1991

Development of an instrument to identify newborns at risk of child abuse or neglect

Denise Huan-hsuan Kung Ihnat
Yale University

DEVELOPMENT OF AN INSTRUMENT TO IDENTIFY NEWBORNS AT RISK OF CHILD ABUSE OR NEGLECT

Denise Huan-hsuan Kung Ihnat

Yale University

1991

DEVELOPMENT OF AN INSTRUMENT TO IDENTIFY NEWBORNS AT RISK OF CHILD ABUSE OR NEGLECT

A Thesis Submitted to the Yale University School of Medicine in Partial Fulfillment of the Requirements for the Degree of Doctor of Medicine

by

Denise Huan-hsuan Kung Ihnat

1991

Table of Contents

Table of Contents (continued)

List of Figures

List of Tables

List of Tables (continued)

Dedication

To Mom and Dad,

for medical school

To Rick,

for this thesis

With love and appreciation

Acknowledgments

This thesis exists only because of the help and encouragement of many friends. Dr. Berne Singsen first started me on the path of scientific inquiry and has been an invaluable source of advice and encouragement. Dr. Alvin Feinstein and the scholars of the Robert Wood Johnson Clinical Scholars Program (especially Jim Wright and David Bruce) were kind enough to teach me some clinical epidemiology and help me sort out innumerable confusions. Dr. David Schonfeld and Dr. James Jekel kindly corrected my navigation when I (not infrequently) began to go adrift. Dr. Herb Goldenring helped me to introduce the study to the pediatric residents and started each of my mornings on the wards with a cheery "hello."

The day to day work was made infinitely easier by Lyla Johnson (thank you for liking to go to the nursery), Jo Anne Mezger (thank you for being SAS-literate), Kathy Toensmeier and Julia Robertson (thank you for never getting tired of telling me 32965), and the kind nurses, pediatric residents, and social workers (especially Andrea Seigerman, Joyce Lewis, Sharon Davis, Anne Murphy, Debra Iverson, Jackie Wilson) who care for the mothers and the newborns on Yale-New Haven Hospital's postpartum floors and who tirelessly filled out hundreds of salmon-colored forms. The practical work of dealing with the computers was aided by my buddy Scott Fritz (despite the 1300 mile gap) and my brother David (despite going to the "other" school).

I am indebted to Julia Hamilton and Diana Wasserman for serving as two of my child maltreatment experts and for being both nice people and easy to work with. I especially wish to thank Diana for helping me to keep the thesis in perspective and knowing when I needed a pep talk. Without Dr. Domenic Cicchetti, I doubt I ever would have understood what my mountains of data meant nor ever have been able to analyze it. I looked forward to my V.A. visits and the chance to work with someone who was always enthusiastic about this project.

Finally, I would like to thank the two most influential people in my life over the past two years. Dr. John Leventhal has been a research advisor extraordinaire, and I also thank him for being an advocate of adopting lost homeless puppies. His experience in clinical research and in advising student research made this project feasible. His friendship made the project fun. My husband Rick saw the importance and value of doing this project before I ever did and, as he is wont to do, kept me from doing stupid things. He also continually woke me up and made me go to work.

To all of my friends, I say Thank You! and Ta Da! It's Done!

ABSTRACT

DEVELOPMENT OF AN INSTRUMENT TO IDENTIFY NEWBORNS AT RISK OF CHILD ABUSE OR NEGLECT

Denise H. Kung Ihnat, Domenic V. Cicchetti,
and John M. Leventhal.
Department of Pediatrics, Yale University,
School of Medicine, New Haven, CT.

The purpose of this study was to develop a rating scale that structures clinicians' evaluations of a newborn's risk of future maltreatment; compared to unstructured judgments, such an instrument should improve the ability of clinicians to recognize those patients who would benefit most from scarce social services. The Clinical Rating Scale (CRS) was composed of 22 risk factors for child maltreatment. Each item was rated on a four-point dichotomous-ordinal scale. On the CRS, a clear description was given for each level of risk for each item. The CRS yielded a binary rating of either High or Low Risk. Thirty-two newborns and their mothers were evaluated by pairs of experts in the field of child maltreatment using their unstructured clinical judgments (of which the consensus rating served as the gold standard of prediction) and then using the CRS. Each child and his/her family were also evaluated by the nurse and pediatrician taking care of them during their post-partum hospital stay. As part of a longitudinal predictive validation study, clinicians using the CRS evaluated 363 consecutive newborns.

The agreements in risk ratings by the pairs of experts using unstructured judgments (kappa=0.80) and the CRS (kappa=0.65) were computed. The sensitivity (SN) and specificity (SP) of the CRS when used by the experts were computed (SN=100%, SP= 51%); the agreement of each expert's CRS rating compared to his/her own unstructured judgment was computed (kappa=0.45). The sensitivity and specificity

of the predictions by the non-experts using both their unstructured judgments (SN_{nurse}=40%, SP_{nurse}=100%; SN_{peds}=63%, SP_{peds}=92%) and the CRS (SN_{nurse}=64%, SP_{nurse}=78%; SN_{peds}=63%, SP_{peds}=92%) were computed. Item by item inter-rater agreement from the longitudinal data using weighted kappa showed good inter-rater agreement.

The CRS structures clinical judgment about risk of maltreatment, improves the identification by non-expert clinicians of those at risk, and yields good inter-rater agreement in judgments. It should be useful to clinicians in identifying children and families at high risk of subsequent abuse or neglect.

CHAPTER ONE:

INTRODUCTION

In 1962, Henry Kempe and his colleagues gave the first formal medical recognition to a medical/social problem that has afflicted families for centuries. Their pioneering work in defining the "battered-child syndrome" laid the groundwork for the acceptance of child maltreatment as a real problem that clinicians must be able to recognize and to treat.

The article by Kempe and colleagues triggered intense research interest in the causes, modes, symptoms, and effects of child maltreatment. While the first article focused on identifying those children who had already suffered maltreatment, the question soon arose whether it would be possible to identify those children who are as yet unharmed but who are at risk of being abused. Armed with the ability to recognize such children, health professionals could enlist support services (such as parenting classes, social work follow-up, support groups, or a visiting nurse for in-home visits) for those families who need it and thus prevent child maltreatment from occurring.

Over twenty years of work by many different researchers and clinicians has produced a multitude of approaches and instruments. Some methods utilize an intense psychosocial interview of parents (or parents-to-be) while others rely on a check-list of socio-economic issues considered to be risk factors. Still others have attempted to predict maltreatment using only unstructured clinical judgment rooted in

clinical experience. These different approaches also have produced a multitude of results ranging from excellent to poor predictive accuracy.

In order for a predictive instrument to be useful in the routine screening of families it must be non-threatening to the patients and fairly easy to complete with high predictive accuracy, sensitivity, and specificity. None of the instruments previously developed fulfills these criteria. Those with excellent statistical indices require lengthy interviews or psychological profiles, sometimes over several visits, which make them unfeasible for routine screening and clinical use. Simpler instruments have poor statistical indices because of reliance on items not highly predictive of child maltreatment and omission of those items that are highly predictive. The reason for this is that most instruments have been developed for research purposes with goals other than routine screening in mind.

The purpose of this study was as follows:

1) To develop a predictive instrument designed specifically to screen all families of newborns in order to identify early those at risk. This instrument would standardize clinical judgment about parenting ability and would be convenient enough to use on all patients during the postpartum period.
2) To refine the instrument through pilot testing.
3) To develop a scoring method for the instrument.
4) To test the sensitivity and specificity of the instru-

ment compared to a gold standard of prediction (concurrent validity).

5) To examine inter-rater agreement among expert clinicians.

6) To assess which items on the instrument are difficult for non-experts to judge.

Since both preventive services and in-depth psychological evaluations for abuse potential are expensive and difficult to offer to an entire population, this new instrument will be a valuable first-step screening instrument that should identify those families in need of the more intensive evaluations and preventive services.

Along with a critical review of past research on risk factors for child maltreatment and past research on predictive instruments, this thesis gives a perspective on the problem of child maltreatment and the need for a way to identify those at high risk. This thesis also examines the area of clinical scales, the measurement of "soft" data, and the problems that must be overcome to develop such an instrument. These topics are all covered in Chapter Two. Chapter Three details the methods employed in conducting this study, Chapter Four contains the results, and Chapter Five contains a discussion of the results, a discussion of the ethical issues involved in a routine screening program, and a proposal for the minimum criteria that a screening instrument must meet in order to be useful.

CHAPTER TWO:

BACKGROUND

An Historical Perspective on Child Maltreatment

Child maltreatment can be defined as any form of child care that physically or psychologically harms a child. Kempe and colleagues defined "the battered child syndrome" in 1962 as a clinical condition (fractures, bruises, soft tissue swelling, failure-to-thrive, sudden death) that resulted from physical abuse (Kempe et al., 1962). Maltreatment goes beyond physical abuse, though. Fifty percent of children with failure-to-thrive have no organic cause; theirs is a problem in parenting and is estimated to affect one in one hundred American children (Harris, 1982). The effects of a lack of sensory or social stimulation on growth and development are well documented in institutionalized children. The effect on children who are deprived of such stimulation by their parents is the same, and these children can experience non-organic failure-to-thrive (Harris, 1982). In 1981, the Select Panel for the Promotion of Child Health recognized blows, burns, sexual assault, starvation, confinement, exposure to unsafe environments, and absence of affection or attention all as forms of child maltreatment (U.S. Dept. HHS, 1981). Child maltreatment, then, includes physical or verbal violence, sexual exploitation, neglect of basic needs, and abandonment.

Acceptance of this definition is predicated on an acceptance of children as people possessing human rights. Historically, children have not been viewed as such. Nei-

ther the abandonment of children in ancient Rome (as in the story of Romulus and Remus), nor the sacrificing of children in Biblical times (as in the stories of Abraham and Isaac, and the killing of male babies by Herod) were viewed as heinous crimes. In colonial America, it was considered within parental rights to whip, castrate, or kill misbehaving children. This concept is reflected in the adage "spare the rod and spoil the child" (Straus, Gelles, Steinmetz, 1981).

Legal protection for children from their parents is a recent historical development. There existed in the United States a Society for the Prevention of Cruelty to Animals before there was a Society for the Prevention of Cruelty to Children; thus, the first case of child maltreatment was prosecuted in 1874 under the SPCA laws (Straus, Gelles, Steinmetz, 1981). It took the Social Security Act to fund the first public services in 1935 "for the protection and care of homeless, dependent and neglected children and children in danger of becoming delinquents" (U.S. Dept. HHS, 1988). It was ten more years before physicians began to consider child maltreatment a national problem; this occurred when radiologists started to notice a recurrent pattern of healing bone that was characteristic of fractures resulting from intentional blows rather than accidental injury. It was the work of Kempe and colleagues, though, that highlighted and clearly defined the problem in 1962, stimulating both widespread public and medical concern (Straus, Gelles, Steinmetz, 1981). From 1963 to 1966, 49 states en-

acted laws requiring the reporting of suspected cases of child abuse or neglect to a designated public agency, and by the end of the 1960's all 50 states had such laws (U.S. Dept. HHS, 1988; Straus, Gelles, Steinmetz, 1981). In 1974, the Child Abuse Prevention and Treatment Act created the National Center for Child Abuse and Neglect (NCCAN) to "support state and local efforts" at prevention and treatment (U.S. Dept. HHS, 1988).

Without recognition of child maltreatment as a problem, much less reporting laws, it is difficult to estimate the extent of child maltreatment prior to the mid-1960's. It is known that in 1968, more children less than five years of age died from parental injuries than from tuberculosis, whooping cough, polio, measles, diabetes, rheumatic fever, and appendicitis combined (Straus, Gelles, Steinmetz, 1981). One of the tasks undertaken by the NCCAN was to clarify the extent of the problem. The NCCAN initiated a Study of the National Incidence and Prevalence of Child Abuse and Neglect (a.k.a. National Incidence Study or NIS). The original study, which was completed in 1980, was updated in a second study, completed in 1986. This second study, NIS-2, counted those children "who experienced <u>demonstrable harm</u> as a result of maltreatment" (the core estimate) as well as a "supplementary estimate" of children endangered (at risk but not harmed yet). NIS-2 included children in both categories who were known to protective services or any third party (e.g. day care or hospital personnel) (U.S. Dept. HHS, 1988).

NIS-2 found that in 1986, 16.3 out of every 1000 children, or more than 1.02 million children, experienced "demonstrable harm" (the core estimate). Fifty-six percent were in the form of abuse and forty-eight percent were in the form of neglect. This represented a 66% overall increase from the NIS-1 data of 1980, with a 74% increase in abuse alone, and a 200% increase in sexual abuse. There was no change in neglect. The overall fatality rate was 0.1% and was more common in younger children. When both children harmed and children endangered were counted (core estimate plus supplemental estimate), NIS-2 reported a rate of 25.2 out of every 1000 children, or 1.5 million children affected. That translates into a lifetime prevalence of from 10 to 40% of all adults who experienced some form of maltreatment as a child. Sixty-three percent of these cases involved neglect and forty-three percent involved abuse.

NIS-2 also examined reporting patterns and found that noninvestigatory agencies (e.g. schools, hospitals) discovered more than five times the number of cases than investigatory agencies (e.g. police, public health services); schools reported the most followed by hospitals and social service agencies. Of those children who had actually experienced harm, only 40% were reported to child protective services (U.S. Dept. HHS, 1988).

The American Association for Protecting Children (AAPC) found similar results for 1986. The AAPC data were compiled from reports made to individual state child protective ser-

vice agencies. They reported that 1.7 million, or 32.8 out of every 1000, children were affected involving 1.3 million families. This represents a 212% increase over the time period 1976 to 1986. Of the reported cases, approximately 40% were substantiated, the average age of the victim was 7.2 years, 52.5% of victims were female, the average age of the perpetrator was 31.7 years, 55.9% of the perpetrators were female, and 48.9% of the families were on public assistance. The racial profile of maltreated children was essentially parallel to that of all children. The age profile of maltreated children was skewed towards younger children; 43% of affected children were aged 0-5 years yet this age group made up only 34% of all children. Children aged 6-11 years made up 31% of all children but represented 33% of the maltreated population while those aged 12-17 made up 35% of the population but only 24% of the maltreated population (American Association for Protecting Children, 1988).

The extent of the problem and the steady increase in reported cases over the past decade can be attributed in part to an emphasis on community awareness of the problem and thus, increased willingness of people to report suspected cases (U.S. Dept. HHS, 1988). However, the data probably also reflect a real increase in child maltreatment that is a part of the overall national problem of domestic violence. It is estimated that on average yearly in America, in one out of six households one spouse strikes the other spouse. In three out of five households with children at least one

parent strikes a child. In three out of five households with more than one child, there is violence among the children. Overall, half of all households experience some form of domestic violence once a year. During the entire course of a marriage, in more than one out of four couples one partner will strike the other (Straus, Gelles, Steinmetz, 1981; Straus, Gelles, 1986).

In outlining its Objectives for the Nation Concerning the Promotion of Health/Prevention of Disease in 1980, the U.S. Department of Health and Human Services ranked control of violent behavior as one of 14 priority areas. Ten specific objectives for control of violence were later outlined and included:

> By 1990, the proportion of the primary care physicians who take a careful history related to personal stress and psychological coping skills should be more that 60%....By 1990, injuries and deaths to children inflicted by abusing parents should be reduced by at least 25% (Silver, Goldston, Silver, 1984).

The Role of Prediction/Prevention

Child maltreatment in the form of physical abuse is clearly a danger to the child's life, growth, and development. However, all forms of maltreatment threaten a child's well-being. Neglect can lead to death from accidents or exposure to harmful elements. Non-organic failure-to-thrive in infancy and psychosocial dwarfism in childhood are manifestations of deprivation of food or nurturance (Harris, 1982). Victims of maltreatment also suffer from psychologi-

cal harm and will be prone towards violence themselves. Studies support the impression that victims of maltreatment will subsequently maltreat their own children (Altemeier, O'Connor, Vietze, 1982; Council on Scientific Affairs, 1985; Hunter et al., 1978; Oates et al., 1979; Widom, 1989). One study estimates that as many as 30% of child maltreatment victims will become perpetrators as parents (Egeland, 1988). The costs of child maltreatment are enormous and far-reaching and include both human costs and such costs to society as medical/psychological treatment for victims and perpetrators, court time, use of the penal system, and the lost productivity of directly affected members of society. In 1976, the cost to society of caring for children brain damaged from abuse was approximately 4.2 billion dollars annually (Rosenberg, Meyers, Shackleton, 1982).

Once a child is identified as having been maltreated, it is important that an intervention take place to protect the child and that the whole family receive treatment. Without such precautions, 50% of physically abused children will experience more maltreatment and 10% will die from it, a death rate 100 times higher than the overall death rate from maltreatment estimated by NIS-2 (Rosenberg, Meyers, Shackleton, 1982; U.S. Dept. HHS, 1988). Ideally, treatment in child maltreatment cases involves counseling for the entire family, not just the perpetrator. In order to identify cases and implement proper protection and treatment, a team effort by pediatricians, nurses, child psychiatrists,

psychologists, social workers, teachers, attorneys, and child care workers is necessary (U.S. Dept. HHS, 1981). Given proper intervention and therapy, it is estimated that as many as "90% of child abuse and neglect cases respond" (Robert Wood Johnson Foundation, 1977).

Even with successful therapy that prevents repeat episodes, once a child is maltreated he or she is likely to carry deep physical and/or emotional scars. The only way to prevent this from happening is to prevent the first episode of child maltreatment itself (Altemeier et al., 1984). Dubowitz wrote,

> Both financial and human costs associated with child maltreatment, although crudely estimated, are staggering. Prevention is, therefore, attractive as a way of reducing these costs of child maltreatment. In addition, there is the possibility that early efforts to enhance family functioning could be more effective than interventions after maltreatment has already occurred (Dubowitz, 1989).

Altemeier et al. agreed that prevention is better than treatment.

> Actually, preventing abuse before it starts may be easier than stopping it. Many of the factors which apparently predispose to parenting disorders are likely to be increased because of the maltreatment (Altemeier et al., 1979).

The Select Panel for the Promotion of Child Health concurred and advocated prevention as one of their goals.

> Because therapy of this kind is expensive, especially if it involves residential treatment, most health and social service experts stress the need for better preventive programs, based on early assessment of family risk, home health visiting by public health nurses, social workers or lay visi-

> tors, and vigorous community-based campaigns of education and crisis management (U.S. Dept. HHS, 1981).

Prevention programs fall into three broad categories, primary, secondary, and tertiary. Primary prevention targets the population at large and takes the form of community service announcements about resources for parents, tapes and demonstrations on good parenting, and efforts to heighten public awareness about the dangers of maltreating children. Secondary prevention programs target people deemed to be at high risk of maltreating their children. These programs strive to improve parenting skills and family functioning. Examples of these include latch-key children programs, programs for pregnant teenagers, and crisis intervention services like hot-lines. Tertiary prevention involves rehabilitation of those known to have maltreated their children along with psychological treatment of the victims in order to foster their growth and decrease the likelihood that they, in turn, will harm their own future children. Child protective services, foster care, and legal prosecution fall into the realm of tertiary prevention (Dubowitz, 1989).

Of the three types of prevention programs, primary prevention is used the least in the United States. These types of programs have not been carefully evaluated but it is expensive and difficult to effectively reach an entire population (Dubowitz, 1989). Most programs fall into the category of secondary prevention. These programs try to identify those families in their population that are at high risk and

then provide these families with support services. These programs report a decrease in child abuse and neglect (Dubowitz, 1989). Helfer concluded after a review of such programs, however, that thorough evaluation research into these programs are still needed in order to judge their effectiveness (Helfer, 1982). Tertiary programs are also common in the United States. They encompass actions to punish the perpetrator, protect the victim, and provide therapy for both, but these programs need thorough evaluations as well (Dubowitz, 1989).

Evidence does exist, though, that supports secondary prevention as an effective means of deterring maltreatment. A prospective study in New Zealand by Monaghan, Gilmore, Muir, et al. concluded that interventions decreased the rate of maltreatment. The Stage I group consisted of 200 families, none of whom received any form of intervention. At follow-up two years later, 52% of those families judged to be High Risk at the start of the study had experienced an adverse outcome characterized by removal of a child from the home for more than 6 months. Nine percent of those characterized as being at Moderate Risk experienced an adverse outcome while 5% of those at Low Risk and 0% of those at No Risk experienced the same outcome. The Stage II group consisted of 300 families, all of whom received such interventions as a support group lead by a social worker, access to day care facilities, support from volunteers, and "consultative resources." In this intervention population, 20% of

the High Risk group and 2% of the Moderate Risk group experienced an adverse outcome while none of the Low or No Risk groups had such an experience. The difference in the rates of adverse outcome between the intervention and non-intervention groups was significant at $p<0.01$ (Monaghan, Gilmore, Muir, et al., 1986).

The best study done to date in the United States on secondary prevention was done by Olds and colleagues in the Appalachian region of New York state. Four randomized treatment groups, each consisting of from 90 to 116 first-time mothers, were studied. Group 1 was a no-treatment control group. Group 2 received free transportation to regular prenatal and well-baby visits. Group 3 received the free transportation along with regular home visits by a nurse during the pregnancy. The fourth group received the same interventions as Group 3 but also received nurse home visitation during the first two years of the newborns' lives. In all four groups, the mothers were interviewed at registration to gather demographic and background information; the children were weighed and measured at 6, 12, and 24 months; the children were screened by an infant specialist at 1 and 2 years of age for developmental and/or sensory problems; and the state child abuse registries were searched for any reports of maltreatment on these children. In this study, 19% of the High Risk mothers (poor, unmarried teens) in the comparison group (Groups 1 and 2) were reported for abusing their babies in the first two years of life compared

to 4% of the High Risk group who received prenatal or prenatal/postpartum nurse visitation (p=0.07). Among all of the women in the study, those in Treatments 3 and 4 showed more concern for their babies' problems (p=0.05) and reported that their babies had better dispositions (p=0.04), while their babies were brought into the emergency room less often (p=0.04) and had fewer accidents and poisonings (p=0.03) (Olds, Henderson, Chamberlin, et al., 1986a).

