Psychological Methods
2000, Vol. 5, No. 1, 102–124

1082-989X/00/$5.00 DOI: 10.1037//1082-989X.5.1.102

Neural Network Modeling of Risk Assessment in Child Protective Services

David B. Marshall and Diana J. English
Office of Children's Administration Research

The advantages of using neural network methodology for the modeling of complex social science data are demonstrated, and neural network analysis is applied to Washington State Child Protective Services risk assessment data. Neural network modeling of the association between social worker overall assessment of risk and the 37 separate risk factors from the State of Washington Risk Assessment Matrix is shown to provide case classification results superior to linear or logistic multiple regression. The improvement in case prediction and classification accuracy is attributed to the superiority of neural networks for modeling nonlinear relationships between interacting variables; in this respect the mathematical framework of neural networks is a better approximation to the actual process of human decision making than linear, main effects regression. The implications of this modeling advantage for evaluating social science data within the framework of ecological theories are discussed.

This article reports the use of neural network methods to model risk assessment by Washington State Child Protective Services (CPS) social workers in cases of child abuse and neglect. The background of and implications for risk assessment in CPS are covered in greater detail elsewhere, where we have reported results of the application of neural network modeling (English, Marshall, Brummel, & Coghlan, 1998; English, Marshall, Coghlan, Brummel, and Orme, in press). Our emphasis in this article is primarily an exploration of the utility of neural networks for the analysis of social science data, with a detailed explanation and interpretation of the neural networks constructed from risk assessment data as an example and, secondarily, a comparison with standard regression approaches. Following a brief outline of general issues in the modeling of human decision making, we summarize the context of our CPS risk assessment data and then discuss the neural network methodology used to analyze it. (After this article was submitted, a book by Garson, 1998, that introduces neural networks to social scientists came to our attention. We did not have the opportunity to review this book before our resubmission deadline, but readers of this article may also find Garson's book worth inspection.)

Only a few features of the vast research on human decision making germane to this study can be addressed here. (English et al., 1998, cover the decision-making literature in the field of child welfare risk assessment.) In the context of professional judgments

David B. Marshall and Diana J. English, State of Washington Department of Social and Health Services, Office of Children's Administration Research (OCAR).

This work could not have been accomplished without the generous support of the National Center for Child Abuse and Neglect and the taxpayers of the United States through the U.S. Department of Health and Human Services, Administration for Children, Youth and Families Discretionary Funds Grant 90-CA-1563. We also gratefully acknowledge the essential help of Tom Johnson of the CAMIS programming team and Matthew Orme of OCAR for providing the electronic download and unduplicated data set from CAMIS, and we acknowledge the invaluable assistance of Sherry Brummel of OCAR and Robert Novicky of Child Protective Services Tacoma, Region 5, in connecting abstract CAMIS variable names to actual CPS social worker practice. David S. Cordray provided very helpful and detailed comments on an earlier version of this article. We also thank Tim Clark of OCAR for providing feedback on a draft of the introduction and for enthusiastically supporting the application of neural network methodology to CPS data.

Correspondence concerning this article should be addressed to David B. Marshall, Department of Social and Health Services, Children's Administration, P.O. Box 45710, Olympia, Washington 98504-5710. Electronic mail may be sent to mdav300@dshs.wa.gov.

or diagnoses of human illness and behaviors, statistical models of collective judgments or experience ("actuarial" models) have been shown to be superior to clinical judgments in a wide variety of applications (Dawes, Faust, & Meehl, 1989). These models are often quite simple; a linear or logistic regression model or even a simple weighted scale with a handful of explanatory variables will correctly classify new cases at rates significantly higher than those achieved by individual judges (Dawes & Corrigan, 1974). Decision-making situations that may on the surface seem quite difficult in practice, such as the assessment of risk of future child abuse, reduce to simple prescriptions of a small number of statistically independent indicators such as size of the family or the presence of caretaker substance abuse (cf. Baird, 1988).

Dawes and Corrigan (1974) have reviewed linear decision-making models and outlined the conditions under which such models are expected to be superior to clinical judgements: when each input or explanatory variable has an independent, monotonic relationship to the outcome or decision variable, when deviations from the statistically optimized weights do not make much practical difference, and when there is error or noise in the measurements. However, the fit of a simple linear model does not preclude the possibility that a more complex model might fit better (provide higher classification accuracies) and thus provide a more complex picture of collective decision making. When interactions and nonlinear relationships between variables are expected or cannot be ruled out, it behooves the analyst to consider more complex models. These more complex details, as indicated by Dawes and Corrigan (1974), may well be "washed out" in the presence of high levels of measurement error or noise, and a simple linear description may well be the only one justifiable by the data. However, this possibility should not inhibit the analyst from testing more complex models, particularly in the absence of any sound theoretical justification for treating underlying physical, biological, or psychological processes as simple linear mechanisms. To this end, if more complex relationships do in fact govern decision-making processes, then taking steps to reduce noise, use clean data, and use large sample sizes in analysis will help avoid unwarranted oversimplifications of complex phenomena.

The Child Protective Services Washington Risk Matrix (WRM) is a 37-item list of risk factors intended to aid caseworkers in their assignment of an overall level of risk, following investigation of an allegation of child abuse or neglect. The overall risk rating is intended to represent the risk to the child of future abuse or neglect in the absence of a "successful" agency intervention. Each individual risk factor and the overall risk rating is defined on a six-point ordinal (Likert-type) scale ranging from 0 (*no risk*) to 5 (*high risk*). The risk factors were selected by a CPS task force with representation from line social workers, researchers, supervisors, and CPS policy makers, all of whom were informed by current CPS practice, clinical experience, and current research. The WRM was intended as an ecological, multidomain perspective on the complex phenomenology of child maltreatment; it was designed to guide but not override clinical judgment. As such, the WRM was deliberately not intended as any sort of simple actuarial or scoring tool: The scores of the individual risk factors are not mathematically combined in any way to arrive at an overall risk level. Instead, the WRM is intended as a mnemonic and holistic tool to aid the social worker in clinical judgment of overall risk.

Like other risk assessment instruments in current use nationally, reliability and validity (construct and concurrent) have not been thoroughly investigated. Early attempts to validate risk assessment models were confounded by poor implementation of these models in actual practice (Doueck, English, DePanfilis, & Moote, 1993). When assessing interrater reliability of the WRM and other models, Fluke (1993) found that the overall scale reliability of the WRM was generally considered good, but the actual performance of the model did not live up to the potential.

English and Aubin (1991) found that CPS staff were accurate in assessing risk at the initial stage of referral screening: Referrals judged low risk at intake had a 6-month rereferral rate of less than 5%, with fewer than 1% of the children actually harmed. Other evaluators of the WRM model have produced conflicting reliability results. McDonald & Marks (1989) found that the WRM correctly classified 80% of nonrecidivists and 86% of recidivists. This rate compared favorably with reliability measures of the actuarial models (Baird, 1988; Johnson & L'Esperance, 1984). However, Camasso and Jagannathan (1995) found that the WRM predicted case closing, case substantiation, and recidivism at a greater level than chance, but they concluded that the error rate was too high to recommend exclusive reliance on this or any other risk assessment instrument for predictive purposes.

The risk factors are therefore not expected to formally or systematically predict overall risk, and in our analysis we make no efforts or claims to build predictive models. Our research interest is in using whatever multivariate techniques are most appropriate to model the associations between the individual risk factors and the overall risk rating in an effort to learn which risk factors or combinations of risk factors in what circumstances might weigh most heavily in caseworkers' assignments of overall risk ratings. We have investigated a variety of categorization and classification models to accomplish this end, including cluster analyses, principal-components analysis (classical and nonlinear), discriminant analyses, logit and probit regression, loglinear analyses, classification and regression trees (CART), and neural networks. Because neural networks and logistic regression are the methods most suitable for the data distributions and modeling issues discussed here and because they give theoretically sensible results, we restrict our attention in this article primarily to a comparison of these two methods.

There are CPS policy and practice reasons why the risk factors may have both ordinal and categorical properties. (From a measurement and statistical analysis point of view, this mixture of properties is unfortunate.) As noted above, the risk factors are defined as a 6-point ordinal scale, from no risk to high risk. However, the zero risk level is also given special emphasis in policy and training as a separate category called "family strengths" or "protective factors" that are supposed to represent a reduction in the risk from other risk factors that would otherwise obtain in the absence of these family strengths. In addition, there are policy guidelines that promote a dichotomous division between overall risk levels 0–2 and risk levels 3–5. This may naturally lead to all risk factors being treated in a dichotomous, categorical fashion by caseworkers (see English et al., 1998, for details).

For dichotomous outcomes, logistic regression (Hosmer & Lemeshow, 1989) is the usual statistical method of choice, particularly when the explanatory variables are categorical or nonnormally distributed. The logistic equation expresses the probability of an event (outcome) as

$$\text{Prob(outcome)} = e^Z/(1 + e^Z),$$

with Z equal to a linear combination of parameters (weights) and p explanatory variables:

$$Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_p X_p.$$

The resulting model is easily interpreted. The logistic function $e^Z/(1 + e^Z)$ transforms a binary outcome into a sigmoidally shaped probability defined on the continuous interval 0 to 1. With this transformation, a regression framework can be applied to binary outcomes. The exponential of each of the parameters $\beta_i(e^{\beta_i})$ is the odds ratio with respect to the outcome variable; for example, the ratio of the probability of observing the outcome given that the explanatory variable is rated high (or the factor is present) to the probability of observing the outcome given that the explanatory variable is rated low (or the factor is absent). Odds ratios (βs) for a given explanatory variable give the effect of that variable on the outcome while controlling for the effects of all other explanatory variables in the model. (The values of β for a given explanatory variable will change if different explanatory variables are included; the odds ratios are model-dependent.) Interaction terms involving more than one explanatory variable can also be tested in the models (e.g., multiplicative terms of the form $\beta_{23}X_2X_3$), but typically extremely large numbers of cases must be studied to provide sufficient statistical power to detect such interactions (McClelland & Judd, 1993). The logistic regression approach, like multiple linear regression, thus emphasizes the main effects of individual variables in explaining observed outcomes and assumes a simple additive relationship between those main effects.

Because of this basic construction, multiple regression (logistic or linear) in general has weak power for the detection of interactions in the presence of main effects. The situation is even worse for field studies in the social sciences; interactions are rarely detected, although interactions are frequently found in experimental studies. McClelland and Judd (1993) described this phenomenon and reviewed reasons for it such as measurement error and nonlinearities. In addition, they demonstrated that tests of interactions with field data will typically have less than 20% of the detection efficiency of optimal experimental tests solely because of the differential residual variances of interactions once the main effects have been partialed out. The general phenomenon of "variance stealing," the loss of detection power even for causally superior variables as a result of partialing out of (i.e., controlling for) other effects, has been reviewed and explored for child abuse effects research by Briere (1988).

