# An Evaluation of the Minnesota SDM Family Risk Assessment

## Final Report

**Conducted for the**

**Minnesota Department of Human Services**

by

**L. Anthony Loman, Ph.D. and Gary L. Siegel, Ph.D.**

**A Report of the**

**Institute of Applied Research**

**St. Louis, Missouri**

**December 2004**

111 N Taylor
St. Louis, Missouri 63122
(314) 966-5101
email: **iar@iarstl.org**
website: **http://www.iarstl.org**

# Table of Contents

# Executive Summary

The Structured Decision Making (SDM) Family Risk Assessment (FRA) determines the probability that a family will continue to abuse or neglect their children. The FRA categorizes families as low, moderate, high or intensive risk of future child abuse and neglect. The charge for the present research was to determine for the *entire population of families served through CPS* and for *specific subpopulations* the *reliability, validity and effects on services* of the FRA. Five subpopulations were to be considered: Caucasian, African American, American Indian, Southeast Asian and Hispanic.

**A note on reading this report**: Those willing to read the entire report should skim this summary and begin with Chapter 1. An alternative for readers desiring more detail but unwilling to wade through the technical details of the report would be to read Chapter 1 and then skip to Chapter 7 which is a longer version of this summary, with fuller conclusions and recommendations. The numbering of the sections of Chapter 7 corresponds to the numbers of preceding chapters (2 through 6) to permit easy movement between the summary and more detailed materials, as desired.

## Predictive Validity

- Analysis showed that the FRA has predictive validity in regard to new reports of child maltreatment and new cases opened for families following such reports. Low-risk families have fewer new reports than moderate-risk families. Similarly, moderate-risk families have fewer new reports than high and intensive risk families.

  - Like all tools intended to predict future human behavior, however, the FRA involves error. Analysis indicated that the scale misclassified approximately one in three families.

  - The larger portion of predictive error arose from families with low and moderate risk scores that were reported later. This may reflect conditions and circumstances of families and family members *that were not present at the time the FRA was administered.* It may also reflect the failure of workers to accurately complete all FRA items.

- All the individual items of the FRA showed predictive validity, including demographic characteristics of families, such as number of children and age of the caregiver. Some risk factors can be addressed and changed while others cannot. Together they point toward the families in greatest need of attention.

## Reliability

- The FRA is composed of two subscales—one for risk of neglect and the other for risk of abuse. Both scales demonstrated internal consistency slightly below the lower range of what is generally considered acceptable.

- Analysis of a vignette survey in which workers determined the risk of family in a written description showed that workers tended to use the two subscales consistently. This finding is mitigated somewhat because the vignette methodology could not take into account differences that might arise from encounters between workers and real-life families.

  - There was also evidence that consistency among workers dropped in producing the final categorical rankings (low through intensive) on the FRA because relatively minor variation in subscale scores can produce substantive variation in final risk categories.

## Other Practice Issues

In surveys and interviews of workers:

- Most workers recognized that the FRA introduced positive features, such as consistency, into the family assessment process.

- Differences were found in the extent to which the FRA affected decision-making about services to families. Some respondents said it was a minor factor or unimportant in responding to families. For others it was a major factor.

  - This seemed to be a function of local offices rather than differences among workers within offices. The larger the county and CPS office, the more importance the FRA assumed in decision-making.

- Responses to families with lower risk scores also varied in the same way. Low-risk families were less often provided with post-assessment services in the larger urban offices compared to other counties.

- The FRA was completed at different points in the assessment process. FRA scores in some cases reflect the state of the family during or shortly after the first visit by the worker while in other cases the score represents the family at the end of the assessment process.

- Workers were also concerned that *certain characteristics of the FRA push families to higher risk levels* than should be the case. These included the following:

- Events from long ago may be scored the same as events that occurred recently (e.g., very old past cases versus cases that just closed)
- Some risk factors may be present but mitigating factors reduce their significance (e.g., coping skill or extended family support)
- Some items may be more risky for one subculture than another (e.g., the number of children in the family).
- Some other *items may need to be modified* to be accurate (e.g., the age of children is a factor in risk).

- On the other hand, *some risk-related items may be missing leading families to be rated as lower risk* than should be the case (e.g., mental health).

- One factor mentioned by workers impinges directly on reliability. Workers indicated that there was *no way to indicate a lack of knowledge*. Missing information is scored as no risk on the FRA. And, items are sometime left blank when *workers suspect but cannot prove that a risk factor is present*.

## Minority Subpopulations

- While the study of the five racial and ethnic subpopulations indicated some differences in the application of individual FRA items, many of these were evened out in the final categorization of families into the fourfold classification of low, moderate, high and intensive risk.

  - The exceptions to this rule were Southeast Asian families that received overall lower risk scores and American Indian families that received overall higher risk scores.

- The FRA showed levels of predictive validity for the subpopulations similar to the entire study sample, with the same exceptions. It was more accurate with Southeast Asian families and less accurate with American Indian families.

  - The lack of predictability of the FRA for American Indian families was examined in greater detail. The primary problem was one of false negatives—low-risk families with new reports. The problem appeared to occur among neglect subscale items having to do with parenting skills, harmful relationships of parents, substance abuse, financial problems, and motivation and cooperation. Examination of worker narratives showed that such risk factors (and some others, such as mental health, not included in the FRA) were present in low-risk American Indian families with report recurrence, either at the time of the original report or in later reports. This may indicate changes in families, mistakes by workers in completing the FRA, or both.

- Perusal of narratives indicated that this same problem was present, perhaps to a lesser extent, across the spectrum of low-risk families with recurrence.

- An experimental design was employed to permit workers to respond to the same family (in a descriptive vignette) but with different minority subpopulation identifications. No evidence of racial/ethnic bias could be detected in this analysis.

## Services and the FRA

A substantial minority of families in the study were involved in the Alternative Response (AR) project, which involved a new approach to families reported for child maltreatment. In comparison to traditional investigations, AR family assessments are non-adversarial family-friendly visits that aim at engagement and fuller family participation in the assessment process. Comparing the traditional and AR approach, several differences were found.

- Significantly more services, particularly services addressing basic financially-related and household needs, were delivered to low-risk families under the Alternative Response approach than under the traditional approach.

- Data were utilized for this analysis from the Alternative Response evaluation. FRA risk levels, services and report-recurrence were considered. Experimental-control comparisons revealed that recurrence was lower for AR families generally under these conditions and, specifically, *that services to low-risk families made a difference in outcomes*.

  - While identification of high-risk and intensive-risk families through the FRA can be used as a means to determine families in need, it should not be the exclusive method of determining need. Services to low risk families improved family outcomes and were, in the long-term, cost effective for the CPS agency.

## Recommendations

Certain recommendations are offered in the full report. They are outlined here. Refer to Chapter 7 for more details.

1. **Change the order of completion of the SDM instruments.** The new order might be SDM safety assessment, assessment of family strengths and needs (FSN), and family risk assessment. Low-risk families with many deficiencies and few strengths on the FSN may be considered for further services. Families with no indications of threats to child safety (or no child maltreatment in an investigation) but with high indications of needs *or* high risk should be invited for services on a voluntary basis.

2. **Improve the FRA Scoring method.** A more sophisticated scoring method might improve the predictability and reliability of the instrument. Furthermore, a new scoring method might provide fuller information to practitioners.

3. **Empirically Test Changes to the FRA.** Changes should be considered to the FRA and empirically tested:

    a. Add a "do not know" category to each FRA item to create an *uncertainty score*.

    b. Permit workers to check an item when they *strongly suspect* but cannot prove the presence of the risk factor.

    c. Create an alternative risk factor list to accompany the FRA that would be the basis of increasing the risk level of families.

    d. Create a list of mitigating or strength-based factors.

4. **Change in Practice in Larger Counties.** If the FRA is used in some large offices as the primary means of excluding low-risk families from response by the agency, consideration should be given to modifying this practice. Because the FRA, as currently employed, may misclassify some families as low-risk, additional criteria should be employed to determine whether post-assessment services are appropriate.

# 1. Introduction

This report concerns research conducted for the Minnesota Department of Human Services (DHS) of the Structured Decision Making (SDM) Family Risk Assessment instrument. (We will refer to the instrument throughout this document as the FRA.) The FRA is one in a battery of SDM instruments in a comprehensive assessment system developed by the Children's Research Center (CRC).[1]

SDM was designed to assist Child Protection Services (CPS) agencies in serving families reported for child abuse and neglect and has been adopted in whole or in part by several states. The system includes instruments designed to simplify and standardize decision making for CPS workers and supervisors. These include: 1) screening criteria, 2) response priority, 3) safety assessment, 4) risk assessment, 5) child needs and strengths assessment, 6) family needs and strengths assessment, 7) case planning and service standards, and 8) case reassessment.[2] The version of SDM adopted in Minnesota includes 2, 3, 4, 6, 7 and 8. In addition, Minnesota uses a family risk *re*assessment tool as well as tools for assessment of the risk and safety associated with reunifying children in out-of-home placement with their families. Each of the SDM instruments includes standardized items that are scored numerically—in some cases as 0 or 1 (no or yes) and in others as rankings (e.g., -1 to +3). In the FRA the numeric rankings are summed to provide scale scores.

The FRA, then, is one tool among many within the SDM system. However, the FRA has a pivotal role. Risk assessment determines the probability that a family will continue to abuse or neglect their children. The FRA categorizes families as low, moderate, high or intensive risk of future child abuse and neglect. CRC presents the FRA as a tool to guide the allocation of resources.[3] There is a recognition that service standards will vary from state to state, but it is assumed that riskier families require increased worker contacts and increased allocation of resources.[4] In a system in which the FRA-based risk classifications are used as the primary or the central criteria in determining the level of services to families or whether the agency continues to work with families at all, the accuracy of the classifications is critically important.

---

[1] CRC is located in Madison, Wisconsin and is a division of the National Council on Crime and Delinquency. Website: http://www.nccd-crc.org/crc/crcindex.html

[2] See: http://www.nccd-crc.org/crc/c_sdm_about.html

[3] "Armed with this critical information [i.e., risk levels], agencies are well-positioned to make decisions about how resources should be differentially allocated across clients." (http://www.nccd-crc.org/crc/c_sdm_risk.html).

[4] See: http://www.nccd-crc.org/crc/c_sdm_caseplan.html

This study did *not* compare the FRA to other risk assessment systems or to CPS systems with no formal risk assessment process. The reader interested in this question is referred to Appendix B for a review of research literature on the SDM FRA and certain other risk assessment tools utilized in CPS. The approach was to consider the tool on its own using standard measures of validity and reliability and statistical analyses of possible racial and ethnic bias and of service implications.[5]

**General Research Tasks.** The charge of DHS to the evaluators was to determine for the *entire population of families served through CPS* and for *specific subpopulations* the *reliability, validity and effects on services* of the FRA. Five subpopulations were to be considered: Caucasian, African American, American Indian, Southeast Asian and Hispanic.

## The Structure of the FRA

The FRA consists of 25 risk assessment *items*, of which 13 are grouped into a risk of *neglect subscale* and 12 into a risk of *abuse subscale*. Each item is scored, that is, it receives a separate *weight* in determining the final risk score. Weights range variously from 0 to 3. Some items are of the present-absent type and receive a score either 0 or 1. Others provide for a measure of intensity with weights of 0-1-2, 0-1-3, or 0-1-2-3. The final *scale scores* are determined by adding these weights. The items and their corresponding weights are shown in the Table 1.1. The numbering for items adheres to that of the original paper-and-pencil instrument utilized in Minnesota. This numbering will be used throughout this report. Neglect-scale numbers range from N1 through N11, although item N6 consists of three separate items (N6b, N6c and N6d) for a total of 13 items. Abuse-scale numbers range from A1 through A12. In late 2001, Minnesota implemented an automated version of the FRA within the Social Services Information System (SSIS) that retained this numbering.

The two subscales are completed for all cases, that is, risk of neglect and risk of abuse are determined without regard to the presenting problem. Items N1 and A1 refer to the presenting problem(s), and either one or both may be completed. Neglect subscale summated scores range from 0 to 20. Abuse summated subscale scores range from 0 to 16.

The scale scores are then used to assign neglect and abuse risk levels: Neglect risk levels are assigned as: low = 0 to 4, moderate = 5 to 7, high = 8 to 12, intensive = 13 to 20. Abuse risk levels are assigned as: low = 0 to 2, moderate = 3 to 5, high = 6 to 9, intensive = 10 to 16. The neglect and abuse risk categories are then compared and the higher of the two is selected to determine the overall family risk level. For example, a neglect score of 5 coupled with an abuse score of 7 would result in a high-risk determination. Similarly, an abuse score of 1 and a neglect score of 6 would result in a moderate-risk determination.

[5] In the CRC literature, the term *equity* is sometimes used to refer to the absence of racial/ethnic bias.

## Table 1.1. FRA Items and Assigned Weights

| | Weights= | 0 | 1 | 2 | 3 |
|---|---|---|---|---|---|
| | **Neglect subscale** | | | | |
| N1 | Whether the current report is for neglect. | No | Yes | | |
| N2 | Number of prior assigned reports. | None | One | Two or more | |
| N3 | Number of children in the home | Two or fewer | Three or more | | |
| N4 | Number adults in the home at the time of the report | Two or more | One or none | | |
| N5 | Age of the primary caregiver | 30 years or older | 29 years or younger | | |
| N6 | Characteristics of primary caregiver | | | | |
| | b. Lacks parenting skills | Not applicable | Applies | | |
| | c. Lacks self-esteem | Not applicable | Applies | | |
| | d. Apathetic or hopeless | Not applicable | Applies | | |
| N7 | Primary caregiver involved in a harmful relationship. | No | Yes, but not victim of domestic violence | Yes victim of domestic violence | |
| N8 | Primary caregiver has a current substance abuse problem. | No | Alcohol only | Other drugs (w. Or w/o alcohol) | |
| N9 | Household is experiencing severe financial difficulty. | No | Yes | | |
| N10 | Primary caregiver is motivated to improve parenting skills. | Motivated and realistic | Unmotivated | Motivated but unrealistic | |
| N11 | Cooperation of caregiver. | A. Viewed seriously and cooperated | B. Viewed less seriously than investigator | C. Failed to cooperate | Both B and c |
| | **Abuse subscale** | | | | |
| A1 | Whether the current report is for abuse. | No | Yes | | |
| A2 | Types of prior abuse reports (physical or sexual). | A. None | B. Physical abuse report(s) | C. Sexual abuse report(s) | Both B and c |
| A3 | Prior CPS service history. | No | Yes | | |
| A4 | Number of children in the home. | One | Two or more | | |
| A5 | Whether the caregivers were abused as children. | No | Yes | | |
| A6 | Secondary caregiver (SC) has a current substance abuse problem. | No, or no SC | Alcohol and/or drug abuse problem | | |
| A7 | Any caregiver employs excessive and/or inappropriate discipline. | No | | Yes | |
| A8 | Caregiver has a history of domestic violence. | No | Yes | | |
| A9 | Caregiver is a domineering parent. | No | Yes | | |
| A10 | Child in the home has a development disability or history of delinquency. | No | DD or delinquent | | |
| A11 | Secondary caregiver motivated to improve parenting skills. | Yes, or no SC | | No | |
| A12 | Primary caregiver views incident less seriously than agency. | No | Yes | | |

A final option in the FRA permits overrides of the family risk level for various reasons. Regardless of the calculated risk, the risk categorization is bumped to intensive for four specific reasons: sexual abuse with continued perpetrator access to the victim; non-accidental physical injury to an infant; injury requiring hospitalization or medical treatment; and death of a sibling. In addition, a "discretionary override" is possible to raise the risk by one categorical level (for example, from moderate to high). This kind of override is open-ended, although the worker must provide a written reason for the override. *No override is permitted for the purpose of reducing the risk level.*[6]

## Research Questions

The research was organized to answer multiple questions:

1. *What is the level of validity of the FRA?* Because risk of child abuse and neglect is a predictive concept, this question was addressed by determining the predictive power and the level of error in prediction of reports of child abuse and neglect.

2. *How reliable is the FRA?* Reliability was examined in two ways. The internal consistency of each of the FRA subscales (neglect and abuse) was determined. Then, the scoring consistency among workers completing the FRA was examined. Broader issues were also addressed by examining variations in the attitudes and practices of workers who used the FRA in Minnesota.

3. *Can any differences be detected in the way the FRA is applied to subpopulations of interest and the validity or reliability of the FRA when used with subpopulations?* The fivefold breakdown of Minnesota families—Caucasian, African American, American Indian, Southeast Asian and Hispanic—was utilized.

4. *What is the relationship between level of risk, service provision and later outcomes for families?* Portions of the analysis for this question used experimental-control comparisons drawn from the evaluation of the Minnesota AR program.

Each of these questions was addressed, along with new questions that arose during the analyses.

## The Context

**Staggered Adoption of the FRA in Minnesota Counties.** Minnesota began implementing the SDM Family Risk Assessment in 1999 as a part of the SDM system.

[6] The manual for the system (*State of Minnesota CPS Structured Decision Making System, Policy and Procedures Manual.* September 1999) provides instructions for scoring and interpretation of the FRA.

SDM was utilized early in several of the larger Minnesota counties and was gradually adopted in other counties over the next four years.

Late in 2000, the state began the Minnesota Alternative Response (AR) Project. The project was piloted in 20 counties that agreed to participate. The 20 AR counties were each required to use the SDM safety assessment and family risk assessment tools for families screened as appropriate for AR. Families that were AR-appropriate constituted a large portion of families reported to CPS. Consequently, the FRA was utilized for many families in the participating counties. Because of this and because the present evaluators conducted the evaluation of the AR Project, data from that project were also analyzed in this study. The AR project is described in greater detail below.

The FRA was completed manually until September 2001, when SDM was automated within the Minnesota Social Services Information System (SSIS). At that point, 18 of the AR counties back-entered into SSIS FRAs that had been completed from February through September 2001.[7] From this time forward, counties adopting SDM entered the FRA directly into SSIS.

The samples of families selected for analysis in this study were drawn from SSIS. The large sample with cases beginning in 2001 (described below), therefore, is skewed somewhat toward the early adopting counties and the 20 AR counties. The sample contains fewer cases from many other counties and no cases from some counties (the smallest in the state) that did not begin SDM until late 2002 and 2003.

**Distribution of Population and Subpopulations in Minnesota.** The map in Figure 1.1 is to be used as a reference for the maps that follow. It shows county populations grouped by size. The high-population counties on the mid-eastern portion of the state are those in the Minneapolis (Hennepin County) and St. Paul (Ramsey County) metropolitan area. The cities of Duluth and Rochester are in St. Louis County and Olmsted County, respectively. The city of Mankato is in Blue Earth County and St. Cloud is in Sterling County.

The map illustrates the variation in population density in Minnesota. Upwards of half of counties have populations of less than 20,000. Only 13 counties have populations of 50,000 or more. This has consequences for the nature of CPS offices and client populations. The CPS system is Minnesota is county-administered. Individual county offices vary in size from a handful of CPS workers to several hundred workers in the large urban counties. This may have consequences for risk assessment in that workers from thinly population rural areas are more likely to know about families that are reported for child abuse and neglect. This is less true of workers in denser urban areas. Whether prior knowledge makes risk assessment more or less accurate is an important question, although not addressed in this study.

[7] Two counties, Hennepin (Minneapolis) and Dakota, that adopted SDM early were also part of the AR project. These counties were using separate automated systems for SDM. FRA data from these systems were made available and were converted for use in this analysis.

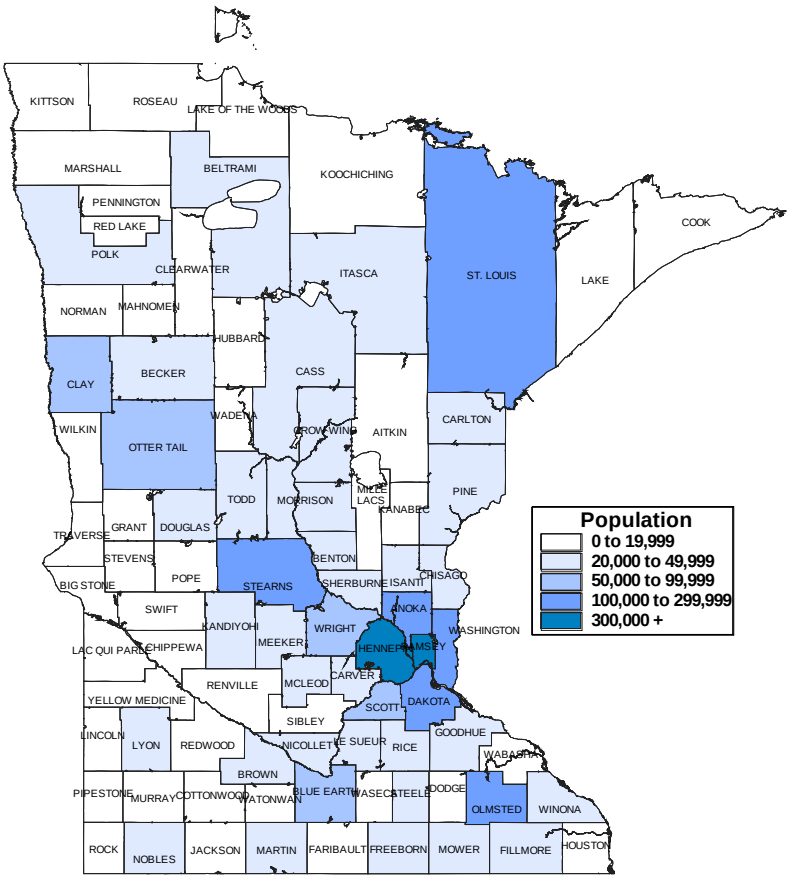


**Figure 1.1. Population of Minnesota Counties**

Figure 1.2 contains four maps showing the distribution of the four minority subpopulations included in this study. The African American and the Southeast Asian populations are more heavily concentrated in urban areas, and there are many counties that contain only tiny numbers of either of these groups.

The American Indian population, by contrast, is distributed more widely in the state. Relatively large numbers are found in some smaller counties that overlap with or are adjacent to reservations. On the other hand, large populations can also be found in urban parts of the state.

The Hispanic population is also more widely distributed in Minnesota. Like other minority populations, they are found in urban areas. But they also live in mid-sized and small rural counties. In some instances, concentrations of Hispanics in smaller counties reflect the presence of migrants working in industries, such as meat processing.

The uneven distribution of minorities suggests the need for sub-state analysis when comparisons are made to the majority Caucasian population. This type of analysis was performed in the present study.

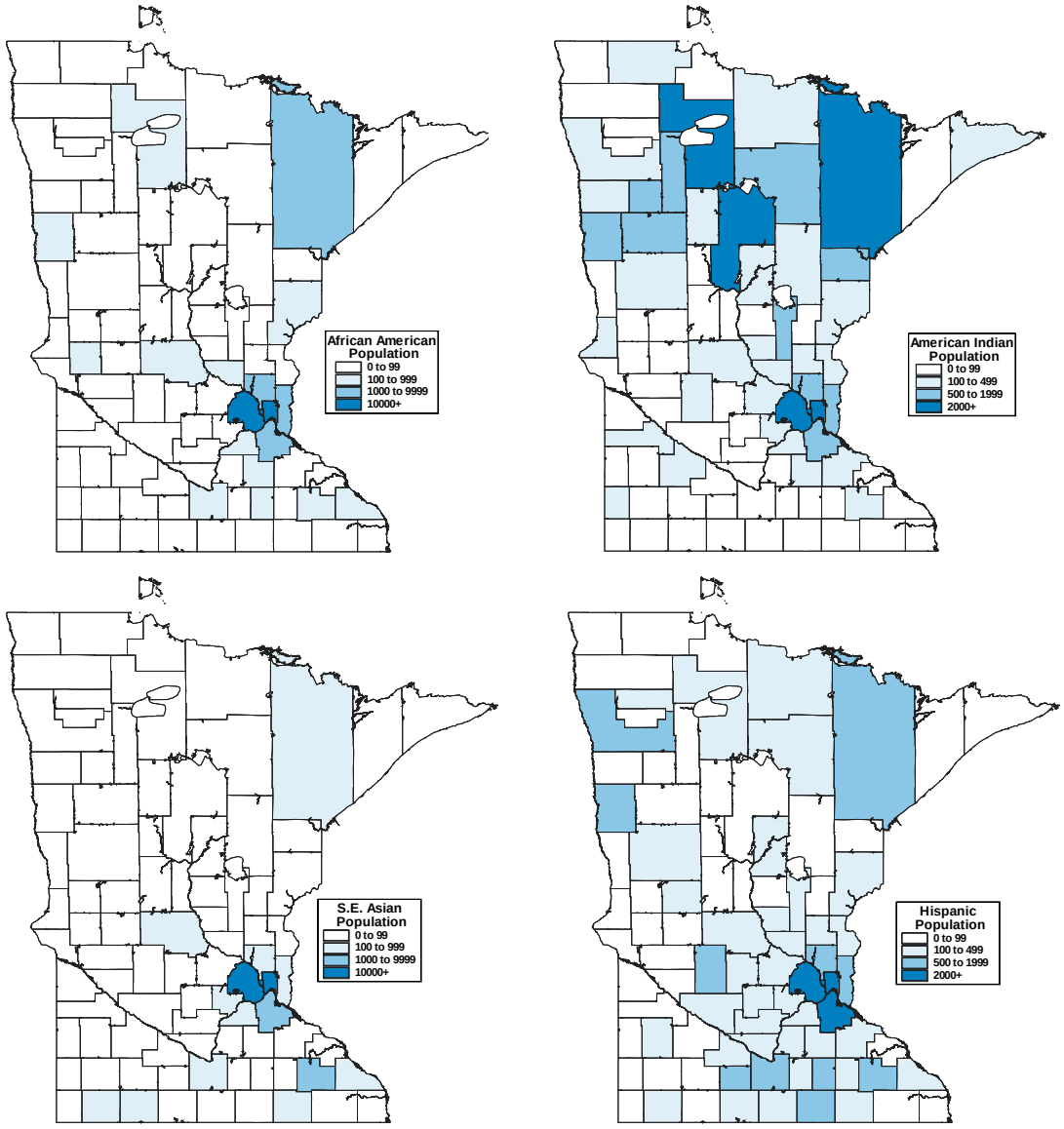


**Figure 1.2. Distribution of Minority Subpopulations in Minnesota: African American, American Indian, Southeast Asian, and Hispanic**

**The Alternative Response Project Evaluation.** As noted, Minnesota began the AR Project in 2000. By February and March of 2001 the program was operating in all 20 counties and the project evaluation had begun. Part of the evaluation involved random assignment of families to experimental or control group status.[8] The control group was

[8] See Minnesota Alternative Response Evaluation: Final Report, Institute of Applied Research, November 2004: http://www.iarstl.org/papers/ARFinalEvaluationReport.pdf. Control cases were assigned in 14 of the 20 AR counties. Chapter 8 of the evaluation report contains a detailed comparison of the experimental and control groups utilized in the impact analysis, showing that the two groups were composed of similar types of families.

provided with a traditional CPS approach—investigation of the allegations of the report, findings regarding maltreatment, and opening of a minority of cases for post-assessment services. The experimental group received an alternative response.[9] Experimental families were not investigated but received a family assessment that was family-friendly, non-adversarial, and focused on broader needs of the family. *The FRA was used for each experimental and control family in the AR project.* Figure 1.3 shows the 20 counties that participated in the AR evaluation.

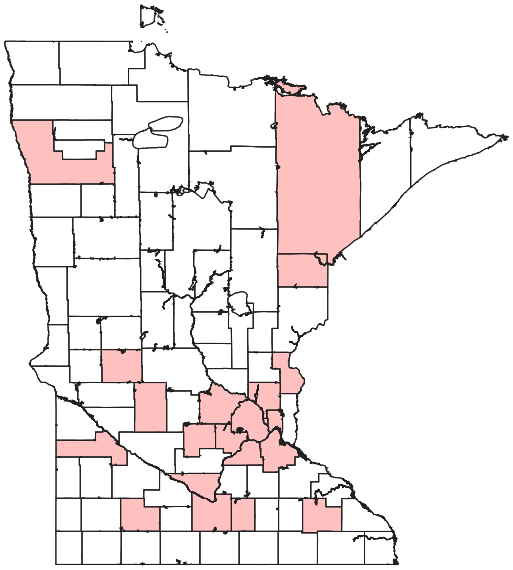

**Figure 1.3. Counties that Participated in the Alternative Response Demonstration Project**

Under the traditional approach to child protection that evolved after the 1960s, CPS concentrated on families most in need, usually families in crisis, on the assumption that this represented the best use of limited agency resources. The FRA has been promoted as a means of improving the accuracy of CPS in identifying high-risk families so that they can be targeted for further intervention and services, while at the same time steering the agency away from low-risk families. The AR approach represents a broadening of the traditional CPS approach. AR workers seriously consider provision of services to lower-risk families—even to families in which no maltreatment of children can be substantiated, if they are willing to participate. The existence of an experimental and control group of essentially similar families that embodied these two approaches presented a unique opportunity to assess the relationship between family risk and services—one the tasks of the present study.

## Variables and Data Collection

Variables are considered in greater detail in later chapters, but it will be useful to say a little about them in this introduction.

1. **Risk Scores.** Final FRA scores were described above as low, moderate, high and intensive family risk. We also used at various points in the analysis: a) *scores on individual items* of the FRA, b) total scores of the FRA *abuse subscale and neglect subscale*, and c) *separate categorical risk scores for abuse and for neglect*.

[9] A more common term for this approach is "differential response." It is being tested in a number of other states and has been adopted in several, most notably Missouri, in the last decade (see the summary of our evaluation of the Missouri demonstration: http://www.iarstl.org/papers/MoFamAssess.pdf).

