Winter 1-1-2012

Applying Neural Network Models to Predict Recurrent Maltreatment in Child Welfare Cases with Static and Dynamic Risk Factors

Jennifer Marie Jolley
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

George Warren Brown School of Social Work

Dissertation Examination Committee:
Brett Drake, Chair
John Constantino
Jeff Gill
Melissa Jonson-Reid
Patricia Kohl
Shanta Pandey

Applying Neural Network Models to Predict Recurrent Maltreatment in Child Welfare Cases with Static and Dynamic Risk Factors

by

Jennifer Marie Jolley

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2012

Saint Louis, Missouri

Table of Contents

List of Figures

Figure

List of Tables

Table

Acknowledgments

The completion of this dissertation study would not have been possible without the intellectual guidance, financial support, and devoted assistance so generously provided by a veritable army of mentors, colleagues, and loved ones. I would like to extend my deepest gratitude to my advisor and Dissertation Chair, Dr. Brett Drake. His feedback has been invaluable in that he consistently assesses and helps to improve the conceptual clarity and the methodological rigor of my scholarship. His passion for transforming preventive service delivery systems, his expertise in analyzing administrative data, and his intuitive grasp of how best to relate statistical findings to daily child welfare practice have made my dissertation possible.

Additionally, I would like to thank all of my Dissertation Committee members for their generous support and their willingness to offer insightful critiques on all things related to my dissertation. Dr. Melissa Jonson-Reid sets the bar for rigorous applied research and is a pioneer in scholarship that assesses the effects of cross-sector service use on child safety and well-being. She inspired me to constantly ask myself, “what is the take-home message and how does this make a difference in child welfare policy and practice?” Dr. John Constantino is an extraordinary scholar who blends his expertise in intervention research with renowned clinical skills. He has provided me with important insights that have helped me to re-conceptualize how risk assessment can best be used to inform child welfare practice. Dr. Trish Kohl and Dr. Shanta Pandey have been generous in their willingness to provide substantive critiques of my scholarship. Dr. Jeff Gill made it possible for me to pursue the application of a very complex set of statistical procedures that required me to stretch well beyond my comfort zone.

I would also like to thank the members of the National Quality Improvement Center on Differential Response in Child Protective Services (QIC-DR) to include Dr. John Fluke, Dr. Donald Baumann, Dr. Ying-Ying Yuan, Ms. Lisa Merkel-Holguin, and Ms. Amy Hahn. The financial support provided to me as a recipient of a doctoral dissertation award funded through the Administration of Children and Families, Department of Health and Human Services, federal CFDA # 93-670, award # 90-CA-1762 (October 1, 2010-September 30, 2012) and administered through QIC-DR has been invaluable. I also deeply appreciate and have benefitted from the multiple opportunities to present findings from my evolving dissertation study and to receive feedback from experts who literally set the tone for advancements in child welfare research. Moreover, I would like to acknowledge the pre-doctoral training and funding that I have received to include my selection as a Centers for Disease Control-funded Brown Center for Violence and Injury Prevention (BCVIP) Ph.D. Scholar (R49CE001510-03) (August 1, 2010-July 31, 2011), and my selection as a National Institute of Mental Health Pre-Doctoral Fellow (T32MH019960) (August 1, 2007-July 31, 2010).

I would like to thank Mr. Ira Schwartz and Dr. Zoran Obradovic for their generous support in providing me assistance in conceptualizing and implementing a study that uses configural and neural network analyses to improve the accuracy of risk prediction in child welfare. Both Mr. Schwartz and Dr. Obradovic are recognized experts in the area of risk assessment prediction in child welfare, and their combined wisdom and experience helped me to advance a promising technology for improving the delivery of preventive services in child welfare. Furthermore, I would like to thank Mr. Doug Flowers for his stellar technical assistance in all matters related to the functioning of my software and computer, and Dr. Jason Brinkley for his assistance with JMP Pro 9. Mr. Flowers displayed unending and extraordinary compassion in the face of unanticipated

glitches, and his gift for finding the correct solution to an array of problems in record time is unparalleled. Dr. Brinkley also deserves my sincere gratitude for his ability to answer any number of wide ranging questions on a dime, in addition to what is a truly mad set of programming skills.

Finally, I would like to thank my Ph.D. colleagues and friends to include Marsela Dauti (soon-to-be Dr. Dauti), Dr. Gyansh Lama, Dr. Jin Huang, Dr. Ben Alexander-Eitzman, and Dr. Ben Lough. In the end, however, I owe everything that I have achieved to date to my partner, Dr. John Kerbs. My ability to see myself through John's eyes is what has and continues to inspire and sustain me each and every day. His kindness, strength, love, support, brilliance, and commitment to social justice have made it possible for me to truly engage in an intellectual journey that has ultimately led to the completion of my dissertation. John, you are and always will be my true north.

ABSTRACT OF THE DISSERTATION

Applying Neural Network Models to Predict Recurrent Maltreatment in Child Welfare Cases with Static and Dynamic Risk Factors

by

Jennifer Marie Jolley

Doctor of Philosophy in Social Work

Washington University in St. Louis, 2012

Professor Brett Drake, Chairperson

Risk assessment in child welfare has a long tradition of being based on models that assume the likelihood of recurrent maltreatment is a linear function of its various predictors (Gambrill & Shlonsky, 2000). Despite repeated testing of many child, parent, family, maltreatment incident, and service delivery variables, no consistent set of findings have emerged to describe the set of risk and protective factors that best account for increases and decreases in the likelihood of recurrent maltreatment. Shifts in predictors' statistical significance, strength, and direction of effects coupled with evidence of risk assessment models' poor predictive accuracy have led to questions regarding the fit between assumptions of linearity and the true relationship between the likelihood of recurrent maltreatment and its predictors (Gambrill & Shlonsky, 2000, 2001; Knoke & Trocmé, 2005). Hence, this dissertation study uses a distinctly nonlinear approach to modeling the likelihood of recurrent maltreatment by employing a combination of random forest and neural network models to identify the predictors that best explain the risk of recurrent maltreatment.

The risk of recurrent maltreatment was assessed for a cohort of children living in a large Midwestern metropolitan area who were first reported for maltreatment between

January 1, 1993 and January 1, 2002. Administrative child welfare records for 6,747 children were merged with administrative records from income maintenance, mental health, special education, juvenile justice, and criminal justice systems in order to identify the effects that various public sector service system contacts have on the risk of recurrent maltreatment. Each child was followed for a period of at least seven years to identify the risk of recurrent maltreatment in relationship to a second report for maltreatment.

Post-hoc analyses comparing the predictive validity of the neural network model and a binary logistic regression model with random intercepts shows that the neural network model was superior in its predictive validity with an area under the ROC curve of 0.7825 in comparison with an area under the ROC curve of 0.7552 for the logistic regression model. Additional post-hoc analyses provided empirical insight into the four prominent risk factors and four risk moderating service variables that best explain variation in the risk of recurrent maltreatment. Specifically, the number of income maintenance spells received, community-level poverty, the child's age at the first maltreatment report, and the parent's status as the perpetrator of the first maltreatment incident defined 21 risk-based groups where the average probability of recurrent maltreatment was dependent upon values for the four primary risk factors, and the risk of maltreatment was moderated by juvenile court involvement, special education eligibility, receipt of CPS family centered services, and the child's receipt of a mental health/substance abuse service in the community. Findings are discussed within a Risk-Need-Responsivity theory of service delivery (Andrews & Bonta, 2006), which links the empiricism of risk assessment with the clinical implementation of a preventive service delivery plan through the identified modifiable risk factors that drive the likelihood of recurrent maltreatment.

Chapter 1: Improving Service Delivery to Prevent Recurrent Maltreatment

Problem Statement and Focus of Study

Differential response (DR) systems within child welfare seek to prevent future episodes of child maltreatment by matching the delivery of prevention services to family needs. This is an arduous task, given that 5.8 million children were reported to Child Protective Services (CPS) for child abuse/neglect in 2007 across 3.2 million separate referrals (USDHHS, 2009). With studies showing rates of re-referrals for abuse/neglect that range from 29% over a follow-up period of 18 months (English, Marshall, Brummel, & Orme, 1999; Marshall & English, 1999) to 62% over a follow-up period of 7.5 years (Drake, Jonson-Reid, & Sapokaite, 2006), it is imperative that DR systems implement an effective risk assessment and treatment planning protocol to guide the delivery of prevention services. However, the effective delivery of prevention services in DR systems is hindered by the lack of a theoretically-derived and empirically-supported protocol for risk assessment and treatment matching. Risk assessment and treatment matching protocols used in DR systems are compromised in three specific ways. First, risk assessment protocols used in child welfare perform poorly in relationship to differentiating cases at high risk of recurrent maltreatment from cases at low risk of recurrent maltreatment (Knoke & Trocmé, 2005; Schwartz, Jones, Schwartz, & Obradovic, 2008). Second, current risk protocols are compromised in their ability to assist workers in service planning and provision by a heavy reliance on *static* (i.e., unchangeable) items related to past behavior as opposed to identifying *dynamic* (i.e., changeable) factors that are amenable to services and that drive the likelihood of repeat maltreatment (Bae, Solomon, & Gelles, 2007, 2009; DePanfilis & Zuravin, 1999b; Gambrill & Shlonsky, 2000; Taxman, 2006). Third, an accurate assessment of the effectiveness of prevention service delivery is compromised by the lack of connection between risk assessment and service planning and provision (Bae, Solomon, & Gelles, 2009; DePanfilis & Zuravin, 1999b, 2001, 2002; Holder, 2000; Huebner, 2005; Loman, 2006; Loman & Siegel, 2004a; Shlonsky & Wagner, 2005). The effectiveness or responsivity of prevention services rests upon (a) matching the intensity of service delivery (i.e., degree of protective oversight, service type, dosage, and duration) to the overall likelihood that future maltreatment will occur, and (b) matching the interventions to the dynamic factors that drive the likelihood of future maltreatment (Andrews & Bonta, 1998, 2006; Andrews, Bonta, & Wormith, 2006; Ferguson, 2002; Taxman, 2006).

This dissertation study sought to improve the accuracy and utility of risk assessment for treatment planning in DR systems by applying innovative statistical methods to administrative data from Missouri, one of the first states in the US to implement DR. Specifically, this study applied neural network modeling to accurately classify families at risk of recurrent maltreatment[1] and to assist workers in identifying the dynamic risk factors that are most amenable to targeted services/interventions (Bishop, 1995; Cheng & Titterington, 1994; Garson, 1991; Paik, 2000). This study extends previous research that has successfully applied neural network modeling to accurately classify 90% of cases at risk of substantiation in the child welfare system and 80.6% to 97% of juveniles at risk of re-arrest in the juvenile justice system (Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008; Schwartz, Kaufman, & Schwartz, 2004). The approach to this study is also congruent with theoretical aspects of risk assessment and service planning/provision as articulated by the Risk-Need-Responsivity (RNR) model. The RNR model is a criminological theory of rehabilitative service delivery to offenders in the criminal justice system (Andrews & Bonta, 2006). Summarized as the “principles of effective intervention” (Cullen, 2005, p. 16), RNR is an evidence-based system of assessment and intervention that has been shown to statistically reduce rates of recidivism by addressing factors that drive the likelihood of future criminal and antisocial activity (Andrews & Bonta, 1998, 2006; Andrews, Bonta, & Hoge, 1990; Andrews, Bonta, & Wormith, 2006; Andrews & Dowden, 2007). More specifically, the assessment/intervention protocol emphasizes the importance of (a) assessing risk in relation to *both* static and dynamic risk factors, and (b) providing services that are responsive to identified dynamic risk factors that drive legally-liable behavior. Thus, the RNR principles will be used to specify variable selection for this study’s neural network models through the inclusion of relevant static and numerous dynamic factors that could drive the risk of recurrent maltreatment.

The problem of recurrent maltreatment is described below in greater detail to include a description of the number of children at risk of recurrent maltreatment, the devastating effects of maltreatment, and the potential for serious cumulative effects as a child’s exposure to maltreatment continues. The problem of recurrent maltreatment is then placed in the context of the delivery of prevention services within DR systems in child welfare. The delivery of prevention services in DR is examined in relationship to (a) how children and families gain access to prevention services, (b) the proportion of children and families who receive prevention services, and (c) the proportion of children and families who have three or more subsequent reports of maltreatment beyond an initial report for abuse/neglect.

Recurrent Maltreatment: Prevalence, Costs, and Consequences

In 2007, 5.8 million children were reported to Child Protective Services (CPS) for child abuse/neglect across 3.2 million separate referrals (USDHHS, 2009). Costs attached to the delivery of medical and non-medical services to maltreated children are high, with about 94 billion dollars a year or 100,00 dollars per maltreated child spent to provide immediate and longer-term care (Foster, Prinz, Sanders, & Shapiro, 2008; Fromm, 2001). The short-term and long-term effects of maltreatment can be devastating to children as evidenced by outcomes such as permanent physical deformities, neurological impairments, anxiety, depression, posttraumatic stress disorder, impaired attachment and relationship-building capacities, increased antisocial behaviors and aggression, and decreased academic performance (Cicchetti & Toth, 2005; Éthier, Lemelin, & Lacharité, 2004; Rosenberg & Krugman, 1991; Scott, Wolfe, & Wekerle, 2003). Furthermore, a high percentage of families that have been reported to CPS for abuse or neglect are subsequently re-reported for maltreatment. Rates of re-referral have ranged from 29% over a follow-up period of 18 months (English, Marshall, Brummel, & Orme, 1999; Marshall & English, 1999) to 62% over a follow-up period of 7.5 years (Drake, Jonson-Reid, & Sapokaite, 2006). Not only is the rate of maltreatment re-referral high, but the likelihood of repeat maltreatment increases with each subsequent occurrence of maltreatment (Fluke, Yuan, & Edwards, 1999). Moreover, as the exposure to maltreatment continues, the effects are cumulative. Studies have shown that *recurrent maltreatment* presents substantial and cumulative risk to a child's future through such negative outcomes as *aggression* (Bolger & Patterson, 2001; Dodge, Pettit, & Bates, 1997; Graham et al., 2010; Manly, Cicchetti, & Barnett, 1994), *peer rejection* (Bolger & Patterson, 2001; Bolger, Patterson, & Kupersmidt, 1998), *impaired social and daily living skills* (English et al., 2005), *anxiety and depression* (Éthier, Lemelin, & Lacharité, 2004), *posttraumatic stress* (English, Graham, Litrownik, Everson, & Bangdiwala, 2005), and *delinquency* (Ireland, Smith, & Thornberry, 2002; Smith & Thornberry, 1995).

The Delivery of Prevention Services in DR Systems

As noted by Fluke (2008), “the reduction of reentry is most likely to be achieved by attending to how a CPS agency intervenes with children and families” (p. 750). One of the more recent and important CPS reforms that aims to improve CPS effectiveness and efficiency is the implementation of differential response systems to accepted reports of maltreatment (Kaplan & Merkel-Holguin, 2008; Waldfogel, 2008). DR systems are premised on the assumption that reentry will be reduced if workers tailor service planning and provision to the incident-, child-, and family-level characteristics in each case of reported maltreatment (Christenson, Curran, DeCook, Maloney, & Merkel-Holguin, 2008; Conley & Duerr Berrick, 2008; Kaplan & Merkel-Holguin, 2008; Waldfogel 1998, 2000a, 2000b, 2008). In contrast to traditional CPS investigations that determine if an act of abuse/neglect occurred, workers in DR systems tailor their response to each case by assessing for risk at two critical decision points. These two critical decision points occur when (a) workers decide how to respond to accepted reports of maltreatment by assigning cases to a particular service pathway (also referred to as track assignment), and (b) when workers prepare to deliver prevention-oriented services after cases have been assigned to a particular service pathway or track (Loman & Siegel, 2004a, 2004b, 2004c, 2005; Yuan, 2005).

Risk of substantiation, track assignment, and the delivery of prevention services.

The *first critical decision point* where risk is assessed in DR systems is at the point of track assignment. Track assignment determines the degree of CPS oversight and the system’s response in relationship to the worker’s assessment of the child’s protection needs at the time the allegations were reported (Christenson, Curran, DeCook, Maloney, & Merkel-Holguin, 2008; Johnson, Sutton, & Thompson, 2005; Loman & Siegel, 2005;

Sawyer & Lohrbach, 2005a; Schene, 2005; Yuan, 2005). Workers assess the risk or likelihood that the report of maltreatment will be substantiated and is likely to require court intervention in order to protect the child from imminent harm (Johnson, Sutton, & Thompson, 2005; Loman & Siegel, 2005; Sawyer & Lohrbach, 2005a; Schene, 2005). If the risk of substantiation is high, cases are assigned to the *investigation track* (Johnson, Sutton, & Thompson, 2005; Loman & Siegel, 2005; Sawyer & Lohrbach, 2005a; Schene, 2005). A forensic investigation is conducted and evidence of a prior act of maltreatment is collected and vetted according to a standard of evidence defined in state statute (Waldfogel, 2008). Workers who conduct investigations are engaged in the delivery of protection-oriented services. Evidence of a past act of maltreatment along with a formal determination that identifies a child victim and an adult perpetrator are necessary when workers seek court involvement to protect children from imminent harm. Protection-oriented services may include out-of-home placement or forced compliance with services that prioritize immediate harm reduction in relationship to a set of characteristics or conditions that have been identified as endangering the child's immediate safety (Conley, 2007; Drake & Jonson-Reid, 2000; Schene, 1998).

In contrast to the investigation track, if the risk of substantiation is moderate to low, cases are assigned to the non-investigation or family *assessment track* (Johnson, Sutton, & Thompson, 2005; Loman & Siegel, 2005; Sawyer & Lohrbach, 2005a; Schene, 2005). An assessment of the family's need for supportive services is conducted; service plans are tailored to the worker's professional judgment in combination with the family members' input (Christenson, Curran, DeCook, Maloney, & Merkel-Holguin, 2008; Conley, 2007; Conley & Duerr Berrick, 2008; Connolly, 2005). While the child's immediate or short-term safety is assessed, the primary purpose of the family assessment track is the provision of prevention-oriented services to children and families. These services promote the child's long-term safety and well-being by reducing the risk for future maltreatment (where the future is framed in the long term and risk of future maltreatment measures the likelihood that a child will be subsequently abused/neglect in the future if services are not delivered) (Loman & Siegel, 2005; Schene, 2005). Services support family functioning with the assumption that improvements in basic family functioning and the members' well-being will prevent future maltreatment by decreasing the factors that contribute to the family's risk of future maltreatment (Loman, 2006; Loman & Siegel, 2004a, 2004c; Schene, 2005).

Risk of future maltreatment and the delivery of prevention services in both tracks.

The *second critical decision point* where workers determine service delivery based on risk assessment occurs after the family has been assigned to a track as described above. Workers assess the risk or likelihood that a child will be maltreated in the future (Loman, 2006; Loman & Siegel, 2004a, 2004b, 2004c). In theory, regardless of track assignment or the presence or absence of a DR system, the assessment of the risk of future maltreatment is used to determine the family's need for prevention-oriented services (DePanfilis & Zuravin, 2001; Doueck, English, DePanfilis, & Moote, 1993; English & Graham, 2000; Jagannathan & Camasso, 1996). These services influence the underlying causes of maltreatment and reduce the likelihood of future abuse/neglect by changing the dynamics of family functioning (Schene, 2005). Nonetheless, in a national study of states with and without DR systems, families that are investigated may or may not receive prevention-oriented services following the completion of the forensic investigation (Fluke, 2009).

More specifically, evidence suggests that post-investigation service delivery apart from out-of-home placement is low for families being investigated (English, 1998; Jonson-Reid, 2002; Loman 2006; Loman & Siegel 2005; Ortiz, Shusterman, & Fluke, 2008; Schene, 1998, 2005; Waldfogel, 1998; Yuan, 2005). For example, using data from the National Child Abuse and Neglect Data System, Fluke (2009) reported that 59% of families with substantiated cases received post-investigation services, while 30% of families with unsubstantiated cases received post-investigation services. However, once placement services were removed from the analysis, just 38% of families with substantiated cases received post-investigation services, and 28% of families with unsubstantiated cases received post-investigation services. Furthermore, numerous studies on the predictors of repeat maltreatment have reported that post-investigation services increased the likelihood that the child would experience repeat maltreatment (Bae, Solomon, & Gelles, 2007, 2009; DePanfilis & Zuravin, 1999a, 2001; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fluke, Yuan, & Edwards, 1999; Lipien & Forthofer, 2004; Marshall & English, 1999; Ortiz, Shusterman, & Fluke, 2008).

Families assigned to the assessment track predominantly receive prevention-oriented services following the implementation of a family needs assessment (Kaplan & Merkel-Holguin, 2008; Loman & Siegel, 2005; Shusterman, Fluke, Hollinshead, & Yuan, 2005; Yuan, 2005). However, as noted by the National Quality Improvement Center on Differential Response in Child Protective Services (2009), literature on service delivery within the assessment track is sparse. In particular, there is a lack of research as well as policy and practice guidelines on the process of using an assessment to inform service planning and provision. For example, a study of DR in Minnesota compared outcomes between families assigned to an investigation or an assessment track; the study employed an experimental design in 14 counties where reports of maltreatment were accepted for response, deemed appropriate for the assessment track, and then randomly assigned to either an investigation or an assessment response (Loman & Siegel, 2004c). The evaluators (Loman & Siegel, 2005) noted the important increase in both the number of preventive family support services offered to families in the assessment track and the proportion of families that received services in the assessment track (36%) as compared to the investigation track (15%). Family support services were summarized as being targeted towards families' basic economic needs. Caregivers were asked to report on the status of the following outcomes one year after their final contact with CPS following the initial maltreatment incident and DR response: (a) their children's aggressive and uncontrolled behavior, (b) their children's relationships in school and academic progress, (c) their own disciplinary methods, (d) their own ability to care for their children, (e) their living arrangements, and (f) their emotional and financial support from friends and relatives. Surprisingly, there were no statistically significant differences in these outcomes when comparing responses from caregivers in families assigned to either the assessment or investigation track (Loman & Siegel, 2005). This is noteworthy because the assumption underlying the delivery of supportive services to families in the assessment track is the belief that families who receive supportive services to improve family functioning and child well-being will in turn benefit from a decreased likelihood of future maltreatment. Thus, child, caretaker, and family-related outcomes such as the ones measured above should improve before the likelihood of future maltreatment can decrease. Prior to providing services, workers implemented the Minnesota Structured Decision Making (SDM) Family Risk Assessment (FRA) instrument (Loman & Siegel, 2004a, 2005), but no information was provided as to how workers used the scores from the FRA to guide service delivery in terms of intensity (i.e., service type, dosage, and duration as well as the degree of CPS oversight) and/or identifying targets for treatment (i.e., the areas of family functioning that are strongly associated with the likelihood of future maltreatment and that will be influenced by particular interventions).

Incidence and costs of repeat maltreatment in the context of DR systems.

Families referred to an assessment track have been shown to have the same or slightly lower re-report rates as families referred to an investigation track (English, Wingard, Marshall, Orme & Orme, 2000; Loman & Siegel, 2005; Ortiz, Shusterman, & Fluke, 2008; Shusterman, Fluke, Hollinshead, & Yuan, 2005). However, there is little evidence to show that service delivery within either track decreases rates of re-reports or improves child and family outcomes substantively. Rates of re-reports are still relatively high for both tracks, where almost one-third of the families were re-reported for a subsequent incident of abuse or neglect (English, Wingard, Marshall, Orme & Orme, 2000; Loman & Siegel, 2005). Furthermore, evaluation studies of DR systems in Missouri and Minnesota addressed the importance of defining and identifying those families that are frequently encountered in the DR system (Loman, 2006; Loman & Siegel, 2004b). Thus, these families have multiple accepted reports for maltreatment to which the DR system responds with an investigation or family assessment beyond an initial accepted report for abuse/neglect.

In the case of Missouri, for example, the evaluators used a quasi-experimental design that included 14 small and medium-sized counties and selected zip codes in St. Louis City and County that implemented DR systems (the demonstration sites) and 14 small and medium-sized counties and selected zip codes in St. Louis City and County that did not implement DR systems (the comparison sites). The demonstration counties and areas denoted by zip code and the comparison counties and areas denoted by zip code were similar according to population characteristics and child welfare caseload characteristics (Loman & Siegel, 2004b; Siegel & Loman, 2000). Families were selected from the demonstration and comparison sites from July 1995 to June 1997 and then followed in the child welfare data system for a period of five years until November 2002. When the cut-off point was set to three subsequent hotline reports of maltreatment beyond an initial hotline report of maltreatment, just over one-third or 34.2% (n = 2,637) of 7,711 families were engaged in chronic abuse/neglect.

In order to describe the costs of delivering in-home prevention-oriented services and out-of-home protection-oriented services to families that were re-reported to CPS at least three subsequent times, classes of "costly" families were created (Loman & Siegel, 2004b, p. 14). Costly was calculated in relationship to each family's service duration and total service expenditures. Service duration was tracked by identifying the longest period of service delivery for each family and then separately summing (a) the number of days each family had an open family-centered services (FCS) case, and (b) the number of days each family had a child in out-of-home placement (i.e., alternative care). Additionally, total service expenditures were calculated by summing expenses related to each family's services across the length of the study period. The cut-off point used to identify costly families was the 80th percentile. Thus, families that were at or above the 80th percentile for days marked as having a case open for FCS services, days with a child in alternative care services, or total expenditures were defined as costly (Loman & Siegel, 2004b). Finally, families were classified into three types of chronic groups by the total number of subsequent hotline reports for maltreatment each family received and by defining the family's service use as costly (as per the aforementioned definition of costly service delivery). In sum, just under 1 in 10 families or 9.3% (n = 720) of 7,711 families fell in one of the following categories: (a) the family had three or four subsequent hotline reports for maltreatment and met the definition of costly service use, (b) the family had five or six subsequent hotline reports for maltreatment and met the definition of costly service use, and/or (c) the family had seven or more subsequent hotline reports for maltreatment and met the definition of costly service use.

While chronic and costly families accounted for 9.3% of the total study population, these families accounted for 41.9% of all service-related expenditures across alternative care (e.g., placement in foster care), daycare (e.g., daycare for protective services cases), children's treatment services (e.g., prevention-oriented in-home services such as counseling and family therapy), and residential treatment (e.g., placement in a residential treatment facility). Total expenditures for the five-year study period amounted to 67.7 million dollars, with 28.4 million dollars spent on the 720 chronic and costly families. The average cost of service delivery to chronic and costly families was seven times the cost of service delivery to families that were not defined as chronic and costly. Specifically, the average cost of services to chronic and costly families was 39,542 dollars as compared to 5,630 dollars for families not identified as chronic and costly (Loman & Siegel, 2004b).

Clearly, the costs and consequences of recurrent maltreatment are high and the proportion of children and families who are re-reported for abuse/neglect is substantial. While the delivery of prevention services is a key objective that informs DR systems in general, the effective delivery of prevention services is compromised by the inherent flexibility of DR systems. The inherent flexibility of service delivery in DR systems has been particularly problematic in relationship to the lack of a standardized protocol that integrates risk assessment with treatment planning as discussed below.

Barriers to the Implementation of Effective Prevention Services in DR: The Flexible Nature of DR Service Delivery

Conceptually and pragmatically, DR is more of a philosophy of intervention as opposed to being a theoretically-derived and empirically-tested intervention (Conley, 2007). At the core of DR philosophy is a privileged notion of flexible service delivery. Flexibility is what drives DR and its potential for reforming the ways in which child welfare systems and workers respond to accepted reports of maltreatment. How local agencies in various counties and states implement DR is highly variable because the driving force of DR – flexibility – deliberately introduces variation in local policies and worker practices (Kaplan & Merkel-Holguin, 2008). This is particularly true as workers are encouraged to be flexible in their approach to assessment and treatment planning for the delivering of prevention services; family needs are co-determined and service plans are based upon family members' input, workers' assessment skills, and the availability of community resources (Conley & Duerr Berrick, 2008; Waldfogel, 2000a, 2000b).

In fact, Kaplan and Merkel-Holguin (2008) described the lack of a standardized approach to decision-making by workers in DR systems in a summary of the qualitative and quantitative findings of the *National Study on Differential Response in Child Welfare,* a national survey designed to identify DR practices implemented in child welfare agencies across the US (Merkel-Holguin, Kaplan, & Kwak, 2006). While agency-based guidelines may exist, flexibility is emphasized as a core component of service delivery within DR programming. Thus, Kaplan and Merkel-Holguin (2008) described the inherent flexibility of DR systems as existing in part due to the strong influence of clinical judgment and worker discretion outside of a standardized assessment and intervention protocol:

> It became readily apparent that workers' clinical judgment and discretion were of great importance in the implementation of differential response. There are few *hard and fast rules* that cannot be altered given the practice wisdom of a specific worker and the approval of a supervisor….While intake and screening systems have discrete guidelines for assigning cases to the response pathways, many of these systems also support case-level decision making in determining the appropriate response. (Kaplan & Merkel-Holguin, 2008, p. 11)

Further evidence of the inherent flexibility in DR systems and the lack of a standardized assessment and intervention protocol is provided in the findings from the *National Study of Child Protective Services Systems and Reform Efforts*; this study included a review of state CPS policy (as of 2002) and a survey of the standard practices employed (as of 2002) by a nationally representative sample of local CPS agencies serving 300 counties (USDHHS, 2003a, 2003b, 2003c). Findings from the policy analysis on the purpose of service delivery during an assessment response emphasized the provision of a family-based assessment and the provision of services, but the purpose of service delivery was not predominantly based upon preventing future abuse and neglect (USDHHS, 2003c). The analysis of state policies was limited to 20 states that identified the existence of an alternative or assessment response, where only 11 out of the 20 states reported the existence of a statewide assessment response. Specifically, 70% of the states required the provision of family assessments, and 65% of the states required that services be delivered as a part of the CPS response. An analysis of state policies revealed that the purpose of the assessment response varied where 55% of the states specified that protecting child safety was the purpose of an assessment response, 45% of the states specified that strengthening the family was the purpose of an assessment response, and *only 20% of the states specified that preventing child abuse and neglect was the purpose of the assessment response* (USDHHS, 2003c).

In reference to required standard practices workers must complete during an assessment, (a) 53% of the agencies reported making an assessment of the immediate service needs of the family, (b) 50% of the agencies reported removing the child if immediate safety was an issue, (c) 42% of the agencies reported making a determination of whether the child had been maltreated, and (d) 41% of the agencies reported making an assessment of the service needs of the child. In contrast, *only 32% of the agencies reported assessing the underlying causes of the maltreatment incident*, 27% of the agencies reported providing short-term services if needed, and 28% reported referring the family for further services if needed (USDHHS, 2003b). *Furthermore, there was a low proportion of agencies that reported using standardized assessment tools to aid in identifying underlying causes of maltreatment and in planning for appropriate service delivery*. Specifically, (a) only 30% of the agencies reported using a formal risk assessment instrument, (b) 14% reported using a standardized family support assessment, (c) 12% reported using a standardized substance abuse assessment instrument, (d) 11% reported using a standardized domestic violence assessment instrument, (e) 9% reported using a standardized child development inventory, and (f) 6% reported using a standardized parenting skills assessment (USDHHS, 2003b).

While flexibility gives workers the freedom to move beyond a narrowly prescribed set of actions that may not be linked to family needs, the flexibility that allows agencies and workers to locally implement DR practices and policies in a myriad of ways also makes it more difficult to evaluate service delivery systems and intervention effectiveness. In sum, nothing in the DR movement appears to be standardized in relation to policy and/or practice protocols for linking risk assessment to service planning and the provision of prevention services. Indeed, the variation in service delivery is so great that no study to date has described (a) which types of families (e.g., by overall risk level, by dynamic risk factors, by family characteristics, etc.) receive what services post-investigation or post-assessment; (b) how workers use assessments to inform service planning and provision; and (c) the characteristics of services to include type, dosage, duration, and quality (Fluke, 2009; National Quality Improvement Center on Differential Response in Child Protective Services, 2009). This high level of variation makes it very difficult for workers to compare their performance to one another as well as to gold standards for assessment and service provision practices. Because there are so many options for assessment and service provision, workers have no stable reference points against which they can evaluate their effectiveness.

Given such variability, the first step in understanding and improving DR rests upon the need to start controlling for a good portion of the variation at the initial decision points that require standardized protocols for integrating risk assessment with service planning and the provision of prevention services. The implementation of such protocols would establish a common baseline within and across local DR systems. Drawing from the literature on evidence-based practice in child welfare (see e.g., Barth, 2009; Chaffin & Friedrich, 2004; Luongo, 2007; Maher, Jackson, Pecora, Shultz, Chandra, & Barnes-Proby, 2009), the implementation of a standard approach to intervention, one that is preferably rooted in a clearly explicated theoretical framework, is a key component of effective practice. The flexibility of DR would be shifted. Instead of being housed in a wide-open field where almost anything goes, flexibility would be housed in a standardized, theoretically-derived, and empirically-tested assessment and treatment protocol. Workers would have the flexibility to provide services that best meet family needs, but within an evidence-based context that is designed to decrease the likelihood of future abuse/neglect.

The Lack of a Protocol for Matching Services to Prevention Needs

In addition to the high percentage of families that had three or more subsequent reports for maltreatment (just over one-third) and the spiraling cost of service delivery to families that repeatedly return to the child welfare system for reports of abuse/neglect, Loman and Siegel (2004b) noted that service delivery in both the investigation and assessment tracks was not effective in preventing ongoing reports of maltreatment for chronic families: "It appears that they [chronic child abuse and neglect families] are unaffected whether they are approached with traditional investigations or with the new family assessment approach" (p. 13). Thus, in both tracks, there appears to be a failure to match (via treatment matching protocols) prevention needs with effective services that subsequently decrease rates of recurrent maltreatment. Furthermore, it was noted that current SDM tools used to assess risk were not effective in accurately classifying recurrent maltreatment families from non-recurrent maltreatment families. Specifically, Loman and Siegel (2004b) reported that the identification of chronic families will be dependent not only on initial assessments based on history of DFS and ratings on tools like the SDM scales (considered below) but also on *full assessments of families that examine underlying problems that may be related to Chronic CA/N*. (p. 16)

Yet, the evaluators did not go on to identify how workers have or should consistently apply assessments to identify the factors driving the risk of chronic abuse/neglect.

Improving the delivery of prevention services to children and families in both tracks in DR systems depends upon (a) an accurate assessment of the overall risk of future maltreatment, and (b) the identification of dynamic risk factors that drive the likelihood of future maltreatment. As discussed in the following sections, risk assessment in child welfare has been plagued with problems of inaccuracy. Furthermore, risk assessment has not been successfully integrated with treatment planning, and problems with integration have been exacerbated by the lack of a theoretical framework guiding the development of risk assessment instruments and the link between assessing risk and decisions related to service provision.

Risk Assessment in Child Welfare: Predictive Accuracy and Treatment Matching Issues

The absence of a theoretical framework.

The development and implementation of risk assessment instruments is nothing new. In fact, English and Graham (2000) noted that states began developing risk assessment models in the mid-1980s for the purpose of creating explicit decision-making guidelines to improve CPS efficiency and effectiveness. Specifically, the CPS system and the children and families served by the system supposedly benefitted from a decision-making protocol that assisted workers with (a) the efficient identification of cases that were most in need of limited resources (Ryan, Wiles, Cash, & Siebert, 2005; Rycus & Hughes, 2008), and (b) the effective clarification of options to reduce the likelihood of future abuse and neglect via case planning and service provision (Doueck, English, DePanfilis, & Moote, 1993; Shlonsky & Wagner, 2005). In addition, structured decision-making guidelines utilized (among other things) risk assessment protocols to improve the consistency of workers' actions and to promote children's safety in a least restrictive manner (English & Graham, 2000; Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008).

While decision-making guidelines were developed for the purpose of improving case outcomes by working to establish a link between a child's risk of future maltreatment and the worker's provision of services to prevent future abuse/neglect, no specified theoretical model was used to guide the development of risk assessment instruments from the mid-1980s to the present (Camasso & Jagannathan, 2000; Doueck, English, DePanfilis, & Moote, 1993; Jagannathan & Camasso, 1996; Wald & Woolverton, 1990). As stated by English and Graham (2000), "the concept of comprehensive assessment of risk was based on the underlying theoretical assumption that child abuse and neglect is a multi-dimensional problem that requires multi-dimensional explanations" (p. 898). It is of note, however, that the presence of a theoretical assumption regarding the multi-dimensional nature of child abuse/neglect does not constitute a fully-developed and articulated theory of repeat maltreatment. The absence of a well-developed theoretical model guiding risk assessment and intervention research is congruent with a study conducted by English, Marshall, Brummel, and Orme (1999). These authors noted that research on the predictors of repeat maltreatment as well as interventions to prevent future maltreatment lacked an explicitly stated theory of recurrent abuse/neglect. Specifically, the authors noted the following:

> Risk factor research is primary atheoretical, focusing on the identification of correlates associated with different CPS decision points that are based on data available in case records. Although there may be theoretical underpinnings for CPS work, the theory is not explicitly articulated as part of the research design in most studies. Research on interventions is more

> likely to explicate a theory, that is, to identify the factors believed to be associated with the occurrence of child maltreatment and why the intervention should ameliorate those factors. However, typically in intervention research, the outcome of recurrence is not explained in theoretical terms. (English, Marshall, Brummel, & Orme, 1999, p. 298)

In addition to the lack of a well-specified theoretical framework that informs risk assessment instruments, the scholarly literature lacks a theoretical framework that also connects risk assessment to treatment planning. For example, Shlonsky and Wagner (2005) described the importance of using an actuarial risk assessment tool to assess the likelihood of future abuse/neglect at the close of an investigation. This would help to (a) determine if a service case should be opened, and (b) determine if a family should received prioritized access to services. As noted by the authors, “the actuarial risk assessment tool is used to help establish the intensity of the CPS response” (Shlonsky & Wagner, 2005, p. 422). However, the family’s targets for treatment should then be assessed with a separate tool (e.g., the California Family Strengths and Needs Assessment) that supports the standardized collection of data to assist in the worker’s clinical assessment of the family. Although both the actuarial risk assessment tool and the clinical assessment tool could be packaged within a system of instruments referred to as structured decision making (SDM), there was no explication as to how the various tools were linked theoretically or empirically to (a) each other, (b) a reduction in the likelihood of future maltreatment, and/or (c) worker decisions related to service planning and provision. Furthermore, Shlonsky and Wagner (2005) reported that the use of a separate standardized clinical assessment tool (i.e., the California Family Strengths and Needs Assessment) to supplement the actuarial risk assessment tool was only the first step in the provision of prevention-oriented services. Workers should also implement a comprehensive clinical assessment that may include additional screening approaches to evaluate specialized substance abuse and mental health needs.

Lack of predictive accuracy.

Most risk assessment tools used by states are either consensus-based (i.e., predictors are selected for the tool on the basis of expert consensus) or actuarial (i.e., predictors are included on the basis of their strong empirical association with repeat maltreatment, having been identified through the application of a statistical procedure on a particular dataset) (English & Graham, 2000). In either case, they explain little variance in the likelihood of recurrent maltreatment (~15%) (Baird & Wagner, 2000). Furthermore, risk assessment tools used in child welfare have high rates of false positives (14-29%) and false negatives (22-45%); thus, as many as 1 in 3 children have been falsely identified as being likely to suffer subsequent maltreatment, and as many as 1 in 2 children have been falsely identified as being unlikely to suffer subsequent maltreatment (Knoke & Trocmé, 2005). The consequences of misclassifying a child's/family's risk of future maltreatment are very serious because risk levels are to be used by workers to determine a child's/family's access to prevention-oriented services to include service type, dosage, and duration as well as the degree of agency oversight and court involvement that will be invoked (Camasso & Jagannathan, 2000; Gambrill & Shlonsky, 2001; Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008). Thus, false positives may lead to the invasive application of services that are unnecessary, costly, and counterproductive to children and their families. False negatives may lead to a lack of service delivery that could have prevented another incident of abuse/neglect had appropriate interventions been matched to the child's/family's overall risk level and the specific factors driving the likelihood of future maltreatment.