This problem is particularly acute for situations where a model must choose a smaller subset of variables from a larger list, particularly when multicollinearity exists. Multiple regression is particularly vulnerable to this problem (Bieber, 1988), but all regression methods that rely on examining covariance in a pairwise fashion tend to overselect explanatory variables, emphasize main effects at the expense of interactions, and often pass over causally related variables for secondary or redundant ones (Spirtes, Glymour, & Scheines, 1993). The general problem of variable or feature selection (where features can be a single variable or some combination of variables constructed via a method such as nonlinear principal-components analysis) remains a very difficult one and is the subject of intense research (cf. Dash & Liu, 1997).

The neural network approach represents a conceptual departure from modeling methods that fit data to a preselected mathematical form. Instead of assuming a particular form of relationship between explanatory and outcome variables then using a fitting procedure to adjust the size of the parameters in the model, neural networks construct a unique mathematical relationship for a given data set on the basis of the desired match of explanatory variables to outcomes. How is this achieved? At a fundamental mathematical level, neural networks are an example of the representation of complex functions via a set of simpler functions. Many such methods exist in all fields of the natural sciences and statistics (basis set representations, projections, etc.); what distinguishes the neural network approach is a conceptual and computational framework for readily accomplishing this representation. This framework, inspired by rather simple analogies with the operation of neurons in biological systems (Peretto, 1992), has proven to be a very general and powerful way of thinking about and solving problems in data modeling and pattern recognition (cf. Holden & Kryukov, 1991a, 1991b, 1991c). Neural networks are also used as primitive maps of the human cognitive process (Holden & Kryukov, 1991b).

The computational strategies for implementing this framework and the computational goals borrow heavily from statistics and other techniques. Indeed, by controlling the architecture of a neural network (discussed below), one can constrain the neural network to produce an output for a particular problem that is identical to that produced by multiple linear regression, or logistic regression, or loglinear analysis, or discriminant analysis, or principal-components analysis, or nonlinear principal-components analysis, or robust regression, or *k*-means clustering, or probit regression, and so forth. Alternatively, using one of the many available adaptive neural network algorithms that construct an architecture from characteristics of the data allows the researcher faced with a complex, unknown situation to rapidly search and screen a very large number of potential models and relationships between input (independent) and output (dependent) variables. The resulting neural network solution can then often suggest which more formal, mathematically defined method would be most suitable to apply. We now routinely use neural network analysis to identify potential interaction terms for entry into generalized linear and additive models and have recently detected theoretically critical interaction terms that would have been otherwise overlooked (unpublished data). For balanced and insightful presentations of the relationships between neural networks and statistics, see Cheng and Titterington (1994), Bishop (1995), or Ripley (1996).

No general definition of neural network exists in the literature. This is in part due to the explosive growth of the field and the enormous variety of networks and algorithms that have been and are being developed. However, some typical characteristics of most neural networks include extensive connectivity of simple processing elements, adaptive construction of these processing elements and connections via learning from examples with known results (called a training set), and easy implementation of massively parallel processing for the computations. Not all neural networks share all these features nor are any of these features unique to neural networks. For alternative definitions and conceptions and an overview of the vast array of available methods, see Sarle (1997b). The two most common variants are multilayer perceptron networks (MLP NNs) and radial basis function networks (RBF NNs). RBF NNs are an example of kernel or nearest-neighbor (distance) classification methods and are useful when different input–output relationships are expected for different regions (subsets) of the data. Many other applications and variants of neural networks also exist, some without common statistical analogs, for example, Kohenon self-organizing maps (cf. Sarle, 1997b). MLP NNs are the most generally useful for constructing regression-like relationships, where one wishes to describe a single relationship between inputs and outputs. An MLP NN approach was used to describe the decision-making data in this article; thus, we will limit further description of neural networks to MLP NN.

An example of MLP NN architecture is shown in Figure 1, where inputs are connected through two hidden nodes in a single hidden layer to a single output, as well as each being directly connected to the

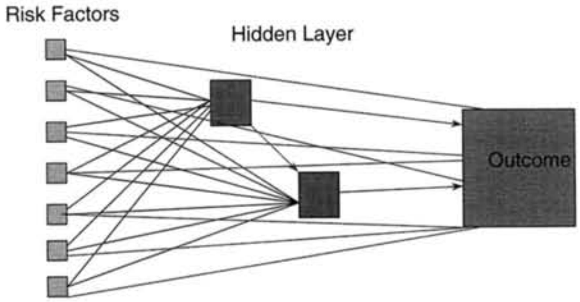

Figure 1. Example of multilayer perception neural network architecture.

output. (For simplicity, bias terms are not shown.) If there were no hidden nodes, that is, all inputs were just directly connected to the output, then the neural network would be identical to a multivariate regression model with main effects and no interactions. If, in addition, the output transfer function was the identity (× 1), then the network would correspond to a linear regression; if the output had a logistic transfer function, then it would correspond to logistic regression; and inputs with fixed transformations or smoothing functions (different for each input) would correspond to a generalized linear model (GLM), with an output activation function equal to the inverse of the GLM link function.

The direct connections are analogous to the main effects in regression, and the hidden nodes to interaction terms; however, the interaction represented by hidden nodes can be far more complex than the simple two-way multiplicative interaction terms that are usually included in regression equations: Each one of the input variables contributes (with its individual weight) to each hidden node. Further interaction results from the weighted contributions of each previous hidden node to subsequent hidden nodes and to the outcome node. (Detailed examples are given in the Results.) If the data variable relationships are, for example, dominated by a single two-way interaction, then the resulting network will show a single hidden node with large weights for the two interacting input variables and much smaller weights for the remaining variables. In effect, the network reduces to a simple regression model with main effects and a two-way interaction.

An MLP NN with a single hidden node (unit) can represent any quadratic relationship between inputs and output; an MLP NN with two hidden nodes can represent any cubic relationship. In fact, it has been shown that an MLP NN with a single hidden layer is capable of representing any continuously differentiable, arbitrarily complex function (cf. Bishop, 1995). This mathematical flexibility, in a conceptually different framework, is shared by the statistical technique of projection pursuit regression (Friedman & Stuetzle, 1981); also see Ripley (1996) for a very informative comparison of projection pursuit regression and neural networks. Because of the underlying architecture and mathematical flexibility of methods such as neural networks (or projection pursuit regression), which are based on multiple weighted contributions from all input variables, they are capable of capturing far more complex variable interactions than simpler regression models, if they exist in the data.

There are some implications of this flexibility compared with regression for the modeling of human decision making. The use of simple linear models of decision making, though demonstrably useful, is somehow not particularly satisfying given our current state of knowledge concerning the nonlinear dynamic character of many macroscopic biological and physical processes. Complex underlying mechanisms will appear simple in the presence of high noise levels, and a successful fit of a simple model to noisy data does not in itself indicate that the relationships between variables are simple. A general, flexible modeling technique such as neural networks can help test for the possibility of underlying complexity and will reduce to simpler models if warranted by the data or if the true, more complex mechanisms are masked by high noise. The exclusive use of simple linear models for the evaluation of social science data becomes a self-fulfilling prophecy that human decision making is simple. As noted earlier, the close correspondence between neural network architecture and the functional behavior of biological neurons and simple cognitive models (Holden & Kryukov, 1991b, 1991c; Peretto, 1992) suggests that the linear approach does not adequately capture the underlying mathematical structure of human cognition.

The actual calculation and construction of MLP NN uses standard numerical optimization techniques and thus shares many computational details with statistical and other methods. MLP NN are so-called feed-forward networks (information is passed from inputs to outputs), and the weights are usually calculated via the method of back-propagation (the difference between actual and predicted outputs (error) is used to adjust the weights, moving backwards from outputs to hidden node weights and then to input weights). An iterative procedure is used until some fitting criterion is optimal or reaches a chosen stopping point. Although a full consideration of these issues is well beyond the scope of this article, it is important to understand here one basic aspect of these methods: how MLP NN differ from statistical regression to a fixed mathematical form (such as the logistic regression equation) in fitting data in the presence of noise.

All methods that model relationships between variables in the presence of noise attempt to strike a balance between bias (the difference between the true relationship in the population and the modeled or fitted relationship in the sample) and variance (the difference between the model and the sample data points). Generalization, or the ability of a model derived from a sample to describe new samples, is a function of both bias (squared) and variance. Both overfitting (low bias, high variance) and underfitting (high bias, low variance) result in poor generalizability. Simple statistical regression techniques can over- or underfit if too many or too few explanatory variables are included in a model, but given a fixed set of variables and a chosen functional form, the parameters are adjusted in the model to minimize the variance. Thus, bias is addressed in the model and variable selection stage, and variance is minimized in the model fitting stage. It is appropriate in a standard statistical modeling context to discuss the goodness of fit and to compare goodness-of-fit criteria of nested models (with the bias implicit in the model selection process usually ignored).

Similar considerations apply to neural networks, of course; some error criterion must be chosen to construct the network, and a variance or some similar measure can be calculated. The bias–variance tradeoff has a rather different flavor in the case of neural networks, and somewhat different computational considerations apply (Gieman, Bienenstock, & Doursat, 1992). Because MLP NNs are capable of approximating any (continuously differentiable) function of an arbitrary degree of complexity, they are capable of severe overfitting, and many strategies have been used to avoid this (cf. Bishop, 1995, or Ripley, 1996). As a consequence, although goodness-of-fit criteria are used in the network construction process, and these criteria can be used to compare networks, the notion of a goodness-of-fit in neural networks is rather arbitrary, because the operator or analyst typically has more control over how good the fit will be. It is thus not appropriate to directly compare goodness of fit for a neural network and a logistic regression model of the same data, because the neural network solution can always be manipulated to provide a superior fit. A more honest and direct means of comparison is between the classification or prediction accuracies of the models, with the ultimate practical test being the differences in model generalizability, that is, which model more accurately classifies new cases (the procedure followed here).

Selection of a data-modeling method inevitably involves a trade-off between accuracy and the complexity and interpretability of the model solution. For descriptive purposes, and in keeping with the scientific principle of parsimony, one usually prefers the simplest representation of the data that provides acceptable accuracy. This additionally tends to ensure good generalizability of the model to new samples from the same population. Studies of the performance of neural networks versus statistical methods and classification trees in many different fields indicate that neural network models are usually more accurate, have higher internal validity, and generalize better to new data (cf. Bishop, 1995), including social science data (Garson, 1991; Wilson & Hardgrave, 1995). The increased accuracy and increased ability to accommodate complicated relationships and interactions comes at the price of easy interpretability. One cannot examine the transfer functions and weights of a neural network solution and reduce them to any simple prescription such as "If risk factors *X*, *Y*, and *Z* are rated high, there is an 80% probability that the abuse allegation will be substantiated" or "If risk factor *X* is rated high, the odds of placement increase by a factor of two over those cases where *X* is rated low." (Although this common sort of interpretation is also subject to error: Logistic regression odds ratios are model-dependent and thus should not be reported or transported to new situations without reference to the entire model.)