2. **Recurrence.** Recurrence of child abuse and neglect was measured indirectly as recurrence of child abuse and neglect *reports*. It was not possible to use the other more commonly used measure of recurrence, namely *substantiated reports*, because with the introduction of AR in Minnesota the majority of reports are not investigated in the traditional way (and thus cannot be substantiated). Recurrence of formal CPS *cases* (termed "case-management workgroups") was also used as a recurrence variable in the study. More detail can be found in Chapter 2.

3. **Subpopulation Identification.** Race and ethnicity is assigned to *individuals* in SSIS. The unit of analysis of this study was the family, and therefore, the race/ethnicity of the head of the cases was used to assign subpopulation identification to families. The process is discussed in greater detail in Chapter 5.

4. **Services.** Service information was collected for certain sample families in the present study. It was not available through SSIS, however, and for analyses of the large sample of 15,100 families, opening of a formal service case was utilized as a proxy variable for services. In Minnesota, these are referred to as case-management workgroups.

**Data Collection.** Several methods of data collection were used. First SSIS data were available on families that had received reports in the system during the period from January 2001 through September 2004. A sample of 15,100 families was selected from the period January 2001 through September 2002 that had received at least one report of child abuse and neglect with a subsequent FRA classification. Each family was tracked for 24 months following the initial report.

Second, a case-specific survey of Minnesota workers was conducted. The survey referred to specific families who had received an FRA, and because we wanted workers to provide further information on family risk, a later sample was drawn of families with a report and an FRA during the final quarter of 2003. Families were selected in such a way that workers would be required to provide information on no more than two families. The final sample consisted of 590 families served by 320 workers. Of these workers, 236 responded (final response rate of 73.7 percent) with information on 412 families. Workers were asked specific questions about families and family members. They were also asked to respond to general questions about the use of the FRA.

Third, interviews of workers and supervisors were conducted. These were confined to the 20 counties involved in the Minnesota Alternative Response (AR) evaluation. Workers were in interviews in the following counties: Anoka, Blue Earth, Carlton, Carver, Chisago, Cottonwood, Dakota, Hennepin, Kandiyohi, McLeod, Nicollet, Olmsted, Polk, Pope, Ramsey, Scott, St. Louis, Waseca, Wright, and Yellow Medicine. Together these counties served the large majority of families entering CPS in Minnesota.

A fourth method involved an experimental design within a vignette survey. Four vignettes were constructed (see Appendix A). These consisted of two versions of a case meant to present risk of neglect and two version of a case meant to present risk of abuse.

The two versions were designed to vary the risk from lower to higher. The five racial/ethnic identifications of families (subpopulations) were varied across the four vignettes for a total of 20 separate combinations. There were 10 neglect combinations (a higher-risk and a lower-risk neglect vignette for each of the five subpopulations. There were 10 abuse combinations (a higher-risk and a lower-risk abuse vignette for each of the five subpopulations). Each worker in the survey responded to *two* vignettes: one of the ten neglect categories and one of the ten abuse categories. Workers were randomly assigned to each set. This insured that workers would be evenly distributed across experimental treatment groups that varied the type of maltreatment, risk level and subpopulation identification. Of the 686 workers who were sent surveys, 459 (67.2 percent) responded in time to be included in the analysis.

The case-specific survey and the vignette survey were both web-based. They were conducted in March 2004 and October 2004, respectively. After permission was obtained from local directors or supervisors, each worker was sent an email explaining the purpose of the survey. In the case-specific survey, ID numbers of families and a combination of first name and initial of last name was used to identify families. Workers were asked to review the case(s) and then click on a link in the email. The link opened a survey form in their Internet browser. In the vignette survey the link contained a code that directed the worker to the two vignette conditions to which the worker had been randomly assigned. After reading the vignettes, workers completed the FRA on each.

As noted, data from the AR evaluation were utilized. No new data collection on AR families was conducted for the present analysis.

## Organization of the Remainder of the Report

Chapter 2 addresses the issue of predictive validity by examining how accurately the final risk ratings and the individual parts of the FRA predict new child abuse and neglect reports and new case-management workgroup openings for families. This is followed in Chapter 3 by analyses of reliability. These include the internal consistency of the FRA items and consistency of scoring among workers. In Chapter 4, certain issues of practical utility are considered. Feedback from workers is presented. An analysis of data from the AR evaluation is presented that focused on variation in scoring of items. Question of the reliability and validity of the FRA for minority subpopulations are considered in Chapter 5. The relationship between FRA ratings of risk and services to families is considered in Chapter 6. Chapter 7 contains conclusions and implications of the preceding analyses. Finally, there are two appendices. Appendix A presents the vignettes utilized in the inter-rater reliability survey. Appendix B contains a brief review of literature on risk and past research on the FRA.

# 2. Predictive Validity of the Family Risk Assessment Instrument

The basic idea underlying the validity of an instrument is whether it actually measures the thing it is claimed to measure. But what is risk of child abuse and neglect a measurement of? It refers to certain characteristics of families, family members and the environment of families that are *correlated* with incidents of child maltreatment. If the characteristics are present child abuse and neglect are more likely in the future. Risk can be thought of as a dimension that underlies the characteristics, or perhaps, as a dimension that summarizes the characteristics. The concept *implies* that certain characteristics of families either *cause* or *are necessary conditions* for child maltreatment. It also implies that child maltreatment develops systematically. Causation and etiology, however, do not have to be explained, or even expressed, to assert that certain families are riskier than others. And, there are different theories about the relationship of risk characteristics to child abuse and neglect. This would seem to rule out *construct* validation of the FRA.[10]

Similarly, no "gold standard" of family risk assessment exists upon which most theoreticians and practitioners agree. If such a generally accepted standard existed we might test the FRA against it to determine validity. Thus *concurrent validity* is eliminated as well.[11]

Another type of validity is *predictive validity*. In this case, the construct being measured is thought to be related positively or negatively to certain future events. Validation is possible by observing whether such events occur or not and determining whether occurrence is related to the measurement. Instruments that measure risk are concerned primarily with future events. The purpose of a child maltreatment risk assessment instrument is to rank or segregate families based on their likelihood of maltreating their children in the future. Testing accuracy of prediction, therefore, is most

[10] *Construct validity* refers to generalizations to other theoretically related constructs. A measure of a construct (variable) A that is valid should also be correlated with construct B if B is in theory causally related to A. It should be uncorrelated with construct C if C has in theory no causal connection to A. Correlation of a measure of A to a measure of B is termed *convergent validity*. The absence or divergence of correlation between A and C is termed *discriminant validity*. Construct validity refers to the consistency of a measure with the propositions of a larger theory.

[11] *Concurrent validity* usually involves comparing the measure to other accepted measures of the same phenomenon at the same time, that is, concurrently. If both measures produce similar results for the same population (of families in this case) concurrent validity of the new measure could be claimed.

appropriate in validating a measure like the FRA. Validity hinges on the power of the FRA to predict.

## The Power of the FRA to Predict Recurring Reports and Cases

We begin at the heart of the matter: Are the four final risk categories of the FRA (low, moderate, high and intensive risk) positively associated with future child maltreatment? Since we cannot measure future child maltreatment directly but only reported child maltreatment, the question must be rephrased as: Is the FRA positively associated with recurrence of accepted reports of child maltreatment?

In this chapter, we leave aside considerations of reliability, interpretation and potential strengths and defects of the FRA. We treat the scoring of individual items on FRA as if they were perfectly reliable. Issues of reliability and potential bias are reserved for later chapters.

When reports of child abuse and neglect are accepted in Minnesota, they appear as intake records in the Social Services Information System (SSIS). Each of the 15,100 families in our follow-up study sample had such a report during the period from January 2001 through September 2002. In each case a SDM FRA was conducted for the family and was also entered into SSIS and these risk assessment scores are considered in this analysis. Each family was then tracked for exactly 24 months from the date of the first report (termed here the *initiating report*). Various events were tallied during this period, including new accepted reports of child maltreatment.

Two events may intervene to make new reports less likely. One is that children may be removed and placed in substitute care. During the time that all children are removed from a family, reports of child maltreatment will not be received. A second factor is that all the children in a family may reach their 18th birthday and are no longer subject to laws governing CPS. To deal with the first problem, we tallied all the days in placement of each child in each family and determined minimum periods in which no child was present in the home. Regarding the second, periods of children aging out of CPS were determined by calculations based on the dates of birth of children. These were combined for each family to create the number of days of *decreased opportunity for new reports of child maltreatment.*[12] This variable was used as a weight across the population of families in the study.[13]

[12] We call these *decreased opportunity* days (dod) rather than *no opportunity* days, because such a method is inherently inexact. Children are removed for overlapping dates that are very difficult to coordinate; removed children return home for trial visits while still in the custody of the court; new children are born or enter families through marriage that were unknown at the time of the initiating report; not all children are listed in investigations, family assessments or case management workgroups and may appear later; children of low-income families sometimes stay with relatives for some period and are in and out of the home. These are just a few of the sources of error.

[13] For the analysis of "any new accepted report" the weight was calculated as 1 – (dod)/730. In fact, the use of this weight turned out to be largely superfluous. Children are removed and placed at all family risk levels, and because more families appear in lower than in higher risk categories, more children in absolute

Table 2.1 shows the proportions of families with new accepted reports for each risk level. The percents of interest are shown in bold in the table. We see that the percent of new reports are indeed in the correct direction. A little over one-fifth of families rated originally as low-risk had new accepted reports. A little over two-fifths of families rated as high-risk had new reports. Moderate-risk families fell in between. Intensive risk families had slightly lower recurrence rates than high-risk families, but generally in line with them. In Minnesota offices, the intensive classification occurred for only 3.6 percent of all families, and even in a large sample, this small a number of families in either category (no reports or at least one report) may have a relatively large influence on percentages. For example, a swing of 40 families (less than .3 percent of the total sample) would raise the percentage of intensive-risk recurrence from 38.2 to 65.3 percent.

**Table 2.1. Risk Assessment by Recurrence of Any Accepted Maltreatment Report during 24 Months**

| *Risk Level* | *No new report* | | *At least one report* | | |
|---|---|---|---|---|---|
| | *Number* | *Percent* | *Number* | *Percent* | *Total* |
| Low | 4,549 | 77.7% | 1,260 | **22.3%** | 5,809 |
| Moderate | 3,899 | 65.1% | 2,023 | **34.9%** | 5,922 |
| High | 1,696 | 59.6% | 1,125 | **40.4%** | 2,821 |
| Intensive | 340 | 61.8% | 208 | **38.2%** | 548 |
| Total | 10,484 | 68.6% | 4,616 | **31.4%** | 15,100 |

Chi Square = 392.7, p < .0001 Tau-b = .144, p < .0001
Somers' d = .116, p < .0001

The relationship between risk level of new reports is statistically significant and indicates that the *FRA has general predictive validity*. On the other hand, Somers' D was .116, an indication of overall weak association between risk level and new reports.[14] (Associational measures can range from –1 to +1, with +1 being perfect direct association and 0 being no association; values from .1 through .25 are usually said to be weak.) Both these conditions—predictive validity and weak positive association—are possible in a measure of this kind. The weak association, however, is an indication of errors in prediction, that is low risk families that returned and high/intensive risk families that did not.

The predictive approach may be expanded by asking whether risk level is related to the number of recurring reports during the follow-up period. In this case, all

---

numbers are removed in lower risk than higher risk families. At the same time, new accepted reports are significantly more likely to occur in families with a child removed during follow-up. The overall effect of weighting is to raise the relative proportions of families with new reports in about equal measure across all risk categories.

[14] Somers' D is an asymmetric version of Kendall's tau-b statistic of ordinal association. Unlike the latter, D corrects for tied ranks, but only on the independent variable, in this case the family risk level.

subsequent reports were counted.[15] The percents are shown in Table 2.2. Again, the differences are consistent with predictive validity in that the largest percentages are in the lower left portion of the table (for example, 22.3 percent under 1 new report) and the smallest are in the upper right. However, the association increased only slightly under this condition (from .116 to .122).

**Table 2.2. Risk Assessment by Number of Accepted Recurring Maltreatment Reports during 24 Months**

| *Risk Level* | *Number of New Reports* | | | | | | |
|---|---|---|---|---|---|---|---|
| | *None* | *1* | *2* | *3* | *4* | *5* | *6 or more* |
| Low | 77.7% | **14.5%** | **4.5%** | **2.0%** | **0.8%** | **0.3%** | **0.2%** |
| Moderate | 65.1% | **20.1%** | **8.6%** | **3.3%** | **1.7%** | **0.8%** | **0.4%** |
| High | 59.6% | **21.4%** | **10.3%** | **4.8%** | **2.2%** | **1.0%** | **0.7%** |
| Intensive | 61.7% | **22.3%** | **10.8%** | **2.9%** | **1.1%** | **0.9%** | **0.3%** |

Chi Square = 472.8, p < .0001
Tau-b = .143, p < .0001
Somers' d = .122, p < .0001

Mean number of new reports per family in each category were, low risk: .35; moderate risk: .61; high risk: .71; intensive risk .63. (F = 121.9, p < .0001). The measure of association in this case (Eta) was .151, a slight comparative increase using parametric statistics.[16]

**Case-Management Workgroups.** Another measure is the occurrence of new cases (Table 2.3). In Minnesota new cases are referred to as case-management workgroups. The number of workgroups is not as relevant in this case since separate workgroups are sometimes opened for individual children who have been removed from the home, but 95 percent of the total population had three or fewer workgroups opened during the two-year follow-up. The pattern of relationships was similar to that for recurring reports and was statistically significant.

[15] The average (mean) number of days to a new report and intake was 254 days. New intakes occurred slightly earlier for intensive-risk cases (227 days) and high-risk cases (229 days) and slightly later for moderate (259 days) and low-risk cases (276 days).

[16] Recurrence rates presented throughout this report may differ from those of other studies and from statistics generated by the state. This analysis is family oriented. The FRA is a *family* risk scale and requires recurrence to be defined as any report on any child in the family by any perpetrator. This includes children who appear later in families through new births or blending of families. Rates measured in this way will be higher than rates for particular children or particular perpetrators *and rates limited only to new reports on the individuals mentioned in the initiating report.* In addition, we combined data for Minnesota counties by adding routines to track families that moved from one county to another during the follow-up period, something that is not done in the current county-based version of SSIS in Minnesota. (There are no statewide IDs for families.) This procedure also leads to higher recurrence rates.

**Table 2.3. Risk by Occurrence of Any Case-Management Workgroup during 24 Months**

| *Risk Level* | *Case-management workgroup* | *Total* |
|---|---|---|
| Low | 9.3% | 5,809 |
| Moderate | 15.1% | 5,922 |
| High | 33.7% | 2,821 |
| Intensive | 34.1% | 548 |
| Total | 17.4% | 15,100 |

$p < .0001$

## Predictive Validity of Subscales and Individual Items of the FRA

**Subscales.** As previously described, the FRA is composed of two subscales—separate groups of scale items for predicting the risk of neglect and the risk of abuse. Scores are obtained by adding together the weights for the individual scale items. Neglect scores range from 0 to 20, and abuse scores from 0 to 16. Rate of recurrence is shown for each subscale score in Table 2.4.

**Table 2.4. Risk Subscale Scores by Recurrence of Accepted Maltreatment Reports**

| *Risk Score* | *Neglect subscale* | *Abuse subscale* |
|---|---|---|
| 0 | 17.2% | 23.0% |
| 1 | 18.7% | 24.2% |
| 2 | 23.5% | 27.2% |
| 3 | 29.5% | 33.0% |
| 4 | 34.1% | 34.8% |
| 5 | 38.8% | 38.9% |
| 6 | 41.5% | 40.8% |
| 7 | 42.5% | 42.0% |
| 8 | 40.2% | 45.1% |
| 9 | 44.5% | 43.4% |
| 10 | 45.6% | 46.0% |
| 11 | 42.5% | 29.4% |
| 12 | 43.6% | 37.0% |
| 13 | 38.8% | 64.3% |
| 14 | 36.0% | 42.9% |
| 15 | 42.2% | 23.0% |
| 16 | 28.9% | -- |
| 17 | 40.7% | |
| 18 | 87.5% | |
| 19 | -- | |
| 20 | -- | |
| | Somers' D = .114 | Somers' D = .084 |

The rates of recurrence increase for both neglect and abuse as the scores increase. Fluctuations among the highest scores were due to smaller numbers of families in these categories. The overall trend within the table again supports the predictive validity of the FRA. The measure of association (.114) indicates that the neglect scores were in line with the overall (weak) predictive power of the FRA. Abuse scores showed a somewhat weaker association (.084).

**Predicting Specific Types of Presenting Problems.** Although no analyses have been published, to our knowledge, on the creation of the FRA subscales, the generation of neglect and abuse scores implies that the former is related to increased reports of child neglect and the latter to increased reports of child abuse (both physical and sexual). Because the present analysis is concerned with accepted reports rather than determinations or substantiation of child abuse and neglect, we turned to the presenting problem contained in SSIS as a means of testing this question.[17]

For this analysis correlations were calculated between the subscale scores and the number of later reports in each of three categories of presenting problems: physical abuse, sexual abuse and child neglect.[18] The results are shown in Table 2.5.

**Table 2.5. Correlations of Risk Subscale Scores with Presenting Problems**

| *Reports of* | *Total neglect subscale* | *Total abuse subscale* |
|---|---|---|
| Physical abuse | .033 | .121 |
| Sexual abuse | .033 | .077 |
| Child neglect | .173 | .059 |

all statistically significant, $p < .0001$

Similar to previous analyses, this analysis confirms the general predictive validity of the FRA subscales. The neglect side is indeed more highly correlated with new reports of neglect and the abuse side with new reports of physical abuse. There was cross-correlation of neglect-abuse and abuse-neglect for the subscales, and sexual abuse appears to fall in between but was more strongly correlated with the abuse subscale.

Recall that Pearson's r statistic ranges from –1 to +1, where minus one is a perfect inverse correlation (the greater the x, the less the y), plus one is a perfect direct

[17] As noted in Chapter 1, the introduction of the Alternative Response (AR) approach in Minnesota in 2001 has reduced the frequency of substantiation or findings of child maltreatment to a small minority of cases. Under AR, there is no investigation and thus no findings of child maltreatment.

[18] Physical abuse included physical abuse, threatened physical abuse and mental injury. Sexual abuse included sexual abuse and threatened sexual abuse. Child neglect included neglect of basic needs (food, clothing and shelter), prenatal exposure, infant medical neglect, endangerment, inadequate supervision, educational neglect, medical neglect, emotional neglect and abandonment.

correlation (the greater the x, the greater the y) and zero represents no correlation. The coefficients in the table are all direct or positive correlations. All are in the correct direction but all are relatively weak, less than .20, that is, they are closer to zero than to one.

The division of the FRA into separate subscales was presumably introduced to improve the predictability of the tool on the assumption that risk of abuse and risk of neglect represent mutually independent or orthogonal dimensions. The method of producing the final risk rating combines the scales by generating a *low-moderate-high-intensive* rating for each subscale and then by selecting the higher risk level of the two. Thus, if a family has a low abuse rating but a moderate neglect rating the resultant overall risk rating is moderate. A low abuse rating with an intensive neglect rating yields an overall intensive-risk rating. This makes scoring of the risk assessment relative easy and leads to the simplest possible categorization of families. The price for this is a *loss of information*. For example, a final high-risk score could result from any of the following five conditions: low neglect/high abuse; moderate neglect/high abuse; low abuse/high neglect; moderate abuse/high neglect; high neglect/high abuse. A question for practice might be whether retaining the separate abuse and neglect risk ratings would be helpful in working with families? The integration of the FRA into local automated systems (SSIS) may make it possible to use a more complex scoring system with results that could better inform decision-making.

**Individual FRA Items.** Workers answered individual items within the FRA at very different rates. These can be seen in Figure 2.1. The figure is shown to give a sense of the frequency with which FRA items are applied to families. As noted earlier, most items are of the no-yes (0-1) type, and for these, the percentages checked as 1 are shown in the figure. A few items involved different risk levels (and different weighted scores). These are shown as stacked bars in the chart, with the options described in the row headings. Items A6a and A6b as well as A10a and A10b are included for information purposes and do not influence the scores on items A6 and A10, that is, the score can be only 0 or 1 on these items. Risk items are utilized for as few as 6 percent and as many as 69 percent of families.

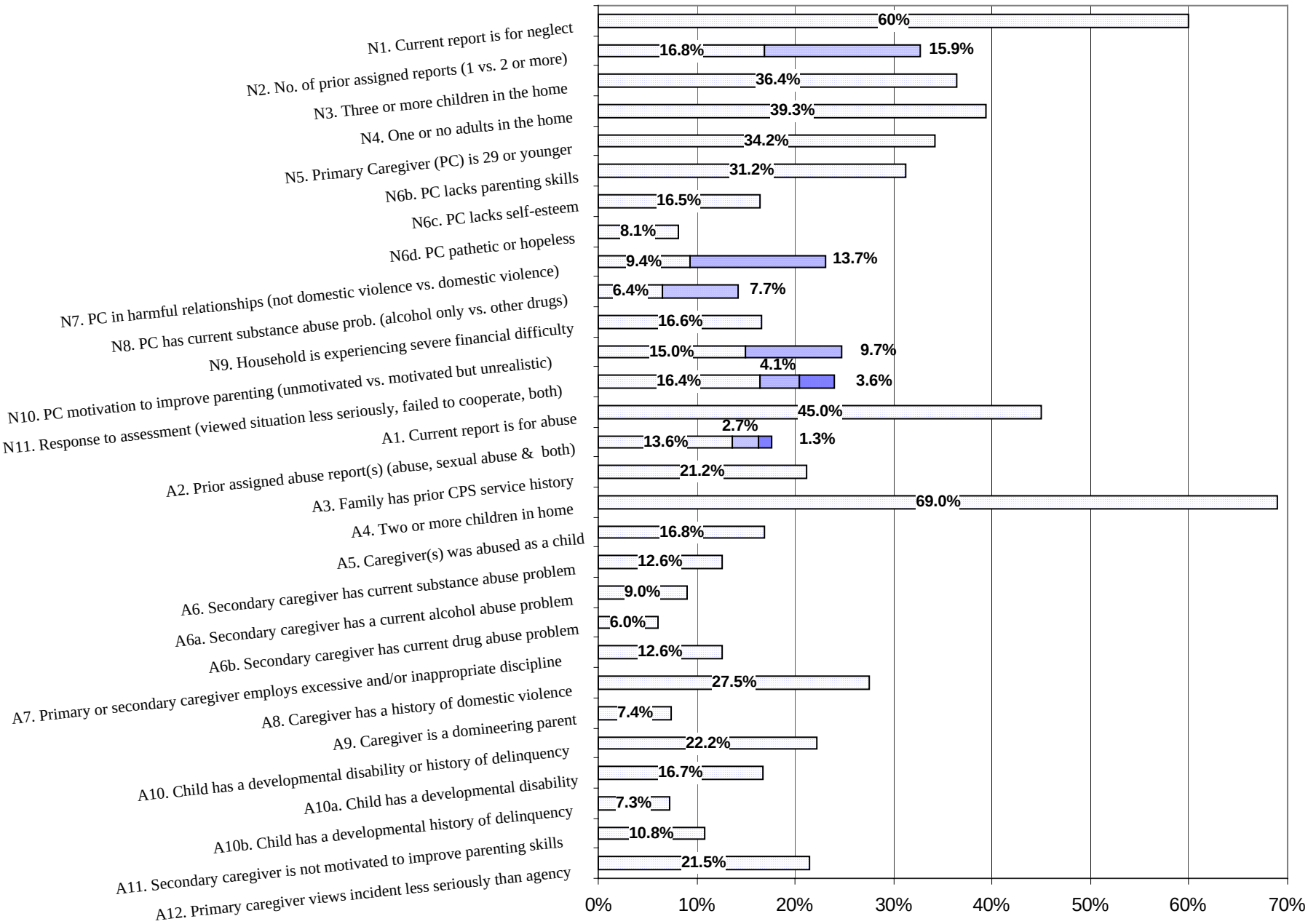


**Figure 2.1. Proportions of Families with Risk Ratings for Each FRA Item**

The predictability of individual items is considered next (Table 2.6). This table contains correlation coefficients like those in Table 2.5, and descriptions of those apply equally here. Correlations of .04 or more and -.04 or less are shown in shaded cells in the table to highlight the stronger relationships.

All the larger correlations are positive except for two items: whether the current report is for neglect (N1) or abuse (A1). A current neglect report predicts fewer future abuse reports while a current abuse report predicts fewer future neglect reports.

**Table 2.6. Correlations of Individual FRA Items with Presenting Problems of Accepted Maltreatment Reports and Total Reports**

| ***Risk items-neglect Subscale*** | *Physical abuse* | *Sexual abuse* | *Neglect* | *Total reports* |
|---|---|---|---|---|
| N1. Current report is for neglect | **-0.07** | **-0.04** | **0.14** | **0.06** |
| N2. Number of prior assigned reports | **0.10** | **0.06** | **0.19** | **0.21** |
| N3. Number of children in the home | **0.06** | **0.05** | **0.08** | **0.10** |
| N4. Number of adults in home at time of report | -0.01 | -0.01 | **0.05** | 0.03 |
| N5. Age of PC | 0.00 | -0.01 | **0.07** | **0.05** |
| N6b. Characteristics of PC: Lacks parenting skill | **0.06** | **0.05** | **0.07** | **0.10** |
| N6c. Characteristics of PC: Lacks self-esteem | 0.01 | 0.03 | **0.05** | **0.06** |
| N6d. Characteristics of PC: Apathetic or hopeless | 0.01 | 0.01 | **0.05** | **0.05** |
| N7. PC involved in harmful relationships | -0.02 | 0.01 | **0.06** | **0.04** |
| N8. PC has a current substance abuse problem | -0.03 | -0.01 | **0.05** | 0.02 |
| N9. Household is experiencing severe financial difficulty | 0.02 | 0.02 | **0.07** | **0.08** |
| N10. PCs motivation to improve parenting skills | 0.03 | 0.03 | **0.07** | **0.08** |
| N11. CG(s) response to assessment | 0.02 | 0.00 | **0.06** | **0.05** |
| ***Risk items-abuse Subscale*** | | | | |
| A1. Current report is for abuse | **0.09** | **0.06** | **-0.13** | **-0.04** |
| A2. Prior assigned abuse reports | **0.07** | **0.06** | **0.05** | **0.10** |
| A3. Prior CPS service history | **0.06** | **0.05** | **0.13** | **0.14** |
| A4. Number of children in the home | **0.07** | **0.04** | **0.05** | **0.07** |
| A5. CG(s) abused as child(ren) | **0.04** | 0.02 | **0.06** | **0.07** |
| A6. SC has a current substance abuse problem | -0.03 | -0.01 | 0.01 | 0.00 |
| A7. PC or SC employs excessive/inappropriate discipline | **0.09** | 0.03 | -0.03 | 0.03 |
| A8. CG(s) has a history of domestic violence | 0.02 | 0.02 | **0.08** | **0.07** |
| A9. CG(s) is a domineering parent | **0.06** | 0.03 | -0.03 | 0.01 |
| A10. Child in the home has DD or history of delinquency | **0.07** | **0.04** | 0.03 | **0.07** |
| A11. SC motivated to improve parenting skills | 0.01 | 0.02 | 0.03 | **0.04** |
| A12. PC views incident less seriously than agency | 0.03 | 0.00 | **0.06** | **0.06** |

The reasons for the cross-correlation (abuse scale with neglect and neglect scale with abuse) in Table 2.5 are apparent. Several of the neglect items (N2, N3 and N6b) are positively related to physical and sexual abuse. At the same time, a larger number of abuse items are positively related to neglect (A2, A3, A4, A5, A8, A12). This appears to

be the reason behind the redundancy between the two subscales in that several pairs of items are very similar and are highly intercorrelated: N2 and A2 (r = .47), N2 and A3 (r = .58), N3 and A4 (r = .47), N7 and A8 (r = .56), N11 and A11 (r = .36), N11 and A12 (r = .72), N10 and A12 (r = .47).[19]

Variables concerning past reports and CPS cases (N2, A2 and A3) were positively correlated, as well. The number of prior assigned reports (none, one, or two or more) is the strongest predictor of new accepted reports. The best predictor of a new child maltreatment report is a past report.

Skepticism is often expressed about demographic items in risk scales. We see, however, that the number of children in the home (N3, A4), number of adults in the home (N4), and the age of the primary caregiver (N5) are all related. These variables are related to family poverty. Families with several children, a single parent or a young parent are more likely to be financially stressed than families without these characteristics and financial stress (N9) is also a predictor of future reports. They are also related to difficulties in parenting—young single mothers with several children, for example, may have greater difficulties in supervision and care of children than older parents, adults in two-parent families and in smaller families. As a rule, these characteristics are more likely to be associated with future neglect, such as failure to provide for food, clothing, shelter and health care as well as lack of supervision.

Caregiver characteristics and history are related to future abuse, neglect or both kinds of reports (N6b, N6c, N6d, N7, N8, A5, A8 and A9). The exception in this analysis may have been the secondary caregiver's substance abuse problems (A6).[20] Similarly, a child with a developmental disability or delinquency (A10) is predictive.

Finally, factors that are directly related to interaction between workers and families, such as cooperative attitudes and behaviors (N11), view of the alleged abuse or neglect (N11, A12) and perceived motivation (N10, A11) are also related.

## Error in Prediction

In studies of test validity, *sensitivity* refers to accuracy of the test in identifying subjects that are the objects of the test—for example, children identified as at risk of dropping out of school who actually drop out or children designated as at risk of developmental and learning disabilities who actually have or develop such disabilities. For the FRA, sensitivity refers to its accuracy in identifying and labeling families that actually have recurring reports of child maltreatment. *Specificity* refers to how well the test rules out (accurately predicts the absence of) subjects that are not the objects of the

[19] Correlations were calculated for the full sample of families.