Programs for the population at large, primary prevention, are not only expensive and difficult to implement, but they tend to be more superficial. The intensive, expensive secondary prevention programs are more likely to have a positive impact on families. However, it would be an inefficient use of resources to try to implement them on a global basis as primary prevention since most families do not need them. Before secondary prevention can be used effectively, high risk populations who will need the interventions must be identified. Altemeier and colleagues have summarized prevention as a three step process:

1) identification of risk factors.
2) identification of high risk families with the risk factors.
3) correction of deviant elements in the families that are at the root of the risk factors (Altemeier et al., 1979).

The question next arises as to when the optimal time is to screen families and intervene on behalf of the high risk

ones. Helfer has identified three times in the lives of parents when they are most accessible for mass screening: during schooling, during prenatal care or delivery of a child, and when a child is first entering the school system (Helfer, 1976b). Since 43% of maltreated children are 5 or younger, to wait until a child enters the school system means failing to prevent almost half of the cases. To intervene during a parent's schooling means to intervene on parents who have children at various ages, including school-age and pre-conception. To try to help people who have yet to start their families is of questionable effectiveness. The ideal time to screen families then is during the perinatal period (Lynch, Roberts, 1982). Concluded Gray and co-workers,

> "Perinatal assessment and early consistent intervention with families identified as high risk for abnormal parenting practices significantly improves the infants' chances of escaping serious physical injury (Gray et al., 1976)."

Prior Attempts to Predict Maltreatment

The success of prevention programs for a large part depends on proper screening and accurate identification of those families who are at high risk. Prior research indicate that the development of such an instrument is possible. One study of families who presented with young children to a pediatric emergency room found that "a simple, brief, objective assessment may be made in the emergency room setting to determine which patients are at increased risk for being

abused in the future (Rosenberg, Meyers, Shackleton, 1982)." Since an abusing parent has more incentive than a non-abusing parent to be less than truthful when questioned, data obtained from families before any adverse event occurs may be more valid than those obtained after maltreatment has begun, and prediction of those at risk might actually be more accurate than identification of those parents already maltreating their children but who as yet have not been identified (Altemeier et al., 1979).

With a general consensus that identification of those at high risk is possible, several attempts have been made in the past to develop such an instrument. Helfer outlined three typical methods: the self-administered questionnaire, the standardized interview, and observational checklists (Helfer, 1987). In reviewing past research on prediction, Leventhal found that the eleven studies he examined fell into one of four categories: a checklist of socioeconomic factors, a structured interview, unstructured clinical judgment and structured clinical judgment (Leventhal, 1988). These two classification schemes can be combined as follows: socioeconomic evaluation, the self-administered questionnaire, the semi-structured interview, unstructured clinical judgment, and structured clinical judgment.

In this combined classification scheme, the five categories are defined by two features: 1) by how the data are gathered (since this element can greatly limit the usefulness of an instrument) and 2) by how judgments of level of

risk are made.

Socioeconomic evaluations are those screening instruments whose contents are based solely on social and economic factors without regard to past experiences, psychological make-up, or attitudes. The data for these instruments are usually gathered in a questionnaire or as part of routine medical care. The judgments about level of risk are based on a specific set of criteria that define High Risk families as those with poor social supports and/or a lower economic class.

The self-administered questionnaire is a list of questions that parents are given and allowed to answer by themselves. Also included in this category are interviews where the interviewer asks a set list of questions and records the responses verbatim with no value judgments made about a patient's veracity and no prompting for responses; this type of structured interview could also be called an oral questionnaire and could be used for illiterate parents.

A semi-structured interview is defined as an evaluation that requires an evaluator (usually a social worker, physician, or nurse trained in evaluating high risk families) to conduct a special interview with the parents. This interview is not part of the family's routine medical care but is conducted specifically to assess the risk of child maltreatment. After the interview, which is open-ended as opposed to a questionnaire, the evaluator has a specific set of criteria to follow in assigning a level of risk.

The data for both structured and unstructured clinical judgments are gathered during a clinician's routine care of the patient. Thus, the evaluator must be a clinician (e.g. nurse, pediatrician, obstetrician/gynecologist, or social worker) caring for a member of the family; this clinician need not be specially trained to evaluate families for risk of maltreatment. No specific "child maltreatment" interview is conducted. Rather, all data are obtained during the course of a continuing medical relationship. In unstructured clinical judgment, the overall level of risk is based on the clinician's experience without specific guidelines to follow. In structured clinical judgment, the clinician must assess the family on specific items and follow established guidelines in scoring the items in order to determine a level of risk.

Socioeconomic evaluation

Although some research has discounted the value of basing a family's risk status solely on socioeconomic factors (Steinberg, Catalano, Dooley, 1981), Garbarino and Sherman did just that in proposing to identify families at high risk of maltreatment by the neighborhood in which they live. With the premise that child maltreatment is a social problem, whole neighborhoods (and the residing families) are classified as "'low risk' if they help support families, and 'high risk' if they work against families" (Table 1) (Garbarino, Sherman, 1980). Incidents of child abuse and neglect were plotted on a map to identify neighborhoods with a

high density of maltreatment events. Their research concluded that compared to Low Risk neighborhoods, High Risk neighborhoods were less tranquil, had lower rent, had no screens on the windows of homes, had a dichotomy between home ownership and home rental, were perceived by the residents to be filled with more unfriendly neighbors, had a higher involvement of families but a weaker family unit, and were experiencing a lot more change and deterioration (Garbarino, Sherman, 1980).

This form of screening is clearly not applicable to screening an inner city population since most families would be labeled as High Risk, although statistically and clinically, most of the families would not experience maltreatment. Furthermore, this technique categorizes people not individually but in large groups, ignoring individual variation and mitigating factors. This method also will misclassify as Low Risk every middle- or upper-class family at risk of maltreatment despite any existing family pathology. While this study may be interesting as descriptive research, its utility as a screening technique is minimal at best.

The self-administered questionnaire

The second method of assessing parents for risk of maltreatment is to ask the parents (usually the mother) to complete a questionnaire. This self-administered questionnaire technique has been employed several times by various researchers. Schneider, Hoffmeister, and Helfer used a 74-item questionnaire in the peripartum period. The questions

clustered into the six categories of self-esteem, social isolation, childhood experience, depression/crisis, expectations of children, and expectations of parenthood (Schneider, Hoffmeister, Helfer, 1976). Dean and others at 3 to 4 months postpartum used both a questionnaire that examined maternal attitudes along with an in-home visit by a social worker who gauged mother-child interactions (Dean, MacQueen, Mitchell et al., 1978). Gabinet administered the Minnesota Multiphasic Personality Inventory to try to outline a personality profile of the child abuser. In a comparison of High Risk parents (chosen by unstructured clinical judgment), known child abusers, and psychiatric outpatients with no past history of abusing, there was a striking similarity between the three groups leading the researcher to conclude, "There is no one abusive personality...(Child abuse is) predictable more from history and other samples of behavior rather than by personality testing (Gabinet, 1979)." The 160-item Child Abuse Potential Inventory developed by Milner, Gold, and Wimberly is designed not as a perinatal screening instrument of those at risk, but as a tool to identify those parents who have already abused their children (Milner, Gold, Wimberly, 1986).

Aside from the individual shortcomings discussed above, questionnaires have some intrinsic difficulties. Like interviews, they require the parent(s) to be cooperative and actively participating in the screening. Those with abusive tendencies, however, may be the ones least willing to be

screened. Questionnaires also offer the easiest opportunity for parents to lie; with a questionnaire, people have the time to carefully study the responses and mark the ones that they feel reflect good parents rather than those responses that reflect themselves. In addition, questionnaires cannot detect behaviors indicative of untruthfulness or violence which a one-to-one interaction between clinician and parent may detect. The value of a questionnaire is completely at the mercy of the willingness of the parent to reveal herself (or himself) on paper.

The semi-structured interview

Many different forms of the semi-structured interview have been developed. An early one was devised by Monaghan and Couper-Smartt in New Zealand and involved an in-depth interview of mothers conducted during the eight-day postpartum hospital stay. The interview was conducted by a pediatrician, a social worker, and a family psychiatrist, members of a Child Care Unit who specifically worked with women having trouble with parenting. The patients interviewed were those referred by their obstetrician after exhibiting some form of parenting distress, a subjective judgment on the obstetrician's part. The evaluation by the Child Care Unit included asking about the mother's own childhood, her social supports, and her expectations; a parent was judged to be at high risk of maltreating a child if she had two or more of the ten examined risk factors (Monaghan and Couper-Smartt, 1977).

While the researchers report that their criteria are "predictive," their method does not have widespread application. First, no maternity ward in the United States has the luxury of an eight-day postpartum observation period during which parenting and bonding can be gauged. At Yale-New Haven Hospital, the average postpartum stay is two days for vaginal births and five days for cesarean births. Second, the application of this interview required three highly trained family experts, expertise which all programs are not likely to have available to them. Third, this method requires that the patient be cooperative and willing to participate in an in-depth interview. Finally, Monaghan and Couper-Smartt's interview is not really a screening instrument. Their interview was not given to all parents (perhaps because of some of the drawbacks discussed above) but rather just to those already judged to be somehow troubled in their parenting. This judgment of troubled parenting in need of follow-up was done by obstetricians in an unstructured way and is really the initial screening step.

This approach to the interview technique was later modified by Monaghan, Gilmore, Muir, et al. They had a team of experts interview pregnant women and then complete a nine-item screening questionnaire afterwards (Table 1). In testing the instrument over a two year follow-up period, a negative outcome was measured as relinquishment of custody of the child, referral to the Department of Social Welfare, referral to Child Protection, or the involvement of family

court or a social worker. This study found a positive predictive accuracy of 90% for those labeled High Risk and 67% for those labeled Moderate Risk, and a negative predictive accuracy of 36% for those labeled Low Risk and 88% for those labeled No Risk (Monaghan, Gilmore, Muir, et al., 1986).

Monaghan et al.'s instrument suffers from the same problem of needing a team of experts to evaluate the patient and a patient willing to be interviewed. There also was poor discrimination between the Moderate and Low risk categories; 67% of the former and 64% of the latter went on to have a negative outcome. These rates may be even higher if, as often happens, not all cases of maltreatment were reported to agencies. Finally, this instrument was developed using a population in New Zealand that is "economically advantaged and has well-developed health and social services" (Monaghan, Gilmore, Muir, et al., 1986). While the screening instrument may identify those at high risk of parenting difficulty in this middle- to upper-class population with socialized medicine, it is not safe to extrapolate the conclusions to an inner city population in the U.S. that has limited access to health care and social services. Identifiable risk factors in one population may not be the same in the second population.

Altemeier et al. developed a 45-minute interview designed to be given to mothers in the prenatal period. This interview examined eight areas including the mother's own nurture as a child, her feelings about the pregnancy, and

any substance abuse problems (Altemeier, Vietze, Sherrod, et al., 1979). This format was later revised to a 35 minute interview with the data clustering into six predictive areas: subjective impression of the interviewer (most predictive), residency transience, untruthfulness, disturbed childhood nurturance, unwanted pregnancy, and increased parent-child exposure (Table 1). When tested on 1400 pregnant women, Altemeier and colleagues found that they had correctly predicted 53% of the abused children with a 94% false positive rate and with prediction good for up to 24 months (Altemeier, O'Connor, Vietze, et al., 1984). This method uses a long interview that is not practical for mass screening; it was developed more with research than screening in mind. The results also yielded an unacceptably high false positive rate. It is of note, though, that the most predictive category was the unstructured judgment of the interviewer, raising the possibility of using that alone as the screening tool.

At the University of Colorado, Murphy, Orkow and Nicola tried screening pregnant women using the Family Stress Checklist originally developed by Schmitt and Carroll (Table 1). This interview was administered by an experienced social worker and had a positive predictive accuracy of 52.6% and a negative predictive accuracy of 96.6% (Murphy, Orkow, Nicola, 1985). The true merits of this checklist are hard to discern, however. Those judged to be High Risk received interventions such as parenting classes during the follow-up

period (Orkow, 1985). Thus, it is possible that if no intervention had been implemented, the positive predictive accuracy may have been higher. On the other hand, the definition of child abuse and neglect used in the follow-up included such criteria as cradle cap and diarrhea which may have incorrectly categorized some well-cared for children as maltreated (Murphy, Orkow, Nicola). A final pitfall of this instrument is that, again, it requires a trained social worker to administer it.

Avison, Turner, and Noh developed a 20-question screening interview for mothers that looked for "parental maladaptation" (Table 1) (Avison, Turner, Noh, 1986). The questions looked at social supports and parenting attitudes. When tested on a group of 87 known maltreaters (maladaptive mothers) and 100 controls (presumed to be well adapted), the interview correctly identified 96% of the maladapted mothers and 90% of the comparison mothers for an overall accuracy of 93%. However, a test of the predictive validity of this instrument was not conducted for "severe ethical and practical difficulties" and the authors go on to warn that this instrument "cannot be regarded as a clinical or diagnostic instrument or used for such purposes" (Avison, Turner, Noh, 1986).

Unstructured clinical judgment

Unstructured clinical judgment goes on daily when physicians and nurses refer families to any type of social work evaluation or intervention because of their own sense, based

on experience, that a family is in need of help. Leventhal, Garber, and Brady studied such judgments at Yale-New Haven Hospital by doing a retrospective, longitudinal cohort study of babies referred at birth (based on unstructured clinical judgments) by pediatricians, nurses, and social workers to the hospital's child abuse committee, known as the DART (Detection, Assessment, Reporting, Treatment) Committee. Compared to the matched control group, by the fourth birthday more of the referred children experienced actual child maltreatment than the control group (23% vs. 8% with a matched odds ratio of 3.1) (Leventhal, Garber, Brady, 1989). While unstructured clinical judgment is the most common method of screening, and is actually a good method, it relies heavily on a clinician's experience and intuition. Younger, less experienced clinicians are less likely to do as well in detecting which families are at risk.

Structured clinical judgment

To capitalize and improve on the effectiveness of unstructured judgments, several researchers have tried to create instruments that structure clinical judgment. This approach benefits from the clinician's personal interactions with the family yet guides the clinician as to which areas of family life to pay particular attention. Rosenberg et al. used this approach in the emergency room to evaluate all children younger than 24 months who came in and their families. An 8-item evaluation to be used by nurses examined the state of the child's care, the parents' behaviors, and

the quality of parent-child interactions with a 42% rate of later maltreatment in the High Risk group (Rosenberg, Meyers, Shackleton, 1982). Since this instrument looks for signs of past maltreatment (bruises, poor child care) it is more a tool for tertiary prevention, rather than secondary prevention. Nonetheless, it is one successful application of the principles of structuring clinical judgment.

Murphy et al. and Gray et al. each developed predictive instruments that structure clinical information in the postpartum period. The instrument by Murphy and colleagues is based on a review of objective information coded by ward clerks into the birth records of all children and is composed of 11 factors, such as social class, age of the mother, and birth weight of the infant, found by the researchers to be correlated with later abuse (Table 1) (Murphy, Jenkins, Newcombe, et al., 1981). While this instrument screens families in the ideal postpartum period, it does not make use of clinical interactions or impressions but is wholly reliant on objective biographical data, thereby not maximizing the full potential of clinical judgment. Gray and colleagues began to develop a screening instrument that also evaluated families in the postpartum period but utilized information on the families' interactions with hospital staff. This instrument included such items as disappointment/pleasure with the baby's gender, parental reactions to infant crying, parental in-hospital care of the baby, and attempted manipulation of the staff by the parents

(Gray, Cutler, Dean, et al., 1976). While this instrument appeared very promising, it was never fully developed or tested (written communication, Aug. 7, 1989).

Lealman, Haigh, Phillips, et al. took a similar approach to Gray and devised a checklist of behaviors and characteristics of parents to screen for risk of later maltreatment. This checklist was composed of 10 items available from maternity notes and was applied during the discharge of mothers from a maternity ward. Risk factors included maternal age, prenatal care, and marital status (Table 1) (Lealman, Haigh, Phillips, et al., 1983). This instrument was never properly tested, either. It was developed and used as part of an prevention program; thus, most of those identified as High Risk received some sort of intervention during the follow-up period, making it impossible to calculate the true sensitivity, specificity, and positive predictive accuracy of the scale.

In their research into risk factors for child maltreatment, Browne and Saqi concluded, "Our findings suggest that family stress is not a sufficient explanation for child abuse (Browne, Saqi, 1988)." They constructed a 12-item checklist designed to be used by nurses after interviewing the mother during the postpartum period to identify those at high risk of later maltreating their newborn baby. This instrument included parental age, history of family violence, and financial problems in the family and was designed with the population of Surrey, England in mind (Table 1)

(Browne, Lowton, 1987). A study of the predictive validity of the scale on 14,283 newborns and their families yielded a sensitivity of 81%, a specificity of 94% but a positive predictive accuracy of only 7% (Browne, Saqi, 1988). The poor predictive accuracy of the instrument may have been due in part to the low prevalence of child maltreatment, but also may have resulted from the inclusion of some questionable risk factors in the scale like bottle (rather than breast) feeding, prematurity (Leventhal, Egerter, Murphy, 1984), and socioeconomic status (Steinberg, Catalano, Dooley, 1981; Egeland, 1979; Daro, 1988) and the exclusion of risk factors felt to be strong predictors, like psychiatric stress (Krugman, Lenherr, Betz, et al., 1986) and whether the pregnancy was planned (Lealman, Haigh, Phillips, et al., 1983; Lynch, 1976). This instrument represents a good beginning in the structuring of clinical judgment, but may benefit in terms of improved positive predictive accuracy if the items were refined and expanded.

Prediction of those at risk of future maltreatment is an important first step in secondary prevention. However, a good scale for screening purposes has yet to be developed. Wrote Rosenberg et al.:

> Most methods of detecting child abuse are very time consuming and retrospective in nature. A good screening procedure has not yet been devised that will enable busy pediatric facilities to detect children at special risk for abuse (Rosenberg, Meyers, Shackleton, 1982).

The ideal time to screen families is during the postpartum period, before any adverse event can occur and before the families feel they have anything to hide. Lynch and Roberts retrospectively reviewed the birth records of abused and non-abused children and identified five risk factors present at birth with five times as much frequency in children later abused as compared to those not abused (Table 1). From this they concluded that families at high risk can be identified with the data readily available and routinely collected on postpartum hospital wards (Lynch, Roberts, 1978). The work of Leventhal, Garber, and Brady also supports the idea that screening in the postpartum period can be done successfully (Leventhal, Garber, Brady, 1989).

The ideal method to screen families is through the use of structured clinical judgment based on the routine interactions of patients and clinicians. As discussed above, evaluations based on specially designed interviews are hindered by their length, the difficulty in gaining access to all parents, the difficulty in gaining the participation of all parents (especially those at high risk), the necessity of training many experts to conduct the interviews, and the limitations on screening every parent based on cost and available labor. Self-administered questionnaires also require active parental participation that may be hard to obtain and more readily lend themselves to deception on the part of the parents. The validity of socioeconomic scales is in dispute; inherently they cannot screen well a low-

income inner-city population. Finally, unstructured clinical judgments have been shown to be the most common and convenient method to use, yet they are in need of improvement so that less experienced clinicians can also predict accurately. Thus, a screening instrument based on structured clinical judgment that focused the clinician's attention on valid, readily decipherable risk factors would be the ideal instrument.

Until now, a screening instrument based on structured clinical judgment designed for use in the postpartum period had yet to be developed and tested.

Clinimetrics and the Development of Rating Scales

A screening instrument for risk of child maltreatment based on structured clinical judgment would fall into the category of clinical rating scales. "A useful clinical scale," point out Hutchinson and colleagues, "must fulfill the basic scientific requirements of a measurement yielding results that are reproducible and valid (Hutchinson, Boyd, Feinstein, et al., 1979)." Unlike scales for laboratory measurements and other "hard" data, the measurement of so-called "soft" data has not enjoyed enough scientific scrutiny to lead to the establishment of rigorous criteria for their construction and use. Hutchinson feels:

> (This is due to the) belief that the subjective information required to assess function is too unreliable to merit serious scientific consideration....No general standards have been established

to appraise rating scales for clinical phenomena (Hutchinson, Boyd, Feinstein, et al., 1979).

Feinstein has made major strides in the field by defining the issues and suggesting preliminary guidelines. He proposes the term "clinimetrics" for "the measurement of clinical phenomena," subdividing the activity into mensuration (the acquisition and labeling of data) and quantification (clinical epidemiology) (Feinstein, 1987). The development of a clinical rating scale would fall under the mensuration subcategory of clinimetrics. Feinstein points out that even though formal standards have been lacking, clinical observations have long been recorded, described, categorized, and rated (Feinstein, 1987).

Since clinical observations are more complex than "hard" data like laboratory values, their measurement is more complex, the method used often depending on the purpose of the measurement. A rating scale can indicate presence/absence, magnitude such as none/mild/moderate/severe, or more complex descriptions such as tumor/node/metastasis. The function of a clinimetric scale has been divided into four general types by Feinstein. The first is to describe the status of a disease; this would include diagnostic criteria and ratings of clinical conditions. The second function is to measure change in a disease, sign, or symptom. The third function is to describe prognosis, and the fourth function is to describe a treatment protocol (Feinstein, 1987).

After clarifying the function of the clinimetric instrument, the outline for the development of a clinimetric scale as proposed by Feinstein is as follows. First, possible variables to be included on the scale are selected; some are retained and some are excluded. Each component variable retained is then described in its own scale. These component variables must then be combined to produce an output (a score for the entire instrument) with its own output scale.

