



What Is R2P?

The Child Welfare League of America's Research to Practice (R2P) Initiative provides relevant research information to the practice community, especially individuals at agencies providing services to children and families. R2P

- reviews research literature on the well-being of children, youth, and families;
- disseminates information about pertinent research and its implications for practice;
- supports effective planning and implementation of evidence-based programs and practices; and
- encourages program evaluation and outcome measurement.

R2P disseminates information through seminars, workshops, and conference presentations; R2P teleconferences; onsite comprehensive consultation; and the R2P website.

How Can You Become Involved?

If your agency has an effective program and an *evaluation component* to support your track record of success, you can become involved in the R2P Initiative. You have an opportunity to share your successful program development and implementation experiences with others! R2P has tools to examine a program's existing research and works closely with the sponsoring agency to determine the appropriateness of the program and research for the R2P Initiative.

R2P Wants to Hear from You!

If you are familiar with a program or practice suited for R2P, we want to hear from you! This is a great opportunity to demonstrate what is working in your community so others can benefit from your experiences. E-mail your program descriptions and evaluations, questions, comments, or suggestions to R2P@cwla.org.

The Empirical Basis of Risk Assessment in Child Welfare: The Accuracy of Risk Assessment and Clinical Judgment

Alan W. Leschied, Debbie Chiodo, Paul C. Whitehead, Dermot Hurley, and Larry Marshall

The importance of risk assessment is juxtaposed with the lack of empirical support regarding the validity of risk inventories. This study compared risk ratings of one risk assessment tool to decisions made by case managers. The researchers sampled 450 children and compared predictive utility of risk assessment to child protection decisions. Risk assessment was consistent with clinical judgment in 74% to 81% of cases, more than previously reported in studies of risk assessment validity. Further analyses identified discriminate functions at the instrument's category and individual-item levels. The results have implications for the validity of the instrument and its utility in child welfare.

Alan W. Leschied, PhD, CPsych, is Professor, and Debbie Chiodo, MA, is Graduate Student, Faculty of Education; Paul C. Whitehead, PhD, is Professor, Department of Sociology; and Dermot Hurley, MSW, is Assistant Professor, School of Social Work, University of Western Ontario, London, ON, Canada. Larry Marshall, MSW, is Director of Family and Children's Services, London and Middlesex Children's Aid Society (CAS), London, ON, Canada. This study is part of a larger investigation of fac-

Risk assessment, both as a concept in describing client assessment for service allocation and as the application of relevant inventories and procedures, has become increasingly popular in human services. In a social service context, risk relates to the probability of requiring a specific intensity and nature of service. In the case of child welfare, risk relates to probability of out-of-home placement.

In part, expansion of risk assessment is a reaction to reduced funding of human services, which has resulted in a need to provide an efficient means of assessing individuals to prioritize those who require scarce social services. Services for all children and youth during the past two decades have seen a rise in the use of measures of risk in the children's mental health and young offender systems. Hoge, Andrews, and Leschied (1986) noted that much of the risk assessment literature relates to the nature and intensity of services offered to youth in either residential or community contexts.

Measures of risk are composed of empirically based predictors of given events or disorders. The primary assumption underlying the development of a measure of risk is that knowing the predictors of a given disorder can lead to appropriate targeting combined with a specified intervention. Administrators can use risk assessment to shape the organization of services and allocation of resources. The potential for clinicians to benefit from risk assessment lies in the opportunity to maximize the efficacy of a given intervention, because risk assessment can help create a match between the intervention and the individual's area of need.

Despite the encouraging research context within which risk assessment has emerged, the clinical application of the concept has not matched its potential. Hoge and Andrews (1996) reviewed two decades of research on the application of the risk concept

tors surrounding the increase in referrals to the London and Middlesex CAS funded by the United Way of London and Middlesex. The authors are indebted to CAS of London and Middlesex for its support of and cooperation with this research. The authors particularly note the assistance of John Liston, Judy Van Leeuwen, Joe Czikk, and Greg Lewis. The authors also acknowledge the contributions of the research team:

and concluded that most risk judgments in children's services are unsystematic, intuitive procedures with questionable validity. The research literature on risk assessment has shifted from describing risk prediction to developing more rigorous procedures for measurement and connecting the outcomes of risk assessment with decisions about the delivery of human services.