[20] This item is marked as "no" (0) when the substance abuse is not present and when there is no secondary caregiver. A positive correlation might be found if the analysis were limited to two-caregiver families. This was not possible because "do not know" and "not present" produce the same score in the FRA. (See discussion in Chapter 4.)

test—children identified as not at risk of dropping out who do not drop out. This is shown for CPS risk assessment in the following table. The two shaded cells represent errors. The two unshaded cells represent accurate predictions.[21]

| Risk Assessment Result | Recurrence | |
|---|---|---|
| | Yes | No |
| At Risk | sensitivity rate | false positive rate |
| Not at risk | false negative rate | specificity rate |

For the FRA, specificity refers to its accuracy in identifying and labeling families that do not return. These could easily be determined for an instrument that divided families into two groups, as is suggested by table: those predicted to have recurring reports and those predicted not to have recurring reports. The FRA does not do this. The categorization of families is not *risk/no risk* but *ranked risk categorizations* from low to intensive. The implication would be that no family is without risk of recurrence and that some *accurately identified low-risk families* will be seen again by the agency and the some *accurately identified high-risk families* will not be encountered again.

Another approach is possible, however. As noted, the FRA has been promoted as a systematic guide to focus the limited services of the agency where they are likely to do the most good. The assumption is that services are likely to have the most effects in reducing recurrence (and by implication promoting the welfare and safety of children) among high-risk families. On the other hand, services to low risk families are likely to have less effect. The instrument can be used to segregate families into those that should always be served (high and intensive risk), those that should not be served (low risk) and those that presumptively should not be served but in which further consideration is needed to make that decision (moderate risk). This is reflected in Minnesota Structured Decision Making Policy and Procedures Manual, although as we will see, the practice regarding risk and services varies at the local level.[22] The introduction of the concept of services suggests that service provision must be factored into the calculation of prediction error. The most conservative approach will take into account whether services were offered to families and set those families aside in the calculation of the rates of sensitivity and specificity.

As we have noted, specific services are not recorded in SSIS. However, just as recurrence of maltreatment reports is a rough measure of new child maltreatment,

[21] This table was modeled after one in the book *Calculated Risks: How to Know When Numbers Deceive You*. Gerd Gigerenzer. New York: Simon and Schuster, 2002, p 47. This book offers many examples of risk tools and risk calculations in a lucid style for the non-technical reader.

[22] Specifically, the manual says: "Low-risk cases will be closed. Moderate-risk cases should be considered for closure." Minnesota Department of Human Services representative point out that the actual policy expressed in official state policy documents is more flexible than the statements in this manual.

opening of a formal case after a risk assessment is a proxy for service to families. Services cases, that is, *case-management workgroups* are the condition for the possibility of services and generate direct contacts with families and collateral contacts of ongoing services workers.

Using this method, we limited the count of false negatives to low- and moderate-risk families that received no services and yet had recurrence. The unweighted count of families determined to be low or moderate risk that did not have a case-management workgroup opened immediately following the assessment phase of the case *and also* had one or more recurring reports was 2,863. This yields a specificity rate of 77.2 percent for the families with no services.

We limited the count of false positives to high- and intensive-risk families that received no services and had no recurrence. The unweighted count of families determined to be high or intensive risk that did not have a case-management workgroup opened immediately following the assessment phase of the case *and also* did not have a recurring report was 1,294. This yields a sensitivity rate of 89.7 percent for families with no services.[23]

Combining false positives and false negatives, the overall error rate was 33.1 percent. This percentage applies to the portion of the population that *received no services* (83.2 percent of the 15,100 families in the study sample). The error rate for all families was 27.5 percent. This means that the FRA failed to predict correctly for slightly less than one in every three families.

[23] If no attention were given to services, the specificity rate would be 77.7 percent and the sensitivity rate would be 86.6 percent.

# 3. Reliability of the Family Risk Assessment

Reliability concerns the consistency of an instrument. Are the parts of the instrument consistent with one another? Is it used consistently in practice? We can also ask whether it is used consistently for different categories of families. The latter is considered in Chapter 5 in reference to racial and ethnic subpopulations. In this chapter we are concerned with FRA reliability generally.

We examined two kinds of reliability: internal consistency of FRA items and consistency among workers using the FRA (inter-rater). While other methods of insuring reliability are sometimes used in creating scales, these methods can be used for existing scales.

Internal consistency is concerned with how well the items work together. The idea behind scales, like the FRA subscales, is that individual items measure different aspects of the same thing (in this case, family risk) and that they are intercorrelated. Looking at many families, this means that when one item is marked (as present) for a given family, other items are often (but not always) marked. For example, we might focus on two items in the FRA: say, number of children in the family (N3) and parental substance abuse (N8). In this case we would want to know whether parental substance abuse was checked more often in families with two or more children than in families with less than two children, something that would seem to be implied by the inclusion of the items in the same neglect scale.

Inter-rater reliability is a different concept. It does not refer to the consistency among items in the scale but to the consistency *among workers* using the scale. The question asked in this case is not whether the items hang together (although internal consistency is, in fact, an end result of agreement among individual raters) but whether workers generally interpret and apply the items in the same way. Reliability as consistency among workers (inter-rater) requires a comparison of how two or more worker react to the same family. In this case, it is necessary to find a way to have many workers complete the FRA on a single family or several families.

**Validity and Reliability.** We chose to examine predictive validity before looking at reliability. It is possible to have a reliable instrument, in particular one with good inter-rater reliability, which nevertheless has marginal validity. However, scales of marginal reliability can generally be expected to have less than optimal validity as well.

## Internal Consistency of the FRA

The FRA is composed of individual items, each of which is considered to be a correlate of risk of child maltreatment. The neglect items are proposed as correlates of neglect recurrence and the abuse items are proposed as correlates of abuse recurrence. We saw in Chapter 2 that the data support this contention. However, each item also has a degree of error associated with it, that is, it is not perfectly associated with recurrence. For example, the neglect subscale contains the item "number of children in the home." The prediction is that significantly more of the families with larger numbers of children will experience neglect recurrence. If all families with three or more children had subsequent reports of neglect and none of the families with two or fewer children had such reports this item would be perfectly associated with recurrence. In fact, this is not the case and there is error associated with the item and all other items in the two scales.

We can ask whether the error associated with individual items of each scale is great in comparison to the overall variation explained by combining all the items in the scale. This is the approach that underlies Cronbach's Alpha, the commonly used measure of reliability as internal consistency. This statistic tests whether combining items reduces the overall error, that is, whether the items in combination produce less error than items considered separately.[24] Alpha can be said to measure the homogeneity of a scale and the complementarity of individual items. Alpha must be applied separately in the analysis to the abuse subscale and the neglect subscale. If the items in the neglect subscale each measure some aspect of risk of neglect and, as a rule, they do not contradict one another a high alpha should result. The same holds for the abuse subscale.

An assumption of Alpha is that measures are independent. This precludes the use of alpha for our entire sample of 15,100 families with FRAs because the same workers completed many of these assessments. The largest possible sample for the analysis was 477 families in which each family had one unique worker associated with it.[25]

Before looking at alpha it will be helpful to review how the items of each subscale are intercorrelated, because positive intercorrelation is necessary for a high alpha measure. In Table 3.1, the intercorrelations of neglect and abuse subscale items are shown. To aid in interpreting the table, positive correlations between .1 and .2 are shown in **bold** and those higher than .2 are in ***bold italics***. Several of the items in each scale appear to be interrelated.

As a rule, the relationships in the scale are positive, but some are not and tend to have (positive or negative) values near zero. This appears to be true of item N3 (number

[24] For Cronbach's Alpha, the question is how the variance of individual scale items is related to the variance and covariance of all items. If the summed unique variance (error) is low in relation to the summed variance and covariance a high alpha will result. In addition, a high alpha implies that scale items are positively intercorrelated.

[25] In this procedure, we took the first or earliest FRA of each worker during the 18-month period in which the sample was selected (January 2001 through September 2002). It turned out that the particular risk assessment selected was not a factor in this analysis. Multiple analyses were conducted on other slightly smaller samples with little change in the results.

of children in the home) and to a lesser extent of N4 (Number of adults in the home). Recall from the last chapter (Table 2.6) that each of these items had a certain predictive validity. The lack of inter-item correlation means that when they are high, the other items are low, and when they are low, the other items are high.[26] (Note: the absence of intercorrelations *does not prove* that these items are invalid—validity was illustrated in the previous chapter—but that they do not contribute strongly the final scale score.)

**Table 3.1. Inter-correlations of Neglect and Abuse Subscale Items (477 Families Assessed by 477 Unique Workers)**

| *Neglect Subscale Items* | *N1* | *N2* | *N3* | *N4* | *N5* | *N6B* | *N6C* | *N6D* | *N7* | *N8* | *N9* | *N10* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N1. Current report is for neglect | | | | | | | | | | | | |
| N2. Number of prior assigned reports | .07 | | | | | | | | | | | |
| N3. Number of children in the home | .01 | **.16** | | | | | | | | | | |
| N4. Number of adults in home at time of report | **.14** | .08 | -.05 | | | | | | | | | |
| N5. Age of PC | **.14** | -.07 | -.16 | .06 | | | | | | | | |
| N6b. Characteristics of PC: Lacks parenting skill | .04 | ***.27*** | .07 | .05 | **.10** | | | | | | | |
| N6c. Characteristics of PC: Lacks self-esteem | .03 | **.16** | .00 | .03 | .08 | ***.38*** | | | | | | |
| N6d. Characteristics of PC: Apathetic or hopeless | .07 | ***.20*** | .00 | .07 | .00 | ***.28*** | ***.55*** | | | | | |
| N7. PC involved in harmful relationships | .04 | **.13** | .00 | -.05 | **.16** | **.16** | ***.28*** | **.18** | | | | |
| N8. PC has a current substance abuse problem | **.12** | **.15** | -.03 | .09 | .01 | ***.20*** | ***.20*** | ***.22*** | **.13** | | | |
| N9. HH experiencing severe financial difficulty | **.14** | **.16** | .02 | .06 | .06 | ***.31*** | ***.28*** | ***.28*** | **.18** | ***.21*** | | |
| N10. PCs motivation to improve parenting skills | .01 | ***.28*** | .05 | .08 | -.07 | ***.47*** | ***.32*** | ***.32*** | .09 | ***.22*** | ***.33*** | |
| N11. CG(s) response to assessment | .09 | **.16** | .02 | .04 | -.03 | ***.34*** | ***.23*** | ***.32*** | **.17** | ***.29*** | ***.22*** | ***.40*** |
| *Abuse Subscale Items* | *A1* | *A2* | *A3* | *A4* | *A5* | *A6* | *A7* | *A8* | *A9* | *A10* | *A11* | |
| A1. Current report is for abuse | | | | | | | | | | | | |
| A2. Prior assigned abuse reports | **.14** | | | | | | | | | | | |
| A3. Prior CPS service history | -.01 | ***.40*** | | | | | | | | | | |
| A4. Number of children in the home | .03 | .07 | **.13** | | | | | | | | | |
| A5. CG(s) abused as child(ren) | .01 | -.02 | .09 | .00 | | | | | | | | |
| A6. SC has a current substance abuse problem | .02 | **.12** | **.15** | -.05 | **.14** | | | | | | | |
| A7. PC/SC employs excessive/inapprop. discipline | ***.29*** | ***.23*** | ***.22*** | **.12** | ***.22*** | **.18** | | | | | | |
| A8. CG(s) has a history of domestic violence | .05 | **.16** | .07 | -.02 | ***.31*** | ***.21*** | ***.22*** | | | | | |
| A9. CG(s) is a domineering parent | ***.21*** | **.18** | **.19** | .03 | ***.23*** | **.16** | ***.37*** | ***.21*** | | | | |
| A10. Child has DD or history of delinquency | .08 | ***.25*** | ***.25*** | **.12** | .03 | .02 | .08 | .02 | **.17** | | | |
| A11. SC motivated to improve parenting skills | .04 | ***.22*** | **.18** | .07 | .05 | ***.33*** | **.16** | .07 | **.15** | .07 | | |
| A12. PC views incident less seriously than agency | -.10 | **.13** | ***.24*** | .02 | **.12** | **.15** | **.19** | **.10** | ***.25*** | .04 | ***.27*** | |

Turning to Cronbach's alpha, the alpha scores for the two scales were: Neglect Subscale = .68; Abuse Subscale = .65. As a rule of thumb, alpha values of .70 and higher

[26] This finding illustrates an assumption of the FRA that may be incorrect—that risk of neglect and risk of abuse are each one-dimensional. Under this assumption risk of neglect is one thing on which families can be ranked from low to high. Perhaps risk of neglect (and abuse) must be broken into further subscales. We suspect that this is the case, and that greater consistency and subsequent predictive power might result.

are considered acceptable. These values are slightly less than this generally accepted cut point and indicate marginal internal consistency for the subscales. (Analysis of the neglect subscale with items N3 and N4 removed raised the alpha to .71.)

These findings do not invalidate earlier work in other states on other child welfare populations. Rather, it confirms that reliability is not an immutable characteristic of scales.[27] Further examination of the application of the scale revealed a possible source of inconsistency.

**Reliability Among Workers Conducting Large Numbers of Risk Assessments.** The method of selecting the above sample of 477 assessments, each by a unique worker, was applied again and again to the entire study sample of 15,100 families. Each new sample of assessments excluded assessments selected for previous samples. The effect of this process was to generate samples of workers who were more and more experienced in administering the FRA. For example, the 25th sample included workers who had conducted 25 or more assessments during the 18-month period during which families were sampled for this study. At the same time, the sample size diminished. Little difference was noticed in the first few samples, but at the level of workers who had conducted 10 or more FRAs, a peculiar variation was noted: the reliability of the abuse subscale seemed to decline. At the level of workers who conducted 10 or more assessments, the alpha was .61; at 20 it was .56; at 30 it was .51; at 40 it was .56; and, at 50 it had declined to .46. No corresponding decline was noted in the neglect subscale—for example, among workers who had conducted 50 or more assessments, the alpha remained at exactly .68. The 50th sample, therefore, was examined in further detail to see what might explain this decline.

The 50th sample included only 115 families, because only 115 workers throughout the state had administered 50 or more family risk assessments during the 18 months from 1/2001 through 9/2002. These workers (and families) came from the following counties:

> Anoka-11, Benton-1, Blue Earth-2, Carver-3, Chisago-2, Cottonwood-1, Dakota-12, Hennepin-42, Kandiyohi-2, McLeod-1, Nicollet-2, Olmsted-4, Polk-2, Ramsey-19, Scott-2, Sherburne-1, St. Louis-1, Washington-5, Wright-1, Unknown-1

Among these families, 73.0 percent came from four counties: Anoka, Dakota, Hennepin and Ramsey. Further analysis showed, however, that the decline in abuse-subscale reliability was not a function of families generally from these counties or from larger urban and suburban counties. Nor was it a function of the inclusion of Alternative-Response cases in the sample (since these four counties were a part of the AR Project

[27] Reliability is not an unchanging characteristic of scales because it involves more than the wording of individual items. Reliability is also a function of 1) the interpretation of items by practitioners, 2) the diligence and care of practitioners in using the instrument, 3) cultural differences in populations in which the instrument is used, and especially for risk assessment tools, 4) difference in the socio-economic environment in which the instrument is used. The idea that reliability, once proven, is true forever is a myth. As an instrument is applied beyond the population in which it was originally developed, reliability and validity must be addressed anew.

beginning in 2001).[28] The number of abuse cases was about 10 percent higher among these families compared to the larger sample of 477 families considered above (43.8 to 53). At the same time, scoring of risk on several items declined—for example, families with past CPS service history (A3) declined from 28.7 percent to 8.7 percent. The reliability problem would appear to be a function of either 1) the judgments of workers that conducted many risk assessments or 2) the kinds of families encountered by workers conducting large numbers of assessments.[29]

## Consistency of Scoring among Workers: Inter-rater Reliability

The research plan called for a study of inter-rater reliability. This type of reliability refers to consistency in the application of an instrument by multiple judges. In the case of the FRA, it concerns the consistency of worker judgments and scoring in the application of the instrument. Within the present study the goal was to determine inter-rater reliability of the FRA generally (the analysis in this chapter) and to determine whether the instrument was used consistently across subpopulations of interest (see Chapter 5).

To study this experimentally, workers must be presented with essentially the same family. The method we used to accomplish this was the presentation of *written case vignettes*. Vignettes are descriptions of families, family situations and behaviors such as an assessment worker might encounter. The basic method involves presenting many workers with the same vignette and asking them to use the FRA to assess the risk of the family described.[30]

To accomplish this, two case descriptions (written vignettes) were created—one that included several child-neglect risk characteristics and the other with several physical-abuse risk characteristics. The characteristics were selected to coincide with those utilized by the FRA. Two versions of the each description were created—one designed to reflect lower-risk conditions and the other higher-risk conditions. In this way four separate vignettes were created: lower-risk neglect, higher-risk neglect, lower-risk abuse

[28] Reliability of the FRA among AR families was generally lower. AR represents the lower-risk portion of the CPS population. The FRA was developed to apply to the entire CPS population.

[29] Another possibility considered but impossible to verify was that the staff IDs in SSIS for these assessments represented more than one worker—for example, IDs of supervisors or of data entry personnel. If true, this might result in duplication of workers in the sample. This could not be verified, but in any event, would be unlikely to have accounted for such a decrease in internal consistency.

[30] There are certain strengths and weakness of the vignette method. The advantage of using vignettes is that many workers can respond to exactly the same set of family and family-member characteristics. Each worker reads the same written description. The disadvantage lies in the artificiality of written descriptions, particularly for an instrument like the SDM Family Risk Assessment. The FRA is not used to determine the risk level of case descriptions of families but of actual families that workers visit and observe and with whom they interact. Observation and interaction cannot be reproduced via a written vignette. However, vignettes can be used as purely cognitive tools to reveal biases in judgments, and that is the intent of this study.

and higher-risk abuse.[31] The four vignettes were further modified to reflect the five racial/ethnic groups that were the focus of this study: Caucasian, African American, American Indian, Southeast Asian and Hispanic. This is further described in Chapter 5, but in the present analysis we averaged across subpopulation identification. *The vignettes can be found in Appendix A, and the reader should turn to them and read them before proceeding further.*

All Minnesota workers who had completed a Family Risk Assessment during the period from October 2003 through March 2004 were surveyed. The survey was web-based. Responses were received from 459 Minnesota workers, whose staff ID was associated with at least one FRA completed during this period. Workers responded by clicking on a hyperlink contained in the email. This link directed workers to a set of Internet-survey pages containing a neglect vignette followed by the FRA risk questions and an abuse vignette followed by the same FRA questions. The links contained codes that had been randomly assigned. The codes controlled the particular combination of family characteristics that workers read about. Each worker responded to two vignettes as follows:

| | Vignette Combination[32] | | Number of Workers |
|---|---|---|---|
| | 1 | 2 | |
| Pair 1. | Neglect lower risk – | Abuse lower risk | 131 |
| Pair 2. | Neglect lower risk – | Abuse higher risk | 108 |
| Pair 3. | Neglect higher risk – | Abuse lower risk | 112 |
| Pair 4. | Neglect higher risk – | Abuse higher risk | 108 |

By asking workers to respond to two vignettes, standard statistics designed to test inter-rater reliability could be used effectively. Each worker completed the FRA two times—first for vignette 1 and then for vignette 2. Recall that the FRA contains a neglect subscale and an abuse subscale. Two subscale scores were calculated for each vignette, and thus, two neglect subscale scores and two abuse subscale scores were generated for each worker. As mentioned earlier the neglect subscale ranged from 0 to 20 while the abuse subscale ranged from 0 to 16. The frequencies of subscale scores are listed in Table 3.2.

Table 3.2 is complex, but is shown to illustrate the distribution of risk and neglect subscale scores for various combinations of the vignettes presented to workers. The numbers in the cells are percents of the total number of workers (second from bottom row). For example, looking at the first number in the upper left, 4.6 percent of 131

[31] Why only four? Four vignettes permitted certain the basic risk dimensions of the FRA to be varied. This was adequate for study of racial and ethnic bias. A more comprehensive approach that systematically varied all 25 FRA neglect and abuse items (see Figures 5.3 and 5.4) was precluded by two factors. Vignettes had to be relatively short because workers have limited time for reading them. The number of variations in items had to be kept small because the study was limited to a few hundred Minnesota workers. These factors militated against large numbers different and exhaustively detailed vignettes.

[32] See Chapter 5 (Table 5.4) for the number of workers per racial/ethnic identification within each type of vignette.

workers or 6 workers produced a score of 2 for the neglect subscale score of the neglect lower-risk vignette. Most workers produced final scale scores within about a 4-point range of one another. Looking again at the first neglect column, nearly 9 out of 10 workers had produced scores from 3 to 6 (87.8 percent).

**Table 3.2. Responses of Four Groups of Workers to Vignette Pairs**

| | *Pair 1* | | | | *Pair 2* | | | | *Pair 3* | | | | *Pair 4* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Vignette →*** | *Neglect lower risk* | | *Abuse lower risk* | | *Neglect lower risk* | | *Abuse higher risk* | | *Neglect higher risk* | | *Abuse lower risk* | | *Neglect higher risk* | | *Abuse higher risk* | |
| ***Subscale →*** | *A Neg.* | *B Abu.* | *C Neg.* | *D Abu.* | *E Neg.* | *F Abu.* | *G Neg.* | *H Abu.* | *I Neg.* | *J Abu.* | *K Neg.* | *L Abu.* | *M Neg.* | *N Abu.* | *O Neg.* | *P Abu.* |
| ***Score ↓*** | % | % | % | % | % | % | % | % | % | % | % | % | % | % | % | % |
| 0 | | 16.0 | 1.5 | | | 16.7 | | | | | | | | | | |
| 1 | | 56.5 | 6.9 | | .9 | 65.7 | | | | 1.8 | 11.6 | | | .9 | | |
| 2 | 4.6 | 21.4 | 6.1 | 2.3 | 6.5 | 15.7 | | .9 | | 10.7 | 19.6 | .9 | | | | 1.9 |
| 3 | 21.4 | 3.1 | 14.5 | 2.3 | 19.4 | 1.9 | 1.9 | 1.9 | | 8.0 | 10.7 | 3.6 | .9 | 4.6 | | |
| 4 | 38.9 | 1.5 | 19.8 | .8 | 32.4 | | 7.4 | 1.9 | 1.8 | 20.5 | 20.5 | | 2.8 | 31.5 | 3.7 | .9 |
| 5 | 11.5 | 1.5 | 13.7 | 3.1 | 17.6 | | 3.7 | 2.8 | 10.7 | 17.9 | 17.0 | 1.8 | 2.8 | 33.3 | 7.4 | .9 |
| 6 | 16.0 | | 15.3 | 3.1 | 14.8 | | 11.1 | | 8.0 | 15.2 | 12.5 | 10.7 | 8.3 | 25.0 | 13.0 | .9 |
| 7 | 4.6 | | 9.2 | 11.5 | 8.3 | | 15.7 | .9 | 20.5 | 9.8 | 3.6 | 14.3 | 15.7 | 3.7 | 16.7 | 7.4 |
| 8 | 1.5 | | 6.9 | 20.6 | | | 17.6 | 12.0 | 17.9 | | 2.7 | 21.4 | 10.2 | .9 | 21.3 | 10.2 |
| 9 | .8 | | 3.8 | 19.1 | | | 18.5 | 14.8 | 15.2 | 8.9 | 1.8 | 25.0 | 18.5 | | 22.2 | 16.7 |
| 10 | .8 | | 2.3 | 23.7 | | | 13.0 | 23.1 | 9.8 | | | 16.1 | 14.8 | | 8.3 | 25.9 |
| 11 | | | | 13.7 | | | 7.4 | 23.1 | 8.9 | | | 6.3 | 10.2 | | 3.7 | 22.2 |
| 12 | | | | | | | 2.8 | 15.7 | 3.6 | | | | 7.4 | | 1.9 | 11.1 |
| 13 | | | | | | | | 2.8 | 1.8 | | | | 5.6 | | | 1.9 |
| 14 | | | | | | | .9 | .9 | 1.8 | | | | 2.8 | | .9 | |
| 15 | | | | | | | | | | | | | | | .9 | |
| 16 | | | | | | | | | | | | | | | | |
| 17 | | | | | | | | | | | | | | | | |
| 18 | | | | | | | | | | | | | | | | |
| 19 | | | | | | | | | | | | | | | | |
| 20 | | | | | | | | | | | | | | | | |
| Number of workers | 131 | 131 | 131 | 131 | 108 | 108 | 108 | 108 | 112 | 112 | 112 | 112 | 108 | 108 | 108 | 108 |
| Mean Score | 4.4 | 1.2 | 4.8 | 8.5 | 4.4 | 1.0 | 8.0 | 9.6 | 8.2 | 4.5 | 3.9 | 8.1 | 9.0 | 4.9 | 7.9 | 9.6 |

These ranges of scores are not unusual for instruments like the FRA. The ranges indicate some disagreement (error) but general agreement because they are small and the distributions are not excessively skewed or bipolar. The average (mean) scores in each case are printed in the bottom row of the table. Inspection reveals that the means are consistent across the different groups of workers. For example, the neglect-lower-risk vignette produced a mean neglect score of 4.4 and a mean abuse score of 1.2 from the first group of 131 workers. It produced a mean neglect score of 4.4 and a mean abuse score of 1.0 from the second group of 108 workers.

The method of comparing can be explained by looking again at Table 3.2. The reader should focus on the letters (A, B, C…) at the top of each subscale column). We compared A with C, B with D, E with G, F with H, and so on. Thus, for example, the neglect subscale scores (A) for the neglect-lower-risk vignette were compared to the neglect subscale scores (C) for abuse-lower-risk vignette. This involved comparisons of scoring by 131 workers. *This approach considers whether workers use the FRA in such a way that it distinguishes low-risk and high-risk families and whether they generally classify together families of the same risk level.*

The statistics used for this are called intraclass correlation coefficients (ICC).[33] These were as follows:

| | Vignette Combination | | ICC | ICC |
|---|---|---|---|---|
| | 1 | 2 | Neglect | Abuse |
| Pair 1. | Neglect lower risk – | Abuse lower risk | .65 | .99 |
| Pair 2. | Neglect lower risk – | Abuse higher risk | .99 | .99 |
| Pair 3. | Neglect higher risk – | Abuse lower risk | .99 | .98 |
| Pair 4. | Neglect higher risk – | Abuse higher risk | .93 | .99 |

The ICC scores represent averages across workers and are high in most cases, indicating good agreement. This shows that in spite of the variations in Table 3.2, workers used the FRA in such a way that it produced the same kinds of differences in risk when two different vignettes were considered.

Some qualifications are in order. These findings concern *the abuse and neglect subscales* of the FRA, not the final risk assessment, which is considered below. The high ICCs demonstrate that workers are interpreting *written descriptions of families* in the same way. It *implies* that workers share common interpretations of individual items on the FRA. This analysis *does not demonstrate* whether two workers encountering the same family in a real setting would interpret what they observed and heard in the same way. When investigators and family assessment workers visit families, they do not find written vignettes describing the family history and behavior posted on the front door. Rather, they must observe and interpret and, in effect, write their own vignettes. The variability that would occur in that process *is not* considered in this analysis. It would, of course, be possible to study interrater reliability in reference to real families by sending two workers out who compete the FRA separately. This procedure was considered but would have been prohibitively expensive and beyond the scope of the present study.

Finally, we show the final risk scores assigned to each vignette by the four groups of workers. The percentages in each risk category can be seen in Table 3.3. In each case

[33] A two-way random effects model was used that assumed absolute agreement among the ratings. This assumes that the comparative scores between any two vignettes should be in actual agreement, not simply relative agreement. The model also assumes that the workers are a random sample of all workers who could use or will use the FRA and that the vignettes are random samples of the kinds of families that would be encountered by workers. The latter is a stretch, since the four vignettes could hardly be considered to be representative the total array of families and family problems encountered by Minnesota CPS workers. This is the weakness, unavoidable in the present study, of limiting the survey to only four vignettes.

there is a clear modal category, that is, a cell in which the percent is much higher than the others. For example in the first column, 61.8 percent of 131 workers rated the family described in the neglect-lower-risk vignette as low risk. However, in each case a second category can be seen that is also substantial, ranging from one-fifth to two-fifths of the workers responding. Given that we have already seen in Table 3.2 that actual subscale scores of workers were generally within about 4 points of one another, why does such variation occur in the final assignment of risk?

**Table 3.3. Final Risk Levels Assigned to Families Described in Vignettes**

| | *Pair 1* | | *Pair 2* | | *Pair 3* | | *Pair 4* | |
|---|---|---|---|---|---|---|---|---|
| ***Vignette →*** ***Final Risk Score*** ↓ | *Neglect lower risk* | *Abuse lower risk* | *Neglect lower risk* | *Abuse higher risk* | *Neglect higher risk* | *Abuse lower risk* | *Neglect higher risk* | *Abuse higher risk* |
| **Low** | 61.8% | .8% | 57.4% | | 1.8% | .9% | .9% | |
| **Moderate** | 35.1% | 7.6% | 42.6% | 6.5% | 33.9% | 5.4% | 26.9% | 1.9% |
| **High** | 3.1% | 54.2% | | 28.7% | 60.7% | 71.4% | 63.9% | 36.1% |
| **Intensive** | | 37.4% | | 64.8% | 3.6% | 22.3% | 8.3% | 62.0% |
| Number of workers | 131 | 131 | 108 | 108 | 112 | 112 | 108 | 108 |

The answer lies in the scoring method used for the FRA. The method produces scale scores for abuse and neglect. These are summative scores that can be treated as interval-level measures from 0 to 20 and 0 to 16. Then, as noted earlier, the scores are converted into cruder categorical measures—twice. First come the categorical risk of abuse and categorical risk of neglect scores (low through intensive) and then comes the final risk score (the highest of the abuse and neglect categories), which is shown in Table 3.3. The final measure is a rank order measure with many tied scores. The scoring process proceeds from the refined to the rough.