There are two basic methods to selecting the component variables for a clinimetric scale. One is to examine what is to be measured and then gather the intrinsic and extrinsic evidence that the clinical phenomenon exists. Intrinsic evidence would describe the phenomenon, such as the existence of enlarged nodes or a symptom like shortness of breath. Extrinsic evidence is a result of the phenomenon, such as the ability to care for oneself. The second method to selecting component variables is to review past research in the field and gather those variables felt to be significant markers for the phenomenon under study (Feinstein, 1987).

In clinimetrics, the original scales for the individual variables are often ordinal (ranked but without equal magnitude between adjacent ranks) rather than dimensional (each rank is of the same magnitude) since clinical phenomena are often with nondimensional outcomes (e.g. the difference between severe and moderate pain may not be the same as the

difference between moderate and mild pain). When ordinal scales are combined to form the outcome scale, the result is what Feinstein calls quasi-dimensional, an ordinal scale with unequal ranks yet with the illusion of having equal magnitude between ranks (Feinstein, 1987).

Guidelines do exist for the use of ordinal scales for clinical measurements. First, the elements of the scale must be clearly defined and mutually exclusive; any event to be measured on the scale must fit into one and only one rank on the scale. Second, the ranks on the scale must exist in a hierarchical order; if not, the scale is nominal, not ordinal. Third, the scale must be constructed in a meaningful way so that a change in rank on the scale reflects a clinically meaningful difference. Fourth, the scale must be symmetrical so that improvement and deterioration can both be measured. Finally, if other related measured outcomes exist, the scale must produce a result consonant with the other outcomes. In addition, it is important to know the clinical significance of differences in scores and the expected variation in scores (MacKenzie, Charlson, 1986).

There are several methods to combine component variables into an outcome measure. One is to sum the variables. Although the easiest method, this is not always the best. As Browne and Saqi point out:

> Unfortunately, most screening procedures using a checklist format add the number of risk factors present and obtain a simple summation score that,

> in effect, treats all risk factors the same.... This is, of course, illogical and limits the usefulness of the checklist (Browne, Saqi, 1988).

Another method is to use Boolean clusters; outcome categories are made up of different combinations of each of the component variables. A tandem profile that lists the result of each component variable in the outcome variable (such as in the TNM staging system) is another possible method. Finally, a hierarchical system can be used where each component variable is analyzed individually in a specified order. If any variable exceeds a specified cut-off then an extreme is reached such that the remaining variables can be ignored. This is how cancer staging works where metastases are evaluated first, then nodes, then tissue pathology (Feinstein, 1987).

"Despite the general scientific prejudice against soft data," writes Feinstein, "clinicians (have) often gone ahead and created indexes for the soft phenomena investigated in their research (Feinstein, 1987)." Classic examples include the Apgar score, the Glascow Coma Scale, the Trauma Index, the Yale Observation Scale, and the Dubowitz score for gestational age.

Apgar's scale for describing the condition of a newborn was an early clinimetric scale. Her scale observes a newborn at one minute and five minutes of life and rates the baby on five component variables, each of which has an individual three point ordinal scale. The component variables were selected from Apgar's observations of intrinsic evi-

dence of newborn condition and were those she felt were easy to determine and which she "considered useful (Apgar, 1953)." The components are combined by direct summation. Although Apgar found one component variable, heart rate, more prognostic of how a newborn will fare, she chose to keep her scale simple and not weight the variables (Apgar, 1953).

The Glascow Coma Scale is another instrument used to describe a patient's condition, in this case, level of consciousness. There are three component variables each of which is recorded on an ordinal scale. However, the component scales have different maximums reflecting relative weighting being given to the variables (Teasdale, Jennett, 1974). Although the "motor" variable measuring movement is alone considered the best indicator of level of consciousness (Jagger, Jane, Rimel, 1983), the Glascow Coma Scale incorporates two other variables since the motor variable cannot always be measured (e.g. if the patient is in traction). The output score is obtained by a summation of the component variables. This scale is an excellent example of how structured clinical judgment can produce more inter-rater agreement than simple unstructured clinical judgment (Teasdale, Jennett, 1974). By offering a standard way to describe a patient's condition, the scale allows doctors to avoid "ambiguities and misunderstandings (that result) when groups of patients treated by alternative methods are compared, or reported from different centres (Teasdale, Jen-

nett, 1974)."

The Trauma Index by Kirkpatrick and Youmans is another screening instrument designed to describe a patient's condition. The variables were selected after a review of those items that clinicians have in the past cited in their hospital notes when assessing a patient's level of trauma. Those component variables easy to assess were retained, and the composite score was derived from the summation of the individual variables (Kirkpatrick, Youmans, 1971). This is the same approach taken by McCarthy in developing the Yale Observation Scale for degree of illness in the febrile child. McCarthy and colleagues initially selected their component variables from those which experienced pediatricians stated they used in evaluating a child with a fever. After testing, only those which were "independent and significant predictors" of serious illness based on multiple regression analysis were retained (McCarthy, Sharpe, Spiesel, et al., 1982). The output variable is again a summation of the component variables. This method, concluded McCarthy and colleagues, "can be used to study clinical judgment in other areas of pediatrics (McCarthy, Sharpe, Spiesel, et al., 1981)."

An interesting rating scale is the Dubowitz score for gestational age as modified by Sweet. This instrument assesses 10 neurologic features and 11 physical features in newborns less than 24 hours old. The summary score is an estimate of the newborn's gestational age (Dubowitz, Dubo-

witz, Goldberg, 1970; Sweet, 1979). This clinimetric scale is of particular interest because unlike many other scales of clinical judgment, there are "hard" data with which to compare the result. When compared to gestational dating by the mother's last menstrual period, Dubowitz found a correlation coefficient of 0.93 with a 95 per cent confidence limit of ± 2 weeks (Dubowitz, Dubowitz, Goldberg, 1970). When 2 independent evaluations of the same baby were compared, the 95 per cent confidence limit of the average score was ± 1.4 weeks. This study is important in highlighting the point that although considered "soft" data, clinical judgment can be measured in a reliable, reproducible fashion.

In order to develop a screening instrument to assess a newborn's risk of subsequent maltreatment, the principles of clinimetrics discussed above must be followed. The function of this instrument will be secondary prevention, to screen the population to identify those at high risk. The component variables will be selected by a combination of selecting those intrinsic and extrinsic factors felt to be important by clinicians and by a critical review of the research literature on risk factors associated with child maltreatment. The next step will be to make individual scales for each component variable and then to combine the variables into an output score. This process will be discussed in the next section.

CHAPTER THREE:

METHODS

This research was reviewed and approved as Protocol #5157 by the Human Investigation Committee of Yale University School of Medicine and Yale-New Haven Hospital. The data were analyzed on an IBM PS-2 using the SAS statistical program. The weighted kappa values were computed using the RATCATA Computer Program for Assessing Rater Agreement and Bias from Contingency Tables.

The development and testing of a new instrument using structured clinical judgment to screen for risk of future maltreatment was conducted in 4 phases. Phase I was the selection of the component variables and the construction of the clinimetric instrument. Phase II was the piloting of the instrument on 176 newborns in order to refine the variables and individual scales. Phase III involved the use of the instrument by both experts and non-experts in the field of child maltreatment. In addition, the experts gave their overall assessments of each subject's risk. From these data, a scoring method was developed, the agreements among expert clinicians in both their structured (CRS) and unstructured (GRS) clinical judgments were examined, and a test of concurrent validity was conducted. The final phase, Phase IV, was the evaluation of 363 consecutive newborns by non-expert clinicians using the new instrument. From these data, inter-rater agreement on each item in the instrument was assessed, and patterns of responses to detect which items on the instrument were difficult to rate were studied.

This phase also began the prospective validation of the instrument.

Phase I: Development of the Screening Instrument

The function of this instrument is to screen newborns during the postpartum hospital stay for risk of child maltreatment in the future. This instrument is not designed specifically as a research tool to be used on a select population. Rather, once fully tested, it is hoped that this instrument will be used on every newborn as a first step in screening; those identified as at high risk would then receive further in-depth evaluation by Social Services to more completely determine the type and extent of preventive services required. With this in mind, the desired attributes of the new instrument include:

1) It should be used during the postpartum period. As discussed in the last chapter, this is an ideal time since the family is available for evaluation, it is before any possible form of maltreatment could have occurred to the newborn, it is a time of stress for the family so the parents' reactions to stress can be evaluated, and evaluations made at this time have been shown to be predictive by Gray, Cutler, Dean, et al. (1976).
2) It should be simple and efficient to use so that time constraints will not prohibit clinicians from evaluating all patients.

3) It should not require a special, formal interview or self-administered questionnaire. Rather, the clinician should be able to gather the necessary data during the course of regular patient care. This will eliminate both the problem of uncooperative parents, a subpopulation likely to have many high risk families, along with the problem for clinicians of finding the time to do an extra 30 to 60 minute interview with each patient. This requirement would also minimize the risk of parents deliberately lying in response to questions they know are part of an evaluation of them as parents.

4) It should be fairly simple to apply and score. If specially trained staff are required to implement the screening instrument, that would be another obstacle, both in terms of manpower and cost, to its widespread use.

5) Finally, the instrument should take the form of structured clinical judgment. This format would help fulfill the preceding criteria. This instrument could be used in the postpartum period since obstetricians, nurses, pediatricians, and social workers see the family often at that time. No special, time consuming interview would be required. Rather, the clinician's response to each item would be based on observations and data normally gathered regularly during the hospital stay. Also, since the evaluation is based on the entire stay, there is less of a chance that families

will be able to lie or cover-up risk factors.

Construction of the instrument

A prototype method of how to develop an observational scale was demonstrated by McCarthy in the development of the Yale Observation Scale for determining level of illness in the febrile child. In McCarthy's study, 262 febrile children were observed by pediatricians, pediatric residents, and nurses. The observers listed those observations that he/she felt were important in making an overall assessment of the child's severity of illness. The most frequently mentioned variables were selected and from those 20, six were found by multiple regression analysis to be "independent predictors of serious illness (McCarthy, Sharpe, Spiesel, et al., 1982)." These six variables were then combined into one instrument (McCarthy, Jekel, Stashwick, et al., 1981; McCarthy, Sharpe, Spiesel, et al., 1982).

McCarthy's model of instrument development was followed in the development of a screening instrument for risk of child maltreatment. A study of residents at Yale-New Haven Hospital (Y-NHH) by Leventhal, Fearn, and Stashwick found that pediatric residents relied on observations more than the medical interview to judge quality of parenting. Important variables that went into making this judgment included how the mother uses the medical system; observations of the mother-child interaction, the mother, and the child; and information from the medical history (Leventhal, Fearn, Stashwick, 1986). Altemeier concurs that a clinician's sub-

jective impression is a very good predictor of who will or will not maltreatment his/her child. Altemeier writes:

> A major question in our minds is whether...subjective impressions could be used in objective, consistent, and reproducible fashion as part of a second generation interview: Their effectiveness when listed as specific observations rather than open-ended impressions remains to be determined experimentally (Altemeier, O'Connor, Vietze, et al., 1984).

The original Leventhal study was later expanded. Leventhal, Garber, and Brady reviewed the records of every infant referred during the postpartum period to the hospital's child maltreatment (DART-Detection/Assessment/Reporting/Treatment) committee in order to study whether those judged to be at high risk by unstructured clinical judgment were subsequently maltreated more often than a low risk non-referred control group. In this study, a list was compiled of the reasons for referral stated by clinicians in their referrals of newborns to the DART committee (Table 2) (Leventhal, Garber, Brady, 1989). Since Leventhal's earlier research found that unstructured clinical judgment could identify those later maltreated, the list of reasons for referrals of newborns was used as the initial set of risk factors to be included in the new screening instrument developed in this current study.

A similar list of variables used in making unstructured clinical judgments about quality of parenting was compiled by Ounsted and colleagues (Table 3). This list also is a mixture of observational variables and variables from the

medical history. Like Leventhal and Garber, Ounsted included maternal behaviors in the hospital, maternal attitude, previous abuse of children, and family conflict issues (Ounsted, Roberts, Gordon, et al., 1982).

An alternative method to use in gathering variables to include in a screening instrument would have been to select those variables shown in previous research to be independent predictors of child maltreatment. The major problem with this method, however, is that a valid profile of the child abuser has yet to be constructed (Starr, 1987). In evaluating the literature on risk factors for child maltreatment, it is clear that there is yet to be overwhelming agreement about which variables are definite risk factors (Table 4). Several possible reasons for the conflicting results of the various studies include differences in how the outcome (child maltreatment) was measured, the use of different types of control groups (including the use of unmatched control groups), and detection bias (Leventhal, Egerter, Murphy, 1984).

Because of the conflicting results among the various studies, the current development of a child maltreatment screening instrument started with McCarthy's method of including only those variables actually used by clinicians. The risk factors considered for the instrument were those variables identified in the Leventhal, Garber, and Brady study as being used by clinicians in their unstructured judgments of risk. In addition, to be included in the in-

strument, rigorous prior research must support a variable as a risk factor for future maltreatment. Further, in the process of testing the new screening instrument, only those variables shown to be obtainable and predictive of maltreatment were to be retained.

Version 1 of the newly developed screening instrument (the Clinical Rating Scale, CRS) is shown in Figure 1. Each of the 22 variables was derived from the Leventhal, Garber and Brady study and is expressed in a dichotomous-ordinal scale ranging from three to five categories. Category 1 for each variable represents "good" parenting while categories 2 through 5 represent worsening degrees of "bad" parenting. Also included for each item is a box to indicate an "unknown" response.

The transformation of the list of variables in Table 2 into the Clinical Rating Scale instrument in Figure 1 was accomplished after a careful review of the literature. How each variable has been described in the past (the range of behaviors or traits), and the associated degree of risk of maltreatment (low, mild, moderate, high) for each specific behavior or trait were studied. Each variable was then operationalized into a dichotomous-ordinal scale. All no risk behaviors or traits were grouped into category 1; thus, category 1 versus all other categories created the dichotomous (Absence of Risk Factor versus Presence of Risk Factor) nature of the scale. The low to high risk behaviors or traits were then expressed in an ordinal scale ranging from

category 2 to category 5. All variables did not readily fall into a 5-point scale. Those with only 3 or 4 categories thus had an empty category somewhere on the scale.

A Global Rating Scale (GRS) was also constructed (Figure 3). This scale measures a clinician's unstructured clinical judgment of a family's risk of subsequent maltreatment. This scale was developed in order to be able to compare clinicians' unstructured clinical judgments to their structured clinical judgments (using the Clinical Rating Scale) of the same families and to test whether the structured clinical judgment is an improvement in predictive ability over the unstructured judgment.

Phase II: Pilot Testing of the Instrument

The next step in development was to pilot test Version 1 of the newly developed Clinical Rating Scale.

Subjects

There were two classes of subjects in this study. The first class consisted of the clinicians (nurses, pediatric residents, and social workers) who cared for postpartum women and newborns. These clinicians were asked to evaluate the families of newborns using both the Clinical Rating Scale and the Global Rating Scale. The clinicians were considered subjects since their abilities to evaluate families were examined in this study. The criteria for inclusion were that the clinician a) must have been a registered nurse, a pediatric resident, or a social worker b) must have

regularly worked on a postpartum floor and c) must have cared for one of the study's enrolled newborns or newborn families for the duration of their hospital stay. Most families in the study were evaluated by the mother's postpartum nurse and the pediatric resident who cared for the baby in the well-baby nursery. Only those families who were visited by a social worker as part of their medical care (about 30% of all postpartum mothers who receive care at Yale-New Haven Hospital's Women's Center are seen by a social worker at the request of a nurse, obstetrician, or pediatrician) were also evaluated for this study by a social worker. Those babies who were cared for in the Newborn Intensive Care Unit and were never transferred to the well-baby nursery prior to discharge did not receive a pediatric evaluation because of time constraints on these residents. This amounted to less than 3% of all births. Consent for participation in the study was obtained orally from the clinicians on a case by case basis.

The second class of subjects in this study consisted of the families of newborns who were evaluated for risk of child maltreatment. Inclusion criteria were that the families must a) have delivered a viable infant at Yale-New Haven Hospital b) have stayed for more than 48 hours on the postpartum floor (those mothers who went to a surgical or intensive care floor after delivery were excluded) c) have spoken English fluently d) have planned to bring their newborn to Yale-New Haven Hospital's Primary Care Center for

routine pediatric care (this would allow the medical records of the newborns to be reviewed in the future to determine if child maltreatment had occurred).

Families were enrolled in the study if after review of the mother's and newborn's current hospital chart it was determined that all of the inclusion criteria were met. Consent was not obtained for the following reason. Since 1967, clinicians at Yale-New Haven Hospital have been obligated to report those children they feel are at high risk of subsequent maltreatment to the hospital's child abuse committee. This is routinely done using unstructured clinical judgments without the parents' knowledge or consent. This research did not alter this procedure nor subject the children or their parents to more scrutiny than they would have normally experienced. In fact, all information obtained during the study was strictly confidential. The responsibility to report children at risk remained with the clinicians; none of the information obtained in this study was used to register children with the child abuse committee.

Revision of the instrument

In the pilot, 176 families were evaluated by their clinicians. For each family, the mother's nurse, the pediatric resident, and the social worker (if there was one) were asked to complete the Clinical Rating Scale and the Global Rating Scale based on information they had gathered during the course of their care of the mother and infant;

the clinicians were asked not to conduct a special interview of the family. In this way, the testing of the instrument would most simulate how the instrument would be used if incorporated into routine postpartum care. Stapled to the end of each form was a sheet asking the clinicians for their feedback on: a) the variables selected (Were there some they felt they could never answer? Were there variables they felt should be added in?) b) the individual rating scale for each of the 22 variables (Were the examples given for each rating appropriate and clear? Were there better examples? Were the categories helpful?) c) how the entire Clinical Rating Scale was constructed d) any other suggestions.

During the course of the piloting, the Global Rating Scale was revised once (Figure 4). This revision softened the tone of the examples given for each rating. In the first version, clinicians were hesitant to give a family a rating of four, even if they felt the family was at very high risk, because the tone in version 1 implied more certainty than most clinicians were willing to ascribe to their judgments.

The Clinical Rating Scale underwent five revisions. For the most part, the revisions were rewordings of titles or of the examples in order to clarify them. The section on parental background (items 2-9) was changed to refer only to the background of the mother. The item on maternal childhood was divided into two items, one on childhood stability

and one on presence or absence of maltreatment. One question asking whether the father has any of the risk factors listed earlier on the instrument was added; since the clinicians were able to gather little information on the father, they felt that one general question was sufficient. If more specific questions were asked, they would have been marked "unknown" over 80% of the time. The item about the mother's cooperativeness in the hospital was combined with the item on the mother's threatening behavior. A shortened version of the Global Rating Scale was added to the end of the Clinical Rating Scale. By putting both scales on the same sheet the visual presentation of the study was improved and thus less intimidating to the clinicians; presented with fewer papers, clinicians were more apt to complete and return the forms.

The major revision of the Clinical Rating Scale was the conversion of all the individual scales to a four-point scale. The original format that included blank squares was confusing with some clinicians inappropriately marking the blank areas. The original format also contained a weighting to the categories. A variable that only had three categories was given less weight in the scoring (since the highest category was a three) than a variable that had five categories. Some items with three categories had blank squares in the middle of the scale in order to increase the scoring of the last categories. These weightings were originally made without rigorous scientific support. To correct this, all

variables were changed to a four-point scale with equal weighting and no blank squares. If analysis of the follow-up data reveals that certain items are more predictive than others, appropriate weights will then be added into the scoring. Version 6 of the Clinical Rating Scale (the final version) is shown in Figure 2.

Phase III: Tests of Inter-Expert Agreement and Concurrent Validity

Since child maltreatment is the result of a combination of personality, environmental, and situational factors, all of which are undergoing constant change, it is impossible to predict with 100% certainty who will and who will not maltreat their children. There is no absolute gold standard of prediction. However, as has been discussed earlier, it is possible to identify those who are at high risk of maltreatment. The most predictive evaluation would be expected from those experienced in working with families where maltreatment has occurred. By combining the evaluations of two or more such experts, a consensus evaluation can be derived and used as a gold standard. This method of establishing a reference standard is called consensual (or concurrent) validation (Feinstein, 1987).

The purpose of the test of concurrent validity was to compare non-experts using the Clinical Rating Scale to the gold standard (the level of risk of child maltreatment assigned by experts using unstructured clinical judgment,

their GRS) in their abilities to differentiate high risk from low risk newborns. In this part of the study, the agreement between pairs of expert clinicians was also studied. The experts consisted of two pediatricians and one social worker, each of whom has conducted research in the area of child maltreatment and has cared for families that have experienced maltreatment.

The weighted kappa statistic was planned to be used to compare item by item agreements in risk ratings between pairs of experts. Cicchetti has shown that the minimum sample size needed in order to use kappa can be estimated by $2k^2$ where k is the number of categories in a scale (Cicchetti, Sparrow, 1981). Since the Clinical Rating Scale has 4 categories for each individual scale, the minimum sample size needed for this part of the study is approximately 32. Thirty-two interviews were done by the experts, eleven by each of two possible pairings of the three experts, and ten by the third pairing. Thus, each of the three experts was involved in the evaluation of 21 or 22 families.

For this phase of the study, the experts evaluated families already enrolled for Phase IV of the study. So in addition to the experts' evaluations, each postpartum mother's nurse, newborn's pediatrician, and family's social worker (if assigned) were also asked to complete the CRS. The families to be interviewed were preselected to include an even balance of those determined by the non-expert clinicians to be high risk and low risk. Since the overall na-

tional incidence of child maltreatment is estimated at less than 10%, this was necessary to avoid interviewing only low risk families. The experts were unaware of the risk status assigned by the non-expert clinicians.