Although workers may be able to select individual risk factors that are associated with an outcome such as repeat maltreatment, they lack the ability to use clinical judgment to appropriately select, weight, and combine the particular risk factors that best predict outcomes to include recurrent maltreatment (Baird & Wagner, 2000; Jagannathan & Camasso, 1996; Gambrill & Shlonsky, 2000). It is for this reason that actuarial tools have been found to outperform individual and group-based (consensus-based) expert judgment (Baird & Wagner, 2000; Baird, Wagner, Healy, & Johnson, 1999; Gambrill & Shlonsky, 2000; Shlonsky & Wagner, 2005). That said, the typically used actuarial algorithm assumes a linear relationship among the predictors where a constant and additive weight is multiplied by each particular predictor (typically *static* predictors) such that for every one-unit of increase in x, there is a constant increase in y. Yet, some of the most useful predictors of future maltreatment are *dynamic* and therefore amenable to intervention. Moreover, these factors may not have a linear relationship with repeat maltreatment. As noted by Gambrill and Shlonsky (2000):

> Risk may not be additive (i.e., adding deficits and subtracting strengths), but may be multiplicative (i.e., a specific combination of risk factors modifies their individual effect, increasing or decreasing risk in different ways) or have some non-linear function. The possibility of this type of interaction among predictor variables highlights the strength of actuarial models as it is highly unlikely that an unassisted individual could accurately carry out these types of calculations, especially given time constraints. Interaction should be explored in greater detail to take

advantage of the benefits of statistical models for estimating risk. The availability of larger data sets should enable more accurate assessment of interactive terms. (p. 828)

Overall, actuarial risk assessment instruments have demonstrated better reliability and validity in comparison to clinical judgment and consensus-based risk-assessment instruments. However, actuarial instruments have yet to demonstrate a *satisfactory* level of predictive accuracy (Gambrill & Shlonsky, 2000, 2001; Schwartz, Jones, Schwartz, & Obradovic, 2008; Shlonsky & Wagner, 2005). To date, the literature that specifically focuses on evaluating the performance of actuarial risk assessment instruments is limited (Cash, 2001). The Michigan Structured Decision Making (SDM) System's Family Risk Assessment of Abuse and Neglect was described by Baird and Wagner (2000) as "the most widely used actuarial-based approach" in child welfare (pp. 844-845). At the time of publication, Baird and Wagner noted that Michigan's Family Risk Assessment (FRA) was being implemented in Minnesota (among other states).

Baird and Wagner (2000) examined the Michigan FRA's predictive validity and determined that it was substantively better than the predictive validity of two widely-used consensus-based instruments (the Washington Risk Assessment Matrix and the California Family Assessment Factor Analysis). In short, the FRA was stronger in its ability to (a) produce risk classifications of families with significantly different re-investigation rates, and (b) produce risk levels that increased in conjunction with re-investigation rates (Baird & Wagner, 2000; Shlonsky & Wagner, 2005). Unfortunately, no information was provided in relationship to the Michigan FRA's utility for treatment planning. Moreover, the authors did not provide information regarding true positives, false positives, true negatives, false negatives, the instrument's sensitivity (sensitivity is calculated as the proportion of true positives divided by the sum of true positive and false negatives), and/or the instrument's specificity (specificity is calculated as the proportion of true negatives divided by the sum of true negatives plus false positives). Of course, none of this bodes well for an evidenced-based approach to risk assessment.

Application of the Family Risk Assessment instrument in Minnesota's DR systems.

Issues related to reliability and validity.

The Family Risk Assessment (FRA) was evaluated as part of Minnesota's larger DR evaluation; beginning in 2000, workers in participating counties were required to use the FRA (Loman & Siegel, 2004a). Evaluation of the FRA found problems with the tool's reliability and validity. Reliability was assessed using Chronbach's alpha to measure the internal consistency among the items comprising the subscale for neglect and the subscale for abuse; additionally, inter-rater reliability was assessed by determining the degree to which workers scored the FRA items, subscales, and global scale in the same way when using written case vignettes. Internal consistency for the neglect subscale (α = .68) and for the abuse subscale (α = .65) was below the generally accepted cut-off point of .70 (Drake & Jonson-Reid, 2008). Inter-rater reliability was poor in that a large amount of variation occurred as workers moved from the process of scoring each subscale item (i.e., neglect subscale and abuse subscale items) to (a) creating a summative score for each subscale, and (b) identifying an overall risk score (category) for the family. Interestingly, *small differences* in the total summed subscale score led to substantial (i.e., inappropriate) categorization differences as workers used ordinal-level categories to classify written case vignettes as "low," "medium," "high," or "intensive" risk (Loman & Siegel, 2004a).

Using data from the Minnesota Social Services Information System, Loman and Siegel (2004b) also examined the FRA's criterion-related validity through its ability to correctly predict which families (N = 15,100) were re-reported for maltreatment during a period of 24 months following their initial report for maltreatment in January of 2001 through September of 2002. In total, the FRA misclassified about one in three families. Similar problems with reliability and validity were noted in the 2004 evaluation of Missouri's DR project (Loman & Siegel, 2004b).

Lack of integration with treatment planning.

In addition to problems with the FRA's reliability and validity, the 2004 evaluation of Minnesota's DR project and the performance of the FRA in the DR systems did not specify how workers used the scores to determine the intensity of service delivery, the targets for treatment, and/or the specific components of service plans. Serious implications for service effectiveness surround the lack of connection among the assessment of the risk of future maltreatment, the needs driving the risk of future maltreatment, and the selection of responsive services that should reduce the likelihood of future maltreatment. Findings from the 2004 evaluation of Minnesota's DR project and the performance of the FRA in Minnesota's DR systems provided further evidence of a lack of fit between assessment and service delivery with implications for children's safety (Loman & Siegel, 2004a). For example, in order to understand the importance the FRA played in workers' daily practices, a sample of DR workers who had completed an FRA (N = 236) for assigned families (N = 412) in the final quarter of 2003 responded to a survey with questions regarding the workers' case-specific practices. Responding to a question about the extent to which the FRA affected whether and how the agency responded to the family, 64.2% of the workers reported either "not at all" or "a minor factor." In contrast, only 33.3% of the workers responding to the survey reported the FRA as a "a major factor" and just 2.4% reported it as the "most important factor." Interestingly, variation in the worker's use of the FRA in conjunction with daily practice decisions was related to the region in which the agency was located (i.e., size of the county and level of urban development) and the point at which the worker chose to implement the FRA (Loman & Siegel, 2004a).

Finally, the lack of a clear fit among how workers used the FRA to assess the overall risk of future maltreatment, the needs that drive the risk of future maltreatment, and the responsivity of delivered services can be seen by comparing the proportion of families with maltreatment re-reports cross-classified by FRA risk level and track assignment. Examination of Table 1.1 (Loman & Siegel, 2004a, p. 64) below provides evidence that services delivered to families in the investigation and assessment tracks did not substantively reduce the risk of future maltreatment, regardless of the FRA-determined risk level. In other words, regardless of families' initial FRA-determined risk of future maltreatment and the prevention services that were delivered to families in both tracks, there were still substantive proportions of families in both tracks that were re-reported for abuse/neglect. Rates of maltreatment re-reports were calculated by following families from their last day of contact with CPS for the family's initial maltreatment report through a period of time that ranged from six months to two years.

In the investigation track, re-reporting was higher for families that received post-investigation services as compared with families that did not receive post-investigation services. This occurred regardless of risk level. In the assessment track, re-reporting was lower for families that received post-assessment services as compared with families that did not receive post-assessment services; this also occurred regardless of risk level.

Table 1.1

Comparing Percentages of Re-Reports for Families Classified by Risk Level and Track Assignment

	Families with a Low or Moderate FRA Risk Score	
	Investigation Track	**Assessment Track**
Subsequent Report, No Services	29.9%	26.7%
Total Number of Families	947	1618
Subsequent Report, Received Services	32.3%	25.5%
Total Number of Families	59	815

	Families with a High or Intensive FRA Risk Score	
	Investigation Track	**Assessment Track**
Subsequent Report, No Services	28.6%	34.5%
Total Number of Families	161	206
Subsequent Report, Received Services	34.8%	29.9%
Total Number of Families	138	221

Note. Adapted from "An Evaluation of the Minnesota SDM Family Risk Assessment," by L. A. Loman and G. L. Siegel, 2004a, St. Louis, MO: Minnesota Department of Human Services, p. 64.

However, the difference in the proportion of families with a subsequent report of maltreatment as compared between those who did and did not receive post-assessment services was quite small. One would have expected to observe a greater difference in re-reports of maltreatment given that post-assessment services were designed to improve family functioning and therefore reduce the likelihood of future abuse/neglect. Furthermore, the evaluators performed a Cox regression analysis to predict the likelihood that a child would be re-reported for maltreatment following the initial report while controlling for FRA risk level (low/moderate vs. high/intensive), services (yes vs. no), track assignment (investigation vs. assessment), and an interaction term compromised of services multiplied by track assignment. Risk level, services, and the interaction term were not statistically significant (given the large sample size, an alpha of .05 is reasonable); only track assignment was statistically significant (RR = 1.427, $p < .05$). Clearly, there is a difference in the orientation of service delivery between the investigation and assessment tracks. However, the effect of services (measured dichotomously as yes/no) is questionable.

The delivery of prevention services to families in both tracks of DR systems has been compromised by (a) an approach to risk assessment that has poor predictive accuracy, and (b) an approach to service planning and provision that is not well integrated (theoretically and/or empirically) with risk assessment. Furthermore, due to the inherent flexibility that drives DR systems, there is no standardized protocol that has been consistently applied for the purpose of assessing risk of future maltreatment and then matching children's/families' needs with the best available services. As noted in a summary of the literature on DR by the National Quality Improvement Center on Differential Response in Child Protective Services (2009), there is a large gap in the literature regarding the practices that workers use to match services to family needs as well as the effects of prevention services versus the approach to service delivery on maltreatment re-reports. Specifically, studies have yet to address whether or not reductions in re-reports of maltreatment are influenced more heavily by the prevention services themselves or by the substitution of positive family engagement in the assessment track for a forensic investigation in the investigation track. Given the high level of variation in the policies and practices employed in DR systems, particularly in relationship to the protocols used to assess risk and need before assigning services, it would be nearly impossible to isolate the effects of services from the effects positive family engagement without clarity on risk assessment protocols. Furthermore, a rigorous comparison of the effects of prevention services with the effects of positive family engagement would be greatly enhanced if the delivery of prevention services is informed by an accurate and fully integrated risk assessment and treatment matching protocol.

In an effort to address these important gaps in the knowledge base regarding the delivery of prevention services in DR, this dissertation study used the Risk-Need-Responsivity (RNR) model as a theoretical basis for (a) improving the accuracy in risk assessment and (b) identifying dynamic risk factors that can be integrated from the assessment of risk into treatment planning. The RNR model is a criminological theory of rehabilitative service delivery to offenders in the criminal justice system (Andrews & Bonta, 2006). Summarized as the "principles of effective intervention" (Cullen, 2005, p. 16), RNR is an evidence-based system of assessment and intervention that has been shown to statistically reduce rates of recidivism by addressing factors that drive the likelihood of future criminal and antisocial activity (Andrews & Bonta, 1998, 2006; Andrews, Bonta, & Hoge, 1990; Andrews, Bonta, & Wormith, 2006; Andrews & Dowden, 2007). More specifically, the assessment/intervention protocol (as discussed below) emphasizes the importance of (a) assessing risk in relation to both static and dynamic risk factors, and (b) providing services that are responsive to identified dynamic risk factors that drive legally-liable behavior. Thus, this dissertation study represents a first step in improving the delivery of prevention-oriented services in DR systems by combining the RNR model with innovative statistical methods to inform the critical practice of matching children/families' needs to prevention services.

Risk, Need, and Responsivity: A Theory of the Effective Principles of Intervention

Although risk assessment and treatment protocols tend to be atheoretical in child welfare systems, this is not to say that there is a complete dearth of theories that might have utility in these arenas. For example, the Risk-Need-Responsivity (RNR) model as described below represents one theory with promise for informing both risk assessment and treatment matching protocols in DR systems. As explained by noted criminologist Francis Cullen (2005), the Risk-Need-Responsivity (RNR) model is a theory of correctional rehabilitation applied through “principles of effective intervention” (p. 16). The best available science suggests that effective treatment programs for offenders should operate with three evidence-based principles concerning (1) offender risk, (2) offender need, and (3) treatment responsivity (Andrews & Bonta, 1998; Andrews, Bonta, & Hoge, 1990). These principles (as defined below) are deliberately linked to theoretically and empirically-informed service delivery; no component operates in isolation from the others. Thus, targets for treatment are identified from the outset and based upon a valid and reliable identification of the factors that are associated with criminal recidivism; these targets for treatment are then matched to a specific level and type of service delivery (Ferguson, 2002). The literature indicates that programs adhering to RNR principles have significantly reduced recidivism rates by 25% to 60% (Gendreau, 1996). Programs not adhering to RNR principles have not reduced recidivism rates, regardless of the available treatment components (Andrews & Dowden, 2007).

As noted below in Figure 1.1’s RNR-informed treatment protocol, the *risk principle* (please see the box at the top of Figure 1.1 labeled “Risk Assessment”) requires that all treatment interventions begin with risk assessments that identify the offender’s risk of (i.e., propensity for) future criminal activity (Andrews & Dowden, 2007). Risk of future criminal activity is comprised of both the individual factors that have been theoretically and empirically associated with criminal recidivism in addition to the cumulative effect of the total number of identified factors (Andrews & Bonta, 2006). Programs should use

RISK ASSESSMENT

Assessed with an actuarial instrument that includes both past-oriented or static factors and future-oriented or dynamic factors.

Provides overall score indicating likelihood of future maltreatment and tracks the characteristics and conditions that drive the overall likelihood of future maltreatment.

More intensive service delivery (restrictiveness of service structure, type of service, dosage, and duration of service) is planned for children and families assessed as having higher levels of risk of future maltreatment.

NEEDS ASSESSMENT

A need is comprised of a subset of dynamic risk factors that form treatment targets.

Likelihood of future maltreatment decreases as the dynamic risk factors are addressed.

RESPONSIVITY

Interventions that have been shown to effectively alter targeted dynamic risk factors (i.e., needs) will be applied for the purpose of reducing the likelihood of future maltreatment.

Responsive services calibrate the restrictiveness of service structure, type of service, dosage, and duration of service in response to (a) the overall likelihood of future maltreatment, and (b) the dynamic characteristics and conditions that drive the probability of future maltreatment.

Figure 1.1. Risk-Need-Responsivity model as it applies to child welfare

valid and reliable risk assessment tools that measure an offender's risk level to include both *static* predictors of recidivism that do not change or change in only one direction (e.g., age, gender, past criminal history, etc.) and *dynamic* predictors of recidivism that are changeable such as substance abuse (Gottfredson & Moriarty, 2006).

The risk principle links the offender's propensity for future criminal activity with the proper intensity of service delivery (levels related to service type, structure, dosage, duration, and the number of services) (Andrews & Dowden, 2007). Hence, high-risk offenders (i.e., those with a high propensity for future criminal behavior) should receive more intensive and extensive services as compared with low-risk offenders (Andrews & Bonta, 2006; Thanner & Taxman, 2003). In sum, failure to properly match level of risk with the proper level of service intensity can aggravate recidivism rates.

As noted in Figure 1.1, RNR-informed programs should always follow risk assessments with need assessments. By definition, "criminogenic needs are a subset of an offender's risk level. They are dynamic risk factors that, when changed, are associated with changes in the probability of recidivism" (Andrews & Bonta, 2006, p. 281). If the risk principle identifies the overall likelihood of criminal recidivism for the purpose of linking the offender to the appropriate level of service intensity, then the *need principle* (please see the box in the middle of Figure 1.1 labeled "Needs Assessment") unpacks the overall level of risk by identifying the particular factors that can be altered through targeted intervention for the purpose of reducing the likelihood of criminal recidivism. The risk principle is used to alert the criminal justice practitioner to the fact that an offender is more or less likely to recidivate. The need principle bridges the gap between the task of identifying the extent to which recidivism is likely occur and the task of identifying the characteristics and conditions that drive the likelihood of recidivism

(Andrews, Bonta, & Wormith, 2006; Ferguson, 2002). Hence, *criminogenic needs* form the bull's eye towards which services are sighted. Research suggests that eight criminogenic needs (factors) are associated with recidivism and, when treated, with reductions in recidivism (Andrews, 2006; Andrews & Bonta, 2006). These eight factors include the following: (a) a history of antisocial behavior, (b) antisocial personality patterns, (c) antisocial cognition, (d) antisocial associates, (e) family and marital problems, (f) problems with poor performance and satisfaction in school and/or work, (g) low levels of involvement and satisfaction with anti-criminal leisure activities, and (h) alcohol and drug abuse (Andrews & Dowden, 2007).

Once level of risk and criminogenic needs are identified, services are linked to address the magnitude and the domains of risk through level of service intensity and type of service. The *responsivity principle* (please see the box at the bottom of Figure 1.1 labeled "Responsivity") requires that services have a base of evidence demonstrating the general effectiveness of the interventions in reducing the likelihood of recidivism and that services be tailored to meet the specific needs of the offender. Hence, the third RNR principle concerns two types of *service responsivity*: *general* and *specific* responsivity (Andrews & Dowden, 2007). For the purpose of simplifying Figure 1.1, the two types of responsivity are represented by the box at the bottom of Figure 1.1 labeled "Responsivity." Effective treatment programs should be *generally responsive*, which is to say that programs should match offenders' criminogenic needs with *evidence-based programs and interventions* (Ferguson, 2002). This general match presumes the use of statistically proven modes of treatment (behavioral, social learning and cognitive-behavioral strategies) that address criminogenic needs as evidenced by substantive decreases in rates of recidivism (Andrews, Bonta, et al., 2006; Andrews & Dowden, 2007). In contrast, *specific responsivity* refines generally-responsive services by tailoring generally-responsive treatment modes and strategies for interventions to fit with offenders' demographics, learning styles, motivations, personalities, and strengths (Andrews & Dowden, 2007).

Using the Principles of RNR to Improve the Delivery of Prevention Services in DR

The principles of RNR, as discussed previously, can be conjoined to form a framework in which effective service delivery in DR systems follows a logical argument from beginning to end. RNR answers the question, *"What works?"* by identifying and connecting the *who*, *what*, and *why* of service delivery. Families contain caretakers *who* are at risk of perpetrating future acts of maltreatment; these individuals need preventive services that can lower their risk of perpetrating future acts of abuse and neglect. Additionally, families contain children *who* are at risk of being victimized by future acts of maltreatment; these individuals need protective services to reduce the likelihood that they will be victimized by maltreatment in the future. Within the RNR framework, *who* is defined by the assessment of the static risk factors (i.e., factors that cannot be changed through intervention) that separate families with a high likelihood of repeat maltreatment from families with a moderate to low likelihood of repeat maltreatment. *Who* is important because the accurate classification of families into high versus moderate to low risk groups assists workers in planning for the appropriate level of service delivery intensity (i.e., service type, service dosage, service duration, and the degree of CPS oversight).

What refers to the drivers of the overall likelihood of repeat maltreatment; drivers are the characteristics and conditions that cause the overall risk of repeat maltreatment to increase or decrease. Within the RNR framework, *what* is defined by the assessment of dynamic risk factors (i.e., factors that can be changed through intervention) that form the targets for treatment. *What* is important because the accurate identification of the specific characteristics and conditions that drive the risk of repeat maltreatment assist workers in selecting the interventions that are capable of addressing each family's specific combination of dynamic risk factors.

Finally, *why* refers to why there should be links among who, what, and service delivery for the purpose of reducing maltreatment recidivism. Within an RNR framework, *why* is the element that holds the components of service delivery together. All of the policies, practices, and tools that comprise the assessment and treatment protocol should be theoretically and empirically connected to serve a common goal: the reduction of repeat maltreatment by simultaneously addressing the overall likelihood of future maltreatment and the specific combination of factors that drive the likelihood of future maltreatment.

Figure 1.2 (see below) proposes a way to improve the fit between DR practices in the investigation and assessment tracks and the RNR principles of effective intervention. More specifically, this figure integrates the delivery of prevention- oriented services for cases in both tracks by using a standardized risk assessment and service planning protocol. The integration of protection and prevention can be followed from (a) risk assessment (please see the box at the top of Figure 1.2 labeled "Risk Assessment"), to (b) needs assessment (please see the box in the middle of Figure 1.2 labeled "Needs Assessment"), and (c) to two types of responsivity (please see the four boxes at the bottom of Figure 1.2). Responsivity is simultaneously determined for each child's protective needs and the family's support needs. Addressing the child's protective needs improves his/her immediate safety while addressing the family's support needs improves

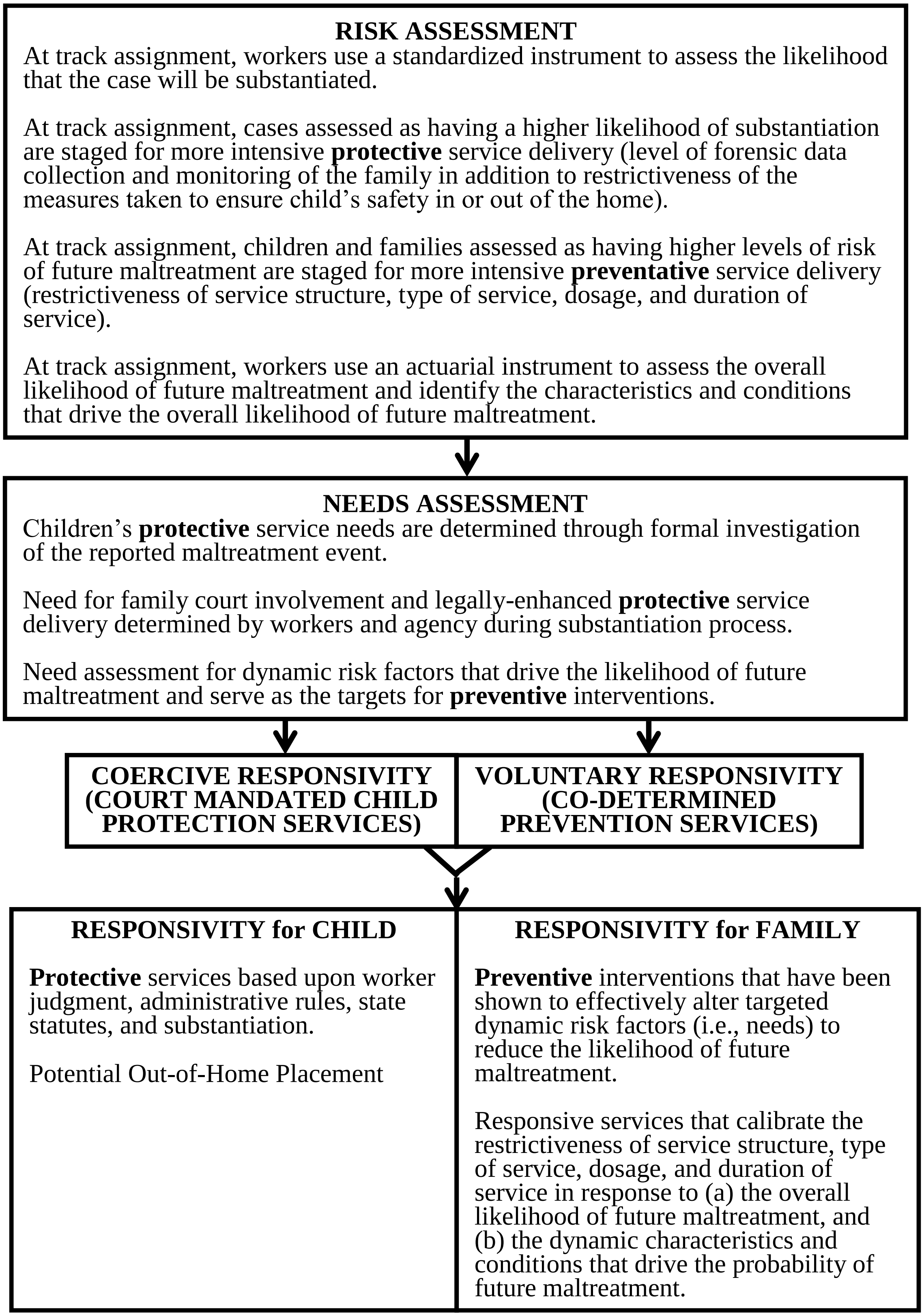


Figure 1.2. Integrated investigation/assessment Risk-Need-Responsivity model

family functioning and child well-being, which reduce the likelihood of future maltreatment. To this end, the next section will outline a study to advance the state of the art in relation to risk assessment (based upon static and dynamic risk factors) from within an RNR perspective.

Chapter 2: Research Objectives

Introduction

This chapter highlights several key methodological and statistical limitations that have driven research on risk factors of repeat maltreatment to date. For example, limitations include the assumption that repeat maltreatment is a linear function of predictors that are predominantly static (Gambrill & Shlonsky, 2000). This assumption has curtailed opportunities for theory building and the integration of risk assessment with treatment planning. Following the analysis of key limitations, this chapter presents the aims for this dissertation study.

Unresolved Methodological and Statistical Issues

Gambrill (2008a) described the challenges related to decision-making in child welfare and highlighted two critical elements that can improve accuracy in decision-making: theory and pattern recognition. An examination of 19 key studies that have identified a variety of statistically significant predictors of repeat maltreatment revealed several things (please see Table 2.3 beginning on page 57). *First*, the majority of significant predictors are static as opposed to dynamic. Thus, most of the significant predictors cannot be used to inform the development and testing of theoretically-derived preventive services because static risk factors are not responsive to treatment. Only dynamic risk factors can be influenced by services. While static factors such as a prior report of maltreatment may explain the largest amount of variance in the likelihood of repeat maltreatment, the inclusion of such static factors in an actuarial risk assessment tool does nothing more than provide a global assessment of the likelihood of subsequent maltreatment (Austin, 2006; Gambrill & Shlonsky, 2001; Gottfredson & Moriarty, 2006; Shlonsky & Wagner, 2005; Taxman, 2006). The assessment of a global risk score is only the first step in service planning and provision; overall risk level informs the intensity of treatment, but it cannot help a worker identify targets for treatment and monitor the factors that account for the largest amount of change in a family's overall risk of repeat maltreatment (Taxman, 2006). Thus, the identification of dynamic risk factors is essential for improvements in assessment and service planning that lead to responsive service delivery.

Second, an overwhelming majority of the statistical analyses in the 19 key studies have assumed that the likelihood of repeat maltreatment is a linear function of its predictors. Hence, none of the 19 key studies included higher-order polynomial terms and very few studies included interaction terms (Bishop, 1995; Fox, 2000; Gujarati, 2003). Those that did include interaction terms limited the interaction terms by including (a) interactions between two predictors wherein each predictor was a first-degree polynomial term, and/or (b) interactions between static as opposed to dynamic risk factors. False assumptions about the functional form of the relationship between repeat maltreatment and its predictors ultimately affect the decision boundary that is created to separate and differentiate between children who will experience subsequent maltreatment and children who will not experience subsequent maltreatment (Bishop, 1995). Furthermore, linear models are not well suited to detecting complex interaction effects among predictors that are being used to explain variance in an outcome that is both noisy (i.e., subject to a great deal of measurement error) and difficult to predict (e.g., is the result of complex human behavior, has predictors measured at multiple levels within a broad ecological model, and has an unknowable true base rate) (Gambrill & Shlonsky, 2000; Marshall & English, 1999, 2000).

Third, the theoretical model that researchers typically use to explain repeat maltreatment is the ecological model (Cash, 2001; DePanfilis & Zuravin, 1999b, 2002; English & Graham, 2000; Jagannathan & Camasso, 1996). This framework is not well-specified, and constructs are not clearly defined except to the extent that variables are assumed to represent a wide range of constructs measured at the individual-, family-, agency-, and community-levels. Additionally, relationships among the variables are not well-specified and typically do not include reference to (a) temporal ordering; (b) distinctions among main effects, joint effects, and mediated effects; and (c) magnitude of effects. Studies of the predictors of maltreatment have not generally contributed to the advancement and refinement of a theory that explains repeat maltreatment. Furthermore, studies that have identified significant predictors of repeat maltreatment have typically focused on identifying the partial effects of various predictors on the likelihood of subsequent maltreatment, but they have not developed patterns of predictors that could be used to classify families. Therefore, rather than focusing on the problematic task of using a model to predict the likelihood of subsequent maltreatment for a particular child (this is not feasible for a variety of reasons to include the fact that the probability of repeat maltreatment for an individual is based upon the combination of the specific values for each predictor included in the model), efforts may be better spent on identifying patterns or combinations of predictors that can assist workers in grouping families into higher or lower categories of risk. Thus, classification as opposed to prediction can help workers in understanding what families with similar values for a particular combination of predictors have a higher or lower likelihood of engaging in subsequent maltreatment (Baird & Wagner, 2000; Gambrill & Shlonsky, 2000; Shlonsky & Wagner, 2005). Identifying combinations of dynamic risk factors may be more useful to workers in the assessment phase as they must use their clinical expertise to integrate an overall risk score with identified targets for treatment while selecting the best possible combination of services to address the dynamic risk factors.

The unresolved methodological and statistical issues described above have led to one of the great challenges in research that focuses on the likelihood of repeat maltreatment: Namely, repeat maltreatment risk assessment studies lack both a theoretical anchor and a "consistency in effects" anchor. Specifically, the ground underneath the collection of studies on the risk of recurrent maltreatment is continually shifting because there is no strongly held theory of repeat maltreatment and there is no consistent set of effects that provide a sound explanation as to what increases or decreases the likelihood of recurrent maltreatment. In general, there is a wide array of predictors that have been included in studies that typically use large administrative data sets to provide information about the potential child, primary caregiver, perpetrator, family, maltreatment incident, service delivery, and community characteristics that influence the risk of recurrent maltreatment. The number of explanatory variables included in the average risk assessment model is typically rather high, and the best studies distinguish themselves by using large samples, long follow-up periods, and an ever-increasing number of predictors in an attempt to find statistically significant parameter estimates that more effectively explain variation in the risk of repeat maltreatment. However, as the number and type of predictors increase, there does not appear to be a concurrent increase in the overall predictive accuracy of risk assessment models and/or increase in the degree to which parameter estimates can be used to improve the delivery of preventive services (Gambrill & Shlonsky, 2000, 2001; Knoke & Trocmé, 2005; Shlonsky & Wagner, 2005). Given the lack of consistency in findings across risk assessment studies to date, coupled with the low predictive accuracy that characterizes risk assessment models in general, this dissertation study focused on asking and answering the question, “What about Door B?”

In essence, “Door B” assumes that the likelihood of recurrent maltreatment is not necessarily a linear function of its predictors. Hence, a combination of random forest and neural network analyses were used to identify the best possible way of relating the risk of repeat maltreatment to a rich selection of predictors that have been tested across the 19 key risk assessment studies in Table 2.3 (Bishop, 1995, 2006; Breiman, 2001; Garson, 1998; Marshall & English, 1999, 2000). The traditionally used statistical techniques (e.g., binary and multinomial logistic regression and Cox regression) assume that the likelihood of a maltreatment re-report is a linear combination of selected predictors: The mathematical operations used to relate recurrent maltreatment to the selected explanatory variables are pre-set and the parameter estimates are adjusted to improve the model’s fit to the data. In contrast, the statistical techniques used in this dissertation study made no assumptions about the mathematical operations that were used to relate recurrent maltreatment to selected predictors. Instead of adjusting parameter estimates to optimize the fit of a pre-selected model (i.e., a mathematical structure that relates Y to X), the analyses used for this dissertation study adjusted the functional form of the estimated relationships between repeat maltreatment and its predictors to best predict which children would be re-reported and which children would not be re-reported (Abdi, Valentin, & Edelman, 1999; Beck, King, & Zeng, 2000, 2004; Cheng & Titterington, 1994; Garson, 1991; Marshall & English, 1999, 2000; Paik, 2000). The neural network model created for this dissertation study was used to identify the combination of risk factors that best differentiate families that will engage in repeat maltreatment from families that will not engage in repeat maltreatment. The ability to estimate a classification scheme that separates returning families from non-returning families as a function of values for key risk factors has substantive implications for improving the delivery of effective preventive services. A generalizable classification scheme can be used to assist workers in identifying families' targets for treatment while preparing for service planning and provision. Moreover, the identification of empirically-supported risk-based groups of children and families, where families are classified as likely or unlikely to be re-reported as a function of specific predictors, could assist in building a theory of repeat maltreatment for the purpose of designing prevention-oriented interventions.

The application of neural network models in child welfare has been briefly explored by Marshall and English (1999, 2000). Specifically, the authors (Marshall & English, 1999) applied neural networks to administrative child welfare data for the purpose of identifying predictors of repeat maltreatment that would otherwise be missed in statistical models that assumed the relationship between repeat maltreatment and its predictors was linear. Additionally, the authors were interested in identifying predictors that could be applied in interaction with other risk factors in binary logistic and Cox regression models. The data set contained a sample of families that had a report accepted for investigation (i.e., referral) to the Washington State CPS system. All families were initially referred to CPS between July 1, 1994 and June 30, 1995. The criterion variable, re-referral, was defined as a subsequent accepted report to CPS that occurred after the summary assessment was issued following the family's initial referral (i.e., an indication that the investigation had already occurred in response to the family's initial referral); the criterion variable was tracked for each family for up to a period of two and one-half years. The predictors were comprised of a range of variables to include 37 individual risk factors from the consensus-based Washington Risk Matrix (WRM) (English, Marshall, Brummel, & Orme, 1999), and additional *child demographic characteristics* (i.e., age, gender, and race/ethnicity), *incident-based characteristics* (i.e., type of abuse and type of reporter), *agency characteristics* (i.e., state administrative region, office size, and population size relative to rural, urban, and metro categories), and *CPS response-based characteristics* (i.e., response time, intensity of investigation, worker's assessment of the overall risk of future abuse/neglect, determination of substantiation, and determination of placement).

Marshall and English (1999) noted that the neural network model "correctly classified the number of referrals (0 to 12 for this data set…) to within ±10% of the actual value for 88% of the cases" (p. 293). However, it was unclear as to how the authors introduced cut points to determine which families were classified by the neural network model as highly likely to be re-referred for maltreatment and which families were classified by the neural network model as unlikely to be re-referred for maltreatment. Furthermore, upon comparing the actual values for the outcome of interest with the predicted values for the outcome generated by the neural network, it was unclear as to how a given predicted value could be within plus or minus 10% of the correct answer. The authors did not report the percentage of cases that were correctly classified within each category. For example, of those families classified as highly likely to be re-referred for maltreatment, the percentage of families that were actually re-referred for maltreatment during the study period is an indication of the model's predictive validity in relationship to sensitivity (i.e., sensitivity is a measure of the model's ability to correctly classify families that are re-referred for maltreatment). Likewise, of those families classified as unlikely to be re-referred for maltreatment, the percentage of families that were not, in fact, re-referred for maltreatment during the study period is an indication of the model's predictive validity in relationship to its specificity (i.e., specificity is a measure of the model's ability to correctly classify families that are not re-referred for maltreatment). Moreover, the authors did not provide any other measures of predictive accuracy for the neural network.

Beyond attempting to increase the predictive accuracy of risk assessment in child welfare, Marshall and English (1999) did not indicate a specific interest in identifying predictive *dynamic* risk factors for the purpose of establishing a link between risk assessment (i.e., the overall likelihood that a family will be re-referred for maltreatment) and treatment planning (i.e., identifying dynamic risk factors that drive the likelihood of future maltreatment and can be influenced through service provision). Overall, the final neural network model included eight risk factors that were used to distinguish families that were re-referred for maltreatment from families that were not re-referred for maltreatment (i.e., caretaker's history of CA/N as a child, caretaker's ability to protect the child, caretaker's victimization of other children, chronicity of the CA/N, hazards in the home, substantiation, abuse type, and length of CPS service). These risk factors were largely static and different from the combination of individual-level WRM risk factors that workers typically used to determine a child/family's overall risk of future maltreatment (Marshall & English, 1999, 2000). Upon including the predictors identified by the neural network as being predictive of repeat maltreatment in a Cox regression model along with predictors identified as being statistically significant in bivariate analyses, the predictive accuracy of Marshall and English's (1999) final risk assessment model explained comparatively little of the variation in the risk of recurrent maltreatment (R^2 = 0.12) (p. 294).

The administrative data set utilized by Marhsall and English (1999) was rich in terms of containing a variety of variables that measured child characteristics, caretaker characteristics, dimensions of the relationship between the caretaker and the child, the caretaker's socioeconomic stress, and dimensions of the maltreatment. However, the data set did not include information regarding (a) child and caretaker mental health as measured by service use, (b) child and caretaker involvement in the juvenile justice and criminal justice systems, (c) dimensions of CPS service delivery to children and families to include the timing and types of services received (other than one broad measure of length of CPS service), and (d) community-level characteristics. As noted by Gambrill and Shlonsky (2001), accurate and responsible risk assessment takes into account the effects that community-based structural conditions have on a child's/family's likelihood of being re-referred for maltreatment in addition to the effects that service delivery has on a child's/family's likelihood of being re-referred for maltreatment.

As noted earlier by Fluke (2008), the ways in which the CPS system and its workers intervene with children and their families is the mechanism through which a reduction in maltreatment recurrence can and should be achieved. Additionally, DR practice and policy emphasize the critical role that community-based service delivery plays in being able to provide a much larger proportion of children and families referred to CPS with access to a greater number of services (Waldfogel, 2000a, 2000b). Inherent in this dependence on a coordinated community response to preventing the recurrence of child abuse/neglect is the need to develop a universal language of risk factors that service providers from a variety of public and private agencies can use to coordinate the assessment, treatment planning, and service provision for CPS-referred children and families (Schene, 2005). If the delivery of prevention-oriented services is going to be effective in reducing recurrent maltreatment, then service providers across a multitude of participating agencies need to know (a) what risk factors should be tracked across the life of a child welfare case by every participating agency, and (b) what agencies are doing in response to the dynamic risk factors most strongly associated with recurrent maltreatment. In short, the ability to accurately identify a set of risk factors across the organizations most likely to coordinate services for CPS-involved children and families is essential for the development of the most effective preventive service delivery plan. Data on community-based characteristics is also important because community context substantively influences family needs and access to family support services (Conley, 2007; Conley & Duerr Berrick, 2008; Gambrill & Shlonsky, 2001; Johnson, Sutton, & Thompson, 2005).

In a separate study, Marshall and English (2000) applied neural networks to identify the combination of individual-level WRM factors that best classified workers' overall risk assessment scores (i.e., risk of future maltreatment). The combination of risk factors that best classified workers' overall risk assessment (i.e., caretaker's parenting skills, caretaker's recognition of the problem, chronicity of the CA/N, stress on the caretaker, and the extent of emotional harm) did not map onto the combination of risk factors that best classified families' re-referral status.

Neural networks have also been applied in two additional studies for the purpose of increasing the predictive accuracy of classification schemes designed to identify (a) juvenile court-involved adolescents who were most and least likely to engage in future acts of delinquency, and (b) the maltreatment reports that were most and least likely to be substantiated. In both cases, the neural network analyses produced models that demonstrated very high predictive validity. In the study by Jones, Schwartz, Schwartz, Obradovic, and Jupin (2006), where juvenile re-arrest was the outcome variable of interest across the entire sample of adolescents (*N* = 8,239), 97% of the juveniles who were predicted as being in the low risk category of recidivism were actually not re-arrested, and 80.6% of the juveniles who were predicted as being in the high risk category of recidivism were actually re-arrested. When applied to a subsample of female juvenile delinquents (*n* = 1,024), 97.5% of the female juveniles in the low risk category were not re-arrested and 81.8% of the female juveniles in the high risk category were re-arrested. When applied to a subsample of male juvenile delinquents (*n* = 7,215), 96.9% of the male juveniles in the low risk category were not re-arrested and 80.5% of the male juveniles in the high risk category were re-arrested. Additionally, when applied to a subsample of White juvenile delinquents (*n* = 1,078), 98.6% of the White juveniles in the low risk category were not re-arrested and 87.7% of the White juveniles in the high risk category were re-arrested. Finally, when applied to a subsample of Non-White juvenile delinquents (*n* = 7,154), 96.7% of the Non-White juveniles in the low risk category were not re-arrested and 79.8% of the Non-White juveniles in the high risk category were re-arrested.

In the study by Schwartz, Kaufman, and Schwartz (2004), substantiation by the Harm Standard of abuse was the outcome variable of interest where cases of maltreatment in the NIS-3 were classified as being at high risk of substantiation according to the Harm Standard or at low risk of substantiation according to the Harm Standard. The total sample of case records (*N* = 1,767) was split into a training set that contained 1,150 cases (65% of the sample) and a test set that contained 617 cases (35% of the sample). After approximating the target function by modeling the relationship between the likelihood of substantiation and its predictors on the training set of case records, the target function's classification accuracy was tested on a new set of case records with the test set. Overall, the model's predictive accuracy was high, where 553 cases in the test set (90%) were accurately predicted as either meeting the Harm Standard or not meeting the Harm Standard. Conversely, 64 cases in the test set (10%) were not accurately predicted; of those cases that were not accurately predicted, there were 4 false positives, 12 false negatives, and 48 cases that were indeterminate (could not be classified). While both studies demonstrated the high level of predictive accuracy that can be achieved with a neural network model, neither study included post-hoc analyses that examined the relative contribution of each predictor in the model to include (a) an assessment of the degree of nonlinearity in the functional form of the outcome variable's relationship with specific predictors, and/or (b) the relative superiority in prediction achieved by the neural network in comparison with a standard linear model (where the same predictors are included in each model).