Looking back at Table 3.2, this loss of information can be easily seen. In the first column of percent (A), we see that 64.9 percent (4.6+21.4+38.9) have scores of 4 or less. Since 4 is the cut-point for low risk of neglect, all these scores produced a low risk of neglect score. But the next set of scores (11.5 percent) is in the moderate risk of neglect category (scores of 5 to 7). One point difference on the neglect subscale kicked a family from low to moderate risk of neglect—the traditional problem associated with turning real numbers into ranked categories.

These findings and our interpretations of them do not mean that the FRA is not useful. It has a measure of validity and reliability. Its usefulness hinges on how it is used with families, as noted at the end of Chapter 2. Utilization is considered in the next chapter.

# 4. The Role of FRA in Practice

In the case-specific survey, workers were asked a small number of questions about the role played by the risk assessment instrument and risk scores in social work practice in their agency. Interviews were also conducted in 12 of the 20 counties that participated in the Alternative Response project, and questions about the Family Risk Assessment instrument and its role in practice were put to CPS workers and supervisors. Below is a summary of what was learned from the surveys and interviews. This is followed by two analyses of the extent of agreement and disagreement of FRA items when the instrument was applied to experimental and control families in the Alternative Response evaluation.[34]

**Survey Questions.** In the survey, workers were asked, "How much did the family risk assessment affect whether and how the agency responded to the family?" Table 4.1 summarizes their responses. As can be seen one in five workers responded that the FRA score had no effect on the agency's response to the family. About twice that number said it was a "minor factor" and one in three said it was a major factor, with a small percentage (2.4 percent) saying it was the "most important factor."

**Table 4.1. Degree of Effect FRA Scores on Agency Response to Families**

| | |
|---|---|
| Not at all | 20.2% |
| A minor factor | 44.0% |
| A major factor | 33.3% |
| Most important factor | 2.4% |

In order to understand more about the impact FRA scores have on practice, counties were divided into three groups: 1) the metro counties Hennepin (Minneapolis) and Ramsey (St. Paul), 2) the other 18 counties that participated in the AR project, and 3) the remaining 67 counties in the state.[35] Table 4.2 shows how workers in each of these three groups of counties answered the question and the data are informative.

[34] See Chapters 1 and 6 for a fuller discussion of the nature of the Alternative Response Project and evaluation.

[35] The 20 counties that participated in the AR project beginning in February 2001 are listed in Chapter 1. This distinction is useful because there were significant changes in CPS practice under AR.

Workers from the large urban counties of Hennepin and Ramsey were more likely (55.7 percent) than the others to say that an FRA score was a major factor or the most important factor in determining agency response to the family. From interviews in Hennepin County we know that the FRA score was relied on in the screening of reports for an Alternative Response or traditional investigation—only low or moderate scores could be screened for AR. On the other hand, it was workers in other AR counties who more often than other workers said the FRA score did not have any influence on the agency's response, placing a higher priority, apparently, on worker judgment based on the extended and holistic assessment interview. However, even these other AR counties were more likely to say the score played a minor or even major role in shaping the response. Workers in the other non-AR project counties (almost all with small to very small caseloads) were least likely to say the score played a major role or was the most important factor in guiding CPS response.

**Table 4.2. Degree of Effect FRA Scores have on Agency Response to Families by County Groups**

| | Ramsey & Hennepin | Other AR project counties | Non-AR project counties |
|---|---|---|---|
| Not at all | 14.3% | 26.8% | 17.0% |
| A minor factor | 30.0% | 37.9% | 54.3% |
| A major factor | 48.6% | 33.3% | 27.7% |
| Most important factor | 7.1% | 2.0% | 1.1% |

During interviews, workers were asked which of the following most shapes whether families receive services and what types of services they might receive: the FRA score, the results of the Strengths and Needs Assessment, or the worker's judgment independent of either the FRA or the S&N. Workers were not of one mind on this. Many said they relied more of the SDM Family Strengths and Needs instrument than the FRA for guiding their practice, although most said their judgment remained the deciding factor. One worker noted that the results of the FRA generally validated her judgment about families. Another said she used the FRA as a guide, while a third said, "I don't have time to question it." But a number do question it, and they do not rely on it or they keep the results in the background. One said, "The risk assessment will sometimes produce moderate or high scores (for example if the mom was in a foster home herself) based on information that isn't relevant anymore."

There was not a significant difference in the response of workers of greater or lesser experience (defined as more than 5 years and less than 5 years CPS experience) across all counties. There was some difference, however, between workers who only did traditional investigations and those who only did AR. The former were more likely to say they did not use the FRA score at all (20.7 percent) in determining whether or how they would respond than AR workers (10.2 percent of whom said they did not use it at

all). The latter were more likely to use it in as "minor" factor (52.5 percent) in guiding practice compared with 39.0 percent of traditional (TR) workers.

In the survey, workers were also asked: What is the expectation in your office regarding the opening of ongoing cases for families with low-risk scores? As can be seen in Table 4.3, just over half of the workers responded that cases were rarely opened for these families. On the other hand, a third of workers indicated that ongoing cases were sometimes opened on families with low risk scores and a few (3 percent) said they were often opened.

Workers in Hennepin and Ramsey counties were more likely than workers in other counties to say that cases were rarely opened if risk scores were low. Workers in the 18 other counties involved in the AR project were somewhat more likely to indicate that ongoing cases might be opened for low-risk families. (See Table 4.4) As was the case above, the experience of the worker had little to do with whether or not families with low-risk scores had ongoing cases opened. On the other hand, AR workers were significantly more likely than TR workers to say such families might have cases opened. This indicates that workers are following agency policies rather than simply deciding for themselves independent of county policy whether or not to open ongoing cases.

**Table 4.3. Action Taken for Families with Low Risk Scores**

| Action Taken for Families with Low-Risk Scores | % |
|---|---|
| Cases rarely opened | 53.4% |
| Cases sometimes opened | 34.3% |
| Cases often opened | 3.0% |
| Risk scores aren't significant consideration | 5.9% |
| Other | 3.4% |
| | 100.0% |
| *Most common "other" responses:* | |
| *Traditional investigation, no; AR, sometimes yes* | *1.3%* |
| *There are broader considerations beyond the risk assessment* | *0.8%* |
| *If family desires/requests/wants* | *1.3%* |

**Table 4.4. Action Taken for Families with Low Risk Scores In Different County Groups**

| | Ramsey & Hennepin | Other AR project counties | Non-AR project counties |
|---|---|---|---|
| Low-risk cases rarely opened | 69.2% | 47.1% | 52.7% |
| Low-risk cases sometimes opened | 20.5% | 37.9% | 36.4% |
| Low-risk cases often opened | 2.6% | 5.7% | 0.9% |
| Risk scores not play a sig. role in whether we open cases | 2.6% | 4.6% | 8.2% |
| Traditional investigation, no; AR, sometimes yes | 2.6% | 2.3% | 0.0% |
| There are broader considerations beyond the risk assessment | 0.0% | 1.1% | 0.9% |
| If family desires/requests/wants | 2.6% | 1.1% | 0.9% |

When is the FRA Completed? During interviews conducted as part of site visits for the AR evaluation, AR workers were asked when they completed the FRA instrument. Some said they took the tool with them into the home and completed it during the interview. Some found the instrument a convenient explanation for why certain questions were being put to the family. Many said that after using the instrument over a number of assessment visits with families, they were able to complete the instrument at a later point (in their car following the first visit or back in their office) and did not need to have the questions in front of them when they interviewed families. But there was a larger difference in timing, and this varied from one office to another and among workers in the same office. For some workers the assessment represented the state of affairs within the family as the worker initially found them. For others it represented the family at a later point in time but prior to the end of the assessment period—in other words, after some intervention efforts may have occurred. And for still others, the FRA score represented the state of the family at the end of the assessment process—and after, at least in some cases, some substantial efforts were made by the worker to assist the family.

**Table 4.5. When was the FRA Completed by the Worker?**

| | Ramsey & Hennepin | Other AR project counties | Non-AR project counties | Total |
|---|---|---|---|---|
| During or shortly after the first home visit | 43.6% | 35.6% | 40.0% | 39.0% |
| Later but before the end of the assessment | 33.3% | 34.5% | 40.0% | 36.9% |
| At the end of the assessment or shortly after | 23.1% | 29.9% | 20.0% | 24.2% |

**Worker Responses in Interviews.** In interviews conducted during site visits to county CPS offices, researchers asked county workers and supervisors their views on the Family Risk Assessment instrument. Many described a generally positive attitude toward the tool. Many would agree with workers who said the instrument was "fairly accurate" and "works in a majority of cases." And many saw the instrument as "better and less subjective than what we used to do" and believed it had "increased consistency" in the assessment process among workers because it "standardizes questions that are asked by workers." As noted above, some workers said simply that they trusted it and used it to guide their case planning, and some spoke of using it primarily in combination with other SDM tools, particularly the family strengths and needs assessment tool.

While a majority of the workers interviewed expressed generally positive views about the FRA instrument, many had specific problems with it. Some saw the tool as "too automatic" in its scoring. A number commented that high scores were produced in family situations in which they did not believe risk to be high. (A supervisor said, "It happens all the time that workers say there's no risk, but the score is high.") Some of this was due to the fact that the instrument does not take into account when something may have occurred. (One worker said, "The risk assessment will sometimes produce moderate or high scores—e.g. a mom was in a foster home—based on information that

isn't relevant anymore.") A prior report or open case "within the last six months is significantly different from one 10 years ago." Others said that the instrument does not take into account mitigating circumstances or protective factors that may reduce risk—"The score may come out high even if there are counter-balancing factors, such as coping skills or extended family support." The instrument in such instances was viewed by social workers as producing too many false positive predictions.

A number of workers noted that there were certain combinations that always yield higher scores—"A woman who is 29 with 3 children and a victim of domestic violence a long time ago will get a higher risk score than a woman who is 30 but was a victim of violence last week."[36] Of all the items on the instrument, the ones most often mentioned as problematic by workers were number of children (N3 and A4), and age of parents (N5).

The weight given to certain items and the automatic nature of the scoring was seen by some as introducing sub-cultural biases—"Minorities get higher scores because of lower income and more children." And "we can't override because of other information that would lower risk; that's a problem." One worker said she thought that "cultural validity may be a problem on certain items, like number of children and age. These items can skew the score." Another said "there are some differences in community standards, but these are not raised by the risk assessment." And another commented that there had been complaints about the instruments from the Native American community."

At the same time there are issues seen as implicated in risk to children not captured by the instrument. One social worker said, "Mental health and chemical dependency are overriding issues that aren't addressed in the tools."[37] Another said, "The assessment tool is incomplete. It doesn't fully capture risk for all populations."

A number of workers indicated the need to take the age of the child into account when assessing risk. The instrument, one said, "doesn't take into account the age of children. The same issue may be different depending on the age of children." Not taking the age of children into account was seen by some as producing both false positive and false negative predictions of risk.

The instrument does not distinguish between a worker's lack of knowledge on an item and a "no" response to a question. One worker noted that the assessment he fills out "has a lot of zeroes in it because I don't know or can't find out some of the information." More than one worker noted that they would like to be able to record "suspicions"

[36] This quote contains a misstatement. Both cases should be scored as 1 on item A8 (domestic violence), although the worker's primary point is valid that a one-year difference in age and one more child can lead to a higher risk score.

[37] Two items in the instrument (N8 and A6) address substance abuse. This worker was referring in part to the need to give more weight to severe conditions, such as methamphetamines addiction and psychiatric disorders when these are encountered. It should also be remembered that the FRA provides for workers to override risk to the next risk level (e.g., moderate to high) if they can justify that action to a supervisor (see Chapter 1).

someplace when they are not convinced a caregiver is being honest in answering a question.

A number of workers said they thought there was too much subjectivity in the items, issues where workers might see things differently. Most often mentioned in this regard were items on caregiver cooperation and motivation (N10, N11, A11 and A12).

## Difference in FRA Item Scores Resulting from Different Approaches to Families

The Alternative Response (AR) Project was described in Chapter 1. As part of the evaluation conducted by IAR, an analysis was carried out on the FRA item scores given to experimental and control group families. Both groups of families were judged appropriate for AR by project county screeners. Control families received a traditional investigation, with the primary emphasis on determining whether the allegations of the child maltreatment report were correct. Experimental families received a non-adversarial family assessment, with an emphasis on family engagement and participation in decision-making and with a focus on a broader array of family needs. Investigations ended with findings or no findings of maltreatment. Family assessments involved no findings of maltreatment. Continuing casework with families after investigations was generally mandatory but after family assessments it was voluntary. As a rule family assessments were more family-friendly, and systematic surveys of families and workers as part of the AR evaluation revealed significantly more positive attitudes and opinions of families regarding the CPS assessment.[38]

Another change that was found in the research was that AR workers sometimes initiated interventions *during the assessment process* before a case-management workgroup (ongoing service case) was opened for the families.

The percents of experimental and control families that received risk ratings on each of the FRA items are shown in Figure 4.1. There are three different types of items in the FRA. Some involve information that is clearly *objective*, such as whether or not the current report is for neglect and the number of children in the home. Other items require *subjective* judgments to be made by the social worker, such as whether or not a caregiver is cooperative or unmotivated. For some items this judgment depended on the quality of *interaction* between the worker and family members. In addition, there are items somewhere in between—they could be considered to be objective but can only be determined if sufficient *probing* is done by the social worker during the assessment interview and even then requires an element of judgment. In the chart, the items are

[38] Minnesota Alternative Response Evaluation: Final Report, Institute of Applied Research, November 2004, www.iarstl.org. See in particular Chapter 4, where family satisfaction, attitudes toward workers, sense of participation in decision-making and emotional responses were analyzed. In the same chapter, workers perceptions about families were analyzed as well. The similarity of the experimental and control groups is discussed in the Chapter 8 of that report. The discussion of the data represented in Figure 4.1 can be found in Chapter 9.

rearranged into these three types. As one might expect, the largest differences occurred between experimental and control families for the subjective items and in particular those that were dependent on the nature of the interaction that took place with the family.

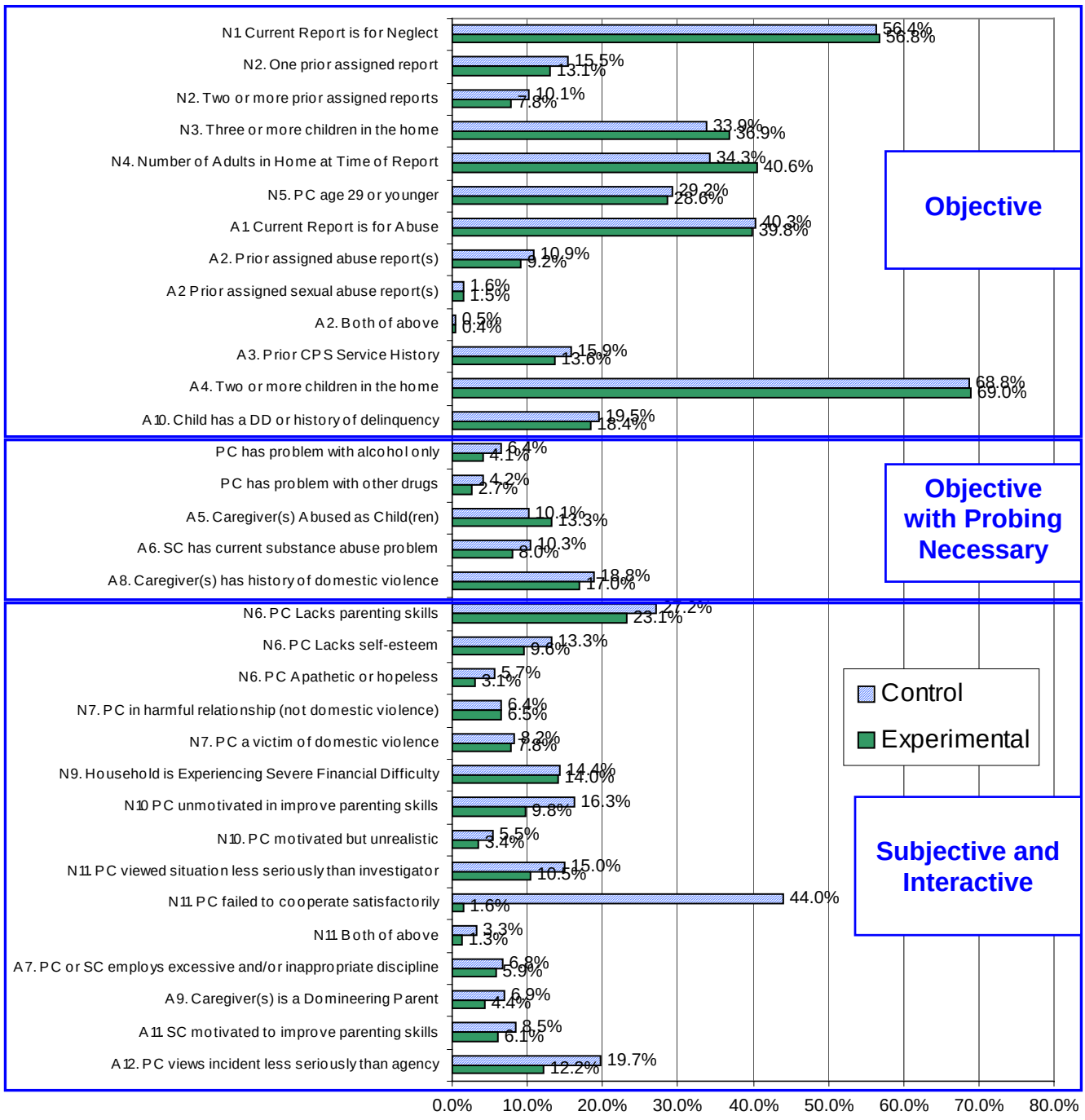


**Figure 4.1. FRA Item Assessment of Experimental and Control Families in AR Study**

*Increased family engagement* under the AR approach is one possible explanation of the differences observed on items. For instance, experimental families were considered to be more cooperative, more motivated, viewing the incident as seriously as the agency, as having better parenting skills—all differences that may have been affected by the more positive and friendly approach of AR family assessments.

Another possible explanation for the differences may have been early intervention in families by AR workers. As indicated above, workers did not all complete the instrument at the same time. In some cases, the assessment reflected the state of the family when first encountered by the worker. In other cases, the assessment reflected the state of the family at a later time. *Assessments by AR workers after they had intervened in families* are another possible explanation of the difference observed on items.

Finally, workers using the two approaches may not focus identically on all the material covered by the FRA tool. In an investigation, a worker must determine whether there is evidence that the report should be substantiated. This may lead her to probe more extensively in certain areas. Under AR, a worker does not make a finding on the report but can attend immediately and directly to issues of child and family well-being leading, in turn, to a focus on factors the worker believes are more relevant to that process.

These differences might be considered to show errors in assessment of risk, but only if risk is considered to be a characteristic of families that is difficult to change. An alternative explanation is that better scores (lower-risk on subjective and interactive items) by experimental families *accurately reflect lowered conditions of risk.* If cooperation, motivation, and situational attitudes generally are indeed measures of risk, then improvements in these areas resulting from the approach to families represent true reduction of risk of future child maltreatment.

## FRA Item Scores and Study Vignettes

Another way to learn about the use of the FRA instrument in practice is to examine the scores given on individual items to the families presented to workers in the four vignettes that we introduced in Chapter 3. This method can reveal differences in the attention of workers to various pieces of information presented to them about a case and in the judgments they make about it as recorded on the FRA tool.

On some FRA items workers were in nearly unanimous agreement. On others a minority disagreed about the presence or the absence of risk. In Table 4.6, the items of greatest disagreement are shown ranked from highest to lowest for each of the four vignettes. The table shows the minority percentages of respondents for each item that *disagreed*. The larger the minority percentage, the more disagreement was present among workers. The table shows only disagreements greater than 20 percent of workers. Percentages ranged as high as 48 percent. Those higher than 30 percent are shown in bold. The type of item is shown in the left column (O = objective; OP = objective, proving necessary; S/I = subjective and interactive).

## Table 4.6. Minority Disagreement on FRA Risk Items (Vignette Survey)

| Type | Item | | Type of Vignette | | | |
|---|---|---|---|---|---|---|
| | | | Lower-risk neglect | Higher-risk neglect | Lower-risk abuse | Higher-risk abuse |
| O | N1 | Whether the current report is for neglect. | | | | |
| O | N2 | Number of prior assigned reports. | | 24% | | 28% |
| O | N3 | Number of children in the home | | | | |
| O | N4 | Number adults in the home at the time of the report | | | **36**% | **36**% |
| O | N5 | Age of the primary caregiver | | | | |
| S/I | N6b | Primary caregiver lacks parenting skills | 26% | | 23% | 28% |
| S/I | N6c | Primary caregiver lacks self-esteem | | **37**% | **38**% | 24% |
| S/I | N6c | Primary caregiver apathetic or hopeless | | | | **47**% |
| S/I | N7 | Primary caregiver involved in a harmful relationship. | | | **40**% | **37**% |
| OP | N8 | Primary caregiver has a current substance abuse problem. | | | | |
| S/I | N9 | Household is experiencing severe financial difficulty. | | | **38**% | |
| S/I | N10 | Primary caregiver is motivated to improve parenting skills. | 26% | **32**% | 21/8% | **32**% |
| S/I | N11 | Cooperation of caregiver. | | 25% | **35**% | **35**% |
| O | A1 | Whether the current report is for abuse. | | | | |
| O | A2 | Types of prior abuse reports (physical or sexual). | | | | **31**% |
| O | A3 | Prior CPS service history. | | **37**% | | **32**% |
| O | A4 | Number of children in the home. | | | | |
| OP | A5 | Whether the caregivers were abused as children. | 21% | 22% | **43**% | **44**% |
| OP | A6 | Secondary caregiver has a current substance abuse problem. | | | | |
| S/I | A7 | Any caregiver employs excessive and/or inappropriate discipline. | | | | |
| OP | A8 | Caregiver has a history of domestic violence. | | | **35**% | |
| S/I | A9 | Caregiver is a domineering parent. | | 27% | **46**% | **48**% |
| S/I | A10 | Child in the home has a development disability or history of delinquency. | | | | |
| S/I | A11 | Secondary caregiver motivated to improve parenting skills. | | | 20% | |
| S/I | A12 | Primary caregiver views incident less seriously than agency. | | **31**% | **31**% | **42**% |

Some variation can be seen for objective items, which may reflect misreading of the vignette materials. However, the greatest variation occurred on the subjective items that involve judgments on the part of social workers. This suggests that there are differences in how workers reach conclusions about certain items or, perhaps, the extent

to which they attend to different issues. This kind of variation may reflect differences in the relative weight workers give to some of the items, another kind of scoring.

At the same time, these are responses to vignettes and not flesh and blood people. In the vignettes, information available to workers was limited to what researchers gave them. Some workers may have been unwilling to reach certain conclusions based on the limited information that can be provided in a vignette. Nonetheless, the extent of disagreement illustrates potential sources of variation in the application of the risk scale.

# 5. Differences Among Minority Subpopulations

This chapter concerns the utilization of the Family Risk Assessment in specific racial and ethnic subpopulations of Minnesota families. Three major topics are considered: 1) how the FRA is used and scored, 2) differences in predictive validity, and 3) possible bias among workers who utilize the instrument.

In its request for proposals, the state indicated that five racial/ethnic subpopulations should be compared: Caucasian, African American, American Indian, Southeast Asian and Hispanic. These categories are utilized in the following analyses.

Racial and ethnic identifications were derived from fields within the Social Services Information System (SSIS). This study is primarily concerned with families and only secondarily with individual family members. SSIS contains no family racial designation but only codes for individuals within families. In general, we depended on the racial designation of the head of the family as the primary indicator of race or ethnicity. The family head is usually listed as the head of the assessment workgroup (the working case of an investigation or family assessment). Occasionally, a child is designated as the head of a workgroup. This was always checked against family-relationship codes and age in related records.

The race field in SSIS contains several general designations and is accompanied by a sub-race field that provides greater detail. For example, if the race field is American Indian, the sub-race field may be the particular tribe or if the race field is Asian, the sub-race field provides a particular country of origin or other Asian ethnic identity. A separate field is used in SSIS to designate Hispanic individuals.

Of the 15,100 families in the study, 10,217 were Caucasian, 3,221 were African American, 729 were American Indian and 377 were Southeast Asian. In 556 families, race could not be determined or fell into other racial categories (e.g., other Asian). In addition, there were 781 families in which the primary caregiver was Hispanic. Among the latter, 693 (94.0 percent) were designated as Caucasian with the remainder distributed proportionately across the other racial groups. In some instances, children in families had different racial designations than caregivers. There were 588 Caucasian families with an African American child, 142 with an American Indian child and 17 with a Southeast Asian child. Among African American families, 80 had a Caucasian and 14 an American Indian child. Among American Indian families, 84 had a Caucasian child, 37 had an African American child, and 1 had a Southeast Asian child. Among Southeast Asian

families, 10 had a Caucasian, 2 had an African American and 1 had an American Indian child. In general, these numbers were too small for separate analyses.

## Risk Assessment Scores by Racial/Ethnic Subpopulations

When the FRA was examined for the five subpopulations, the following final risk categorizations were obtained (Figure 5.1). These are the final risk scores on the *initiating* reports, that is, the first report on each family during the 18-month target period. The overall risk assessment scores are rather similar for Caucasian and African American. Southeast Asian and Hispanic families appeared slightly more often in the low to moderate risk categories. American Indian families were higher risk, primarily because of greater proportions of families in the high and intensive risk categories.

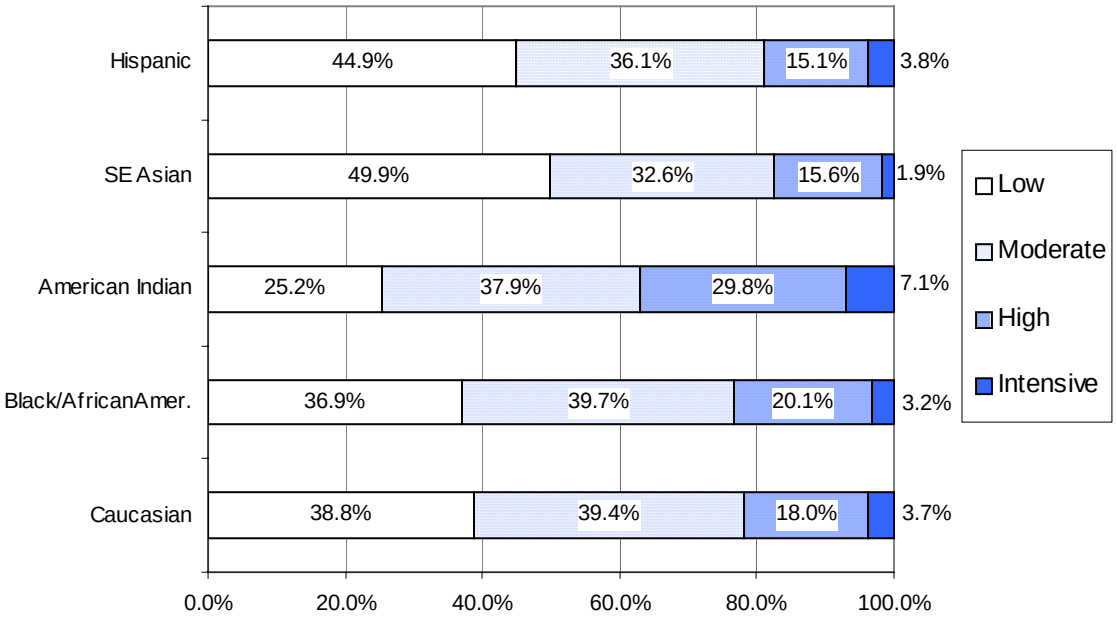


**Figure 5.1. Final Family Risk Categorization by Race/Ethnicity**

The differences as well as the similarities in final risk can be understood through an analysis of the FRA subscales and individual risk items. As noted previously, the subscale scores consist of totals of the neglect risk items and the abuse risk items. If the worker marks no risk items in a particular group the score is zero. The maximum neglect score possible after adding up the neglect items is 20; the maximum abuse score is 16. The neglect and abuse averages (means) for each subpopulation are shown in Figure 5.2. Each group, of course, had a higher neglect average than abuse average, simply because the neglect score range is greater.

Several things are apparent in Figure 5.2. Hispanic and Caucasian averages were very similar for both neglect and abuse. Southeast Asian families scored lower on

neglect-risk but were very similar to Hispanic and Caucasian on abuse-risk. African American families and American Indian Families averaged higher average neglect-risk scores but African American had a comparable abuse-risk average. American Indian families were on average rated higher on both abuse-risk and neglect-risk.

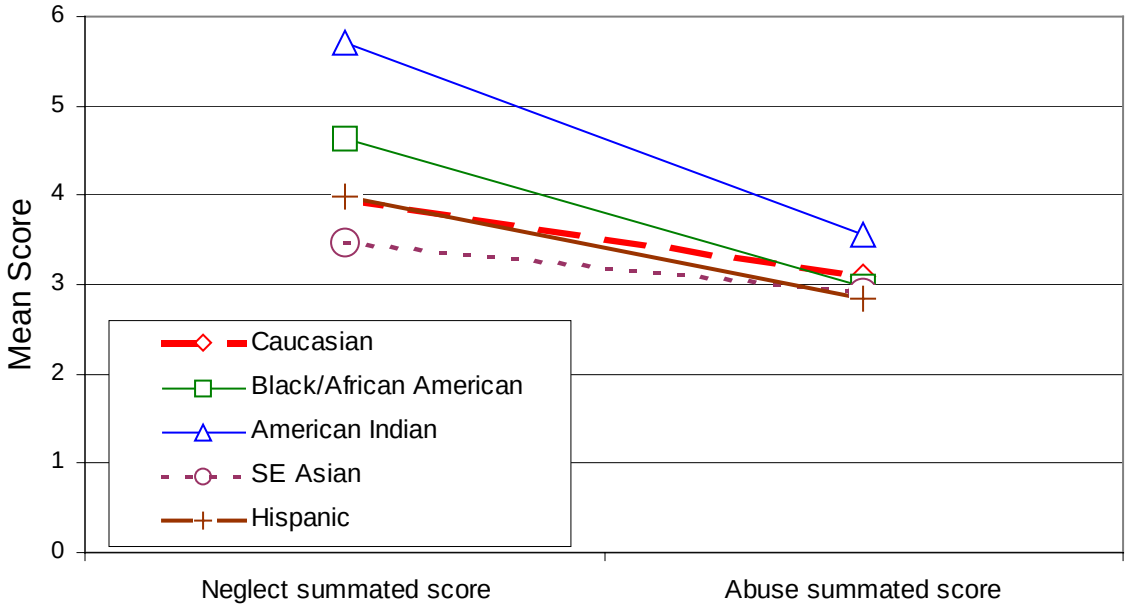


**Figure 5.2. Summated Scores of Neglect and Abuse Subscales by Race/Ethnicity**

The symbols in Figure 5.2 represent averages. The distribution of risk scores varied for each group. While American Indians had the highest risk scores, they also had the most variation in risk scores (Standard Deviation = 3.63 for neglect and 2.44 for abuse), that is, families within this subpopulation were more spread out across the full range from low to intensive risk on the two scales.