To conduct this phase of the study, oral consent to participate was obtained from each mother interviewed. The experts explained the purpose of the study, gave the mother an information sheet that also explained the study, and then obtained oral consent prior to beginning the interview. In each case, the mother was interviewed by one of the experts while the second expert observed. Each interview lasted from 15 to 30 minutes and, making certain to cover the 22 items on the CRS, consisted of the medical history and family history that the experts normally employ when evaluating families in clinic. After the interview, the experts reviewed the mother's prenatal chart and hospital chart, and spoke to the mother's nurse. Then each separately rated the family using version 6 of the Clinical Rating Scale and using their unstructured clinical judgments. For their unstructured judgments, the experts agreed to the definition of each risk rating as in Figure 5. These definitions are comparable to version 2 of the GRS.

Development of a scoring method

To develop a scoring method, the instrument will have to be tested on a large population of newborns. Once the outcomes for the children are known (abuse or non-abuse), the instrument can undergo multiple regression analysis to

determine appropriate weights for each item on the instrument (Feinstein, 1985, Ch. 10), and the instrument can be analyzed to determine the most sensitive and specific cut-off score. The initial work of evaluating a large newborn population with the instrument has begun (see Phase IV below). In the meantime, a preliminary scoring method was developed from the evaluations completed by the experts in child maltreatment. Five possible scoring methods were considered:

1) The arithmetic sum. Each item on the instrument was given a value of 0 if "unknown" was marked or a response was omitted, 1 if category 1 (no risk trait) was marked, 2 if category 2 (low risk trait) was marked, 3 if category 3 (moderate risk trait) was marked, or 4 if category 4 (highest risk trait) was marked. The sum of each of the 22 items was used as the score. The higher the score, the more at risk the subject. This method was considered the simplest and a good starting place.

2) The average score. This score is the result of the arithmetic sum (method 1, above) divided by the number of questions answered (that is, those questions where the response was not omitted and was not "unknown"). This method takes into account those items left unmarked or marked as "unknown." Method 1 would give a higher score to an evaluation where all 22 items were marked 2 (low risk) than to an evaluation where 20 items were marked 2 and two were marked "unknown."

Method 2 would give both evaluations the same score, thereby not "penalizing" an evaluation for having fewer unknowns.

3) The frequency of items given a value of 2, 3, or 4. This method simply counts the number of items where a value of 2 or greater was assigned. In the dichotomous-ordinal scale used in the instrument, category 1 indicates "no risk" behaviors or traits while categories 2, 3, and 4 indicate "at risk" behaviors with a progression from "low risk" in category 2 to "high risk" in category 4. This method counts the frequency of "at risk" behaviors or traits. This method also takes into account items left blank or marked "unknown." By method 2, an evaluation with 1 item marked in category 4 and 22 items marked in category 1 has a lower score than an evaluation with 1 item in category 4, 19 items in category 1, and 3 items marked unknown. Method 3 would give each of the two evalua tions the same score.

4) The number of items given a value of 3 or 4. This method counts the number of items where a value of 3 or 4, the higher risk behavior or traits, were assigned. This method is very similar to method 3 but only counts the more "at risk" categories.

5) The arithmetic sum of those items marked 2 or higher. Like method 1, this method adds the values assigned to each item, but only if the value is 2 or higher. This

method only counts behaviors or traits marked as "at risk." As opposed to method 3, this method weights the severity of the value given to each item. By method 3, an evaluation with 1 item marked in category 4, 1 item marked in category 2, and 21 items marked in category 1 has the same score as an evaluation with 2 items marked in category 2 and 21 items marked in category 1. By method 5, the first evaluation would have the higher score since it contained a more severe rating of an item.

For each of the five possible scoring methods, a histogram was made of each expert's score on the Clinical Rating Scale compared to the same expert's assessment using the Global Rating Scale (Figures 6-10). Based on these graphs, it was apparent that Method 5 produced the most differentiation between the No/Low Risk group and the Moderate/High Risk group. A cutoff score of 9 was chosen to give a margin of error that erred on the side of overestimation of the number at risk. Since this instrument is a screening tool intended to lead to further evaluation of those at risk, overestimation in order to prevent possible harm is desired over underestimation. By this method, the CRS yields a binary rating of either High Risk or Low Risk. Those whose scores equaled or exceeded 9 were given a CRS rating of High Risk while those whose scores were less than 9 were given a CRS rating of Low Risk.

Inter-expert agreement

Before the consensual validation could be analyzed, how well the experts agreed in their unstructured clinical judgments had to be determined and the gold standard established. To simplify the analysis, the GRS ratings were dichotomized into Low Risk (GRS ratings of 1 or 2) and High Risk (GRS ratings of 3 or 4). Four types of inter-rater agreements were analyzed.

1) The agreement for the 32 pairs of unstructured judgments was computed in several ways. Percent overall agreement, percent agreement on just the High/Moderate Risk cases, and percent agreement on just the Low/No Risk cases were computed first. The percent agreements on High/Moderate Risk and Low/No Risk are proportions of specific agreement. They analyze separately the agreements on the two categories of risk and represent the probability that a second rater will chose a specific risk category if an earlier rater did so already (Fleiss, 1981).

The two experts' overall GRS ratings for each of the 32 subjects agreed with each other in 29 instances. This GRS rating represented the gold standard of prediction. In the three instances of disagreement, the higher rating was taken as the gold standard (again, this was done to err on the side of over-prediction).

The 32 pairs of unstructured assessments were also tested for agreement using the kappa statistic. Kappa

is a measure of agreement between 2 observers that corrects for agreements expected by chance (Fleiss, 1981, Ch. 13). Kappa ranges from -1 (complete disagreement) to +1 (complete agreement) with 0 representing chance agreement.

2) To test the experts' agreements using the CRS, the percent overall agreement, the percent agreement on High/Moderate Risk, the percent agreement on Low/No Risk, and the kappa statistic was used to compare the 32 pairs of CRS overall ratings of risk.

3) CRS item by item agreement among the experts was then examined. For each item, weighted kappa was calculated. While weighted kappa corrects for chance-expected agreements, like kappa does, it also weights the severity of disagreements. A dichotomous-ordinal scale like the ones used on the CRS can yield disagreements in both the existence of a problem and the severity of a problem. Weighted kappa takes into account both possible types of disagreements (Fleiss, 1981, Ch. 13). The weights used are given in Figure 11.

4) In addition, using each expert's own binary GRS rating as the gold standard of prediction, the sensitivity and specificity of the experts' CRS ratings were calculated. Since the GRS rating served as a standard of prediction, the sensitivity and specificity evaluated the performance of the CRS as a "diagnostic test" (Fleiss, 1981, Ch. 1). Overall percent agreement, percent

agreement on High/Moderate Risk cases, percent agreement on No/Low Risk cases, and the kappa statistic for agreement comparing an experts' CRS rating to his/her own GRS rating were done.

Non-expert evaluations versus the gold standard

The actual consensual validation took place by comparing the non-experts' CRS ratings to the gold standard, the consensus expert GRS rating. Of the 32 subjects, 24 had CRS evaluations returned by their nurses, 20 had both CRS and GRS evaluations returned by their nurses, 22 had CRS evaluations returned by their pediatricians, 20 had both CRS and GRS evaluations returned by their pediatricians, and only 9 were assigned an in-hospital social worker, in every instance of which both CRS and GRS evaluations were returned. The sensitivity, specificity, percent overall agreement, and proportions of specific agreement of just the nurses' CRS evaluations and just the pediatricians' CRS evaluations were calculated. (Social work data were not analyzed individually since the number was too small.) Finally, the test indices of the instrument using the highest rating given by any of the 3 clinicians to a given subject were calculated. For all of these computations, only those CRS ratings that had a corresponding GRS rating were used.

To examine whether the CRS improved the prediction of the non-experts, their GRS predictions were compared to their CRS predictions. The sensitivity and specificity of both the nurses' GRS ratings and CRS ratings when compared

to the gold standard were calculated and compared. The same was done for the pediatric evaluations and for the highest evaluation given each subject by any non-expert. An increase of the CRS sensitivity over the GRS sensitivity would mean that more at-risk subjects were identified with the new instrument. Ideally, an increase in CRS specificity over GRS specificity should occur. However, allowing for a margin of error that errs on the side of caution (overestimation of those at risk) means that the specificity would be poorer than the sensitivity. A specificity below 50% would not be desired, though because that would signal very poor discrimination of those who are not at risk.

Phase IV: Evaluation of the Instrument

This phase of the study involved the use of the new instrument on a large population of newborns. Non-expert clinicians used the instrument to evaluate 363 consecutive newborns and their families over a one-year period. The inclusion criteria for both the clinicians and the infants were the same as described for Phase II. In addition, the mothers' medical charts were abstracted to obtain information on demographic characteristics (e.g. age, race, parity). The data gathered in Phase IV were used to develop a profile of the study population, to determine those items about which non-experts had difficulty making judgments, and to test inter-rater agreement on the 22 CRS items.

The demographic profile of the entire enrolled study

population was composed. Although 363 subjects were enrolled in the study, not all subjects had evaluations returned. To compare if the subpopulation of newborns without any evaluations differed from the subpopulation that had at least one evaluation returned, the demographic profiles of the two subpopulations were compared. If there were no significant differences, the chances were good that the subpopulation without any evaluations does not differ much from the population that was evaluated. If there were significant differences in the demographic profiles, it is more likely that the unevaluated subpopulation is different from the rest of the study group; most worrisome would be the possibility that the unevaluated subpopulation represented High Risk families that went unrecognized because they were able to keep clinicians from getting to know them well.

Using the scoring method developed in Phase III, the subpopulation of newborns that was evaluated by a clinician was divided into High Risk and Low Risk categories based on the highest rating given to the subject by any clinician. The percentage of newborns labeled High Risk was computed.

To learn which items were most difficult for non-expert clinicians to assess, the percentage of times each item was marked "unknown" or left blank was calculated for each CRS item for each type of clinician. A ranked ordering of items by most difficult to least difficult was compared among the three types of clinicians. This ranked ordering of the non-experts' was then compared to a ranked ordering of the items

based on the experts' percentage of unknowns for each item.

The final analysis of this study was to test item by item agreement among the clinicians. For all of the infants where an evaluation was returned by both a nurse and a pediatrician, the weighted kappa statistic was calculated for each of the 22 items on the Clinical Rating Scale to determine the level of agreement between nurse and pediatrician. The process was repeated on the evaluations where both the nurse and a social worker returned evaluations and then again on the evaluations where both the pediatrician and a social worker returned evaluations. High inter-rater agreement would signal that the three types of clinicians were consistently evaluating families using the same set of standards. The purpose of these analyses was to determine if the evaluations from the three types of clinicians could be pooled to render one rating of risk. Pooling of data would be desirable because of the high number of "unknowns" marked on individual evaluations. By pooling the data, these unknowns could be reduced. In routine use of the CRS, the ability to pool data means that different types of clinicians could collaborate in the evaluation of a family.

In the future, these data will be part of a study of the predictive validity of the instrument. When the newborns have passed their second birthdays, and again after their fourth birthdays, their medical records will be reviewed to determine if maltreatment has occurred in the in-

tervening years. From these data, the predictive ability of the Clinical Rating Scale will be measured and the length of optimal prediction of the instrument will be determined. Because of changing family, social, and psychological states, it is expected that the instrument will be maximally predictive up to the second year of life. As discussed earlier, these data will also allow the scoring method to be refined.

CHAPTER FOUR:

RESULTS

Phases I and II

The results of Phase I and Phase II can be found in Figures 1 through 4. These two phases involved the development and refinement of both the Global Rating Scale (GRS, overall assessment based on unstructured clinical judgment) and the Clinical Rating Scale (CRS, assessment based on structured clinical judgment). Version 6 of the CRS (Figure 2) incorporated the final GRS into the final CRS and was the version used in Phases III and IV.

Phase III

The families of 32 newborns were evaluated by both expert and non-expert clinicians in this phase of the study. While all 32 were evaluated by two experts using both the GRS and the CRS, the number of non-expert evaluations varied from 0 to 3. There was also a tendency for the non-experts to not complete the GRS portion of the evaluation. Only 6 subjects were evaluated on the CRS by all three non-experts and 2 of these subjects were missing one of the GRS evaluations; 15 subjects were evaluated on the CRS by only 2 non-experts, and 3 of these were missing one GRS; 7 subjects were evaluated on the CRS by only 1 non-expert, and 1 was missing the GRS; 4 subjects were not evaluated on the CRS by any non-experts. Thus, of the 32 subjects, 28 had at least one CRS evaluation, and 27 had at least one complete GRS and CRS evaluation, for a pooled total of 55 CRS's re-

turned and 49 CRS/GRS's turned in.

Development of a scoring method

As discussed in the Methods section, the preliminary scoring system for the CRS was developed by analyzing the expert evaluations of the 32 subjects. The histograms in Figures 6 through 10 each represent the five different scoring methods proposed. Each histogram groups each expert's CRS scores by the same expert's GRS ratings of risk. The best scoring method would yield the least overlap of No/Low Risk subjects to Moderate/High Risk subjects both within each rater and among all three raters.

Figure 6, the arithmetic sum, was the starting place. For experts 2 and 3, there was too much overlap of the Moderate/ High Risk group with the No/Low Risk group. Figure 7, the average score, improved on this by clearly separating out the High Risk group from the Low Risk group of expert 3. However, the degree of overlap in expert 2 worsened, and the 2 risk groups, while still separated, were brought closer for expert 1. Figure 8, the frequency of items marked 2 or higher, was a further improvement by clearly separating out the No/Low Risk group for expert 2, keeping the 2 groups separated for expert 1, and shrinking the amount of overlap for expert 3. Figure 9, the frequency of items marked 3 or higher, did not change expert 3's groupings, but increased the overlap of experts 1 and 2 and was clearly a step backwards as far as discriminating between the two risk groups. Figure 10, the arithmetic sum of those items marked 2 or

higher provided the best resolution of the two risk groups. The two groups were distinct for experts 1 and 2 with a very small overlap for expert 3. In addition, this method maintained the wide separation between groups for expert 1.

This method also made intuitive sense. Rather than just count the number of items marked with some degree of risk (ratings 2, 3, and 4), this method weights those items so that an infant with 2 high risk ratings (4's) gets a higher risk score than an infant with 2 low risk ratings (2's). While a cut-off score of 10 rendered all of the Moderate/High Risk subjects (as judged by the experts' GRS ratings) classified as High Risk by the CRS, a score of 9 or higher was chosen as the cut-off. A margin of error was built in to decrease the chance of missing an infant at risk. As a first step screening tool, the preference is to over-predict rather than miss an infant in need of intervention.

Inter-expert agreement

1) When the agreement on the pairs of experts' GRS ratings was examined (Table 5), the percent observed agreement (P_o) was 91% and kappa was 0.80. The proportions of specific agreement were 88% for High/Moderate Risk cases and 92% for Low/No Risk cases. This is considered excellent inter-rater agreement (Cicchetti, Sparrow, 1981).

Since the composition of each of the 32 pairs of experts could be any of three possible combinations,

the assignment of which expert in each pair was labeled "Expert 1" and which was labeled "Expert 2" for the contingency table was random. This is acceptable so long as the number of disagreements between the experts remained small. As can be seen in Table 5, there were only 3 disagreements. Changing the assignment of which expert was "Expert 1" and which was "Expert 2" would not change P_o and would change kappa by no more than 0.01.

2) Table 6 shows the experts' contingency table of CRS ratings. Again, the assignment of which expert in each of the 32 pairs was labeled "Expert 1" and which was labeled "Expert 2" was random. The P_o for the overall risk rating was 84%, agreement on High/Moderate Risk cases was 88%, agreement on Low/No Risk cases was 76%, and kappa was 0.65, a level considered good (Cicchetti, Sparrow, 1981).

3) The experts' item by item agreements were calculated using P_o and weighted kappa. The results, ranked in order from most to least agreement is shown in Table 7. A P_o of greater than or equal to 0.70 and a kappa of greater than or equal to 0.40 is considered fair agreement (Volkmar, Cicchetti, Dykens, et al., 1988). All but two of the items met these requirements for fair agreement. The two items that did not meet the criteria for fair agreement had low kappas but high (86%) observed overall agreements. The reason for the low

kappa is that the expected agreements (P_e) on these two questions were also very high (greater than 80%); since kappa considers and adjusts for agreements expected by chance, a high P_o coupled with a high P_e yields a low kappa (Cicchetti, 1988).

4) The contingency table of each expert's CRS rating compared to his/her own GRS rating for each subject is shown in Table 8. Using the GRS as the predictive standard, the evaluation by the CRS as a diagnostic test had a sensitivity of 100% and a specificity of 51%. This indicates that the CRS overestimates the number at high risk; all of the errors occurred in the direction of misidentifying Low Risk subjects as High Risk subjects. As discussed above, this was intentionally done in the process of developing a scoring method. When intra-rater agreement (each expert's CRS rating compared to his/her own GRS rating), was calculated, P_o=70% and kappa=0.45, fulfilling the criteria for fair agreement. The proportion of specific agreement for High/Moderate Risk cases was 72% compared to 68% for Low/No Risk cases. Thus, the agreement was higher for those subjects felt to be at High/Moderate Risk.

Non-expert evaluations versus the gold standard

The results of the consensual validation are shown on Tables 9 through 14. The nurses' GRS evaluations (Table 9) when compared to the gold standard had a sensitivity of 40% and a specificity of 100% meaning that the nurses' unstruc-

tured judgments overestimated the number not at risk and misidentified the majority of those subjects the experts felt were at risk. The nurses' CRS evaluations (Table 10) had a sensitivity of 64% and a specificity of 0.78%. The CRS thus improved the nurses' identification of those at risk. This occurred with a decrease in specificity, yet without lowering the specificity to an unacceptably low level. The CRS also improved the nurses' proportion of specific agreement, raising it from 57% to 70%. This means that if the experts felt a subject was at High/Moderate Risk, there was a 57% probability that the nurses would concur using their unstructured judgments but a 70% probability that the nurses would concur if they used the CRS to help them in their evaluation.

The pediatricians' GRS evaluations (Table 11) had a sensitivity of 63%, a specificity of 92%, and a kappa of 0.57. The use of the CRS (Table 12) did not change any of the test indices for the pediatricians. This indicates that the non-expert pediatricians did a good job of discriminating who is at risk of child maltreatment with and without the Clinical Rating Scale.

Table 13 compares the highest GRS rating given to each subject by a non-expert (social work data included) to the gold standard. The sensitivity was 54% and the specificity was 86%. As shown in Table 14, the highest CRS rating given to each subject by a non-expert had a sensitivity of 77%, a specificity of 79%, a kappa of 0.55, and a proportion of

specific agreement of the High/Moderate Risk cases of 77%. Overall, these indices are better than for the nurse evaluations considered singly or the pediatrician evaluations considered singly.

In summary, experts using their unstructured clinical judgments have good inter-rater agreement in assessing an infant's risk of subsequent child maltreatment. Using the newly developed Clinical Rating Scale, the experts maintain good agreement with each other and with their own original unstructured evaluations of risk. Using the experts' unstructured judgments as the standard of prediction, the CRS has high test indices when used by non-experts and shows promise as a screening instrument for newborns.

Phase IV

Of the 363 subjects enrolled for this phase, 26 (7%) had no evaluations returned by a non-expert nurse, pediatrician, or social worker. Table 15 shows the distribution of returns.

Five of the 363 (1%) maternal medical charts were unavailable for review. Table 16 gives a demographic profile of the entire population, the subpopulation that had at least one evaluation returned, and the subpopulation that had no evaluation returned. The 2 subpopulations had similar profiles. There were some minor differences between the 2 groups, but since one subpopulation is 23 times the size of the other (N=337 vs. N=26) it is difficult to draw the

conclusion that there is a significant difference. The fact that the 2 subpopulations are fairly similar in race, gender of the infant, type of delivery, and marital status signal that the two subpopulations are probably identical.

In Phase IV, the subjects were given a rating of High Risk or Low Risk based on the CRS by each clinician who returned a CRS. Table 17 gives a profile of how each type of clinician rated the population. The social work evaluations differed from those of the nurses and pediatricians in that the social workers did not evaluate all of the subjects; due to limited resources, the social workers only visited those patients referred by nurses and doctors. Thus, it is expected that a higher percentage of this group would be identified as high risk as compared to the entire population seen by the nurses and pediatricians.

To study how difficult certain items were for clinicians to evaluate, the percentage of times each item was left blank or marked "unknown" by each type of clinician was computed and is shown in Table 18. Table 19 shows the same data but ranks the questions from highest to lowest percentage of unknowns. The non-expert clinicians were similar in which items were more difficult although, not surprisingly, the individual percentages of unknowns were lower for the social workers than for the other types of clinicians. For the most part, the difficult questions concerned background information about the mothers (e.g. history of violence, criminal record, psychiatric history, care as a child, and

family conflict) while the least difficult questions concerned in-hospital behaviors of the mother.

Of note, the item on past care of children, an item one would intuitively expect to be an indicator of risk of maltreatment, was one clinicians were reluctant to evaluate. It is unclear whether clinicians had difficulty talking to mothers about child care and their past care of children or whether they had difficulty judging the honesty of the responses obtained. That the experts marked "unknown" up to 23.4% of the time (mode=4.7%, median=4.7%, only 2 items were never marked unknown) despite interviewing the mothers with the CRS items in mind, indicated that difficulty in forming judgments about people played a significant role in the high percentages of "unknowns" marked.

Since each group of clinicians evaluated a different number of subjects, it is possible that the rankings would be different if the exact same subjects were evaluated by each type of clinician. To test this, Table 20 ranks the items again by percentage of unknowns but only for the 190 patients evaluated by both a nurse and a pediatrician. This table is very similar to Table 19 and suggests that Table 19 is a fairly accurate reflection of how much difficulty each type of clinician had with each item.