Risk Assessment in Child Welfare

Doueck, English, De Panfilis, and Moote (1993) noted that in the 1980s, the application of structured risk assessment in child welfare increased. Since that time, risk measures have proliferated, for example, the Illinois CANT 17B, the Washington Assessment of Risk Matrix, and the Child at Risk Field System. Doueck et al. noted that most risk assessment systems do not reflect rigorous scientific efforts and are plagued with serious theoretical and methodological flaws. Gambrill and Shlonsky (2000, 2001) also pointed out that methodological problems in risk assessment continue to pose challenges for child welfare decisionmakers. These challenges include the predictive accuracy and the ability of structured risk assessment systems to contribute to child welfare decisions beyond the information that can be gained in other contexts. Gambrill and Shlonsky noted that decisions in child welfare are predicated on the validity of the risk assessment tools used.

Risk Instruments

Risk instruments vary in type, content, and applicability. Murphy-Berman (1994) noted that no uniformity or agreed-on procedures are involved in their use. An overview of the development of risk assessment by English and Pecora (1994) suggested that fewer than 50% of the variables contained in risk instruments had been empirically tested, which raises questions about the instruments' validity. Baird and Wagner (2000) suggested that although many child protection service agencies have used risk assessment instruments for a decade or more, the validity of these instruments had not been the subject of serious psychometric inquiry. Despite

this rather pessimistic note, the agencies have continued to profess the need for increased efficiency and valid decisionmaking, arguing that the theoretical and empirical supports for these systems need to be strengthened.

Predictive Accuracy

Among the more serious problems facing risk assessment in child welfare is the challenge of improving predictive accuracy. De Panfilis and Zuravin (2001) noted the low predictive accuracy in risk assessment procedures. They suggested the possibility of abandoning formal risk assessment and contended that it may be more useful to identify a set of key factors to help workers determine which families require further intervention. In partial support for this view, early British work cited by Anglin (2002) suggested that child protection practice in the United Kingdom rarely exceeded 70% in accuracy of predictions in protecting children from future abuse, further noting that the effectiveness of risk assessment would likely not exceed 80%.

Improving the measurement properties of risk instruments is of considerable importance. The introduction of risk assessment practices typically occurs as the result of a tragic outcome, when a child was not removed from a dangerous situation, or the inability of child welfare agencies to prioritize services to those children and families most in need. Anglin (2002) noted, however, that one of the unintended effects of risk assessment may be to increase the number of children admitted to care. These issues are of critical importance in Ontario, Canada, where researchers conducted the present study.

Adoption of Risk Assessment in Ontario's Child Welfare System

Ontario is the largest province in Canada, representing almost 40% of the country's population. Since the mid-1990s, referrals to the province's child welfare agencies and admissions to care have dramatically increased. The number of children who agencies

suspected of being abused or neglected and investigated increased by 44% between 1993 and 1998 across 52 children's aid societies (CASs). The number of substantiated investigations of neglect doubled between 1993 and 1998, rising from 4,400 to 8,900. The number of children actually taken into care increased from 10,419 in 1996 to 15,792 by 2001 (Rivers, Trocme, Goodman, & Marwah, 2002).

In response to this crisis in the demand for services, the provincial ministry responsible for CAS programs decided to use a risk assessment model. A research team from the University of Toronto's Faculty of Social Work, along with the cooperation of numerous CASs, modified early work from the Child Well-Being Scales by Magura and Moses (1986). This produced the Risk Assessment Model for Child Protection in Ontario (Ontario Association of Children's Aid Societies [OACAS], 2000).

Developers have trained social workers in the use of the instrument as a standard part of clinical practice since 2000. The instrument is now part of the provincial strategy to assess children within the CAS mandate. Despite the ministry's commitment to test the instrument's reliability and validity, these studies have not yet been done. This study is the first to report on the measurement properties of the instrument.

Risk Information

This risk assessment summary includes five distinct areas for assessment, along with a rating scheme that extends along a five-point, Likert-type scale for each area assessed. Table 1 identifies each of the five areas and the subareas. The result yields both an overall assessment of risk (low to high) and a cumulative risk assessment score composed of a total of the ratings from each of the five areas.