The differences among the subpopulations are shown in greater detail in the following two charts (Figures 5.3 and 5.4). The first contains line graphs for each of the racial/ethnic groups for the 11 neglect items while the second shows a similar graph for the 12 abuse items. In these graphs the height of each point on the line represents the percent of families that were rated as having particular risk characteristic. (Risk items with more than one level have been collapsed to simplify the graph).

The line for the majority Caucasian subpopulation is bold and dashed (with solid diamond points) in the graphs so that it may be used as a reference line. The lines for minority populations can be compared to Caucasian line. The minority subpopulation will now be discussed.

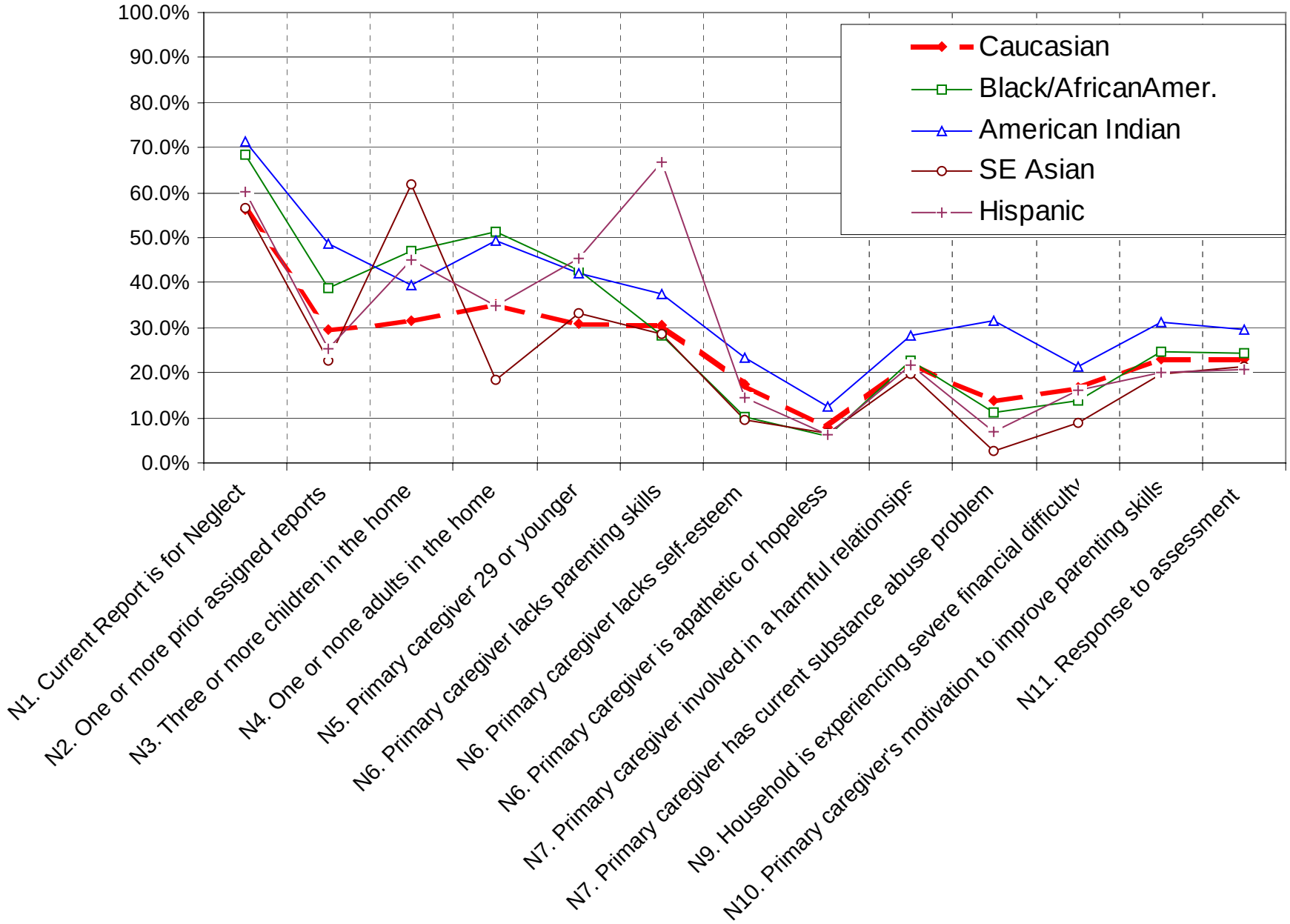


**Figure 5.3. Individual Neglect Risk Items by Race/Ethnicity**

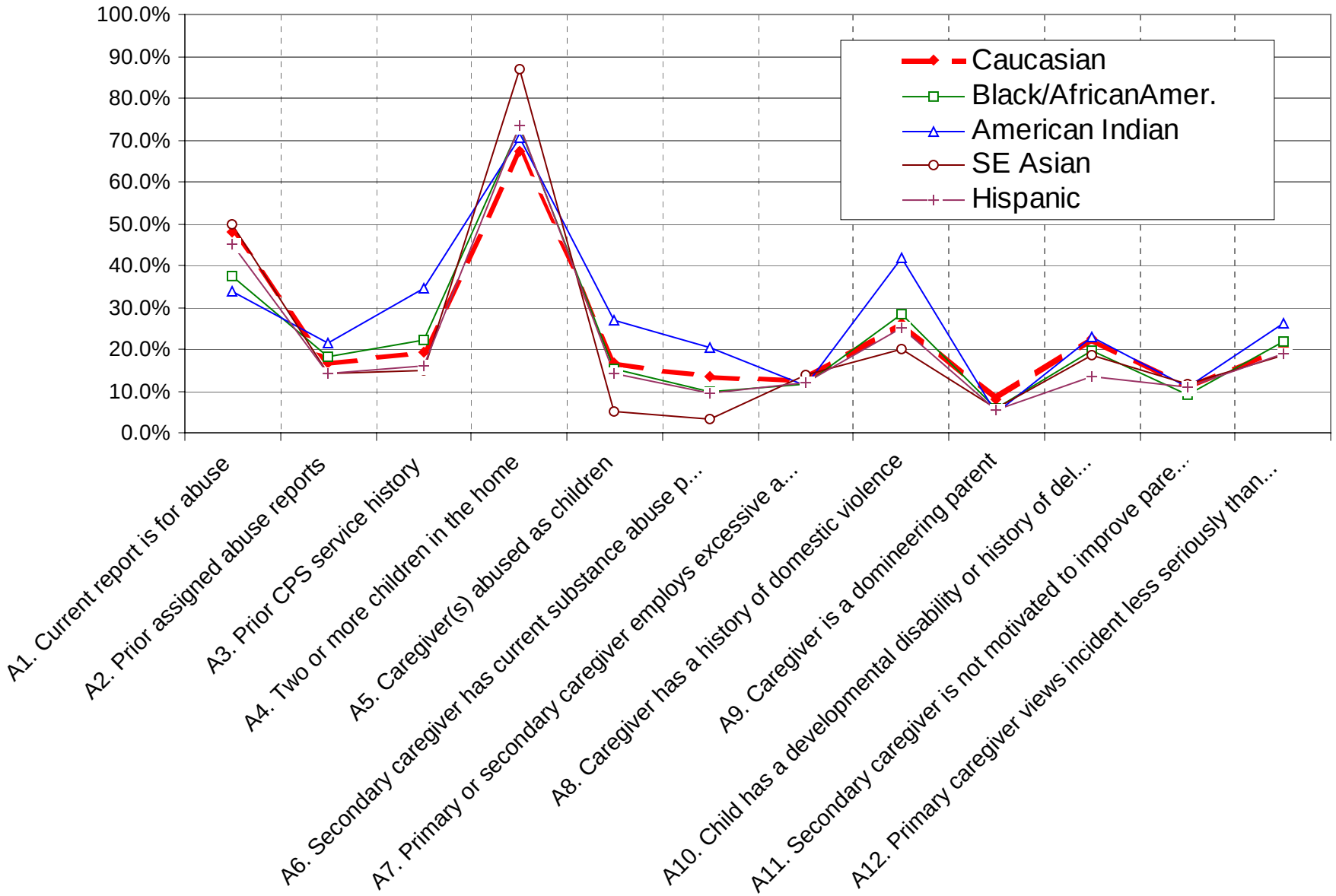


**Figure 5.4. Individual Abuse Risk Items by Race/Ethnicity**

**Hispanic.** Families in this subpopulation differed substantially from majority Caucasian families on only three items. They more often included a primary caregiver younger than 30 years of age (N5), had three or more children in the home (N3), and they were more likely to be considered to lack parenting skills (N6). Otherwise, Hispanic scores on individual items were quite similar to but slightly lower than Caucasian, which accounts for lower overall proportions among low and moderate risk families compared to the majority population.

**Southeast Asian.** As noted, Southeast Asian families were rated at overall lower risk than other subpopulations (Figure 5.1), and this would seem to have arisen from the lower risk ratings for child neglect (Figure 5.2). However, Southeast Asian families were rated higher on two items that referred to the number of children in the home at the time of the report (N3 and A4). This reflects larger family size and perhaps extended family living arrangements because these families were less often rated as having only one adult in the home (N4). Southeast Asian families were also less often rated (compared to Caucasians) as having a primary (N7) or secondary (A6) caregiver with a substance abuse problem, a history of domestic violence (A8), having severe financial difficulties (N9), caregiver lacking self-esteem (N6), or caregivers being abuse as children (A5). Less frequent checks on these items account for lower overall risk scores for Southeast Asian families.

**American Indian.** Families in this subpopulation were more often considered to show risk (compared to Caucasian families) in all the neglect risk items. Indeed, these families were more often considered to show risk compared to all other subpopulations on most of the neglect risk items—the only exceptions being fewer children in the home and more adults in the home compared to African American families. On risk of abuse items, risk ratings were more variable in comparison to Caucasian families. However, visibly larger percentages of families were rated with a prior CPS history (A3), caregivers abused as children (A5), secondary caregiver with a substance abuse problem (A6), and domestic violence history (A8). Primary caregivers were considered to have substance abuse problems (N7) at about twice the rate of Caucasian and African American families and secondary caregivers at a somewhat lower relative difference (A6). Within the FRA, this item for the primary caregiver is broken into alcohol only or other drugs with or without alcohol. It was the first of these two items (alcohol abuse only) on which American Indian families differed from other subpopulations (for example, 16.9 percent versus 9.9 percent for Caucasians).

**African American.** The higher overall risk of African American families on the neglect risk subscale was due exclusively to first five items in the scale. They were more likely to have a neglect report (N1), to have prior neglect reports (N2), to have more children (N3), to have only one adult in the home (N4), and to have a young primary caregiver (N5). Workers checked about the same proportions (or less) on other neglect items as Caucasian families. Past reports for child neglect is contingent on the kinds of reports received more frequently on African American families. The other three factors are demographic in nature, and the predictive power of demographic items was verified in Chapter 2. On the abuse subscale the percentages follow those of Caucasian families

very closely, except for the first item. Because they were reported more often for neglect, they were reported less often for abuse (A1).

**Risk Determination Limited to Counties with Substantial Minority Subpopulations.** Figures 5.1 through 5.4 compare the minority subpopulations with majority Caucasian population throughout the state. Some counties have only tiny proportions of one or the other types of minority families (see Chapter 1), it can be argued that they should be set aside in analyses. For example, many Minnesota counties have very few African American families. Perhaps a better analysis would compare African American families with Caucasian families only in counties with detectable populations of both groups. To examine this we limited the analysis to counties in which the client population (in this analysis) contained 5 percent or more of African American, American Indian or Hispanic families and, because the number of SE Asian families were small, to counties with *any* SE Asian families. The results are shown in Table 5.1

Each of the four sections of Table 5.1 reflects a separate analysis based on different combinations of counties. The size of the entire client populations varied considerable among these counties. The sets of counties were:

**Hispanic (5 percent or more):** Blue Earth, Brown, Carver, Chippewa, Clay, Cook, Freeborn, Hennepin, Jackson, Kandiyohi, McLeod, Meeker, Nicollet, Nobles, Norman, Polk, Ramsey, Red Lack, Scott, Sibley, Waseca.

**SE Asian (any):** Anoka, Benton, Carver, Chisago, Cottonwood, Dakota, Hennepin, Jackson, Nobles, Olmsted, Ramsey, Scott, St. Louis, Swift, Washington, Winona.

**American Indian (5 percent or more):** Big Stone, Carlton, Cass, Clearwater, Cook, Hennepin, Hubbard, Itaska, Mahnomen, Pine, Polk, Ramsey, Scott, St. Louis, Yellow Medicine.

**African American (5 percent or more):** Anoka, Benton, Blue Earth, Dakota, Hennepin, Olmsted, Ramsey, Scott, Washington.

This analysis had the potential to show whether the rating of risk relative to the majority population was more or less extreme *in counties in which workers were likely to encounter both minority and majority families.* This can be determined by comparing the percentages in this table with the corresponding percentages in Figure 5.1. As can be seen, the percentages changed only slightly in this analysis. The only pattern visible was a slight shift (about 2 to 3 percent) of American Indian risk toward lower risk while Caucasian families shifted toward slightly higher risk. If this shift is a valid measure it may indicate that American Indian and Caucasian families in the same office caseload tend to be somewhat more similar than in a statewide comparison. Nonetheless, the rates of high and intensive risk among American Indian families remained significantly and substantially higher.

**Table 5.1 Final Risk Determinations Limited to Counties with Substantial Minority Populations of Each Type**

| Comparisons | | *Final Risk Determination* | | | |
|---|---|---|---|---|---|
| | | *Low* | *Moderate* | *High* | *Intensive* |
| Non-Hispanic | Number | 3328 | 3270 | 1541 | 265 |
| | Percent | 39.6% | 38.9% | 18.3% | 3.2% |
| Hispanic | Number | 292 | 226 | 93 | 26 |
| | Percent | 45.8% | 35.5% | 14.6% | 4.1% |
| Caucasian | Number | 3148 | 3200 | 1290 | 211 |
| | Percent | 40.1% | 40.8% | 16.4% | 2.7% |
| SE Asian | Number | 187 | 129 | 54 | 5 |
| | Percent | 49.9% | 34.4% | 14.4% | 1.3% |
| Caucasian | Number | 1520 | 1432 | 604 | 116 |
| | Percent | 41.4% | 39.0% | 16.4% | 3.2% |
| American Indian | Number | 150 | 210 | 156 | 37 |
| | Percent | 27.1% | 38.0% | 28.2% | 6.7% |
| Caucasian | Number | 2733 | 2718 | 1172 | 211 |
| | Percent | 40.0% | 39.8% | 17.1% | 3.1% |
| African American | Number | 1159 | 1248 | 617 | 96 |
| | Percent | 37.1% | 40.0% | 19.8% | 3.1% |

## Validity of the FRA for Subpopulations

The issue of the validity of the FRA was addressed in Chapter 2 by examining the predictive power of the instrument. Family risk of child maltreatment is a predictive concept—the higher the risk, the greater the chance of future maltreatment. The same analysis was conducted for each of the five subpopulations in the study.

Percentages of families with any recurring reports during the 24-month follow-up are shown for each category of the risk assessment in Figure 5.5.[39] The statistic of association (Somers' D) for each type was Caucasian: .113 ($p < .0001$); African American: .103 ($p < .0001$); American Indian: .034 (not significant); Southeast Asian: .180 ($p < .0001$); Hispanic: .165 ($p < .0001$). Recall that Somers' D ranges from –1 to +1. Each racial group shows a statistically significant association in the range of that for the general population (.116) except American Indian.

Examining the differences in percents for each subpopulation in the chart reveals that substantial differences exist between recurrences of low-risk and moderate-risk families in most cases. The differences were 12.0 percent for Caucasian, 13.5 percent for African American, 20.1 percent for SE Asian, and 14.5 percent for Hispanic. However, it was only 8.2 percent for American Indian families. Low-risk American Indian families returned at a higher rate that those of other subpopulations and the difference between recurrences of low-risk and moderate-risk was less.

[39] See Table 2.1.

The difference between moderate and high risk existed but was less dramatic in most cases. In addition, intensive risk families in three cases (Caucasian, African American and American Indian) returned less often than high risk. As indicated in Chapter Two, this may be due to the small numbers of families in these categories.

The FRA appears to best distinguish low, moderate-high, and intensive among SE Asians and low, moderate, and high/intensive among Hispanic. It is evident that it does a poorer job among American Indian families, as reflected in both the chart in Figure 5.5 and the non-significant associational statistic (Somers' D = .035). The problem for this group appears to be the lack of predictability among low-risk families. Among American Indian families the rate of "false negatives" was high since two of every five low-risk families had a new report within two years. We return to this below.

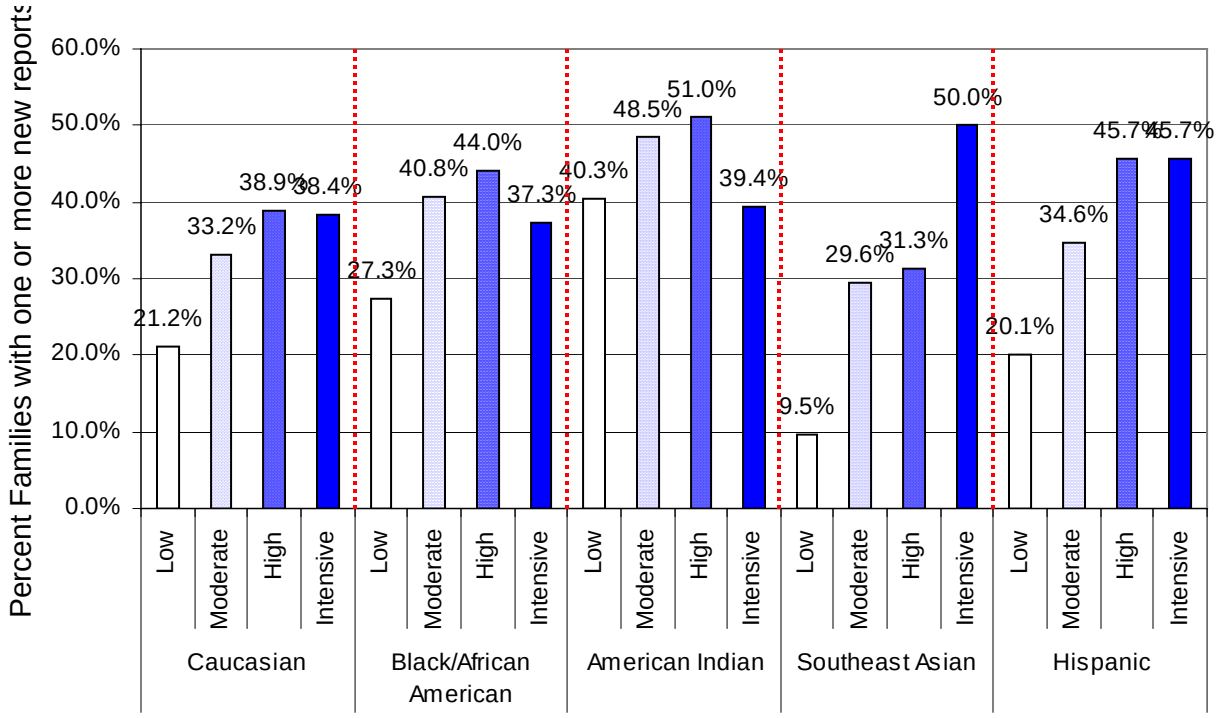


**Figure 5.5. Risk Assessment by Recurrence of Any Accepted Maltreatment Report during 24 Months for Subpopulations**

The abuse and neglect subscales of the FRA were also analyzed following the method described in Chapter 2 (correlational statistics).[40] The same types of correlations are shown in Table 5.2. All five sub-tables show that higher scores on the abuse subscale predict higher recurrence of abuse reports. Likewise the neglect subscale generally predicts more new neglect reports. The exception is American Indian families where the

[40] See Table 2.5.

statistic is weakest (r = .027). This is also a clue to the predictive weakness of the FRA for this subpopulation.

**Table 5.2. Correlations of Risk Subscale Scores with Presenting Problems of Accepted Maltreatment Reports for Subpopulations (Pearson r)**

| *New Reports of:* | | *Neglect subscale* | *Abuse subscale* |
|---|---|---|---|
| Caucasian | Physical Abuse | 0.042 | 0.124 |
| | Sexual Abuse | 0.034 | 0.081 |
| | Neglect | 0.191 | 0.053 |
| African American | Physical Abuse | -0.012 | 0.097 |
| | Sexual Abuse | 0.029 | 0.064 |
| | Neglect | 0.099 | 0.068 |
| American Indian | Physical Abuse | 0.010 | 0.118 |
| | Sexual Abuse | 0.003 | 0.048 |
| | Neglect | 0.027 | -0.024 |
| SE Asian | Physical Abuse | 0.068 | 0.220 |
| | Sexual Abuse | 0.112 | 0.089 |
| | Neglect | 0.335 | 0.217 |
| Hispanic | Physical Abuse | 0.053 | 0.094 |
| | Sexual Abuse | -0.035 | 0.065 |
| | Neglect | 0.206 | 0.097 |

## Individual FRA Items

It is unnecessary to present the analysis of the predictability of each individual item in the FRA for each of the five subpopulations. In general, the relationships resembled those shown for the entire sample of study families in Chapter 2.[41]

A closer examination would appear to be in order, however, for American Indian families. Do any items stand out as different for these families than for other subpopulations and are any items particularly weak predictors of recurrence? The *weak relationship of the neglect subscale with recurrence* of neglect reports was the first clue. Examination revealed no major differences from other subpopulations among the abuse subscale items. On the neglect subscale, several items were found which appeared to be weak predictors for these families. They are shown in Table 5.3. The differences in percentages were small and non-significant. Each of these items involves a judgment call by the worker completing the form:

- Whether the caregiver has adequate parenting skills

[41] See Table 2.6.

- Whether and in what way the primary caregiver is involved in harmful relationships
- The presence and type of substance abuse problems
- The degree of financial difficulty the household is experiencing
- Caregiver motivation and cooperation

**Table 5.3 FRA Items that Distinguished Recurrence Weakly for American Indian Families**

| FRA Neglect Items | *Responses and Percent Families with Recurring Reports* | | |
|---|---|---|---|
| N6b. Characteristics of PC: Lacks parenting skill | *No* | *Yes* | |
| | 45.6% | 47.6% | |
| N7. PC involved in harmful relationships | *No* | *Yes, not domestic violence* | *Yes, domestic violence* |
| | 46.1% | 45.6% | 49.3% |
| N8. PC has a current substance abuse problem | *No* | *Yes, alcohol only* | *Yes, other drugs w/wout alcohol* |
| | 46.0% | 48.3% | 47.0% |
| N9. Household is experiencing severe financial difficulty | *No* | *Yes* | |
| | 45.6% | 49.4% | |
| N10. PCs motivation to improve parenting skills | *Motivated and realistic* | *Unmotivated* | *Motivated but unrealistic* |
| | 46.1% | 52.9% | 39.3% |

It is possible that these items simply are not related to new incidents of child abuse and neglect for the American Indian population. This is an unlikely explanation because these characteristics are known to be predictors of new child maltreatment reports across many different CPS populations. *A more likely explanation for the lack of predictability of these items is that the "no" responses represent errors. This would mean that these characteristics were actually present in families but workers did not indicate their presence on the risk scale.*

To test this hypothesis, we examined intake narratives, maltreatment narratives and, when placements occurred, the reasons for placement of children for *low-risk* American Indiana families *with new reports*. There were 185 American Indian families rated as low-risk at the time of the initiating report. Of these, 70 had new reports. Workers narratives were available in SSIS for 41 of the latter families.[42] A content analysis was conducted of both initial narratives and those that were associated with later reports. The analysis was focused on risk characteristics. Of these 41 low-risk families with subsequent maltreatment reports the following was found:

[42] Narrative information is also added to written files and is not consistently available in SSIS across all Minnesota counties. Most of the narratives that were present recounted events that occurred prior to and during the family assessment, including circumstances of families and characteristics of family members.

From the 41 narratives, families were found with:

- Adult drug abuse = 16 (in initial case: 5, in later case: 11)
- Adult alcohol abuse = 11 (in initial case: 2, in later case: 9)
- Adult or mental illness = 7 (in initial case: 3, in later case: 4)
- Domestic violence = 8 (in initial case: 5, in later case: 3)
- Child Behavior Problems = 10 (in initial case: 5, in later case: 5)

In addition other possible sources of risk were present: in 3 cases a parent was later incarcerated for criminal activity, 1 parent was later noted to be a prostitute, and 3 families were later found to have unstable residences. Child behavior problems were in most cases unspecified reasons for removal and placement, and therefore, were probably either status offenses or delinquency.

Looking at the initial FRA, these 41 families had the following checked:

- Drug abuse = 1
- Alcohol abuse = 1
- Domestic violence = 4
- Child emotional problem = 1
- Child delinquency = 1

In each instance, other items were checked on the FRA, but these usually were the demographic items (number of children, age of parent, etc.) or counts of past reports or cases.

As is apparent, in many instances we discovered these characteristics in later narratives created sometime months *after the initial FRA and case.* It is possible that these represent new risk factors that were not present at the time of the FRA. They may also represent characteristics that were present initially but were overlooked or ignored.

There are no items in the FRA referring to adult mental illness, residential problems, or criminal activities. During interviews (see Chapter 4), workers indicated that the first of these might be included in the risk assessment. All three have possible relations to child neglect. Concerning the remainder, the more likely reasons for why the items were not checked in the risk scale are: 1) the worker completed the risk scale before discovering the problem and did not revise the scale, 2) the worker suspected the problem was present but was not certain enough to indicate it on the risk scale, 3) the circumstances of the assessment were such that they worker did not learn of these problems. The second and third of these were both mentioned by workers during interviews as sources of difficulty in completing the FRA (see Chapter 4).

While these issues were found in American Indian families rated as low-risk, our perusal of narratives revealed that the same thing occurred across the spectrum of subpopulations and may account for some proportion of recurrence among families scored as low-risk. If we assume that factors we have suggested are the sources of error in predictive validity three possible remedies suggest themselves. First, workers might be encouraged to raise the risk level when they suspect the presence of risk factors, including known risk factors not included in the FRA. Second, workers should be encouraged to revise the assessment when new information is discovered or they should wait to complete the risk assessment when the investigation or family assessment is over.

Third, some indication should be made of risk assessments based in part on unknowns so that *low risk is not confused with unknown risk.*

Another consideration is whether families are treated differently based on the risk assessment scale. The special problem noted among American Indian families is that more families rated as low risk had later reports of child maltreatment. Another predictive criteria is opening of case-management workgroups at a later date. This is considered next.

**Case-Management Workgroups.** In Chapter 2, we also examined the predictability of the FRA in reference to case-management workgroups opened subsequently. The same analysis is shown in Table 5.4 but broken by the five subpopulations. In this case, unlike the previous examples of recurring maltreatment reports, the differences are statistically significant ($p < .0001$) for all subpopulations, including American Indian.

**Table 5.4. Risk Assessment by Occurrence of Any Case-Management Workgroup during 24 Months by Subpopulation**

| *Risk Level* | *Caucasian* | *African American* | *American Indian* | *Asian American* | *Hispanic* |
|---|---|---|---|---|---|
| Low | 9.2% | 9.7% | 11.5% | 7.4% | 7.9% |
| Moderate | 16.4% | 10.8% | 14.1% | 17.6% | 19.2% |
| High | 34.4% | 32.0% | 29.4% | 39.1% | 35.7% |
| Intensive | 34.9% | 32.8% | 30.8% | 50.0% | 40.0% |
| Total | 17.8% | 15.7% | 19.5% | 16.8% | 17.8% |

While American Indian families that were scored as low-risk have more than predicted later reports, new post-assessment services cases following these reports are opened in roughly the same proportion as other populations. A greater "equity" appears to have resulted in the case-management and service response to subpopulations.