The final question was how well the non-experts agreed with each other on each item of the CRS. Tables 21, 22, and 23 list the weighted kappa, observed percent agreement (P_o), and expected percent agreement (P_e) for each item for each

of the three possible pairings of non-experts. The items are ranked from highest to lowest weighted kappa values. Only 5 of the 22 items met the criteria for fair agreement ($P_o > 0.70$ and weighted kappa > 0.40). The other items yielded a low weighted kappa in the face of high P_o values. As occurred when the experts' item by item agreements were examined, the low weighted kappas resulted from very high chance-expected agreements. For the most part on these items, subjects were rated as No Risk or Low Risk, ratings the majority of the population would be expected to receive. Since kappa evaluates agreements correcting for chance, on items where there is a low prior probability of being rated Moderate/High risk (and, thus, a high P_e), kappa will be low even if the observed agreement is very high.

Tables 21, 22, and 23 show that despite low values for weighted kappa, the non-expert clinicians had excellent observed agreements for the vast majority of the items. This indicates that the non-expert clinicians completing the CRS evaluations individually were using similar standards to rate subjects. This further indicates that the data from the different types of clinicians could be pooled in order to decrease the number of unknowns on a subject before a final score was tallied. Another application of this analysis would be to have the different types of clinicians confer with each other and complete the CRS as a joint effort. The item by item agreement analysis indicates that complet-

ing the CRS by consensus would not lead to conflicting conclusions about a subject's risk for each item.

CHAPTER FIVE:

DISCUSSION

Prior studies have shown that expert and non-expert clinicians can identify those at risk of child maltreatment. Global scales of unstructured clinical judgment are simple, direct, and easy to construct and use. These global scales, however, have unspecified components with unspecified demarcations (Feinstein, 1987). Such vagueness leads to wide variances in evaluations among different clinicians and inconsistencies within clinicians.

This study standardized judgments about a newborn's risk of subsequent maltreatment and thus improved the ability of clinicians to identify those families who would most benefit from social service support. This was accomplished by first identifying the risk factors that experienced clinicians evaluate in making their judgments about a child's risk of subsequent maltreatment. The measurement of each of these risk factors was then standardized. The resulting Clinical Rating Scale (CRS) was pilot tested, refined, and then tested on a large population of newborns. This study has shown that the CRS has good inter-rater and intra-rater reliability. Used as part of the routine care of all newborns, the CRS should help non-expert clinicians focus their evaluations and use specific criteria in their judgments.

Strengths and Limitations

The CRS is practical for use as a screening tool on all newborns. As discussed in Chapter Two, the ideal screening

tool would be used in the postpartum period, would use structured clinical judgment, and would not require a specially designed interview. The CRS meets all of these criteria. Because it can be completed expeditiously, is simple to use, and does not require special training for clinicians, it has strong potential for widespread clinical application.

The consensual validation revealed good sensitivity and specificity of the CRS as a screening test when compared to the current gold standard of prediction. There was overall improvement of the non-experts' judgments over their unstructured clinical judgments. In particular, the judgments of nurses were significantly improved. There was also good inter-rater agreements on the 22 items in the rating scale indicating consistency in the use of the instrument by different evaluators.

The CRS improves on prior instruments that screen for risk of maltreatment in several respects. The CRS takes less than five minutes to complete at the end of a mother's postpartum hospital stay. The instruments of Monaghan and colleagues (Monaghan, Gilmore, Muir, et al., 1986) and Altemeier and colleagues (Altemeier, O'Connor, Vietze, et al., 1984) both required interviews of about one hour. Altemeier's interview yielded a sensitivity of 65% compared to a sensitivity of 64% for nurses using the CRS and 63% for pediatricians using the CRS. Browne developed a 12-item postpartum checklist that was reported to have a sensitivity of

81% and a specificity of 94% but a positive predictive accuracy of only 7% (Browne, Lowton, 1987; Browne, Saqi, 1988). This low positive predictive accuracy resulted in part because of a low prior probability. Whether the CRS can improve on this positive predictive accuracy awaits the longitudinal follow-up phase. However, it is expected that there will be a higher positive predictive accuracy because the population in this study has a higher incidence of child maltreatment.

A major problem in this study was the high percentage of times that items were marked "unknown" by the non-expert clinicians. Phase IV showed that clinicians (especially the nurses and pediatric residents) had difficulty gathering sensitive data and making judgments about the mother's past. Other studies on the self-report of drug use have found that 23% of illicit drug users deny their use when questioned (Zuckerman, Amaro, Cabral, 1989). In any type of evaluation, judging the veracity of the data obtained is difficult.

Improvement of this data gathering would require a two-tiered approach. First, clinicians need to be encouraged to broach sensitive subjects with their patients. Clinical training needs to stress the importance of past actions and behaviors on the current/future health and welfare of the patient and his/her family. Clinicians are in the unique to position of being able to help patients identify and address areas of their lives that threaten their health (e.g. drug/

alcohol use, violent behavior). Clinicians thus need to be taught that it is part of their care of patients to ask these questions. This type of increased experience at gathering sensitive information would make it easier for clinicians to ask these questions. Second, clinicians need to hone their abilities to judge the quality of the information that they do gather. Increased experience asking sensitive questions would help here also. Improving the quality of the relationship between clinician and patient would also be a significant aid. The non-expert clinicians in this study all had a very short amount of time (from two to five days) in which to get to know the families studied. That they were not completely comfortable evaluating these families is understandable. However, they had the same amount of time in which to get to know these families as do the clinicians who routinely care for these families in the postpartum period and who routinely judge the safety of sending a newborn home with his/her parents. A system of continuity of health care would improve on the clinician-patient relationship. This issue is important for the entire practice of medicine. Such an improved relationship would help clinicians both evaluate risk factors and quickly help those judged to be at risk.

Another issue is the high percentage of families identified as High Risk by the CRS. The nurses and pediatricians identified 32% and 36% of the population as being at high risk. When the highest rating given by any clinician

was used, 50% of the population was identified as at high risk. One reason for this was that the cut-off score was chosen to be lower in order to over-estimate the number at risk. Since the CRS is designed to be a first step screening tool with those identified as High Risk going on to more in-depth social service evaluation, it is better to err on the side of safety. If the cut-off score were to be raised one point to a score of 10, nurses would identify 27% of the population as at high risk, pediatricians would identify 32% as such, and the use of the highest rating would identify 45% of the population as at high risk.

Another possible reason for the high number of families identified as High Risk is that the individual risk factors are not weighted. All of the CRS items are counted as equally important despite the fact that, intuitively, past abuse of children would seem to put a family more at risk than would cooperativeness with the hospital staff. The results of the prospective study will be necessary in order to refine the scoring system.

Finally, it is difficult to say whether these high percentages reflect to some degree the nature of the population included in this study. As discussed in the Background section of Chapter Two, estimates of the lifetime prevalence of child maltreatment range from 10 to 40%. Until the longitudinal follow-up of the population is conducted, it is not possible to identify how much the true characteristics of the specific population studied contributed to this high

percentage.

Ethical Considerations

There are several ethical issues entangled with the concept of screening families for risk of child maltreatment. The first is the question of whether screening should be voluntary or mandatory. The whole purpose of screening is to protect children, yet they would not be the ones to give consent, their parents would. Parents who are at high risk are less likely to volunteer for such screening. With mandatory screening, a conflict then arises between the parents' right to (or desire for) privacy and freedom in parenting and the children's right to life and physical integrity. Regardless of the choice that society makes between these two conflicting liberties on purely ethical grounds, society benefits from preventing maltreatment because of the high cost to society of child maltreatment (see Chapter One).

Another major area of concern is that of labeling people. Although it is not meant to be, being identified as "High Risk" may be equated with actually being a child maltreater. This labeling is even more harmful for those who are mislabeled. With an incidence of 10%, and a sensitivity and specificity of 75%, fully 25% of the population will be misclassified when compared to actual outcomes of parenting. For each high risk family correctly identified, three low risk families will be misidentified as being at high risk

(Kotelchuck, 1982; Avison, Turner, Noh, 1986). Even with very sensitive and specific instruments the magnitude of misclassification will still be large.

The best ways to protect families from being stigmatized include:

1) the careful protection of confidentiality.
2) the careful use of terms. Families should be recognized as those who would most benefit from social services and should not be called "abusive" or "failures."
3) an avoidance of coercion. Consent and full disclosure should be part of all programs. Without evidence of actual or imminent danger to the children, help for these families should be voluntary (Brody, Gaiss, 1976).

The large number of misclassifications also includes a large number of false negatives. This misidentification can lead to a false sense of security and to ultimate harm coming to a child (Brody, Gaiss, 1976). However, this is a weak argument against screening. Without any screening program in place, the situation that occurs in many health centers, all children at risk will lack preventive measures and even more children would meet harm.

Restricting or prohibiting screening programs would not ameliorate all of the issues associated with misidentification of families. As Daniel points out, most health centers already perform some sort of screening, all of which are associated with false positive and false negative events

(Daniel, Newberger, Reed, 1978). The screening in these programs often consists of informal, unstructured clinical judgment that results from routine clinic visits. As the current study has shown, such unstructured judgments have poorer test indices than structured instruments like the CRS. Not implementing structured screening programs would ultimately yield more mislabeled families than the use of the CRS.

Conclusions

Phase III evaluated the use of the CRS in the hands of both expert and non-expert clinicians when compared to the current gold standard of prediction. When compared to the gold standard, the CRS had a good level of sensitivity and specificity for the non-experts as well as the experts.

When compared to actual outcomes (whether the subjects are actually maltreated), it is anticipated that the CRS will retain a good level for these test indices. However, it is unlikely that either the sensitivity or the specificity will exceed 90% for many reasons. First, there is no "typical" abuser, so all High Risk cases will not be identified at birth (Roberts, 1988). Second, social pediatric illnesses, (e.g. trauma, failure to thrive, accidents, ingestions) are on a continuum. There are not always cut-and-dried demarcations between maltreatment and non-maltreatment (Kotelchuck, 1982). Also, issues related to potential for sexual abuse are more complex than can be assessed in a rou-

tine clinical relationship. Thus, the CRS does not attempt to gauge those issues although sexual abuse would be counted in any follow-up that identified which subjects were subsequently maltreated. Further, not all patients identified as being at high risk will have a negative outcome since child maltreatment is a result of interactions between many personal and environmental factors. Schneider et al. found that 20% of the population have attitudes and experiences similar to those of known abusers, and 20% of the population have unusual child rearing practices, yet most of these people will not maltreat their children (Schneider, Hoffmeister, Helfer, 1976). Finally, since child maltreatment is the result of many factors, this instrument is not necessarily looking for those who will maltreat their children but those who are at risk of maltreatment. Those found to be at high risk would benefit from social service interventions. This over prediction in the interest of prevention leads to a low specificity.

Phase IV evaluated the use of the instrument by non-expert clinicians. This evaluation demonstrated that the CRS is used similarly by the different types of clinicians. The three types of clinicians who were studied had difficulty collecting data and making judgments on the same set of items. On those items that they did collect data and make judgments, the clinicians had good inter-rater agreement. These results suggest that the information from the different types of clinicians could be pooled in completing the

CRS, thus decreasing the number of unknown items.

The predictive validity of the CRS remains to be tested. However, once these data are collected, it will still be difficult to assess how well the instrument has performed since there have yet to be established in the literature standards for screening instruments. As a first step in the establishment of such standards, the following criteria are proposed:

Sensitivity	False Negative Error	Clinical Evaluation
<70%	>30%	Poor
70%-79%	21%-30%	Fair
80%-89%	11%-20%	Good
90%-100%	0%-10%	Excellent

The above criteria, coupled with a specificity of at least 50%, are suggested as the minimum acceptable test indices for a screening instrument.

With most screening instruments, those who exceed a specified cut-off score will be further evaluated for the trait/disease being sought. The false positives on a test represent those who will erroneously have further, in-depth evaluations performed. The false negatives represent those who will "pass" the screen and will, thus, receive no further evaluation. While it is desirable to minimize the false positives, the commission of this error has less of a negative outcome than the commission of false negative error. Those who pass the screen, but should not have, will receive no further evaluation and may consequently suffer

physical harm/illness. The above criteria thus emphasize the false negative error and set as the limit of acceptable missed cases 30%. A higher rate of missed cases would, in practical terms, mean the screening instrument is ineffective in identifying the true positives.

A specificity of at least 50% is desirable because this index indicates how many of the true negatives are correctly identified. If the specificity were less than 50%, then those who are truly without the trait/disease being sought would have a higher probability of being falsely identified as positive than correctly identified as negative. While the only resulting harm would be that too many subjects would undergo further, in-depth evaluations, such a situation would mean that the screening instrument is less useful in identifying the true negatives than a flip of a coin.

Whether the CRS meets these criteria when compared to the outcomes of the subjects remains to be seen. When the evaluation of the experts is used as the standard of measurement, the CRS (highest rating given by an evaluator, Table 14) meets the criteria for fair evaluation.

The initial evaluation using the CRS on 363 subjects was completed in Phase IV; the long-term follow-up of these subjects remains to be conducted. From these data, multiple regression analysis will be performed, and each item in the scale will be weighted to reflect its importance on overall predictive ability. With appropriate weights, the scoring of the instrument will be refined which will improve the

sensitivity and specificity of the CRS. The regression analysis may even lead to a new method of scoring. Rather than using an arithmetic sum, Boolean clusters or a hierarchical system might prove to be a better scoring system (refer to "Clinimetrics and the Development of Rating Scales in Chapter Two).

Those found to be at risk of abuse should be closely followed for psychosocial stressors that may lead to parenting failure. Like prior attempts at identification of those at high risk, it is expected that the CRS will be predictive for 2 to 4 years with a decline in predictive ability over time (U.S. Dept. HHS., 1988; Altemeier, O'Connor, Vietze, et al., 1982). This is because the parents will change over time and because over the years, social and personal circumstances will change.

No matter what the results of the predictive validation study, the CRS needs to be retested on a new sample of patients. Despite the large size of the sample population used in Phase IV, the sample may not reflect the total population at large (Soper, Cicchetti, Satz, et al., 1988). A difference in prevalence of maltreatment between the sample population and the total population can greatly affect the test indices of sensitivity and specificity.

As discussed in Chapter One, perinatal screening for risk of maltreatment is important; despite intervention after abuse has occurred, even after just one event, the maltreatment can lead to permanent physical and psychologi-

cal damage to the child.

> The way forward for us seems to be further development of screening programmes at the time of birth when vulnerable parents can be given appropriate help before any serious damage, physical or emotional, has been inflicted on the baby. Our ultimate aim must be prevention (Lynch, Roberts, 1982).

Along with perinatal screening, ongoing assessments of families also must occur. Since all people change over time, those who are found initially to be at low risk should not be overlooked. They should also be observed by clinicians during routine health care for changes in personal or social circumstances or for stressful life events that may turn a Low Risk family into a High Risk family.

> Because of our complex dynamic society with such factors as unemployment, inflation, infidelity, etc., we feel that the child and parent should be observed on a continuing basis in order that the potential for child abuse might be detected whenever it develops (Rosenberg, Myers, Shackleton, 1982).

The ability to be a good parent is partly innate and partly a skill developed by the experience of trial and error. Just because a parent, especially a new parent, does not show complete confidence or finesse at the birth of a child, it does not portend a fateful outcome for the child. Thus, the CRS does not examine parenting style and does not seek to have parents prove that they will be good parents. Rather, the CRS assesses risk factors that could cause even good parents to experience parenting failure. The purpose of the CRS is to identify those who would benefit most from

scarce social service resources so that their parenting abilities can be maximized.

FIGURES

Figure 1

Parenting Skills Screening Instrument - Version 1

CARE OF OLDER SIBS Past history	Sibs well cared for	Sib unkempt, miss school often, or miss appts. often	Documented hx of Failure-to-Thrive	Suspicion of Abuse or Neglect	Documented abuse or neglect	Unable to score question/ Unknown
PARENTS Intelligence	No concern—both parents able to function well	Parents slow—unable to interact well	Documented in chart that one parent has IQ<70			Unable to score question/ Unknown
Psychiatric Hx	No prior hx or current sx	Mild illness (mood d/o)—under control and tx	Mild illness (mood d/o)—not being tx'ed	Major illness (psychosis, personality d/o)		Unable to score question/ Unknown
Illicit Drug Use	Never used drugs	<2 per week prior to pregnancy—none now	≥2 per week prior to pregnancy—none now	Multiple drugs in past or any use during pregnancy		Unable to score question/ Unknown
Alcohol Use	Never used alcohol	<2 drinks per wk prior to preg.—none now	≥2 drinks per wk prior to preg.—none now	Known alcoholic or drank during pregnancy		Unable to score question/ Unknown
Criminal Record	No prior criminal record	Has been charged >2 of crimes—never convicted	Currently has charges pending	Has been convicted/jail record		Unable to score question/ Unknown
Violent Personality	No violent verbal or physical outbursts	Has violent verbal outbursts—never physical	Has rare violent verbal and physical outbursts	Has frequent (>2 per mo.) physical outbursts	Has injured someone enough to require hosp.	Unable to score question/ Unknown
Parents' Childhood	Happy memories of childhood	Changed guardians >1—always stayed with family	Was in foster care	Changed foster homes	Was abused or neglected	Unable to score question/ Unknown
Maternal Age	Mother >18 y.o. at birth of first child	Mother now >18 but was ≤18 at birth of 1st kid	Mother now ≤18			Unable to score question/ Unknown
CURRENT FAMILY Family Violence	No family violence-domestic peace	Frequent (>2/mo) bitter (yelling) arguments			Physical violence among family members	Unable to score question/ Unknown
Lifestyle	Stable—family intact	Family members (not parents) move in/out often	No father figure	Mother's boyfriend changes often	Children shuttled between different family members	Unable to score question/ Unknown
Housing	Safe, clean home with adequate room—move <1/yr	Moves >1/yr, but between safe, roomy homes	Live in unsafe, overcrowded home	Live in temporary shelters	Live on street/ abandoned bldgs.	Unable to score question/ Unknown
Provisions for Baby	Have necessary clothes and furniture ready	Have crib, car seat, baby food—seeking others	Actively seeking necessary items—can borrow items		Unprepared—parents unaware of needs	Unable to score question/ Unknown
Social Supports	Someone can care for baby at all times	Usu. someone to care for baby—no one in emerg.	Frequently (>1/wk) times with no one for baby care		No provisions for baby care on a daily basis	Unable to score question/ Unknown
CURRENT PREGNANCY Attitude Towards	Preg. planned and wanted	Unplanned preg. but baby wanted	Unplanned preg.—abortion/adoption considered		Parents remain ambivalent towards baby	Unable to score question/ Unknown
Prenatal Care	Regular care; began ≤8 weeks	First appt >9wks or missed ≥ 3 appts.	Mother had no prenatal care but delivered in hosp	No prenatal care and delivered at home		Unable to score question/ Unknown
CURRENT BEHAVIORS Care of Newborn	Mother is attentive to baby's needs;is gentle	Mother is attentive but handles baby roughly	Mother ignores baby's needs but interacts		Mother ignores baby	Unable to score question/ Unknown
Visiting Child	Mother with child constantly	Mother with child at least 1/day	Mother visits <1/day or only when encouraged to		Mother refuses or reluctant to see child	Unable to score question/ Unknown
Cooperativeness	Mother cooperates with hospital staff	Mother needs to be coaxed into cooperating	Mother resists medical care or advice	Mother tries to leave hosp. with baby AMA		Unable to score question/ Unknown
Obstructive Behavior	Mother cooperates with baby's med. care	Mother argues with staff caring for baby	Mother obstructs medical care for baby			Unable to score question/ Unknown
Danger to Child	Mother mindful of baby's safety	Mother not careful with baby—unintentional	Mother places baby in obvious danger		Mother harms baby	Unable to score question/ Unknown
Threats	Mother pleasant with hosp. staff	Mother ambivalent towards staff	Mother argues with staff	Mother threatens staff with physical harm	Mother tries to actually harm staff	Unable to score question/ Unknown

Ver. 1

Figure 2

Parenting Skills Screening Instrument - Version 6

For each of the numbered items on the left, please mark the statement to the right which, in your best judgment, is most accurate. If more than one statement applies, mark the statement which is furthest to the right.

