few risk-assessment instruments contain variables that were derived from studies of cases where children had been killed, but those variables are not given extra weight in the instruments.
26. Two studies have developed cut points. See Weadon, *supra* note 18, and Baird, *supra* note 13. As discussed in note 18, the evidence from Alaska, Alameda County, and Vermont indicates that even when high cut points are used—that is, a large number of risk factors must be present to classify a case as high risk—there are still a substantial number of false positives, that is, cases where predicted reabuse did not occur.
27. One of the implications of this problem can be seen in an example comparing the risk profiles of two families. In the first case, the family may have low income and a parent may be unemployed. If both income level and parental employment are included on an instrument as risk factors, this family would receive a rating of high risk for both factors. Similarly, a family in which the father was alcoholic and the child was weakly bonded to the mother might receive a high-risk rating on two factors. However, if income level and parental unemployment are highly correlated with each other, then it is possible that their contributions to risk are not additive. That is, knowing the second of these two factors may not significantly increase our ability to predict risk beyond what we know from the first factor. Labeling the family as high risk on both of the factors would therefore lead to an inflation of their risk level. In contrast, if we assume that alcoholism and extent of bonding are uncorrelated, then each factor would increase the risk. Thus, a risk-assessment instrument based solely on an additive model, without taking into account intercorrelation of factors, would incorrectly equate the risk levels present in these two families.
28. See English, D., "Risk Assessment in Child Protection: A Brief Review of Issues and Concerns," in *Research Issues*, *supra* note 1 and Pecora, *supra* note 21.
29. Even studies that controlled for treatment would have to be used cautiously. The quality of the treatment must be considered. People who are at low risk for reabuse if involved in a good treatment program, may be at high risk if not given the right services. And people who reabuse after receiving poor services might respond better to good services. Thus, risk profiles drawn from any given study will not necessarily be valid under other conditions.
30. Agencies justify use of the same instrument because it supposedly allows workers to assess change in the parents over time. Change should be monitored, since it will be important in making decisions regarding reunification, or possibly, in making an initial disposition following court adjudication. The need to monitor change, however, does not justify relying on factors that do not predict risk with respect to the decision that must be made at each point in time.  
We do not object to using the same instrument over time if only a subset of the factors are used at each decision point, the appropriate subset having been identified through research.
31. Even though the child may not have already suffered harm, the past behavior must have been legally abusive or neglectful, that is, it was just a matter of fortuity that harm did not occur.
32. A possible exception to our claim that risk factors, without behavior, do not justify coercive

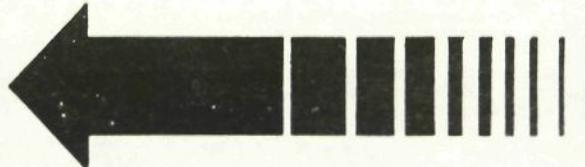
intervention might exist when a parent has harmed another child who no longer lives with the family, and there is an allegation that a second child is endangered. For example, an agency may receive an allegation from a 21-year-old that she had been sexually abused by a stepfather some years before. The woman is concerned that her stepfather may now be abusing her younger sister. Can we investigate such a family, determine the number of risk factors, and use that as a basis of sexual abuse so that protective services intervention is justified? We believe that court jurisdiction is inappropriate in such situations, although the issue is by no means clear-cut. We are reluctant to authorize court intervention absent research that adequately tells us the probability that the younger child will be sexually abused in such a situation. Absent such research, there is not enough evidence or basis for determining, by preponderance of the evidence, that the child is in danger of imminent sexual abuse. It might be that in such cases investigation is appropriate, and if there are a high number of risk factors, agencies might try to convince the parents to accept voluntary services.