## Inter-rater Reliability of Assessments of Families in Subpopulations

This topic was considered in Chapter 3. As indicated, this type of reliability refers to consistency in the application of an instrument by multiple raters or judges. In the case of the FRA, it concerns the consistency of worker judgments and scoring in the application of the instrument. Within the present study the goal was to determine inter-rater reliability of the FRA generally (Chapter 3) *and to determine whether the instrument was used consistently across subpopulations of interest.* Do Minnesota assessment workers apply the FRA similarly for Caucasians, African-American, American Indians, Southeast Asians and Hispanics? *The following section repeats some of the text from Chapter 3, because some readers may not wish to backtrack as they read this chapter.*

To study inter-rater reliability experimentally workers must be presented with essentially the same family in which the only difference is the racial and ethnic identity. The method we used to accomplish this was *written case vignettes*. Vignettes are written description of families, family situations and behaviors such as an assessment worker might encounter. The basic method involves presenting many workers with the same vignette and asking them to use the FRA to assess the risk of the family described.[43]

To accomplish this, two case descriptions (written vignettes) were created—one that included several child-neglect risk characteristics and the other with several physical-abuse risk characteristics. The characteristics were selected to coincide with those utilized by the FRA. Two versions of the each description were created—one designed to reflect lower-risk conditions and the other higher-risk conditions. In this way four separate vignettes were created: lower-risk neglect, higher-risk neglect, lower-risk abuse and higher-risk abuse.[44] The four vignettes were further modified to reflect the five racial/ethnic groups that were the focus of this study: Caucasian, African American, American Indian, Southeast Asian and Hispanic. This was accomplished by varying the designation of "race" in a data table associated with each vignette and by using names that might be associated with the subpopulations. *Readers who have not yet referred to Appendix A to examine the vignettes should do so now before reading the remainder of this section.*

Studying the reactions of workers to subpopulations using vignettes seemed feasible because only one aspect of the vignette had to be varied—the race or ethnicity of the family. Creating the four versions of the vignettes just described permitted the general research question to be expanded to whether the instrument might be consistently used across subpopulations under conditions of alleged abuse and neglect and low-risk versus high-risk conditions. Workers reading any one of four versions of the vignette all read exactly the same thing except for this racial-ethnic identification. If bias existed in the minds of workers it might be manifested in *overall differences in scoring of the FRA.*

The questions that can be asked and answered through this method are:

- Can statistically significant difference be found in the final scoring of the FRA and final ratings of risk when racial/ethnic identity of the family is varied?

[43] There are certain strengths and weakness of the vignette method. The advantage of using vignettes is that many workers can respond to exactly the same set of family and family-member characteristics. Each worker reads the same written description. The disadvantage lies in the artificiality of written descriptions, particular for an instrument like the SDM Family Risk Assessment. The FRA is not used to determine the risk level of case descriptions of families but of actual families that workers visit and observe and with whom they interact. Observation and interaction cannot be reproduced via a written vignette. However, vignettes can be used as purely cognitive tools to reveal biases in judgments, and that is the intent of this study.

[44] Why only four? Four vignettes permitted certain the basic risk dimensions of the FRA to be varied. This was adequate for study of racial and ethnic bias. A more comprehensive approach that systematically varied all 25 FRA neglect and abuse items (see Figures 5.3 and 5.4) was precluded by two factors. Vignettes had to be relatively short because workers have limited time for reading them. The number of variations in items had to be kept small because the study was limited to a few hundred Minnesota workers. These factors militated against a larger number of different and exhaustively detailed vignettes.

- Are ratings of this kind different in low-risk cases versus high-risk cases and cases of risk of physical abuse versus risk of neglect?

To answer these questions, all Minnesota workers who had completed a Family Risk Assessment during the period from October 2003 through March 2004 were surveyed. The survey was web-based. Emails were sent to 683 Minnesota workers, whose staff ID was associated with at least one FRA completed during this period. Workers responded by clicking on a hyperlink contained in the email. This link directed workers to a set of Internet-survey pages containing a neglect vignette followed by the FRA risk questions and an abuse vignette followed by the same FRA questions. The links contained codes that had been randomly assigned. The codes controlled the particular combination of family characteristics that workers read about. Each worker responded to two vignettes—one from each group in Table 5.5

**Table 5.5. Stimulus Conditions of Vignettes and Number (n) of Workers Responding to Each Condition**

| ***Group 1*** | | | |
|---|---|---|---|
| *Neglect lower risk* | *n* | *Neglect higher risk* | *n* |
| 1. Neglect-lower-risk-Caucasian | 47 | 6. Neglect-higher-risk-Caucasian | 48 |
| 2. Neglect-lower-risk-African-American | 53 | 7. Neglect-higher-risk-African-Amer. | 46 |
| 3. Neglect-lower-risk-American Indian | 48 | 8. Neglect-higher-risk-American Indian | 49 |
| 4. Neglect-lower-risk-Southeast Asian | 44 | 9. Neglect-higher-risk-Southeast Asian | 47 |
| 5. Neglect-lower-risk-Hispanic | 47 | 10. Neglect-higher-risk-Hispanic | 53 |
| ***Group 2*** | | | |
| *Abuse lower risk* | *n* | *Abuse higher risk* | *n* |
| 1. Abuse-lower-risk-Caucasian | 42 | 6. Abuse-higher-risk-Caucasian | 43 |
| 2. Abuse-lower-risk-African-American | 46 | 7. Abuse-higher-risk-African-American | 40 |
| 3. Abuse-lower-risk-American Indian | 45 | 8. Abuse-higher-risk-American Indian | 41 |
| 4. Abuse-lower-risk-Southeast Asian | 37 | 9. Abuse-higher-risk-Southeast Asian | 44 |
| 5. Abuse-lower-risk-Hispanic | 50 | 10. Abuse-higher-risk-Hispanic | 48 |

Responses were received from 459 workers in time for analysis. The method resulted in a random distribution of these 459 workers across group one and a random distribution of the same 459 workers across group two. For example, a particular worker might have responded to the neglect-lower-risk-Caucasian cell (1) for group one but in the abuse-higher-risk-American Indian cell (8) for group two. This procedure insured that random groups of workers responded in roughly the same numbers to each condition. The numbers of workers responding are also shown in Table 5.5.

The simplest analysis was to conduct an analysis of variance (ANOVA) of the FRA subscales for each vignette version. The subscales yield neglect scores and abuse scores. The means for each of the eight ANOVA analyses are shown in Table 5.6.

**Table 5.6. Mean Neglect and Abuse Subscale Scores by Subpopulation Identity for Each Study Condition**

| *Group 1. Lower-Risk Neglect Vignette* | *Mean neglect risk* | *Mean abuse risk* |
|---|---|---|
| Caucasian | 4.4 | 1.2 |
| African American | 4.2 | 0.9 |
| American Indian | 4.4 | 1.1 |
| Southeast Asian | 4.7 | 1.2 |
| Hispanic | 4.4 | 1.2 |
| ANOVA | F=0.85, p=0.49 | F=1.53, p=0.20 |
| *Group 1. Higher-Risk Neglect Vignette* | *Mean neglect risk* | *Mean abuse risk* |
| Caucasian | 8.5 | 4.4 |
| African American | 8.7 | 4.8 |
| American Indian | 8.2 | 4.6 |
| Southeast Asian | 9.1 | 4.9 |
| Hispanic | 8.5 | 4.9 |
| ANOVA | F=0.75, p=0.56 | F=1.79, p=0.13 |
| *Group 2. Lower-Risk Abuse Vignette* | *Mean neglect risk* | *Mean abuse risk* |
| Caucasian | 4.6 | 8.4 |
| African American | 4.0 | 7.9 |
| American Indian | 4.6 | 8.8 |
| Southeast Asian | 4.0 | 7.9 |
| Hispanic | 4.5 | 8.6 |
| ANOVA | F=0.96, p=0.43 | F=2.15, p=0.08 |
| *Group 2. Higher-Risk Abuse Vignette* | *Mean neglect risk* | *Mean abuse risk* |
| Caucasian | 7.7 | 9.9 |
| African American | 8.0 | 9.4 |
| American Indian | 8.3 | 10.1 |
| Southeast Asian | 7.9 | 9.6 |
| Hispanic | 7.9 | 9.5 |
| ANOVA | F=0.42, p=0.79 | F=0.85, p=0.49 |

While the mean values in each condition vary to some degree, the differences are not great enough to be distinguished from chance variation. The F values in each case are relative small and the probabilities are generally large indicating no significant differences in the assignment of risk by racial/ethnic identity of the family described in the vignette. Only in the case of the mean abuse score for lower-risk abuse vignette was the F value larger (2.15). The probability may be considered a trend, but because eight different ANOVA analyses were conducted, such a variation would itself be expected by chance alone. Workers did not seem to be seriously influenced in their completion of FRA items simply by the verbal designation of race or ethnicity.

On the other hand, descriptions of conditions in families influence workers' assessments of the risk of abuse and neglect. This was determined by average across racial and ethnic variation for each condition. The following eight means correspond to

the eight sections of Table 5.6. The mean neglect-risk subscale score for the neglect vignettes, disregarding race/ethnicity, was 4.4 for the lower-risk condition compared to 8.6 for the higher-risk condition. For the same vignettes, the mean abuse-risk subscale score was 1.1 for the lower-risk condition compared to 4.7 for the higher-risk condition. Turning to the abuse vignettes, the mean neglect-risk subscale score was 4.4 for the lower-risk condition compared to 7.9 for the higher-risk condition and the mean abuse-risk subscale score was 8.3 for the lower-risk condition and 9.7 for the higher risk condition. In each case the comparison of means (t-test) was statistically significant ($p < .0001$).

The four versions of the vignettes were designed to produce differences in ratings of risk scales. This was indeed the result, as was also shown in Chapter 3. Each of the four vignettes was varied by race to determine whether difference in risk ratings would occur by race. They did not. These relationships may be slightly more apparent in the following bubble chart (Figure 5.6).

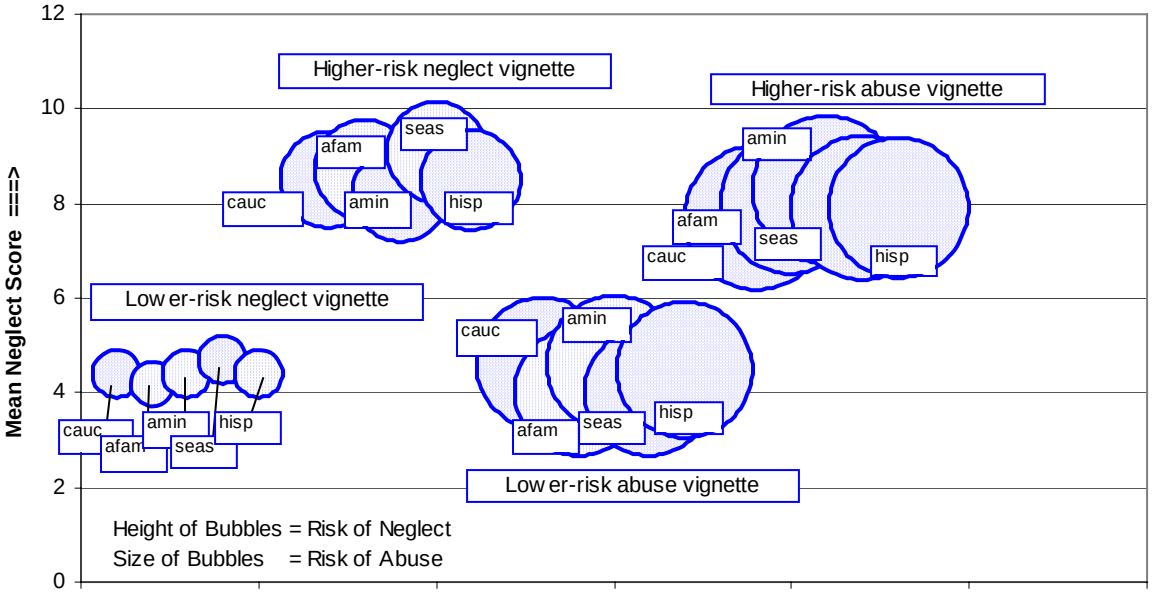


**Figure 5.6. Mean Neglect and Abuse Subscale Scores by Subpopulation Identity for Each Study Condition (Type of Vignette)**

The size of the circles (bubbles) in Figure 5.6 represents the mean risk of abuse and their height represents the risk of neglect. The slightly different sizes and heights represent the differences in means listed in Table 5.6. The groups of vignettes are highly separated and vary in size but within the groups little difference can be seen. This graphic representation shows that the racial/ethnic designation had little effect on the overall scores assigned by workers.

The final risk categorization for each vignette may be instructive as well. These are shown graphically in Figure 5.7. First the slight differences between subpopulation

risk within each vignette type were not large enough to be significantly different. It is evident that for any vignette description the rating tended to fall mainly across two risk categories. Starting from the bottom of the figure, the family described in the lower-risk neglect vignette tended to be rated as either low or moderate risk. Moving up, the description within the higher-risk neglect vignette produced risk rating in the moderate to high range. The lower-risk abuse vignette produced ratings in the high to intensive range and the higher-risk abuse vignette also had this effect but with more intensive ratings.

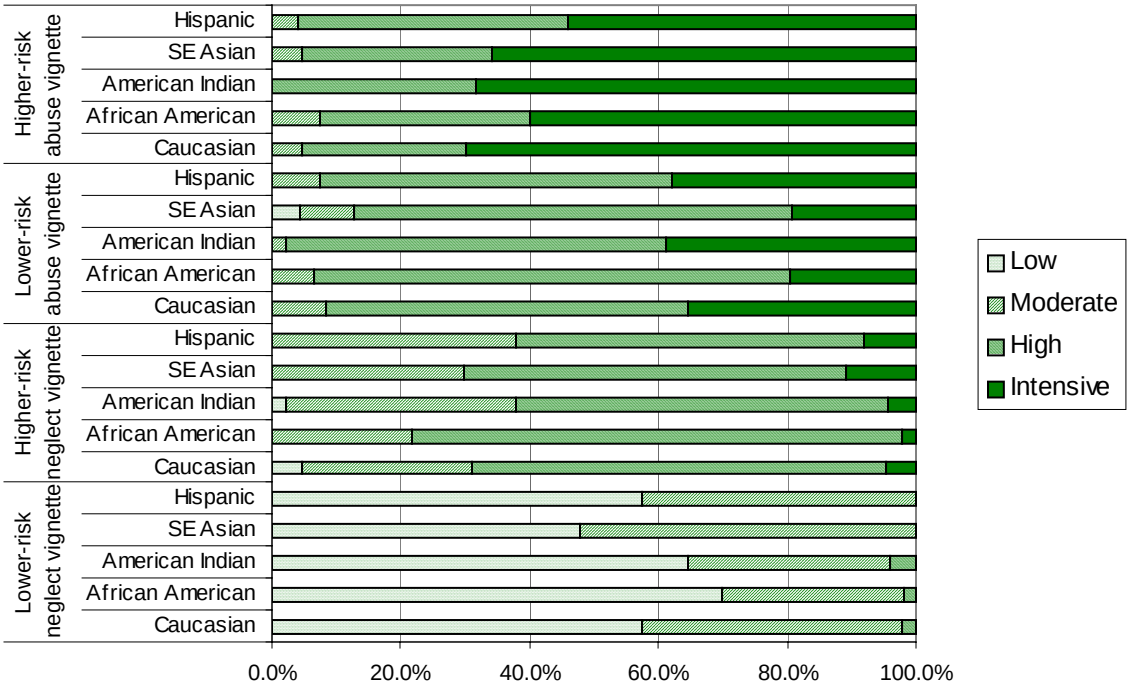


**Figure 5.7. Final Family Risk Categorization by Subpopulation for Each Type of Vignette**

Little difference is apparent by racial/ethnic subpopulations. However, the variability within the risk categories requires some explanation. For example, in the bottom bar of the graph, in the Caucasian condition, 57 percent received low-risk ratings while 40 percent received moderate risk ratings. It is possible that the restricted information provided in the vignettes produced this variation. However, it is also possible that such difference reflect the kind of variation that occurs when workers assess real families. As explained at the end of Chapter 3, because each category—low, moderate, high, intensive—represents a specific cut point in abuse and neglect subscale scores, a family can shift from low to moderate or moderate to high on the basis of an answer to a single question. As we noted in Chapters 2 and 3, the FRA is not a precise assessment tool. Rather it is a rough indicator—a screening tool—that can point the CPS worker in a general direction but should not be used as a sole basis for critical decisions about families. Assuming this general approach, however, the present methods revealed no racial or ethnic biases in the application of the instrument.

# 6. Services and the FRA

Another topic of interest to the state was the effects of the FRA on provision of services. We noted in Chapter 1 that the Structured Decision Making system and in particular the Family Risk Assessment has been promoted as a tool for identifying families in need of services. Introduction of a systematic and accurate assessment of risk, like the FRA, permits the CPS agency to identify families that are likely to be most in needs of and most responsive to services. By directing resources to the high and intensive risk families (and when needed, to moderate risk families) the agency services will be more cost-effective. Implicitly, this assumes that low-risk families benefit less from services. This approach fits nicely with the traditional residual approach to child protection in which limited service resources are directed primarily to families in crisis. Under this assumption the FRA can form the basis of a more efficient and presumably cost-effective CPS system.

At the same time in Minnesota and in several other states a fundamentally different approach is being tested. The Minnesota Alternative Response (AR) program was also described in Chapter 1. AR is directed toward (the majority of) CPS families in which child safety threats are less severe. AR family assessments are non-adversarial family-friendly visits that aim at engagement and fuller family participation in the assessment process. There is no determination of child abuse and neglect. While traditional investigations focus narrowly on potential abuse and neglect, AR is concerned with a broader array of family strengths and needs, and this emphasis begins at the time of the first visit with the family. Services are premised not on substantiation of child abuse and neglect but only the welfare of the family. Further contacts with families are voluntary. One of the consequences of this approach is that a broader array of services tend to reach families—*regardless of the families risk level*. The evaluation of AR showed that significantly and substantially more services were offered to low-risk families than under the traditional CPS system.[45]

The existence of AR and data collected during the AR evaluation offered an opportunity to test the service hypothesis associated with the promotion of the FRA. Under AR, family needs rather than risk level drive service delivery. Services are seen as preventive rather than simply remedial. Does this shift lead to positive outcomes?

[45] *Minnesota Alternative Response Evaluation: Final Report*, Institute of Applied Research, November 2004, www.iarstl.org. See in particular, Chapters 5 and 10.

## Services to Low-Risk Families

The case-specific survey in the present study collected worker responses about specific families with which they had worked during the final quarter of 2003. Each of these families had an FRA completed on them by the worker. Workers were asked a series of questions about each family, including the level and types of services offered.

Data on assessment workgroup track from SSIS were used to categorize the families into traditional versus AR. There were 225 families in the traditional track, involving a CPS investigation and the traditional events ensuing after an investigation. There were 187 families in the AR track. These families received a family assessment and the different service emphasis under AR. The proportions of new services offered to families assessed at different risk levels within the two tracks are shown in Figure 6.1.

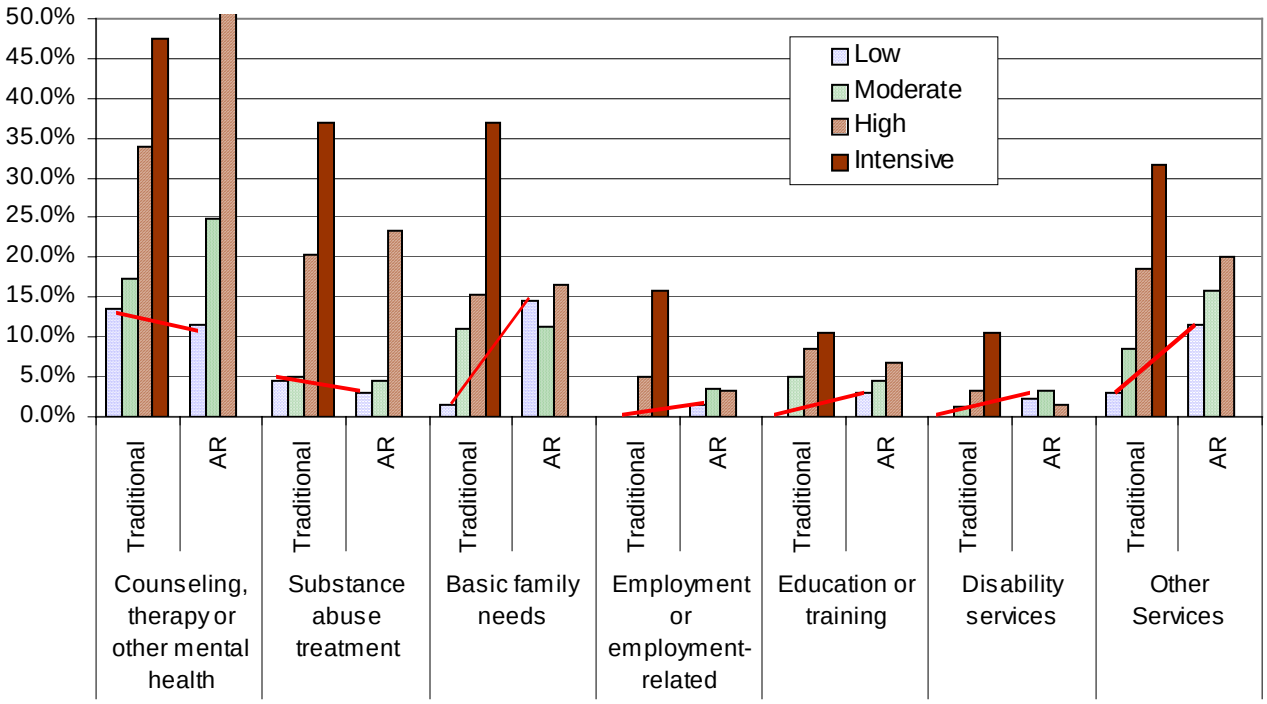


**Figure 6.1. New Services Offered to AR versus Traditional Families of Different FRA Risk Levels (Case-specific Survey of Workers, 225 Families screened Traditional, 187 screened AR)**

The findings correspond to those of the AR evaluation. That report, which was focused on AR families during 2001 and 2002, and the present data show increased services of various kinds to low-risk families. (Connecting lines are included in the chart to direct the reader's eye to comparisons.) The biggest differences occurred in the basic services (financially-related, family and household needs) and in other services, many of which could be defined as basic.

The table also shows that low-risk families received an array of services on the traditional side, indicating that Minnesota workers often respond to low-risk families in spite of their FRA category. (The flexibility of workers and local offices in this regard was discussed in Chapter 4.)

It would have been useful to track these families in order to determine whether differences in outcomes occurred. This was not feasible because less than 10 months of follow-up data were available for many.[46] A better approach was to return to the data from the original AR evaluation. The evaluation design involved random assignment of families, all of which had been screened as *appropriate for AR*, to an experimental or a control group. This is diagramed on the left side of Figure 6.2. Under this evaluation, 2,860 families were assigned to an experimental group and 1,305 to a control group. As noted previously, experimental families were highly similar, as a group, to control families.[47] This is important because it insured that the provision of services to families did not hinge so much on initial differences in reports and family characteristics but on the approach taken to families—traditional or AR.

Each experimental family received an AR family assessment, while each control family received a traditional assessment (investigation).[48] Each family also received a Family Risk Assessment as part of the assessment process. In the diagram in Figure 6.1 we have collapsed families into two risk categories—low-moderate and high-intensive.

The diagram also shows that a choice of post-assessment services or no post-assessment services was made for each family. This is a crude proxy measure for the categories shown in Figure 6.1, but it was the best available for this large data set because the SSIS does not contain records of specific services for families.

The service difference under AR and traditional become immediately apparent when the numbers in Figure 6.2 are compared. The numbers refer to families. Only 59 of 1,006 low and moderate risk control families had service cases opened compared to 815 of 2,433 similar experimental families. Part of the difference is due to the fundamental difference in the assessment process. The investigations of many of the control families ended with no findings of maltreatment, and traditionally unsubstantiated investigations do not result in service cases. AR family assessments are not concluded with findings of maltreatment but with invitations for services, if needed. *The control side of the experiment resembles more closely the system advocated by the designers of the FRA. The experimental side represents a fundamental shift in direction from protection only to prevention. The question is: Were there long-term consequences of this difference in approach to services?*

[46] The sample was selected from reports with FRAs during October, November and December 2003. Data collection ended as of 9/30/2004.
[47] *Minnesota Alternative Response Evaluation*, Chapter 8.
[48] Families were selected and assigned during the period from February 2001 through December 2002. Therefore, most were included among the 15,100 families in this study, which were selected during the period from January 2001 through September 2002.

To examine this we tracked child maltreatment report recurrence. This was done in a slightly different way than described earlier in this report. In accordance with the experimental design, new reports were counted only after the final contact of the agency with the family in the initial case. Thus, reports that might have been received during an open case on a family were ignored.[49] Second, families were not tracked for a standard period of two years, as they were in the present study. Tracking periods for families varied from as little as six months to over two years.[50] Differences in tracking periods require an alternative analytic approach. Before explaining that approach we will examine a simple cross-tabulation.

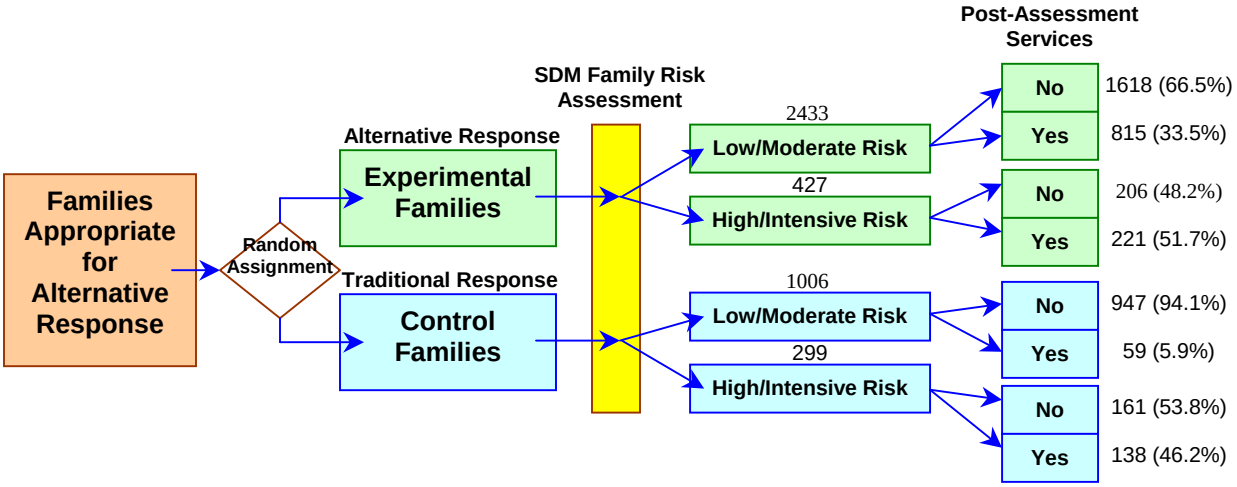


**Figure 6.2. Flow of Families in the AR Project Evaluation, Risk Assessments and Service Case Choices**

In Table 6.1, recurrence is shown for low/moderate-risk and for high/intensive-risk experimental and control families. The percentage cells in this table correspond to the 8 boxes on right side of Figure 6.2, and as can be seen, the numbers of families in the table correspond to the numbers in the figure. Recurrence was reduced among AR families that received *no services*, an effect of the difference in

> Although outside the purview of this analysis, some reviewers of this report expressed alarm that so many high and intensive risk families received no post-assessment services (48.2 percent of experimental families and 53.8 percent of control families). Control families were all investigated. Typically, post-assessment services are not provided unless the investigation is substantiated. Of the 161 high/intensive-risk control families with no services, only 58 (36.0 percent) had substantiated investigations. Among these 58, workers determined in 45 cases that no *child protection services* were needed. Remember that services addressing risk are family-welfare services. Among experimental families (quite similar to control families), the key was the willingness of the family to participate in services. AR assessment workers did not conduct investigations but at the end of the family assessment, the worker and family jointly made the decision whether to continue with services. In either the control or experimental group, high-risk families could either be disregarded or opt out of services. Workers in the traditional system have little power to force families to accept services except in the minority of cases that enter the court system. Workers under AR must persuade families to participate.

[49] Analysis, however, showed no significant difference in the proportion of new reports for experimental and control families during the initial assessment or case management workgroups. Thus, the exclusion of these reports was of no consequence in the analysis.

[50] Data collection for the current phase of the AR evaluation ended as of March 31,2004.

approach to families described above: non-adversarial, family-friendly, participatory, and so on.[51] Recurrence was also reduced among families in which a *service case was opened*.

**Table 6.1. Recurrence Among Experimental and Control Families Risk Level and Opening of Service Case**

| ***Low and Moderate-Risk Families*** | | |
|---|---|---|
| | *Control* | *Experimental* |
| Recurrence, No Services | 29.9% | 26.7% |
| Families | 947 | 1618 |
| Recurrence, Service Case Opened | 32.2% | 25.5% |
| Families | 59 | 815 |
| | | |
| ***High and Intensive-Risk Families*** | | |
| | *Control* | *Experimental* |
| Recurrence, No Services | 28.6% | 34.5% |
| Families | 161 | 206 |
| Recurrence, Service Case Opened | 34.8% | 29.9% |
| Families | 138 | 221 |

This table suggests that services are worthwhile to lower-risk families. Higher-risk control families that were offered services had *higher* rates of recurrence (34.8 vs. 28.6 percent). The finding is not unusual because, under the traditional CPS system, families that are offered services are more likely to be crisis. Such families, as a group, tend to return to the system more often *regardless of the service intervention.* On the AR side, this pattern is reversed, presumably because among higher-risk families provided AR broader criteria are applied in determining service provision. This difference is described as an *interaction effect* in analytic parlance, and we take that into account in the next analysis.

The figures in Table 6.1 can be regarded as illustrative but not as definitive for the reasons stated above—differing follow-up periods for families. The proper analysis in this case is *survival analysis*. Survival analysis refers to a family of statistical techniques that can be used when follow-up varies. In this analysis, the outcome question is: Do experimental families "survive" for longer periods before a new report of child maltreatment than control families? If they do, it shows that report recurrence is less likely for experimental families (something indicated in Table 6.1). Recurrence reduction is a positive outcome.

The analysis first confirmed the findings of the AR evaluation—that experimental families that had experienced AR had significantly less recurrence that control families ($p = .05$).[52] This is shown in the following graph (Figure 6.3).

[51] AR workers sometimes provided services to families under this condition without opening a case-management workgroup, that is, a formal service case. In addition, AR workers immediately approach families as service workers, rather than investigators. We are not considering in this analysis direct services and other assistance offered by CPS workers.