Sensitivity analysis (described below) does provide magnitudes and variances that allow one to state (approximately) which risk factors in a given neural network have relatively larger effects and can therefore be designed as more important, just as dichotomous factors that are significant in a logistic regression equation with odds ratios above two are typically considered important. However, there is no single parameter such as an odds ratio or node of a decision tree that describes the magnitude of the effect as simply and quantitatively, independent of other factors in the model. One also abandons a standard mathematical prescription for the relationships between explanatory and outcome variables that may (or may not) have some theoretical justification. However, it is likely that the mathematical forms traditionally assumed for social science and child welfare models are overly simplistic for much real-world data. Given the nature and purpose of CPS risk assessment, it is possible that the neural network approach is a closer approximation to the complex process of human decision making than the simpler models of linear or logistic regression equations, particularly if as in Washington State, multiple risk factors are intended by policy and training to be assessed holistically to guide CPS decision making. The superior results claimed for simpler linear models of human decision making in the literature (Dawes & Corrigan, 1974; Dawes, Faust, & Meehl, 1989) may in part be due to the lack of comparison with more mathematically sophisticated models or an artifact of the extremely high noise levels in small-sample social science data.

Method

Data Set

The CPS decision-making data set used in this report consists of all cases of child abuse and neglect reported to CPS in the State of Washington from July 1, 1994, to June 30, 1995, that also met certain selection criteria: accepted for investigation; a summary risk assessment completed by September 30, 1995, or not longer than 240 days after the referral received date; reasonably complete risk factor information; reported for only one type of (listed) abuse or neglect; not reviewed by supervisory review or transferred to other agencies. Restriction to cases with completed summary assessments essentially indicates that these cases are higher risk, that is, that are considered among the 40,000+ referrals received annually by CPS to warrant a full CPS investigation. After unduplication of records and removal of spurious referrals, referrals with large amounts of missing data, and certain administrative review or other special status referrals, the final number of actual CPS-investigated referrals for the data set was 12,978. Extensive bivariate tests at each stage of record removal of the demographic composition and decision status showed no substantial differences with the total number of referrals. A complete description of the selection criteria is given in English et al. (1998).

Analysis for different types of abuse was also conducted but not reported here; a separate study involving the reading and hand-coding of referral narrative information revealed that the type of abuse variable as reported in the Washington data system is incomplete and unreliable. Modeling with complete type of abuse information, however, reveals that there is little that distinguishes the relationship between overall risk and the risk factors for different types of abuse, other than the appearance of the obvious evidentiary risk factors in the corresponding models—for example, sexual abuse/exploitation for sexual abuse cases, physical injury/harm for physical abuse cases, (lack of) supervision for physical neglect cases. Otherwise, the key risk factors are the same across all types of abuse (English et al., 1998; English, Marshall, Coghlan, Brummel, & Orme, in press). One possible explanation of this commonality, despite the great emphasis in the maltreatment literature on specifying and studying separate types of abuse and neglect, is our frequent observation that the vase majority of children who are repeated victims of maltreatment suffer from a wide variety of forms of maltreatment; for example, neglect is not an isolated phenomenon in a subset of abuse victims but is very frequently accompanied or followed by physical, emotional, and sexual abuse (unpublished data).

Description of Variables

The input variables considered in this study are the 37 separate risk factors assessed by social workers for each investigated CPS case (see Appendix A for a complete list), the overall risk rating whose assignment the risk factors are intended to guide, and a case status variable that can take on one of two categories: initial or closure. A total of 2,170 cases, or 17% of the total data set, have initial case status, and 10,808, or 83%, have closure status. The actual meaning of this variable as used in the field is unclear; most cases with an initial assessment do not ever receive another one; most cases with closure assessments do not have previous initial assessments, and there is little correspondence between closure assessment status and actual case closure (which we have only recently been able to track; most cases in this data set do not have a formal indication of case disposition). Case status and reasons for its inclusion in analysis are discussed further in English et al. (1998). It was found that the multivariate models were not improved with the inclusion of any other CPS practice variable, intake information, or demographic information such as gender, age, or ethnicity; the risk factors fully contain the variance described by these factors.

Selection of Variables, Imputation, and Variable Collapses

The actual empirical use and statistical properties of 6-point ordinal risk scales as used in field risk assessment are almost completely unknown. Therefore, exploratory neural network and logistic regression modeling was performed using a number of different forms of the WRM variables to determine which collapse or combination of information would yield the best outcome classification accuracy (best agreement with the known assignment of overall risk level). These included ordinal and categorical collapses into 3- and 4-point ordinal scales, categorical collapses into dichotomous variables (3–5 risk vs. everything else), and 3-, 4-, and 5-category variables. The not applicable label was variously collapsed with either zero risk or low (1–2) risk. The insufficient information to assess label was collapsed with low (1–2) risk or retained as a separate category. The higher risk ratings were either collapsed together (3–5) or moderate risk was kept as a separate category or ordinal point, but still collapsing moderately high and high risk ratings together. All of these combinations were compared with the complete 6-point ordinal scale (0–5 risk), which includes imputed values for missing data and variables labeled not applicable or insufficient information to assess. The rationale for and details of the imputation strategies used and their relative efficacy (assessed by the classification accuracy and goodness-of-fit statistics of the models containing them) are given in English et al. (1998).

On the basis of considerations detailed in English et al. (1998), modeling was performed with both the 6-point ordinal scale and (in separate models) the four-category categorical version. A simpler 3-point ordinal scale was also used for classification models for dichotomous overall level of risk. This 3-point scale was constructed as 0 = *no risk*; 1 = *1–2 risk, not applicable, missing, and insufficient*; and 2 = *3–5 risk*. This simpler ordinal scale preserves the separate function of the zero risk level designation and also has the great advantage of avoiding all problems of imputation, allowing for the possibility of simple field deployment of the classification model as a caseworker decision-making aid, at the disadvantage of the usual loss of information when collapsing scales and a slight deterioration in classification accuracy. English et al. (1998) also discussed the use of risk factor summary measures as input features for modeling: the mean of all risk factors, the number of risk factors rated 3–5 risk, scores from oblique rotation factor analysis, and scores from nonlinear principal-components analysis.

Neural Network Construction

Neural network models were constructed using the PREDICT software program (1997). This program constructs a feed-forward MLP NN that uses back propagation to adjust the weights on the inputs and the nodes of a single hidden layer for the best classification accuracy of the outputs. A genetic programming algorithm is used to prepare subsets of variables for testing, selecting those variables that have the highest correlation with the outputs and selecting a data transformation for each variable that also optimizes the correlation. In the genetic selection algorithm, random subsets are tested for suitability as classifying variables. The more successful subsets are allowed to "reproduce" (grow or change by the addition or deletion of variables in each subsequent generation) and the less successful are eliminated from subsequent testing, or "killed." Random changes ("mutations") are also used to avoid "genetic drift" into variable sets that are the most successful only in a local-optimum sense (cf. Dash & Liu, 1997).

There are two basic ways in which PREDICT performs the transformation and selection at each generation: using logistic regression as a preprocessing stage, or constructing an actual neural network at each evolutionary stage of the genetic algorithm. After either of these, the final neural network is constructed using adaptive gradient descent or Kalman filtering (for the highest noise levels) to find and select the neural network weights and the number of nodes in the hidden layer. Preprocessing using logistic regression is much faster (1–4 hours versus 10–40 hours for each network on a 100 MHz Pentium computer), but preprocessing with neural networks is required if one is faced with a highly nonlinear problem.

Greater model stability (assurance of convergence to a global minimum) was obtained by adding a stochastic component to the direction of the gradient search procedure; this helps avoid local minima (nonglobal solutions) in the gradient descent algorithm. All neural network models were constructed using this option. All neural network models were also constructed using the most comprehensive set of available variable transforms, and the final models are the "winners" in an exhaustive network search method that constructs and compares several competing networks. Data were declared "very noisy" to avoid overfitting and thus poor generalization. The PREDICT program splits the total data set into training and test portions (the default percentages of 70 and 30, respectively, were used). A composite test score is computed across three (default) cross-validation partitions of the training data. Finally, for classification problems using dichotomous outcome variables, Gaussian and hyperbolic tangent transfer functions were included as possible transfer functions for the hidden layer nodes; Gaussian functions often give better classification results when the outcome variable is limited to a small number of categories.

PREDICT applies sensitivity analysis to help determine the relative importance of input variables in neural networks. Sensitivity analysis proceeds by testing the final model for each case, by "jittering" or imposing small fluctuations on the value of an input while holding all other inputs constant and observing the effect on the output(s). This procedure is performed for each case and each input variable. Average (and average squared) output fluctuations are calculated over all cases for each input, as well as the variances of these averages. These average fluctuation magnitudes correspond to the partial derivatives of the output with respect to each input, with all other inputs constant. The partial derivatives are more interpretable than the network weights (Sarle, 1997a) because they approximate the overall influence of each input on the output through both its direct connection (weight) and its indirect connections via the hidden nodes. However, it should be kept in mind that they are an approximation: To enable the calculation of partial derivates, the magnitudes of the imposed jitter must be kept rather low, often smaller than the natural range of the input variables in the data set. In addition, it is an inherently artificial situation that only one input variable would change in going from case to case; a principal motivation for the use of neural networks in the first place is to model situations where many input variables change and interact in complex ways. As in the case of β coefficients or odds ratios in linear or logistic regression, it is dangerous and unwarranted to transfer or report individual coefficient values out of the full model context. For an overview of the interpretive pitfalls of various measures of input importance in linear regression and neural networks, see Sarle (1997a).

Results and Discussion

Frequencies and Distributions

A typical distribution of risk factor values for this data set is shown in Figure 2, and Figure 3 shows the distribution for overall risk level.

All of the risk factors have a similar distribution, with the exception of child age risk level and self-protection, which are skewed in the opposite direction, access to/responsibility for child, which is bimodal with peaks at risk levels 0 and 5, and stress on caretaker, which has nearly equal and higher frequencies for risk levels 0–3 than the frequencies for risk levels 4–5. When the risk factors were used in linear regression calculations (see below), they were entered as raw values or as appropriately transformed values (log, inverse square root) to more closely approximate normal distributions (the neural network software tests and performs these transformations automatically).

The actual frequencies for overall risk level are reported in Table 1. As a dichotomous variable, overall risk 0–2 represents 10,283, or 79%, of all cases; overall risk 3–5 represents 2,695, or 21%, of all cases.