Loss of Predictive Accuracy As a Result of Failing to Walk Through Door B: Placing False Negatives and False Positives in Context

As noted earlier, the costs of maltreatment are very high for medical and non-medical services at 94 billion dollars a year or 100,000 dollars per maltreated child who receives immediate and long-term care (Foster, Prinz, Sanders, & Shapiro, 2008; Fromm, 2001). Excluding medical care, a recent survey of 50 states estimated that federal, state, and local expenditures in child welfare aggregated to 22.2 billion dollars in state fiscal year 2002 (Scarcella, Bess, Zielewski, Warner, & Geen, 2004). A breakdown of these costs can be seen in Table 2.1 (taken from Scarcella, Bess, Zielewski, Warner, & Geen, 2004, p. 10) directly below. Out-of-home placements and adoptions claim a relatively large (56.6%) portion of child welfare spending, while in-home prevention services (under "Other") claim a relatively small portion (14.0%) of child welfare spending. Furthermore, an out-of-home placement is clearly a more invasive method of protecting a child from harm; in contrast, in-home prevention services are less invasive and can be used to promote child safety and well-being while supporting family functioning.

If funds are to be saved and/or properly allocated to those in need of prevention services, then DR systems in child welfare need to employ a risk assessment protocol that properly classifies (a) those who are likely to be re-reported for maltreatment and (b) those who are not likely to be re-reported for maltreatment. In addition to being accurate, a risk assessment protocol needs to be integrated with service planning and provision. Thus, risk assessment should also (a) assist the worker in identifying the appropriate level of service intensity (i.e., degree of CPS oversight, service type, dosage, and duration) and, (b) assist the worker in identifying appropriate targets for treatment.

Failure to properly classify children and integrate the classification with service planning and provision can have serious financial implications. Loman and Siegel (2004a) evaluated the predictive validity of the Family Risk Assessment (FRA) in 20 counties in Minnesota that were required to utilize the FRA in conjunction with the Alternative Response Project. The evaluators described the purpose of the FRA in assisting workers determine who will receive services:

> The FRA has been promoted as a means of improving the accuracy of CPS in identifying high-risk families so that they can be targeted for further intervention and services, while at the same time steering the agency away from low-risk families. The FRA approach represents a broadening of the traditional CPS approach. AR workers seriously consider provision of services to lower-risk families—even to families in which no maltreatment of children can be substantiated, if they are willing to participate. (Loman & Siegel, 2004a, p. 8)

Table 2.1

State Fiscal Year 2002 Child Welfare Spending by Use ($ in Millions) [a]

	Total	Federal	State	Local
SYF 2002 Expenditures	$22,156	$11,304	$8,206	$2,646
Out-of-Home Placements	9,955	6,082	2,806	1,066
Support services	1,238	791	422	25
Room and board	3,522	2,546	812	164
Administration	2,588	2,254	329	5
Uncategorized out-of-home Placements [b]	2,606	492	1,243	872
Adoptions	2,580	1,419	1,033	129
Administration	1,727	708	1,006	13
Other	3,103	1,802	944	356
Uncategorized Expenditures	4,792	1,293	2,417	1,081

Note. Adapted from "The Cost of Protecting Vulnerable Children IV: How Child Welfare Funding Fared during the Recessions," by C. A. Scarcella, R. Bess, E. H. Zielewski, L. Warner, and R. Geen, 2004, Washington, DC: The Urban Institute, p. 10.
[a] Numbers may not total because of rounding.
[b] The variety of accounting methods states use to track their spending means that some states were not able to categorize all expenditures according to the Urban Institute's uniform categories.

Additionally, the evaluators reported that the FRA manual directs workers to close low-risk cases and to consider closing moderate-risk cases. At the same time, the evaluators noted that as per Minnesota state policy, workers had a greater flexibility in how they are directed to apply service planning and provision decisions in accordance with FRA scores. That said, no further information was provided in regards to how workers applied policy-based, agency-based or even individually-based decision-making guidelines when using FRA scores to inform the delivery of prevention services.

As can be seen upon examining Table 2.2 directly below (taken from Loman & Siegel, 2004a, p. 13), a high proportion of families (22.3%) scored as low risk had at least one re-report for maltreatment (*n* = 1,260 out of 5,809 cases). As stated earlier, false negatives occur when families classified as being unlikely to be re-reported for maltreatment are subsequently re-reported for maltreatment. But what exactly is the cost of a false negative? A false negative increases the likelihood that children and their families will not receive prevention services. While a worker can always decide to deliver services regardless of the FRA score, this practice is not particularly useful in the case of a false negative because there is little information that can assist the worker in linking elements of the risk score to sound treatment planning. After all, if the child/family has been inappropriately classified as low risk, what information can the worker use to appropriately determine the intensity of service delivery and the targets for treatment?

Table 2.2

Risk Assessment by Recurrence of Any Accepted Maltreatment Report during 24 Months[a]

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

Note. Adapted from "An Evaluation of the Minnesota SDM Family Risk Assessment," by L. A. Loman and G. L. Siegel, 2004a, St. Louis, MO: Minnesota Department of Human Services, p. 13.
[a] Chi Square = 392.7, $p<.0001$
Tau-b = .144, $p<.0001$
Somer's d = .116, $p<.0001$

In terms of the costs of a false negative, one has to consider the effects of withholding prevention services and/or the effects of delivering prevention services that were not properly matched to the child/family's overall risk level and dynamic risk factors. Failure to provide responsive prevention services allows the conditions that drive the likelihood of future maltreatment to continue unabated. Over time, the child may experience an increasing level of danger; as the conditions supporting the likelihood of future maltreatment worsen, the child may require a more costly and invasive form of protection such as out-of-home placement, which, as we have seen above, constitutes the majority of child welfare expenditures. The risk of maltreatment could materialize into real events of maltreatment, and the child is likely to suffer physical and psychological trauma as a result of being abused/neglected (Cicchetti & Toth, 2005; English, Graham, Litrownik, Everson, & Bangdiwala, 2005; Éthier, Lemelin, & Lacharité, 2004; Rosenberg & Krugman, 1991). If the maltreatment continues, the consequences of the maltreatment could worsen and the child may experience long-term difficulties with aggression, peer rejection, impaired social and daily living skills (Bolger & Patterson, 2001; English et al., 2005). Finally, the danger to the child may increase to the point where a worker sees the immediate need for protection and removes the child from his/her home. An accurate risk assessment tool that is integrated with a treatment planning protocol could be used to deliver less expensive prevention services designed to address the conditions that were likely to lead to future maltreatment.

Conversely, as can be seen upon examining Table 2.2 (taken from Loman & Siegel, 2004a, p. 13), a high proportion of families (61.8%) scored as intensive risk had no re-reports for maltreatment (*n* = 340 out of 548 cases). As stated earlier, false positives occur when families classified as being highly likely to be re-reported for maltreatment are subsequently not re-reported for maltreatment. But what exactly is the cost of a false positive? A false positive increases the likelihood that children and their families will receive prevention services that are both unnecessary and overly intensive. Therefore, the degree of CPS oversight is likely to be high in addition to service dosage and duration; furthermore, the family could be required to participate in unnecessary interventions that could potentially disrupt the normal daily functioning of the family (Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008). Increased surveillance is likely to accompany increased CPS oversight as well as service dosage and duration; thus, an excessively monitored family would be more likely to be re-reported for maltreatment (Bae, Solomon, & Gelles, 2007; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Hélie & Bouchard, 2010; Marshall & English, 1999). Forced compliance with unnecessary and overly intensive services would be very traumatic for family members and costly to the agency. Valuable resources to include worker time and attention would be diverted away from children/families who need intensive CPS oversight and service delivery. An accurate risk assessment tool that is integrated with a treatment planning protocol could be used to avoid the costly and invasive delivery of unnecessary prevention services.

Specific Aims of the Dissertation Study

The exploratory study conducted for this dissertation was designed for the purpose of improving the accuracy of risk prediction for repeat maltreatment within an RNR perspective. A neural network analysis was used to explore the possibility that risk prediction could be improved by incorporating predictor variables that have been used in previous risk assessment studies within a flexible approach to function approximation. In other words, the exploratory study conducted for this dissertation utilized a risk classification method that makes no assumptions about the ways in which the likelihood of repeat maltreatment is related to variables that are typically included as potential predictors of recurrence. Instead, neural networks "allow the data to speak for itself" and therefore provide the opportunity to discover underlying structure that may have been wholly overlooked. Hence, a neural network analysis allows researchers to capitalize on what they do not currently know because neural network analyses are designed to look for the kinds of nonlinear and interaction effects that are not included in the standard linear models typically used for risk assessment.

Due to the exploratory nature of the study, no pre-specified criterion was used to define what constitutes an acceptable level of predictive accuracy. Clearly, a higher level of predictive accuracy and subsequently a lower level of misclassification error is desirable given the consequences that follow type I (i.e., a false positive) and type II errors (i.e., a false negative). Instead, the neural network in this dissertation study was built with the utmost care wherein effort was spent on (a) pre-processing, and (b) the specification of the network's typology and architecture for the purpose of building a model with the highest achievable level of predictive accuracy (please see Chapter 3 for details on the form and functions of neural networks to include information about pre-processing, typology, and architecture). Moreover, JMP Pro 9 software was used because of its unique ability to provide opportunities for the post-hoc visual analysis of a neural network model. In addition to assessing the neural network's ability to achieve classification accuracy, it was important to assess for evidence of nonlinearity to include evidence of higher order polynomial and interaction terms. JMP Pro 9 provides opportunities for post-hoc visual analysis that are akin to sensitivity tests, where the relative effects of predictors in the model can be visually inspected and assessed for the shape, magnitude, and direction of the slope representing each variable's relationship to the average level of risk for an individual with specified values for each predictor. This kind of post-hoc visual analysis gives the researcher the opportunity to evaluate the relative contributions of each predictor and is especially helpful when comparing the sensitivity of static versus dynamic predictors. Moreover, a careful examination of the relative contributions of each predictor allows for the opportunity to take a first step in creating a meaningful link between the art and science of accurate risk prediction and the art and science of daily practice where child welfare workers in DR systems are responsible for creating, implementing, monitoring, and evaluating effective preventive service delivery plans. Previously conducted and described neural network studies in child welfare and juvenile justice were used as inspiration for the analyses conducted herein.

Before proceeding to Chapter 3 for a full discussion of the methods applied in this dissertation study, Table 2.3 summarizes the effects of a wide array of predictors included in 19 key child maltreatment risk assessment studies. Findings from these studies are discussed in relationship to the findings from the neural network and post-hoc analyses contained within this dissertation study in Chapter 5. The predictors included in the neural network analysis that follows were based on (1) variables commonly included in child maltreatment risk assessment studies, and (2) variables that were often neglected by the extant literature but promising for this analysis.

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**[a]
Bae, Solomon, & Gelles (2007)	Re-referral for maltreatment	Cox non-proportional regression No Interaction Terms No Higher Order Terms	*Type of maltreatment with neglect as the reference group --Sexual abuse (RR = 0.84) --Multiple forms (RR = 1.11) *Age of the child (RR = 0.98) *Race/ethnicity of the child with other as the reference group --Black (RR = 1.83) --Latino (RR = 1.50) --White (2.24) *Family structure with both parents as the reference group --One parent (RR = 1.16) --Non-biological (RR = 1.12) *Number of dependents (RR = 1.21) *Reporter type with non-mandatory as the reference group --Mandatory (RR = 0.92) *Frequency of contacts (RR = 1.02) *Investigation level (RR = 1.18) *Length of intervention (RR= 1.01)

Note. When a study provided a separate model for a re-report and a separate model for a subsequent substantiated report, the model estimating the likelihood of a re-report was chosen.

[a]Dynamic factors were defined as broadly as possible to include those characteristics or conditions that could be altered with an intervention. For example, while a child’s documented disability cannot be ameliorated, the child’s level of functioning and/or the parent’s ability to cope effectively with the demands of providing care for a child with special needs can be altered through a planned intervention. Additionally, service delivery was defined as representing a dynamic characteristic or condition if the service was targeted toward (a) a specific condition such as a child’s need for mental health/substance abuse services, or (b) a specific set of behavioral issues such as the child’s receipt of a juvenile court petition. Service delivery from the child welfare system was not defined as representing a specific set of modifiable and underlying conditions that drive the likelihood of recurrent maltreatment because the services generally target the outcome of interest (i.e., promoting child safety by preventing recurrent maltreatment) as opposed to the dynamic factors driving the outcome of interest.

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Bae, Solomon, & Gelles (2009)	Subsequent substantiated report for maltreatment	Multinomial logistic regression No Interaction Terms No Higher Order Terms	*Multiple vs. Single Recurrence* *Age of child (OR = 0.97) *Family structure with both parents as the reference group --Mother only or living with other (OR = 1.22) --Stepparents (OR = 1.36) *Number of dependents (OR = 1.16) *Service type with court-ordered permanency as the reference group --General CPS services (OR = 1.38) *Multiple vs. No Recurrence* *Type of maltreatment with neglect as the reference group --Sexual abuse (OR = 0.59) --Physical abuse (OR = 0.71) *Age of child (OR = 0.94) *Race/ethnicity of child with Black as reference group --Latino (OR = 0.64) --Other (OR = 0.19) *Family structure with both parents as the reference group --Mother only or living with other (OR = 1.62) --Father (OR = 1.64) --Stepparents (OR = 1.46) *Reporter type with non-mandatory as reference group --Medical (OR = 0.74) --Law enforcement (OR = 0.68) *Service type with court-ordered permanency as the reference group --Custody or foster care (OR = 1.36) --General CPS services (OR = 1.46)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Connell, Bergeron, Katz, Saunders, & Tebes (2007)	Re-referral for maltreatment	Cox proportional regression Interaction terms including (1) maltreatment type by substantiation, and (2) post-investigation services by substantiation No Higher Order Terms	*Age of child at index report with under 1 as the reference group --6-10 yrs (RR = 0.87) --11-15 yrs (RR = 0.73) --16-18 yrs (RR = 0.37) *Child race/ethnicity with Caucasian as the reference group --African-American (RR = 0.80) --Hispanic (RR = 0.83) ***Disability (RR = 1.33)** ***Family sub abuse history (RR = 1.50)** ***Family poverty (RR = 3.26)** *Prior substantiated report (RR = 1.09) *Type of maltreatment with neglect as the reference group --Sexual abuse (RR = 0.82) *Maltreatment type by substantiation -- Physical abuse by substantiation (RR=1.22) *Post-investigation services by substantiation (RR=1.30)
DePanfilis & Zuravin (1999a)	Subsequent substantiated report for maltreatment	Survival analysis No Interaction Terms No Higher Order Terms	*The survival experience among three groups was significantly different: --Families that received no services following substantiation of CA/N report had the lowest rate of recurrence --Families that received services following substantiation of CA/N report --Families that received continuing services following the substantiation of CA/N report with at least one prior confirmed report had the highest rate of recurrence

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
DePanfilis & Zuravin (1999b)	Subsequent substantiated report for maltreatment	Cox proportional regression 1 Interaction Term No Higher Order Terms	***Child vulnerability (RR = 1.37)** ***Family conflict (RR = 1.51)** ***Family stress (RR = 1.22)** ***Social support deficits (RR = 1.45)** *Placement (RR = 1.93) ***Family stress by social support deficits (RR=0.84)**
DePanfilis & Zuravin (2001)	Subsequent substantiated report for maltreatment	Survival analysis No Interaction Terms No Higher Order Terms	*The survival experience between two groups was significantly different: --Families that received no post-investigation services following substantiation had a lower rate of recurrence --Families that received post-investigation services following substantiation had a higher rate of recurrence
DePanfilis & Zuravin (2002)	Subsequent substantiated report for maltreatment	Cox proportional regression No Interaction Terms No Higher Order Terms	*Placement (RR = 1.97) ***Child vulnerability (RR = 1.39)** ***Family conflict (RR = 1.44)** ***Family stress (RR = 1.25)** ***Survival stress (RR = 1.16)** ***Social support deficits (RR = 1.45)** *Service attendance (RR = 0.68)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Drake, Jonson-Reid, & Sapokaite (2006)	Re-report for maltreatment	Cox proportional regression 13 interaction terms for (1) child characteristics by time, (2) foster care by time, and (3) substantiation by services (FCS, FPS, Foster Care). Most were significant at the .05 and .01 levels. No Higher Order Terms	*Child age at index event (RR = 0.97) *Child is a person of color (RR = 0.83) *Type of maltreatment with neglect as the reference group --Physical abuse (RR = 0.85) --Sexual abuse (RR = 0.74) *More than one victim in the index report (RR = 1.22) *Number of children in family (RR = 1.16) ***Caregiver graduated from high school (RR = 0.88)** ***Caregiver with MHSA before index event (RR = 1.58)** ***Permanent AFDC exit before index event (RR = 0.88)** *** Permanent AFDC exit after index event (RR = 0.68)** *Service type with the reference group as no service need indicated or received --FCS only (RR = 0.72) --FPS or FPS and FCS (RR = 1.44) --Foster care (RR = 2.46) --Service need but no services (RR = 1.47)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
English, Marshall, Brummel, & Orme (1999)	Re-referral for maltreatment	Bivariate analysis No Interaction Terms No Higher Order Terms	***Caregiver history of domestic violence** *Chronicity of CA/N *Child's age *Caregiver's history of CA/N as a child ***Caregiver employment status *Caregiver impairments** ***Caregiver substance abuse *Stress on caregiver** ***Parenting skills** *Victimization of others in family ***Social support *Protection of child**
Fluke, Shusterman, Hollinshead, & Yuan (2008)	Two outcome variables: Re-report for maltreatment substantiated re-re-report for maltreatment	Cox proportional regression 4 interaction terms including (1) victimization by post-investigation services and (2) victimization by foster care placement. All interactions were significant at the .05 and .001 levels. No Higher Order Terms	*Source of initial report with social and mental health services as the reference group --Medical (RR = 0.87) --Law enforcement (RR = 0.88) --Non-professional (RR = 1.14) *Child age at the initial report with infants as the reference group --All other age categories through 18 (-) *Child sex with female as the reference group --Male (RR = 0.95) *Child race/ethnicity with White as the reference group --Asian/Pacific Islander (RR = 0.60) --African-American (RR = 0.84) --Hispanic (RR = 0.87) ***Child with disability (RR = 1.47)** ***Caretaker with alcohol abuse (RR = 1.12)** *Post-investigation services provided (RR = 1.35) *Placement (RR = 2.19)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Fluke, Yuan, & Edwards (1999)	Subsequent substantiated report	Survival analysis No Interaction Terms No Higher Order Terms	*The survival experience between the following groups was significantly different: --12 to 17 year olds had a lower rate of recurrence compared with 6-11 year olds, 3-5 year olds, and 0-2 year olds --Asian/Pacific Islander group had the lowest rates of recurrence as compared with White and African-American groups --Cases with neglect as the index event were most likely to recur, followed by cases with physical abuse as the index event; cases with sexual abuse as the index event were least likely to recur --Significant difference in survival distributions among groups that experienced single recurrence, second recurrence, and third recurrence; likelihood of recurrence increased following each subsequent event --Children who received post-investigation services were at higher risk of recurrence as compared to children who did not receive post-investigation services
Fryer & Miyoshi (1994)	Subsequent confirmed case of maltreatment	Survival analysis No Interaction Terms No Higher Order Terms	*The survival experience between the following groups was significantly different: --Younger children were at higher risk of recurrence as compared to older children --Children by maltreatment type

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Fuller, Wells, & Cotton (2001)	Subsequent indicated report	Binary logistic regression No Interaction Terms No Higher Order Terms	*Investigation Sample* *Age of the youngest child with 6-18 yrs old as the reference group --0-2 years old (OR = 3.03) *Type of maltreatment with sexual abuse as the reference group --Physical abuse (OR = 5.39) --Neglect (OR = 5.04) *Case disposition with no services needed as the reference group --Referral to community agency (OR = 4.63) --Assessment/family maintenance services (OR = 1.68) *Household structure with all other arrangements as the reference group --Single parent (OR = 2.00) ***Number of child problems (OR = 1.84)** ***Number of caretaker problems (OR = 1.31)** *Number of previous reports (OR = 1.33) *Intact Family Sample* *No CERAP completed (OR = 4.09) *Prior reports (OR = 2.56) ***Number of caretaker problems (OR = 1.33)** *No services during first 60 days (OR = 1.99)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Jonson-Reid (2002)	Re-report for maltreatment	Binary logistic regression 10 interaction terms including (1) race by investigation status, (2) race by age groups, and (3) maltreatment type by investigation status. No Higher Order Terms	*Child's age at first report with 1-6 years of age as the reference category --11-14 yrs (OR = 0.88) --15-16 yrs (OR = 0.53) *Report reason with neglect as the reference group --Sexual abuse (OR = 0.61) *Service level with opened for services as the reference group --Investigated, not served (OR = 1.22) *Significant Interactions (reference groups = Caucasian, age 1-6 at first report, and neglect for interactions below, respectively) --Hispanic by investigated only (OR=0.70) --Hispanic by ages 14-16 (OR=1.32) --Physical abuse by investigated only (OR=0.80)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Jonson-Reid, Emery, Drake, & Stahlschmidt (2010)	Subsequent report of maltreatment for progressive stages	Cox regression No Interaction Terms No Higher Order Terms	Findings below for significant predictors of 1st to 2nd reports *Child characteristics --Age at start of stage (HR=0.97) **--Medical risk at infancy (HR=1.19)** *Parent characteristics --Age at child's birth (HR=1.01) **--Less than HS education (HR=1.29)** --Hx foster care (HR=1.13) **--Never AFDC (HR=0.48)** **--Tract income (HR=0.99)** *Service prior to ever having report for maltreatment (reference group = no service use) --**AFDC (HR=0.87)** **--Parental mental health treatment (HR=1.43)** *Characteristics of 1st report in each stage (reference=neglect/other) --Physical abuse (HR=0.85) --Sexual abuse (HR=0.69) -- Substantiated (HR=1.41) *Child welfare services during current stage (reference group = no service receipt) --FCS (HR=0.50) --FCS and IIS (HR=0.74) --Foster care (HR=0.82) *Other services during stage (reference group=no service use) --Injury (HR=1.14) **--Mental health (HR=1.81)** **--Special education (HR=0.43) --AFDC (HR=1.12)** **--Parent mental health treatment (HR=0.65)** **--Child mental health treatment (HR=0.65)**

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Lipien & Forthofer (2004)	Subsequent substantiated report	Binary logistic regression No Interaction Terms No Higher Order Terms	*Child race/ethnicity with White as the reference group --Nonwhite (OR = 0.88) *Child's age with 0-3 as the reference group --4-7 yrs (OR = 0.85) --8-11 yrs (OR = 0.79) --12-15 yrs (OR = 0.77) *Maltreatment type at index event with neglect as the reference group --Physical abuse (OR = 0.74) --Sexual abuse (OR = 0.69) *Service disposition with no services as the reference group --Short-term services (OR = 1.22) --In-home services (OR = 1.70) --Relative foster care (OR = 0.81)
Marshall & English (1999)	Re-referral for maltreatment	Cox regression No Interaction Terms No Higher Order Terms	*Maltreatment type with sexual abuse as the reference group --Physical abuse (RR = 1.32) --Physical neglect (RR = 1.52) *Administrative region with other region as the reference group --Region 2 (RR = 1.37) --Region 4 (RR = 0.80) *Determination with substantiated as the reference group --Inconclusive (RR = 1.37) --Unfounded (RR = 1.27) ***Child developmental disability (RR = 1.08)** *Chronicity of CA/N (RR = 1.16) *Victimization of others in the family (RR = 1.05) *Caregiver history of CA/N as a child (RR = 1.04) *Child age (RR = 1.06) *Length of service (RR = 1.002)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Ortiz, Shusterman, & Fluke (2008)	Re-report for maltreatment	Cox proportional regression 2 interaction terms No Higher Order Terms	*Child's sex with male as the reference group --Female (RR = 1.04) *Child's age with 0-3 years as the reference group --All age categories from 4-16 yrs or older (-) *Child's race/ethnicity with White as Caucasian/white as the reference group --Asian/Pacific Islander (RR = 0.64) --African-American/black (RR = 0.84) --Hispanic (RR = 0.94) *Prior victim (RR = 1.50) ***Child disability (RR = 1.26)** *Report Source with social services/mental health as the reference group --Medical Personnel (RR = 0.92) --Law enforcement or legal personnel (RR = 0.88) --Educational personnel (RR = 1.24) --Non-professional (RR = 1.15) *Disposition with non-victim as the reference group --Victim (RR = 0.84) *Type of response with investigation as the reference group --Assessment (RR = 0.92) *Received post-investigation services (RR = 1.59) *Received foster care services (RR = 0.93) *Interaction Terms (reference groups = not alternative response and victim, respectively) --Alternative response* foster care (RR=1.50) --Non-victim * foster care (RR=2.02)

Table 2.3

A Summary of Statistical Methods and Static and Dynamic Predictors of Repeat Maltreatment (Continued)

Author (year)	**Outcome Variable**	**Statistical Analysis**	**Significant Predictors with Dynamic Factors in Bold Font**
Way, Chung, Jonson-Reid, & Drake (2001)	Re-report for maltreatment	Cox proportional regression No Interaction Terms No Higher Order Terms	*Sexual Abuse* *Perpetrator's initial report was substantiated (RR = 0.67) ***Neighborhood mean income (RR = 0.97)** *Perpetrator gender with male as the reference group --Female (RR = 1.32) *Physical Abuse* ***Neighborhood mean income (RR = 0.99)** *Perpetrator ethnicity with other as the reference group --Caucasian (RR = 0.91) *Perpetrator gender with male as the reference group --Female (RR = 1.28) *Neglect* *Perpetrator's initial report was substantiated (RR = 1.27) ***Neighborhood mean income (RR = 0.99)** *Perpetrator ethnicity with other as the reference group --Caucasian (RR = 0.94) *Perpetrator gender with male as the reference group --Female (RR = 1.35)

Chapter 3: Method

Improving Risk Assessment by Exploring Underlying Structure

As noted in previous chapters, research on risk assessment for recurrent maltreatment has identified a rather wide array of child-, parent-, perpetrator-, family-, maltreatment event-, and service-level variables that have been used to predict the likelihood of being re-reported for maltreatment (see e.g., Bae, Solomon, & Gelles, 2007, 2009; DePanfilis & Zuravin, 2001, 2002; Drake, Jonson-Reid, & Sapokaite, 2006; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). However, the findings across studies have not been consistent; moreover, the low predictive accuracy of risk assessment tools (Gambrill & Shlonsky, 2000, 2001; Knoke & Trocmé, 2005; Schwartz, Jones, Schwartz, & Obradovic, 2008; Shlonsky & Wagner, 2005) combined with the lack of theory guiding the development of risk assessment measurement (English, Marshall, Brummel, & Orme, 1999; Jagannathan & Camasso, 1996) have created a bit of a standoff. On the one hand, it could be inferred that higher levels of predictive accuracy are not possible given the substantive level of noise in child welfare data and the inherent difficulties in predicting an outcome that is in and of itself rife with measurement problems not to mention the additional complexities of trying to predict future-oriented human behavior among a diverse population of children and families. Underlying this perspective is the idea that researchers have come as far as they can based on what is known. On the other hand, it could be that the right data set and/or the right predictors have not yet been found. Underlying this alternate perspective is the idea that researchers are facing an impasse because of the large amount of information that is not known. An exploratory neural network analysis falls in between the two perspectives by applying in practice the following edict from Beck, King, and Zeng

(2004): “When we know something, we assume it; when we don’t know, we estimate it” (p. 381).

Given the substantive body of literature regarding recurrent maltreatment and the approaches to assessing for the risk of recurrent maltreatment, researchers do know which variables should be included in a risk assessment analysis. Moreover, based on thorough reviews of the state of risk assessment in child welfare, researchers do know that past and current approaches to the study of risk assessment have typically failed to include nonlinear forms of predictor variables such as higher order polynomial and interaction terms (see e.g., Gambrill & Shlonsky, 2000; Marshall & English, 1999, 2000). Finally, previous studies designed to improve the accuracy of risk prediction for child welfare and juvenile justice populations have used neural networks to identify which reported cases of maltreatment were likely to be substantiated (Schwartz, Kaufman, & Schwartz, 2004), which children were likely to be re-reported for maltreatment (Marshall & English, 1999), and which adolescents in the juvenile justice system were likely to recidivate (Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008) with levels of accuracy ranging from 80.6% to 97.5%. That said, none of the previous studies conducted neural network analyses within an RNR framework that provided an explicit strategy for connecting the results of the neural network analyses to the delivery of preventive services. Moreover, not one of the previous studies included post-hoc analyses of the ways in which the neural network results could be used to differentiate children’s and families’ treatment needs.

Based on what is and is not known about risk assessment for repeat maltreatment, an exploratory neural network analysis provides the best option for improving risk prediction both in terms of maximizing predictive accuracy and maximizing the utility of risk prediction in informing the delivery of preventive services. As described below, neural networks are a general class of function approximators that provide a highly flexible environment in which to improve risk prediction. Ultimately, neural networks learn how values for the predictor variables (i.e., the inputs) map onto values for the response variable (i.e., the outputs or target values) for the purpose of developing an algorithm that can predict target values for new cases (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001). A neural network's ability to learn the mapping process is partially based on what is known and therefore on what is included in the model. However, the beauty of applying neural networks to this area of research also lies in what is not known. Because neural networks are flexible function approximators, a network's ability to learn the underlying mapping process can increase the predictive accuracy of risk assessment precisely because of what is not known and therefore what is not assumed *a priori* (Beck, King, & Zeng, 2000, 2004). The sections below provide details as to why a lack of prior knowledge about the functional form of the relationship between repeat maltreatment and its predictors coincides beautifully with the strengths of a neural network analysis. Figure 3.1 provides a visual representation of how a neural network analysis not only takes advantage of what is not known but estimates it in contrast to applying a regression model that assumes the relationship between the probability of maltreatment and its predictors is linear.

Neural Networks as Function Approximators

Neural networks are a general class of flexible and even universal approximators that estimate the target function responsible for generating values for an outcome variable in relationship to values for given independent variables. Neural networks are referred to as a general class of approximators because each neural network's topology (i.e., the

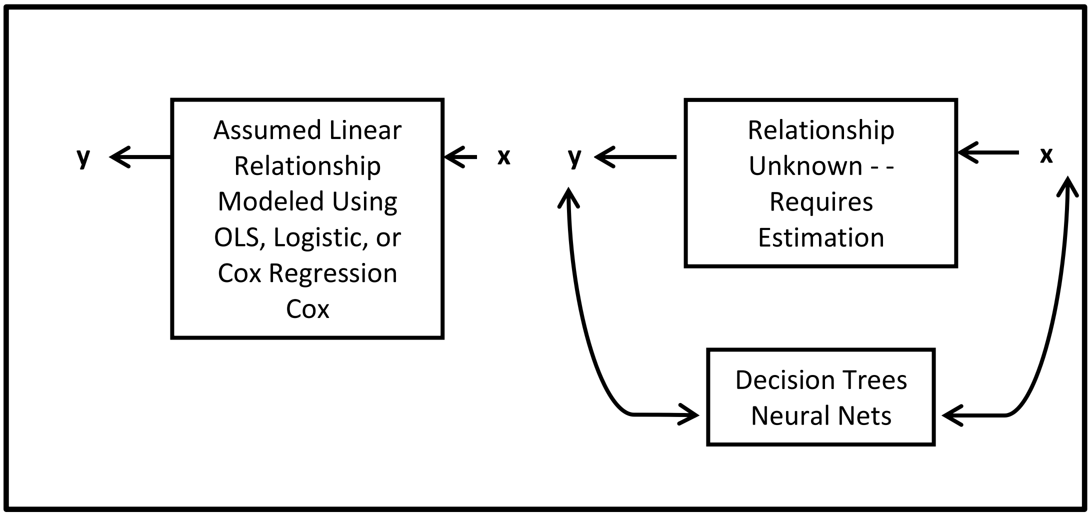


Figure 3.1. In contrast to a linear regression model with a specific functional form, a neural network needs to model the relationship between x and y by learning the function that maps x onto y. Adapted from "Statistical Modeling: Two Cultures," by L. Breiman, 2001, *Statistical Science, 3,* p. 199.

number of layers containing various types of nodes, the number of nodes within each layer, and the connections among the nodes) and architecture (i.e., the system of computational processes) can be specified in a number of ways for the purpose of modeling the underlying functions that generate any given data set (Garson, 1998). In fact, neural networks are so flexible that with a correctly specified topology and architecture, a given neural network is capable of approximating any smooth target function that maps values for the independent variables (i.e., input values) onto values for the outcome variable (i.e., target values) (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001).

Among the number of technical terms just provided to generally describe neural networks and their utility, the most important phrase is "target function." A target function can be defined as an unknown set of mathematical processes that generate values for the outcome variable given values for the independent variables within a particular data set (Garson, 1998; Smith, 1993). For example, Breiman (2001b) defines the estimation of a target function through the application of neural networks as the process of identifying a function $f(x)$ that operates on a vector of values x to generate a vector of values for the outcome y. To more clearly specify what is meant by a function $f(x)$, Gill (2006) describes $f(x)$ as constituting a type of mapping process where each respective value of x is transformed into a unique and new value, and the particular steps taken to modify x are encoded in $f()$. That said, a target function can also be defined through the modeling of an underlying conditional probability mass function that describes the probability or relative likelihood of observing specific values (y) for a discrete random variable (Y) given values for predictor variables (x) and unknown parameters (θ) (Bishop, 1995; Gill, 2006).

Probability and likelihood: Providing a foundation for function approximation.

Defining a target function in terms of modeling its conditional probability mass function provides a starting point from which to estimate the unknown set of mathematical processes (i.e., the algorithm) that generate values for the outcome variable given values for the independent variables and unknown parameters. Essentially, neural networks allow the researcher to learn how best to model the functional form of the relationship between the response variable and its predictors by working backwards from the observed values of the random response variable and the fixed values of the predictors. In theory, the target functions estimated by neural networks are wholly data driven because the estimation process is not constrained by *a priori* assumptions regarding (a) the mathematical operations that associate x with y, and/or (b) the conditional probability mass function that describes the long-run relative likelihood of observing specific values (*y*) for a discrete random response variable (*Y*) (Bishop, 1995; Breiman, 2001b; Hastie, Tibshirani, & Friedman, 2001). Freedom from these constraints is the primary reason neural network analyses were selected for the exploratory analyses that follow. To date, scholars interested in improving the accuracy of recurrent maltreatment risk assessment have consistently used linear models to relate values of the predictor variables to the likelihood of recurrent maltreatment (see e.g., Bae, Solomon & Gelles, 2007, 2009; DePanfilis & Zuravin, 1999a, 2001, 2002; English, Marshall, Brummel, & Orme, 1999; Fluke, Yuan, & Edwards, 1999; Fuller, Wells, & Cotton, 2001; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Lipien & Forthofer, 2004). The traditional use of linear models may be a factor that is undermining this area of research by producing inconsistent findings and low predictive accuracy (Gambrill & Shlonsky, 2000, 2001; Jones, Schwartz, Schwartz, Obradovic, & Jupin, 2006; Schwartz, Jones, Schwartz, & Obradovic, 2008; Schwartz, Kaufman, & Schwartz, 2004).

The neural networks used for the analyses that follow are multilayer (i.e., one layer of input nodes, one layer of hidden nodes, and one layer of output nodes) perceptron feed-forward networks with a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm (Gotwalt, 2011). Target functions were estimated using penalized maximum likelihood; choices regarding the networks' topology and architecture were made either (a) by the researcher in conjunction with the extant literature (see e.g., Beck, King, & Zeng, 2000, 2004; Bishop, 1995; Faraggi & Simon, 1995a, 1995b; Garson, 1998; Hastie, Tibshirani, & Friedman, 2001; Haykin, 1999; Mitchell, 1997; Smith, 1993), or (b) in accordance with the pre-specified settings of the JMP9 Pro software used to conduct the neural network analyses (Cox, Gaudard, Ramsey, Stephens, & Wright, 2010; Gotwalt, 2011). Before proceeding to a detailed description of the design and mathematical operations that characterize the neural networks created for this study, this particular section focuses on the use of maximum likelihood to estimate the target functions (a future discussion of the penalty term used to regularize or constrain unnecessary complexity in the model fitting process will follow).

As noted above, in the course of exploring the possibility that neural networks could be used to improve the accuracy of recurrent maltreatment risk prediction, no assumptions were made about the functional form of the relationship between the probability of maltreatment recurrence and selected predictors. Hence, the neural network analyses that follow did not begin with an equation that models the specific ways in which the probability of maltreatment recurrence is generated by a specific set of mathematical functions as applied to values of the predictor variables. For example, fitting a binary logistic regression model to data assumes that the log odds of maltreatment recurrence are linearly related to selected predictors and can be represented by the equation as follows:

$$\log(\frac{p_i}{1-p_i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... \beta_k x_{ki} .$$

Note the absence of mathematical functions for the parameters and the variables in the equation above (e.g., the parameters and variables appear with a power of 1 and are not multiplied or divided by another parameter or variable); hence, the functional form of the relationship between the log odds of recurrent maltreatment and selected predictors is specified as linear, and is said to be linear in the parameters and the variables (Gujarati, 2003). For example, if x_1 represented child age, for every one year of increase in the child's age, the rate of change in the log odds of recurrent maltreatment would be considered constant. The rate of change would not be constant if β_1 or x_1 were squared or cubed (Pampel, 2000). The assumption of linearity imposed by the logistic regression model constrains the probability of occurrence (*p*) as expressed in the form of a logit to either increase or decrease monotonically (i.e., in the form of a relatively straight line) as a function of a given predictor x_k with no possibility that *p* might depend on x_k in a curvilinear manner (e.g., in the form of a U) (Elkan, 2012; Hastie, Tibshirani, & Friedman, 2001).

With a pre-specified functional form in hand, the binary logistic regression model is then fitted to the data for the purpose of obtaining consistent and efficient beta coefficients that are evaluated for their statistical significance and role in explaining the outcome. The researcher assumes that the specified functional form is capable of producing unbiased parameter estimates and is the correct form to test the explanatory power of the predictor variables in relationship to established theoretical constructs. The analysis is not conducted for the purpose of exploring the functional form of the data-generating model and/or for estimating relationships not established through prior theory (Beck, King, & Zeng, 2004; Hastie, Tibshirani & Friedman, 2001).

While neural networks provide an assumption-free framework for estimating the functional form of the data-generating model, there is inherently a challenge in estimating such a model. Namely, without prior assumptions in hand, where does one begin the estimation process? Good starting points are maximum likelihood, the conditional probability mass function, and the conditional likelihood function (Gelman & Hill, 2007). The probability mass function, $f(x; \theta)$ is a distribution of realized or observed values (x) for a random variable (X) and the associated probability of observing x given fixed values for unknown parameters (θ) (Bishop, 1995). The probability mass function (PMF) describes a data-generating process based on the relative likelihood of observing specific values (x) for X. As noted by Gill (2006), the data-generating process as described by the PMF is a "probabilistic description of the underlying structure that determines the observed phenomenon" (p. 336). The PMF can be further specified as a conditional probability mass function that accounts for the fact that y (realized values of a random response variable Y) follows a different probability distribution for different fixed values of predictor variables x and a fixed set of parameters θ (Elkan, 2012). Hence, the conditional PMF is

$$f(y|x;\theta),$$

where specific values of x and parameter estimate $\hat{\theta}$ can be used to predict values for y. Given data consisting of $\langle x_i, y_i \rangle$ pairs (i.e., observations or case records) the principle of maximum likelihood is used to select a value for parameter estimate $\hat{\theta}$ for which y_i has the highest probability of occurrence; for each $\langle x_i, y_i \rangle$ pair, the best value for parameter estimate $\hat{\theta}$ is the one that maximizes the product of the probabilities across the $\langle x_i, y_i \rangle$ pairs (i.e., across the case records comprising the data set),

$\prod_i f(y_i \; |x_i;\theta)$ (Elkan, 2012, p. 4).

The conditional likelihood function is algebraically the same as the conditional PMF. However, the conditional PMF is a function of the observed values (y) for Y and expresses the relative plausibility or likelihood of observing y given different values for x and θ, while the conditional likelihood is a function of θ given different values for y and x. Specifically, Elkan (2012) noted that

Conditional PMF = $f(y|x;\theta)$, and

Conditional Likelihood Function = $L(\theta; y|x)$,

where for the conditional PMF, values for y are unknown or varying and values for x and θ are fixed, but for the conditional likelihood function, values for θ are varying and values for y and x are fixed. Likelihood is a tool for linking values of θ with the probabilities of observing specific values of y and therefore summarizes the evidence (i.e., the probabilities associated with values for y) supporting the choice of values for $\hat{\theta}$ (Bishop, 1995; Bolker, 2008; Elkan, 2012; Wilks, 2011). Approximating a neural network's target function with maximum likelihood centers on the estimation of the set of parameter estimates $\hat{\theta}$ by maximizing the conditional likelihood function where "the most reasonable values for $\hat{\theta}$ are those for which the probability of the observed sample is the largest" (Hastie, Tibshirani, & Friedman, 2001, p. 31).