Please be sure to answer the last question concerning your overall rating of risk. *THANK YOU!*

CARE OF OLDER SIB: Past History 1.	Unknown / Primip		No problems with care	Poor well-child care (e.g. missed appts.)	Suspicion of abuse or neglect	Verified abuse, neglect, or failure to thrive
MOTHER Intelligence 2.	Unknown		No concerns; functions well	Slow; limited thinking	Mildly retarded	Moderately/severely retarded
Psychiatric (Ψ) 3. History	Unknown		No prior history or current symptoms	Mild (e.g. anxiety)	Past Ψ hosp. admission or current Ψ drug use	Major (psychosis, major depression, suicide try)
Illicit Drug Use 4.	Unknown		No history of drug use	Used before pregnancy, none during pregnancy	Used occasionally during pregnancy	Used regularly during pregnancy
Alcohol Use 5.	Unknown		No history of alcohol use	Drank before pregnancy, none during pregnancy	Drank occasionally during pregnancy	Drank reg during preg. or untreated alcoholic
Motivation 6.	Unknown		Uses resources; seeks help as needed	Seeks help once encouraged to do so	Slow to address problems	Denies problems; resists help
Criminal Record 7.	Unknown		No criminal record	Has been charged but never convicted	Currently has charges pending	Has been convicted OR jail stay now or in past
History of 8. Violence	Unknown		Verbal outbursts only	Rare physical outbursts	Frequent physical outbursts	A past victim has needed medical care
Care as a Child 9.	Unknown		No history of abuse of neglect	Poor nurturance (e.g. freq. change in care)	History of abuse or neglect	Spent time in foster care
Age 10.	Unknown		>18 at birth of first child	Now >18, but was ≤18 at birth of first child	Now ≤18 years old, but >15 years old	Now ≤15 years old
CURRENT PARTNER Gen. Character 11.	Unknown		No concerns--is appropriate	No man is involved	Man involved has past prob. in any areas #2-10	Man involved has current prob. in any areas #2-10
CURRENT FAMILY Family Conflict 12.	Unknown		None or mild	Discord with some bitter arguing	Strong discord but no violence	Physical violence in family
Stability of 13. Couple	Unknown	No partner	Stable--couple/family intact	Couple's relationship unstable	Boyfriends change often	Absence of stable relationships
Housing 14.	Unknown		Safe and adequate	Unsafe or overcrowded home	Live in temporary shelter	Live on street/ abandoned buildings
Provisions for 15. Baby	Unknown		Have necessities ready	Have food, crib, car seat; need the rest	Still seeking necessities	Parents unaware of needs for baby
Social Supports 16.	Unknown		Family or friends available regularly	Family or friends available occasionally	Family/friends avail. only in emergency	No family/friends avail. to help, ever
CURRENT PREGNANCY Attitude 17.	Unknown		Pregnancy planned and wanted	Unplanned pregnancy, but baby wanted	Abortion or adoption was considered	Remain ambivalent (e.g. drug induced delivery)
Prenatal Care 18.	Unknown		Regular and began in the first trimester	Regular and began in the second trimester	Began in third trim. OR irreg. attendance	No prenatal care
CURRENT BEHAVIORS Care of Newborn 19.	Unknown		Mom attentive and appropriate with baby	Attentive but rough with baby	Ignores baby's needs often but interacts	Usually ignores baby OR interacts seldom
Visiting with 20. Baby	Unknown		Mom w/ baby constantly; daily if in NBSCU	Mom w/ baby >½ the day; every 1-2 days if NBSCU	Visits <½ the day; 1-2 times a week if NBSCU	Reluctant or refuses to see baby
Cooperativeness 21.	Unknown		Cooperates with hospital staff	Resists medical care or advice	Tries to leave with baby AMA	Threatens or tries to harm staff
Danger to Child 22.	Unknown		Mom mindful of baby's safety	Unintentionally not careful	Places baby in obvious danger OR not careful	Mother harms baby

** *Please complete the following statement by circling the phrase in bold which you feel is most accurate.*

I think that this baby is at *HIGH RISK / MODERATE RISK / LOW RISK / NO RISK* for being abused or neglected sometime in the future.

Figure 3

Global Rating Scale - Version 1

Based on your clinical judgment and your interactions with this newborn's family, please mark the one statement below which you feel is most accurate.

_____ I am fairly certain that this child will be abused or neglected and needs to be protected by separation from the parents.

_____ I am concerned that this child might be abused or neglected, and I feel that this family will need to be closely followed and given assistance by Social Services.

_____ I feel that there's a chance this child will be abused or neglected, but I don't feel that intervention is warranted at this time.

_____ I feel that this child will not be abused or neglected.

Figure 4

Global Rating Scale - Version 2

Based on your clinical judgment and your interactions with this newborn's family, please mark the one statement below which you feel is most accurate.

_____ I am fairly certain that this child will be abused or neglected; this family definitely needs assistance from Social Services (e.g. parenting classes, support groups, assistance).

_____ I am concerned that this child might be abused or neglected, and I feel that this family will need some assistance from Social Services.

_____ I feel that there's a chance this child will be abused or neglected, but I don't feel that intervention is warranted at this time.

_____ I feel that this child will not be abused or neglected.

Figure 5

Definitions of Risk Ratings

High Risk--There is a strong possibility the baby will be maltreated. During the hospital stay, the parents have exhibited significant risk factors. The family needs definite social work evaluation, a DART referral, and some sort of intervention.

Moderate Risk--There is the possibility that this baby will be maltreated. The parents have revealed risk factors during the hospital stay, but they have support systems intact in the family. A social work evaluation is in order and careful social work follow-up after discharge is necessary.

Low Risk--The rater feels that the baby is safe, but there is a slight possibility of maltreatment. While no social work follow-up or intervention is mandatory at this time, the pediatrician should follow this family closely.

No Risk--A strong feeling that the baby is not at risk for maltreatment. The evaluator feels there are no outstanding risk factors and feels comfortable with this rating.

Figure 6

Scoring Method 1: Sum of All Items

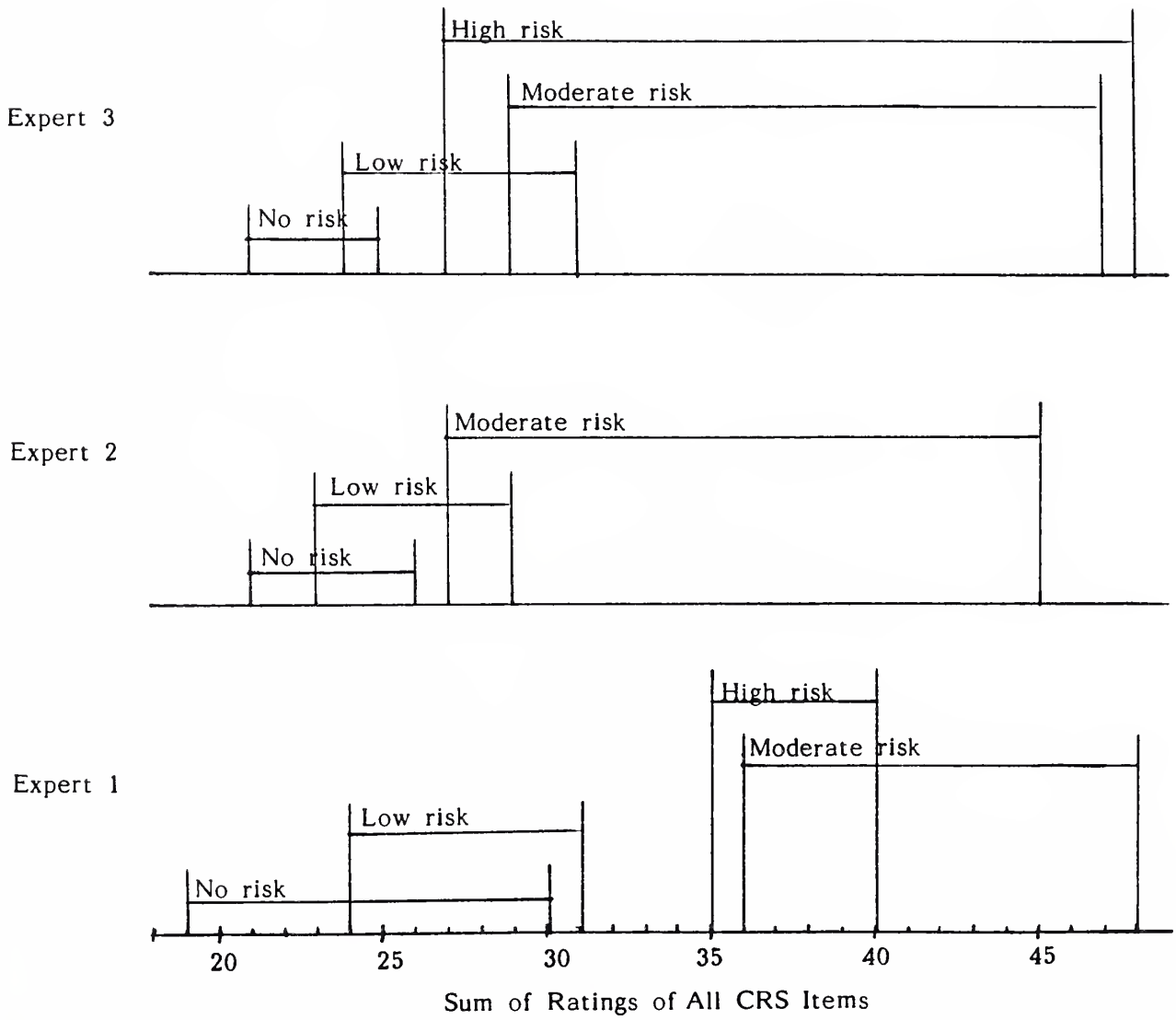

Figure 7

Scoring Method 2: Average Score of All Items

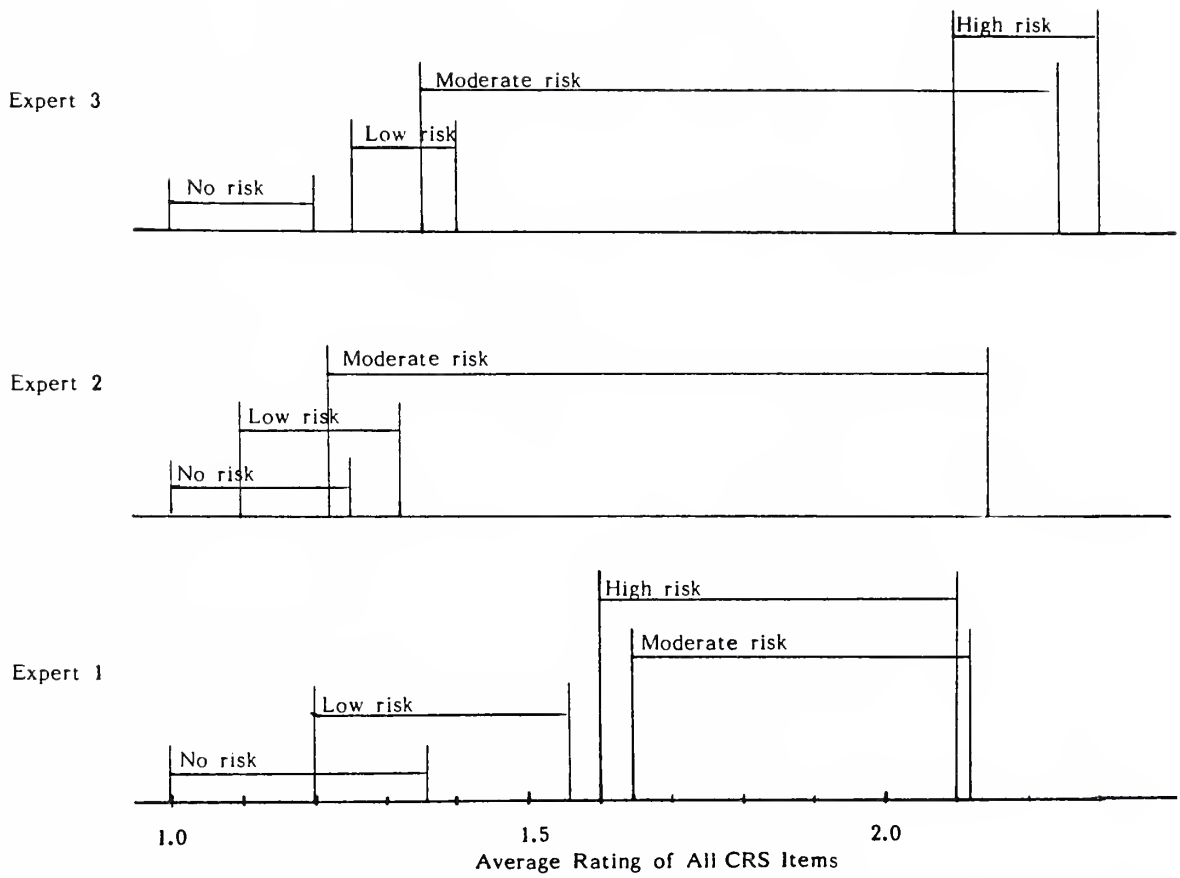

Figure 8

Scoring Method 3: Frequency of Ratings 2 or Higher

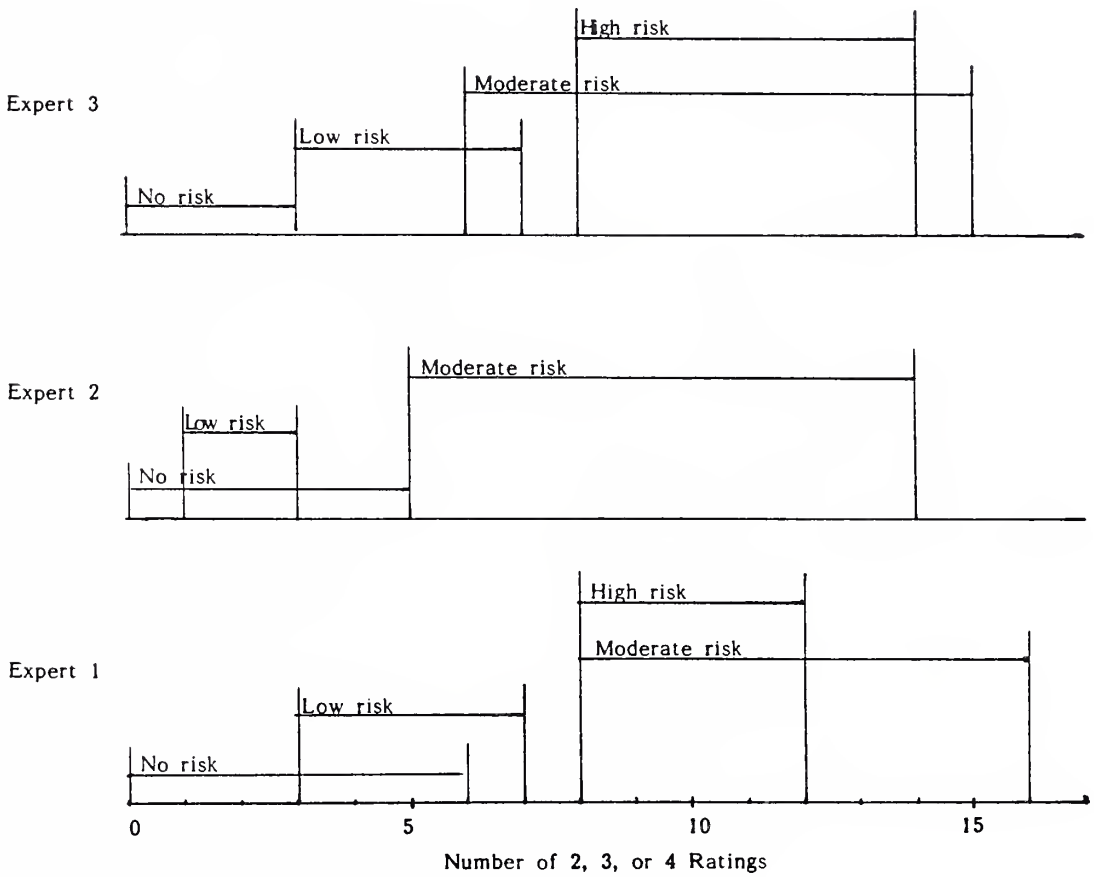

Figure 9

Scoring Method 4: Frequency of Ratings 3 or Higher

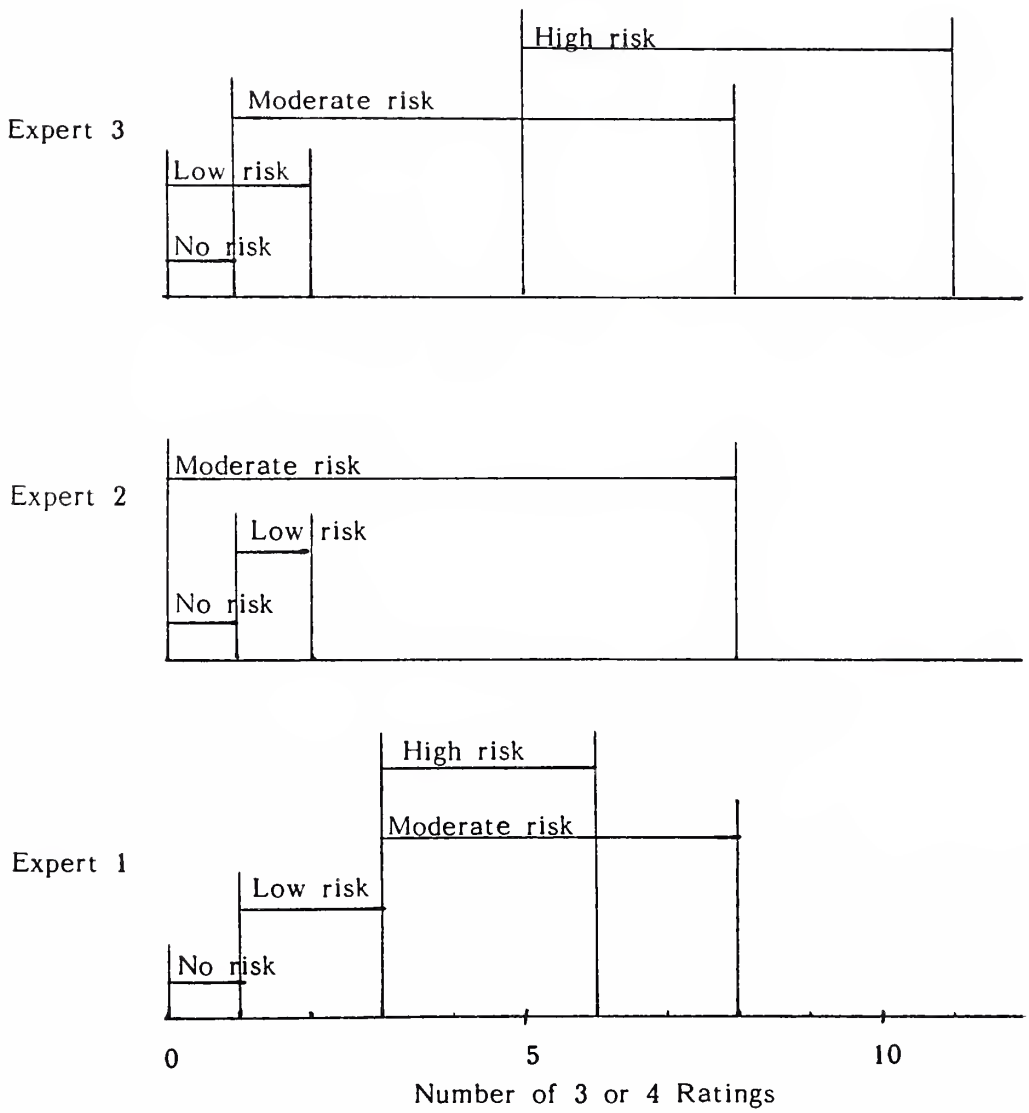

Figure 10

Scoring Method 5: Sum of Items Rated 2 or Higher

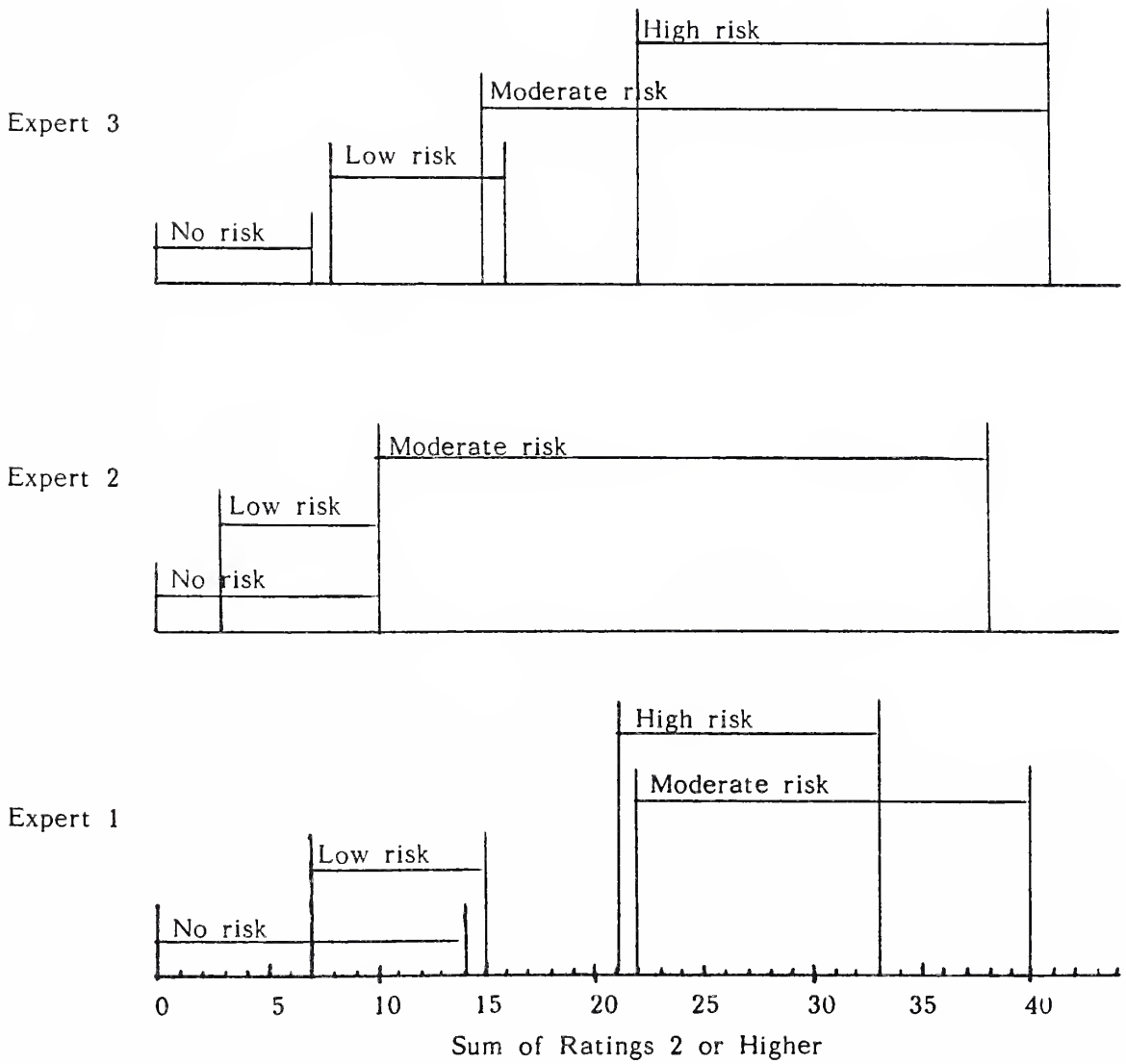

Figure 11

Weights Used in Analysis of CRS
Item By Item Agreements

		Expert #2 Response			
		No Risk	Low Risk	Mod. Risk	High Risk
Expert #1 Rsponse	No Risk	1.0	0.6	0.2	0.0
	Low Risk	0.6	1.0	0.8	0.4
	Mod. Risk	0.2	0.8	1.0	0.8
	High Risk	0.0	0.4	0.8	1.0