Bivariate Associations

Overall risk rating is significantly associated with all the risk assessment matrix variables in the expected positive direction (higher overall risk associated with higher risk for individual risk assessment matrix [RAM] variables). These associations are far stronger than the associations of overall risk with any of the other CPS variables with the exception of substantiation, case disposition, and case status. Association between overall risk rating and the RAM variables was investigated in several different ways: comparing the entire ordinal scales of overall risk and the RAM variables (with insufficients and not applicables recoded to missing) using an unsaturated loglinear association model; through cross-tabulations using the saturated loglinear method with dichotomized overall risk as the dependent variable and the original RAM variables treated as categorical variables (to explicitly include missings, insufficients, and not applicables as distinct categories) and with the RAM variables treated as ordinal scales (with insufficients and not applicables recoded to missing); in addition, via 2 × 2 cross tabulations of dichotomized overall risk and dichotomized RAM variables (all dichotomized as 0–2 vs. 3–5) to allow calculation of odds ratios. All these methods gave consistent results; all RAM variables have a significant positive association with overall risk and those with the strongest associations were roughly the same regardless of method. (This suggests that insufficient and not applicable labels do not have a major effect on the as-

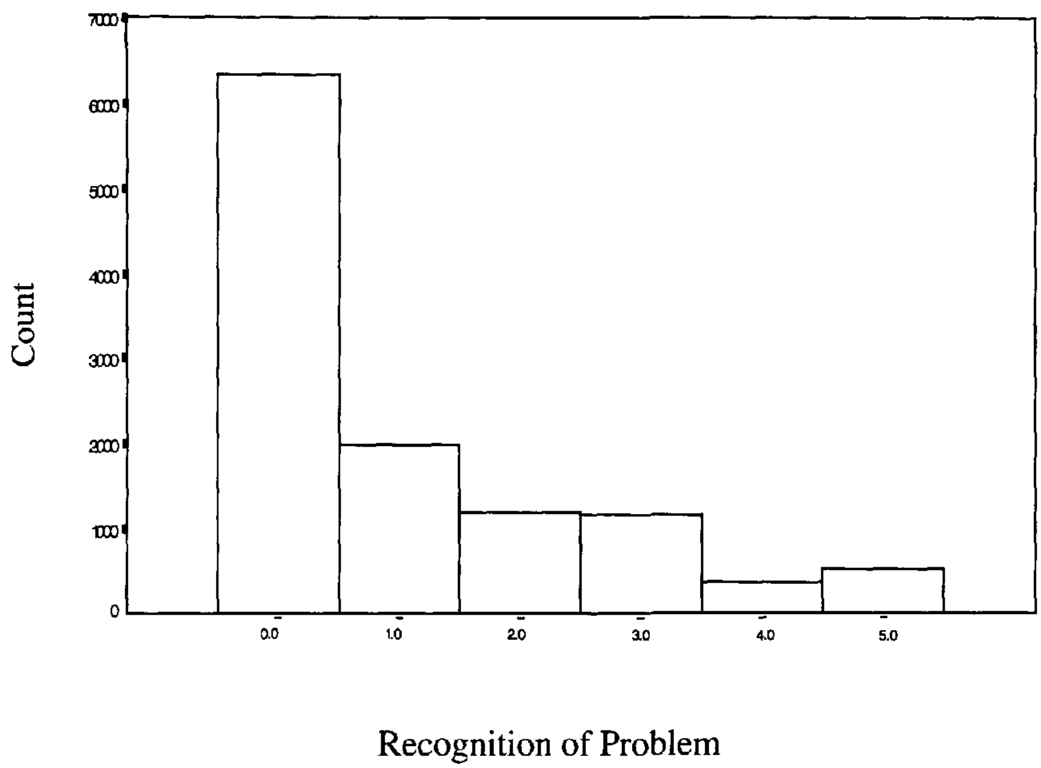

Figure 2. Representative histogram of Washington risk matrix variables.

signment of overall risk levels.) The risk factors are also highly intercorrelated: See Appendix B for the Spearman correlation matrix of the imputed, 6-point ordinal scale risk factors.

Given the very large number of cases (N = 12,978), it is not surprising or informative that all 37 risk factors had a statistically significant association with overall risk. Only the risk factors with the largest odds ratios and correlations with (dichotomized) overall risk (treated as the dependent variable) are presented in Table 2. Positive predictive value (PPV), or "true positives," is the percentage of cases rated 3–5 for an independent variable (the predictor) that are also rated 3–5 in overall risk; 1 − NPV (negative predictive value), or "false negatives," is the percentage of cases rated 0–2 for a predictor but rated 3–5 in overall risk. For example, changing the risk level of protection of child from 0–2 to 3–5 increases the percentage of cases rated 3–5 in overall risk from 15% to 79%, an increase by a factor of 22. For simplicity of presentation, only the percentages of 3–5 overall risk cases are given; percentages of cases rated 0–2 in overall risk are equal to 100 minus the percentage of 3–5 risk cases.

Neural Network Results

For tutorial purposes, Tables 3 and 4 show a portion of the output for the neural network model of overall risk level using all 37 risk factors in 6-point ordinal form plus the case status variable as possible inputs and the overall risk level as the model output. Table 3 shows some results from testing the model fit on the data, for the training (70% of all data) and test (30% of all data) sets, and the combined results for all data. In order, the output columns show the linear correlation between the actual inputs and actual (observed) output (R); the network classification accuracy, which is the fraction of the predicted outputs that are within ±10% of the observed outputs, and 95% confidence intervals (predicted overall risk is within ± confidence limits of the observed value). The correlation coefficient is quite high (R = .697; R^2 = .486), and the prediction accuracy is quite good: 81% of all cases are predicted to within ±10% of their actual overall risk values. The confidence interval indicates that this prediction is accurate to within ±1.6 risk level units.

Table 4 shows the results of a sensitivity analysis of

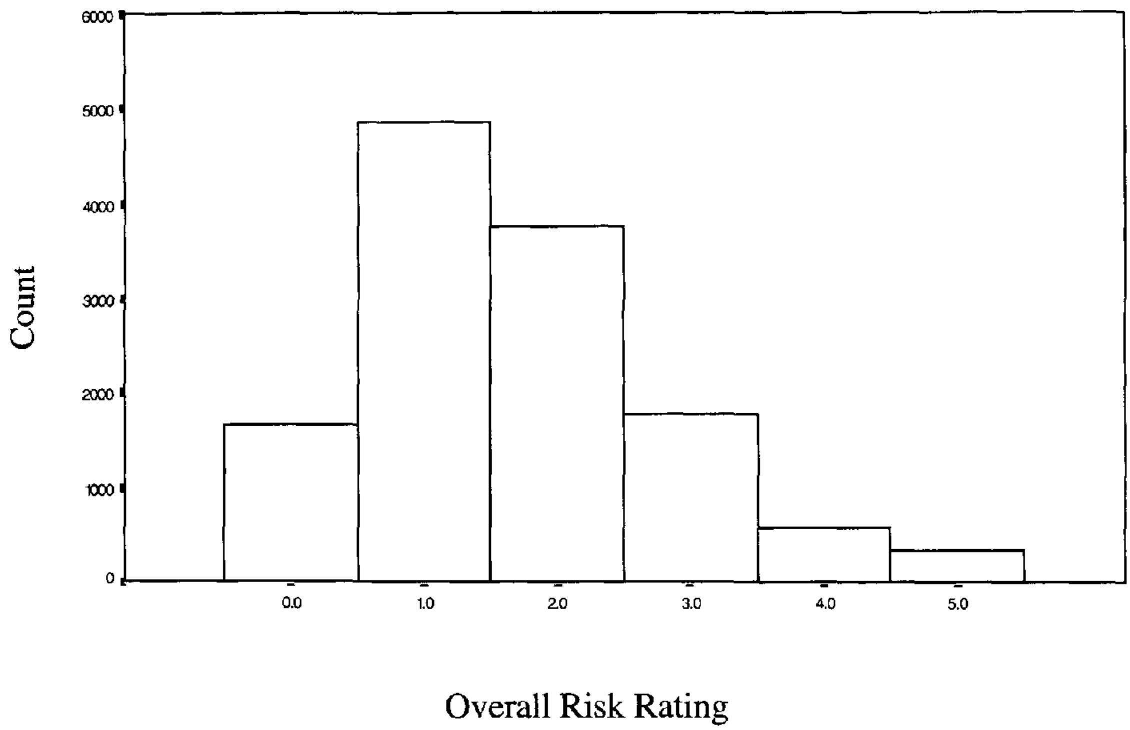

Figure 3. Histogram of overall risk rating.

the neural network model for all cases. The sign on the average sensitivity value indicates the overall direction of the variable effect: Positive values indicate that higher risk for the risk factor is correlated with higher overall risk. A negative sign indicates the possible use of a risk factor as a protective factor, an influence that can be confirmed by analysis of the categorical forms of the risk factors. To underscore the analogy with regression coefficients and to avoid confusion with the epidemiological utility measure of "sensitivity" which we will use to characterize classification models, average sensitivity values for each of the neural network model input variables are here called B_{nn}, and variances of the B_{nn} are given the symbol V_{Bnn}. (All of the sensitivity values reported in this article have been normalized to real-world units, that is, rescaled to the original scales of the model variables.)

Table 1
Frequencies for Overall Risk Levels

Overall risk	*N*	Percentage
0	1,675	13
1	4,854	37
2	3,754	29
3	1,772	14
4	582	5
5	341	3

Chronicity of child abuse or neglect (CA/N) has the largest B_{nn} and the largest main effect in the network, followed by (failure to recognize) recognition of problem, (poor) parenting skills, and stress on caretaker. (Recall that all risk factors are defined such that a higher score or risk level corresponds to a less favorable situation.) However, the lower variance of recognition of problem indicates that higher risk for recognition of problem is related more reliably and consistently to higher overall risk level than is chronicity of CA/N. The negative sign for response to disclosure indicates that it might be operating as a protective factor. The low magnitude for emotional harm–abuse indicates that this risk factor is of much less practical importance. Supervision and dangerous acts have large B_{nn} but also large variances, making them less consistent as indicators of higher overall risk level across a wide variety of cases. The neural network model contains only four risk factors with substantial and consistent effects on overall risk across all cases when predicted on a scale of 0–5. Different results are obtained when a different ques-

Table 2
Odds Ratios and Related Measures for Overall Risk 3–5

Independent variable	Odds ratio	PPV (%)	1 – NPV (%)	R^2
Protection of child	22.0	79	15	.215
Recognition of problem	13.1	65	12	.225
Parenting skills	12.0	65	14	.196
Chronicity of CA/N	9.3	56	12	.187
Cooperation with agency	10.8	66	15	.154
Caretaker impairments	8.7	62	16	.128
Emotional harm	8.4	61	16	.128
Nurturance	15.0	76	17	.119
Dangerous acts	6.7	54	15	.119
Supervision	7.7	59	16	.117
Response to child's behavior	9.3	65	16	.116
Social support	8.0	60	16	.115
Basic needs	18.7	80	18	.114
Victimization of others	9.1	64	17	.112
Substance abuse	6.7	55	16	.108
Response to disclosure	12.0	72	17	.104
Child's role in family	12.9	73	18	.103
Attachment–bonding	14.7	76	18	.092
Fear of caretaker	8.7	65	18	.083
History of domestic violence	4.9	49	16	.076
Medical care	11.9	73	19	.070
Hazards in the home	10.2	70	18	.070
Physical injury–harm	6.0	57	18	.059

Note. N = 12,978. The independent variables are dichotomized: 3–5 versus 0–2, insufficient, not applicable, missing.
Goodman and Kruskal (G-K) tau is used as a measure of the fraction of the total dependent variable variance captured by the independent variable. G-K tau ranges from zero (no correlation) to one (perfect correlation). PPV = positive predictive value; NPV = negative predictive value; CA/N = child abuse or neglect.

tion is asked, namely, What risk factors are most important in deciding between overall risk levels 0–2, cases where further CPS intervention is not typical, versus risk levels 3–5, where further CPS intervention is typical? This question is addressed in the classification models for dichotomous overall risk level reported below.