In short, the PMF and the conditional likelihood function provide a solid foundation from which to estimate a target function using maximum likelihood because the most plausible values for the target variable (i.e., recurrent maltreatment) can be obtained in relationship to different values for the inputs (i.e., selected predictors) and the network weights (i.e., the set of parameter estimates $\hat{\theta}$, which are analogous to beta coefficients in a regression model) (Gelman & Hill, 2007). Moreover, values for the weights can be selected for the purpose of maximizing the probability of the observed target values, thereby improving the accuracy of the approximated target function while reducing the gap between the observed/actual target values and expected/estimated target values. In this way, the neural network learns from the patterns of input and output values and the probabilistic structure underlying the data in order to estimate a target function that is theoretically capable of generating estimated target values with a high probability of occurrence.

The neural network analyses for this dissertation study were conducted using JMP Pro 9 where the neural platform assumes that categorical response variables have a multinomial distribution[2] (Gotwalt, 2011). The multinomial probability mass function for

y_i, the outcome (i.e., recurrent maltreatment) for child i in one of c categories (i.e., re-reported for maltreatment or not re-reported for maltreatment) is as follows:

$$f(n_1, n_2 \ldots n_{c-1}) = \left(\frac{n!}{n_1!n_2!\ldots n_c!}\right) p_1^{n_1} p_2^{n_2} \ldots p_c^{n_c},$$

where n = the number of independent trials (i.e., number of children who were either re-reported or not re-reported for maltreatment), n_c = the number of trials having the outcome in category c, and p_c = the probability of responding in category c (wherein p remains constant across each trial) (Allen, 1990; Elkan, 2011). The binomial probability mass function is a special case of the multinomial probability mass function, where c = 2, and is represented as follows:

$$f(n_2) = \left(\frac{n!}{n_1!n_2!}\right) p_1^{n_1} p_2^{n_2}.$$

Maximum likelihood was used to approximate the target function that estimates the conditional probability of each class *k* (i.e., maltreatment recurrence and no recurrence) for the response variable *G*, given *X* and the parameter vector θ,

$$\Pr(G = k|X = x_i; \theta),$$

where for the two-class g_i, $y_i = 1$ when g_i = 1 and $y_i = 0$ when g_i = 2. For computational ease, the logarithm of the joint conditional likelihood function was maximized as represented by

$logL(\theta) = \sum_{i=1}^{N} logp_{g_i} (x_i; \theta)$ (Hastie, Tibshirani, & Friedman, 2001, p.98).

The maximum likelihood estimator of θ are those values for which the observed data have the highest probability of occurrence and therefore best conform with the data as follows:

$$\hat{\theta} = argmax_{\theta} = \sum_{i=1}^{N} logp_{g_i} (x_i; \theta),$$

$\hat{\theta} = argmin_{\theta} = -\sum_{i=1}^{N} logp_{g_i} (x_i; \theta)$ (Bishop, 1995; Elkan, 2012; Gotwalt, 2011; Hastie, Tibshirani, & Friedman, 2001). Figure 3.2 provides an illustration of the maximum likelihood estimation of θ.

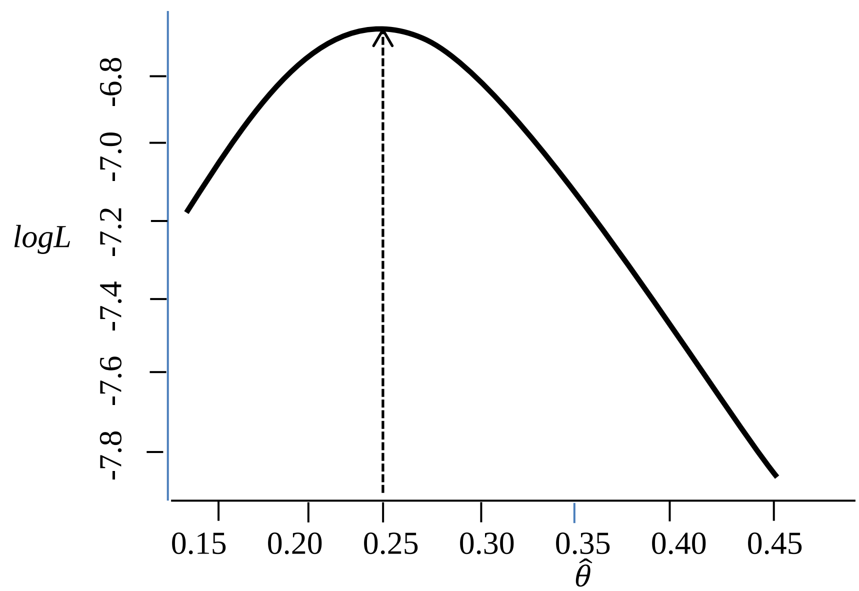


Figure 3.2. The maximum likelihood estimate for $\hat{\theta}$ is the value that makes the data most likely ($\hat{\theta}$ = 0.25) and is the point on the plot where the log- likelihood takes the largest value.

As noted in the equations above, maximizing the log joint conditional likelihood function is equivalent to minimizing the negative log joint conditional likelihood function. Moreover, minimizing the negative log-likelihood function is equivalent to minimizing the cross-entropy or deviance error function (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001; Mitchell, 1997). The accuracy of the estimated target function improves in relationship to the minimization of the cross-entropy error function, a process that is more easily understood by visualizing the reduction in the error function as dependent upon the search for the lowest point (i.e., global minimum) of the error surface sitting above weight space. Error is minimized as a function of the vector of values for the network weights, $E(w)$. A learning algorithm such as the BFGS quasi-Newton algorithm conducts a search of the error surface in order to find the vector of weights associated with the lowest point (i.e., the global minimum) of the error surface and therefore the smallest value for E as well as the values for the weights that make the data most likely (Bishop, 1995; Haykin, 1999; Mitchell, 1997; Smith, 1993).

Figure 3.3 depicts the learning process that occurs when the minimization of the error function is used to provide the best possible fit between the estimated target function and the conditional likelihood of the observed values for Y. By exploring the error surface, the network learns the values of the network weights that minimize the error function, which is equivalent to learning the vector of network weights for which the observed data have the highest probability of occurrence (Bishop, 1995; Smith, 1993). Each point in weight-space corresponds with the coordinates for a set of possible values for the network weights, and the dashed line with an arrow pointing upward represents the fluctuating value of the error function (*E)* measured as the height over the point in weight space that corresponds with the coordinates for a given set of values for the network weights (Bishop, 1995; Smith, 1993). The symbol ∇E represents the gradient of the error function with respect to the weights, and the negative gradient of *E* is a vector of values that reduce *E* by providing a direction for the search of the lowest point of the error surface (Pandya & Macy, 1996). The placement of ∇E at point C in Figure 3.3 is used to symbolize the search for $\nabla E = 0$ or the point at which a continued search along the error surface would not produce a decrease in *E* (Bishop, 1995; Smith, 1993). The gradient is equal to zero at this point because a continued search of the error surface would only increase *E*. Judging from the position of point C in the figure, it is clear that the search for the lowest value for *E* would need to continue; however, points A and B symbolize two points that could satisfy $\nabla E = 0$ and would therefore be local minima of the error function (Bishop, 1995). The true minimization of *E* occurs when the global minimum has been identified, and it is at this point that the fit is best between the estimated target function and the conditional likelihood of the observed values for Y (Haykin, 1999; Mitchell, 1997).

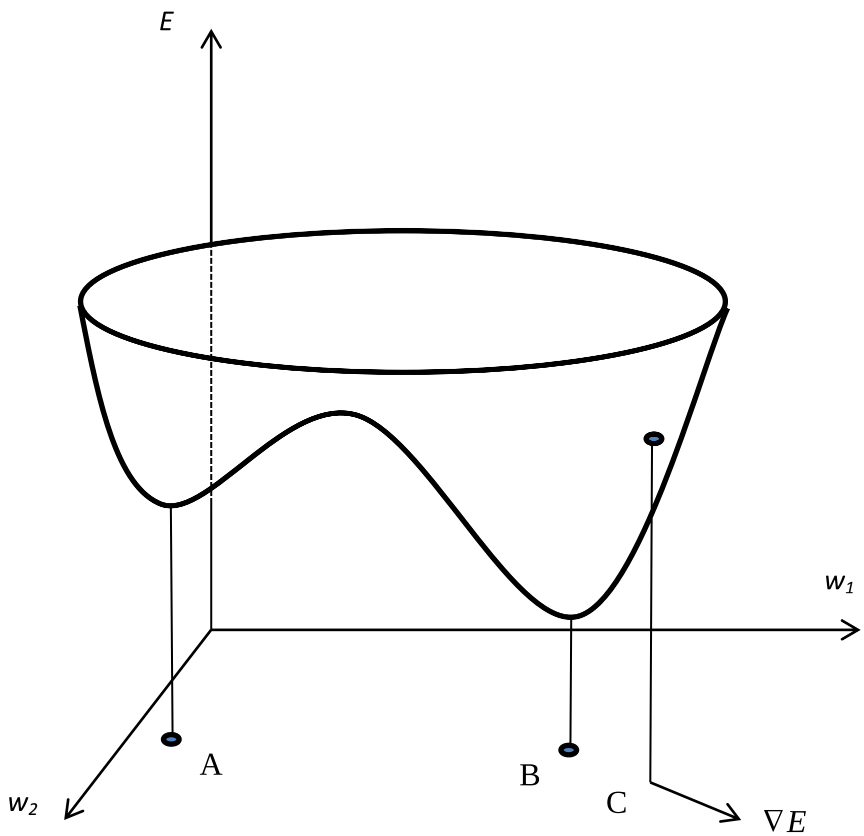


Figure 3.3   Depiction of the minimization of the error function as a search for the lowest point along the error surface sitting above weight space. Adapted from *Neural Networks for Pattern Recognition*, by C. M. Bishop, 1995, p. 254.

To summarize the points made above, it is helpful to imagine that before deciding to conduct a neural network analysis, the researcher decided to fit a binary logistic model to a given data set in order to estimate the likelihood of recurrent maltreatment to a set of static and dynamic risk factors. A binomial distribution of the data was assumed and the regression model was fit by maximum likelihood using the log joint conditional likelihood (*LCL*) as follows:

$$LCL = \sum_{i=1}^{N} logL(\theta; y_i\,|x_i) = \sum_{i=1}^{N} log\, f(y_i\,|x_i; \theta)$$ (Elkan, 2012, p. 6).

The maximum likelihood estimate for $\hat{\theta}$ is the value that makes the data most likely; maximizing the LCL links the relative plausibility of values for Y with given values for X with the vector of parameter estimates $\hat{\theta}$. Values for Y, X and the vector of parameter estimates $\hat{\theta}$ are the key "ingredients" in a "recipe" that relates the probabilities of recurrent maltreatment class membership (the likelihood of being in the group that was re-reported for maltreatment and the likelihood of being in the group that was not re-reported for maltreatment) to a combination of linear functions in *x* as follows:

$$\Pr(Y = 1|X = x) = \frac{e^{\beta_0 + \beta_k x_k}}{1 + e^{\beta_0 + \beta_k x_k}},$$

$$\Pr(Y = 0|X = x) = \frac{1}{1 + e^{\beta_0 + \beta_k x_k}},$$

$$log\frac{\Pr(Y=1|X=x)}{\Pr(Y=0|X=x)} = \beta_0 + \beta_k x_k$$ (Hastie, Tibshirani, & Friedman, 2001, p. 80).

The following equations put all of the "ingredients" together in the linear regression model's "recipe" in order to solve for the vector of parameter estimates $\hat{\theta}$ (i.e., the set of regression coefficients $\hat{\beta}$) by maximizing the log joint conditional likelihood:

$$LogL\,(\beta) = \sum_{i=1}^{N}\{y_i logp(x_i; \beta) + (1 - y_i)\log(1 - p(x_i; \beta))\}$$

$$= \sum_{i=1}^{N}\left\{y_i \beta^T x_i + \log(1 + e^{\beta^T x_i})\right\},$$

where $p_1(x; \theta) = p(x; \theta), and\ p_0(x; \theta) = 1 - p(x; \theta)$ (Hastie, Tibshirani, & Friedman, 2001, p. 98).

Fitting a neural network model to the same given data set to approximate a function that relates the probabilities of recurrent maltreatment class membership to a set of static and dynamic risk factors follows a similar but more general and flexible set of steps as detailed above. The difference can be seen in the equations for maximizing the log joint conditional likelihood. The equation for the logistic regression model has been specified with a functional form that explains how values for X and $\hat{\beta}$ predict values for Y, while the equation for the neural network has not been specified with a particular functional form,

Logistic Regression = $LogL\ (\beta) = \sum_{i=1}^{N}\{y_i logp(x_i;\beta) + (1-y_i)\log(1-p(x_i;\beta))\}$

Neural Network = $LogL(\theta) = \sum_{i=1}^{N} logp_{g_i}(x_i;\theta)$,

where $p_k(x_i;\theta) = \Pr(G = k|X = x_i;\theta)$ (Hastie, Tibshirani, & Friedman, 2001, p. 98).

The "take-home message" can be found in the following sentiments. *First*, the estimation of the target function is based on maximizing the conditional likelihood function -- figuring out how the patterns of input values map onto the target values is greatly facilitated by understanding how the network weights can be chosen to increase the probability of observing the target values given the inputs. *Second*, reducing the difference between the actual target values and the network-generated target values (i.e., the problem of misclassification) is based on decreasing an error function that leads right back to maximizing the conditional likelihood function. Error is decreased by choosing network weights that make the observed target values given the inputs most likely. *Third*, maximizing the conditional likelihood function is nothing new; in fact, these same principles are used when carrying out a binary logistic regression. Hence, binary logistic regression and the neural network analysis carried out in this dissertation study share the same assumptions about the distribution of the response variable and the best way of estimating θ. What is different is the functional form that is used to model the relationship between the inputs and the target variable; neural networks do not assume that Y is equal to a linear combination of X. The following section will provide specific details about the form and function of a multilayer perceptron feed-forward neural network. It is within this section that the critical role of the "hidden space" is introduced; specifically, a parallel set of logistic regressions is contained within the layer of hidden nodes and it is within this "hidden space" that Y can be modeled as a nonlinear function of X (Bishop, 1995, 2006; Cheng & Titterington, 1994; Paik, 2000).

Neural Network Form and Functions: Specifics About How a Neural Network Operates

While the preceding sections provided a theoretical rationale for a neural network's ability to estimate a target function based on the conditional probability of observing specific values for the target variable given values for the inputs and network weights, this section describes in more concrete detail the specifics of a multilayer perceptron (MLP) feed-forward neural network. All information about neural networks will be based on the type of model used for this dissertation study; that said, as indicated earlier, neural networks are a class of estimators and the inherent flexibility that characterizes a neural network extends to the myriad of ways that a network's topology and architecture can be specified (Bishop, 1995; Garson, 1998; Haykin, 1999; Mitchell, 1997). For the purpose of this dissertation study, a multilayer network is defined as including one layer of input nodes, one layer of hidden nodes, and one layer of output nodes. As Garson (1998) notes, the language used to characterize neural networks can vary and this includes how a multilayer network is described. Differences in the description of what constitutes a multilayer network occur in relationship to the number and type of processing entities that are included. In short, the number of nodes that perform mathematical calculations – i.e., the number of perceptrons -- is sometimes used to define the number of layers within a given neural network (Garson, 1998). Moreover, some scholars limit the description of what constitutes a multilayer network even further by focusing only on the number of layers that contain hidden nodes (Bishop, 1995, 2006). Again, for the purpose of this dissertation study, a multilayer network will always refer to a network with one layer of input nodes, one layer of hidden nodes, and one layer of output nodes.

Networks are described as being fully connected when each node in a preceding layer is connected to a node in the following layer, where the connections between a pair of nodes can be thought of as a synapse or pathway that facilitates the transmission of information from one node to another. The strength of each connection and therefore the relative influence of the inputs in estimating the target function are determined by each respective connection weight (Garson, 1998; Smith, 1993). The neural network has two layers of perceptrons (also referred to as nodes or neurons) that perform mathematical functions to include the hidden nodes and the output nodes. Input nodes merely store and forward propagate the input values for each observation or case record to the hidden nodes (Haykin, 1999; Mitchell, 1997).

Information in a feed-forward network can only be passed forward and is described as being forward propagated from (a) a given input node to each hidden node, and (b) from a given hidden node to the output node. Values for the explanatory variables are forward propagated to each hidden node for parallel processing. Hence, the same set of input values are forward propagated to each hidden node for processing because each input node is connected to every hidden node. Each hidden node is equipped with a summation and an activation function for the purpose of creating a linear combination of the inputs that is then transformed into a nonlinear term. Each input value is multiplied by its respective connection weight and the products are summed together. This combination of weighted input values is then transformed into a nonlinear term through the activation function. Calculations in the hidden nodes occur simultaneously and nonlinear terms produced by each hidden node are unique because the input values are multiplied by a different set of connection weights that lead to each hidden node. The nonlinear terms that are created by each hidden node are forward propagated to the output node for final processing. Hence, each hidden node is connected to the output node (Cheng & Titterington, 1994; Paik, 2000).

Similar to the hidden nodes, the output node is also equipped with a summation and an activation function. The nonlinear terms that are forward propagated by each hidden node are multiplied by their respective connection weights and then summed. This linear combination of nonlinear terms is then transformed via the activation function. Different activation functions may be employed within the hidden nodes and the output node (Garson, 1998). The neural network in this dissertation study was equipped with the hyperbolic tangent function (tanh) in all of the hidden nodes and the logistic function in the output node.

While the logistic and tanh functions are similar in that they are rescaled versions of the other, use of the tanh function in the hidden nodes is recommended (Blackwell & Chen, 2009; Garson, 1998; Haykin, 1999) while use of the logistic function for neural networks with a categorical response variable is recommended for the output node (Bishop, 1995). Both activation functions serve the purpose of defining the relative amplitude or strength of the signal that is comprised of the information extracted from the pattern of input values for each observation or case record. Similar to a biological neuron, each artificial neuron or node in a neural network sends (input nodes and hidden nodes) and/or receives (hidden nodes and output nodes) a signal that serves to excite or inhibit processing activity in the following node. The activation function defines the degree (i.e., magnitude) to which and the form (i.e., linear or nonlinear functions to include interaction terms) in which the input nodes influence the estimation of the target function (Abdi, Valentin, & Edelman, 1999; Cheng & Titterington, 1994; Mitchell, 1997; Paik, 2000).

The logistic function places values for the nonlinear term on a scale bound by 0 and 1, while the tanh function places values for the nonlinear term on a scale bound by -1 and 1 (Garson, 1998; Haykin, 1997). Both nonlinear activation functions are described as being sigmoidal because both produce S-shaped curves that place values for the nonlinear term anywhere between the lower and upper bounds of the activation function's range. Use of the antisymmetric (i.e., $f(x) = -f(x)$) tanh function in the hidden nodes is recommended because of the reduction gained in the amount of training time needed for the network to iteratively learn the input-output mapping process through the adjustment of the estimated network weights (Blackwell & Chen, 2009; Garson, 1998; Haykin, 1997). Use of the logistic activation function in the output node is recommended when the response variable is categorical because the range of output running from 0 to 1 allows the researcher to estimate the conditional probabilities of class membership (Bishop, 1995).

Figure 3.4 below provides a visual comparison of a biological "neural network" with an artificial "neural network"; in both cases, the basic components of each system are highlighted as opposed to providing a full picture of a particular system of neurons. In short, the biological system that inspired the creation of artificial neural networks provides some insight into the key ideas that support the utility of an artificial neural network. Specifically, like the biological system at the top of Figure 3.4, the artificial neural network does not operate according to a pre-specified set of rules that dictate how inputs are believed to be associated with outputs. Information is what fuels the way the system operates; relationships are approximated on the basis of what the neural network learns from the information that is sent, processed and received among parallel structures.

The forthcoming Figure 3.4 provides a detailed description of the form and function of a simple neural network -- i.e., one with two input nodes, two hidden nodes and one output node. Every layer, node, connection weight, and node-based calculation is labeled. That said, before proceeding to the description, it is important to understand how the neural network acts as a classifier and how the use of fixed nonlinear basis functions allow the neural network to classify cases that are not linearly separable (Bishop, 1995, 2006).

One of the key features of the neural network created for this dissertation study is the two- stage use of fixed nonlinear basis functions (i.e., the tanh and logistic functions) to transform the inputs in a linear combination into basis functions -- i.e., representations of the original inputs that allow for the application of a linear decision boundary in the new "feature space" to separate patterns of inputs that are not linearly separable in the original input space (Bishop, 2006; Haykin, 1999). In short, one of the main goals of a neural network analysis with a categorical response variable is the creation of decision boundaries that divide the input space (i.e., x-dimensional space defined by the number of predictor variables) into c decision regions (where c= the number of response categories).

By working backwards from the data in order to estimate a target function that

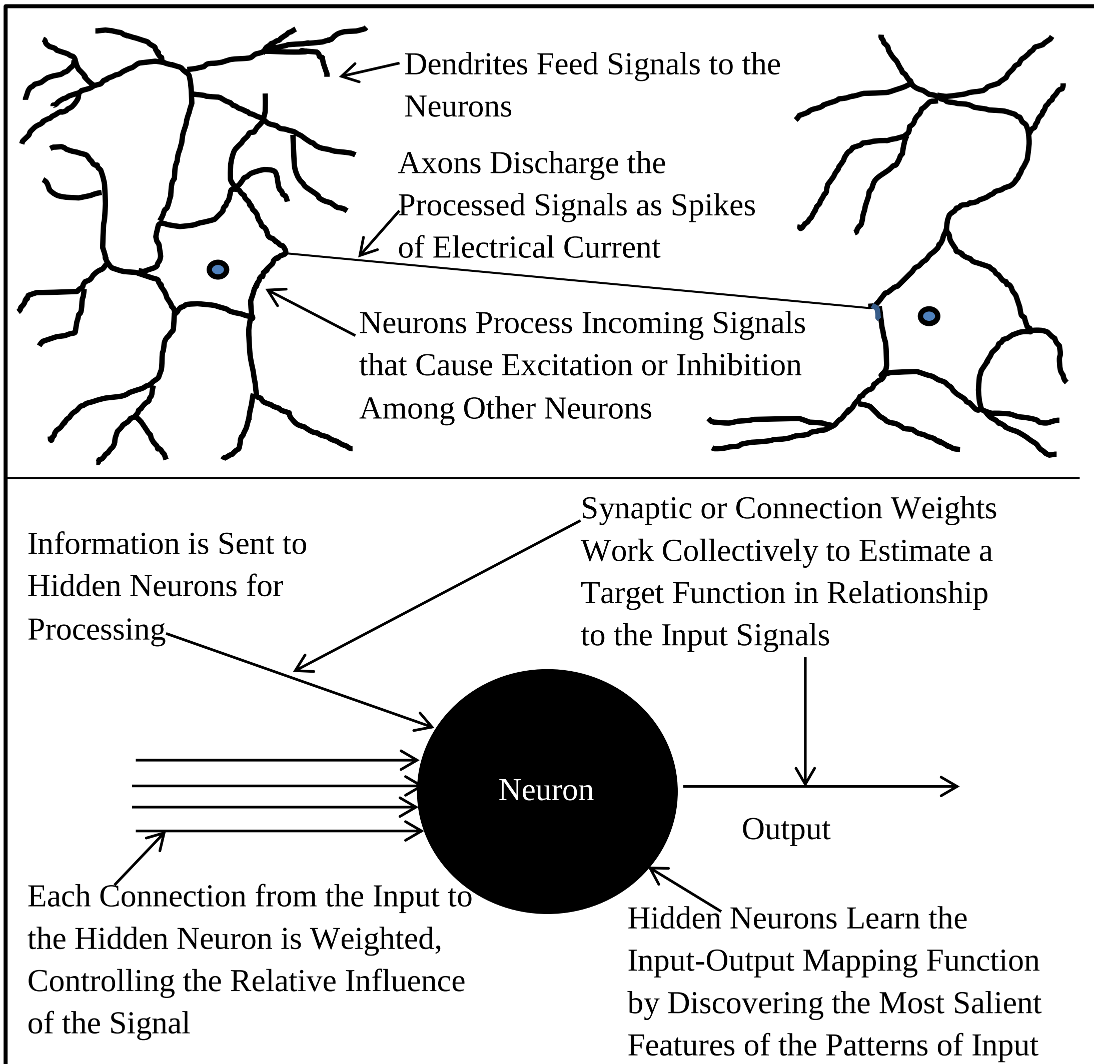


Figure 3.4 Depiction of the biological inspiration for the creation of an artificial neural network. Adapted from *Neural Networks: An Introductory Guide for Social Scientists*, by G. D. Garson, 1998, p. 24.

mathematically defines how input variables map onto values of the output variable, researchers use neural networks for the purpose of learning how to build decision boundaries that predict how patterns or combinations of values for the input variables predict output values (Hastie, Tibshirani, & Friedman, 2001). The estimated target function describes in algebraic terms what the decision boundaries look like in geometric terms because "a pattern classifier [neural network] provides a rule for assigning each point of feature space [transformed input space] to one of c classes" (Bishop, 1995, p. 24). Essentially, the target function learns how patterns (i.e., observations or case records) of inputs can be used to predict into which response category each new observation or case record will fall.

The purpose of estimating a target function is ultimately the creation of a rule that allows for the future prediction of new cases -- i.e., to be able to predict which children are likely to be re-reported and which children are unlikely to be re-reported on the basis of their patterns of inputs. Hence, the neural network learns how to classify or assign cases to a decision region by maximizing the conditional likelihood function in order to estimate the network weights for which the probability of a response outcome is the highest. With known values for the network weights in place, the conditional probability mass function $P(c_k|x;\theta)$ gives the probability that the child should be assigned to a response category of k (re-reported) given the child's vector of input values and the vector of network weights; additionally, the conditional PMF $P(c_j|x;\theta)$ gives the probability that the child should be assigned to a response category of j (not re-reported) given the child's vector of input values and vector of network weights. The possibility of misclassification is minimized by choosing to assign the child to the class that has the highest probability of occurrence (Bishop, 1995).

The classification process makes intuitive sense, but the possibility of misclassification is very high if the estimation of the decision boundaries is limited to a linear classifier where the only type of decision boundary that can be created is a straight line. Figure 3.5 provides an example of the type of decision boundary that is created by a linear classifier. In this case, the input space is limited to two dimensions where each of the two predictors is placed on an axis. The input space is defined by the combination of predictors used to classify each case by its outcome. As noted above, for every combination of values on the predictors, each pattern of inputs is assigned a point in input space that ultimately falls within a decision region. Each decision region represents the group of patterns or cases belonging to an outcome response category. The ability to correctly predict case outcomes on the basis of the case input values depends on the accurate estimation of the decision regions in relationship to a decision boundary that separates the regions. Figure 3.5 provides an example of the high level of misclassification that can occur when it is assumed that the best classifier is linear. In contrast, Figure 3.6 provides an example of the reduction in misclassification error that can be achieved by using a nonlinear classifier that is capable of creating a decision boundary that is curved. A linear functional form that characterizes linear regression, binary logistic regression, and Cox regression can only produce linear (i.e., straight line) decision boundaries. As discussed previously, neural networks do not assume that a particular functional form exists, and in fact, the neural network estimates the functional form (Bishop, 1995, 2006; Breiman, 2001b; Cheng & Titterington, 1994; Hastie, Tibshirani, & Friedman, 2001; Paik, 2000).

The ability to estimate a target function that creates something other than a straight line (if called for based on the mapping of the inputs onto the outputs) is facilitated by the creation of nonlinear basis functions. In the case of a neural network, a sigmoidal activation function, to include the logistic and tanh functions, can be used to transform the original inputs included in a linear combination (Bishop, 2006). For example, in a network with two input nodes, two hidden nodes, and one output node (please see

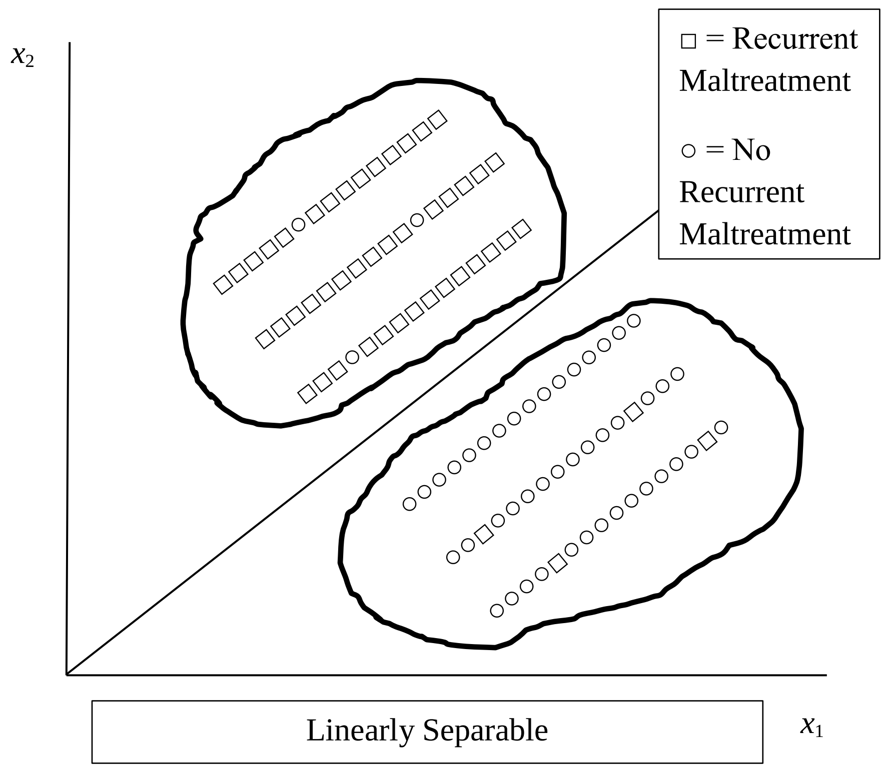


Figure 3.5 Depiction of the classification of outcome values for a binary dependent variable as a function of the input values for x_1, x_2 and the network weights; the clear majority of outcome values can be separated as a function that is linear in both the network weights and the inputs (on the basis of Bishop, 1995, 2006; Hastie, Tibshirani, & Friedman, 2001).

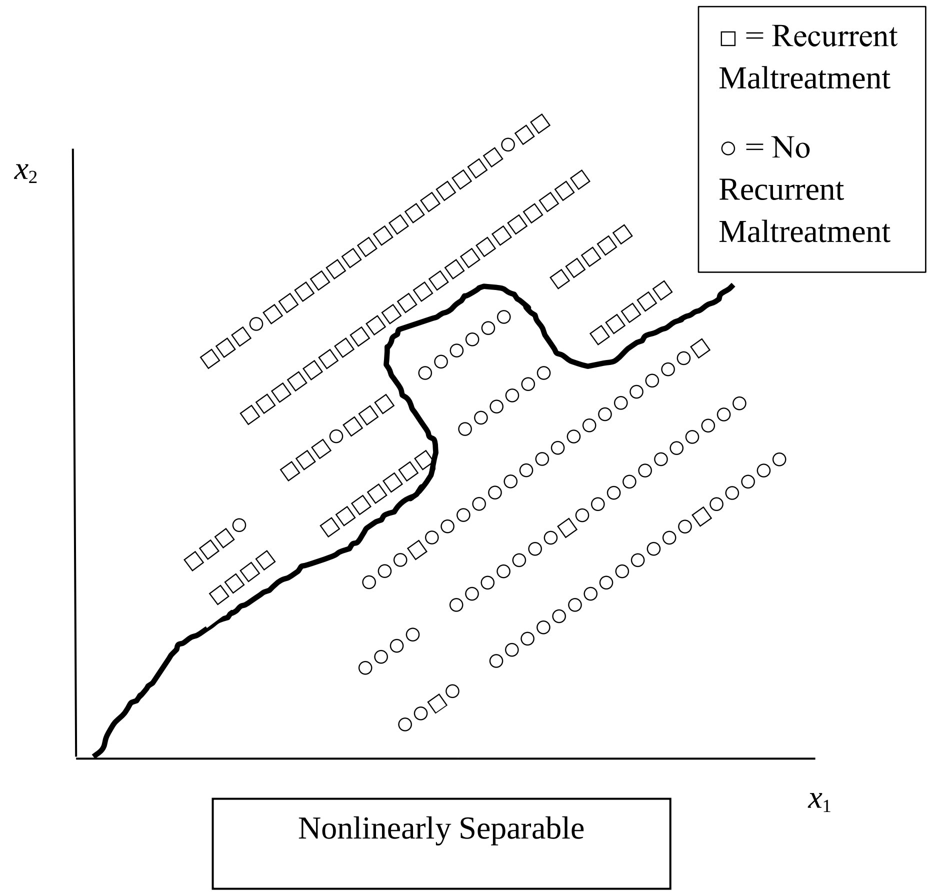


Figure 3.6   Depiction of the classification of outcome values for a binary dependent variable as a function of the input values for x_1, x_2 and the network weights; the clear majority of outcome values cannot be separated as a function that is linear in both the network weights and the inputs (on the basis of Bishop, 1995, 2006; Hastie, Tibshirani, & Friedman, 2001).

Figure 3.7), a linear combination of the inputs is created and network weights are estimated in relationship to the original inputs within both hidden nodes as follows:

$$u_1 = a_0 + a_1 x_1 + a_2 x_2$$

$$u_2 = b_0 + b_1 x_1 + b_2 x_2,$$

where $u_1 = u_2 = \Pr(G = k|X = x; \theta)$, a_0 and b_0 are bias weights (intercepts) for hidden nodes one and two, a_1 and a_2 are connection weights between the inputs and hidden node one, and b_1 and b_2 are connection weights between the inputs and hidden node two. A fixed nonlinear function, the tanh function, is used to transform the inputs into basis functions, $\varphi(x)$, in order to create nonlinear functions of *x*,

$$g(u_1) = a_0 + a_1\varphi_1(x_1) + a_2\varphi_2(x_2)$$

$$g(u_2) = b_0 + b_1\varphi_1(x_1) + b_2\varphi_2(x_2),$$

where $g(u_1)$ and $g(u_2)$ are basis functions or features that are used to represent the original inputs in a binary logistic regression that is conducted in the output node.

Ultimately, the neural network creates a regression model that is linear in the parameters and nonlinear in the inputs; moreover, the *linear* function (i.e., linear combination of transformed inputs) of the nonlinear basis functions (i.e., transformed inputs) in the *feature* space (i.e., projection of the transformed input space) becomes a *nonlinear function* in the *original input space* (Bishop, 2006; Hastie, Tibshirani, & Friedman, 2001). The inputs are projected into feature space so that a linear decision boundary can be used to separate case observations by their features. Linear decision boundaries in feature space $\varphi(x)$ correspond to nonlinear decision boundaries in the original input space *x*, and classes that can be linearly separated in feature space $\varphi(x)$ do not need to be linearly separable in input space *x* (Bishop, 2006). Transforming the inputs into basis functions to be used as inputs in a linear regression model is what gives neural networks their inherent flexibility in estimating a more complex decision boundary. Constructing parallel logistic regressions in the hidden nodes allows the neural network to learn different parts of the input-output mapping process that can be used to create local decision boundaries (Smyth, 2007). Haykin (1999, p. 248) summarizes the flexibility gained in creating local decision boundaries through each hidden node as follows:

> Each neuron is responsible for producing a hyperplane of its own in decision space. Through a supervised learning process, the combination of hyperplanes formed by all the neurons in the network is iteratively adjusted in order to separate patterns drawn from different classes not seen before [new cases], with the fewest classification errors on average.

Figures 3.7 and 3.8 put into visual play all of the previously discussed details regarding the form and functions of a multilayer perceptron feed-forward neural network. Figure 3.7 walks the reader through the process of a fairly simple network with two input nodes, two hidden nodes, and one output node, while Figure 3.8 provides a detailed picture of a more complex neural network with ten input nodes, four hidden nodes, and one output node. Each of these figures is followed by a detailed set of bullet points that describe the function of each component included in the diagram to include the provision of all mathematical calculations that take place within each processing entity.

- a_0 = the bias weight for the first hidden node; the bias weight functions as an intercept and is added to the weighted sum of input values
- b_0= the bias weight for the second hidden node; the bias weight functions as an intercept and is added to the weighted sum of input values
- c_0= the bias weight for the output node; the bias weight functions as an intercept and is added to the weighted sum of the hidden node outcome values
- x_1 = the input values for the first independent variable
- x_2 = the input values for the second independent variable

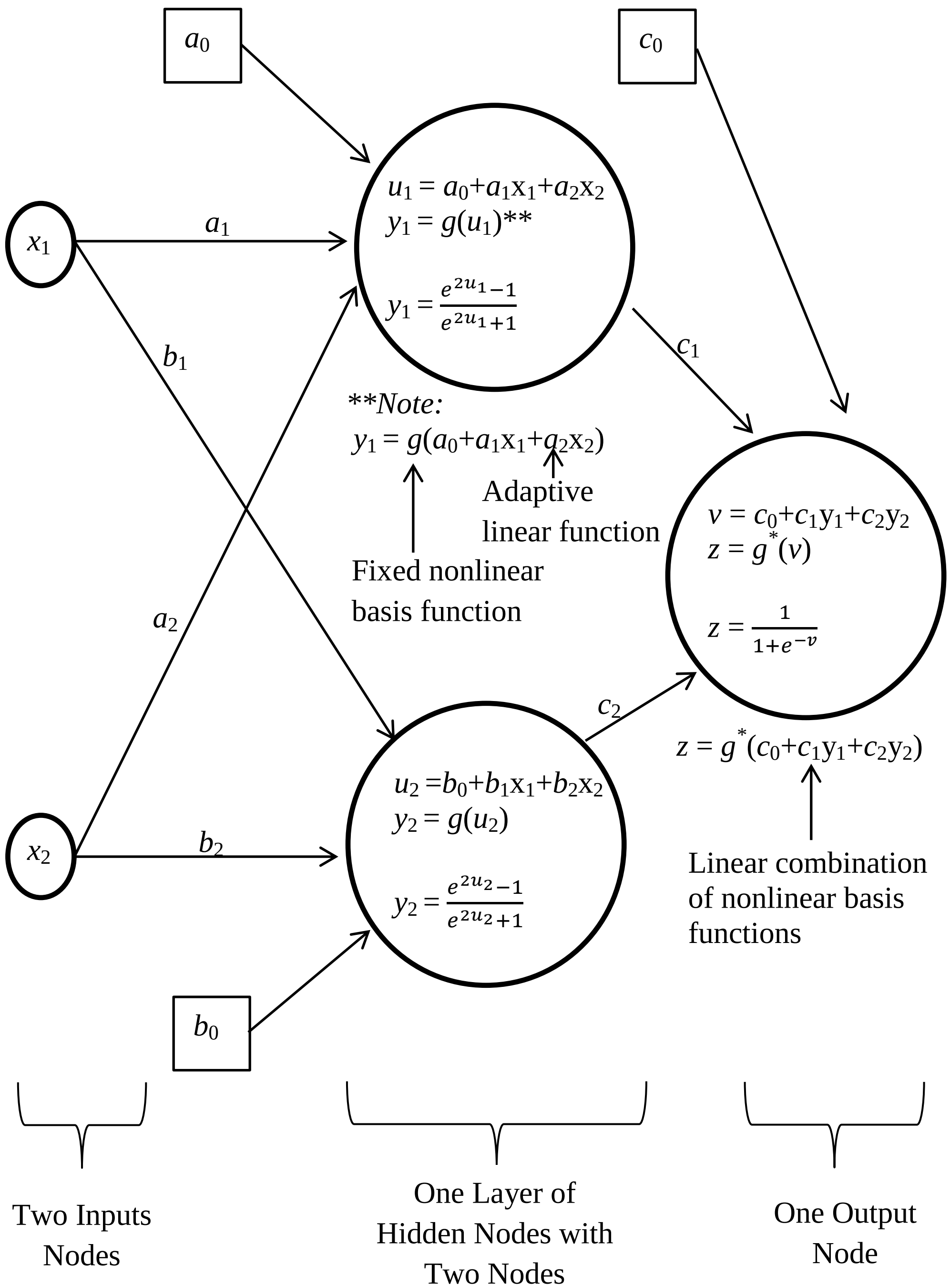


Figure 3.7. Depiction of the parallel calculations and forward propagation of information within a multilayer perceptron feed-forward neural network with two inputs, two hidden nodes, and one output node (on the basis of Bishop, 1995; Garson, 1998; and Smith, 1993).