33. Several studies have found that workers focus on arbitrary factors, such as economic status. See Katz, M. et al., "Returning Children Home: Clinical Decision Making in Cases of Child Abuse and Neglect," *American Journal of Orthopsychiatry* 56 (1986): 253.
34. Bator, P., "The State Courts and Federal Constitutional Litigation," *William and Mary Law Review* 22 (1981): 605, 625.
35. Wald, *supra* note 3.
36. See Johnson, W., and Clancy, T., "A Study to Find Improved Methods of Screening and Disposing of Reports of Child Maltreatment in the Emergency Response Program in Alameda County, California," in *Validation Research*, *supra* note 8.
37. We would not base decisions on the likelihood of confirmation. For example, it might be preferable to investigate all cases alleging serious physical injuries before investigating all cases alleging failure to send a child to school, even though the substantiation rate of serious physical injury is much lower than the rate of in-school neglect cases. The seriousness of the harm and the potential for irreparable damage probably should outweigh the likelihood of substantiation.  
Studies of substantiation rates may nonetheless be useful. For example, they may enable administrators to identify inadequate reporters. These reporters could then be given special training.
38. Workers might be given leeway to argue against intervention.
39. See sources cited in note 18 *supra*. Diana English informs us that in Washington State between 33% and 50% of the diverted families were again reported to CPS by 20 months, depending on the diversion program. The ratio varied by program.
40. See Baird, *supra* note 13. See also DePanfilis, D., *Final Report: Determining Safety in Child Protective Services and Child Placement Decisions*. Charlotte, NC: ACTION for Child Protection, 1988. The state of Michigan is now researching the applicability of the NCCD instrument for its cases. This approach, doing research before using an instrument to make decisions, should be followed elsewhere.
41. This rationale has been stated at several conferences attended by one of the authors (Wald). See also Stein, T., "Risk Assessment," *Protecting Children* (Winter 1986-87): 2.
42. See Wilson, *supra* note 1.
43. We would like to see services that would promote their positive development, not just protect them from terrible outcomes.

44. As previously indicated, we do not believe that all substantiated cases require ongoing CPS involvement. Many such cases are appropriately diverted to alternative systems; in some cases the risk of reinjury may be so low that subjecting the family to any intervention may be inappropriate.
45. We are concerned about another claim being made for risk assessment. We have heard administrators state that risk assessment is a tool for avoiding foster care. It is argued that by using risk-assessment instruments, limited resources can be concentrated on those cases at higher risk. There is a fallacy in this reasoning, however. Concentrating resources on those families at highest risk may not be the best use of resources. Risk assessment tells us nothing about the likely success of various services. It might be more sensible to focus services such as family preservation on the families most amenable to services, thereby preventing escalation of risk. Higher-risk situations might, in fact, require foster placement and, in some situations, termination of parental rights. We are not arguing against family preservation, even in those cases believed to be at high risk. We are stating only that without proper research, the use of risk assessment tells little about how resources are best allocated among families requiring intervention.
46. See, for example, Wilson, *supra* note 1; Au Claire, P., "Status Report: Hennepin County Community Service Department Risk Assessment Guidelines for CPS Decision-Making" and Torti, T., "Status Report: Vermont Division of Social Services Family Risk Assessment Project," in *Second National Roundtable on CPS Risk Assessment*, *supra* note 8.

(Address requests for a reprint to Michael S. Wald, Stanford University Law School, Stanford, CA 94117.)

**MOVING?**  
Make sure  
**CHILD WELFARE**  
goes along...



1. For FASTEST service attach old mailing label in space below

If mailing label is not available, print your OLD name and address in this box

2. Print your NEW address here

Name \_\_\_\_\_ Title \_\_\_\_\_

Address (if student, give name of school) \_\_\_\_\_

City \_\_\_\_\_ State \_\_\_\_\_ Zip \_\_\_\_\_

Check here to renew your subscription. Attach mailing label above

3. Mail to Child Welfare League of America

c/o Transaction, Inc., Rutgers University, New Brunswick, NJ 08903



Please allow six weeks for change to take effect

**Copyright of Child Welfare is the property of Child Welfare League of America and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.**