Recall that sensitivity values measure the effect of a model variable throughout the entire model, through the direct connection and hidden node weights for that variable. The network in this case was rather simple, containing just one hidden node, so a detailed interpretation of network weights is reserved for a later, more complicated model. To provide a convenient measure of model complexity, we define C_{nn} as the number of model inputs, p, multiplied by $h + 1$, where h is the number of hidden nodes in the network. Thus, for this example, $C_{nn} = p(h + 1) = 9(1 + 1) = 18$. As discussed in the Introduction, the hidden layer

Table 3
Test Results for Ordinal Scale Overall Risk Model Using Imputed Risk Factors

Data set	R	Accuracy (±10%)	95% CI
All (N = 12,978)	.697	81	1.62
Train (N = 9,084)	.692	80	1.63
Test (N = 3,894)	.710	82	1.59

Table 4
Sensitivity Analysis for All Cases (N = 12,978)

Model variable	B_{nn}	V_{Bnn}
Recognition of problem	.150	.001
Stress on caretaker	.109	.001
Parenting skills	.133	.002
Chronicity of CA/N	.166	.006
Response to disclosure	–.012	.001
Dangerous acts	.323	.114
Supervision	.334	.136
Emotional harm–abuse	.002	.031

Note. B_{nn} = average sensitivity value for model variable; V_{Bnn} = variance of B_{nn}; CA/N = child abuse or neglect.

contribution to the output node is roughly analogous to interaction terms in regression and ANOVA models, but in this case a single value describes a composite interaction between all variables in the model, rather than, for instance, a simple multiplicative term involving just two variables. The weights for the direct connections between inputs and outputs are analogous to main effects. The hidden layer has an output weight of .26. This value is comparable in magnitude to the individual main effect or direct connection risk factor weights, indicating that this composite interaction term has an influence on the output roughly comparable in magnitude to any one of the larger main effects.

Although some qualitative information can be gained by this sort of direct interpretation of network weights, it is not usually recommended (cf. Sarle, 1997a). Adjustments must be made for the different variable transforms used and the influence of other weights and hidden nodes in the model. The overall importance of the risk factors, accounting properly for both main and interaction effects, is more accurately given by the sensitivity analysis results, where risk factors with high average sensitivities and low sensitivity variances have major impacts on the outcome variable that are relatively consistent and reliable across many different cases.

Sensitivity values are reported in normalized units, allowing direct comparisons to be made of variables within the same model. However, the magnitudes will systematically vary depending on the type of variable used (ordinal scale or categorical). Thus, magnitudes are not directly comparable across different types of models; for example, magnitudes for the categorical variable model reported below are smaller than those for the ordinal model. The magnitudes indicate the relative importance of variables within a model, and rankings of variables using the magnitudes can be compared from model to model.

A multiple linear regression model without any interaction terms would correspond to a neural network model with no hidden nodes (only direct connections) and linear transformations for each of the input variables. Because this neural network model is close to that (linear transformations for some of the variables and only one hidden node with a moderate output weight), it is instructive to compare the neural network model with a corresponding linear regression model. Such a regression model, with each of the neural network variables entered as a main effect, gave a multiple linear correlation of .696 (R^2 = 0.485), essentially the same as the neural network model. All variables were significant to 99.99% confidence except for response to disclosure (75% confidence). The regression coefficients for each of the risk factors were as follows: recognition of problem = 0.172, chronicity of CA/N = 0.171, parenting skills = 0.137, stress on caretaker = 0.110, dangerous acts = 0.083, supervision = 0.071, emotional harm–abuse = 0.057, and response to disclosure = 0.010.

The relative order is very similar to the neural network order, and the goodness of fit is similar. However, the model prediction accuracy is much worse: The overall risk level is predicted to within ±10% of actual for only 34% of the cases, compared with the prediction accuracy of the neural network model of 70% to within ±10%. Transforming the input and output variables to more closely approximate normal distributions does not improve these results. The linear regression approach also suffers from the redundancy and instability problems discussed in the Introduction: Entering all risk factors into a backwards elimination regression model results in the retention of 30 out of the 37 risk factors and an R^2 of .514, only marginally better than the much more parsimonious neural network model shown above. The beta values for the risk factors are small, and different random subsets of the data give regression models with different risk factors. The ±10% prediction accuracy improves to 66%, much better but still substantially less than the accuracy of the neural network model.

Ignoring nonlinearities and interactions results in regression models that are unrealistically overinclusive, where small main effects of many variables are used to attempt to explain the variance accounted for by nonlinearities and interactions among just a few key variables. The situation would be only partially improved by transforming the risk factors with the functions chosen by the neural network model before entry into a linear regression model, because the regression model would still not account for the interactions represented by the hidden node in the neural network. Once again, the lower accuracy of regression versus neural network models indicates that the relationship between risk factors and overall risk is modeled more realistically by allowing for these interactions and nonlinearities.

The relatively simple form of the neural network model for this example allowed a straightforward comparison with linear regression. However, most of the neural network models constructed in our CPS decision-making research are more complex, containing multiple hidden nodes, and defy easy comparison with simple models.

We turn now to a consideration of neural network classification models for dichotomous overall risk. Tables 5, 6, and 7 show the neural network results for the classification of all cases as 0–2 overall risk or 3–5 overall risk (called Riskhilo). Table 5 shows the results from classification of all cases (classification rates for the separate training and test subsets were very similar and are therefore not listed). The first row shows the results for all cases with observed (actual) overall risk of 0–2; an 86% correct classification rate, 8,868 cases predicted 0–2 and observed 0–2; and 1,415 cases predicted 3–5 overall risk but observed 0–2 risk. The second row shows the corresponding figures for all cases with an observed overall risk of 3–5. The third row gives the composite value for both categories: a total classification accuracy of 79%.

The last row gives the Komolgorov–Smirnov (K-S) statistic and threshold values. The K-S statistic is a measure of how well the neural network model has separated the two output classes: Cumulative histograms of the two classes are calculated and overlaid, and the K-S statistic is the greatest distance between corresponding bin values in the two histograms. A value of 0 indicates the model has output random values, and a value of 1 indicates a perfect model. The K-S threshold is the point where the histogram separation occurs. Because the internal recode of the variable recodes overall risk 3–5 = 1 and 0–2 = 0, the threshold value of .575 indicates a separation near the midpoint of the predicted range of 0 to 1. The K-S statistic value of .585 here indicates a moderately good separation.

The sensitivity analysis in Table 6 is interpreted as previously described. Table 7 shows the actual network weights, transfer functions, bias terms, and input variable transforms. This neural network model is more complex than the previous example, containing three hidden nodes instead of one hidden node. Each hidden node represents a different interaction term, each with different contributions from each of the model variables. Larger numbers of hidden nodes represent a higher degree of explanatory variable interactions and greater nonlinearities in the relationship between outcome and explanatory variables. This network has a complexity measure, C_{nn} = 6 input variables × (3 hidden nodes + 1) = 24. Note also the use of the Softmax output function for this dichotomous classification model, which is a threshold function (equivalent to the logistic function) that converts the continuously scaled predicted values to the categorical output classes of 0–2 or 3–5 overall risk.

Table 5
Test Results for Dichotomous Overall Risk Classification Model Using Imputed Risk Factors

Overall risk (observed)	Classification rate (%)	Overall risk (predicted) 0–2	3–5	Total *N*
0–2	86	8,868	1,415	10,283
3–5	72	748	1,947	2,695
All	79	9,616	3,362	12,978

Note. Komolgorov-Smirnov (K-S) statistic = 0.585; K-S threshold = 0.575.

Table 6
Sensitivity Analysis for Dichotomous Overall Risk Classification Model Using Imputed Risk Factors

Model variable	B_{nn}	V_{Bnn}
Stress on caretaker	.254	.013
Fear of caretaker	.227	.012
Parenting skills	.367	.031
Recognition of problem	.433	.039
Physical injury–harm	.529	.229

From the classification matrix shown in Table 5, we can, as in logistic regression models, calculate epidemiological measures of model utility such as sensitivity (not to be confused with the neural network sensitivity analysis), specificity, PPV and NPV. Sensitivity is simply the percent correct classification of the 3–5 risk cases, and specificity is the percent correct 0–2 risk cases. The values of these measures for the above example (Riskhilo classification of all cases using the ordinal scale imputed risk factors) are as follows: sensitivity = 72%, specificity = 86%, PPV = 58%, and NPV = 92%.

Returning to the sensitivity analysis values, we observe that stress on caretaker, recognition of problem, and parenting skills are important in distinguishing 0–2 risk cases from 3–5 risk cases and in predicting overall risk along the entire 6-point scale. However, chronicity of CA/N is not important for the categorical distinction of overall risk across all cases. Instead, fear of caretaker and (to a lesser extent, given its large variance) physical injury–harm emerge as important variables for this distinction.

The classification problem is traditionally ad-

Table 7
Network Weights

Weight type	Hidden 1	Hidden 2	Hidden 3	Output
Node transfer function	tanh	tanh	tanh	SoftMax
Bias term	.324	−.127	−.0457	.786
Input Variables (transforms)				
Case status (dummy for initial status)	−.363	−.227	−.078	.807
Fear of caretaker (linear)	−.295	−.110	−.064	.550
Physical injury–harm (inverse)	.123	−.408	.230	.566
Parenting skills (linear)	−.597	.139	−.126	1.010
Recognition of problem (linear)	−.303	.081	.153	1.280
Stress on caretaker (linear)	−.032	.162	.321	.837
Hidden Node 1	0	0	0	−.913
Hidden Node 2	0	0	0	−.692
Hidden Node 3	0	0	0	.161

dressed using logistic regression. A logistic regression model entering the above variables (plus case status as a categorical variable) gave a total classification accuracy of 87%; however, the classification accuracy (sensitivity) for 3–5 risk cases was only 51%. Entry of various interaction terms (e.g., fear of caretaker × physical injury–harm was significant in the logistic regression model) does not significantly improve this situation. Odds ratios for the risk factors are as follows: recognition of problem = 1.7, parenting skills = 1.5, stress on caretaker = 1.4, physical injury–harm = 1.3, and fear of caretaker = 1.3.