- a_1= the synaptic weight that determines the strength of the connection between x_1 and the first hidden node as well as the relative influence -- in terms of magnitude and direction – that x_1 has on the estimation of the target function; a_1 functions as the regression coefficient for x_1 in what amounts to a logistic regression in the first hidden node
- a_2= the synaptic weight that determines the strength of the connection between x_2 and the first hidden node as well as the relative influence -- in terms of magnitude and direction – that x_2 has on the estimation of the target function; a_2 functions as the regression coefficient for x_2 in what amounts to a logistic regression in the first hidden node
- b_1= the synaptic weight that determines the strength of the connection between x_1 and the second hidden node as well as the relative influence -- in terms of magnitude and direction – that x_1 has on the estimation of the target function; b_1 functions as the regression coefficient for x_1 in what amounts to a logistic regression in the second hidden node
- b_2= the synaptic weight that determines the strength of the connection between x_2 and the second hidden node as well as the relative influence -- in terms of magnitude and direction – that x_2 has on the estimation of the target function; b_2 functions as the regression coefficient for x_2 in what amounts to a logistic regression in the second hidden node
- u_1= the linear combination of the bias weight and the input values in the first hidden node, where each input value is multiplied by its respective connection weight
- u_2= the linear combination of the bias weight and the input values in the second

hidden node, where each input value is multiplied by its respective connection weight

- $g(u) = \frac{e^{2u_1 - 1}}{e^{2u_1 + 1}}$ = the hyperbolic tangent (tanh) function is a fixed nonlinear basis function φ(u) that is used to transform the linear combination of inputs into a nonlinear function of the inputs called a basis function; forward propagation of the basis function to the output node for processing then allows the output node to create a linear combination of the nonlinear basis functions across all hidden nodes
- y_1= the output produced by the first hidden node where the linear combination of the bias weight and input values is transformed into a nonlinear basis function via the tanh activation function; y_1 is forward propagated to the output node and represents a feature the network has extracted in order to identify the most salient aspects of the patterns of inputs
- y_2= the output produced by the second hidden node where the linear combination of the bias weight and input values is transformed into a nonlinear basis function via the tanh activation function; y_2 is forward propagated to the output node and represents a feature the network has extracted in order to identify the most salient aspects of the patterns of inputs
- c_1= the synaptic weight that determines the strength of the connection between the first hidden node and the output node; c_1 functions as the regression coefficient for the hidden node-produced nonlinear term that represents aspects of the input data that are most relevant in predicting values for the response variable
- c_2= the synaptic weight that determines the strength of the connection between the

second hidden node and the output node; c_2 functions as the regression coefficient for the hidden node-produced nonlinear term that represents aspects of the input data that are most relevant in predicting values for the response variable

- v= the linear combination of the bias weight and the hidden-node produced nonlinear basis functions, where each nonlinear term is multiplied by its respective connection weight
- $g^*(v) = \frac{1}{1+e^{-v}}$ = the logistic function is a fixed nonlinear basis function $\varphi(v)$ that is used to transform a linear combination of nonlinear basis functions produced by the hidden nodes
- z= output produced by the output node where the linear combination of the bias weight and nonlinear basis functions is transformed into a second nonlinear term via the logistic activation function; the application of the logistic activation function completes what amounts to a final logistic regression conducted in the output node that yields the estimated probabilities of class membership (either re-reported for maltreatment or not re-reported for maltreatment)
- I=the number of inputs, x_i
- x_i= the input value from the i^{th} input
- a_i = the weights from input i to hidden node 1
- b_i = the weights from input i to hidden node 2
- c_i = the weights from input i to hidden node 3
- d_i = the weights from input i to hidden node 4
- J= the number of hidden nodes, j
- e_j= the weight from hidden node j to the output node

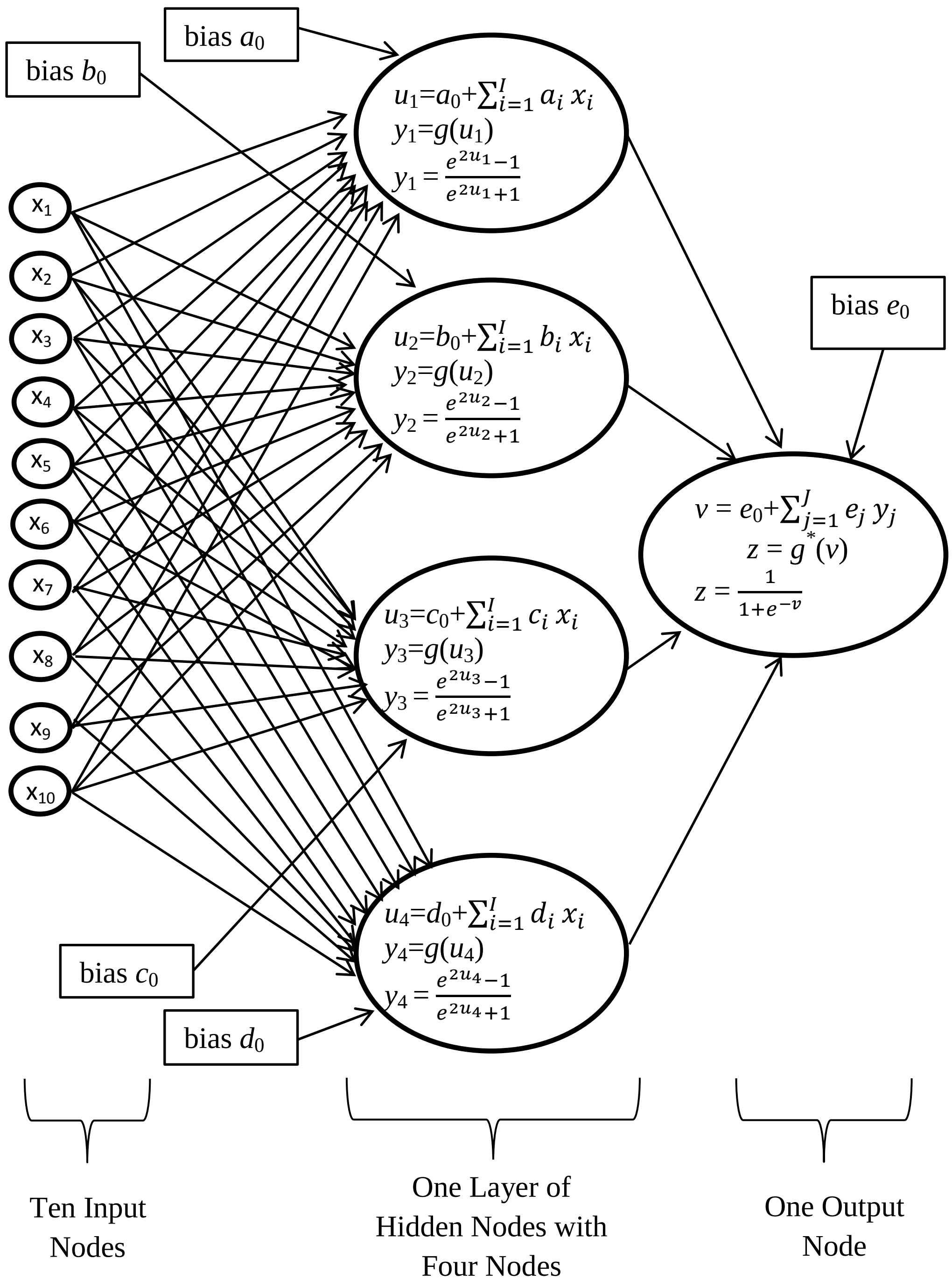


Figure 3.8. Depiction of the increasing complexity of the parallel calculations and forward propagation of information within a multilayer perceptron feed-forward neural network with ten inputs, four hidden nodes, and one output node (on the basis of Bishop, 1995; Garson, 1998; and Smith, 1993).

- y_j= the output value forward propagated from hidden node j to the output node
- bias a_0= the bias weight for hidden node 1
- bias b_0= the bias weight for hidden node 2
- bias c_0= the bias weight for hidden node 3
- bias d_0= the bias weight for hidden node 4
- bias e_0= the bias weight for hidden node 5

For a feed-forward network containing any number of input nodes, any number of hidden nodes (within one layer of hidden nodes), and one output node, the output from the hidden nodes can be represented as

$$y_j = g_{tanh}(a_{0j} + \sum_1^I a_{ij}\, x_i),\ j=1, \ldots, J,\ i=1, \ldots I,$$

where a_{ij} are the weights from each input i to each hidden node j. The output from the output node can be represented as

$$z = g_{logistic}(b_0 + \sum_1^J b_j\, y_j),\ j=1, \ldots, J,$$

where b_j are the weights from each hidden node to the output node.

As the number of hidden nodes increases, the complexity of the neural network increases. Adding a hidden node provides an additional opportunity to extract information from the input variables in a different manner. As noted earlier, the hidden nodes transform each input into a basis function, which is also referred to as feature extraction in the neural network literature (Bishop, 1995, 2006; Hastie, Tibshirani, & Friedman, 2001). While feature extraction should never be used in lieu of careful and strategic data management as well as input selection, it does provide opportunities to identify aspects of the data that might otherwise have been entirely overlooked. For example, neural networks can automatically test for any number of interactions that could be very helpful in predicting class membership but would not otherwise be identified by the researcher prior to analysis or would not otherwise be pragmatic given the following: (a) the number of free parameters required to test for such a large array of interactions; (b) issues related to multicollinearity; and (c) the need to specify the correct form of the interaction prior to analysis (e.g., a simple multiplicative term will only test for a linear interaction) (Beck, King, & Zeng, 2000, 2004; Jaccard, 2001).

The flexibility in estimation provided by the neural network is particularly suited to areas of research that are not well supported by theory and/or are subject to inconsistent findings. Moreover, given the lack of assumptions constraining a neural network analysis, results can be applied to future studies in any number of ways to include (a) assessing the benefits of re-specifying the form that predictors take such as the potential benefits of including higher order polynomial terms, (b) identifying and including key interactions terms, and (c) assessing the degree to which assumptions about monotonically increasing or decreasing relationships have precluded researchers from developing and testing new approaches to measurement.

Neural Network Form and Functions: A Comparison with Logistic Regression

In short, the neural network in this dissertation study is a nonlinear generalization of a binary logistic regression model (Faraggi & Simon, 1995b; Hastie, Tibshirani, & Friedman, 2001), and because the tanh function used in the hidden nodes is a centered and rescaled version of the logistic function (Haykin, 1997; SAS, 2010), a comparison of the neural network employed in this dissertation study with a binary logistic regression model is applicable. Where the linear functions of the selected predictors ($\beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$) in a standard binary logistic model are transformed into nonlinear basis functions through the logistic activation function, the basis functions are not carried forward for a second regression. In contrast, the nonlinear basis functions produced in the hidden nodes are carried forward,

$$(\theta_0 + \theta_1 z_1 + \cdots + \theta_p z_p),$$

where θ_0 = the output node bias weight, θ_p = the connection weight from hidden node j to the output node, and z_p = the nonlinear basis function forward propagated by hidden node j.

If "the central idea is to extract linear combinations of the inputs as derived features, and then model the target as a nonlinear function of these features," (Hastie, Tibshirani, & Friedman, 2001, p. 347), then the hyperbolic tangent function in each hidden node is the operation that transforms the inputs into nonlinear basis functions (also described as feature extraction) to be entered as the inputs in a linear combination that is again transformed by a fixed nonlinear basis function (in the output node the logistic activation function is used). The nonlinear basis functions in the hidden nodes are estimated as linear functions of *x* before being transformed by the tanh function into nonlinear terms; however, the nonlinear basis functions in the output node are included as terms in a linear combination where the network weights are estimated and adapted in relationship to the nonlinear basis functions (Bishop, 2006). Weights in the output node are estimated in relationship to the transformed and nonlinear version of the inputs where the weights in the hidden nodes are estimated in relationship to the original version of the inputs.

A very nice representation of the similarities and differences between a binary logistic regression model and a multilayer perceptron feed-forward neural network can be seen in the work of Beck, King, and Zeng (2000, 2004). The probability of class membership or π_i (for the class represented by Y_i = 1) is explained by the binary logistic regression model as a combination of linear functions that are then transformed by a logistic link function as follows:

$$Y_i \sim Bernoulli(\pi_i),$$

$$\pi_i = X_i\beta = \beta_0 + \beta_1 X_{1_i} + \cdots + \beta_k X_{k_i},$$

$$\pi_i = logit(X_i\beta) = \frac{1}{1 + e^{-X_i\beta}},$$

$$\pi_i = logit(linear(X_i)),$$

where the vector of regression coefficients β contains a constant term and *k* regression coefficients that are multiplied by each of the *k* explanatory variables, and linear$(X_i) = X_i\beta$. The probability of class membership is modeled as a logit function of a linear function of X_i (Beck, King, & Zeng, 2000, p. 24), producing an S-shaped curve that summarizes the relationship between the probability of class membership and the selected predictor variables.

In contrast, each hidden node within a neural network produces its own S-shaped curve that approximates the relationship between the probability of class membership and its predictors by weighting and summing the inputs differently. This process of extracting features from among various weighted combinations of the inputs followed by a sigmoidal activation function is what allows the researcher to approximate a potentially more complex target function from every relevant point of curvature to include testing for the presence of innumerable interaction effects. As noted by Beck, King, and Zeng (2000, 2004), the neural network model described below has the same distributional assumptions as the binary logistic regression model as described above, but assumes a more complex form that includes the creation of a new set of explanatory variables as linear combinations of the original inputs that are transformed by a nonlinear activation function. For a feed-forward neural network with *M* hidden nodes, the functional form of the relationship between the probability of class membership and its predictors is expressed (see Beck, King, & Zeng, 2000, p. 25) as follows:

$$Y_i \sim Bernoulli(\pi_i),$$

$$\pi_i = logit\left[\gamma_0 + \gamma_1 logit\left(X_i B_{(1)}\right) + \gamma_2 logit\left(X_i B_{(2)}\right) + \cdots + \gamma_M logit\left(X_i B_{(M)}\right)\right],$$

where a logistic activation function is used in all *M* hidden nodes and the output node, γ_0 = the bias weight for the output node, γ_M = the connection weight between the M^{th} hidden node and the output node, and $logit\left(X_i B_{(M)}\right)$ = the nonlinear term created by the M^{th} hidden node that is then forward propagated to the output node to be used as a predictor in the binary logistic regression that is executed in the output node. For a feed-forward network with one hidden node, the functional form of the relationship between the probability of class membership and its predictors is expressed as (Beck, King, & Zeng, 2000, p. 25)

$$\pi_i = logit(linear(logit(linear(X_i)))).$$

Figures 3.9 and 3.10 (see below) each provide a three-dimensional surface plot that models the probability of maltreatment recurrence (where the binary outcome = 1) in relationship to the number of income maintenance spells received before a second maltreatment report and the child's age at the first maltreatment report holding all other variables constant (the models produced for both figures are based on the set of predictors used for the neural network analysis presented in the results section, but for the sake of simplicity, the figures will be discussed in relationship to the three variables represented across all axes). However, the surface plot in Figure 3.9 was derived from a binary logistic regression model and the surface plot in Figure 3.10 was derived from a neural network. As discussed previously, a neural network includes a layer example of a surface plot that would be obtained from one of the hidden nodes in a neural network. In fact, a neural network with a simplified topology can be specified to produce a binary logistic regression; special cases include (a) the absence of a hidden node layer where the

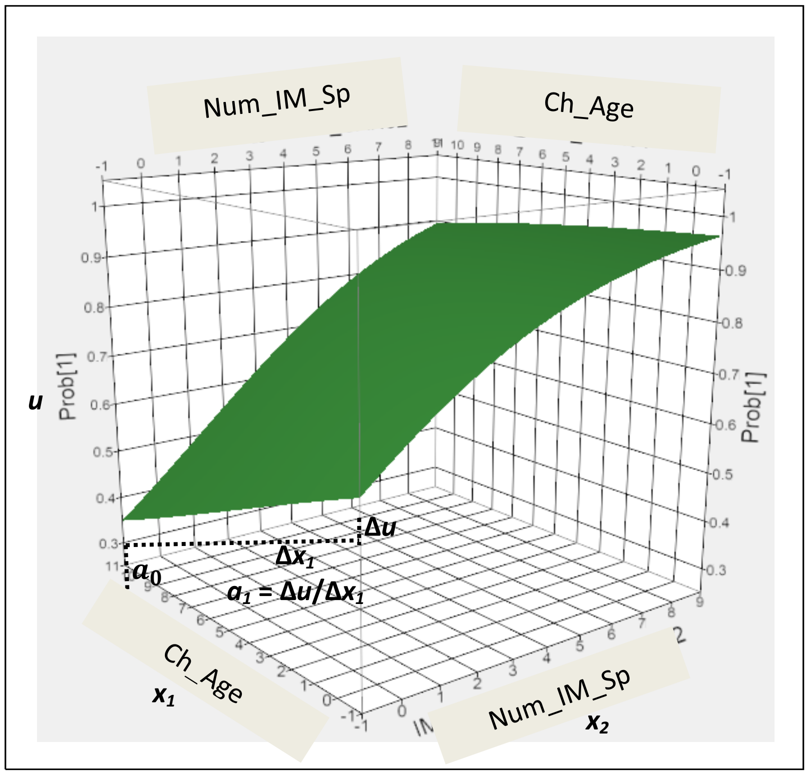


Figure 3.9. A surface plot depicting the probability of maltreatment recurrence as a function of child age at the first maltreatment report, the number of income maintenance spells prior to the second maltreatment report, and the estimated network weights. The function is linear in the network weights and linear in X_i.

inputs are forward propagated directly to the output node for weighting, summation and activation; and (b) the inclusion of one hidden node with a linear (e.g., identity) activation function (Beck, King, & Zeng, 2004; Paik, 2000; Zeng, 1999). Additionally, if the connection weights leading to a hidden node are very small, the linear combination produced by the summation function will fall on the linear portion of the sigmoidal curve, and the lack of curvature following the activation function will persist if the connection weights leading from the hidden node to the output node are very small as well (Bishop, 1995).

Figure 3.9 provides an excellent example of a local decision boundary produced by one hidden node; with the exception of the slight S-shaped curve, the surface is largely flat. For the purpose of illustration, let u = the probability of recurrence where $u = g_{logistic}(a_0 + a_1x_1 + a_2x_2)$, and a_0 is equivalent to the intercept, a_1 is the weight (i.e., the beta coefficient) for predictor x_1 (child age at the first maltreatment report), and a_2 is the weight for predictor x_2 (the number of income maintenance spells prior to the second maltreatment report). The surface in Figure 3.9 represents the set of all points that satisfy the equation $u = g_{logistic}(a_0 + a_1x_1 + a_2x_2)$, given particular values for a_0, a_1, and a_2. As noted by Smith (1993), a_0 functions as an intercept and determines the point at which the surface intersects the u axis measuring the probability of maltreatment recurrence. Weight a_1 determines the direction and slope of the surface that runs along the x_1 axis measuring child age at the first maltreatment report while holding the number of income maintenance spells constant. Hence, a_1 is the ratio of change in the probability of maltreatment recurrence corresponding to a change in child age at the first maltreatment report while holding the number of income maintenance spells constant (Smith, 1993). Similarly, weight a_2 determines the slope of the surface that runs along the x_2 axis measuring the number of income maintenance spells while holding child age at the first maltreatment report constant. Ultimately, each weight determines the shape of the surface's curve in one dimension (Haykin, 1999; Smith, 1993), and all functions are linear in the weights and in *x* (Bishop, 1995, 2006; Gujarati, 2003; Hastie, Tibshirani, & Friedman, 2001).

In contrast, Figure 3.10 provides a visual depiction of what happens when local regression analyses and multiple S-shaped curves are combined into a more global analysis that accounts for the possibility that nonlinearity in *x* exists. The degree of nonlinearity and therefore complexity in the function the neural network estimates to represent the relationship between the probability of maltreatment, child age at the

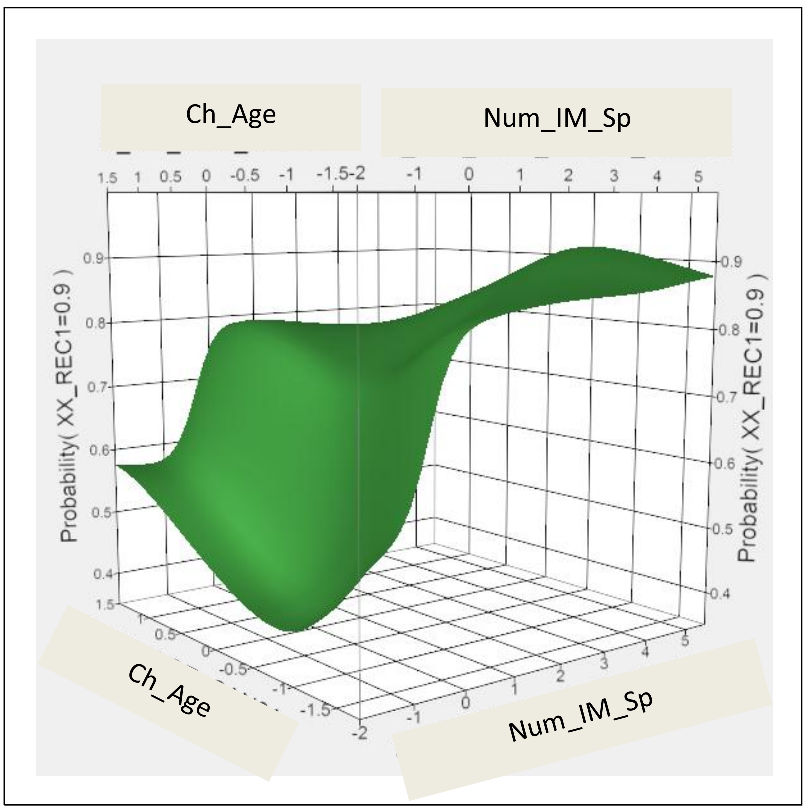


Figure 3.10. A surface plot depicting the probability of maltreatment recurrence as a function of child age at the first maltreatment report, the number of income maintenance spells prior to the second maltreatment report, and the estimated network weights. The function is linear in the network weights and nonlinear in X_i.

first maltreatment report, and the number of income maintenance spells before the second maltreatment report is determined by the following: (1) the form of the underlying data generating process, and (2) the degree to which the neural network can approximate the underlying mechanism. In the section that follows, the neural network's ability to model a complex target function is discussed in relationship to two key concepts: bias and variance. Ultimately, one of the key questions surrounding function approximation is as follows: Is the model complexity a true representation of the input-out mapping process or is it a statistical sleight of hand produced by the hidden nodes?

Generalization of the Target Function: A Balance Between Bias and Variance

As noted above, the estimated target function is a classification tool that is used to predict the values of the outcome variable based on values of the input variables and network weights. The target function is estimated in relationship to a data set that is referred to as the training data set; however, the utility of the estimated target function lies in its ability to generalize a high level of predictive accuracy to data sets that have the same underlying data generating mechanism as the training data set but have different values for the input variables (Bishop, 1995, 2006; Garson, 2998; Haykin, 1999; Mitchell, 1997; Smith, 1993). The difficulty in estimating the target function and in optimizing its predictive accuracy occurs in relationship to two conflicting sources of error: bias and variance. In both cases, the issue revolves around increasing the agreement between the systematic process that generates the observed data and the function that has been created in order to model the underlying process. An estimated target function is said to be biased if the model is too simplistic and therefore lacks the necessary amount of curvature; in this case, the modeled target function has been under-fitted to the data (Bishop, 1995, 2006). On the other hand, an estimated target function suffers from too much variance if the model's predictive performance decreases when applied to new cases because the estimated function is too complex and therefore fitted too closely (i.e., has been over-fitted) to the training data. Optimizing the generalizability of the modeled target function involves reaching a compromise that simultaneously minimizes bias and variance (Bishop, 1995, 2006).

Figure 3.11 (see below) provides a good example of a modeled target function that is characterized by a high level of bias as the estimated data-generating model clearly lacks the curvature that characterizes the true but unknown underlying data generating process. The lack of curvature produced by the estimated target function is due to a lack of flexibility in the neural network's estimation process -- an estimation process that has been made inflexible by a deficient number of input nodes, hidden nodes, and/or connection weights as well as weights that are deficient in strength or magnitude. In contrast, Figure 3.12 provides a good example of a modeled target function that is characterized by a high level of variance as the estimated data-generating model has been so closely fit to the training data that new cases with different values will be unlikely to fall within the elaborate and highly specialized decision regions that result from such a complex classifier. The estimated data-generating model relies too heavily on the specific combination of input values from a particular data set with which to classify outcome values for cases with a new set of data points. Bishop (1995) notes that generalization is optimized in models with an intermediate level of flexibility and that the management of flexibility can be achieved by applying techniques that (a) alter the network's topology, (b) regularize the effective complexity of the model by reducing the magnitude of the network weights, and (c) minimize the error function in relationship to the validation data set as opposed to the training data set. Generalization was optimized

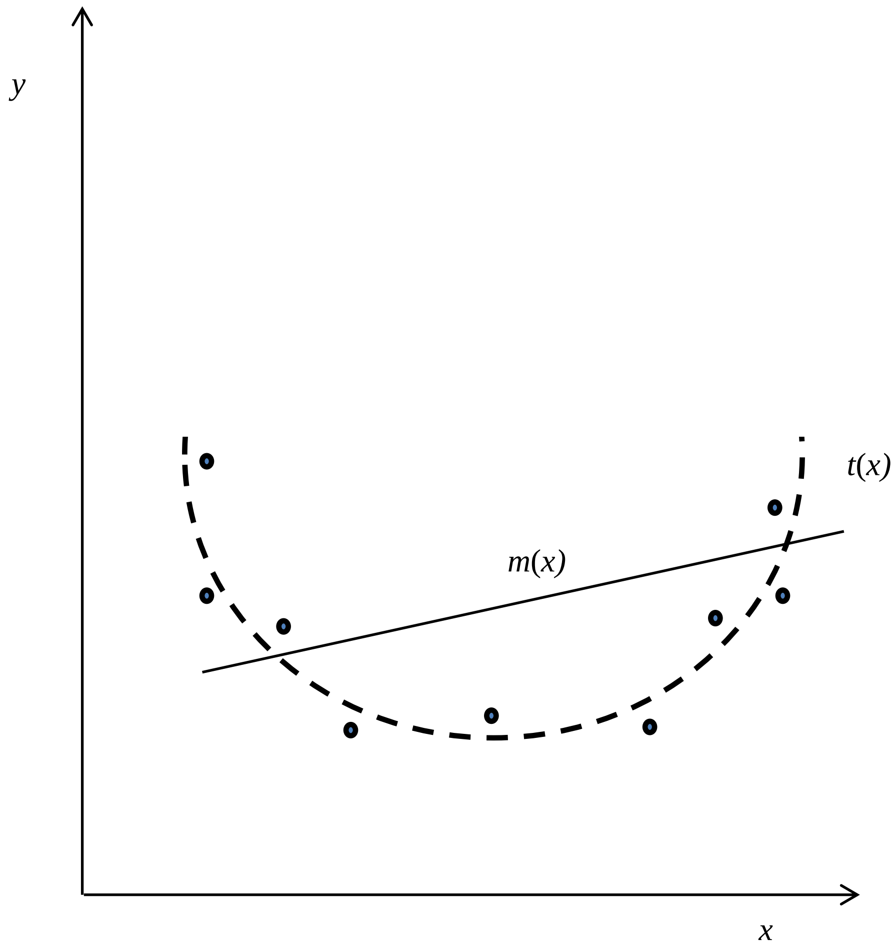


Figure 3.11. A comparison of a model-generated target function, *m*(*x*) (represented by the solid straight line) with a known target function, *t*(*x*) (represented by the curve with dashes). The model-generated target function is biased in its lack of complexity and is therefore a poor representation of the known target function. Adapted from *Neural Networks for Pattern Recognition*, by C. M. Bishop, 1995, p. 336.

for the estimated data-generating model in this dissertation analysis by applying each of the suggested techniques.

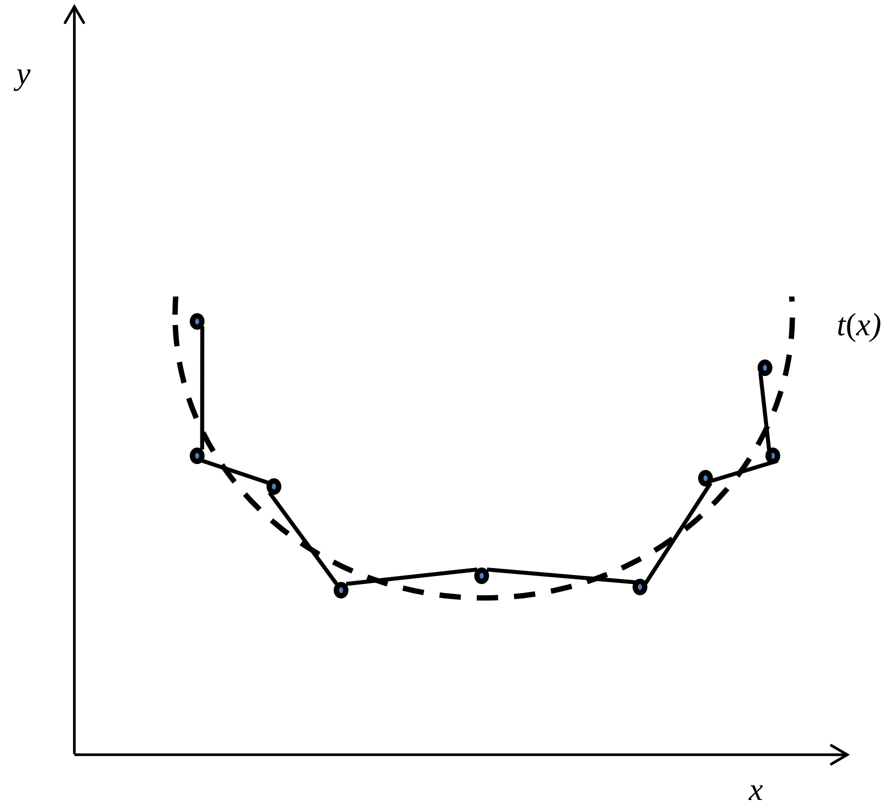


Figure 3.12. A comparison of a model-generated target function (represented by the curve without dashes) with a known target function (represented by the curve with dashes). The model-generated target function has poor generalizability and high variance because it was over-fitted to the data. The known target function is actually a much smoother mapping of the input values onto the output values. Adapted from *Neural Networks for Pattern Recognition*, by C. M. Bishop, 1995, p. 336.

Controlling model complexity through network topology.

The number of hidden nodes was limited to one layer as opposed to two layers and the final number of hidden nodes was selected by minimizing the negative log-likelihood and the misclassification rate for the validation set while simultaneously maximizing the Receiver Operating Characteristic (ROC) curve of the validation data set. The final number of hidden nodes (i.e., eight) fell within the range suggested by Hastie, Tibshirani, and Friedman (2001) as generally being adequate in representing the underlying data-generating mechanism. The selection of one layer of hidden nodes also concurred with heuristic suggestions provided by Hastie, Tibshirani, and Friedman as well as Garson (1998), wherein more than one layer of hidden nodes is only used if supported by discipline-specific theoretical assumptions about the relationship being modeled. Adding a second layer of hidden nodes expands the neural network's ability to capture underlying nonlinear effects by including nonlinear functions of the weights in addition to nonlinear functions of the inputs (Bishop, 1995, 2006).

Controlling model complexity through regularization.

The effective complexity of the model was regularized by including a weight decay parameter to the error function; hence, *penalized* maximum likelihood was used to estimate the network weights in conjunction with estimating a function to represent the underlying data-generating mechanism (Bishop, 1995, 2006; Garson, 1998; Hastie, Tibshirani, & Friedman, 2001; Haykin, 1999; Mitchell, 1997; Smith, 1993). Generally speaking, regularization is used to shrink the values of the weights towards zero, thereby restricting the range of hidden node outputs to the linear portion of the sigmoidal curve; hence, the complexity of the model is restrained and the input functions are more likely to be linear as opposed to nonlinear (Bishop, 1995, 2006; Hastie, Tibshirani, & Friedman, 2001). Regularization works by adding a weight decay term to the error function along with a coefficient that increases or decreases the extent to which the weight decay regularizer shrinks the values of the network weights towards a linear functional form. Hence, the penalized error function is represented by

$$\tilde{E} = E + v\Omega,$$

where $\tilde{E}$ is the penalized error function, E is the cross-entropy error function, v is the weight decay coefficient, and Ω is the weight decay term (Bishop, 1995, p. 338). The specific form of the weight decay term used in the neural platform of JMP Pro 9 is

$$\frac{\beta^2}{1+\beta^2}$$

where β = the vector of network weights (Gotwalt, 2011).

As noted by Bishop (1995), a regularizer that sums the squares of all connection and bias network weights is analogous to the application of ridge regression, which is a method used to adjust linear regression parameters for over-fitting by shrinking the regression coefficients as they are being estimated. Penalized maximum likelihood extends the concept of ridge regression by minimizing the negative log penalized likelihood for the vector of regression coefficients. Penalized regression coefficients depend less on the specific values of the data used to estimate the coefficients and more on the underlying process the coefficients are modeling; hence, the predictive accuracy of the estimated model should be less variable across new cases (Moons, Donders, Steyerberg, & Harrell, 2004).

In addition to estimating the best value for the weight decay coefficient (*v*), the optimization of generalizability through regularization is substantively influenced by the coding of the input values. In short, the scale upon which the input values are measured influences the selection of the initial starting values for the network weights as well as the magnitude of the weights, which in turn influence the process of regularization and ultimately the modeled function's representation of the true underlying function (Bishop, 1995; Garson, 1998; Hastie, Tibshirani, & Friedman, 2001). Hence, *z*-scores were created for all continuous predictors, categorical predictors were recoded so that their values fell slightly inside the range of the activation function used for each hidden node, and output values were recoded so that these values fell slightly inside the range of the activation function used for the output node (Garson, 1998; Haykin, 1999; Mitchell, 1997; Smith, 1993).

Standardizing the scale upon which continuous predictors are measured transforms the original input values by making them more alike; this in turn allows the values for the weights to be more alike as opposed to markedly dissimilar in order to accommodate substantial differences in the values for inputs measured on markedly divergent scales (Bishop, 1995; Garson, 1998). The ability to estimate values for the weights that are similar as opposed to markedly dissimilar depending on the values of the inputs also extends to the process of initially assigning each network weight a random starting value in order to begin the search for a vector of weights that minimizes the error function. The ability to estimate similar weight values facilitates the random assignment of initial weight values from the same distribution (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001). Moreover, if the input values are measured on a standard scale, the inputs can be treated equally during the regularization process (Hastie, Tibshirani, & Friedman, 2001).

Recoding values for categorical predictors to fall slightly within the bounds of the tanh function involves recoding a value of “1” to indicate the presence of some characteristic (e.g., the utilization of a mental health service) into a value of 0.9 and the recoding of a value of “0” to indicate the absence of some characteristic (e.g., the absence of mental health service utilization) into a value of -0.9. In both cases, the recoded values fall within the range of values for the tanh function, which are -1 and 1. According to Garson (1998), in order for the weight decay parameter to work properly, the values of the inputs that are forward propagated (referred to as the “signal” transmitted by each input node) must fall within the range of the values that would be produced by the tanh function. However, values of -1 and 1, just like the values of 0 and 1 for the logistic function, are never actually reached. Hence, it is appropriate to recode values of the categorical variables (both predictor and response variables) so that their signals will fall slightly within the range of values that are produced by the respective activation function (Garson, 1998; Haykin, 1999; Mitchell, 1997; Smith, 1993).

Figures 3.13 and 3.14 depict the application of a weight decay regularizer to the learning process, where the values of the network weights are (a) adjusted to decrease the error function, and (b) adjusted to account for over-fitting. The first function adjusts the weights to minimize error, thereby increasing model complexity and decreasing bias; at the same time, the second function adjusts the weights to account for over-fitting, thereby decreasing model complexity and decreasing model variance. Finding the right balance between these competing functions is critical and the estimation of the value used for the weight decay coefficient influences the degree to which each component of the penalized error term is minimized,

$$\tilde{E} = E + \nu\Omega,$$

where a more complex and curved function that fits the data closely minimizes E, and a less curved, more linear function that does not fit the data as closely minimizes Ω (Bishop, 1995). Figure 3.13 compares two functions: (1) the neural network-generated target function

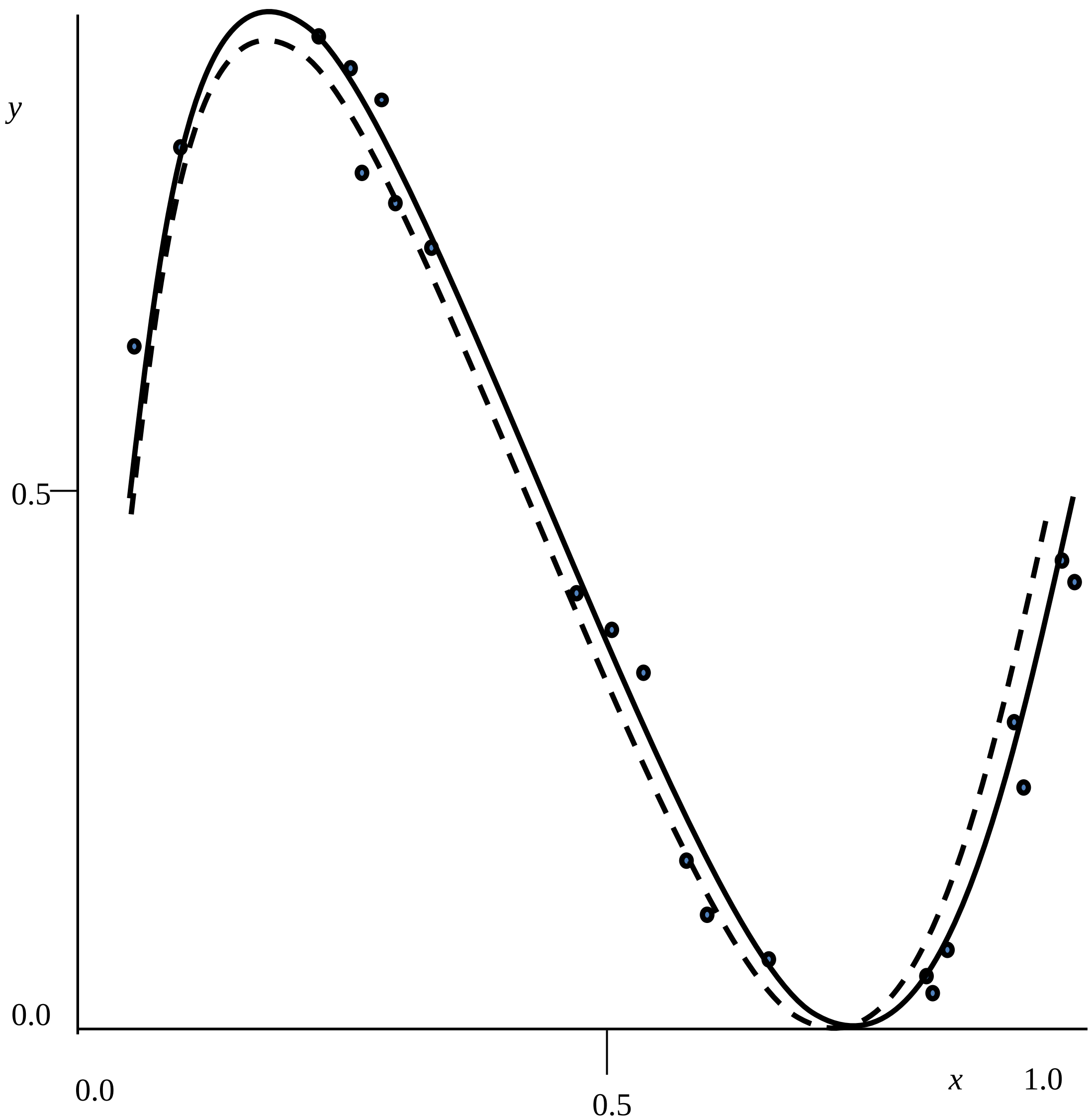


Figure 3.13. A comparison of a model-generated target function (represented by the curve without dashes) with a known target function (represented by the curve with dashes). The weight decay coefficient (v = 35) achieved a satisfactory balance between (a) the desire to minimize the error function in order to decrease bias, and (b) the desire to minimize the penalty term that monitors model complexity. Adapted from *Neural Networks for Pattern Recognition*, by C. M. Bishop, 1995, p. 344.