Risk Assessment in the Context of Case Planning

For the specification of risk assessment to be an integral part of the case plan, the *Ontario Risk Assessment Manual* suggests that

TABLE 1
Summary of Areas for Risk Assessment

1. Caregiver Influence

Abuse/neglect
 Alcohol/drug use
 Expectations of child
 Acceptance of child
 Physical capacity to care for child
 Mental/emotional/intellectual capacity

2. Child Influence

Vulnerability
 Response to caregiver
 Behavior
 Mental health and development
 Physical health and development

3. Family Influence

Family violence
 Ability to cope with stress
 Availability of social supports
 Living conditions
 Family identity and interactions

4. Intervention Influence

Caregiver's motivation
 Caregiver's cooperation with intervention

5. Abuse/Neglect

Access to child by perpetrator
 Intention and acknowledgement of responsibility
 Severity of abuse/neglect
 History of abuse/neglect or neglect committed by present caregivers

information must be provided that relates to the nature of the case management decision (OACAS, 2000). In other words, the instrument must have concurrent and predictive validity. Clinically, the tool is used in all cases in which workers have identified a protection concern and the decision could result in the child being admitted to care. This investigation was also interested in whether a rating based on the risk measure relates to the decision with respect to admitting the child to the care of CAS. Hence, t

researchers completed risk measures not only on the cases that were admitted to care, but also on cases where the decision was to provide in-home services or no further service at all.

The Present Study

Consistent with calls for more empirical input into the development and use of risk assessment in child welfare, this study focuses on the predictive validity of scores on risk with respect to clinical decisions made in one large, urban CAS. Cases reflected the broad spectrum of child welfare services, ranging from closing of the case with no further need for intervention to admitting a child to care. Research investigation is in three areas:

- Examining the degree of congruence between ratings from scores based on the risk measure and decisions made by child welfare case managers, who draw on a broad basis of social history information in arriving at child welfare decisions.
- Identifying the category of items that discriminates between cases that are admitted to care and those that are not.
- Identifying the individual items that contribute to the discriminations made in child welfare decisions.

Method

Participants

The participants in this study were 450 children randomly selected from the 2,645 children who were either referred to or sought assistance from CAS of London and Middlesex* in 1995. The researchers extracted the data from the case files in summer 2001 as part of a larger study examining factors related to recent increases in admissions to care. Because the goal of this study

* The Children's Aid Society (CAS) of London and Middlesex is the third-largest child welfare authority in the province, serving a community of approximately 350,000 people. The London CAS is a comprehensive children's service providing individual and family counseling and residential services (foster and group home care), along with the traditional CAS protection services and adoption.

was to examine the contribution of measured risk to decisions with respect to admissions to care, the outcome was considered binary. That is, the study compared children who were admitted to care and those who stayed in the community. The sample consisted of 234 children who were admitted to care and 216 children who received either no service or some form of service that resulted in the children remaining with their birthparents. The latter type of intervention could range from personal and family counseling to a referral to another agency. The researchers chose cases from 1995 because they predated the introduction of a risk assessment instrument.

To capture sufficient cases of the subgroup of interest (children in care), the researchers oversampled these groups, including a larger proportion of cases in the sample than would have been found in the population of cases in the given year. For analytical purposes, they statistically corrected for this discrepancy by applying a mathematical correction, a statistical weight, to each case. This allowed them to make population inferences from this sample.

Procedure

The researchers trained six raters in the method of risk assessment by an expert trainer at the sponsoring agency. The study provided training in risk assessment so that each rater would meet a criterion of acceptance. The study based all ratings on file reviews. Each risk element received a score of severity ranging from zero to four. The raters coded each level of service predicted as the most intensive service received by the child in the calendar year. Interrater reliability for the cumulative risk was $r = .92$ and for the overall risk was $r = .96$. The study used the calendar year as the reference point for admission because this study also examined the effects of changes in social policy bound by the year in question. To extend the follow-up period beyond the calendar year would have compromised risk as a predictor with the potential effects from changes in social policy that might affect place-

ment independent of risk. The researchers then compared the decision of the case manager with the risk assessment provided by the trained rater.

Statistical Analyses

Analyses involved three stages. First, to address the discriminant validity of the scores on the measure of risk, the researchers performed a direct discriminant function analysis using two risk assessment variables as predictors of membership in two groups. Predictors were the overall assessment of the risk score (low to high) and the cumulative risk assessment score. The possible range was 0 to 88. In the sample, the actual range was 1 to 72.

The researchers computed this by determining a total score from each of the five areas of risk. Cronbach's alpha coefficient is a measure of the internal consistency of the subscales. No generally agreed on cut-off reliability value exists, however, 0.7 and above is acceptable for a small number of test items (Nunnally, 1978). The cumulative risk assessment scores within each area of risk formed a reliable scale: Cronbach's alpha values ranged from .79 to .83.