TABLES

Table 1
A Comparison of the Contents of Instruments that Screen for Risk of Child Maltreatment
(Key to studies at end of table)

Item	Study 1	2	3	4	5	6	7	8	9	10	11
1. Past care of children	*	*	*	*							
2. Mom's intelligence	*			*							
3. Mom's psychiatric history	*		*	*		*	*		*		*
4. Mom's drug use	*								*		
5. Mom's alcohol use	*										
6. Mom's motivation to solve own problems	*			*	*						
7. Mom's criminal record	*		*								
8. Mom's history of violence	*		*			*			*		
9. Mom's care as a child	*	*	*	*	*				*		
10. Mom's age	*					*	*		*	*	*
11. Father's character	*										
12. Family conflict	*	*							*		
13. Family stability	*								*	*	
14. Housing	*	*		*					*		
15. Provisions for baby	*										
16. Social supports	*		*		*				*		
17. Attitude towards current pregnancy	*	*	*	*					*		*
18. Prenatal care	*									*	*
19. Care of newborn	*		*								
20. Visiting with baby	*		*						*		
21. Cooperativeness	*			*					*		*
22. Danger to child	*										
23. Observed abusive tendency		*									
24. New infant in family		*							*		
25. Truthfulness		*									
26. Job change		*							*		
27. Marital status		*				*			*	*	*
28. Parenting judgment		*	*		*				*		
29. Perspective on life		*									
30. Self-esteem		*	*		*				*		
31. Ability to handle stress		*			*						
32. >1 child ≤5 y.o.		*							*		
33. Depression			*						*		
34. Multiple crises or stresses			*		*				*		
35. Unrealistic expectations of child			*	*							
36. Child perceived as difficult			*						*		

(continued)

Table 1, continued

Item	1	2	3	4	5	6	7	8	9	10	11
37. Time between pregnancies						*			*		
38. Post-delivery separation						*	*			*	*
39. Prematurity, low birthweight, handicap						*	*		*	*	
40. SES problems						*		*	*	*	
41. Socialization difficulties						*			*		
42. Social work consult at delivery							*				*
43. Mom concerned with parenting ability							*		*		
44. Neighborhood								*			
45. Mom rigid or domineering									*		
46. Poor self-understanding									*		
47. Observed violence as a child									*		
48. Parenting skills									*		
49. Inconsistent use of discipline									*		
50. Deficit in parenting knowledge									*		
51. Parent/child personality mismatch									*		
52. Child resembles disliked adult									*		
53. Reconstituted family									*		*
54. Many children									*		
55. Death of friend/family									*		
56. Sudden loss/chronic illness									*		
57. Tolerance of physical punishment									*		
58. Sexual stereotypes of children									*		
59. Violence in media									*		
60. Extreme ideas of rights or privacy									*		
61. Prenatal classes										*	
62. Smoking										*	
63. Breast feeding										*	
64. Pregnancy/delivery complications											*

Key to Studies in Table 1

1. Kung/Leventhal Parenting Skills Instrument
2. Altemeier, O'Connor, Vietze, et al. (1984)
3. Murphy, Orkow, Nicola (1985)
4. Monaghan, Gilmore, Muir, et al. (1986)
5. Avison, Turner, Noh (1986)
6. Browne, Lowton (1987)
7. Lynch, Roberts (1978)
8. Garbarino, Sherman (1980)
9. Daro (1988)
10. Murphy, Jenkins, Newcombe, et al. (1981)
11. Lealman, Haigh, Phillips, et al. (1983)

Table 2

Reasons for Referrals to a Child Abuse Registry
(From Leventhal, Garber, Brady, 1989)

Categories	Reasons for Referrals
Care of Previous Children	Sibling abused
	Sibling neglected/failure-to-thrive/history of poor child care
Characteristics of Parents	Mental retardation
	Serious psychiatric problem
	Drug abuse
	Alcohol abuse
	Jail history
	History of violence
	Abused/neglected as a child
	Adolescent parent
Current Family	Violence in family
	Chaotic lifestyle
	Inadequate/overcrowded housing
	No provisions for baby at home
	No social support
Current Pregnancy	Unwanted/adoption considered/strongly ambivalent
	No prenatal care/delivery at home
Current Behaviors	Parents (mother) not providing physical care
	Fails to visit or inquire about infant
	Wants to leave hospital AMA with baby
	Obstructive behavior towards child's medical care
	Places child in dangerous situation
	Threats of violent behavior or actual violent behavior towards hospital staff

Table 3

Reasons for Referrals to a Child Abuse Registry
(From Ounsted, Roberts, Gordon, et al., 1982)

Categories of Reasons for Referrals
Doubt about parenting ability (attitude, actual harm to infant, indifference, rejection, display of temper)
Psychiatric history
Behaviors in hospital
Social and medical problems (housing, income, marriage)
Previous abuse of children
Maternal illness or handicap (includes low IQ)
Social problems alone
Miscellaneous (e.g. adoption considered)

Table 4

A Review of Selected Past Research on Risk Factors for Child Maltreatment

This table compares which studies support and do not support the listed items as risk factors for child maltreatment.

Item	Supports	Does Not Support
Parental Characteristics		
Poor Childhood Nurture	Altemeier et al., 1982 Monaghan,Gilmore, et al.,1986 Oates et al.,1979	
Abused as a Child	Egeland,1988 Council on Sci. Aff.,1985 Hunter et al.,1978	Widom,1989
Young Age	Stier,1989 Zuravin,1988 U.S. HHS,1988	Altemeier et al.,1982
Race		U.S. HHS,1988
Gen. Character	Monaghan,Gilmore, et al.,1986	Egeland,1979
Low Intelligence/ Education	Hunter et al.,1978 Seagull,Scheurer, 1986	Altemeier et al.,1982
Mental Illness	Council on Sci. Aff.,1985	
Alcohol Abuse	Council on Sci. Aff.,1985	Altemeier et al.,1984
Drug Abuse	Council on Sci. Aff.,1985	Altemeier et al.,1984
Low Self-Esteem	Altemeier et al., 1982 Evans,1980	
Aggressive	Altemeier et al., 1982 Evans,1980	
Violent	Avison,Turner Noh,1986	
Impulsive	Hunter et al.,1978	
Depression/Apathy	Evans,1980 Stern,1973 Hunter et al.,1978	
External Locus of Control	Ellis,Milner,1981 Evans,1980	

(continued)

Table 4, continued

Item	Supports	Does Not Support
Parenting Style		
Unreasonable Expectations of Child	Avison,Turner, Noh,1986 Council on Sci. Aff.,1985 Oates et al.,1979	Altemeier et al., 1984
Very Punitive	Avison,Turner Noh,1986	Evans,1980
Family Profile		
Low Income	U.S. HHS,1988 Council on Sci. Aff.,1985 Hunter et al.,1978 Oates et al.,1979	Daro,1988
Recent Job Change /Loss	Krugman et al., 1986	Steinberg,Catalano, Dooley,1981
Lack of Supports	Hunter et al.,1978 Oates et al.,1979	Altemeier et al., 1984
Social Isolation	Council on Sci. Aff.,1985 Stern,1973 Hunter et al.,1978	Altemeier et al., 1984
Life Stresses	Egeland,Breitenbucher,Rosenberg, 1980	
Chaotic Lifestyle	Egeland,Breitenbucher,Rosenberg, 1980 Oates et al.,1979 Evans,1980	
Poor Housing	Council on Sci. Aff.,1985 Oates et al.,1979	
City vs. Rural		U.S. HHS,1988
Spousal Conflict	Altemeier et al., 1985 Hunter et al.,1978	
Family Violence	Council on Sci. Aff.,1985 Widom,1989	
Health Problems	Oates et al.,1979	
>4 Children in Family	U.S. HHS,1988 Zuravin,1988	
Children Spaced Closely Together	Hunter et al.,1978	Zuravin,1988
Inadequate Child Care	Hunter et al.,1978	

(continued)

Table 4, continued

Item	Supports	Does Not Support
Family Profile, continued		
Children Have Different Fathers	Zuravin,1988	
Current Pregnancy		
Unplanned/ Unwanted	Altemeier et al., 1982 Oates et al.,1979 Zuravin,1987 Hunter et al.,1978	Lynch,1976
Poor Prenatal Care	Hunter et al.,1978	
Pregnancy/Delivery Complications	Altemeier et al., 1985	
Newborn		
Health		Egeland,1979
Spent Time in ICU	Hergenroeder et al.,1985	
Prematurity		Leventhal,Egerter, Murphy,1984
Low Birth Weight	Hergenroeder et al.,1985 Vietze et al.,1980 Stern,1973	Leventhal,Egerter, Murphy,1984
IUGR		Leventhal,Berg, Egerter,1987
Behavior/Temperament	Egeland,1979 Vietze et al.,1980	
Gender	U.S. HHS,1988	
"Wrong" Gender	Hunter et al.,1978	
Twins	Groothuis et al., 1982	
Birth Order ≥2	Hergenroeder et al.,1985	
In-hospital Behaviors		
Mother and Baby Separated	Hergenroeder et al.,1985	
Poor Parent-Child Interaction	Egeland,1979 Vietze et al., 1980	Starr,1987
Rejection/ Hostility	Avison,Turner, Noh,1986 Oates et al.,1979	
Poor Parenting Skills	Avison,Turner, Noh,1986 Egeland,1979	

(continued)

Table 4, continued

Item	Supports	Does Not Support
Out-of-hospital Behaviors		
More Exposure Between Parent and Child	Altemeier et al., 1982	

Table 5

Experts' 2 x 2 Contingency Table
of Paired GRS Ratings

		Expert #2 GRS Rating: High/Moderate Risk	Expert #2 GRS Rating: Low/No Risk	
Expert #1 GRS Rating	High/Moderate Risk	a=11	b=0	f_1=11
	Low/No Risk	c=3	d=18	f_2=21
		n_1=14	n_2=18	N=32

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=88%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=92%

Percent overall observed agreement (P_o)=91%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.80

Table 6

Experts' 2 x 2 Contingency Table of Paired CRS Ratings

		Expert #2 CRS Rating: High Risk	Expert #2 CRS Rating: Low Risk	
Expert #1 CRS Rating	High Risk	a=19	b=1	f_1=20
	Low Risk	c=4	d=8	f_2=12
		n_1=23	n_2=9	N=32

Percent agreement (High/Moderate Risk)$=\frac{2a}{n_1+f_1}=88\%$

Percent agreement (Low/No Risk)$=\frac{2d}{n_2+f_2}=76\%$

Percent overall observed agreement (P_o)=84%

Kappa$=\frac{2(ad-bc)}{n_1f_2+n_2f_1}=0.65$

Table 7

Experts' CRS Item By Item Agreements
(Ranked in descending order by $kappa_w$, then by P_o)

CRS	Item	$Kappa_w$	P_o	P_e
21	Cooperativeness-Mother	1.00	1.00	0.98
19	Care of Newborn by Mother	1.00	1.00	0.97
4	Illicit Drug Use-Mother	0.92	0.98	0.69
14	Housing	0.92	0.98	0.76
7	Criminal Record-Mother	0.88	0.98	0.83
17	Attitude	0.84	0.96	0.75
15	Provisions for Baby	0.80	0.95	0.77
9	Care as a Child-Mother	0.74	0.93	0.75
5	Alcohol Use-Mother	0.73	0.92	0.70
11	General Character-Father	0.71	0.88	0.58
18	Prenatal Care	0.69	0.88	0.62
1	Care of Older Sibs	0.62	0.92	0.80
12	Family Conflict	0.62	0.86	0.61
8	History of Violence-Mother	0.61	0.90	0.74
13	Stability of Couple	0.58	0.85	0.64
2	Intelligence-Mother	0.47	0.97	0.95
6	Motivation-Mother	0.44	0.85	0.72
10	Age-Mother	0.43	0.80	0.65
16	Social Supports	0.40	0.89	0.82
3	Psychiatric History-Mother	0.32	0.86	0.80
20	Visiting with Baby	0.15	0.86	0.84
22	Danger to Child in Hospital	*	1.0	*

* 100% agreement, all in cell "a" of contingency table; P_e (percent expected agreement) and kappa not computable

Table 8

Experts' 2 x 2 Contingency Table of CRS and GRS Ratings

		Experts' GRS Rating: High/Moderate Risk	Experts' GRS Rating: Low/No Risk	
Same Expert's CRS Rating	High Risk	a=25	b=19	f_1=44
	Low Risk	c=0	d=20	f_2=20
		n_1=25	n_2=39	N=64

Sensitivity=100%

Specificity=51%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=72%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=68%

Percent overall observed agreement (P_o)=70%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.45

Table 9

Nurses' GRS Ratings vs.
Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Low/No Risk	
Nurses' GRS Rating	High/Moderate Risk	a=4	b=0	f_1=4
	Low/No Risk	c=6	d=10	f_2=16
		n_1=10	n_2=10	N=20

Sensitivity=40%

Specificity=100%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=57%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=77%

Percent overall observed agreement (P_o)=70%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.40

Table 10

Nurses' CRS Ratings vs.
Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Experts' Consensus GRS Rating: Low/No Risk	
Nurses' CRS Rating	High Risk	a=7	b=2	f_1=9
	Low Risk	c=4	d=7	f_2=11
		n_1=11	n_2=9	N=20

Sensitivity=64%

Specificity=78%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=70%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=70%

Percent overall observed agreement (P_o)=70%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.41

Table 11

Pediatricians' GRS Ratings vs.
Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Low/No Risk	
Pediatricians' GRS Rating	High/Moderate Risk	a=5	b=1	f_1=6
	Low/No Risk	c=3	d=11	f_2=14
		n_1=8	n_2=12	N=20

Sensitivity=63%

Specificity=92%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=71%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=85%

Percent overall observed agreement (P_o)=80%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.57

Table 12

Pediatricians' CRS Ratings vs.
Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Low/No Risk	
Pediatricians' CRS Rating	High Risk	a=5	b=1	f_1=6
	Low Risk	c=3	d=11	f_2=14
		n_1=8	n_2=12	N=20

Sensitivity=63%

Specificity=92%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=71%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=85%

Percent overall observed agreement (P_o)=80%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.57

Table 13

Highest Non-expert GRS Ratings vs.
Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Low/No Risk	
Highest Non-expert GRS Rating	High/Moderate Risk	a=7	b=2	f_1=9
	Low/No Risk	c=6	d=12	f_2=18
		n_1=13	n_2=14	N=27

Sensitivity=54%

Specificity=86%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=64%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=75%

Percent agreement (Low/No Risk)=75%

Percent overall observed agreement (P_o)=70%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.40

Table 14

Highest Non-expert CRS Ratings vs. Experts' Consensus GRS Ratings

		Experts' Consensus GRS Rating: High/Moderate Risk	Experts' Consensus GRS Rating: Low/No Risk	
Highest Non-expert CRS Rating	High Risk	a=10	b=3	f_1=13
	Low Risk	c=3	d=11	f_2=14
		n_1=13	n_2=14	N=27

Sensitivity=77%

Specificity=79%

Percent agreement (High/Moderate Risk)=$\frac{2a}{n_1+f_1}$=77%

Percent agreement (Low/No Risk)=$\frac{2d}{n_2+f_2}$=79%

Percent overall observed agreement (P_o)=78%

Kappa=$\frac{2(ad-bc)}{n_1f_2+n_2f_1}$ =0.55

Table 15

Breakdown of Evaluations By Clinician Type

No Evaluations Returned	26 (7%)
At Least One Evaluation Returned	337 (93%)
Only a Nurse's Evaluation	41 (11%)
Only a Pediatrician's Evaluation	44 (12%)
Only a Social Worker's Evaluation	15 (4%)
Both a Nurse's and a Pediatrician's Evaluation	133 (37%)
Both a Nurse's and a Social Worker's Evaluation	28 (8%)
Both a Pediatrician's and a Social Worker's Evaluation	19 (5%)
Nurse's, Pediatrician's and Social Worker's Evaluation	57 (16%)
Total Evaluations by Nurses	259 (71%)
Total Evaluations by Pediatricians	253 (70%)
Total Evaluations by Social Workers	119 (33%)

Table 16
Demographic Profile

	Entire Study Group (N=363)	Those with Evaluations (N=337)	Those without Evaluations (N=26)
Race			
Black	67%	67%	63%
White	17%	18%	8%
Hispanic	16%	15%	21%
Oriental	1%	0%	8%
Religion			
None/Unknown	13%	13%	11%
Catholic	33%	32%	44%
Protestant	49%	51%	22%
Other	5%	4%	22%
Marital Status			
Single	79%	79%	76%
Married	12%	12%	20%
Separated	5%	6%	0%
Divorced	3%	3%	4%
Number of Prior Pregnancies			
None	22%	20%	40%
One or more	78%	80%	60%
Five or more	28%	28%	28%
Number of Prior Deliveries			
None	32%	31%	44%
One or more	68%	69%	56%
Five or more	4%	4%	0%
Current Gestation			
Premature	22%	21%	24%
Term	74%	74%	72%
Post-dates	5%	5%	4%
Number of Babies			
Singleton	96%	96%	92%
Twins	4%	4%	8%
Delivery			
Planned C-sec	6%	6%	8%
Crash C-sec	3%	3%	0%
NSVD	77%	77%	76%
Induced vag.	9%	9%	12%
C-sec after trial of labor	5%	5%	4%
Gender of Baby			
Male	49%	49%	52%
Female	51%	51%	48%
Feeding			
Breast	11%	12%	0%
Bottle	84%	84%	80%
Both	5%	4%	20%
Hospital Billing			
Self-Pay	5%	5%	5%
Insurance	14%	13%	26%
Public Aid	82%	82%	68%

Table 17

Risk Rating Assignments By Clinician Type

Clinician Type	N	No. High Risk	% High Risk
Nurse	259	82	32
Pediatrician	253	91	36
Social Worker	119	99	83
Highest CRS Rating	337	168	50

Table 18

Percentage of Times Each CRS Item
Was Unknown By Clinician Type

CRS Item Number	Expert	Nurse	Pediatrician	Social Worker
1	8.0	67.0	52.6	34.9
2	1.6	13.9	8.3	3.4
3	4.7	64.1	61.7	24.4
4	0	44.0	24.9	8.4
5	0	57.9	35.2	17.6
6	3.1	13.1	26.1	2.5
7	4.7	80.3	76.3	26.1
8	14.1	91.9	95.3	77.3
9	3.1	78.0	85.8	50.4
10	7.8	18.5	17.8	5.0
11	4.7	46.7	51.8	17.6
12	4.7	67.2	64.4	13.4
13	6.3	52.5	62.1	26.1
14	3.1	48.3	42.7	2.5
15	1.6	25.5	23.3	2.5
16	3.1	17.8	24.1	1.7
17	4.7	43.2	53.0	11.8
18	4.7	34.0	25.3	9.2
19	17.2	9.3	7.1	5.9
20	23.4	22.0	13.4	18.5
21	1.6	3.5	2.4	1.7
22	9.4	13.9	10.7	5.0
	N=64	N=259	N=253	N=119

Table 19

Items Ranked by Percentage Unknown By Each Clinician Type

Expert		Nurse		Pediatrician		Social Worker	
Item #	%	Item #	%	Item #	%	Item #	%
20	23.4	8	91.9	8	95.3	8	77.3
19	17.2	7	80.3	9	85.8	9	50.4
8	14.1	9	78.0	7	76.3	1	34.9
22	9.4	12	67.2	12	64.4	7	26.1
1	8.0	1	67.0	13	62.1	13	26.1
10	7.8	3	64.1	3	61.7	3	24.4
13	6.3	5	57.9	17	53.0	20	18.5
3	4.7	13	52.5	1	52.6	5	17.6
7	4.7	14	48.3	11	51.8	11	17.6
11	4.7	11	46.7	14	42.7	12	13.4
12	4.7	4	44.0	5	35.2	17	11.8
17	4.7	17	43.2	6	26.1	18	9.2
18	4.7	18	34.0	18	25.3	4	8.4
6	3.1	15	25.5	4	24.9	19	5.9
9	3.1	20	22.0	16	24.1	10	5.0
14	3.1	10	18.5	15	23.3	22	5.0
16	3.1	16	17.8	10	17.8	2	3.4
2	1.6	2	13.9	20	13.4	6	2.5
15	1.6	22	13.9	22	10.7	14	2.5
21	1.6	6	13.1	2	8.3	15	2.5
4	0	19	9.3	19	7.1	16	1.7
5	0	21	3.5	21	2.4	21	1.7
N=64		N=259		N=253		N=119	

Table 20

Items Ranked By Percentage Unknown:
Only Where Nurse and Pediatric Evaluations Both Returned

Nurse		Pediatrician	
Item #	%	Item #	%
8	90.2	8	98.5
7	75.2	9	87.2
9	72.9	7	77.4
1	64.0	13	65.4
3	60.9	12	63.2
12	60.9	3	60.9
5	55.6	11	55.6
13	54.1	17	53.4
11	48.9	1	50.6
4	48.1	14	42.1
14	45.1	5	36.8
17	41.4	6	32.3
18	40.6	4	30.1
20	27.1	18	26.3
15	25.6	16	23.3
10	17.3	15	18.0
16	15.0	10	15.0
2	13.5	20	10.5
6	10.5	2	8.3
22	8.3	22	7.5
19	3.8	19	5.3
21	2.3	21	0.8
N=190		N=190	