(represented by the solid continuous curve), and (2) the true data-generating function (represented by the dashed curve). The fit between the two functions is very close as

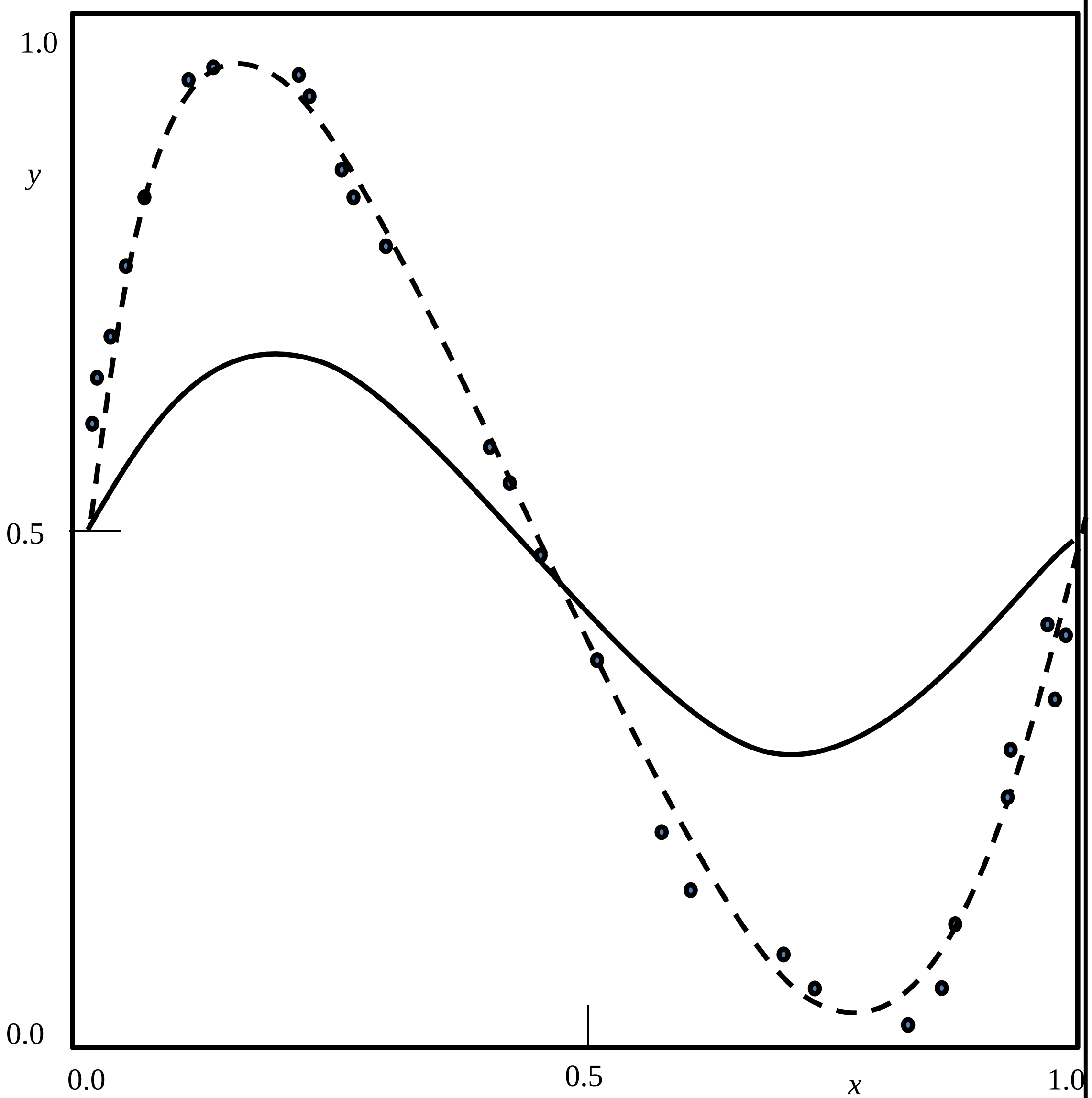


Figure 3.14. A comparison of a model-generated target function (represented by the curve without dashes) with a known target function (represented by the curve with dashes). The weight decay coefficient (v = 1100) minimized the penalty term that monitors model complexity at the expense of the error function. The result is a model-generated target function that is biased in its lack of complexity. Adapted from *Neural Networks for Pattern Recognition*, by C. M. Bishop, 1995, p. 344.

evidenced by the overlapping of the two curves. The high level of agreement between the two functions has been facilitated by the selection of a weight decay coefficient (v = 40) that balances the minimization of E with the minimization of Ω (Bishop, 1995). In contrast, Figure 3.14 shows a poor level of agreement between the two functions because the value selected for the weight decay coefficient (v = 1000) was too large and therefore constrained the effective complexity of the function to the point of creating a biased function. Hence, Ω was minimized at the expense of E.

JMP Pro 9 simultaneously estimates the value of the weight decay coefficient while iteratively searching the error surface in order to minimize the penalized error function (i.e., the negative log-likelihood plus the weight decay penalty). Hence, the quasi-Newton BFGS learning algorithm has an inner loop that estimates the value for the weight decay parameter and an outer loop that maximizes the penalized error function (Gotwalt, 2011). At the beginning of the learning process, randomly selected values from the normal distribution are assigned as starting values for the network weights and the penalty term is set to zero. As the error surface is searched, values for the network weights and penalty parameter are iteratively updated, and candidate values for the penalty and network weight parameters are compared according to the degree to which the parameter estimates minimize the negative log-likelihood for the data (i.e., the values that make the observed data most likely) (Gotwalt, 2011). As described in the forthcoming section on cross-validation and early stopping, the values for the network weights and penalty parameter are selected in relationship to the minimization of the error function in the validation data set (Bishop, 1995; Gotwalt, 2011).

Finally, JMP Pro 9 provides the opportunity to estimate the target function multiple times by beginning from a different set of randomly assigned starting values for the network weights where the final set of parameters that are reported are those that make the observed data most likely and therefore minimize the penalized error function to the greatest degree (SAS, 2010). Each set of starting values for the weights influences the iterative search of an error surface that typically contains many different local minima. As noted earlier, the final form of the estimated target function is determined in concert with the selection of weight values that minimize the negative log-likelihood function. For each local minimum, a unique vector of network weights will be selected in conjunction with the minimization of the error surface at that given point (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001; Mitchell, 1997).

Controlling model complexity through cross-validation and early stopping.

The effective complexity of the model is not only influenced by the number and magnitude of the network weights, but it is also influenced by the extent to which the network learns from the mapping of the input values onto the output values by iteratively readjusting the weights to minimize the error function. Hence, the amount of training time also plays a key role in determining the level of complexity a modeled target function will achieve (Bishop, 1995; Garson, 1998; Hastie, Tibshirani, & Friedman, 2001; Haykin, 1999; Mitchell, 1997; Smith, 1993). The greater the amount of effort that is applied to minimizing the error function relative to the training data set, the greater the model complexity will be as it fits the modeled function to the systematic aspects of the underlying data-generating process *and the idiosyncratic characteristics of the training data* (i.e., "noise"). For example, if *y* is a function of x_1, x_2, β, and ε with a linear functional form, then

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

represents the function that generates observed values of *y* for a given training data set. The goal in using a neural network analysis is to model the systematic aspects of the input-output mapping process that can be reliably applied across many data sets with the same systematic aspects discovered in the original training data set. Hence, the goal is to model y as a function of $\beta_0 + \beta_1 x_1 + \beta_2 x_2$, but not ε. Iteratively updating network weights in order to drive the value for the error function down causes the modeled function to iteratively adjust the representation of the underlying mapping process by continually taking into account the effect of ε.

In addition to adding a penalty term to the error function in order to shrink the network weights and thus reduce model complexity, cross-validation and early stopping can simultaneously be used to determine model complexity in relationship to error in the validation data set as opposed to the training set (Bishop, 1995). Moreover, k-fold cross-validation can be used to gain the best of both worlds by training the neural network on the entire sample of cases (so as to have access to the full complement of examples of patterns of inputs with corresponding outputs) and by fine-tuning model complexity in relationship to *k* equally divided sections of the original data set (so as to have access to validation sets). Specifically, instead of proportionally dividing a given data set into one training set and one validation set, the original sample is divided into *k* equal sections, where *k* is typically equal to 10 (Bishop, 1995; Hastie Tibshirani, & Friedman, 2001). A separate neural network is then built 10 different times in relationship to *k*-1 sections that constitute the training set and the remaining section that constitutes the validation set for that neural network. For each successive neural network that is created, a different section is held out to serve as the validation set. Figure 3.15 (see below) provides a visual depiction of this process (Haykin, 1999).

The benefit of fine-tuning model complexity in relationship to the minimization of the error function for a validation set as opposed to a training set is as follows. Repeated iterations of the learning algorithm are what minimize the error function while maximizing model complexity. With each set of weight changes, the algorithm learns with greater specificity how to model the observed target values as a function of the observed input values. With extended training, the error will monotonically decrease, producing an estimated target function that is a poor representation of the systematic aspects of the true function (Bishop, 1995; Mitchell, 1997). In contrast, repeated weight changes do not continuously decrease error for the validation set because the validation set is comprised of new cases that test the modeled function's generalizability. Hence,

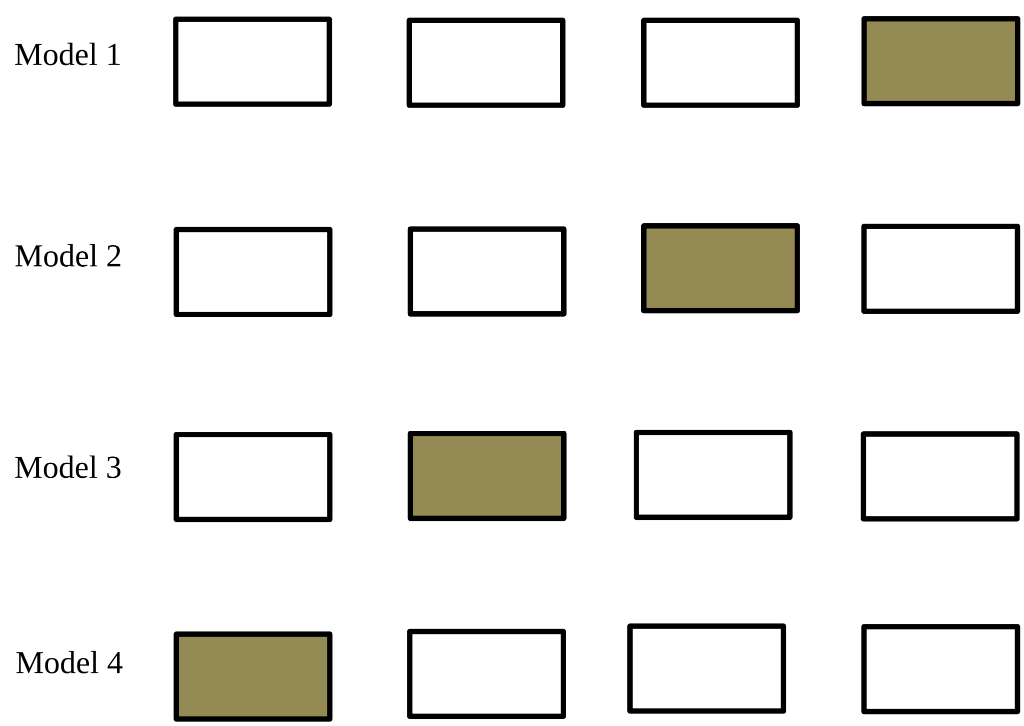


Figure 3.15. Depiction of the k-fold cross-validation method. Adapted from *Neural Networks: A Comprehensive Foundation*, by S. Haykin, 1999, p. 218.

stopping training when the error function has reached its lowest point in relationship to the validation set decreases model complexity and optimizes generalizability (Bishop, 1995; Mitchell, 1997). Even though the error function could be minimized further for the training set, and even though the learning algorithm could continually refine the weights in order to capture the underlying aspects of the data, stopping training at the lowest point of error for the validation set optimizes generalization by limiting the model to a functional form that is based on the systematic aspects of a data-generating process (Bishop, 1995).

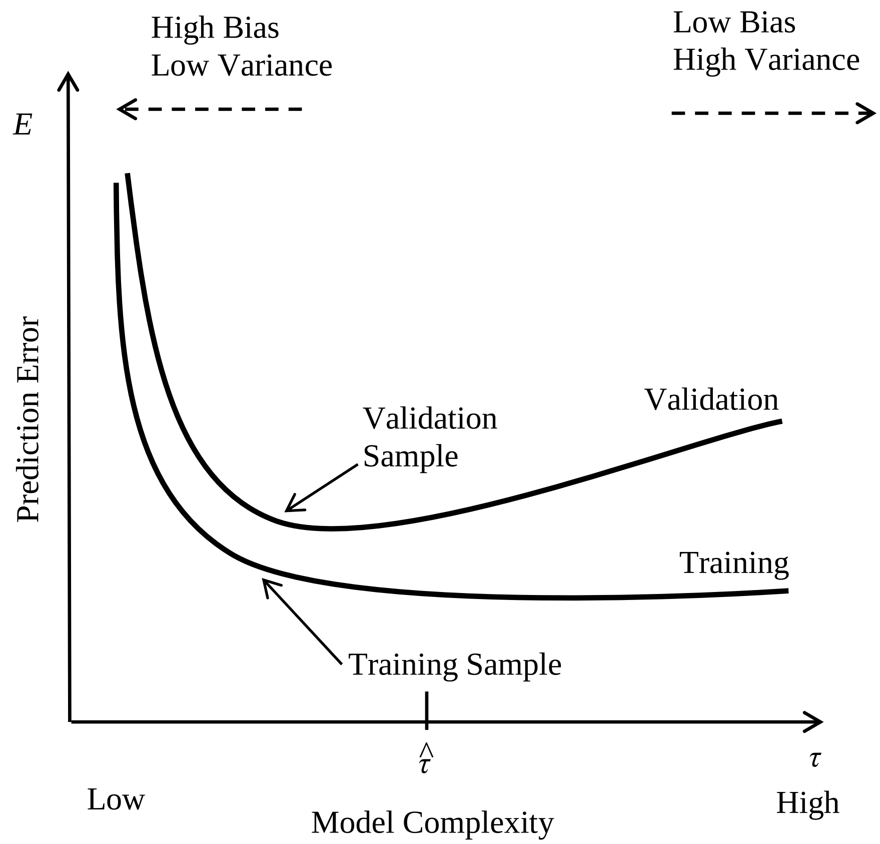


Figure 3.16. Depiction of the minimization of the error function in relationship to the training data and validation data. Minimizing error in relationship to the validation sample optimizes generalizability and model complexity. Adapted from *Neural Networks for Pattern Recognition*, by Christopher M. Bishop, 1995, p. 336, and *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by T. Hastie, R. Tibshirani, and J. Friedman, 2001, p. 194.

Figure 3.16 (above) depicts the utility of applying the early stopping technique and the different trajectories of the error function in relationship to the training and validation sets. The lowest point of error for the validation set corresponds with the point at which training iterations should be stopped ($\hat{\tau}$); however, training (τ) could be continued until the global minimum of the error function is reached for the training data (i.e., the point of convergence) (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001). Continuing training to the point of convergence for the training data set decreases the modeled target function's generalizability as the model complexity increases and the model variance increases (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001). That said, stopping training too early can compromise the learning process and produce a modeled target function that is biased in its lack of complexity (Mitchell, 1997).

JMP Pro 9 uses the validation set to optimize generalizability by determining model complexity in relationship to the monitoring of the error function for the validation sample. Specifically, the software does not pursue convergence for the training data but instead selects the vector of values for the network weights and the penalty parameter in conjunction with the minimization of the negative log penalized likelihood function for the validation sample. The quasi-Newton BFGS learning algorithm is terminated when the negative log penalized likelihood function is no longer decreasing in relationship to the validation data (Gotwalt, 2011). When using k-fold cross validation, the retained vector of values for the network weights and penalty parameter are those values for which the reduction in the error function was largest (and consequently the parameter estimates for which the observed data have the highest probability of occurring) (Gotwalt, 2011).

Adding Inputs to the Neural Network: Pre-Processing Matters

In addition to the specification of the neural network's topology and architecture, the pre-processing of the inputs has a substantial influence on the quality of the model-generated target function. In fact, Bishop (1995) notes that "for practical applications, data pre-processing is often one of the most important stages in the development of a solution, and the choice of pre-processing steps can often have a significant effect on generalization performance" (p. 296). Pre-processing focuses on maximizing the signal-to-noise ratio in order to assist the neural network in learning how to separate and predict values for the outputs as a function of the inputs. Similar to the challenge of finding the right balance of model complexity in order to optimize the estimated target function's generalizability, a challenge during the pre-processing stage is working to find the right balance between the incorporation of information content of the input data and the need to reduce the dimensionality of the input data (Bishop, 1995).

Classification accuracy is a function of the input data and therefore requires that researchers include the inputs that are most capable of separating the values of the outputs in distinct decision regions. Moreover, neural network analyses can be very effective in locating underlying structure, such as interactions and higher order polynomial terms that would otherwise have been missed by the researcher. However, as the number of inputs increases, so too does the dimensionality of the data and what Bishop (1995) and Haykin (1999) describe as the curse of dimensionality. In short, the number of dimensions that constitute the input space increases in connection with the number of inputs as well as the range of values that each input can take. Classification of values for the outputs as a function of the inputs begins with the process of assigning each case record or observation (also referred to as each pattern) to a point in input space. This point represents the location of the case record in relationship to the values taken by each of the inputs; moreover, values for the outputs are located as a function of the placement of the associated inputs. The ability to develop a prediction rule that explains how output values are most accurately classified as a function of input values depends on the distribution of data points throughout input space (Bishop, 1995; Haykin, 1999). As the number of dimensions increases in the input space, it may become increasingly difficult to learn the input-to-output mapping function, especially if the data points do not adequately fill the many dimensions contained in the input space.

The development of an accurate classification scheme depends upon the neural network's ability to model the systematic aspects of the data, which requires a "dense" (Haykin, 1999) distribution of data points across the dimensions (p. 212). A relatively large number of data points is key because each data point provides the neural network with an example of how a particular group of inputs are associated with a particular output value. As noted by Garson (1998), "each example [data point] can be used by the model to reinforce a different input-output relationship principle," where each example of an input-output relationship makes "some conceptual point to the network" (p.89). Moreover, diversity among the input-output examples is important for the neural network to be able to learn the underlying function well; thus, the sample of data points should include input-output patterns with values that fall along the full continuum of the possible values for *x* and *y* (Garson, 1998). Finally, modeling the target function accurately in high-dimensional input space is more difficult due to the greater level of complexity that is likely to characterize such an underlying function (Haykin, 1999). All things considered, Bishop (1995) describes the importance of including information content in the inputs that can classify outcome values; however, there is a limit as to the return on predictive accuracy one can expect with the inclusion of each additional input variable or feature. Specifically, the author notes that "beyond a certain point, adding new features [inputs] can actually lead to a *reduction* in the performance of the classification system" (Bishop, 1995, p. 7).

Several steps were taken in order to expand the information content of the input variables while reducing the dimensionality of input space to include only those variables or pre-processed features that would be instrumental in helping the neural network to learn the underlying classification process as a function of the input values and the network weights (which is really the conditional probabilities of each class of outcome values given values for x and θ). First, information content of the input variables was expanded through the creation of a myriad of new variables based upon the information originally provided; specific details are provided (please see the forthcoming new section on the following page) regarding the original sample of merged administrative records tracking each child who was reported at least once for maltreatment and his/her primary caregiver across multiple public sector service systems. Second, principle components analysis was used to reduce three different sets of correlated predictors -- i.e., a set of correlated census tract variables measuring different aspects of poverty, a set of correlated worker-observed family characteristics measuring different protective factors/strengths, and a set of correlated worker-observed perpetrator characteristics measuring different protective factors/strengths -- into three different principal components (where each set of original predictors was reduced to one principal component). Third, a classification and regression trees (CART) analysis and specifically, a random forest analysis, was conducted in order to identify and extract the explanatory variables that were most instrumental in predicting which cases would be re-reported for maltreatment and which cases would not be re-reported for maltreatment.

Expanding information content for the inputs: A description of the data set.

Data for this dissertation study were originally collected across three federally-funded parent studies by principal investigators Dr. Melissa Jonson-Reid and Dr. Brett Drake to include (a) Child Neglect: Cross Sector Service Paths and Outcomes (Part 1) (NIMH[MH6173302]), (b) Child Neglect: Cross Sector Service Paths and Outcomes (Part 2) (NIMH[MH06173304A1]), and (c) Young Adult Violence: Modifiable Predictors and Paths (CDC#R01CE001190). The data were collected for the purpose of studying the maltreatment-related trajectories of children in relationship to their involvement in public sector service systems, where system involvement can be operationalized as measuring the relative presence or absence of dynamic risk factors that are (a) modifiable, and (b) associated with maltreatment recidivism. The full data set includes 12,409 children from a large mid-western metropolitan area whose original criteria for sample selection included the following: (a) the child had a first-known maltreatment report falling between January 1, 1993 and December 31, 1994, and/or; (b) the child had a concurrent or a previous AFDC/TANF record that was tracked as far back as 1990.

At the time of initial sampling, the statewide child welfare administrative records for children with a first-known maltreatment report in 1993 or 1994 were linked to statewide AFDC administrative records (which later became TANF in 1996) for the purpose of identifying two groups of children. The first group was comprised of children with a first-known report for maltreatment and a record of concurrent or past receipt of income maintenance. The second group was comprised of children with a first-known report for maltreatment and no record of income maintenance receipt. A third group was then created when children with a record of income maintenance receipt but no record of maltreatment were matched by city/county residence and age to the children who had been reported for maltreatment and who had a record of AFDC involvement (please note that the children in this group could have been reported for maltreatment after December 31, 1994). A subset of AFDC-only children from each residence and age-based strata were randomly selected for inclusion in the study. To ensure the independence of observations, one child was randomly drawn from each family. Child welfare and AFDC/TANF administrative records for each child were collected through 2009 (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Jonson-Reid, Drake, & Kohl, 2009).

Child welfare records provided an array of child-, primary caregiver-, event-, and post-investigation service-level information, to include (a) *child* gender, race/ethnicity, and date of birth; (b) *primary caregiver* gender, date of birth, educational status, and history of foster care placement; (c) *event-based* type of maltreatment reported, date of report, case substantiation status, and relationship of the perpetrator to the child; and (d) *post-investigation service delivery* commencement dates for up to four spells of family centered services (i.e., less intensive in-home case management, counseling, and referrals), family preservation services (i.e., short-term, very intensive in-home services), and foster care placement (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). Child welfare records also provided worker observations of up to 26 family characteristics and 21 perpetrator characteristics. These characteristics were not measured with a standardized clinical instrument but represent the worker's assessment of those risk and/or protective factors that were so pronounced in relationship to the investigation of the maltreatment report that they warranted notation. In short, these observations represent the kind of "in the field" assessment and decision-making activities that child welfare workers are required to do on a daily basis and therefore, these observations provide good examples of the patterns of inputs workers are likely to focus on during their interactions with the families reported for maltreatment. Income maintenance records provided additional information regarding the socio-economic context in which the children and their primary caregivers were living. AFDC/TANF records were used to obtain commencement dates for up to 12 income maintenance spells.

Birth record data were used to triangulate and supplement child welfare records in regards to providing information for (a) the child's date of birth, (b) the primary caregiver's date of birth, (c) the primary caregiver's age at the time of the child's birth, and (d) the primary caregiver's level of education. Death record data were used to create a censoring variable for the purpose of identifying and dropping cases where the child was under the age of 18 and died during the course of the study. The "birth problems" variable was created in consultation with a neonatologist who reviewed diagnoses (if present for a given case record) the child was given as a result of hospital-based care he/she received from birth through a follow-up period of 12 months (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010).

Administrative records from a wide range of public sector service systems were collected in order to obtain information about the types and timing of services that were received (e.g., mental health services) as well as the types and timing of mandated system involvement that occurred (e.g., the issuance of a juvenile court petition). The study of public sector service system involvement for each child and his/her primary caregiver extended to (a) statewide department of mental health Medicaid and non-Medicaid records of *inpatient and outpatient mental health and substance abuse service* use (for the child and primary caregiver), (b) metropolitan area *emergency department* records for admission related to *mental health treatment needs* (for the child), (c) metropolitan area *special education screening and eligibility* records (for the child), (d) metropolitan area juvenile court records of *petitions issued for status and delinquency offenses* (for the child), and (e) statewide Highway Patrol records of *criminal arrests* (for the primary caregiver) (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Jonson-Reid, Drake, & Kohl, 2009). A common state-level identifier was used to link administrative case records across a number of the systems providing data. In the event that a common state-level identifier was not used by a particular system providing data, the first four letters of the first and last name in addition to date of birth were used to link case records; moreover, if the match remained uncertain, additional identifying information such as gender or the middle initial were used to link case records. Match rates were hand checked (Drake, Jonson-Reid, & Sapokaite, 2006).

Information regarding community context was also obtained particularly as it related to dimensions of poverty. Specifically, 1990 U.S. census data for 261 unique census tracts was linked with families' addresses provided at the time of their sample inclusion; addresses were geocoded using Arcview and then linked to the census records (Drake, Jonson-Reid, & Sapokaite, 2006). The data provided for this dissertation study were de-identified and therefore contained no unique identifiers and/or no unique information that could be used to deductively determine a subject's identity. Additional steps were taken to prevent deductive identity disclosure to include (a) the application of random error to the data to prevent deductive disclosure, (b) the encryption of all data, (c) the exclusion of categorical variables with fewer than 40 observations in a given response category from random forest and neural network analyses; and (d) the safeguarding of all data on an external hard drive that remained in a locked filing cabinet when not in use. The data were obtained and subsequent analyses were conducted with Washington University Hilltop Institutional Review Board approval.

The "raw" data set obtained for this dissertation study included 12,409 observations and 944 variables. The great majority of the variables were dates for services or mandated system involvement. Child welfare worker observations regarding family and perpetrator characteristics were in multinomial form and contained categorical response levels with sparse observations. Dates for child and primary caregiver mental health and substance abuse service delivery were aggregated into diagnostic groupings, and multiple groupings contained sparse observations. While the "raw" ingredients for the final data set were present, a substantial amount of effort was spent in preparing the variables for analysis. All pre-processing activities were carried out using SAS 9.3.

Expanding information content for the inputs: Steps taken during pre-processing.

The first step taken during the pre-processing information expansion phase included the creation of censoring variables that were subsequently used to pare down the original 12,409 observations into a data set that contained 6,747 observations. Part of the process of expanding information content for the inputs was the creation of a sample of children and primary caregivers that conceivably shared the same underlying data-generating mechanism that could be modeled as a function of the inputs. Exclusion criteria were developed for the purpose of creating a risk pool where each child had a first-known maltreatment report that fell between January 1, 1993 and January 1, 2002 and the child was just under 11 years of age (i.e., no older than 10.99 years of age) at the time of the first maltreatment report. This allowed for a follow-up period of at least seven years from the time of the first maltreatment report; additionally, by capping the child's age to just under 11 years of age, no child will have reached the age of 18 before the study's end (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). If the child was placed in foster care on or after his/her first maltreatment report, the child had to have been returned home before a second maltreatment report (if such an event occurred). In order to be at risk of recurrent maltreatment, children needed to be in an environment where recurrence was a possibility. Following this same line of thought, children who were under the age of 18 years and who died during the course of the study were also excluded (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010).

In order to avoid mixing children with first-known reports of maltreatment that occurred during the timeframe specified above with children who were likely to have already experienced a true first-known report for maltreatment, observations were deleted if the child was documented as having received a post-investigation service (i.e., family centered services, family preservation services, or foster care placement) before the date of his/her first-known maltreatment report. Finally, exclusionary criteria were applied in order to (a) avoid counting a duplicate or echo of the first report as a separate and second event; and (b) mixing reports regarding neglect, physical abuse and sexual abuse with reports of other types of abuse to include emotional abuse, a fatality, or report types categorized as "other." Hence, observations (i.e., subjects) were dropped if the second report did not fall more than 14 days after the first report and if the report reason (for the first and/or second report) was anything other than neglect, physical abuse, and/or sexual abuse. This approach to censoring the data is based upon approaches utilized in other studies (see, e.g., Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Drake, & Kohl, 2009). That said, this study is unique in that the operational definition of general maltreatment for both the first and potentially second reports was based on including only those reports for which the allegations corresponding with reported actions or inactions met criteria for neglect, physical abuse, and/or sexual abuse.

The second step taken during the pre-processing information expansion phase was the creation of 24 service-related variables to identify either potential dynamic (i.e., modifiable) risk factors or service system contacts that could conceivably interact with dynamic risk factors. In either case, great care was taken to create variables that broadly fit within an RNR framework that links the reduction of risk with the identification of characteristics that are likely to provide information about potential treatment need. Dates of service delivery were compared with dates of maltreatment reports using conditional logic for the purpose of creating temporally-ordered indicators of service receipt (or mandated system involvement). Upon creating diagnostically-specific variables that measured the presence or absence of mental health or substance abuse service receipt relative to the timing of maltreatment reports, low variance was discovered. In order to avoid difficulties in pattern recognition that are related to sparse data points in a high-dimensional input space, conditional logic statements with inclusive "or" operators were used to code a child or a primary caregiver as having received a mental health/substance abuse service if they had a service date for any one or more of the diagnostic categories listed. Children were coded as having received a conflict-related service if a date for service delivery had occurred in relationship to a diagnostic code for (a) child abuse, or (b) assault or homicide. Children and primary caregivers were coded as having received a mental health/substance abuse service if a date for service delivery had occurred in relationship to a diagnostic code for (a) general mental health problems, (b) substance abuse problems, (c) mental health problems with psychosis, (d) personality disorders, (e) other mental health problems, or (f) mental delays.

The third step taken during the pre-processing information expansion phase was the creation of 26 dichotomous variables that each represented a response level for the multinomial variable measuring worker observations of family characteristics; similarly, 21 dichotomous variables were created to represent the response levels for the multinomial variable measuring worker observations of perpetrator characteristics. Low variance was discovered for a majority of the newly created dichotomous variables for both worker-observed family and perpetrator characteristics. While the use of conditional logic statements with inclusive “or” operators could reasonably be used to aggregate similar types of conflict-related and mental health/substance abuse related services together, such an approach did not appear to be reasonable in relationship to the diverse array of family and perpetrator characteristics. Hence, principal components analysis was used to extract one component to represent the maximum amount of variance for a selection of predictors within each subset of characteristics (please see the next section for more details).

Finally, several continuous-level inputs were created in order to expand (when relevant) the scale upon which inputs were typically measured given the preponderance of dichotomous inputs. As noted by Garson (1998), the neural network’s ability to learn the input-output mapping function depends in large part upon the diversity of examples presented to the network to include patterns of inputs with values for X that span the full range of *x*. Hence, in addition to creating a dichotomous variable that measured the presence or absence of income maintenance receipt relative to the occurrence of the first and second maltreatment reports, an additional variable was created for the purpose of summing all income maintenance spells that had occurred in relationship to the first and second maltreatment reports. Furthermore, several age-based variables were created for the purpose of identifying child and primary caregiver age in relationship to key events such as the first maltreatment report.

Reducing dimensionality: Steps taken during pre-processing.

Having expanded information content for the inputs to the greatest and most relevant degree possible, principal components analysis was applied as a method for reducing dimensionality in input space while retaining a maximum amount of information (i.e., variance explained by the group of original inputs). Each set of family and perpetrator characteristics was assessed for dependence among any possible combination of the inputs using Pearson chi-square analyses. Statistically significant and strong associations were found among four worker-observed family characteristics that appear to be family strengths/protective factors to include (a) amenable to services, (b) presence of stable family relationships, (c) presence of adequate parenting skills, and (d) presence of adequate living conditions. Specifically, amenable to services and stable family relationships were significantly associated (χ^2 = 1121.03, *df*=1, *N*=6747, $p < 0.001$); amenable to services and adequate parenting skills were significantly associated (χ^2 = 190.42, *df*=1, *N*=6747, $p < 0.001$); amenable to services and adequate living conditions were significantly associated (χ^2 = 545.85, *df*=1, *N*=6747, $p < 0.001$); stable family relationships and adequate parenting skills were significantly associated (χ^2 = 116.65, *df*=1, *N*=6747, $p < 0.001$); stable family relationships and adequate living conditions were significantly associated (χ^2 = 334.38, *df*=1, *N*=6747, $p < 0.001$); and adequate parenting skills and adequate living conditions were significantly associated (χ^2 = 56.80, *df*=1, *N*=6747, $p < 0.001$).

Statistically significant and strong associations were found among three worker-observed perpetrator characteristics that appear to be perpetrator strengths/protective factors to include (a) amenable to services, (b) presence of an adequate support system, and (c) no apparent mental-emotional disturbance. Specifically, amenable to services and adequate support system were significantly associated (χ^2 = 30.78, *df*=1, *N*=6747, $p <$ 0.001); amenability to services and no apparent mental-emotional disturbance were significantly associated (χ^2 = 625.61, *df*=1, *N*=6747, $p < 0.001$); and adequate support system and no apparent mental-emotional disturbance were significantly associated (χ^2 = 191.76, *df*=1, *N*=6747, $p < 0.001$). Finally, statistically significant and weak to moderate associations were found among three worker-observed perpetrator characteristics that appear to be aspects of the perpetrator's caretaking style to include (a) loss of control during discipline, (b) unrealistic expectations of the child, and (c) perpetrator immaturity. Specifically, loss of control during discipline and unrealistic expectations of the child were significantly associated (χ^2=12.18, *df*=1, *N*=6747, $p < 0.001$); loss of control during discipline and perpetrator immaturity were significantly associated (χ^2= 5.32, *df*=1, *N*=6747, $p < 0.05$); and unrealistic expectations of the child and perpetrator immaturity were significantly associated (χ^2=6.26 , *df*=1, *N*=6747, $p < 0.05$).

A principal components analysis (PCA) with varimax rotation was conducted for each of the three sets of family and perpetrator characteristics to determine if each of the subsets could be replaced with one or two composite variables that captured a substantive portion of the common variance shared by the original inputs as well as the variance that was unique to each input (Bishop, 1995; Dunteman, 1989; Gorsuch, 1983; Harman, 1976; Koutsoukos et al., 1994). As a data reduction technique, PCA extracts components that are linear combinations of the original inputs. Alternatively, each original input can be viewed as a function of the linear combination of the weighted component scores as follows:

$$x_{i1} = \mu_1 + \lambda_{11}C_{i1} + \lambda_{12}C_{i2} \ldots + \lambda_{1q}C_{iq} + \varepsilon_{i1}$$

$$x_{i2} = \mu_2 + \lambda_{21}C_{i1} + \lambda_{22}C_{i2} \ldots + \lambda_{2q}C_{iq} + \varepsilon_{i2}$$

$$x_{ip} = \mu_p + \lambda_{p1}C_{i1} + \lambda_{p2}C_{i2} \ldots + \lambda_{pq}C_{iq} + \varepsilon_{ip},$$

where x_{i1} = the value for the original input for the i^{th} person, μ = the intercept, λ = the component loading (measures the correlation between the component and the original input variable), *C* = the component score (the standardized value for the original input variable multiplied by a standardized scoring coefficient to determine the amount of the common and unique variance each observation possesses), *p* = the number of original input variables, and *q* = the number of components.

Before proceeding to PCA, the %POLYCHOR SAS macro was used to compute a tetrachoric correlation matrix for each of the three subsets of family and perpetrator characteristics (SAS, 2011). A one-component solution with an eigenvalue of 2.01 and four component loadings that were greater than .40 was identified for the group of family characteristics that appear to describe a set of family-based strengths or protective factors (i.e., amenable to services, stable family relationships, adequate parenting skills, and adequate living conditions). With an eigenvalue of 2.01 and a set of four original inputs, the component for the four family-based protective characteristics summarized about 50.32% of the variance in the original inputs and was equivalent to about two of the four original inputs. A one-component solution with an eigenvalue of 2.01 and three component loadings that were greater than .40 was identified for the group of perpetrator characteristics that appear to describe a set of perpetrator-based strengths or protective factors (i.e., amenable to services, adequate support system, and no apparent mental-emotional disturbance). With an eigenvalue of 2.01 and a set of three original inputs, the component for the perpetrator-based protective characteristics summarized about 67.06% of the variance in the original inputs and was equivalent to about two of the three original inputs. Finally, a one-component solution with an eigenvalue of 2.00, two component loadings that were greater than .40, and a third loading that was under .40 was identified for the group of perpetrator characteristics that appear to describe a set of caretaking characteristics (i.e., loss of control during discipline, unrealistic expectations of the child, and perpetrator immaturity). With an eigenvalue of 2.00 and a set of three original inputs, the component for the perpetrator-based caretaking characteristics summarized about 66.70% of the variance in the original inputs and was equivalent to about two of the three original inputs.

Rather than substitute component scores for the scores of the original input values, where component scores are a linear combination of the standardized values of the original inputs multiplied by a standardized scoring coefficient obtained through a refined regression method (SAS, 2011), values for each of the original inputs represented by each of the one-component solutions were summed.[3] Hence, if a one-component solution represented four original dichotomous inputs, the score for the new composite variable would range from 0 to 4. Due to the nature of the distribution of the worker observations, each of the new composite variables ranged from 0 to 1, where at most, each case record was noted as possessing one of the family-based protective factors, one of the perpetrator-based protective factors, and one of the perpetrator-based caretaking characteristics. Despite the lack of variance in the values for each new composite variable, it is reasonable to combine the original inputs within the three original subsets to form three respective composite variables given the empirical support provided by the principal components analyses.

In addition to creating new composite variables to summarize the variance in the original inputs for three sets of worker-observed family and perpetrator characteristics, one final PCA with varimax rotation was conducted with a set of original inputs from the 1990 U.S. Census data that appeared to describe various aspects of community-level poverty and stability. The set of census data-derived community characteristics was assessed for dependence among any possible combination of the inputs using Pearson correlation analyses. Statistically significant correlations were found among all of the census tract-based inputs to include (a) median household income, (b) percent of adults (25+) with a high school degree, (c) percent of households that moved within the last five years, (d) percent of children living below the poverty line, and (e) percent of adults without employment. Strong correlations occurred among all of the inputs with the exception of percent of households that moved within the last five years. Table 3.1 summarizes the correlations among all of the census-tract community characteristics.

Table 3.1

Pearson Correlation Coefficients for 1990 Census Tract Variables (N=6,747)

	Z_CENS HHINCOME	Z_CENS HIGH	Z_CENS MOVE	Z_ZZKID PCTBPL	Z_ZZNOT LABOR FORCE
Z_CENS HHINCOME	1				
Z_CENS HIGH	0.77508***	1			
Z_CENS MOVE	-0.07208***	0.11202***	1		
Z_ZZKID PCTBPL	-0.75933***	-0.80161***	0.07971***	1	
Z_ZZNOT LABORFORCE	-0.70820***	-0.76383***	0.00843	0.90434***	1

* $p<0.05$, ** $p<0.01$, ***$p<0.001$

A one-component solution with an eigenvalue of 3.36 and four component loadings that were greater than .80 was identified for the group of census tract characteristics that appear to describe various aspects of poverty (i.e., median household income, percent of adults with a high school degree, percent of children living below the poverty line, and percent of adults without employment). With an eigenvalue of 3.36 and a set of four original inputs, the component for the four aspects of census tract poverty summarizes about 83.96% of the variance in the original inputs, and is equivalent to about two and one-third of the four original inputs. The one-component solution for the measures of census-tract poverty agree with the PCA-derived results reported by Sampson (1997) in his seminal work on collective efficacy theory, where he used PCA to reduce a large number of 1990 U.S. Census indicators into three components that each represented a construct of social disorganization theory to include poverty, neighborhood stability and ethnic heterogeneity. Standardized component scores from the PCA conducted for this dissertation study were substituted for the four original poverty-based inputs, and percent of households that moved within the last five years was retained as a single indicator.

Feature selection: Steps taken during pre-processing.

As described in the preceding section, PCA can be used as a technique to reduce the dimensionality of the input space by mapping a larger number of original inputs onto a smaller number of features that summarize a maximum portion of variance in the original inputs. Feature selection is different in that no mapping processes are used to produce a smaller number of features to represent a larger number of inputs (Bishop, 1995). During the feature selection process, several steps were taken to select the specific features (i.e., inputs) that would be included in the neural network analysis.

The first step in the feature selection process included the elimination of dichotomous variables with such a low amount of variance that less than 2% of the observations were located in the "yes" category (represented by a "1"). For example, following the principle components analyses conducted with the 26 worker-observed family characteristics and 21 worker-observed perpetrator characteristics, the remaining indicators (i.e., those not represented by an extracted component) were assessed for their level of variance. Any indicator that did not have at least 2% of the observations located in the "yes" category was eliminated. When describing their pre-processing stage, Schwartz, Kaufman, and Schwartz (2004) underscored the importance of including a sample of training patterns (i.e., a sample of case records) with a diverse array of input values from which the neural network can learn to predict output values; hence, the authors noted that input variables characterized by low variance were deleted.

Additionally, steps were also taken to combine response levels for categorical variables if (a) the number of observations falling into a given response level was low (e.g., the proportion of observations falling into a given response level ranged from 0.04% to 8%), and (b) if the response levels could reasonably be combined. Examples include the combination of responses for the child's race/ethnicity such that observations fell into one of two categories: White or Non-White. Observations falling into categories to include Black, Hispanic, American Indian, Asian American, or other were combined into the Non-White category. Responses for the parent's status as the potential perpetrator of the first maltreatment incident were combined into one of two categories: the parent was identified as the perpetrator or the parent was not identified as the perpetrator. Observations falling into categories to include the biological parent, the adoptive parent, or the step-parent were combined into the parent was identified as the perpetrator category. Observations falling into categories to include foster parent (children who were not returned home following a foster care placement were not considered to be at risk of recurrent maltreatment and were therefore censored out of all analyses), grandparent, institution, paramour, sibling, or other were combined into the parent was not identified as the perpetrator category.