A backwards elimination logistic regression calculation using all risk factors includes 27 of the 37 risk factors, none with odds ratios greater than 2, and a sensitivity that improves only slightly, to 55% (total classification accuracy = 87%). As with linear regression, logistic regression overemphasizes main effects, and includes a large number of marginally important risk factors in order to model the variance due to nonlinearities and interactions. Entry of multiplicative interaction terms into the logistic regression model did not improve model accuracy.

What, then, accounts for the higher classification accuracy for risk 3–5 cases achieved by the neural network approach? In this example, 21% of the cases are actually rated (observed) at risk 3–5. One possibility is that the neural network procedure is using prior probabilities of group membership in a more effective way than the logistic regression procedure. We thus ran a linear discriminant analysis of these data, with the list of neural-network-selected input variables shown above and using prior probabilities of group membership.[1] The classification results were essentially equivalent to the results for the logistic regression model: About 50% of cases rated 3–5 risk were correctly classified (total classification accuracy = 85%). Entering all risk factors into the discriminant analysis improves this figure to 56% (total = 87%).

Interestingly, running a discriminant analysis with equal prior probabilities of group membership, entering all 37 risk factors into the model and performing stepwise variable selection gave classification accuracy results comparable with those for the neural network: sensitivity = 77%, specificity = 85%, total accuracy = 84%, PPV = 58%, and NPV = 93%. However, it achieves this through a considerable loss in parsimony: As with logistic regression, 27 of the 37 risk factors are selected into the model, compared with 5 risk factors in the neural network model. Calculation of an impoverished neural network, one with no input variables, would show how the neural network classifies on the basis of prior group membership. Unfortunately, limitations in the neural network software prohibited construction of such a network. To approximate an impoverished neural network, a network was constructed with just the case status variable entered. The results were the worst for any of the models; a sensitivity of 33% (total classification accuracy = 60%). This indicates that differences in how prior probabilities of group membership are used

[1] We thank David S. Cordray for suggesting this possibility and the tests with discriminant analysis and the impoverished neural network.

are not the origin of the superior neural network results. Table 8 summarizes each of the models used for the dichotomous overall risk data.

The origin of the superior classification results and greater parsimony of the neural network model is also not due to the transforms the neural network program applies to the input variables: In this example, linear transforms are used for all but one of the risk factors. The origin of the improvement seems most likely to be the better treatment of variable interactions through the three hidden nodes in this network and also possibly the modeling of nonlinear variable relationships. The fact that the 5-variable neural network achieves classification accuracies comparable to the 27-variable discriminant analysis model argues strongly for the presence of these nonlinearities and interactions in the data: The discriminant analysis model is in effect attempting to model an elephant with 27 straight rulers. Factor analysis or nonlinear principal-components analysis prior to factor entry into logistic regression or discriminant analysis models also does not improve the accuracy or parsimony of the results (English et al., 1998).

Tables 9 and 10 show the results from neural network modeling of dichotomous overall risk level using the categorical versions of the risk factors. Recall that the categorical versions used in this analysis categorized the risk factors as no risk = 0; low risk = 1, 2, or not applicable; risk = 3–5; and insf = insufficient information to assess. The selection of the no risk category (and a negative sign on the B_{nn} value) indicates the possible use of the risk variable as a protective factor in risk assessment.

The insufficient category of Risk Factor 3 (behavioral problems) is used in the model with a positive average sensitivity: The "insufficient" label raises the probability of a case being assigned overall risk 3–5. The no risk category of Risk Factor 38 (response to child's behavior) has a negative average sensitivity; rating a case zero risk for response to child's behavior lowers the probability of being assigned 3–5 overall risk. This is further evidence for response to child's behavior acting as a protective factor. For all other risk factors in the network, the 3–5 risk category (risk) increases the probability of a case being rated 3–5 overall risk. Fit, accuracy, and relative importance values for this network are as follows: K-S = .586, C_{nn} = 42, total accuracy = 79%, sensitivity = 81%, specificity = 77, PPV = 48%, and NPV = 94%.

When treated as an ordinal variable, response to child's behavior is not selected in the neural network model. When treated as a categorical variable, zero risk for response to child's behavior emerges as an important variable. This disagreement is likely due to two reasons: first, when faced with a large number of highly interacting and redundant variables, the neural network or any other approach will find it difficult to find a single model that clearly provides the best fit. The neural network approach as used here is more likely to come up with a nearly unique model, but the basic problem remains. The modeling approach in this context is attempting to find a single, common pattern in a data set with wildly divergent cases and a multitude of individual decision-making processes. The model shown above successfully classifies 79% of all cases, but other nearly equivalent models would do nearly as well. The designation of risk factors as "important" should therefore be viewed as a relative judgment based on the data at hand. Secondly, as discussed above, the risk factors as designed and apparently used have a mixture of ordinal and categorical properties, and coding the risk factors as ordinal variables will obscure any potential distinctive use of the zero label as a protective factor. One should then expect the answer (model) to be different depending on which form of the variables is used.

Neural network modeling of dichotomous overall risk with the simpler 3-point ordinal scale version of the risk factors (where 0 = *0 risk*; 1 = *1–2 risk, not applicable, missing and insufficient*; and 2 = *3–5*

Table 8
Comparison of Classification Models for Dichotomous Overall Risk

Model	Sensitivity (%)	Total accuracy (%)	Risk factors
Full neural net	72	79	5
"Impoverished" neural net	33	60	0
Logistic regression	55	87	27
Discriminant analysis, equal priors	56	87	27
Discriminant analysis, group membership priors	77	84	27

Table 9
Test Results for Dichotomous Overall Risk Classification Model Using Categorical Risk Factors

Overall risk (observed)	Classification rate (%)	Overall risk (predicted) 0–2	3–5	Total *N*
0–2	77	7,919	2,364	10,283
3–5	81	502	2,193	2,695
All	79	8,421	4,557	12,978

Note. Komolgorov-Smirnov (K-S) statistic = 0.586; K-S threshold = 0.443.

risk) gave the following results (N = 12,978): K-S = .587, C_{nn} = 24, total accuracy = 79%, sensitivity = 74%, specificity = 85%, PPV = 56%, and NPV = 93%. See also Table 11.

The simpler 3-point ordinal scale gives results for Riskhilo that are comparable in classification accuracy and goodness of fit to the more complicated versions of the risk factors. Apparently, attempting to model all cases together results in a very averaged-over or smoothed-out picture of the relationship between Riskhilo and the risk factors, and the relatively crude representation of the risk factors as 3-point ordinal scales is sufficient to capture this relationship. The relatively small decrease in model accuracy is also an indication that the loss of information due to collapse of the outcome variable is relatively small.

In summary, considering B_{nn} and V_{Bnn} values from all four neural network models, parenting skills, recognition of problem, chronicity of CA/N, stress on caretaker, and emotional harm–abuse emerge as particularly important variables for the determination of overall risk, when considered across all cases. Fear of caretaker and response to child's behavior (as a protective factor) may also be important. The results presented here illuminate the complex considerations that must be addressed in the analysis of CPS decision-making data and introduce the use and interpretation of neural networks as a useful tool for the modeling of these complex relationships. English et al. (1998) covered results from neural network analysis of various subsets of the total data set and addressed in more detail the nature and practice implications of CPS assessment of overall risk, and English et al. (in press) covered neural network results and practice implications for the substantiation decision. One obvious practice implication to draw from these results is that a 37-item risk factor matrix for risk assessment seems a bit too elaborate for how risk assessment is actually performed in the field and creates an unnecessary paperwork burden on caseworkers (who are required to complete the 37-item matrix for each of their investigated cases).

Table 10
Sensitivity Analysis for Dichotomous Overall Risk Classification Model Using Categorical Risk Factors

Model variable	B_{nn}	V_{Bnn}	Category
Response to child's behavior	–.0425	.0024	no risk
Protection of child	.0078	.0013	risk
Parenting skills	.0184	.0038	risk
Recognition of problem	.0221	.0048	risk
Emotional harm–abuse	.0189	.0048	risk
Behavioral problems	.0225	.0069	insf

Note. insf = insufficient; B_{nn} = average sensitivity value for model variable; V_{Bnn} = variance of B_{nn}.

Table 11
Sensitivity Analysis for Dichotomous Overall Risk Using Three-Point Ordinal Version of Risk Factors

Model variable	B_{nn}	V_{Bnn}
Protection of child	.246	.014
Recognition of problem	.176	.057
Emotional harm–abuse	.160	.009
Cooperation with agency	.041	.002
Chronicity of CA/N	.414	.108

Note. CA/N = child abuse or neglect; B_{nn} = average sensitivity value for model variable; V_{Bnn} = variance of B_{nn}.

Generalizability of Neural Network Models

Three additional, separate data sets were used to test the generalizability of the neural network models for overall risk. In this mode of testing a neural network model, the values of each risk factor included in the original model are input for the new set of cases, and the outcome is predicted from those values and compared with the actual outcome for the new cases. Summary statistics are then generated, indicating the accuracy at which the model classifies the new cases. Sensitivity analysis based on the new cases then indicates any changes in the relative importance of the individual (predetermined) model risk factors in performing this classification.

In the data set described above, 1,458 cases with more than one CA/N code indicated were not included. The multiple abuse cases were assigned imputed (6-point ordinal scale) risk matrix variables following the same procedures used for the working data set. The neural network models constructed for the working data set for overall risk and dichotomized overall risk (0–2 vs. 3–5) were then run on the multiple abuse cases. The results for the multiple abuse cases were nearly completely identical to those for the working data set used to construct the original model: The cases were classified by both models to within 1% of the same accuracies obtained for the working data set, and the B_{nn} and V_{Bnn} values for each risk factor were also nearly identical to the values obtained for the working data set. The only exception was an increase in the B_{nn} for emotional harm–abuse, from .002 for the cases in the working data set to .015 for multiple abuse cases, in the model for overall risk as a 6-point ordinal scale versus the imputed risk factors. (Emotional harm–abuse is the least important risk factor in this model.)

A set of 1,930 cases that had all CA/N codes completely missing were also examined. A test of the neural network models for overall risk and dichotomous overall risk on these cases gave classification accuracies and B_{nn} values nearly identical to those for the working data set, except for a 10% reduction in the B_{nn} value for the supervision risk factor.

The neural network models for overall risk and dichotomized overall risk (0–2 vs. 3–5) were also tested on 177 separate cases that represent single male households. These male caretaker cases were not significantly different from the cases in the working data set with respect to demographic composition or CPS outcomes (given the low number of cases, the differences would have to be quite large to be considered statistically significant). Average age of the children was slightly higher (8 vs. 6 years), and the average overall risk rating and risktag at intake were slightly lower (1.5 vs. 1.7, and 3.6 vs. 3.8, respectively). The proportions of abuse types were different: 14% sexual abuse and 44% physical abuse, versus 17% and 34%, respectively, for the working data set. The proportions of physical neglect were essentially the same (36% vs. 37%).