Table 21

Nurse-Pediatrician CRS Item By Item Agreements (N=190)
(Ranked in descending order by kappa$_w$, then by P_o)

CRS Item Number	Kappa$_w$	P_o	P_e
4*	0.76	0.91	0.62
3	0.75	0.97	0.89
1*	0.62	0.88	0.67
12*	0.55	0.95	0.89
11	0.50	0.81	0.62
5*	0.47	0.87	0.76
10*	0.42	0.81	0.68
17	0.40	0.89	0.82
18	0.39	0.77	0.61
13	0.35	0.86	0.78
19	0.29	0.95	0.93
6	0.29	0.79	0.71
15	0.25	0.91	0.88
14	0.21	0.89	0.86
20	0.20	0.83	0.79
2	0.18	0.88	0.86
21	0.13	0.97	0.97
22	0.06	0.96	0.96
9	0	0.89	0.89
8	0	0.80	0.80
7	-	-	-
16	-0.06	0.86	0.87

* Designates items with fair agreement on Tables 21, 22, and 23
P_e=Percent expected agreement

Table 22

Pediatrician-Social Worker CRS Item By Item Agreements
(N=76)
(Ranked in descending order by $kappa_w$, then by P_o)

CRS Item Number	$Kappa_w$	P_o	P_e
4*	0.73	0.88	0.57
5*	0.68	0.86	0.57
21	0.66	0.97	0.91
1*	0.65	0.84	0.53
19	0.64	0.92	0.78
11	0.53	0.80	0.56
12*	0.52	0.84	0.67
10*	0.48	0.83	0.67
18	0.48	0.83	0.68
3	0.43	0.88	0.79
20	0.32	0.80	0.70
13	0.30	0.79	0.70
15	0.30	0.78	0.68
2	0.27	0.88	0.84
6	0.27	0.75	0.65
14	0.24	0.84	0.79
17	0.22	0.86	0.82
9	0.22	0.79	0.73
16	0.20	0.80	0.75
22	0.08	0.91	0.90
7	0	0.96	0.96
8	-0.25	0.67	0.73

* Designates items with fair agreement on Tables 21, 22, and 23
P_e=Percent expected agreement

Table 23

Nurse-Social Worker CRS Item By Item Agreements (N=85)
(Ranked in descending order by $kappa_w$, then by P_o)

CRS Item Number	$Kappa_w$	P_o	P_e
8	1.00	1.00	0.80
10*	0.77	0.92	0.63
4*	0.61	0.82	0.54
1*	0.60	0.83	0.57
9	0.55	0.88	0.74
14	0.51	0.91	0.82
22	0.48	0.95	0.90
5*	0.47	0.80	0.63
12*	0.42	0.88	0.80
20	0.37	0.81	0.69
6	0.33	0.77	0.66
11	0.27	0.71	0.60
18	0.26	0.73	0.64
15	0.23	0.78	0.71
19	0.20	0.87	0.84
21	0.17	0.92	0.91
16	0.17	0.85	0.82
2	0.14	0.82	0.79
17	0.09	0.84	0.82
7	0	0.95	0.95
3	-0.08	0.84	0.86
13	-0.12	0.70	0.73

* Designates items with fair agreement on Tables 21, 22, and 23
P_e=Percent expected agreement

REFERENCES

References

Accardo PJ, ed. Failure to Thrive in Infancy and Early Childhood. Baltimore, Md: University Park Press; 1982.

Altemeier WA, O'Connor SM, Sherrod KB, Vietze PM. Prospective study of antecedents for nonorganic failure to thrive. J Pediatr. 1985;106(3):360-365.

Altemeier WA, O'Connor S, Vietze P, et al. Antecedents of child abuse. J Pediatr. 1982;100(5):823-829.

Altemeier WA, O'Connor S, Vietze P, et al. Prediction of child abuse: A prospective study of feasibility. Child Abuse Neglect. 1984;8:393-400.

Altemeier WA, Vietze PM, Sherrod KB, et al. Prediction of child maltreatment during pregnancy. J Am Acad Child Psychiat. 1979;18:205-218.

American Association for Protecting Children. Highlights of Official Child Neglect and Abuse Reporting 1985. Denver,Colo:The American Humane Association;1987.

American Association for Protecting Children. Highlights of Official Child Neglect and Abuse Reporting 1986. Denver,Colo:The American Humane Association;1988.

American Association for Protecting Children. Highlights of Official Child Neglect and Abuse Reporting 1987. Denver,Colo:The American Humane Association;1989.

Apgar V. A proposal for a new method of evaluation of the newborn infant. Anesth Anal. 1953;32:260-267.

Avison WR, Turner RJ, Noh S. Screening for problem parenting: Preliminary evidence on a promising instrument. Child Abuse Neglect. 1986;10:157-170.

Brody H, Gaiss B. Ethical issues in early identification and prevention of unusual child rearing practices. In: Helfer RE, Kempe CH, eds. Child Abuse and Neglect: The Family and the Community. Cambridge, Mass:Ballinger Publishing Co;1976:372-375.

Browne K, Davies S, Stratton P, eds. Early Prediction and Prevention of Child Abuse. Chichester,Engl:John Wiley and Sons;1988.

Browne K, Lowton M. Checklists for completion by midwives and health visitors to identify children at risk of child abuse. Final Report to the Leverhume Trust, July, 1987.

Browne K, Saqi S. Approaches to screening for child abuse and neglect. In: Browne K, Davies C, Stratton P, eds. Early Prediction and Prevention of Child Abuse. Chichester,Engl:John Wiley and Sons;1988:57-85.

Cicchetti DV. When diagnostic agreement is high, but reliability is low: Some paradoxes occurring in joint independent neuropsychology assessments. J Clin Exp Neuropsychol. 1988;10(5):605-622.

Cicchetti DV, Sparrow SS. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. Am J Ment Def. 1981;86(2):127-137.

Council on Scientific Affairs. AMA Diagnostic and treatment guidelines concerning child abuse and neglect. JAMA. 1985;254(6):796-800.

Daniel JH, Newberger EH, Reed RB, et al. Child abuse screening: Implications of the limited predictive power of abuse discriminants from a controlled family study of pediatric social illness. Child Abuse Neglect. 1978;2:247-259.

Daro D. Confronting Child Abuse: Research for Effective Program Design. New York,NY:The Free Press;1988.

Dash L. When Children Want Children: The Urban Crisis of Teenage Childbearing. New York,NY:William Morrow and Co Inc;1989.

Dean JG, MacQueen IAG, Mitchell RG, et al. Health visitor's role in prediction of early childhood injuries and failure to thrive. Child Abuse Neglect. 1978;2:1-17.

Dubowitz H. Prevention of child maltreatment: What is known. Pediatrics. 1989;83(4):570-577.

Dubowitz LMS, Dubowitz V, Goldberg C. Clinical assessment of gestational age in the newborn infant. J Pediatr. 1970;77(1):1-10.

Egeland B. Preliminary results of a prospective study of the antecedents of child abuse. Child Abuse Neglect. 1979;3:269-278.

Egeland B. Breaking the cycle of abuse: Implications for prediction and intervention. In: Browne K, Davies C, Stratton P, eds. Early Prediction and Prevention of Child Abuse. Chichester,Engl:John Wiley and Sons;1988:87-99.

Egeland B, Breitenbucher M, Rosenberg D. Prospective study of the significance of life stress in the etiology of child abuse. J Consult Clin Psychol. 1980;48(2):195-205.

Ellis RH, Milner JS. Child abuse and locus of control. Psychol Reports. 1981;48:507-510.

Evans AL. Personality characteristics and disciplinary attitudes of child-abusing mothers. Child Abuse Neglect. 1980;4:179-187.

Feinstein AR. Clinical Epidemiology: The Architecture of Clinical Research. Philadelphia:WB Saunders Co.;1985: 177-178, 186-187.

Feinstein AR. Clinimetrics. New Haven,Conn:Yale University Press;1987:ix, 3, 4, 11, 23-25, 33-37, 45-47, 93, 192-193.

Fleiss JL. Statistical Methods for Rates and Proportions. New York:Wiley;1981:4, 6, 212-236.

Franklin AW, ed. Child Abuse: Prediction, Prevention and Follow-up. Edinburgh,Engl:Churchill Livingstone;1978.

Frodi AM, Lamb ME. Infants at risk for child abuse. Infant Mental Health J. 1980;1(4):240-247.

Gabinet L. MMPI profiles of high-risk and outpatient mothers. Child Abuse Neglect. 1979;3:373-379.

Garbarino J, Crouter A. Defining the community context for parent-child relations: The correlates of child maltreatment. Child Dev. 1978;49:604-616.

Garbarino J, Sherman D. Identifying high-risk neighborhoods. In: Garbarino J, Stocking SH, eds. Protecting Children from Abuse and Neglect: Developing and Maintaining Effective Support Systems for Families. San Francisco,Calif:Jossey-Bass Inc;1980:94-108.

Gray J, Cutler C, Dean J, et al. Perinatal assessment of mother-baby interaction. In: Helfer RE, Kempe CH, eds. Child Abuse and Neglect: The Family and the Community. Cambridge, Mass:Ballinger Publishing Co; 1976:377-392.

Groothuis JR, Altemeier WA, Robarge JP, et al. Increased child abuse in families with twins. Pediatrics. 1982;70(5):769-773.

Hamill PVV, Drizd TA, Johnson CL, et al. Physical growth: National Center for Health Statistics percentiles. Am J Clin Nutr. 1979;32:607-629.

Hanson R, McColloch W, Hartley S. Key characteristics of child abuse. In: Franklin AW, ed. Child Abuse: Prediction, Prevention and Follow-up. Edinburgh,Engl: Churchill Livingstone;1978:39-52.

Harris JC. Nonorganic failure-to-thrive syndromes: Reactive attachment disorder of infancy and psychosocial dwarfism of early childhood. In: Accardo PJ, ed. Failure to Thrive in Infancy and Early Childhood. Baltimore,Md:University Park Press;1982:229-241.

Helfer RE. Basic issues concerning prediction. In: Helfer RE, Kempe CH, eds. Child Abuse and Neglect: The Family and the Community. Cambridge Mass:Ballinger Publishing Co;1976a:363-375.

Helfer RE. Early identification and prevention of unusual child-rearing practices. Pediatr Ann. 1976b;5:91.

Helfer RE. A review of the literature on the prevention of child abuse and neglect. Child Abuse Neglect. 1982;6: 251-261.

Helfer RE. An overview of prevention. In: Helfer RE, Kempe RS, eds. The Battered Child, 4th Edition. Chicago,Ill:The University of Chicago Press;1987a:425-433.

Helfer RE, Kempe CH. Child Abuse and Neglect: The Family and the Community. Cambridge, Mass:Ballinger Publishing Co;1976c.

Helfer RE, Kempe RS, eds. The Battered Child, 4th Edition. Chicago,Ill:The University of Chicago Press;1987b.

Hergenroeder AC, Taylor PM, Rogers KD, et al. Neonatal characteristics of maltreated infants and children. AJDC. 1985;139:295-298.

Hunter RS, Kilstrom N, Kraybill EN, et al. Antecedents of child abuse and neglect in premature infants: A prospective study in a newborn intensive care unit. Pediatrics. 1978;61(4):629-635.

Hutchinson TA, Boyd NF, Feinstein AR, et al. Scientific problems in clinical scales, as demonstrated in the Karnofsky Index of Performance Status. J Chron Dis. 1979;32:661-666.

Jagger J, Jane JA, Rimel R. The Glascow Coma Scale: To sum or not to sum? Lancet. 1983;ii:97.

Kempe CH, Silverman FN, Steele BF, et al. The battered child syndrome. JAMA. 1962;181:17-24.

Kirkpatrick JR, Youmans RL. Trauma Index: An aide in the evaluation of injury victims. J Trauma. 1971;11(8): 711-714.

Kotelchuck M. Child abuse and neglect: Prediction and misclassification. In: Starr RH, ed. Child Abuse Prediction: Policy Implications. Cambridge, Mass: Ballinger Publishing Co;1982:67-104.

Kramer MS. Clinical Epidemiology and Biostatistics: A Primer for Clinical Investigators and Decision-Makers. Berlin,Ger:Springer-Verlag;1988.

Krugman RD, Lenherr M, Betz L, et al. The relationship between unemployment and physical abuse of children. Child Abuse Neglect. 1986;10:415-418.

Lealman GT, Haigh D, Phillips JM, et al. Prediction and prevention of child abuse-an empty hope? Lancet. 1983;i:1423-1424.

Leventhal JM. Research strategies and methodologic standards in studies of risk factors for child abuse. Child Abuse Neglect. 1982;6:113-123.

Leventhal JM. Can child maltreatment be predicted during the perinatal period: Evidence from longitudinal cohort studies? J Rep & Infant Psychol. 1988;6:139-161.

Leventhal JM, Berg A, Egerter SA. Is intrauterine growth retardation a risk factor for child abuse? Pediatr. 1987;79(4):515-519.

Leventhal JM, Egerter SA, Murphy JM. Reassessment of the relationship of perinatal risk factors and child abuse. Am J Dis Child. 1984;138:1034-1039.

Leventhal JM, Fearn K, Stashwick CA. Clinical data used by pediatric residents to assess parenting. Child Abuse Neglect. 1986;10:71-78.

Leventhal JM, Garber RB, Brady CA. Identification during the post partum period of infants who are at high risk of child maltreatment. J Pediatr. 1989;114(3):481-487.

Lynch M. Risk factors in the child: A study of abused children and their siblings. In: Martin HP, ed. The Abused Child: A Multidisciplinary Approach to Developmental Issues and Treatment. Cambridge, Mass:Ballinger Publishing Co;1976:43-56.

Lynch M, Roberts J. Early alerting signs. In: Franklin AW, ed. Child Abuse: Prediction, Prevention and Follow-up. Edinburgh,Engl:Churchill Livingstone;1978: 28-37.

Lynch MA, Roberts J. Consequences of Child Abuse. London, Engl:Academic Press;1982.

McCarthy PL, Jekel JF, Stashwick CA, et al. Further definition of history and observation variables in assessing febrile children. Pediatrics. 1981;67(5):687-693.

McCarthy PL, Sharpe MR, Spiesel SZ, et al. Observation scales to identify serious illness in febrile children. Pediatrics. 1982;70(5):802-809.

MacKenzie CR, Charlson ME. Standards for the use of ordinal scales in clinical trials. Brit Med J. 1986;292:40-43.

McLaren DS, Read WCC. Classification of nutritional status in early childhood. Lancet. 1972;ii:146-148.

McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika. 1947;12(2):153-157.

Manciaux M, Deschamps G. Prediction of child abuse. Child Abuse Neglect. 1977;1:59-61.

Martin HP, ed. The Abused Child: A Multidisciplinary Approach to Developmental Issues and Treatment. Cambridge, Mass:Ballinger Publishing Co;1976.

Milner JS, Gold RG, Wimberly RC. Prediction and explanation of child abuse: Cross-validation of the child abuse potential inventory. J Consult Clin Psychol. 1986;54 (6):865-866.

Monaghan SM, Buckfield PM. Obstetrics and the family: Identification of mothers at risk for parenting failure and methods of support. Child Abuse Neglect. 1981;5: 27-32.

Monaghan SM, Couper-Smartt J. Experience of an anticipatory management programme for potential child abuse and neglect. Child Abuse Neglect. 1977;1:63-69.

Monaghan SM, Gilmore RJ, Muir RC, et al. Prenatal screening for risk of major parenting problems: Further results from the Queen Mary Maternity Hospital Child Care Unit. Child Abuse Neglect. 1986;10:369-375.

Murphy JF, Jenkins J, Newcombe RG, et al. Objective birth data and the prediction of child abuse. Arch Dis Child. 1981;56:295-297.

Murphy S, Orkow B, Nicola RM. Prenatal prediction of child abuse and neglect: A prospective study. Child Abuse Neglect. 1985;9:225-235.

Nelson KG. An index of severity for acute pediatric illness. AJPH. 1980;70(8):804-807.

Oates RK, Davis AA, Ryan MG, et al. Risk factors associated with child abuse. Child Abuse Neglect. 1979;3:547-553.

Olds DL, Henderson CR, Chamberlain R, et al. Preventing child abuse and neglect: A randomized trial of nurse home visitation. Pediatrics. 1986a;78(1):65-78.

Olds DL, Henderson CR, Tatelbaum R, et al. Improving the delivery of prenatal care and outcomes of pregnancy: A randomized trial of nurse home visitation. Pediatrics. 1986b;77(1):16-28.

Olds DL, Henderson CR, Tatelbaum R, et al. Improving the life-course development of socially disadvantaged mothers: A randomized trial of nurse home visitation. AJPH. 1988;78(11):1436-1445.

Orkow B. Implementation of a family stress checklist. Child Abuse Neglect. 1985;9:405-410.

Ounsted C, Roberts JC, Gordon M, et al. Fourth goal of perinatal medicine. Brit Med J. 1982;284:879-882.

Robert Wood Johnson Foundation Special Report. Child Abuse: Slowly Answers Begin to Emerge. Princeton, NJ: Robert Wood Johnson Foundation;1977.

Roberts J. Why are some families more vulnerable to child abuse? In: Browne K, Davies C, Stratton P, eds. Early Prediction and Prevention of Child Abuse. Chichester, Engl:John Wiley and Sons;1988:43-56.

Rosenberg NM, Meyers S, Shackleton N. Prediction of child abuse in an ambulatory setting. Pediatrics. 1982;70 (6):879-882.

Rowe DS, Leonard MF, Seashore MR, et al. A hospital program for the detection and registration of abused and neglected children. NEJM. 1970;282(17):950-952.

Schneider C, Hoffmeister JK, Helfer RE. A predictive screening questionnaire for potential problems in mother-child interaction. In: Helfer RE, Kempe CH, eds. Child Abuse and Neglect: The Family and the Community. Cambridge, Mass:Ballinger Publishing Co; 1976:393-407.

Seagull EA, Scheurer SL. Neglected and abused children of mentally retarded parents. Child Abuse Neglect. 1986;10:493-500.

Silver BJ, Goldston SE, Silver LB. The 1990 objectives for the nation for control of stress and violent behavior: Progress report. Public Health Reports. 1984;99(4): 374-384.

Slap GB, Connor JL, Wigton RS, et al. Validation of a model to identify young patients for lymph node biopsy. JAMA. 1986;255(20):2768-2773.

Soper HV, Cicchetti DV, Satz P, et al. Null hypothesis disrespect in neuropsychology: Dangers of alpha and beta errors. J Clin Exp Neuropsychol. 1988;10(2):255-270.

Starr RH. Child Abuse Prediction: Policy Implications. Cambridge, Mass:Ballinger Publishing Co;1982.

Starr RH. Clinical judgment of abuse-proneness based on parent-child interactions. Child Abuse Neglect. 1987;11:87-92.

Steele B. Psychodynamic factors in child abuse. In: Helfer RE, Kempe RS, eds. The Battered Child, 4th Edition. Chicago,Ill:The University of Chicago Press; 1987:81-114.

Steinberg LD, Catalano R, Dooley D. Economic antecedents of child abuse and neglect. Child Dev. 1981;52:975-985.

Stern L. Prematurity as a factor in child abuse. Hosp Prac. 1973;8(5):117-123.

Stier DM. The epidemiology of child abuse and neglect: Is young maternal age a risk factor? New Haven, Conn:Yale School of Medicine;1989. Thesis.

Straus MA, Gelles RJ, Steinmetz SK. Behind Closed Doors: Violence in the American Family. Newbury Park:Sage Publications;1981.

Straus MA, Gelles RJ. Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. J Marriage Fam. 1986;48:465-479.

Sweet AY. Classification of the low-birth-weight infant. In: Klaus MH, Fanaroff AA, eds. Care of the High-Risk Neonate, 2nd Edition. Philadelphia,Penn:WB Saunders Co;1979:74.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness. Lancet. 1974;ii:81-84.

U.S. Congress, Office of Technology Assessment. Healthy Children: Investing in the Future. Washington,DC:US Government Printing Office;1988:167-186.

U.S. Department of Health and Human Services. Better Health for Our Children; A National Strategy: The Report of the Select Panel for the Promotion of Child Health, Vols. I & II, DHHS(PHS)-79-55071. Washington, D.C.:US Government Printing Office;1981:110.

U.S. Department of Health and Human Services. Study Findings: Study of National Incidence and Prevalence of Child Abuse and Neglect: 1988: Washington, D.C.: US Government Printing Office;1988.

Vietze P, Falsey S, Sandler H, et al. Transactional approach to prediction of child maltreatment. Infant Mental Health J. 1980;1(4):248-261.

Volkmar FR, Cicchetti DV, Dykens E, et al. An evaluation of the Autism Behavior Checklist. J Autism Dev Disorders. 1988;18(1):81-97.

Widom CS. Child abuse, neglect and adult behavior: Research design and findings on criminality, violence and child abuse. Am J Orthopsychiat. 1989;59(3):355-367.

Zuckerman B, Amaro H, Cabral H. Validity of self-reporting of marijuana and cocaine use among pregnant adolescents. J Pediatr. 1989;115(5,1):812-815.

Zuravin SJ. Unplanned pregnancies, family planning problems and child maltreatment. Fam Relations. 1987;36:135-139.

Zuravin SJ. Fertility patterns: Their relationship to child physical abuse and child neglect. J Marr Fam. 1988;50:983-993.

HARVEY CUSHING / JOHN HAY WHITNEY
MEDICAL LIBRARY

MANUSCRIPT THESES

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the Medical Library are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but passages must not be copied without permission of the authors, and without proper credit being given in subsequent written or published work.

This thesis by has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

NAME AND ADDRESS DATE