The second step in the feature selection process was defined by the use of configural analysis -- i.e., a regression model that classifies observations into categories of output values as a function of input values -- to select the input variables or features that would be used in the neural network analysis. As noted by Schwartz, Jones, Schwartz, and Obradovic (2008) running a configural analysis prior to their neural network analyses was extremely helpful in identifying the inputs that would be most useful in predicting output values. A classification tree is a regression model with a tree-based structure and, similar to a neural network, is a method of supervised learning that maps a vector of input values *x* onto output values *y* to separate observations into classes defined by the categorical response level (e.g., where 1 = membership in the recurrent maltreatment class and 0 = membership in the non-recurrent maltreatment class) (Hastie, Tibshirani, & Friedman, 2001; Mitchell 1997). Also similar to a neural network, a classification tree predicts class membership by first maximizing conditional probabilities of class membership as a function of values for X, and then predicting class membership (i.e., assigning a target value of 1 or 0) by selecting the class with the highest conditional probability (Shalizi, 2009).

Unlike a neural network, there is no vector of network weights which in concert with values for *x* make the observed values of *y* more likely. Instead, a classification tree maximizes the conditional probability of class membership (Y = *y*) as a function of specific values for input variables (X = *x*) by repeatedly sub-setting the original sample of observations into groups where the target values (*y*) for Y are made increasingly similar as a function of specific values (*x*) for the input variables X. Hence, a classification tree maximizes the conditional probability of class membership by repeatedly splitting the data into smaller and smaller groups that have a high proportion of observations with a target value of *y* as a function as having specific values (*x*) for input variables X (Hastie, Tibshirani, & Friedman, 2001; Mitchell, 1997; Shalizi, 2009). The process of repeatedly splitting the data into smaller groups characterized by a particularly high proportion of observations with a specific target value (*y*) for Y is guided by two principles that are essentially different sides of the same concept. Specifically, entropy is a measure of the expected reduction in disorder (i.e., dissimilarity among observed target values) among a subset of observations as a result of information gained by knowing how the proportion of observed target values can be changed as a function of values for X (Mitchell, 1997; Shannon, 1948). Additionally, the cross-entropy error function can be used to minimize the negative log conditional likelihood function, which is equivalent to maximizing the modeled probability of Y given X:

$$logL(data, Q) = -\frac{1}{n}\sum_{i=1}^{n} logQ(Y = y_i|X = x_i),$$

where $Q(Y = y|X = x)$ is the conditional probability the model predicts (Shalizi, 2009, pp. 21-22).

The following sections describe the splitting algorithm that is used to recursively (i.e., repeatedly) partition the original sample of observations into increasingly smaller and specialized groups (i.e., specialized in relationship to the specific values for Y and X). The point at which a set of observations is split is called a decision point and at each decision point, a classification or prediction rule is produced that describes the succession of input values (*x*) that various predictors X must take on in order to produce a group of observations that have a high conditional probability of class membership (Y = *y*) and are therefore predicted as having the target value that corresponds with the categorical response represented by *y* (Hastie, Tibshirani, & Friedman, 2001; Shalizi, 2009).

Feature selection through the application of CART.

In short, the classification and regression trees (CART) algorithm uses a succession of binary splits to partition the original sample of observations into increasingly homogenous groups, where the observations' target values are increasingly similar as a function of the values for the predictors used to partition the sample. Each recursive binary split resembles a test where all predictors that have been entered into the model are evaluated for their capacity to provide a split point (i.e., a single value) that simultaneously (a) reduces the dissimilarity (i.e., disorder as measured by entropy) of target values within each of the resulting two subsamples, and (b) maximizes the modeled probability of Y given X within each of the resulting two subsamples (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Fox, 2000; Mitchell, 1997; Shalizi, 2009). The structure of the classification tree is wholly derived from the data and is therefore not subject to constraints imposed by a priori assumptions about functional form. In fact, a classification tree is an ideal choice for the pre-processing stage because any nonlinear structure that may exist in the data is likely to be incorporated in the classification rules that are recursively developed as the splitting algorithm seeks the predictors that best separate observations by their target values (Strobl, Malley, & Tutz, 2009).

The splitting algorithm conducts a greedy search that looks for the best immediate predictor for each particular "partition" test without taking into account the ways in which the current split will influence future splits; all predictors become candidates in each search for the best variable and at the best value with which to partition the observations into two groups. Once two observations are split into two different groups, the observations cannot rejoin (Stine, 2011). The structure of the tree is determined by the selected splitting criterion and the selected stopping criterion. As noted earlier, the splitting criterion is used to select the predictor which provides the best reduction in entropy as a measure of disorder among a collection of observations. Mitchell (1997, p. 57) defines entropy as

$$Entropy\ (S) \equiv \sum_{i=1}^{c} -p_i log_2 p_i,$$

where S = a sample of observations containing a mixture of target values as measured by class i, and p_i = the proportion of S belonging to class i. Hence, entropy as described above measures the degree to which a collection of observations have the same target values and therefore belong to the same class (i.e., 1 = the recurrent maltreatment class and 0 = the non-recurrent maltreatment class). A decrease in entropy makes even more sense in relationship to the concept of information gain where a binary split at a given value of predictor (i.e., attribute) *A* partitions the observations into two subsamples. Information about the target values is gained as a function of the value of attribute *A*, and candidates for each split are assessed relative to the expected reduction in entropy that will occur as a result of knowing the particular value of attribute *A* (Mitchell, 1997; Shannon, 1948). Specifically, information gain, $Gain\ (S, A)$ of an attribute *A*, in relationship to a sample of observations S is represented as

$$Gain\ (S, A) \equiv Entropy\ (S) - \sum_{v \in Values\ (A)} \frac{|S_v|}{|S|} Entropy(S_v),$$

where $Values\ (A)$ = the set of all possible values for attribute *A*, and S_v = the subset of S for which attribute *A* has value *v* (Mitchell, 1997, p. 58).

JMP Pro 9 uses the likelihood ratio chi-square G^2 with adjusted *p*-values to assess the degree to which attribute candidates are significantly associated with the target variable and therefore increase the probability that Y = y given the target variable's dependence on X = *x*. Entropy is reduced by splitting the sample of observations into two groups by a value of *x* on which y is known to be dependent, where G^2 is twice the change in Entropy and Entropy is $\sum_{i=1}^{N} -\log(p_i)$, p_i = the probability attributed to the response that occurred for each observation (SAS, 2010). By recursively splitting the observations in this fashion, the input space is partitioned into rectangular decision regions, where each region contains a collection of data points (i.e., observations) that are housed in a terminal or leaf node from which no further splits are conducted. Hence, each terminal node in the classification tree represents a decision region in input space whose location in input space is a function of the values (*x*) of X. A piecewise constant probability model is fitted to all observations within each respective decision region where conditional probabilities of class membership (for each class) are estimated as the proportion of observations in each categorical response level (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Hastie, Tibshirani, & Friedman, 2001; Shalizi, 2009; Stine, 2011). Hence, every observation within a given decision region will have the same probabilities of class membership (for each class) as a function of the values of the inputs that define each decision region.

As noted earlier, a decision region functions as a classifier that provides a fast way of predicting the target values for new observations. Any new observation with a sequence of values for X that agree with the sequence of values that define a given decision region will automatically be placed in that region. Moreover, the predicted probabilities of class membership from that decision region will automatically be assigned to the new observation, and the predicted target value will be determined by selecting the class with the highest probability of membership (i.e., the response level with the largest proportion of observations) (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Hastie, Tibshirani, & Friedman, 2001; Shalizi, 2009; Stine, 2011).

The recursive binary splitting could continue indefinitely, producing a very complex tree that has been over-fitted to the data. Hence, the selection of a stopping criterion is important. A common stopping criterion is the selection of a minimum number of observations that must be included in a given node; this stopping criterion was employed in this dissertation study where a minimum of 25 observations was specified (Stine, 2011; Strobl, Malley, & Tutz, 2009). Setting a ground floor requirement in relationship to the number of observations that must be present in each node assists in improving the predictive accuracy of the tree by avoiding the creation of prediction rules that are based on input values affected by a very small number of observations.

Figure 3.17 provides an example of a simple classification tree. At the top of the tree is the root node where all of the observations are stored before any splitting has occurred. The first branch that extends underneath the root node and runs from the left to the right represents the first test for which a candidate search is executed to find the specific value for a particular predictor variable that will be used to partition the total sample of observations into two groups. At the ends of each branch are the two groups that have been created by the split and each node is referred to an internal node. Each internal node represents a decision point that has been reached after conducting the search to find the best predictor upon which to split the data (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Shalizi, 2009; Strobl, Malley, & Tutz, 2009). Each subsequent search and binary split produces two additional subsamples that flow from a

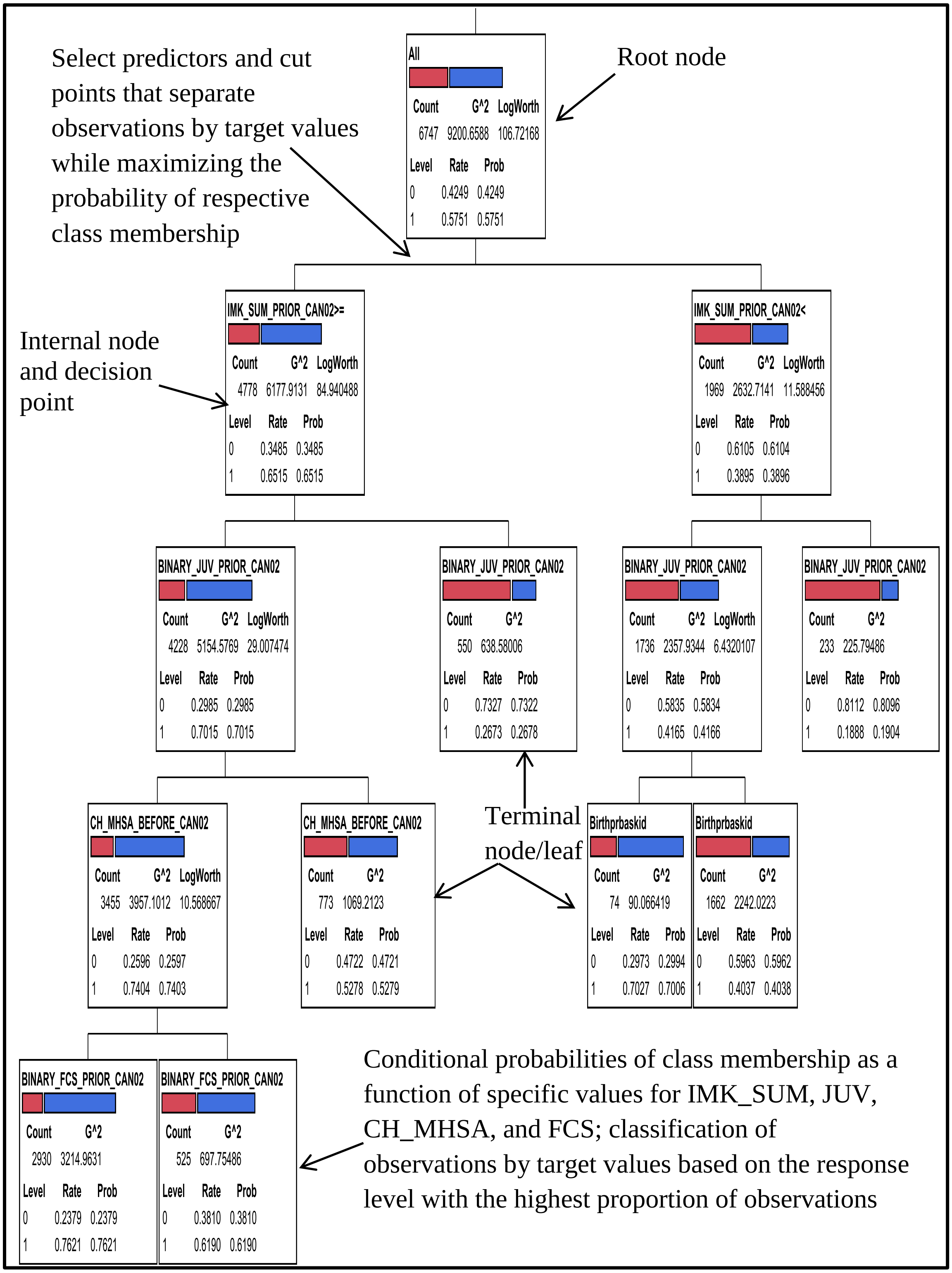


Figure 3.17. Example of a CART classification tree with binary recursive partitioning as produced in JMP Pro 9.

particular internal node (i.e., the parent node) to a set of two additional internal nodes (i.e., children or daughter nodes). Nodes from which no further splitting occurs are referred to as terminal nodes or leaves. Each leaf represents a rectangular decision region in input space (an example of which is depicted by Figure 3.19). All splits leading to a particular leaf represent each successive classification rule that is used to define the location of the corresponding decision region in input space as well as the configuration of values of X that were used to estimate each observation's (in that leaf) probabilities of class membership and predicted target value.

In order to see how a classification tree's terminal nodes map onto the decision regions in input space, please refer to Figures 3.18 and 3.19. Figure 3.18 provides an example of a simple classification tree with five terminal nodes that correspond with the five decision regions in two-dimensional input space found in Figure 3.19. Specifically, tracing down from the root node in the classification tree in Figure 3.18 to the first terminal node labeled as "R_1," it can be seen that observations in R_1 have a value for X_1 that is less than or equal to t_1 and have a value for X_2 that is less than or equal to t_2. Taking this information and looking at Figure 3.19, it can be seen that the R_1 decision region is situated in input space where the region runs from the far left of X_1 up to t_1 and the region runs from the bottom of X_2 up to t_2. The terminal node's position relative to the cut-point values established in the classification tree maps onto its decision region position relative to the same cut-point values along the axes of the input space (Hastie, Tibshirani, & Friedman, 2001).

The complexity of a classification tree is determined by the number of splits that produce corresponding branches and internal nodes. A tree with a high level of complexity has low bias but high variance; similar to the neural network, as the complexity of the tree increases, the ability to fit a given set of data very well actually constrains the generalizability of the tree in being able to predict target values for new observations. Conversely, a tree with too few branches and internal nodes will have less variance but will have increased bias and poor predictive accuracy (Fox; Hastie, Tibshirani, & Friedman, 2001; Mitchell; Stine, 2011; Strobl, Malley, & Tutz, 2009). Thus, an optimum balance is needed to provide a level of complexity (not too much and not too little) that maximizes generalizability. Moreover, classification trees are

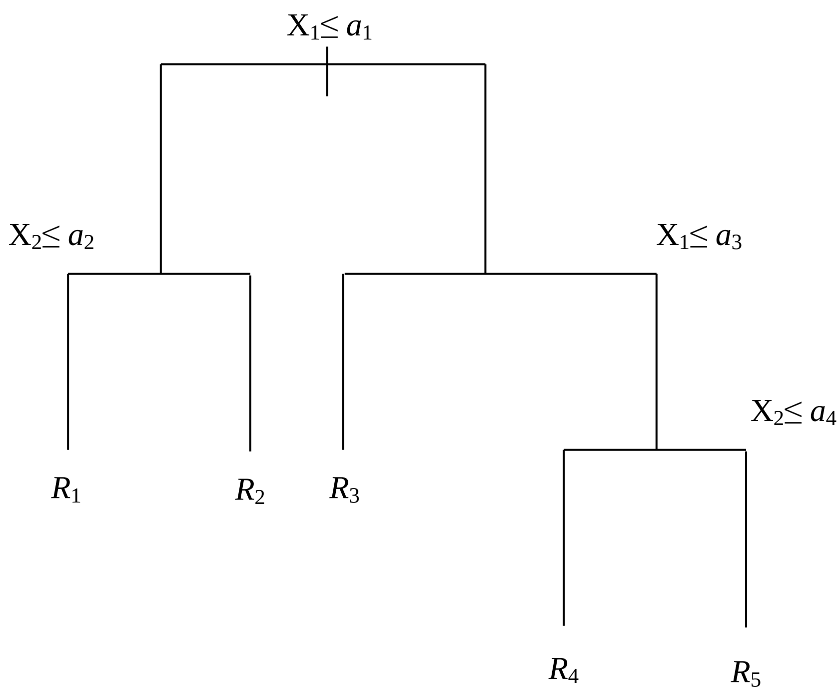


Figure 3.18. CART classification tree with five terminal nodes that correspond with the input space (featured in Figure 3.19) partitioned into five decision regions. Adapted from *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by T. Hastie, R. Tibshirani, and J. Friedman, 2001, p. 268.

very sensitive to the observations in the sample as well as the features used to partition the input space. The series of splits is sequential; thus, the fit of the tree to the data as well as the tree's predictive accuracy is dependent upon a sequential set of local decisions. The inherent lack of stability in a single classification tree has led to the development of ensemble methods that allow the researcher to average the results from a large number of classification trees (Breiman, 1996; Hastie, Tibshirani, & Friedman, 2001; Stine, 2011; Strobl, Malley, & Tutz, 2009). The random forest is a particularly well-supported example of such an ensemble method and is the technique

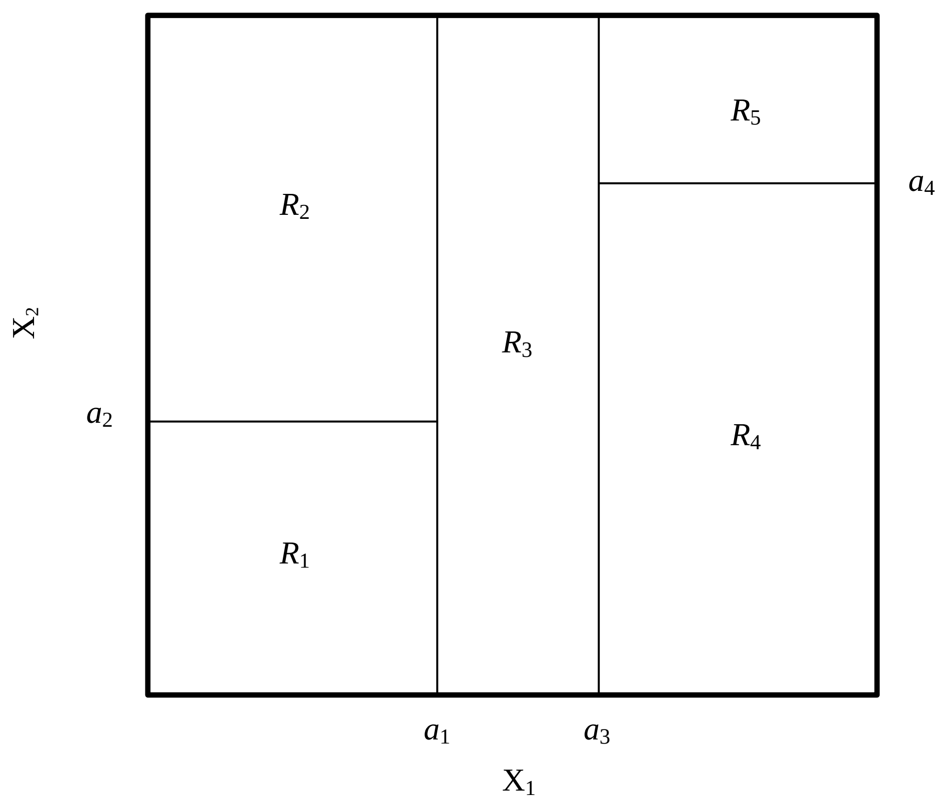


Figure 3.19. Depiction of a two-dimensional input space with five decision regions corresponding to five terminal nodes from a classification tree created with a CART recursive binary splitting algorithm. Adapted from *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, by T. Hastie, R. Tibshirani, and J. Friedman, 2001, p. 268.

that was utilized in this dissertation study (Biau, Devroye, & Lugosi, 2008; Breiman, 2001a; Cutler et al., 2007; Diaz-Uriarte & de Andrés, 2006; Evans & Cushman, 2009; Genuer, Poggi, & Tuleau-Malot, 2010; Lin & Jeon, 2006; Strobl, Malley & Tutz, 2009).

As an ensemble method, random forests use bagging or bootstrap aggregation to produce many classification trees and then average the results across the ensemble of trees in order to reduce variance, smooth decision boundaries, and improve predictive accuracy (Hastie, Tibshirani, & Friedman, 2001). Specifically, bootstrap resampling is used to create a succession of samples based upon the original collection of observations. Treating the observed sample as the population, bootstrap samples with replacement are repeatedly drawn from the population, which typically leaves two-thirds of the population observations in a given bootstrap sample and one-third of the population observations out of the given bootstrap sample. Population observations that are included in the bootstrap sample are described as being "in the bag," while observations that are excluded from the bootstrap sample are described as being "out of the bag." A separate classification tree is fitted to each bootstrap sample, and at each split only a random subset m of the total number of predictors p is considered, where m may equal $\sqrt{p}$ or $log_2 p$. Limiting each split to candidates that are random subsets of the total number of predictors increases the independence of the trees and lowers generalization error (Breiman, 2001a; Hastie, Tibshirani, & Friedman, 2001 Strobl, Malley, & Tutz, 2009).

JMP Pro 9 estimates probabilities of class membership using in-bag observations and then uses a validation sample to fine tune model complexity by varying the number of trees and the number of randomly selected terms for each split. Model complexity is refined in relationship to values for Entropy R^2, which is 1 minus the ratio of the log-likelihoods from the fitted model and the constant probability model (Sall, 2009; SAS, 2010). Similar to refining model complexity for a neural network using the validation data set, the complexity of the ensemble of trees in a random forest is refined using a validation set of observations because generalization can be assessed by evaluating model fit in relationship to new observations (i.e., observations not used for training). Model complexity can be further specified by selecting the minimum number of splits each tree is required to make, and for the purpose of this dissertation study, six was selected as the minimum number of splits that had to be executed (although selecting a minimum number does not constrain trees from executing additional splits in relationship to entropy and the splitting criterion) (Stine, 2011). Likelihood ratio chi-square values for all predictors were sorted in descending order as a measure of variable importance. Predictor variables were not selected for the neural network model if $G^2 = 0$. Additionally, predictors located in the bottom 20% of the ordered G^2 values were not selected for the neural network model (Diaz-Uriarte & de Andrés, 2006; Genuer, Poggi, Tuleau-Malot, 2010).

In sum, the utilization of a random forest as a feature selection technique while preparing for a neural network analysis is supported by the scholarship of Schwartz, Jones, Schwartz, and Obradovic (2008) as well as the feature selection criteria set forth by Bishop (1995). Specifically, the method used for feature selection should be based on a systematic procedure that searches through all candidate features. Additionally, features were selected in accordance with the degree to which they function effectively as classifiers that can separate observations into distinct decision regions. Moreover, the features' effectiveness in separating observations into distinct classes is evaluated with the same measures -- i.e., misclassification through a confusion matrix and overall predictive accuracy by examining the area under the Receiver Operating Characteristic

(ROC) curve -- across both the random forest and the neural network. Finally, the random forest is also ideal for assessing the features' capacity to act as classifiers both as individual inputs as well as two or more inputs that work interactively. Just like a neural network, a random forest builds a model from the data as opposed to fitting a model with a pre-specified functional form to the data. Hence, there are no assumptions or requirements of nonlinearity that could hamper a random forest's ability to identify the presence of underlying nonlinear structure that would typically be overlooked but is essential in obtaining the best classification results. As noted by Strobl, Malley, and Tutz (2009), "an ensemble of trees has the advantage that it gives each variable the chance to appear in different contexts with different covariates, and can thus better reflect its potentially complex effect on the response" (p. 20).

Missing data and multicollinearity: Steps to complete pre-processing.

Upon completing all of the previously described pre-processing tasks to prepare for the neural network analysis, two issues remained. First, the final data set needed to be assessed for the presence of missing data. In all, 193 cases had missing values for parent gender and/or the parent's potential status as the perpetrator of the reported maltreatment event; hence, in a data set with 6,747 total case records, just 2.9% of the case records were missing values for no more than two variables. Binary variables were created for each predictor to account for the presence of missing observations where case records with missing data were coded as 1, and case records without missing data were coded as 0. A Pearson chi-square analysis revealed that the incidence of missing values for parent gender was not dependent on the response variable for recurrent maltreatment [$\chi^2 = .0034\ (df = 1, N = 6747), p = 0.9533$]. Additionally, a Pearson chi-square analysis revealed that the incidence of missing values for parent status as perpetrator was not dependent on the response variable for recurrent maltreatment [$\chi^2 = .0382$ ($df = 1, N = 6747), p = 0.8450$].

As noted by Allison (2002), should multiple imputation or maximum likelihood be selected as a method for producing values for the missing observations, the model that will be used to generate values for the missing data must be specified in advance and the model used to generate values for the missing data must agree with the model used for the analysis. Moreover, both models should accurately represent the data. Due to the nature of the neural network analysis, there is no pre-specified functional form or model that describes the association between recurrent maltreatment and its predictors. As an alternative, listwise deletion was assessed for its viability given the choice of listwise deletion by Schwartz, Kaufman, and Schwartz (2004) for their neural network analysis and given Bishop's (1995) edict that "if the quantity of data available is sufficiently large, and the proportion of patterns affected is small, then the simplest solution is to discard those patterns from the data set" (p. 301). Upon further examination, listwise deletion was selected as the method for handling the missing observations for this dissertation analysis due in large part to the fact that the missing observations in the two predictor variables did not depend on the response variable. Allison (2002) notes that in circumstances when the probability of missing data within a predictor variable is *independent* of the values of the outcome variable, "regression estimates using listwise deletion will be unbiased" (pp. 6-7).

The second and final issue that needed to be addressed prior to running a neural network analysis was based on Haykin's (1999) discussion of multicollinearity. Specifically, Haykin provides a clear set of heuristic guidelines for improving the performance of a neural network model, and eliminating multicollinearity or correlation among one or more predictors is a key suggestion. Multicollinearity is defined in relationship to the degree to which any given predictor (X_i) is correlated with one or more other predictors, and therefore the degree to which X_i can provide unique information in the prediction of the outcome variable Y (Cohen, Cohen, West, & Aiken, 2003). Variance inflation factor (VIF) scores measure increases in the variance of parameter estimates (i.e., standard errors) in relationship to a baseline condition in which none of the predictors are correlated, and VIF scores are calculated as follows

$$\left(\frac{1}{1-R^2_{i.12\ldots(i)\ldots k}}\right),$$

where $R^2_{i.12\ldots(i)\ldots k}$ is the squared multiple correlation between X_i and the other predictors in the regression model (Cohen, Cohen, West, & Aiken, 2003, p. 423). The determination of the point at which multicollinearity is problematic varies, with cut-off points ranging from a conservative value of 2 to a more lenient value of 10. Taking the square root of a VIF value provides a measure against which to interpret the degree to which the standard error of a parameter estimate increases in comparison with the variance expected when none of the predictors are correlated. For example, for a VIF of 10 where the square root of 10 is equal to 3.16, the standard error of a regression coefficient with a VIF 10 will have a standard error that is a little over three times greater than the standard error that would be obtained if no correlation among the predictors was present. In order to promote the highest level of predictive accuracy possible, I elected to use the most conservative cut-off value of 2 for determining the point at which multicollinearity was problematic.

Upon running an ordinary least squares regression model and requesting that VIF scores be provided (estimated using PROC REG in SAS 9.3), I examined each VIF score in relationship to the selected cut-off value of 2. Upon removing three inputs that were highly correlated with at least one other predictor [i.e., (a) worker-observed family characteristic that the primary caregiver was a single parent, (b) receipt of FCS within 45 days of the first maltreatment report[4], and (c) primary caregiver's age at the first maltreatment event], all VIF scores dropped below 2, and a subsequent neural network model run without the three deleted predictors benefitted from a decreased misclassification rate and an increased area under the ROC. Moreover, a visual inspection of the plots of the probability of maltreatment recurrence by each predictor in the neural network revealed changes in the overall shape and steepness of the slopes for multiple inputs.

Table 3.2 provides a full review of each variable within the pre-processed data set before the random forest analysis. This table includes all of the variables that were used to create new features (e.g., worker-observed family characteristics that were mapped onto one component during PCA analyses) as well as the features that were created to represent subsets of original inputs. A full description of each variable is provided to include the variable name, a description of the variable, the variable's level of measurement, and the coding scheme/units of analysis for each variable.

Table 3.3 provides univariate statistics, bivariate statistics, and results from the random forest analysis. Pearson chi-square analyses were used to test for an association between each dichotomous input and recurrent maltreatment, while point biserial analyses were used to test for an association between each continuous input and recurrent maltreatment. As noted earlier, at the point of each recursive binary split, a likelihood ratio chi-square (G^2) analysis is conducted to test for dependence in the distribution of values for recurrent maltreatment on the values of each input variable. The input with the largest G^2 and correspondingly, the smallest adjusted p-value, is selected as the input that will be used to split the observations into two groups. Specifically, the splitting algorithm estimates class conditional values for x for which the observed values y in each class (i.e., in each response category) have the highest probability of occurrence. Feature selection is then based on the sorting of the G^2 values of the inputs in descending order to determine the relative level of contribution each input made to the ensemble of classification trees in the forest. Please note that the adjusted p-values for each G^2 statistic are rescaled in JMP Pro 9 as LogWorth values where LogWorth is calculated as $-log_{10}(p - value)$. Larger LogWorth values indicate smaller p-values. For example, a LogWorth value of 2 is the p-value 10^{-2} which is equivalent to a p-value of 0.01 (SAS, 2010).

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes

Variable Name	**Variable Description**	**Level of Measurement**	**Coding Scheme**
Outcome Variable			
Maltxt	Second maltreatment report	Nominal	1=Yes, 0=No
Child Characteristics			
Ch_Gender	Child gender	Nominal	1=Female, 0=Male
Ch_Race	Child race/ethnicity	Nominal	1=Non-White, 0=White
Ch_Age	Child age at first maltreatment report	Continuous (interval)	In years
Ch_Birth_Prb	Presence of very low birth weight or other birth complication in first year of child's life that is ongoing	Nominal	1=Yes, 0=No
Primary Caregiver Characteristics			
Pt_Gender	Primary caregiver gender	Nominal	1=Male, 0=Female
Pt_Age	Primary caregiver age at birth of child	Continuous (interval)	In years
Pt_Age_CAN01	Primary caregiver age at first maltreatment report	Continuous (interval)	In years
Pt_Educ	Primary caregiver educational status	Nominal	1=High school degree or greater, 0=No high school degree

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	**Variable Description**	**Level of Measurement**	**Coding Scheme**
Mom_Hx_Fost_Care	Primary caregiver foster care history	Nominal	1=Mother was foster care child 0=Mother not foster care child
Maltreatment Event Characteristics			
Substantiation	Substantiation status for 1st maltreatment report	Nominal	1=Substantiated 0=Unsubstantiated
Pt_Perp	Perpetrator was parent, adoptive parent, or step parent for 1st maltreatment report	Nominal	1=Perpetrator was parent 0=Perpetrator was not parent
Worker-Observed Family Characteristics			
Fam_Single_Pt	Single parent household	Nominal	1=Characteristic Present 0=Characteristic Not Present
Fam_Lack_Pt_Skills	Lack of parenting skills	Nominal	1=Characteristic Present 0=Characteristic Not Present
Fam_Amen_Svcs	Family is amenable to services at 1st maltreatment report	Nominal	1= Condition Present 0=Condition Absent
Fam_Stable_Relation	Stable family relationships at 1st maltreatment report	Nominal	1= Condition Present 0=Condition Absent
Fam_Adq_Pt_Skills	Adequate parenting skills at 1st maltreatment report	Nominal	1= Condition Present 0=Condition Absent
Fam_Adq_Liv_Cond	Adequate living conditions at 1st maltreatment report	Nominal	1= Condition Present 0=Condition Absent

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	Variable Description	Level of Measurement	Coding Scheme
Fam_Protective	Presence of a family protective characteristic	Nominal	1=Characteristic Present 0=Characteristic Not Present
Worker-Observed Perpetrator Characteristics			
Perp_Drug_Prb	Drug related problems	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Low_Self_Est	Low self esteem	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Immature	Immaturity	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Loss_Control	Loss of control during discipline	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Unreal_Expect	Unrealistic Expectations of Child	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Amen_Svcs	Amenable to Services	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Adq_Supp_Sys	Adequate Support System	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_No_Emot_Disturb	No apparent mental-emotional disturbance	Nominal	1=Characteristic Present 0=Characteristic Not Present
Perp_Neg_Care_Skills	Presence of a negative caretaking characteristic	Nominal	1=Characteristic Present 0=Characteristic Not Present

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	Variable Description	Level of Measurement	Coding Scheme
Perp_Protective	Presence of a protective characteristic	Nominal	1=Characteristic Present 0=Characteristic Not Present
Cross-Sector Service Characteristics			
FCS[a]	Received 1st FCS spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
FPS[a]	Received 1st FPS spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
ALT[a]	Received 1st ALT (foster care) spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
FCS_Wtn_45Days[a]	1st FCS spell began within 45 days of 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
FPS_Wtn_45Days[a]	1st FPS spell began within 45 days of 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
ALT_Wtn_45Days[a]	1st ALT spell began within 45 days of 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
FCS_Second_Sp[a]	Received 2nd FCS spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
FPS_Second_Sp[a]	Received 2nd FPS spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
ALT_Second_Sp [a]	Received 2nd ALT spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	Variable Description	Level of Measurement	Coding Scheme
Ch_Conflict_Svc_Pr_01[a]	Child received 1st mental health service (specifically related to conflict) prior to 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
Ch_Conflict_Svc_Pr_02[a]	Child received 1st mental health service (specifically related to conflict) on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
Ch_MHSA_Pr_01[a]	Child received 1st general mental health or substance abuse (MHSA) service prior to 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
Ch_MHSA_Pr_02[a]	Child received 1st MHSA service on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
Pt_MHSA_Pr_01[a]	Primary caregiver received 1st general mental health or substance abuse (MHSA) service prior to 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
Pt_MHSA_Pr_02[a]	Primary caregiver received 1st MHSA service on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
IM_Pr_01[a]	Received 1st income maintenance (IM) spell prior to 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition
IM_Pr_02[a]	Received 1st IM spell on or after 1st maltreatment report	Nominal	1=Presence of Service Condition 0=Absence of Service Condition

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	Variable Description	Level of Measurement	Coding Scheme
Num_IM_Sp[a]	Number of IM spells (out of 12 possible) received prior to 1st maltreatment report	Continuous (ratio)	Number of income maintenance spells
Juv_Ct_Pr_01[a]	1st juvenile court petition prior to 1st maltreatment report	Nominal	1=Presence of petition 0=Absence of petition
Juv_Ct_Pr_02[a]	1st juvenile court petition on or after 1st maltreatment report	Nominal	1=Presence of petition 0=Absence of petition
Crim_Pr_01[a]	1st criminal court arrest for primary caregiver prior to 1st maltreatment report	Nominal	1=Presence of arrest 0=Absence of arrest
Crim_Pr_02[a]	1st criminal court arrest for primary caregiver on or after 1st maltreatment report	Nominal	1=Presence of arrest 0=Absence of arrest
Spec_Ed_Pr_01[a]	1st special education screening prior to 1st maltreatment report	Nominal	1=Presence of screening 0=Absence of screening
Sec_Ed_Pr_02[a]	1st special education screening on or after 1st maltreatment report	Nominal	1=Presence of screening 0=Absence of screening
1990 Census Tract Characteristics			
Comm_Inc	Median household income in 1990 census tract	Continuous (interval)	Dollar amount
Comm_Educ	Percent of all adults (all races/ethnicities) 25 years of age and older in 1990 census tract with high school degree	Continuous (ratio)	Percentage

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.2

Variable Names, Variable Descriptions, Levels of Measurement, and Coding Schemes (Continued)

Variable Name	Variable Description	Level of Measurement	Coding Scheme
Comm_Move	Percent of households in 1990 census tract that moved between 1985 and 1990	Continuous (ratio)	Percentage
Comm_Ch_Pov	Percent of all children (all races/ethnicities) in 1990 census tract living below the poverty line	Continuous (ratio)	Percentage
Comm_Unemploy	Percent of all adults (all races/ethnicities) in 1990 census tract not working	Continuous (ratio)	Percentage
Comm_Pov	Factor regression score measuring the degree of concentrated poverty in 1990 census tract	Continuous (ratio)	Standardized score, mean = 0 and SD = 1

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (N = 6,747)

Variable Name	Variable Description	Univariate Statistics	Bivariate Associations	G^2
Outcome Variable				
Maltxt	Second maltreatment report	57.51% Yes 42.49% No	- -	- -
Child Characteristics				
Ch_Gender	Child gender	48.30 Female 51.70 Male	$\chi^2 = 0.04$ $df = 1$	G^2 =628.03 Splits =292
Ch_Race	Child race/ethnicity	63.29% NW 36.71 White	$\chi^2 = 74.07$*** $df = 1$	G^2 =915.77 Splits =217
Ch_Age	Child age at first maltreatment report	M=4.58 Yrs SD=3.14	Pt.BiserialTest R_{pbi}=-0.15841 t = 13.18*** df = 6745	G^2 =3908.48 Splits =420

Note. Splits refers to the total number of splits to which each input variable contributed.
[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt.
* p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	**Variable Description**	**Univariate Statistics**	**Bivariate Associations**	**G^2**
Ch_Birth_Prb	Presence of very low birth weight or other birth complication in first year of child's life that is ongoing	15.44% Yes 84.56% No	χ^2 = 77.02*** df = 1	G^2 =708.13 Splits =119
Primary Caregiver Characteristics				
Pt_Gender	Primary caregiver gender	8.82% Male 91.18% Female	χ^2 = 35.85*** df = 1	G^2 =138.63 Splits =44
Pt_Age	Primary caregiver age at birth of child	M=24.28 Yrs SD=6.27	Pt.Biserial Test R_{pbi}=-0.06172 t = 5.08*** df = 6745	G^2 =1558.89 Splits =295
Pt_Age_CAN01	Primary caregiver age at first maltreatment report	M=28.86 Yrs SD=6.97	Pt.Biserial Test R_{pbi}=-0.12692 t = 10.51*** df = 6745	G^2 =2193.30 Splits =317

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	Variable Description	Univariate Statistics	Bivariate Associations	G^2
Pt_Educ	Primary caregiver educational status	50.21% HS+ 49.79%No HS	χ^2 = 63.15*** df = 1	G^2 =1269.70 Splits =351
Mom_Hx_Fost_Care	Primary caregiver foster care history	3.72% Yes 96.28% No	χ^2 = 24.01*** df = 1	G^2 =150.33 Splits =23
Maltreatment Event Characteristics				
Substantiation	Substantiation status for 1st maltreatment report	19.64% Yes 80.36% No	χ^2 = 0.76 df = 1	G^2 =931.09 Splits =163
Pt_Perp	Perpetrator was parent, adoptive parent, or step parent for 1st maltreatment report	80.36% Yes 19.64% No	χ^2 = 57.49*** df = 1	G^2 =1313.07 Splits =192
Worker-Observed Family Characteristics				
Fam_Protective	Presence of a family protective characteristic	70.24% Yes 29.76% No	χ^2 = 28.59*** df = 1	G^2 =651.06 Splits =192

Note. Splits refers to the total number of splits to which each input variable contributed.
[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	Variable Description	Univariate Statistics	Bivariate Associations	G^2
Fam_Single_Pt	Single parent household	19.71% Yes 80.29% No	χ^2 = 19.95*** df = 1	G^2 =490.62 Splits =152
Fam_Lack_Pt_Skills	Lack of parenting skills	1.51% Yes 98.49% No	χ^2 = 3.56 df = 1	--
Worker-Observed Perpetrator Characteristics				
Perp_Neg_Care_Skills	Presence of a negative caretaking characteristic	6.11% Yes 93.89% No	χ^2 = 0.67 df = 1	G^2 =56.73 Splits =26
Perp_Protective	Presence of a protective characteristic	55.88% Yes 44.12% No	χ^2 = 47.54*** df = 1	G^2 =721.74 Splits =304
Perp_Drug_Prb [b]	Drug related problem	1.05% Yes 98.95% No	χ^2 = 2.22 df = 1	--
Perp_Low_Self_Est [b]	Low self esteem	1.33% Yes 98.67% No	χ^2 = 3.94* df = 1	--

Note. Splits refers to the total number of splits to which each input variable contributed.
[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	**Variable Description**	**Univariate Statistics**	**Bivariate Associations**	G^2
Perp_Immature [b]	Immaturity	1.72% Yes 98.28% No	$\chi^2 = 6.34^*$ $df = 1$	--
Cross-Sector Service Characteristics				
FCS[a]	Received 1st FCS spell on or after 1st maltreatment report	18.10% Yes 81.90% No	$\chi^2 = 75.87^{***}$ $df = 1$	G^2 =1667.96 Splits =159
FPS[a]	Received 1st FPS spell on or after 1st maltreatment report	4.45% Yes 95.55% No	$\chi^2 = 10.01^{**}$ $df = 1$	G^2 =100.69 Splits =21
ALT[a]	Received 1st ALT (foster care) spell on or after 1st maltreatment report	2.61% Yes 97.39% No	$\chi^2 = 0.34$ $df = 1$	G^2 =3.65 Splits =1
FCS_Wtn_45Days[a]	1st FCS spell began within 45 days of 1st maltreatment report	9.54% Yes 90.46% No	$\chi^2 = 39.88^{***}$ $df = 1$	G^2 =419.29 Splits =82
FPS_Wtn_45Days[a,b]	1st FPS spell began within 45 days of 1st maltreatment report	1.85% Yes 98.15% No	$\chi^2 = 12.19^{***}$ $df = 1$	--
ALT_Wtn_45Days[a,b]	1st ALT spell began within 45 days of 1st maltreatment report	1.76% Yes 98.24% No	$\chi^2 = 3.20$ $df = 1$	--

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the “yes” response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	Variable Description	Univariate Statistics	Bivariate Associations	G^2
FCS_Second_Sp [a, b]	Received 2nd FCS spell on or after 1st maltreatment report	0.62% Yes 99.38%	$\chi^2 = 2.60$ $df = 1$	--
FPS_Second_Sp [a,b]	Received 2nd FPS spell on or after 1st maltreatment report	0.27% Yes 99.73% No	$\chi^2 = 0.03$ $df = 1$	--
ALT_Second_Sp [a,b]	Received 2nd ALT spell on or after 1st maltreatment report	0.03% Yes 99.97% No	Fisher's Exact Test *p*-value=1.00	--
Ch_Conflict_Svc_Pr_01[a]	Child received 1st mental health service (specifically related to conflict) prior to 1st maltreatment report	6.14% Yes 93.86% No	$\chi^2 = 24.18$*** $df = 1$	G^2 =78.60 Splits =27
Ch_Conflict_Svc_Pr_02[a]	Child received 1st mental health service (specifically related to conflict) on or after 1st maltreatment report	8.00% Yes 92.00% No	$\chi^2 = 0.02$ $df = 1$	G^2 =77.16 Splits =31

Note. Splits refers to the total number of splits to which each input variable contributed.
[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt.
* $p<0.05$, ** $p<0.01$, and *** $p<0.001$ for all chi-square tests and 2-tail *t*-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	Variable Description	Univariate Statistics	Bivariate Associations	G^2
Ch_MHSA_Pr_01[a]	Child received 1st general mental health or substance abuse (MHSA) service prior to 1st maltreatment report	5.04% Yes 94.96% No	χ^2 = 18.76*** df = 1	G^2 =68.50 Splits =16
Ch_MHSA_Pr_02[a]	Child received 1st MHSA service on or after 1st maltreatment report	18.36% Yes 81.64% No	χ^2 = 120.41*** df = 1	G^2 =2992.09 Splits =177
Pt_MHSA_Pr_01 [a,b]	Primary caregiver received 1st general mental health or substance abuse (MHSA) service prior to 1st maltreatment report	0.46% Yes 99.54% No	χ^2 = 8.86** df = 1	--
Pt_MHSA_Pr_02[a]	Primary caregiver received 1st MHSA service on or after 1st maltreatment report	1.96% Yes 98.04% No	χ^2 = 36.36*** df = 1	G^2 =231.422 Splits =14
IM_Pr_01[a,c]	Received 1st income maintenance (IM) spell prior to 1st maltreatment report	70.82% Yes 29.18% No	χ^2 = 391.66*** df = 1	--
IM_Pr_02[a,c]	Received 1st IM spell on or after 1st maltreatment report	4.08% Yes 95.92% No	χ^2 = 0.79 df = 1	--

Note. [a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the “yes” response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	**Variable Description**	**Univariate Statistics**	**Bivariate Associations**	**G^2**
Num_IM_Sp[a]	Number of IM spells (out of 12 possible) received prior to 1st maltreatment report	*M*=1.46 spells *SD*=1.25	Pt.Biserial Test R_{pbi}=0.1622 t = 13.50*** df = 6745	G^2 =10172.06 Splits =302
Juv_Ct_Pr_01[a,b]	1st juvenile court petition prior to 1st maltreatment report	0.67% Yes 99.33% No	χ^2 = 4.64* df = 1	--
Juv_Ct_Pr_02[a]	1st juvenile court petition on or after 1st maltreatment report	11.61% Yes 88.39% No	χ^2 = 397.48*** df = 1	G^2 =8809.36 Splits =167
Crim_Pr_01[a]	1st criminal court arrest for primary caregiver prior to 1st maltreatment report	3.33% Yes 96.67% No	χ^2 = 5.19* df = 1	G^2 =9.75 Splits =3
Crim_Pr_02[a]	1st criminal court arrest for primary caregiver on or after 1st maltreatment report	1.94% Yes 98.06% No	χ^2 = 8.50** df = 1	G^2 =51.17 Splits =8

Note. Splits refers to the total number of splits to which each input variable contributed.