The results were again very similar, with a few more differences in the magnitudes of the model risk factors than observed for multiple abuse cases. B_{nn} for emotional harm–abuse was again higher, .022 versus .002 for the working data set, and the B_{nn} values for the more important risk factors dangerous acts and supervision were also 20–25% higher. The higher value for dangerous acts is consistent with the higher proportion of physical abuse cases for this sample, but the possible meanings of the higher values for emotional harm–abuse and supervision are not as clear. Despite these differences, the classification accuracies of the models were actually 4–5% higher than those for the main data set of 12,978, indicating again that the neural network models are quite robust in being able to correctly classify a wide variety of cases.

Conclusions

We have shown that neural networks, in particular, MLP NNs, are a useful tool for the analysis of CPS risk assessment data. The models have classification accuracy at least equal to and usually substantially superior to those for linear or logistic regression. This improvement is attributed to the ability of neural networks to model nonlinear relationships between highly interacting variables, a situation that likely is common in risk assessment data obtained from the field, particularly for the WRM, which is intended as a holistic risk assessment tool. The neural network models derived here, for the association between overall risk ratings and individual risk factors, are also shown to generalize well to sets of cases not originally included in the set used to construct the models. Overall risk ratings as used by Washington CPS are of course intended to summarize in some general sense the collective risk assessed through the individual risk factors. Neural network classification models for substantiation show that useful associations with the risk factors can be derived for this decision as well, where the superiority of neural networks over logistic regression is even more dramatic than that shown here for models of overall risk, including situations where neural networks give quite good models but logistic regression completely fails to discriminate one class from another well (English et al., in press).

Social science data, particularly child welfare administrative data from uncontrolled field situations, is notoriously complex, noisy, and difficult to model. Many sophisticated ecological theories are currently proposed to account for the complex phenomenology of child abuse and neglect or to explain and describe the human decision-making process in CPS and other agencies. However, the statistical regression and classification methods that are usually used to test these models do not mathematically reflect this complexity and suffer from serious deficiencies in the ability to model nonlinear relationships among interacting variables. In our opinion, neural network modeling is a method that provides a mathematical framework that is more suitable in some situations for the testing of complex ecological theories of human behavior.

Although neural networks as used here are a step in the right direction, they are far from perfect. The loss of simple interpretability is serious. One author refers to the current situation as the "missing middle" (Ripley, 1996): a methodological gap between easily interpreted models that are unrealistic and models that generalize well but whose details are obscure. Much research remains to be done before a general, mathematically based framework for collective human decision-making and other social phenomena will become available. In this regard, the restriction of MLP NN to a "feed forward" mode, where output values cannot affect input values, is not likely a realistic model of how humans adjust their perceptions in light of experience. Recent research in neural networks that incorporate feedback may hold some promise in more accurately modeling this phenomenon.

Despite these drawbacks, it is clear that neural networks can be a useful tool to the analyst seeking to model complex relationships in risk assessment and other fields. The differences with simpler statistical models are not purely technical: They have implications for interpretation of human decision making and for the practice and policy of risk assessment. To give just one indication, the results described here indicating that a very small number of risk factors were actually being used in caseworker practice (combined with our results for other caseworker decisions) were instrumental in prompting and guiding a major overhaul of risk assessment in Washington State that is currently in progress, an overhaul that will result in a major reduction of the size of the risk assessment matrix, and a reemphasis in caseworker training around certain risk factors (such as a history of domestic violence) that are known to be related to recidivism of child abuse or neglect but are currently not being used by caseworkers in their assessment of overall risk or their decision to place a child in foster care.

References

Baird, C. (1988). Development of risk assessment indices for the Alaska Department of Health and Social Services. In T. Tatara (Ed.), *Validation research in CPS risk assessment: Three recent studies.* Washington, DC: American Public Welfare Association.

Bieber, S. L. (1988). Multiple regression and its alternatives. *Social Science Journal, 25,* 1–19.

Bishop, C. M. (1995). *Neural networks for pattern recognition.* Oxford, England: Clarendon Press.

Briere, J. (1988). Controlling for family variables in abuse effects research: A critique of the 'partialling' approach. *Journal of Interpersonal Violence, 3,* 80–89.

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

Cheng, B., & Titterington, D. M. (1994). Neural networks: A review from a statistical perspective (with discussion). *Statistical Science, 9,* 2–54.

Dash, M., & Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis, 1.* Retrieved October 1997 from the World Wide Web: http://www-east.elsevier.com:80/ida/browse/0103/ida00013/article.html

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95–106.

Dawes, R. M., Faust, D. & Meehl, P. E. (1989, March 31). Clinical versus actuarial judgement. *Science, 243,* 1668–1674.

Doueck, H. J., English, D. J., DePanfilis, D., & Moote, G. T. (1993). Decision making in child protective services: A comparison of selected risk assessment systems. *Child Welfare, LXXII,* 441–452.

English, D. J., & Aubin, S. (1991). *Impact of investigation outcomes for child protective services cases receiving differential levels of service.* Olympia, WA: State of Washington Department of Social and Health Services.

English, D. J., Marshall, D. B., Brummel, S. C., & Coghlan, L. K. (1998). *Decision-making in child protective services: A study of effectiveness. Final report, Phase I: Quantitative analysis.* Olympia, WA: State of Washington Department of Social and Health Services.

English, D. J., Marshall, D. B., Coghlan, L, Brummel, S., & Orme, M. (in press). Causes and consequences of the substantiation decision in Washington State Child Protective Services. *Children & Youth Services Review.*

Fluke, J. (1993). *Evaluation of the Pennsylvania approach to risk assessment.* Englewood, CO: American Humane Association.

Freidman, J. H., & Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association, 76,* 817–823.

Garson, G. D. (1991). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review, 9,* 399–434.

Garson, G. D. (1998). *Neural networks: An introductory guide for social scientists.* London: Sage.

Gieman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation, 4,* 1–58.

Holden, A. V., & Kryukov, V. I. (1991a). *Neural networks: Theory and application.* New York: Wiley.

Holden, A. V., & Kryukov, V. I. (1991b). *Neurocomputers and attention: Vol. 1. Neurobiology, synchronization, and chaos.* New York: Wiley.

Holden, A. V., & Kryukov, V. I. (1991c). *Neurocomputers and Attention: Vol. 2. Connectionism and neurocomputers.* New York: Wiley.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: Wiley.

Johnson, W., & L'Esperance, J. (1984). Predicting the recurrence of child abuse. *Social Work Research and Abstracts, 20*(2), 21–26.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

McDonald, T. P., & Marks, J. (1989). *Predicting recurrences of child maltreatment.* Portland, ME: National Child Welfare Resource Center for Management and Administration, University of Southern Maine.

Peretto, P. (1992). *An introduction to the modeling of neural networks.* Cambridge, England: Cambridge University Press.

PREDICT (Version 2.0) [Computer software]. (1997). Pittsburgh, PA: NeuralWare.

Ripley, B. D. (1996). *Pattern recognition and neural networks.* Cambridge, England: Cambridge University Press.

Sarle, W. S. (1997a). *How to measure importance of inputs?* Retrieved October 1997 from the World Wide Web: ftp://ftp.sas.com/pub/neural/importance.html

Sarle, W. S. (Ed.). (1997b). *Neural network FAQ* [Periodic posting to the Usenet newsgroup comp.ai.neural-nets]. Retrieved October 1997 from the World Wide Web: ftp://ftp.sas.com/pub/neural/FAQ.html

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, Prediction and Search.* Berlin, Germany: Springer-Verlag.

Wilson, R. L., & Hardgrave, B. C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational & Psychological Measurement, 55,* 186–195.

Appendix A

Key to Numbered Risk Factors

A more detailed description of how these risk factors are defined and operationalized is given in English et al. (1998). The (omitted) odd numbered risk factors following nrisk16 are risk factors for secondary caretakers; as most of these are missing for nearly all cases (even when two caretakers are present in the household, social workers in Washington apparently do not commonly assess risk for them), they were not included in most of our analyses. Thus, risk factors following nrisk15 (chronicity of child abuse or neglect [CA/N]) are assessed for the primary caretaker. Each risk factor is defined consistently such that increasing value of the variable corresponds to increasing risk; for example, a higher score for basic needs refers to a more serious lack of provision of basic needs.

Child characteristics

nrisk1 Child age risk level (Younger children at higher risk)

nrisk2 Disability–development (Physical, mental, or social deficiencies, delays, or both)

nrisk3 Behaviorial problems (Child's behavioral disturbances)

nrisk4 Self-protection (Child's age and developmental-stage-related ability to self-protect)

nrisk5 Fear of caretaker (Child's fear of caretaker of home environment)

Severity of CA/N

nrisk6 Dangerous acts (Acts by caretaker that place the child at risk of injury)

nrisk7 Physical injury–harm (Extent of physical harm to the child)

nrisk8 Emotional harm–abuse (Extent of emotional or behavioral impairment)

nrisk9 Medical Care (Degree of lack of provision of routine and crisis medical or dental care)

nrisk10 Basic needs (Degree of lack of provision of food, clothing, and shelter)

nrisk11 Supervision (Degree of lack of supervision that places the child at risk of harm)

(Appendixes continue)

nrisk12 Hazards in home (Physical hazards or dangerous objects in the home environment)
nrisk13 Sexual abuse–exploitation (Extent of sexual abuse, sexual exploitation, or both)
nrisk14 Nonsexual exploitation (Extent of nonsexual exploitation)

Chronicity of CA/N

nrisk15 Chronicity of CA/N (Frequency of abuse, neglect, or both)

(Primary) caretaker characteristics

nrisk16 Victimization of others (Caretaker victimization of other children)
nrisk18 Caretaker impairments (Mental, physical, or emotional impairments of caretaker)
nrisk20 Deviant arousal (Adult is sexually aroused by children)
nrisk22 Substance abuse (Use of alcohol or drugs to the point of impairment of parenting)
nrisk24 History of domestic violence (History of domestic violence or assaultive behavior)
nrisk26 History of CA/N (Caretaker's history of abuse or neglect as a child)
nrisk28 Parenting skills (Unrealistic expectations, gaps in knowledge or skills)
nrisk30 Nurturance (Withholding of affection, interaction, acceptance)
nrisk32 Recognition of problem (Lack of awareness or denial, refusal to accept responsibility)
nrisk34 Protection of child (Lack of ability, reliability or willingness to protect the child)
nrisk36 Cooperation with agency (Lack of cooperation or hostility towards agency or workers)

(Primary) caretaker relationship with child

nrisk38 Response to child's behavior (Inappropriate or abusive response)
nrisk40 Attachment–bonding (Anxious, disturbed, or missing child attachment to parent)
nrisk42 Child's role in family (Inappropriate role that limits or prevents normal development)
nrisk44 Pressuring child to recant (Indirect or direct pressure on child to recant abuse)
nrisk46 Personal boundary issue (Violations of personal or physical boundaries)
nrisk48 Response to disclosure (Lack of belief in child's disclosure of abuse)