[52] Cox Proportional Hazards Regression was used. In the first analysis, only the experimental/control variable was utilized.

The higher the line in the figure the longer the number of days (survival until a new maltreatment report occurred). As can be seen, experimental cases survived for longer without a new report. This means they had fewer new reports over comparable time periods.

The question of services was addressed by a more complex analysis. The survival analysis in this case involved four variables: a) experimental-control membership, b) risk level (low/mod. vs. high/int.), c) services (service case vs. none), and d) the interaction between services and experimental-control group membership. The results are shown in the second graph (Figure 6.4) and the accompanying table (Table 6.2). The difference in Figure 6.4 shows that that differences in recurrence between the experimental and control groups remained when services and risk were controlled.

The final survival model involved stepped entry of the variables into the regression equation. First, risk was entered and that showed that high/intensive risk families were more likely to have report recurrence. Then, services were entered, showing that, in general, service provision reduces recurrence. Third, experimental-control group

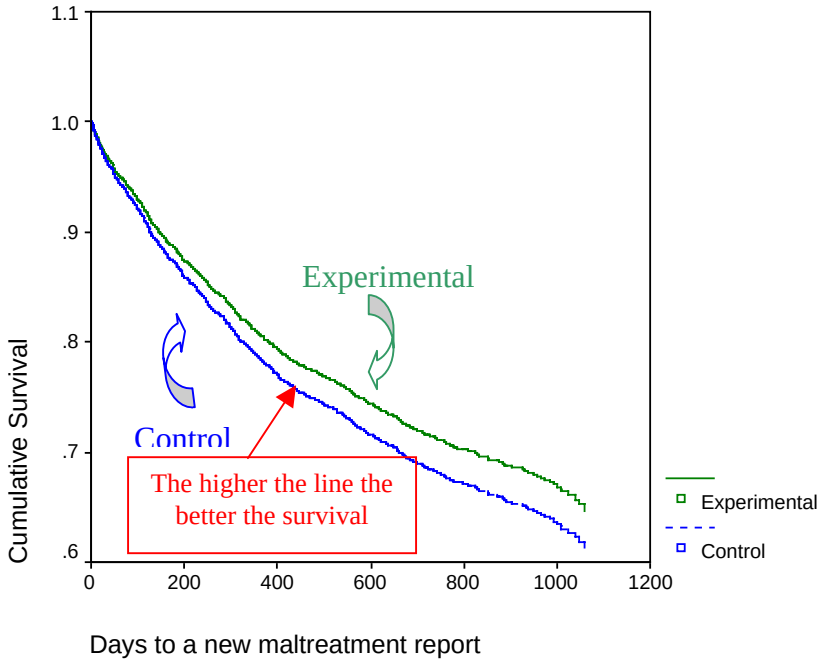


**Figure 6.3. Survival Plot of Days to New Maltreatment Reports for AR Experimental and Control Families**

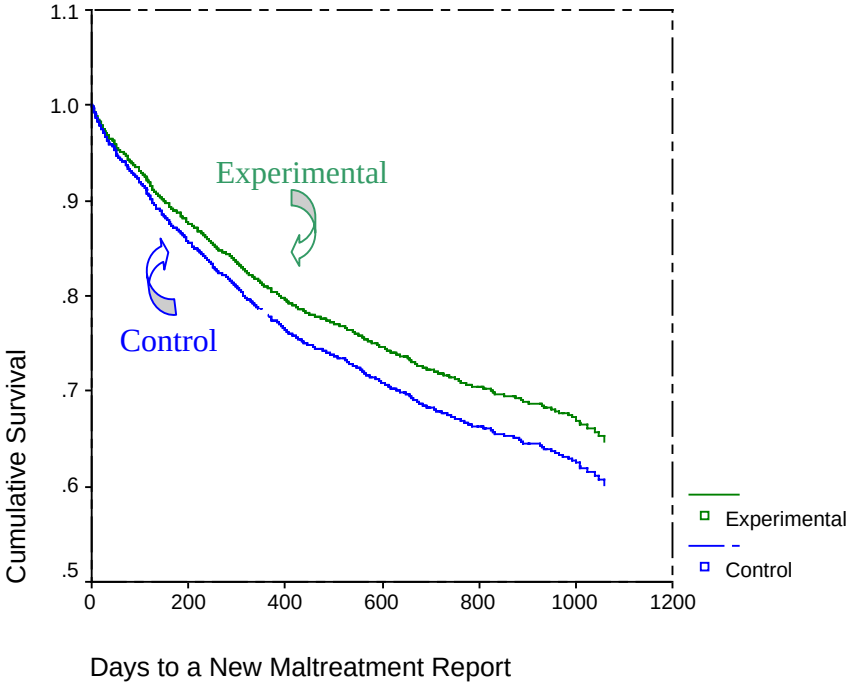


**Figure 6.4. Survival Plot for Experimental and Control Cases Controlling for Risk, Service and Service-Group Interaction**

membership was entered, showing that when risk and services are controlled, AR still reduces recurrence. These findings are evident in the top portion of Table 6.2 (see the level of significance of the variables).

The bottom portion of the table shows what happens to the regression equation when the interaction effect was introduced. The important difference is that the effect of services in general drops from statistical significance (p = .033 in the top portion) to non-significant (p = .225 in the bottom portion) but the interaction effects is significant or at least a trend (p = .092). This analysis shows that the effects of services lie in the difference between the AR and the traditional approach—the kind of difference illustrated in Table 6.1. *It confirms the notion that services across the spectrum of families encountered by CPS is beneficial and that the traditional approach of offering services mainly to higher risk families is less beneficial for families screened as appropriate for AR.*

**Table 6.2. Cox Regression Equation Variables**

| *Variable* | *B* | *S.E.* | *Wald* | *df* | *Sig* | *R* | *Exp(B)* |
|---|---|---|---|---|---|---|---|
| Risk | .16 | .076 | 4.38 | 1 | .036 | .011 | .853 |
| Services | .15 | .068 | 4.56 | 1 | .033 | .012 | .864 |
| Experimental-Control groups | .14 | .064 | 4.74 | 1 | .030 | .012 | 1.151 |
| *Equation after the introduction of the Interaction Effect* | | | | | | | |
| Risk | .13 | .079 | 2.78 | 1 | .095 | .006 | .877 |
| Services | .09 | .076 | 1.47 | 1 | .225 | .000 | .912 |
| Experimental-Control groups | .36 | .142 | 6.29 | 1 | .012 | .015 | 1.427 |
| Interaction Services and groups | .27 | .158 | 2.83 | 1 | .092 | .007 | .767 |

## Changing Levels of Risk

Another way of assessing appropriateness of services is to compare risk levels on the same families at more than one point in time. This was possible for the sample of 15,100 families because the data system (SSIS) provided not only information on new intakes but also scores of the FRA administered at the time of each new intake. In the present analysis there were 3,424 families with recurring reports that had a second report and an FRA. The first and second FRA could be compared for these families.

Families are dynamic entities. They change over time. Family members grow older. Relationships among members change. Various family members disappear. New members appear. Some families disintegrate. Others merge and blend. Some changes occur in response to the changing external environment, including employment, extended-family relationships, living arrangements, and disease. Other changes occur because of changes to members, including mental and emotional conditions, substance abuse, education, religious commitments, and new relationships with individuals outside the family. Families typically seen in CPS are more affected by personal and environmental changes compared to families generally in our society because they are the most vulnerable. CPS families as a group are the poorest of the poor and the least well

educated, the most likely to be socially isolated. It is a profound error, therefore, to assume that characteristics of a CPS family observed at one point in time will remain unchanged into the foreseeable future. Risk characteristics are no exception to this rule.

The initial FRA of the 3,424 families yielded 994 (29.0 percent) low-risk, 1,512 (44.2 percent) moderate-risk, 788 (23.0 percent) high-risk and 130 (3.8 percent) intensive-risk assessments. On the second risk assessment there were 680 (17.8 percent) low-risk, 1,667 (48.7 percent) moderate-risk, 967 (28.2 percent) high-risk and 182 (5.3 percent) intensive-risk assessments. The percentages indicate a decrease in low risk and increases in each of the other risk categories—an unsurprising finding since three items in the risk scale refer to past reports and cases of families. These changes in percent do not give the full picture, however. This is obtained by cross-tabulating the two sets of assessments. This is shown in Figure 6.4.

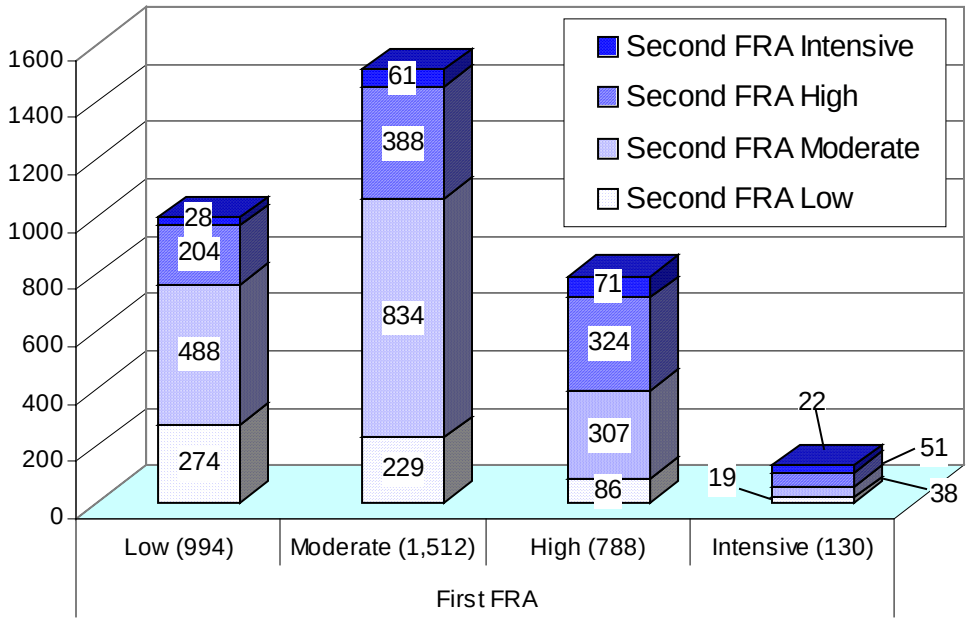


**Figure 6.4. Risk Levels on the First and Second Administration of the FRA for 3,424 families**

It is apparent that substantial shifts occurred in the risk assessments of families. While more families showed an increased risk level on the second administration of the FRA, a substantial minority of families showed reduced risk. Some of this shift has to do with the reliability of the scale. For example, it was apparent in Chapter 3 that families could shift up or down a risk level simply by a change in the score of one risk item. Some may also be due to variations in information obtained on families by different workers and different interpretation of risk items and when they should be checked. The portion of this variation that is due to such errors cannot be determined.

Some part of the change is also the result of changes that occur in families. The average (mean) number of days between the first and second intake for these cases was

366 days. Some of the shifts represent the kinds of changes that may occur in families over a year. Of particular interest for the present analysis are the shifts among families that were rated as low and moderate risk in the initial FRA. Among initial low-risk families, 232 (23.3 percent) were rated as high or intensive risk in the second assessment. And among initial moderate-risk families, 449 (29.7 percent) had shifted into the high or intensive risk categories. This represents more than one-quarter (27.2 percent) of all low- and moderate-risk families.

While the FRA may be a valuable tool in assessing families at one point in time, other factors must come into play as time goes on. The changes suggest that other factors may be important in determining whether CPS should attend to families reported for child maltreatment. The increases in post-assessment services to low- and moderate-risk families that occurred through the AR program were not random events. Rather, they were responses to family needs identified by workers and families in a joint process. The AR evaluation documented that the needs were real and that addressing them led to improved outcomes for families and long-term cost-savings for the agency.

Risk assessment, and in particular the SDM Family Risk Assessment, is a valuable and necessary part of work with families reported for child abuse and neglect. This analysis shows that the FRA might be more valuable if it is used as one tool within a fuller assessment of strengths and needs and joint decision-making by workers and families.

# 7. Summary, Conclusions, and Recommendations

The introduction of Structured Decision Making tools into CPS interventions was an effort to introduce greater accuracy and consistency in the manner in which families and children are assessed and to provide certain uniform guidelines to planning for child safety and family services. As one tool in SDM battery, the FRA was intended to provide accurate and consistent determinations of family risk of future child abuse and neglect and to permit consistent decision-making about further work with families. *The numbering of the following sections matches corresponding chapter numbers.*

## 2. Predictive Validity

The FRA has predictive validity in regard to new reports of child maltreatment and new cases opened for families following such reports. The validity extends to individual items in neglect and abuse subscales and to the subscale scores themselves. Like all tools intended to predict future human behavior, however, the FRA involves error. Our analysis indicated that the scale misclassified approximately one in three families.

This error rate is high for an instrument that is to be used as the *sole or central criterion for decision-making* regarding post-assessment services. However, it is not necessarily high for an instrument that is to be used with other assessment and decision-making tools. High- and intensive-risk scores, in particular, on the FRA are evidence that a family has a number of special needs with the potential to affect child safety in the future. That the FRA requires each worker to examine a consistent set of risk factors for each family is valuable and provides clues to workers concerning the nature of the needs of families. We recommend continuation of the use of these items, although other items also associated with family risk might also be included (see below).

The larger portion of predictive error arose from families with low and moderate risk scores that were reported later. From the standpoint of long-term child safety, this type of error may be the more dangerous. It is always possible that false negatives are not really false, that is, that the risk assessment was accurate and that new reports reflect conditions and circumstances of families and family members *that were not present at the time the FRA was administered.* Alternate explanations are that the errors occur because of the absence of certain risk-related items on the FRA, the failure of workers to accurately complete all FRA items, or to some combination of the two, as we suggest below. The latter issues refer to the reliability and the manner of using the instrument.

*Another issue of concern expressed by some advisors was that certain FRA items were superfluous from the standpoint of risk assessment.* Specifically, there was concern that demographic characteristics of families (number of children, adults and age of caregiver) were not predictors of risk. This was not found to be the case. As a rule, each of the items on the FRA was predictive, alone or in combination with others, of new reports of child maltreatment. There was also concern that such items, even if predictive, provided little direction to workers in working with families—you cannot change a person's age, so why ask about it? In our opinion, this represents a misunderstanding of the nature of risk assessment. Certain risk factors cannot be addressed or changed while others can, but together they point toward the families in greatest need of attention.[53]

## 3. Reliability

The FRA has a marginal internal consistency. To be internally consistent the neglect subscale items should be positively related to the neglect subscale score most of the time. Similarly, abuse subscale items should be related to the abuse subscale score. Based on scoring of items by Minnesota workers, the standard measure of internal consistency for the FRA was at the lower end of the acceptable range for both the abuse and neglect subscales. This means that individual risk items usually supported one another but that a level of inconsistency was found. Some improvement in consistency was found when certain items in the neglect subscale were set aside but the change was, in our opinion, not great enough to warrant modification of the scale.

We also found that the internal consistency of the abuse subscale was lower for certain workers who completed large numbers of risk assessments (50 or more) during the 18-month sampling period. This might be the result of differences in the types of families these workers encountered or to different patterns of completing the FRA. Regarding the latter, it could be explained by failure to complete certain items.

Another measure of reliability was consistency between workers in completing the FRA. Analysis of the vignette survey showed that workers tended to use the subscales consistently, when judging families described in a written case description or vignette. And consistent differences in risk ratings resulted when they considered and compared two different vignettes. The vignette methodology could not take into account differences that might arise from encounters between workers and real-life families.

While high consistency was obtained for the final scores of the neglect and the abuse subscales, there was evidence that consistency dropped in producing the final categorical rankings (low through intensive) on the FRA. This is a natural consequence of going from a scale score to a categorical ranking (numeric scores from 0 to 20 and 0 to 16 to a final rating of low, moderate, high, or intensive risk). The effect of this process is

[53] Many examples could be provided of this from other areas. Auto insurance rates are higher for teens than for older drivers. Other factors may reduce their risk of accidents, such as driver's education classes, good academic records, etc., but rates still remain high because age is a strong predictor by itself. In promoting auto safety, this risk relationship points up the need for special attention to teen drivers.

that relatively minor variation in subscale scores can produce substantive variation in final risk scores. That reliability is reduced is obvious: two workers can score all items the same way save one, and that one item can produce a categorical difference in risk scores. Small difference in rater judgment can produce large differences in outcomes. More sophisticated scoring methods might avoid this outcome.

## 4. Other Practice Issues

Most workers recognized that the FRA introduced positive features, such as consistency, into the family assessment process. However variations were found in the way in which the FRA was applied. In addition, workers expressed a number of relevant criticisms of the instrument.

Differences were found in the *extent to which the FRA affected decision-making about services to families*. Some respondents said it was a minor factor or unimportant in responding to families. For others it was a major factor. This seemed to be a function of local offices rather than differences among workers within offices. The larger the county and CPS office, the more importance the FRA assumed in decision-making.

*Responses to families with lower risk scores* also varied in the same way. Low-risk families were less often provided with post-assessment services in the larger urban offices compared to other counties.

*The point in time that the FRA was completed during the assessment* also varied. It was evident that FRA scores in some cases reflect the state of the family during or shortly after the first visit by the worker while in other cases the score represents the family at the end of the assessment process. This has implications in the context of the Alternative Response approach because some assessment workers provide services to families during the assessment process itself.

Workers were also concerned that *certain characteristics of the FRA push families to higher risk levels* than should be the case. These included the following: events from long ago may be scored the same as events that occurred recently (e.g., very old past cases versus cases that just closed); some risk factors may be present but mitigating factors reduce their significance (e.g., coping skill or extended family support); some items may be more risky for one subculture than another (e.g., the number of children in the family). Some other *items may need to be modified* to be accurate (e.g., the age of children is a factor in risk). On the other hand, *some risk-related items may be missing, leading families to be rated as lower risk* than should be the case (e.g., mental health).

The analysis of predictive validity found false positives (erroneous higher-risk scores) to be somewhat less of a problem than false negatives (erroneous lower-risk scores), suggesting that the latter comment may be a more important issue. However, one factor mentioned by workers impinges directly on reliability. Workers indicated that

there was *no way to indicate a lack of knowledge*. Missing information is scored as no risk on the FRA. In addition, items are sometimes left blank when *workers suspect but cannot prove that a risk factor is present*.

**Differences in FRA Scoring under the Alternative Response Approach.** Many of the families in this study were served through the new AR approach. The hallmarks of this approach are family assessments that are non-adversarial and family friendly and that emphasize family participation and voluntary continuation with the CPS agency. FRA scores changed somewhat under these conditions, especially on items that were more subjective and dependent on worker-family interaction and items that required probing by workers. The best explanation is that the *increased family engagement that we know occurred under AR created differences in families*. FRA items that concerned cooperation, motivation, and situational attitudes were scored more positively, indicating that the worker family-relationship may reduce the risk of future child maltreatment. Another possible explanation for the differences may have been early intervention in families by AR workers. *Assessments by AR workers after they had intervened in families* are another possible explanation of the differences observed on items.

## 5. Minority Subpopulations

The study of the five racial and ethnic subpopulations indicated some differences in the application of individual FRA items. Many of the differences on individual items, however, were evened out in the final categorization of families into the fourfold classification of low, moderate, high and intensive risk. The exceptions to this rule were Southeast Asian families that received overall lower risk scores and American Indian families that received overall higher risk scores. These findings held when the analysis was limited to comparison of minority and majority populations in counties with substantial minority representation.

**Predictive Validity.** The FRA showed levels of predictive validity for the subpopulations similar to the entire study sample, with the same exceptions. It was more accurate with Southeast Asian families and less accurate with American Indian families.

The lack of predictability of the FRA for American Indian families was examined in greater detail. The primary problem was one of false negatives. These are families that were rated as low risk using the FRA but later re-reported for child maltreatment. Such families received new reports at rates only slightly lower than families rated as moderate- or high-risk, and in some counties little difference in report recurrence could be found among risk levels. The problem appeared to occur among neglect subscale items having to do with parenting skills, harmful relationships of parents, substance abuse, financial problems, and motivation and cooperation.

To examine this further, worker narratives of 41 low-risk American Indian families that were re-reported were examined. Content analysis revealed that a number of problems either were present in these families at the time of the first FRA or appeared

in the family at the time of a later report. These included adult drug abuse, adult alcohol abuse, adult mental illness, domestic violence, and child behavior problems. These overlap significantly with the problem items indicated above that were only rarely checked for low-risk families on the original FRA. In some cases, they may have been new problems not present at the time of the original assessment. No FRA items exist to permit workers to register the presence of mental illness in families. For the others, they may have been miscoded on the FRA for these reasons: 1) Workers did not know about them when they completed the FRA, and when they learned of them, they did not correct the FRA. 2) They suspected the problem but were not sure of it and left it blank on the FRA. 3) They simply never learned of the problem at the time of the initial assessment of the family. Perusal of narratives indicated that this same problem was present, perhaps to a lesser extent, across the spectrum of low-risk families with recurrence.

**Reliability and Racial/Ethnic Bias.** An experimental design was employed to permit workers to respond to the same family (in a descriptive vignette) but with different minority subpopulation identifications. No evidence of racial/ethnic bias could be detected in this analysis.

## 6. Services and the FRA

Workers in surveys reported the service response to a sample of families selected from the final quarter of 2003. Significantly more services, particularly services addressing *basic financial and household needs, were delivered to low-risk families under the Alternative Response approach than under the traditional approach.* This emphasis on preventive services to low-risk families runs counter in some ways to the emphasis implicit in the FRA of directing services primarily to higher-risk families.

Data were utilized for this analysis from the Alternative Response evaluation. FRA risk levels, services and report-recurrence were considered. Experimental-control comparisons revealed that recurrence was lower for AR families generally under these conditions and, specifically, *that services to low-risk families made a difference in outcomes*.

This suggests that, while *identification of high-risk and intensive-risk families through the FRA can be used as a means to determine families in need*, it should not be the exclusive method of determining need. *This analysis and the AR evaluation have shown that services to low risk families improved family outcomes and were, in the long-term, cost effective for the CPS agency*.

## Recommendations

1. **Change the order of completion of the SDM instruments.** Currently workers are to complete the child safety assessment and the family risk assessment as first steps. The present analysis suggests that other considerations may be equally important in determining whether work with a family is needed. The new order might be SDM safety assessment, assessment of family strengths and needs (FSN), and family risk assessment. Low-risk families with many deficiencies and few strengths on the FSN may be considered for further services. Families with no indications of threats to child safety (or no child maltreatment in an investigation) but with high indications of needs *or* high risk should be invited for services on a voluntary basis. This approach alone might reveal some of the problems that are leading *apparently* low-risk families back to the CPS system and permit them to be addressed.

2. **Improve the FRA Scoring method.** The FRA is currently scored like the paper and pencil version of the instrument, even though it is contained in a sophisticated computer system. A more sophisticated scoring method might improve the predictability and reliability of the instrument. Furthermore, a new scoring method might provide fuller information to practitioners. We conducted analyses to assign weights to individual items that could be used to generate more accurate neglect and abuse scale scores. It is possible that workers might make different decisions based on combinations of scores. For example, should a family high on risk of neglect and high on risk of abuse be treated differently than a family high on risk of neglect and low on risk of abuse? Both would be high-risk in the current scoring, but the very presence of different scale scores indicates that they are different types of families.

3. **Empirically Test Changes to the FRA.** To improve the usability of the FRA, the following changes should be considered. We suggest they be added systematically to the FRA protocol in a few test county offices. The reliability and validity of the assessments could easily be tested using the methods of this study after a period of a year. The following changes are suggested:

    a. Add a "do not know" category to each FRA item. The final scores may be generated as they are now by counting don't know the same as no risk (i.e., as zero). However, an *uncertainty score* could also be generated equal to the total number of items checked as "do not know." These would provide a supervisor a context for determining the weight to give the FRA score.

    b. Permit workers to check an item when they *strongly suspect* but cannot prove the presence of the risk factor. This would especially apply to substance abuse by any caregiver, domestic violence or domineering relationships between adults, fragile financial situations, and perhaps others.

    c. Create an alternative risk factor list to accompany the FRA that would be the basis of increasing the risk level of families. This might include the following: severe

emotional problems of an adult, adult psychosis, adult or child depression, chronic and serious physical illness of an adult or child, adult criminal behavior, social isolation of the family (including both neighbors and relatives), lack of viable childcare or respite care in situations of high stress, child behavior problems other than delinquency (acting out, truancy, sexual behaviors, extreme hyperactivity, and others). This list could easily be extended by polling workers about other important risk factors in their experience. These items would assume a formal status as additional risk items. Through further analysis a new and expanded risk assessment might be possible. However, these changes should be empirically tested.

d. Create a list of mitigating or strength-based factors that could be cited as justification to ignore certain risk items (i.e., check them as "no risk") or in a modified version of the FRA *may be scored as negative risk factors*. These would include: extended family support in severe financial difficulties, substance abuse problems in consistent and long-term treatment, child developmental disabilities under competent caregiver control and/or service programs, child behavior problems or delinquency under competent caregiver control and/or service programs, prior assigned reports of an absent perpetrator, remote or irrelevant prior CPS service history, level of maturity of a young caregiver, level of immaturity of an older caregiver, older teen children discounted from counts of children, secondary caregiver in temporary relationship with primary caregiver. This list could also be expanded as needed. Final decisions on these items should also be empirically based.

We would suggest that such changes be tested for a sample of 500 or more families (larger if stratified by minority subpopulation). The number of workers should be as large as possible. *The validity analysis would involve simple counts of new reports and new open cases during a one-year follow-up period. Reliability would be determined by a procedure that permits internal consistency measures (such as that in Chapter 3) to be utilized.* Workers would also be surveyed concerning their impressions of the changes.

4. **Change in Practice in Larger Counties.** If, as surveys in the study indicate, the FRA is used in some large offices as the primary means of excluding low-risk families from response by the agency, consideration should be given to modifying this practice. Because the FRA, as currently employed, may misclassify some families as low-risk, additional criteria should be employed, as suggested above, to determine whether post-assessment services are appropriate.

## Appendix A. Vignettes for Inter-rater Reliability Study

The following four vignettes were used in the survey of workers. Phrases in italics were varied to change the risk conditions. The first two vignettes are labeled "neglect" because we constructed them to produce varying scores on the neglect subscale. Similarly, the other vignettes are labeled "abuse" because the intention was to create varying scores on the abuse subscale. They largely accomplished these goals. The data tables were presented and race was designated in the final column. In addition, the names of individuals in the vignettes were modified to reflect common names in racial/ethnic groups. These are shown in the table at the end of this appendix. The Caucasian version of the name is included in the vignettes printed below.

Together the four vignettes with five different racial classifications produced 20 different versions of the vignettes or experimental conditions. Each worker was presented with two of these: a randomly assigned version of vignettes 1 or 2 with a random racial/ethnic group and a randomly assigned version of vignettes 3 or 4 with a random racial/ethnic group. The final experiment, then, included about the same number of workers in each of the 20 experimental conditions.

## 1. Low-Risk Neglect Vignette

A parent educator from a local program called to say she was concerned about Bobby (child), an infant. Bobby (child) appeared to be unwashed when she had visited him and his mother. She said she was also concerned that the mother was not feeding the child properly. She said she thought the mother and father were intellectually limited and that the child was in danger. The reporter said that Bobby (child) was twelve months old and was an only child. His mother and father, Melanie (mother) and Charles Larsen (father), were married and living together. A record check provided the following information on the parents:

| Name | Previous CPS Service Cases | Previous CPS Reports | Criminal Arrests | Convictions | Mental Health Services | Developmental Disabilities | Age | Sex | Race |
|---|---|---|---|---|---|---|---|---|---|
| Melanie Larsen | None | None | None | None | None | Childhood Head Injury/Seizures | 29 | F | White |
| Charles Larsen | CPS child-victim, with foster care | None | None | None | None | Mild Mental Retardation | 30 | M | White |
| Bobby Larsen | No record | No record | - | - | No record | No record | No record | No record | No record |

The assessment worker visited the home. The mother and father and infant were home at the time in their small apartment. She noted that the apartment was messy and looked like it had not been cleaned for some time. The rugs and floor were dirty and cluttered. Bobby (child) was 13 months old. He had a dirty face, arms and legs and at the time needed to have his diapers changed. When asked about feeding, Melanie (mother) said that Bobby (child) still took a bottle but she was also feeding him baby food. Bobby (child) was crawling on the floor and trying to pull up on the sofa.

The worker asked if she could look around the apartment. *The worker found baby food and some formula in the refrigerator.* She noticed that the bathroom stool was stopped up and asked about it. Melanie (mother) said that they had told the apartment owner about it several times in the past three weeks but he would not fix it. So, they were using the bathroom at a nearby service station. She said they were behind on the rent. *Charles (father) said they would start looking for another apartment as soon as possible and Melanie (mother) agreed.*

Melanie (mother) said that she received SSI and food stamps. They said that their money sometimes ran out toward the end of the month and they had to make do until Melanie's (mother) next SSI check came. Melanie (mother) said her parents helped them with food when things really got bad but that they were older and retired and living on Social Security and did not have much to spare. Charles's (father) parents provided no support. When asked about the dirty diaper, Melanie (mother) said that diapers were so expensive that she sometimes waited to change them.