Second, to address the discriminant validity of the specific areas of risk influence, the researchers undertook a stepwise discriminant function analysis using the scores from each of the five areas of risk.

Finally, to address the discriminant validity of all risk assessment items, the researchers performed a stepwise discriminant function analysis on the 22 risk items.

Results

The sample of 450 children was composed of 234 males and 216 females. Children's ages ranged from birth to 19 years ($M = 7.79$, $SD = 5.25$, $n = 448$). The study identified 176 (39%) children who were admitted to care and 274 (61%) who had either no service or some form of in-home service (these reflect weighted numbers).

The two groups did not differ in gender ($\chi^2 = .05, df=1, p > .5$) or age ($F[1, 447] = .55, p > .5$).

Discriminant Function Analyses

Discriminant Function Based on Total Risk Scores

The initial interest was in the possibility of identifying the two groups on the basis of the overall scores on the risk assessment measure. For a two-group discrimination, researchers calculated one function that maximized the differences between groups and the homogeneity within the groups, which are expressed in terms of minimizing Wilks' lambda. The overall risk assessment score and the cumulative risk assessment score served as predictors. The analysis yields Wilks' lambda = .753 (the $F[1, 447]$ approximation = 146.93, $p < .001$) for the overall risk assessment score and Wilks' lambda = .875 (the $F[1, 447]$ approximation = 63.864, $p < .001$) for the cumulative risk assessment score. This indicates that the two groups are significantly differentiated on the basis of these two variables.

The overall risk assessment score produces the largest correlation with the discriminate function ($r = .996$), whereas the cumulative risk assessment score yields a moderate correlation with the discriminate function ($r = .586$). For classification, the study used sample sizes to estimate the probability of group membership to either the in-care or out-of-care group. The discriminant function based on both predictors correctly classified 76% of the cases (canonical correlation = .497, $\chi^2[1] = 126.923, p < .001$). When combined, the overall risk assessment and cumulative risk assessment scores produced extremely high discriminant validity.

The sensitivity of the risk assessment measure—the probability that it classified children in care who actually were in care—was 81%. On the other hand, the specificity of this measure—the probability that it classified children as not in care who were not in care—was 74%. The risk assessment measure accurately classified 88 (50%) of the children in care, compared with the 68 (39%)

who would be correctly classified by chance alone. The predicted model incorrectly classified 20 (7%) of the children as being in care when they actually were not.

Discriminant Function by Category

To address the discriminant validity of the specific areas of risk assessment influence, the researchers performed a stepwise discriminant function analysis on the total scores of the five areas of risk influence. Family influence accounted for the largest difference in group membership (Wilks' lambda = .96, $F[1] = 20.88, p < .001$). Intervention influence accounted for the second largest (Wilks' lambda = .95, $F[2] = 13.0, p < .001$). None of the remaining variables differentiated group membership. The discriminant function based on family influence ($r = .9$) and intervention influence ($r = .8$) together correctly classified 64% of the cases (canonical correlation = .234, $\chi^2[2] = 25.27, p < .001$). Together, these areas of risk influence correctly classified 29 (17%) of children in care. This suggests that the total scores of caregiver influence, child influence, and abuse or neglect influence do not significantly discriminate between children in care and those receiving a less intensive form of intervention.

Discriminant Function by Individual Items

To determine the relative contribution of the individual risk assessment items, the researchers performed a stepwise discriminant function analysis on the scores of the 22 risk items. When they forced these variables into a discriminant function, four items were able to accurately classify 76% of the participants (Canonical correlation = .494, Wilks' lambda = .756, $p < .001$). These four items accurately classified 109 (76%) of the children in care, compared with 68 (39%) who would be correctly classified by chance alone. The discriminant function based on these four items incorrectly classified 15% of the children as being in care when they were not (see Table 2).

TABLE 2
Discriminant Function Item Analysis Related to Children's Aid Society Level

Step	Variable	Wilks' lambda	df	Univariate F	r
1	Family Influence/ability to cope with stress	.81	1	104.5	.85
2	Family Influence/availability of social supports	.79	2	58.36	.31
3	Intervention influence/caregiver's motivation	.77	3	45.09	.71
4	Intervention influence/caregiver's cooperation with intervention	.76	4	35.91	.23

Note: $p < .001$.