Social and economic factors

nrisk50 Stress on caretaker (Significant stresses on or life changes for caretaker)
nrisk52 Employment status (Lack of employment, marketable skills, or both)
nrisk54 Social support (Degree of lack of support, social isolation)
nrisk56 Economic resources (Resources inadequate to meet basic needs)

Perpetrator access to child

nrisk58 Access to and responsibility for child (Access of abusive perpetrator to child)

Appendix B

Risk Factor (Spearman) Correlation Matrix

	NRISK1S	NRISK2S	NRISK3S	NRISK4S	NRISK5S	NRISK6S	NRISK7S	NRISK8S	NRISK9S	NRISK10S	NRISK11S	NRISK12S
NRISK2S	−.0222*											
NRISK3S	−.2165**	.5670**										
NRISK4S	.5909**	.0599**	−.1103**									
NRISK5S	−.1265**	.3349**	.3784**	−.0054								
NRISK6S	.0336**	.2625**	.2632**	.1068**	.4433**							
NRISK7S	−.0637**	.2598**	.3030**	.0192*	.4088**	.5216**						
NRISK8S	−.1173**	.3043**	.3720**	.0239**	.5258**	.4591**	.3961**					
NRISK9S	.0878**	.3721**	.2705**	.1119**	.3729**	.3697**	.3358**	.3514**				
NRISK10S	.0873**	.3410**	.2554**	.1158**	.3836**	.3744**	.2880**	.3727**	.6525**			
NRISK11S	.0938**	.2764**	.2493**	.1230**	.3350**	.4002**	.2320**	.3661**	.4704**	.5502**		
NRISK12S	.0381**	.2714**	.2368**	.0862**	.3466**	.3544**	.2690**	.3450**	.4741**	.5721**	.4552**	
NRISK13S	−.0236**	.1925**	.1798**	.0821**	.2547**	.1773**	.1688**	.3377**	.2330**	.2270**	.2307**	.2474**
NRISK14S	−.0258**	.2222**	.2197**	.0438**	.2882**	.2506**	.2446**	.3289**	.3046**	.3100**	.2835**	.3103**
NRISK15S	−.0059	.2416**	.2294**	.1021**	.3621**	.3760**	.2574**	.4666**	.3250**	.3726**	.3845**	.3169**
NRISK16S	.0360**	.2604**	.2271**	.0935**	.3204**	.3163**	.2381**	.3562**	.3734**	.4065**	.3855**	.3514**
NRISK18S	.0369**	.3168**	.2592**	.1131**	.3581**	.3321**	.2336**	.4253**	.3956**	.4333**	.3784**	.3641**
NRISK20S	−.0113	.1741**	.1549**	.0565**	.2216**	.2110**	.1965**	.2543**	.2702**	.2708**	.2382**	.2755**
NRISK22S	.0996**	.1477**	.0994**	.1236**	.2267**	.2432**	.1202**	.2555**	.3096**	.3520**	.3381**	.2868**
NRISK24S	.0406**	.1459**	.1368**	.0973**	.2420**	.2299**	.1552**	.2937**	.2337**	.2643**	.2440**	.2265**
NRISK26S	.0493**	.1593**	.1438**	.1341**	.1838**	.1772**	.1264**	.2478**	.1999**	.2133**	.2130**	.2041**
NRISK28S	.0253**	.3037**	.2980**	.1175**	.3796**	.3920**	.2716**	.4360**	.4288**	.4650**	.4480**	.3847**
NRISK30S	−.0628**	.3437**	.3450**	.0307**	.4774**	.3827**	.3120**	.4603**	.4695**	.4942**	.4505**	.4164**
NRISK32S	.0161	.2645**	.2320**	.0833**	.3993**	.3707**	.2491**	.4087**	.4135**	.4486**	.4360**	.3647**
NRISK34S	−.0067	.2900**	.2835**	.0776**	.4443**	.4169**	.3140**	.4544**	.4364**	.4703**	.4691**	.3962**
NRISK36S	.0116	.2624**	.2117**	.0794**	.3796**	.3394**	.2638**	.3587**	.4160**	.4480**	.3827**	.3646**
NRISK38S	−.0974**	.3088**	.3944**	.0092	.4444**	.3848**	.3413**	.4662**	.3884**	.3970**	.3728**	.3551**
NRISK40S	−.0703**	.3430**	.3553**	.0109	.4865**	.3677**	.3211**	.4564**	.4544**	.4755**	.4163**	.4043**
NRISK42S	−.0831**	.3307**	.3595**	−.0009	.4606**	.3703**	.3086**	.4713**	.4375**	.4588**	.4377**	.4096**
NRISK44S	.0143	.2398**	.2035**	.0763**	.3383**	.2569**	.2383**	.3018**	.3319**	.3363**	.3013**	.3239**
NRISK46S	.0046	.2786**	.2556**	.0710**	.3425**	.2742**	.2386**	.3455**	.3648**	.3680**	.3381**	.3525**
NRISK48S	−.0053	.2712**	.2352**	.0701**	.3692**	.2971**	.2516**	.3531**	.3622**	.3717**	.3460**	.3304**
NRISK50S	.0511**	.2159**	.2088**	.1329**	.2672**	.2669**	.1567**	.3467**	.2432**	.2918**	.2668**	.2658**
NRISK52S	.1259**	.2268**	.1513**	.1454**	.2046**	.2289**	.1413**	.2176**	.3373**	.3766**	.3043**	.2816**
NRISK54S	.0464**	.2656**	.2147**	.0955**	.3132**	.2928**	.2078**	.3352**	.3716**	.4171**	.3755**	.3195**
NRISK56S	.1029**	.2485**	.1979**	.1541**	.2564**	.2698**	.1973**	.2743**	.3785**	.4335**	.3585**	.3386**
NRISK58S	.0906**	.1546**	.1435**	.1434**	.1906**	.2159**	.1567**	.1757**	.2224**	.2309**	.2204**	.2097**

(Appendix continues)

Appendix B (continued)

	NRISK13S	NRISK14S	NRISK15S	NRISK16S	NRISK18S	NRISK20S	NRISK22S	NRISK24S	NRISK26S	NRISK28S	NRISK30S	NRISK32S
NRISK14S	.5640**											
NRISK15S	.3276**	.2729**										
NRISK16S	.2250**	.2655**	.4403**									
NRISK18S	.2363**	.2876**	.4163**	.4984**								
NRISK20S	.3778**	.5396**	.1854**	.2739**	.2872**							
NRISK22S	.1632**	.2152**	.3559**	.4115**	.4339**	.2297**						
NRISK24S	.1958**	.2046**	.3287**	.3463**	.4016**	.1991**	.4801**					
NRISK26S	.1727**	.1502**	.2822**	.2801**	.3684**	.1661**	.3423**	.4413**				
NRISK28S	.1748**	.2306**	.4612**	.4819**	.5887**	.2530**	.4248**	.3690**	.3587**			
NRISK30S	.2380**	.3002**	.4141**	.4732**	.5589**	.3034**	.4013**	.3436**	.2999**	.6748**		
NRISK32S	.1889**	.2441**	.4470**	.4396**	.5211**	.2392**	.4118**	.3418**	.2875**	.6438**	.6266**	
NRISK34S	.2450**	.2824**	.4561**	.4702**	.5241**	.2829**	.4038**	.3452**	.2882**	.6165**	.6521**	.6848**
NRISK36S	.1914**	.2554**	.3915**	.3854**	.4488**	.2551**	.3797**	.3045**	.2323**	.5235**	.5472**	.6682**
NRISK38S	.2360**	.2738**	.3965**	.4333**	.4912**	.2749**	.3200**	.3133**	.2794**	.6140**	.6374**	.5582**
NRISK40S	.2440**	.3032**	.3871**	.4351**	.5101**	.3010**	.3501**	.2976**	.2546**	.5802**	.7324**	.5481**
NRISK42S	.2833**	.3501**	.3855**	.4440**	.5143**	.3307**	.3558**	.3193**	.2676**	.5589**	.6503**	.5229**
NRISK44S	.3765**	.3369**	.2762**	.3587**	.3547**	.3565**	.2809**	.2604**	.2286**	.3409**	.4396**	.3820**
NRISK46S	.3897**	.3957**	.3090**	.3827**	.4291**	.4358**	.3215**	.2936**	.2695**	.4027**	.4828**	.5954**
NRISK48S	.3138**	.2830**	.3402**	.3826**	.4203**	.2998**	.3088**	.2870**	.2551**	.4433**	.5169**	.5174**
NRISK50S	.1483**	.1461**	.3454**	.2895**	.4126**	.1297**	.2845**	.2766**	.2850**	.4507**	.3567**	.3727**
NRISK52S	.1364**	.1904**	.2581**	.3064**	.3961**	.2069**	.3320**	.2577**	.2226**	.4310**	.3756**	.3454**
NRISK54S	.1837**	.2350**	.3248**	.3654**	.4477**	.2330**	.3460**	.3093**	.3117**	.4939**	.4644**	.4532**
NRISK56S	.1759**	.2229**	.2944**	.3419**	.4242**	.2454**	.3330**	.2844**	.2680**	.4722**	.4265**	.3801**
NRISK58S	.1252**	.1167**	.2442**	.2500**	.2181**	.1342**	.1626**	.1544**	.1586**	.2657**	.2500**	.2429**

	NRISK34S	NRISK36S	NRISK38S	NRISK40S	NRISK42S	NRISK44S	NRISK46S	NRISK48S	NRISK50S	NRISK52S	NRISK54S	NRISK56S
NRISK36S	.5752**											
NRISK38S	.5789**	.4983**										
NRISK40S	.5960**	.5227**	.6462**									
NRISK42S	.5740**	.4830**	.6134**	.6997**								
NRISK44S	.4333**	.3963**	.4292**	.4672**	.4863**							
NRISK46S	.4478**	.3775**	.4652**	5052**	.5583**	.6553**						
NRISK48S	.5188**	.5145**	.5085**	.5260**	.6161**	.6643**	.6100**					
NRISK50S	.3535**	.3035**	.3533**	.3161**	.3314**	.2013**	.2574**	.2775**				
NRISK52S	.3513**	.3232**	.3009**	.3436**	.3404**	.2174**	.2647**	.2656**	.4350**			
NRISK54S	.4604**	.3950**	.4119**	.4121**	.4317**	.2875**	.3350**	.3446**	.4672**	.5176**		
NRISK56S	.3997**	.3506**	.3568**	.3839**	.3964**	.2726**	.3362**	.3264**	.4774**	.6539**	.6118**	
NRISK58S	.2973**	.2436**	.2650**	.2370**	.2323**	.2366**	.2376**	.2678**	.2094**	.1980**	.2276**	.2265**

* $p = .05$. **$p = .01$, two-tailed.

Received June 23, 1998
Revision received October 19, 1999
Accepted October 20, 1999 ■