Melanie (mother) said she took pills for her seizure disorder. Melanie (mother) and Charles (father) said they did not drink. The worker engaged the parents in a discussion about hygiene. *Melanie (mother) and Charles (father) said they wanted help and would do whatever the worker wanted them to do. As the worker was leaving, Charles (father) put on his coat and said he was going out to look for work right now.*

## 2. High-Risk Neglect Vignette

A parent educator from a local program called to say she was concerned about Bobby (child), an infant. Bobby (child) appeared to be unwashed when she had visited him and his mother. She said she was also concerned that the mother was not feeding the child properly. She said she thought the mother and father were intellectually limited and that the child was in danger. The reporter said that Bobby (child) was twelve months old and was an only child. His mother and father, Melanie (mother) and Charles Larsen (father), were married and living together. A record check provided the following information on the parents:

| Name | Previous CPS Service Cases | Previous CPS Reports | Criminal Arrests | Convictions | Mental Health Services | Developmental Disabilities | Age | Sex | Race |
|---|---|---|---|---|---|---|---|---|---|
| Melanie Larsen | None | Neglect report six month ago | None | None | None | Childhood Head Injury/Seizures | 21 | F | White |
| Charles Larsen | CPS child-victim, with foster care | None | One Previous arrest for allegedly hitting wife | No conviction for dom. violenc | None | Mild Mental Retardation | 22 | M | White |
| Bobby Larsen | None | Alleged neglect victim | - | - | No record | No record | 1 | M | White |

The assessment worker visited the home. The mother and father and infant were home at the time in their small apartment. *Charles (father) wanted to know who had reported them and made a comment about nosy people and nosy workers.* The worker noted that the apartment was messy and looked like it had not been cleaned for some time. The rugs and floor were dirty and cluttered. Bobby (child) was 13 months old. He had a dirty face, arms and legs and at the time needed to have his diapers changed. When asked about feeding, Melanie (mother) said that Bobby (child) still took a bottle but she was also feeding him baby food. Bobby (child) was crawling on the floor and trying to pull up on the sofa.

The worker asked if she could look around the apartment. *The worker found no baby food or formula in the refrigerator.* She noticed that the bathroom stool was stopped up and asked about it. Melanie (mother) said that they had told the apartment owner about it several times in the past three weeks but he would not fix it. So, they were using the bathroom at a nearby service station. She said they were behind on the rent. Melanie (mother) said that she received SSI and food stamps. They said that their money sometimes ran out toward the end of the month and they had to make do until Melanie's (mother) next SSI check came. Melanie (mother) said her parents helped them with food when things really got bad but that they were older and retired and living on Social Security and did not have much to spare. Charles's (father) parents provided no support. When asked about the dirty diaper, Melanie (mother) said that diapers were so expensive that she sometimes waited to change them.

Melanie (mother) said she took pills for her seizure disorder. *The worker noticed that Melanie (mother) had a black eye and facial contusions. When she asked Melanie (mother) about it, Charles (father) said she had had a seizure and had hit her face on a table. Melanie (mother) said she could not remember.* The worker engaged the parents in a discussion about hygiene. *Melanie (mother) said she wanted to help and would do whatever the worker wanted them to do. Charles (father) disagreed and said that there was nothing wrong with the apartment and asked what business was it of other people how often they cleaned.*

## 3. Low-Risk Abuse Vignette

A report was received from a third grade teacher concerning one of her students, Carolyn Larsen (child 1). Carolyn (child 1) is eight years old. The teacher asked the child about a bruise/welt on her arm and hand. She first said her arm had been caught in the screen door at home, but after the teacher talked to her she admitted that she received them while trying to shield her legs as her mother's boyfriend Charles (mother's boyfriend) was spanking her with his belt. The teacher asked if she had bruises elsewhere. Carolyn (child 1) said yes and showed her several strap bruises on thighs. She told the teacher that her mother was out and Charles (mother's boyfriend) was watching TV and told her to bring him a beer from the refrigerator but she could not get it out of the plastic holder. He had to get the beer himself and he got mad and spanked her for

being disobedient. Her mother told her to tell anyone who asked that she had caught her arm in the door.

| Name | Previous CPS Service Cases | Previous CPS Reports | Criminal Arrests | Convictions | Mental Health Services | Developmental Disabilities | Age | Sex | Race |
|---|---|---|---|---|---|---|---|---|---|
| Melanie Larsen | None | None | None | None | None | None | 30 | F | White |
| Charles McCune | None | None | Two driving while intoxicated | One DWI conviction, 1 year ago | None | None | 32 | M | White |
| Carolyn Larsen | None | None | - | - | No record | No record | 8 | F | White |
| Brenda Larsen | No record | No record | - | - | No record | No record | No record | No record | No record |

The assessment worker visited Carolyn (child 1) at school where she confirmed the bruising. Carolyn (child 1) told the worker that Charles (mother's boyfriend) did not live with them but sometimes he slept over. She said when he drank beer she was afraid of him because he yelled at her. She said he had whipped her two times before.
The worker visited Melanie (mother) at her home with Carolyn (child 1) and her four-year-old sister Brenda (child 2). Charles (mother's boyfriend) was not present. Melanie (mother) said that Charles (mother's boyfriend) was over the day before yesterday and that he agreed to watch the girls while she went out for a while. She was not aware that Charles (mother's boyfriend) had spanked Carolyn (child 1) until the next day when she saw the bruises. *She said she told Charles (mother's boyfriend) that he was never to use his belt on Carolyn (child 1). When asked about earlier spankings she said that he had swatted Carolyn (child 1) on the butt once.*

Melanie (mother) said she was "really low" for a long time after her divorce from her husband at that time, who had been abusive to her, and that she did not work but lived on welfare, food stamps and occasional child support checks from her former husband. *She said she had found a part-time job earlier in the year and that income was helping. Melanie (mother) said that Charles (mother's boyfriend) had a full-time job and sometimes helped her with some of her bills.*

*Melanie (mother) expressed concern about Charles's (mother's boyfriend) drinking and his temper*. Carolyn (child 1) volunteered that she was afraid of Charles (mother's boyfriend) and that he sometimes called her bad names when her mother was not there. Melanie (mother) seemed surprised at this. *After this, Melanie (mother) agreed not to leave the girls alone with Charles (mother's boyfriend) again.*

Two days later the worker visited again. Charles (mother's boyfriend) admitted spanking the child but said that he had received much worse when he was growing up and that it had done him good. He said it was important to discipline children. He said that Carolyn (child 1) was making noise with her sister while he was trying to watch TV, that he had asked them to be quiet but they would not and that is why he spanked her. The worker said that punishing Carolyn (child 1) with a belt was inappropriate. *Charles (mother's boyfriend) seemed unconvinced but said he would not use a belt again.*

## 4. High-Risk Abuse Vignette

A report was received from a third grade teacher concerning one of her students, Carolyn Larsen (child 1). Carolyn (child 1) is eight years old. The teacher asked the child about a bruise/welt on her arm and hand. She first said her arm had been caught in the screen door at home, but after the teacher talked to her she admitted that she received them while trying to shield her legs as her mother's boyfriend Chuck (mother's boyfriend) was spanking her with his belt. The teacher asked if she had bruises elsewhere. Carolyn (child 1) said yes and showed her several strap bruises on thighs. She told the teacher that her mother was out and Charles (mother's boyfriend) was watching TV and told her to bring him a beer from the refrigerator but she could not get it out of the plastic holder. He had to get the beer himself and he got mad and spanked her for being disobedient. Her mother told her to tell anyone who asked that she had caught her arm in the door.

| Name | Previous CPS Service Cases | Previous CPS Reports | Criminal Arrests | Convictions | Mental Health Services | Developmental Disabilities | Age | Sex | Race |
|---|---|---|---|---|---|---|---|---|---|
| Melanie Larsen | None | One--child witnessed domestic violence | None | None | None | None | 28 | F | White |
| Charles McCune | None | None | Two driving while intoxicated | One DWI conviction, 1 year ago | None | None | 32 | M | White |
| Carolyn Larsen | None | Alleged victim | - | - | No record | No record | 8 | F | White |
| Brenda Larsen | No record | No record | - | - | No record | No record | No record | No record | No record |

The report for a child witnessing domestic violence was received three years ago based on an alleged altercation between Melanie (mother) and her husband at that time. Carolyn (child 1) was the alleged victim of that report.

The assessment worker visited Carolyn (child 1) at school where she confirmed the bruising. Carolyn (child 1) told the worker that Charles (mother's boyfriend) did not live with them but sometimes he slept over. She said when he drank beer she was afraid of him because he yelled at her. She said he had whipped her two times before.
The worker visited Melanie (mother) at her home with Carolyn (child 1) and her four-year-old sister Brenda (child 2). Charles (mother's boyfriend) was not present. Melanie (mother) said that Charles (mother's boyfriend) was over the day before yesterday and that he agreed to watch the girls while she went out for a while. She was not aware that Charles (mother's boyfriend) had spanked Carolyn (child 1) until the next day when she saw the bruises. When asked about earlier spankings she said that he had swatted Carolyn (child 1) on the butt once. Melanie (mother) admitted that Charles (mother's boyfriend) tended to get mean when he was drinking.

Melanie (mother) said she was "really low" for a long time after her divorce from her former husband, who had been abusive to her, and that she did not work but lived on welfare, food stamps and occasional child support checks. She said she had found a part-time job earlier in the year but that it was "too much for her to handle" and she had quit. She said she was in danger of losing her apartment and that it was "a day-to-day thing." She had no help from her family.

Melanie (mother) admitted that Charles (mother's boyfriend) sometime drank too much and had a temper. Carolyn (child 1) volunteered that she was afraid of Charles (mother's boyfriend) and that he sometimes called her bad names when her mother was not there. Melanie (mother) seemed surprised at this.

Two days later the worker visited again. Charles (mother's boyfriend) admitted spanking the child but said that he had received much worse when he was growing up and that it had done him good. He said it was important to discipline children. He said that Carolyn (child 1) was making noise with her sister while he was trying to watch TV, that he had asked them to be quiet but they would not and that is why he spanked her. The worker said that punishing Carolyn (child 1) with a belt was inappropriate. Charles (mother's boyfriend) seemed unconvinced.

## Name Variations in Vignettes

| Race | Neglect | Abuse |
|---|---|---|
| Caucasian | Mother: Melanie Larsen | Mother: Melanie Larsen |
| | Father: Charles Larsen | Mother's Boyfriend: Charles McCune |
| | Child: Bobby Larsen | Child 1: Carolyn Larsen |
| | | Child 2: Brenda Larsen |

| | | |
|---|---|---|
| Black, African American | Mother: Jessica Jackson | Mother: Jessica Jackson |
| | Father: William Jackson | Mother's Boyfriend: William Washington |
| | Child: Jamal Jackson | Child 1: Tamika Jackson |
| | | Child 2: Tonya Jackson |
| | | |
| American Indian | Mother: Maria Whitefeather | Mother: Maria Whitefeather |
| | Father: Jesse Whitefeather | Mother's Boyfriend: Jesse Davis |
| | Child: Dakota Whitefeather | Child 1: Angela Whitefeather |
| | | Child 2: Samantha Whitefeather |
| | | |
| Southeast Asian | Mother: Mai Thao | Mother: Mai Thao |
| | Father: Lee Thao | Mother's Boyfriend: Lee Yang |
| | Child: Paul Thao | Child 1: Jennifer Thao |
| | | Child 2: Nancy Thao |
| | | |
| Hispanic | Mother: Jessica Martinez | Mother: Jessica Martinez |
| | Father: Anthony Martinez | Mother's Boyfriend: Anthony Perez |
| | Child: Joseph Martinez | Child 1: Angela Martinez |
| | | Child 2: Ashley Martinez |

# Appendix B. Review of Literature on the SDM Family Risk Assessment and Other Risk Assessment Instruments

Structured Decision Making (SDM) is a comprehensive risk assessment system first developed by the Children's Research Center (CRC), a division of the National Council on Crime and Delinquency, as a decision making model for child protection services (CPS) agencies. The main goals of SDM are: to provide simple, objective and reliable tools that enable workers to make the best decisions for individual cases from referral to closing, and to provide managers with information for improved planning and resource allocation (Freitag). SDM was designed to assist Child Protection Services (CPS) agencies in serving families reported for child abuse and neglect and has been adopted in whole or in part by several states. The system includes instruments designed to simplify and standardize decision making for CPS workers and supervisors. These include: 1) screening criteria, 2) response priority, 3) safety assessment, 4) risk assessment, 5) child needs and strengths assessment, 6) family needs and strengths assessment, 7) case planning and service standards, and 8) case reassessment. The SDM Family Risk Assessment (FRA) is one tool among many within the SDM system.

At least 42 US states use at least some type of risk assessment tool to aid child welfare decisions (Lyons, Doueck & Wodarski, 1996). Since SDM was developed in 1988, approximately 13 child protection jurisdictions have implemented all or part of this model, namely: Alaska, Michigan, Rhode Island, Vermont, New Hampshire Indiana, Wisconsin, New Mexico, Georgia, California, Minnesota, Florida and Missouri. (See Freitag; CRC, 2003).

## Evaluation of SDM by Researchers Associated with CRC

The Children's Research Center has performed several internal studies on the FRA, including an evaluation of the Michigan FRA (Baird et al., 1995); measurements of reliability and validity (Baird et al., 1999; Baird & Wagner, 2000); and a study of equity (Baird, Ereth & Wagner, 1999). Initial process and impact evaluations of the Michigan FRA by Baird and colleagues (1995) compared cases in counties using the SDM system with those that were not. Researchers reported that in SDM counties, "cases that were closed without services had fewer new reports and substantiations, fewer subsequent removals, and fewer child injuries reported over a 12 month follow up period (Baird et al., 1995). Rates of new substantiations for abuse or neglect in SDM counties were less than half of those in comparison counties. Other FRA evaluation data (Wagner and Bell, 1998) cited in an Office of Juvenile Justice and Delinquency bulletin, "indicate that

accurate identification of families with the greatest potential for subsequent maltreatment, together with appropriate allocation of resources, can play a significant role in protecting children from harm."

The FRA is comparable to other risk assessment models that define risk assessment as the prediction of whether a child will be maltreated in the future—that is, recurrence of maltreatment (Camasso & Jagannathan, 1995; Fuller & Wells, 2001; NCCD, 2003). Alternative definitions of risk assessment include assessing the risk of substantiation or the risk of maltreatment. In a more general synopsis, risk assessment tools were categorized as either empirically based, consensus based or blended instruments (Cash, 2001). Empirically based (or actuarial) instruments include only risk items or factors that have been determined to be correlated with recurrence through research. Consensus based instruments include factors that are selected on the basis of a process and rely heavily on the judgments of trained case workers, while blended instruments consist of items that were originally developed through non-empirical means but later validated through research.

The SDM FRA, which is an actuarial model, underwent further testing when it was compared to two commonly used, consensus-based, risk assessment instruments—the Washington Risk Assessment Matrix (WRAM) and the California Family Assessment Factor Analysis (CFAFA)[54]. The study, funded by the U.S. Department of Health and Human Services' Office of Child Abuse and Neglect (OCAN), examined the reliability and validity of these three risk assessment tools. CRC researchers (Baird et al., 1999; Baird & Wagner, 2000) applied these instruments to a sample of 1,400 families from four different states: Oakland, California; Dade County, Florida; Kansas City, Missouri; and Macomb, Muskegon, Ottawa and Wayne counties in Michigan. These sites were chosen for their geographic representation and ethnic representation of minority populations. Reliability tests indicated a significantly higher rater-reliability for the Michigan SDM instrument than for either the California or the Washington models (Baird et al, 1999). Three out of four raters (75 percent) agreed on 85 percent of the cases for the FRA, compared with 31.3 percent for the Washington instrument and 45.1 percent for California instrument. This pattern remained consistent even after further analyses were carried out to overcome possible biases.

Relative validity of consensus versus actuarial models was measured by "the ability of each system to classify cases into risk groups with significantly different rates of subsequent maltreatment reported," as well as with the Dispersion Index for Risk (Baird & Wagner, 2000). Essentially, the Michigan FRA instrument was found to have significantly higher validity than the Washington or California models. Using the FRA, risk classifications for new investigations increased from 16 percent (low) to 32 percent (moderate) to 46 percent (high). While each of the differences was significant ($p < 0.001$), this was not the case for the Washington and California risk instruments. Both of the latter instruments showed significant differences between the low risk and moderate

[54] This California risk assessment tool was a consensus-based model that was used prior to the implementation of SDM. The California Structured Decision Making instrument has been empirically evaluated.

levels but not between the moderate and high risk designations. Risk classifications for the Washington model ranged from 25 percent (low) to 35 percent (moderate) to 39 percent (high), and the California model classifications were 28 percent (low), 38 percent (moderate) and 38 percent (high). Regarding racial equity, all three instruments effectively classified African Americans—when outcomes were computed by risk level. The Michigan SDM model correctly classified Whites as well, but "there was little relationship between risk ratings and outcomes for Whites" in the Washington and California risk assessments systems (Baird & Wagner, 2000). These findings, however, were limited by small sample sizes after dividing the sample by race.

One of the goals of the CRC in implementing the FRA was to ensure "that differences between races in maltreatment rates recorded at each risk level are minimized" (Baird, Ereth & Wagner, 1999). The rationale for this objective is supported by findings from the National Incidence Study of Child Abuse and Neglect, which found no race differences in maltreatment incidences (Sedlack & Broadhurst, 1996). While these NIS results have been disputed in the literature (Derezotes, 2003), they raise important concerns—when viewed in context of the disproportionate representation of children of color in the child welfare system (AFCARS, 2001; Derezotes, 2003). In order to assess the level of equity inherent in the FRA, CRC researchers (Baird, Ereth & Wagner, 1999) reviewed data from the three largest states using the SDM FRA actuarial system—California, Georgia and Michigan. Results indicated that these actuarial risk assessment models assigned virtually equal numbers of African Americans and Whites to each risk level. In fact, Whites tended to score at slightly higher risk levels than African Americans. For example, in Georgia Whites had an average risk rating of 1.852 while the average rating for African Americans was 1.763. A different independent study for the California Department of Social Services also found no bias in any item or in the SDM instrument as a whole (Johnson, 1999). According to CRC researchers, actuarial risk assessment instruments produce high levels of equity because of their structural components. The separation of SDM instruments into an abuse scale and a neglect scale is "key to achieving equity as different family characteristics are related to the recurrence of different types of maltreatment" (Baird, Ereth & Wagner, 1999). This accounts for differences in racial trends, which indicate, for example, that African Americans have a higher prevalence of neglect, while Whites have higher prevalence of abuse. Another structural advantage of actuarial risk assessment instruments is that the factors used to rate the risk of recurrence were chosen because they relate directly to subsequent reports in families with previous *investigated* allegations. Therefore, if the average score on an item differs between races, this is a reflection of the actual differences in outcomes.

## Other Risk Assessment Research

Numerous other studies have been conducted on risk assessment instruments in general. An extensive literature review by Lyons, Doueck & Wodarski (1996), assessed 10 risk assessment models and categorized them into five basic types: (1) matrix models with tables of risk factors rated in terms of their severity (e.g. Illinois Child Abuse & Neglect Training System), (2) empirical/actuarial predictor models with small sets of

factors found to be predictive of substantiation or recurrence (e.g. Washington Assessment of Risk Model), (3) behaviorally anchored scales that assess levels of child or parent functioning (e.g. Child Well Being Scales), (4) comprehensive ecologically structured scales (such as the Child at Risk Field), and (5) computerized expert systems, which use a blend of CPS expertise and artificial intelligence to derive computer-based decision rules. The authors examined the inter-rater reliability of the five models and concluded that while overall scale reliability was good, the actual performance of the models—measured by mean single-rater correlations—was substandard in comparison. In general, measures of internal consistency and concurrent validity appeared to be at acceptable levels. However, rates of false positives ranged from 66 percent (Vermont) to 14 percent (Washington) and no evidence was found to support the claim that variables that predict occurrence also predict recurrence.

In more recent studies, further tests of reliability and validity have been applied to risk assessment models and the risk factors incorporated therein. English and colleagues (1998) conducted analyses to determine the classification accuracy rates across types of abuse and post investigation level of risk, sensitivity and specificity of the risk factors in the Washington Risk Assessment Model (WRAM). The accuracy rates ranged from 79.1 percent to 84.6 percent, the sensitivities ranged from 78.3 percent to 85.1 percent and the specificities ranged from 80.0 percent to 84.8 percent. Inter-rater reliability coefficients for the Illinois Child Endangerment Risk Assessment Protocol (CERAP) reportedly ranged between .83 and .90 (Fuller, Wells & Cotton, 2001). The same model was evaluated for predictive validity by comparing rates of recurrence of maltreatment within 60 days of an initial report for the one-year period prior to CERAP implementation and the one-year following implementation. Results indicated recurrence of abuse or neglect in the 60 days after a child's first report decreased by 28.6 percent. Furthermore, cases that did not utilize the CERAP were at higher risk (four times as high) for short-term recurrence than those that did.

Many researchers have studied risk factors typically included in risk assessment tools and documented their impact on the recurrence of abuse and neglect. Examples include studies exploring risk reduction at different points in the life of a case (Lyle & Graham, 2000; Fuller & Wells, 2001), trying to determine which variables are significant predictors of recurrence (Rittner, 2002; Cash, 2001; Baird et al., 1995), and measuring convergent validity of several risk factors (English & Graham, 2000). Lyle & Graham (2000) conducted an outcome study of the Illinois Child Abuse and Neglect Training System (CANTS) in which data on initial and closing risk levels were compared. Results were highly significant and indicated a decrease in risk from the initial administration of the instrument to its use at closing. While the researchers posited that maltreatment appeared to be reduced by CPS intervention, they also warned that these differences might have been largely due to the artificial inflation of risk scores by caseworks (to ensure children's acceptance for ongoing services). Similarly, Fuller and Wells (2000) measured predictors of maltreatment recurrence in a study examining the usefulness of the Illinois CERAP at two points in the life of a case. The first study looked at risk factors at investigation initiation while the second focused on intact family cases when they were opened for services. Three factors were found to be predictive in both studies:

number of prior indicated reports of maltreatment, number of caretaker problems, and characteristics of service delivery. Other factors were found to be important only at one particular point in the life of the case. At the initial investigation point, for example, two factors—age of the youngest child and a single parent living alone with children—significantly predicted maltreatment recurrence in the presence of other factors such as type of maltreatment, case disposition, number of family problems and prior reports on perpetrator.

Actuarial risk assessment instruments are comprised of factors that have been empirically associated with occurrence or recurrence of maltreatment. The Michigan Structured Decision Making system, for example, used a stratified random sample of 1,896 cases to select qualified factors (Baird et al., 1995). Data was collected from these cases and relationships between family characteristics and case outcomes were examined to determine which risk factors would be included in its risk assessment tools. Rittner (2002) also used a stratified sample of 500 to observe whether commonly used variables were predictive of maltreatment recurrence. Results of the study "offered little support for using variables employed by risk assessment instruments to predict which caretakers were most likely to re-abuse because re-abusers and non-re-abusers shared many features" (Rittner, 2002). Most variables were only somewhat predictive of recurrence, with moderate associations found for the following variables: past history as victim of abuse, substance abuse, mental health problems and poverty. Independent of other environmental factors, substance abuse was not strongly predictive of re-abuse.

In an analytical review of factors empirically known to be associated with child maltreatment, Cash (2001) distinguished between factors that predict occurrence and those that predict recurrence. Contradictory to the results found by Rittner (2002), factors such as past histories as victim of abuse, substance abuse and poverty were notably documented as predictors of maltreatment *occurrence*. Other predictors of occurrence were maternal and parental depression, unemployment, social isolation, lack of social support system, unrealistic expectations of the child, and increased family stress (Cash, 2001). Predictors of *recurrence* were parent's unrealistic expectations of child, perpetrator's access to child, multiple children in home, poor parenting skills, child's fear of caretaker, number of in-person visits to family receiving CPS services (Cash, 2001; English et al, 1994). English & Graham (2000) cautioned that models based on occurrence may not be useful in predicting recurrence. Using data from Longitudinal Studies of Child Abuse and Neglect (LONGSCAN), these researchers studied correlations between CPS worker's ratings of risk on the WRAM and independent measures of the same risk contrasts collected by research interviewers. Convergent validity was measured, rather than recurrence, and significant correlations were found on measures of caregivers' emotional and physical health. The validity of scales such as WRAM was brought into question due to the lack of correlation between workers ratings and child risk factors associated with developmental or behavioral issues or soci-economic factors such as stress or support. Identically, Camasso and Jagannathan (2000) found no support for the severity, chronicity, parent/child and caretaker characteristics on WRAM as predictors of maltreatment recurrence.

Research on the development and evaluation of risk assessment instruments for Child Welfare is ongoing and continues to face many difficulties. Validity of some currently used instruments has not yet been determined and reliability is often low. Statistical results are also minimized by methodological problems such as the inaccurate use of correlation coefficients or estimates of association to rate reliability (Baird, 1999). Many studies measure only predictive validity and do not examine convergent and discriminant validity (English et al, 1999). Gambrill and Shlonsky (2000) name several other challenges such as vague definitions of outcome measures, changes in risk over time, the absence of base-rate data, low predictive capacity for individuals despite high overall predictive validity and the inclusion of risk factors that may not have any predictive validity.

While the SDM FRA is empirical in nature, decisions of workers using this or any other tool may still be influenced by personal characteristics, confirmation bias, limited information and environmental factors (Gambrill and Shlonsky, 2000). Agencies using and evaluating the Structured Decision Making model must take these concerns into account and seek concrete and practical solutions. Further research, combined with insight gleaned from practice wisdom, is necessary in ensuring greater success for child welfare decision-making and the health and safety of the children and families served.

## References


AFCARS (Adoption and Foster Care Analysis and Reporting System). (2001). *The AFCARS Report*. Retrieved September 24, 2003, from U.S. Department of Health and Human Services, Administration for Children and Families: http://www.acf.dhhs.gov/programs/cb/publications/afcars/june2001.htm.

Baird, C., Ereth, J., & Wagner, D. (1999). *Research Based Risk Assessment: Adding Equity to CPS Decision Making*. Retrieved September 24, 2003, from Children's Research Center: http://www.nccd-crc.org/crcindex.htm.

Baird, C., Wagner, D., Caskey, R., & Neuenfeldt, D. (1995). *The Michigan Department of Social Services Structured Decision Making System; An Evaluation of it's Impact on Child Protection Services*. Retrieved September 24, 2003, from Children's Research Center: http://www.nccd-crc.org/crcindex.htm.

Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk Assessment in Child Protective Services: Consensus and Actuarial Model Reliability. *Child Welfare, 78*(6), 723-748.

Baird, C., & Wagner, D. (2000). The Relative Validity of Actuarial and Consensus-Based Risk Assessment Systems. *Children and Youth Services Review, 22*(11/12), 839-871.

Camasso, M., & Jagannathan, R. (1995). Prediction Accuracy of the Washington and Illinois Risk Assessment Instruments: An application of receiver operating characteristic curve analysis. *Social Work Research, 19*(3), 174-183.

Camasso, M., & Jagannathan, R. (2000). Modeling the Reliability and predictive Validity of Risk Assessment in Child Protective Services. *Children and Youth Services Review, 22*(11/12), 873-896.

Cash, S. (2001). Risk Assessment in Child Welfare: The Art and Science. *Children and Youth Services Review, 23*(11), 811-830.

CRC (Children's Research Center). (2003). Retrieved September 24, 2003, from National Council on Crime and Delinquency: http://www.nccd-crc.org/crcindex.htm.

Derezotes, D. (2003). *Examining Child Maltreatment and the Impact of Race in Receipt of Child Welfare Services in the United States*. Retrieved September 24, 2003, from Family Violence Prevention Fund: http://endabuse.org/programs/children/files/prevention/ChildWelfareResearch.pdf

English, D., & Pecora, P. (1994). Risk Assessment as a Practice Method in Child Protective Services. *Child Welfare, 73*(5), p 451.

English, D., & Graham, C. (2000). An Examination of Relationships Between Children's Protective Services Social Worker Assessment of Risk and Independent LONGSCAN Measures of Risk Constructs. *Children and Youth Services Review, 22*(11/12), 897-933.

Freitag, R. *Structured Decision Making*. Retrieved September 24, 2003, from Children's Research Center: http://www.socservices.com/best-practices-pdf/childwelfare.childrensresearch.pdf.

Fuller, T., Wells, S., & Cotton, E. (2001). Predictors of Maltreatment Recurrence at Two Milestones in the Life of a Case. *Children and Youth Services Review, 23*(1), 49-78.

Gambrill, E., & Shlonsky, A. (2000). Risk Assessment in Context. *Children and Youth Services Review, 22*(11/12), 813-837.

Johnson, W. (1999). *Race and the California Family Risk Assessment*. San Francisco, CA: Paper presented at the 13th Annual CPS Risk Assessment Roundtable.

Lyle, C., & Graham, E. (2000). Looks Can Be Deceiving: Using a Risk Assessment Instrument to Evaluate the Outcomes of Child Protection Services. *Children and Youth Services Review, 22*(11/12), 935-949.

Lyons, P., Doueck, H., & Wodarski, J. (1996, September). Risk Assessment for Child Protective Services: A review of the empirical literature on instrument performance. *Social Work Research, 20*(3), 143-155.

Rittner, B. (2002). The Use of Risk Assessment Instruments in Child Protective Services Case Planning and Closures. *Children and Youth Services Review, 24*(3), 189-207.

Sedlak, A., & Broadhurst, D. (1996). *Executive Summary of the Third National Incidence Study of Child Abuse and Neglect*. Retrieved September 24, 2003, from National Clearinghouse on Child Abuse and Neglect Information: http://www.calib.com/nccanch/pubs/statinfo/nis3.cfm.

Tatara, T. (1996). A Survey of States on CPS risk assessment practice: Preliminary findings. In E. Gabmrill & A. Shlonsky , *Risk Assessment in Context.* Children and Youth Services Review *22(*11/12), 813-837.

Wagner, D., & Bell, P. (1998). The use of Risk Assessment to Evaluate the Impact of Intensive Protective Service Intervention in a Practice Setting. In (Wiebush, R., Freitag, R., & Baird, C.) *Preventing Delinquency Through Improved Child Protection Services* (Juvenile Justice Bulletin). Retrieved September 24, 2003, from U.S. Department of Justice, Office of Juvenile Justice and Delinquency: http://www.ncjrs.org/html/ojjdp/jjbul2001_7_1/contents.html.