Discussion

This study is a first step in the empirical verification of the risk assessment instrument. This instrument has two functions: It relates to decisions by social workers to take children into care, and it serves as a predictor of future harm.

Overall, this risk scale displays a higher level of agreement with clinical decisions than previously reported in studies identified by Anglin (2002), in which the state of the art in predictive accuracy seldom exceeded 70%. The overall accuracy rate in this study ranged from 74% to 81%.

Of additional interest is the significant relationship between some specific items in the scale and clinical decisions. The strongest contribution from the discriminant function was the caregiver's motivation to cooperate with intervention. In 1995, child welfare policies reflected a strong commitment to maintaining families through a graduated, least-intrusive approach to sustaining children in their families. In part, this may be reflected in the importance placed on the factor of family motivation to ac-

cess treatment (e.g., policies supporting putting families first, making least-intrusive interventions, and maintaining the integrity of families) and less by child protection. It may be that changes to child welfare policy could influence which items relate to decisions regarding admissions to care.

These results suggest that the contribution of the three family influence items is minimal and that the family's ability to cope with stress is responsible for the discriminant power of the subscale. Moreover, the results of the second discriminant function indicate that the caregiver's and child's influence risk area scores do not discriminate between the two groups. In contrast, at an item level, the caregiver's physical and mental capacity to care for the child and the child's mental and behavioral development have important discriminant value.

In general, strong but far from perfect correspondence exists between scores on the risk assessment instrument and decisions actually made by child protection workers. Of special interest are the cases in which the instrument predicted that the child would be taken into care, but the social worker decided not to do so. It could be that the instrument was a correct predictor that the child would be taken into care, but this happened at a later date. Extending the follow-up period of child welfare decisions may account for some of these cases.

It may also be that availability of alternative resources to child welfare intervention may account for some of the variation in decisions about admission. Subsequent research will examine these possibilities by comparing the criterion validity of the instrument and exploring additional items that researchers could add to the scale to improve its predictive accuracy. This study provides some encouragement that a risk measure can be constructed that exceeds previously held expectations for criterion and predictive validity in child welfare. ♦

References

- Anglin, J. P. (2002). Risk, well-being, and paramouncy in child protection: The need for transformation. *Child and Youth Care Forum, 31*, 233-255.
- Baird, C., & Wagner, D. (2000). The relative validity of actuarial-and consensus-based risk assessment systems. *Children and Youth Services Review, 22*, 839-871.
- De Panfilis, D., & Zuravin, S. J. (2001). Assessing risk to determine the need for services. *Children and Youth Services Review, 23*, 3-20.
- Doueck, H. J., English, D. J., De Panfilis, D., & Moote, G. T. (1993). Decision-making in child protective services: A comparison of selected risk-assessment systems. *Child Welfare, 72*, 441-452.
- English, D. J., & Pecora, P. (1994). Risk assessment as a practice method in child protective services. *Child Welfare, 73*, 451-473.
- Gambrill, E., & Shlonsky, A. (2000). Risk assessment in context. *Children and Youth Services Review, 22*, 813-837.
- Gambrill, E., & Shlonsky, A. (2001). The need for comprehensive risk management systems in child welfare. *Children and Youth Services Review, 23*, 79-107.
- Hoge, R. D., & Andrews, D. A. (1996). *Assessing the youthful offender: Issues and techniques*. New York: Plenum Press.
- Hoge, R. D., Andrews, D. A., & Leschied, A. W. (1986). An investigation of risk and protective factors in a sample of youthful offenders. *Journal of Child Psychology and Psychiatry, 37*, 419-424.
- Magura, S., & Moses, B. S. (1986). *Outcome measures for child welfare services: Theory and applications*. Washington, DC: Child Welfare League of America.
- Murphy-Berman, V. (1994). A conceptual framework for thinking about risk assessment and case management in child protective service. *Child Abuse & Neglect, 18*, 193-201.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Ontario Association of Children's Aid Societies. (2000). *Risk assessment model for child protection in Ontario: Eligibility spectrum, risk assessment model for child protection in Ontario*. Toronto, Canada: Ontario Ministry of Child and Family Services.
- Rivers, B., Trocme, N., Goodman, D., & Marwah, I. (2002). Reporting and beyond: ...
...trends in child abuse and neglect call for broader reforms. *Voices for Children Newsletter*.