[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt.

* p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Table 3.3

Univariate Statistics, Bivariate Associations, and G^2 for Random Forest Analysis for General Maltreatment (Continued)

Variable Name	**Variable Description**	**Univariate Statistics**	**Bivariate Associations**	G^2
Spec_Ed_Pr_01[a]	1st special education screening prior to 1st maltreatment report	6.34% Yes 93.66% No	$\chi^2 = 0.17$ $df = 1$	G^2 =100.93 Splits =39
Spec_Ed_Pr_02[a]	1st special education screening on or after 1st maltreatment report	13.19% Yes 86.81% No	$\chi^2 = 156.36$*** $df = 1$	G^2 =3237.19 Splits =166
1990 Census Tract Characteristics				
Comm_Move	Percent of households in 1990 census tract that moved between 1985 and 1990	M=45.47% SD=11.86%	Pt.Biserial Test R_{pbi}=0.013893 t = 1.14 df = 6745	G^2 =1293.86 Splits =229
Comm_Pov	Factor regression score measuring the degree of concentrated poverty in 1990 census tract (standardized)	M=0 SD=1	Pt.Biserial Test R_{pbi}=0.17722 t = 14.79*** df = 6745	G^2 =4920.23 Splits =324

Note. Splits refers to the total number of splits to which each input variable contributed.
[a] Services were coded as having occurred only if they began before the outcome event for those cases where the child had a second maltreatment report. [b]Variables with less than 2% of the observations in the "yes" response level category were excluded from the random forest analysis. [c]Receipt of income maintenance (IM) spells was measured with two dichotomous variables to capture the beginning of IM support in relationship to the timing of the first maltreatment report. Additionally, IM support was measured as a continuous variable capturing the total number of spells received before the second (if existing) maltreatment report. Results from the random forest analysis assigned a higher G^2 to the continuous-level version of the IM variables, but both the dichotomous and continuous measures of IM receipt were in the top 80% of variables to be included in the neural network analysis, which hampered the performance of the neural network due to multicollinearity. Hence, the random forest analysis was re-run using only the continuous-level version of IM spell receipt. * p<0.05, ** p<0.01, and *** p<0.001 for all chi-square tests and 2-tail t-tests listed above.

Accuracy, Nonlinearity, and Utility: Assessing the Results of the Neural Network Analysis

By allowing a neural network to model what is not known about the functional form of the relationship between recurrent maltreatment and its predictors, the accuracy with which risk of repeat maltreatment is assessed may increase substantially. However, aside from assessing the benefits of a neural network analysis in terms of a misclassification rate and an ROC curve, the utility of a neural network can and should be addressed in several other ways. First, the predictive accuracy of a neural network should be compared with the predictive accuracy of a less complex linear model using the same predictors. This comparison is important in answering the following question: In terms of predictive accuracy, is the additional complexity of the neural network worth it? To facilitate this comparison, a binary logistic regression model with random intercepts[5] (to account for systematic unexplained variation among the 261 census tracts) was fitted to the data and the same set of predictors used in the neural network (Gill & Womack, in press; Gelman, 2006; Gelman & Hill, 2007). Results from the neural network were compared with the results from the simpler linear model by examining the misclassification rates and the areas under the ROC curves for both models (Beck, King, & Zeng, 2000, 2004; King & Zeng, 2001).

Second, beyond the potential that a nonlinear model may have for improving predictive accuracy, a nonlinear model with a superior performance as compared with a linear model may provide evidence that a linear functional form is too simple and unable to accurately represent the true relationship between recurrent maltreatment and its predictors. Specifically, in terms of estimating an unbiased model that represents the true relationship between recurrent maltreatment and its predictors, is the added complexity of the neural network model necessary? To facilitate an assessment of the ways in which the neural network's more complex functional form differs from the simpler linear functional form, two- and three-dimensional plots of the probability of recurrent maltreatment by each predictor (while holding all other predictors constant) for each model were compared (Beck, King, & Zeng, 2000, 2004; King & Zeng, 2001). Moreover, the predicted probability of recurrent maltreatment from the linear model was entered into a second neural network model as a predictor along with the original set of predictors (Stine, 2011). In theory, if the neural network has little to offer beyond the estimated relationships that a linear model can produce, then only one strong relationship in the new neural network will occur between the predicted probability of maltreatment as modeled by the neural network and the predicted probability of maltreatment as modeled by the binary logistic regression. All other plots would show a flat horizontal slope. On the other hand, if the relationship between recurrent maltreatment and its predictors is more complex than what a linear model would specify, then at least one or more of the slopes for the original set of predictors will have a shape that is closer to being curvilinear and a pitch that is steeper than what is produced by a horizontal line (Stine, 2011).

Third, beyond the predictive accuracy and utility of a neural network in modeling nonlinearity, how can a neural network inform practice? Key questions within an RNR perspective come down to who needs preventive service delivery, at what level of intensity, and for which dynamic risk factors. Key questions within a pattern recognition perspective come down to which combinations of predictors best separate observations into classes. Extracting information about the predictors as functions can inform the creation of risk-based treatment groups where the delivery of services is targeted towards the ways in which distinctive combinations of risk factors increase or decrease the probability of repeat maltreatment. In order to facilitate an examination of the clinical significance of the neural network results, a series of plots were created for the purpose of modeling the relationship between the probability of recurrent maltreatment and various predictors to determine (a) which predictors were strongest in separating the likelihood of being re-reported into high and low regions, (b) which predictors acted as moderators, and (c) which predictors were curvilinear. Finally, a regression tree was created for the purpose of identifying a set of risk-based groups by entering the neural network's predicted probability of recurrence as the response variable and the original set of predictors from the neural network as the input variables (Faraggi, LeBlanc, & Crowley, 2001). The regression tree used binary recursive splits to create groups with high and low average probabilities of repeat maltreatment as a function of a sequence of predictors (SAS, 2010).

Chapter 4: Results

How Did the Neural Network Do? Results According to Predictive Accuracy, Evidence of Nonlinearity in X, and Utility of Risk Prediction Within an RNR Perspective

As stated earlier, rather than testing explicitly defined hypotheses, this dissertation study is exploratory and is largely based on asking one simple question: How accurate is the specified neural network (i.e., the neural network created for this study) in classifying children into one of two maltreatment groups (i.e., 1 = will be re-reported and 0 = will not be re-reported) as a function of the combination of values for the input variables? That said, the need for post-hoc analyses that go beyond measures of predictive accuracy are important for broadly understanding which inputs explain the largest shifts in the probability of recurrence and how these key inputs are related to the probability of being re-reported for maltreatment. In order to provide a structured scope of inquiry for the post-hoc analyses that follow the reporting of results for the random forest and the neural network analyses, a second decision tree was used to place the predicted probabilities of maltreatment recurrence for each child into context (Faraggi, LeBlanc, & Crowley, 2001). Specifically, a regression tree (similar to a classification tree, but the outcome is continuous as opposed to categorical) was used to better understand how the set of predictors used in the neural network model account for variance in the probability of being re-reported for maltreatment that was estimated for each observation by the neural network model.

As described earlier, a classification tree uses each recursive binary split in order to increase the probability that a group of observations has the same target value for recurrent maltreatment (with observations split into two groups, where 1 = re-reported for maltreatment and 0 = not re-reported for maltreatment) as a function of values for selected input variables. In contrast, a regression tree uses each recursive binary split to reduce the residual sum of squares (i.e., unexplained variance) in the probability of recurrence by creating homogenous groups of observations characterized by low within-group variation in the probability of being re-reported for maltreatment and large between-group variation in the probability of being re-reported for maltreatment (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Fox, 2000; Hastie, Tibshirani, & Friedman, 2001; SAS, 2010; Stine, 2011). Instead of fitting a constant probability model to observations within each partition in input space (defined by the values of the predictor variables used to split observations into homogenous groups), each group of observations is fitted with a constant mean. Hence, each group of observations has an average likelihood of being re-reported for maltreatment as a function of the values for a specific combination of input variables.

As noted by Hastie, Tibshirani, and Friedman (2001), one of the benefits of a regression tree is its ease of interpretability; each partition in input space and hence each group's estimated average likelihood of recurrence is conditioned on an explicit set of input values that sequentially define the characteristics that best explain the expected probability of recurrence. The regression tree provides a set of directions that explain (with varying degrees of accuracy depending on the model's overall predictive accuracy) how a set of observations with an expected probability of recurrence arrived at that particular level of risk. The beauty of using a regression tree to "decode" the results for the neural network lies in the explicit creation of a set of decision rules that make it very clear how a group of observations was estimated as having a particular likelihood of recurrence. Faraggi, LeBlanc, and Crowley (2001) suggest using a regression tree to clarify the relationships between the predictors used in a neural network model and the estimated probability of recurrence that the neural network model produces for each observation as a result of learning how to produce a set of output values as a function of the input values. Hence, this method was employed in order to (a) better understand how the predictors in the neural network model relate to the likelihood of maltreatment recurrence, (b) identify a post-hoc step that links the prediction of recurrent maltreatment to the delivery of preventive services within an RNR perspective, and (c) provide a scope that focuses the post-hoc analyses of the neural network findings on the most important areas for effective treatment planning.

This section reports the results from the random forest and neural network analyses followed by the regression tree analysis just described above. The remaining portions of this section proceed to unpack and assess the utility of the neural network analysis relative to (a) its predictive accuracy compared with a linear model, (b) evidence of nonlinearity in X in comparison with a potentially biased linear model, and (c) curvilinear and interaction effects among key risk factors identified by the regression tree.

Feature Selection: Output from the Random Forest

As noted in the method chapter, a random forest of classification trees was used to identify a subset of inputs that were most effective in separating children into groups of observations for each target value; hence, the random forest was used to identify the inputs that would be most valuable in helping the neural network to learn the values of the target variable and estimate the conditional probabilities of class membership for a given set of input values and network weights. Observations in the total sample (N = 6,747) were assigned a number between 0 and 1 from the random uniform distribution that comprised the variable used to assign observations to the training and validation sets for the random forest analysis (i.e., train_valid); observations with values that fell on or between the smallest value of train_valid and ~0.85 were assigned to the training set while the remaining observations were assigned to the validation set. Hence, approximately 85% of the observations were randomly assigned to the training set (n = 5,725) and 15% of the observations were randomly assigned to the validation set (n = 1,022). Bootstrap samples were drawn with replacement from the training set and probabilities of maltreatment recurrence were estimated from each of the 64 classification trees (and then averaged across all classification trees) using 64 bootstrap samples (one for each tree). The validation data set was used to fine tune the final model by assessing model complexity and fit in relationship to a new set of observations (different from those contained in the bootstrap samples drawn from the training set) (Sall, 2009; SAS, 2010). Cases with missing observations (193 cases that comprised 2.9% of the total sample of observations) do not need to be dropped from classification tree analyses (Hastie, Tibshirani, & Friedman, 2001; SAS, 2010). In the event that parent gender or parent's potential status as the perpetrator (the only two predictors with missing values) are used as splitting variables, cases with missing values for the splitting variable are randomly assigned to each of the two resulting child nodes (SAS, 2010).

Although the JMP Pro 9 partition platform was programmed to create 100 classification trees, 64 trees were produced as a result of electing to use the early stopping rule wherein the total number of classification trees to be built was reduced to the total number of trees that were associated with ongoing improvements in the validation statistic entropy R^2 (where values closer to 1 indicate a better fit). Entropy R^2 fine tunes model-predicted probabilities of recurrence as a function of the inputs by comparing the log-likelihoods of the fitted probability model (probabilities of class membership across the observations fitted by the classification tree) and the constant probability model (the probability of class membership based upon the proportion of observations with a target value of 1 or a target value of 0 for the population of observations in the sample) as follows:

$$Entropy\ R^2 = 1 - \frac{log{-}likelihood_{fitted\ probabilit\ model}}{log{-}likelihood_{constant\ probability\ model}}$$ (SAS, 2010, p. 308).

A total of 32 inputs were entered into the model as predictors where 64 classification trees were fitted to the data; each tree was required to (a) split at least six times, (b) select the best splitting variable from among eight randomly selected predictors at each split, and (c) maintain at least 25 observations per node. The size of the training sample was privileged in order to provide the classification trees with the greatest opportunity to discover the data's underlying structure, and over-fitting was controlled by choosing conservative values for the minimum number of required splits and minimum number of observations per node (to control model complexity) (Stine, 2011). All model fit statistics measure the degree to which the model-fitted probabilities of conditional class membership (where the target value corresponding to the class with the highest probability of membership becomes the target value predicted by the model) agree with the actual target values (Sall, 2009; SAS, 2010).

The entropy R^2 (values closer to 1 indicate a better fit) was 0.18 for the training data and 0.12 for the validation data. The misclassification rate (values closer to 0 indicate a better fit) was 27.72% for the training set and 32% for the validation set; additionally, the area under the Receiver Operating Characteristic (ROC) curve (values closer to 1 indicate a better fit) was just over 0.70 for both the training and validation sets. The misclassification rate measures the proportion of observations with predicted target values (based upon the class with the highest probability of class membership) that do not match the actual target values (a smaller misclassification value indicates better model fit). The area under the ROC curve measures the classification efficiency of the model by plotting *true positives* on the *y*-axis (i.e., where the model predicts that the child was re-reported for maltreatment with a target value of 1 and the child was in fact re-reported for maltreatment with a real target value of 1) by *false positives* on the *x*-axis (i.e., where the model predicts that the child was re-reported for maltreatment and the target value = 1 when the child was not, in fact, re-reported for maltreatment and the target value actually = 0). Perfect model performance is equal to 1 and model performance that is no better than random chance is equal to 0.50 (SAS, 2010).

Confusion matrices (one for the training set and one for the validation set) provide more detail about the model's predictive accuracy through a contingency table that assesses the extent to which model-predicted target values and the actual target values agree and disagree. The cell that spans the top left portion of the contingency table (please see Table 4.1 below for an example of a confusion matrix) contains the number of observations that are *true positives* (TP) -- the number of children whose predicted target value of 1(predicted as having a maltreatment re-report) agreed with the actual target value of 1 (re-reported for maltreatment); the cell that spans the top right portion of the contingency table contains the number of observations that are *false positives* (FP) -- the number of children whose predicted target value of 1 (predicted as having a maltreatment re-report) did not agree with the actual target value of 0 (not re-reported for maltreatment). The bottom left portion of the contingency table contains the number of observations that are *false negatives* (FN) -- the number of children whose predicted target value of 0 (predicted as not having a maltreatment re-report) did not agree with the actual target value of 1 (re-reported for maltreatment); the bottom right portion of the contingency table contains the number of observations that are *true negatives* (TN) -- the number of children whose predicted target value of 0 (predicted as not having a maltreatment re-report) agreed with the actual target value of 0 (not re-reported for maltreatment).

The accuracy of the model's target function as applied to new observations can be estimated by assessing the model's validation set *sensitivity* and *specificity*. In short, *sensitivity* is the model's probability of correctly identifying children who will be re-reported (i.e., the true positive cases) [TP/(TP + FN)], and *specificity* is the model's probability of correctly identifying children who will not be re-reported (i.e., true negative cases) [TN/(FP + TN)] (Hastie, Tibshirani, & Friedman, 2001; Shlonsky & Wagner, 2005). Based on the confusion matrix reported for the validation set, the sensitivity of the model was .80, which means that the model correctly identified 80% of the true positive or true high risk cases. Conversely, the specificity of the model was 0.53, which means that the model correctly identified 53% of the true negative or true low risk cases. Tables 4.1 and 4.2 contain confusion matrices for the training and validation sets.

Table 4.1

Confusion Matrix for the Random Forest Training Data

Predicted	**Training Set**	**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	2722 (True Positives)	993 (False Positives)
Not Re-Reported (0)	594 (False Negatives)	1416 (True Negatives)

Table 4.2

Confusion Matrix for the Random Forest Validation Data

	Validation Set	
Predicted		**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	453 (True Positives)	216 (False Positives)
Not Re-Reported (0)	111 (False Negatives)	242 (True Negatives)

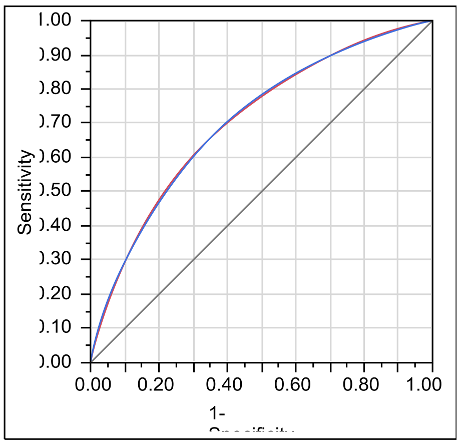
Final measures of model fit include two error-based measures – i.e., the root mean square error (RMSE) and the mean absolute deviation -- where error is measured as the difference between the actual target value and the estimated probability of class membership for the target value (e.g., if the actual target value was 1, what was the observation's estimated probability of having a target value of 1 and therefore being a member of the recurrent maltreatment class). Equations for the RMSE and mean absolute deviation (Mean Abs Dev) are as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} y_i\,[j_i] - p_i[j_i]^2/n}$$

$$Mean\ Abs\ Dev = \sum_{i=1}^{n} |y_i[j_i] - p_i[j_i]|/n,$$

where n = observations divided among the training and validation sets, $y_i[j_i]$ = the actual target value for recurrent maltreatment, and $p_i[j_i]$ = the model-estimated probability of class membership for the recurrent maltreatment group represented by the actual target value (where 1 = the group where children received a re-report, and 0 = the group where children did not receive a re-report) (SAS, 2010, p. 308). The RMSE (smaller values indicate a better fit) was 0.43 for the training set and 0.46 for the validation set; the mean absolute deviance (smaller values indicate a better fit) was 0.40 for the training data and 0.42 for the validation data. Overall, the model fit was adequate; Figures 4.1 and 4.2 summarize all model fit statistics discussed above.

Random Forest ROC Curve for Training Data



Model Fit Statistics for Training Data

Entropy R^2 0.18

RMSE 0.43

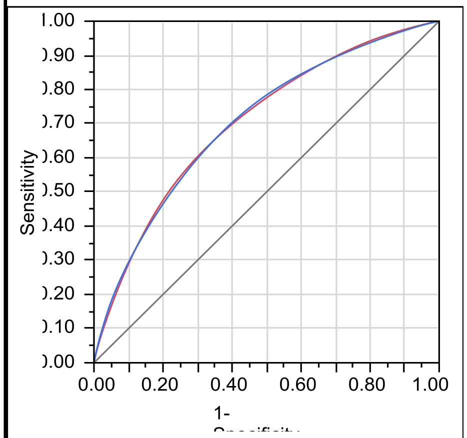
Mean Absolute Deviation 0.40

Misclassification Rate 0.28

Legend	REC1	AUC
	0	0.70
	1	0.70

Figure 4.1. Random forest training data model fit statistics and the area under the ROC curve (AUC).

Random Forest ROC Curve for Validation Data



Model Fit Statistics for Validation Data

Entropy R^2 0.12

RMSE 0.46

Mean Absolute Deviation 0.42

Misclassification Rate 0.32

Legend	REC1	AUC
———	0	0.70
———	1	0.70

Figure 4.2. Random forest validation data model fit statistics and the area under the ROC curve (AUC).

Beyond model fit statistics, input variable importance is measured through the column contribution function. All inputs were ranked ordered (in descending order) by the extent to which each predictor contributed to the estimation of conditional class probabilities of maltreatment recurrence as measured by the magnitude of values for the G^2 and LogWorth statistics (where larger values indicate greater variable importance) (SAS, 2010). Input variables were selected for inclusion or exclusion from the neural network model on the basis of the rank ordered variable importance. Input variables were selected for inclusion in the neural network model if they were included in the top 80% of the ordered G^2 values; conversely, input variables were selected for exclusion from the neural network model if they were in the bottom 20% of the ordered G^2 values (Diaz-Uriarte & de Andrés, 2006; Genuer, Poggi, Tuleau-Malot, 2010). The seven predictors located at the bottom of Figure 4.3 were excluded from the neural network analysis (counting from the bottom of the table upward). Figure 4.3 summarizes the G^2 for each predictor in descending order as well as the number of times each predictor acted as a splitting variable.

In addition to excluding seven predictors from the neural network analysis, three additional predictors [i.e., (a) worker-observed family characteristic that the primary caregiver was a single parent, (b) receipt of FCS within 45 days of the first maltreatment report, and (c) primary caregiver's age at the first maltreatment event] were dropped from the neural network analysis in order to avoid problems associated with multicollinearity (Haykin, 1999). Single parent was strongly associated with family protective factors [$\chi^2 = 3909.55$ (df =1, N = 6747), $p < .001$]. FCS receipt within 45 days of the first maltreatment report was strongly associated with FCS receipt on or after the first maltreatment report and before the second report (if a second report occurred)

Column Contributions

Term	Number of Splits	G^2
Num_IM_Sp	302	10172.063
Juv_Ct_Pr_02	167	8809.355
Comm_Pov	324	4920.234
Ch_Age	420	3908.483
Spec_Ed_Pr_02	166	3237.185
Ch_MHSA_Pr_02	177	2992.086
Pt_Age_CAN01	317	2193.304
FCS	159	1667.959
Pt_Age	295	1558.892
Pt_Perp	192	1313.073
Comm_Move	229	1293.855
Pt_Educ	351	1269.697
Substantiation	163	931.087
Ch_Race	217	915.771
Perp_Protective	304	721.736
Ch_Birth_Prb	119	708.130
Fam_Protective	192	651.062
Ch_Gender	292	628.033
Fam_Single_Pt	152	490.616
FCS_Wtn_45Days	82	419.292
Pt_MHSA_Pr_02	14	231.422
Mom_Hx_Fost_Care	23	150.327
Pt_Gender	44	138.629
Spec_Ed_Pr_01	39	100.926
FPS	21	100.690
Ch_Conflict_Svc_Pr_01	27	78.600
Ch_Conflict_Svc_Pr_02	31	77.156
Ch_MHSA_Pr_01	16	68.504
Perp_Neg_Care_Skills	26	56.734
Crim_Pr_02	8	51.165
Crim_Pr_01	3	9.751
ALT	1	3.652

Figure 4.3. Random forest input variable importance measured by rank ordering in descending order each input's G^2 statistic.

[χ^2 = 3222.17 (*df* =1, *N* = 6747), *p* < .001]. The primary caregiver's age at the first maltreatment report was moderately associated with the child's age at the first maltreatment report [*r* = 0.44 (*N* = 6747), *p* < .001] and strongly associated with the primary caregiver's age at the birth of the child [*r* = 0.89 (*N* = 6747), *p* < .001]. Single parent family, FCS receipt within 45 days of the first maltreatment report, and the primary caregiver's age at the first maltreatment report appeared to be providing redundant information despite earlier attempts to (a) use principal components analysis to represent the largest portion of variance among worker-observed family characteristics with a smaller set of features, and (b) to use a random forest analysis to identify a subset of predictors whose combination of input values are most effective in separating case records by their observed target values (Bishop, 1995; Cohen, Cohen, West, & Aiken, 2003; Haykin, 1999; Schwartz, Jones, Schwartz, & Obradovic, 2008). Moreover, no additional attempts to combine any of the three redundant predictors with a subset of the remaining 22 predictors appeared to be feasible given the general lack theoretical support and even basic face validity (i.e., a conceptual grasp of how certain predictors could fit together). Hence, single parent family, FCS receipt within 45 days of the first maltreatment report, and the primary caregiver's age at the first maltreatment report were excluded from the neural network model.

Neural Network Analysis: Output Regarding Predictive Accuracy

A neural network model with 22 input nodes, 8 hidden nodes (one layer of hidden nodes), and one output node estimated a target function that classifies children into one of two possible risk of recurrent maltreatment groups as a function of the input values. The target function's predictive accuracy is determined by the degree to which the neural network was able to identify patterns of input values that differentiated children who were likely to be re-reported for maltreatment from children who were not likely to be re-reported for maltreatment. As noted by Bishop (1995), conditional probabilities of class membership are estimated and compared

$$P(C_k|\boldsymbol{x};\boldsymbol{\theta})$$

$$P(C_j|\boldsymbol{x};\boldsymbol{\theta}),$$

where $k \neq j$; k = a target value of 1, which signifies membership in the recurrence group;

j = a target value of 0, which signifies membership in the no recurrence group; $\boldsymbol{x}$ = a vector of input values or the pattern of input values for a given observation; and $\boldsymbol{\theta}$ = the vector of network weights that relate patterns of input values to an estimated target value. The group with the highest probability of membership is the group to which the observation is assigned; the model is then used to classify new observations by assigning each observation to a class of target values that is associated with the observation's pattern of input values (Bishop, 1995).

In cases where the neural network is not able to learn patterns of inputs values that pull the classes of target values far apart, the classes of target values overlap. In this case, the decision boundary is placed between the decision regions (where each respective class of target values and their corresponding patterns of input values are located) runs right through the point at which the classes intersect (placing the decision boundary in this location minimizes the probability of misclassification) (Bishop, 1995). A high degree of overlap increases the probability of misclassification, where the class conditional probabilities of group membership might differ by as little as 2% (e.g., where the estimated probability of C_k = 49% and the estimated probability of C_j = 51%). In the case of the neural network created for this dissertation study, the decision boundary equals 0.5, and the class conditional probability of group membership that surpasses 0.5 is the group of target values to which the observation is assigned; hence, the cut point used to differentiate high risk from low risk cases is equal to 0.5. Measures of predictive accuracy are then used to assess the neural network's classification performance relative to a cut point that does not inherently require an observation's probability of being a high risk case to differ substantively (numerically speaking) from its probability of being a low risk case.

In order to explore the effects of three different cut points on the neural network's classification performance, measures of predictive validity are examined in relationship to three different cut points. The first model set the cut point at 0.5. The second model adjusted the prediction formula created by the neural network where the formula predicts the probability that the target value is equal to 1 (maltreatment recurrence); cases that had a model-estimated probability running from 0.1 to 0.45 were assigned a predicted target value of 0, and cases that had a model-estimated probability from 0.55 to 0.99 were assigned a predicted target value of 1. The third model adjusted the prediction formula created by the neural network where the formula predicts the probability that the target value is equal to 1 (maltreatment recurrence); cases that had a model-estimated probability running from 0.1 to 0.40 were assigned a predicted target value of 0, and cases that had a model-estimated probability from 0.60 to 0.99 were assigned a predicted target value of 1. While the cut points changed across the three models presented below, the target function that was estimated by the neural network and the class conditional probabilities of group membership produced by the target function were not altered.

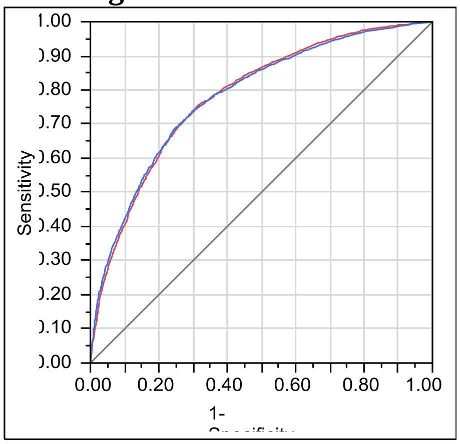
The estimated target function and the final specified form of the neural network (i.e., the number of hidden nodes) were iteratively produced by selecting the smallest number of hidden nodes that yielded the largest area under the ROC curve (i.e., one that did not appreciably deteriorate across the training and validation sets), the lowest misclassification rate, and the smallest negative log-likelihood for the *validation data*. Concurrently, different sets of random starting values for the network weights were requested, resulting in the generation of a number of models that were then examined in conjunction with the criteria described above (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001). A full range of model fit statistics was used to evaluate the neural network's ability to correctly predict the actual values of the target variable in relationship to the estimated conditional probabilities of class membership. Model fit statistics include the *entropy* R^2 (values closer to 1indicate the neural network's contribution to estimating probabilities of class membership beyond the constant probability model), the area under the *ROC curve* (values closer to 1 indicate the model's ability to differentiate true positive or high risk cases from false positive or low risk cases), the *negative log-likelihood* (smaller values indicate a minimization of the cross entropy error function and maximization of the log-likelihood of the data), the *misclassification rate* (values closer to 0 indicate higher agreement between the predicted probabilities of class membership and the actual target values), the *root mean square error* (smaller values indicate higher agreement between the predicted probabilities of class membership and the actual target values), and the *mean absolute deviation* (smaller values indicate higher agreement between the predicted probabilities of class membership and the actual target values).

The total number of cases spanning the training and validation data sets was 6,554, where 193 cases (which comprised 2.9% of the total number of observations) were dropped due to missing values for no more than two variables (parent gender and parent's potential status as the perpetrator). Observations were randomly assigned into 10 equal portions or folds and 10-fold cross validation was used to (a) estimate the target function across 10 different training sets (each created by training a different neural network model on 10-1 folds), and (b) fine-tune model complexity and estimate the target function's generalizability across 10 different validation sets (validation occurred in relationship to the one fold that was held back from each of the 10 different training sets) (Bishop, 1995; Hastie, Tibshirani, & Friedman, 2001; Haykin, 1999). The model with the best performance across the 10 different samples (where nine folds were used for training and one fold was used for validation) is reported by JMP Pro 9.

Neural network model with a cut point of 0.5.

As summarized by Figures 4.4 and 4.5 below, the neural network demonstrated an acceptable level of predictive accuracy with an area under the ROC curve of 0.78 for both the training and validation data and a misclassification rate of 27.7% for both training (n = 5,935) and validation (n = 638) data. The confusion matrices for the training and validation data as summarized by Tables 4.3 and 4.4 below help to pinpoint where the misclassification occurred. In cases where the pattern of inputs values did not sufficiently differentiate children who were re-reported from those who were not re-reported, the neural network was more likely to over-estimate the probability of being re-reported (obtaining a higher proportion of false positives) as opposed to under-estimating the probability of being re-reported (obtaining a smaller proportion of false negatives). In cases of risk over-estimation (i.e., false positives), the class conditional probability of being re-reported, where $y = 1$, reaches a value that is greater than 0.5, when in fact the estimated probability of being re-reported should have been a value that was less than 0.5. Conversely, in cases of risk under-estimation (i.e., false negatives), the class conditional probability of not being re-reported, where $y = 0$, reaches a value that is greater than 0.5, when in fact the estimated probability of not being re-reported should have been a value that was less than 0.5. Based on the confusion matrix reported for the validation set, the sensitivity of the model was .75, which means that the model correctly identified 75% of the true positive or true high risk cases. Conversely, the specificity of the model was 0.67, which means that the model correctly identified 67% of the true negative or true low risk cases.

Neural Network ROC Curve for Training Data



Legend	REC1	AUC
red line	0.1	0.78
blue line	0.9	0.78

Model Fit Statistics for Training Data

Entropy R^2 0.19

RMSE 0.43

Mean Absolute Deviation 0.37

Misclassification Rate 0.28

-Log-Likelihood 32840.81

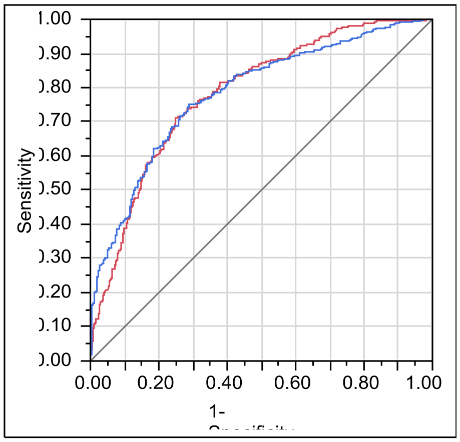
Figure 4.4. Neural network training data model fit statistics and the area under the ROC curve (AUC).

Table 4.3

Confusion Matrix for the Neural Network Training Data

	Training Set	
Predicted		**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	2662 (78.53%) (True Positives)	728 (21.47%) (False Positives)
Not Re-Reported (0)	913 (36.14%) (False Negatives)	1613 (63.86%) (True Negatives)

Neural Network ROC Curve for Validation Data



Legend	REC1	AUC
red line	0.1	0.78
blue line	0.9	0.78

Model Fit Statistics for Validation Data

Entropy R^2 0.19

RMSE 0.43

Mean Absolute Deviation 0.37

Misclassification Rate 0.28

-Log-Likelihood 350.25

Figure 4.5. Neural network validation data model fit statistics and the area under the ROC curve (AUC).

Table 4.4

Confusion Matrix for the Neural Network Validation Data

Predicted	**Validation Set**	**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	300 (79.37%) (True Positives)	78 (20.63%) (False Positives)
Not Re-Reported (0)	99 (38.08%) (False Negatives)	161 (61.92%) (True Negatives)

Neural network model with cut points of 0.45 and 0.55.

In order to determine if altering the cut points would reduce some of the ambiguity and therefore misclassification caused by cases that were located in overlapping distributions of class conditional probabilities, cut points were reassigned to cases where those with an estimated probability of recurrence from 0.1 to 0.45 were predicted as having a target value of 0 and cases with an estimated probability of recurrence from 0.55 to 0.99 were predicted as having a target value of 1. Cases that fell in between the cut points were deleted from the subsequent ROC analysis and confusion matrices that follow (n = 699). Cut points were altered in relationship to the estimated target function -- i.e., the formula that was used to predict the probability of recurrence, where y = 1 for all cases (N = 5,855). By altering the cut points, the sensitivity of the model increased to 0.77, and the specificity increased to 0.71. The model correctly identified 77% of the true positive or true high risk cases and 71% of the true negative or true low risk cases. Moreover, the predictive validity of the model increased as evidenced by the area under the ROC curve that increased to 0.79. Figure 4.6 contains the confusion matrix and ROC curve for observations with predicted values based upon the new cut points of 0.45 and 0.55.

Predicted	Validation Set	Actual
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	2800 (81.30%) (True Positives)	644 (18.70%) (False Positives)
Not Re-Reported (0)	850 (35.26%) (False Negatives)	1561 (64.74%) (True Negatives)

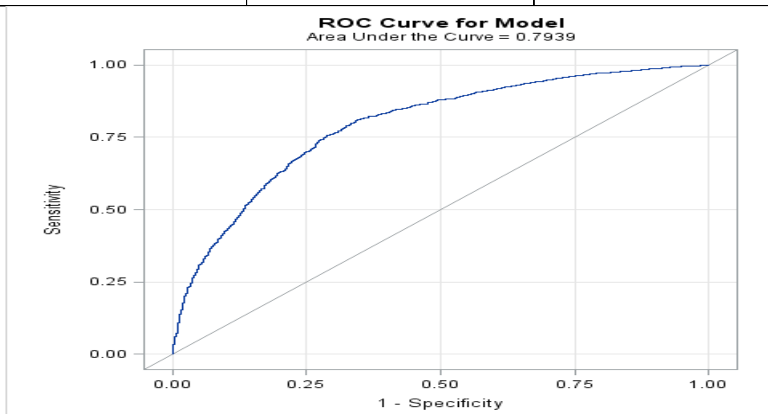


Figure 4.6. Confusion matrix and AUC for altered cut points, where predicted probabilities of 0.1 to 0.45 = a predicted target value of 0 or no maltreatment recurrence and predicted probabilities of 0.55 to 0.99 = a predicted target value of 1 or maltreatment recurrence.

Neural network model with cut points of 0.40 and 0.60.

In order to determine if altering the cut points would reduce some of the ambiguity and therefore misclassification caused by cases that were located in overlapping distributions of class conditional probabilities, cut points were reassigned to cases where those with an estimated probability of recurrence from 0.1 to 0.40 were predicted as having a target value of 0 and cases with an estimated probability of recurrence from 0.60 to 0.99 were predicted as having a target value of 1. Cases that fell in between the cut points were deleted from the subsequent ROC analysis and confusion matrices that follow (n = 1,450). Cut points were altered in relationship to the estimated target function -- i.e., the formula that was used to predict the probability of recurrence, where

y = 1 for all cases (N = 5,104). By altering the cut points for a second time, the sensitivity increased again to 0.79, and the specificity increased as well to 0.73. The model correctly identified 79% of the true positive or true high risk cases and 73% of the true negative or true low risk cases. Moreover, the predictive validity of the model increased for a second time as evidenced by an area under the ROC curve that increased to 0.81. Figure 4.7 contains the confusion matrix and ROC curve for observations with predicted values based upon the new cut points of 0.40 and 0.60.

Validation Set		
Predicted		**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	2586 (84.07%) (True Positives)	490 (15.93%) (False Positives)
Not Re-Reported (0)	691 (34.07%) (False Negatives)	1337 (65.93%) (True Negatives)

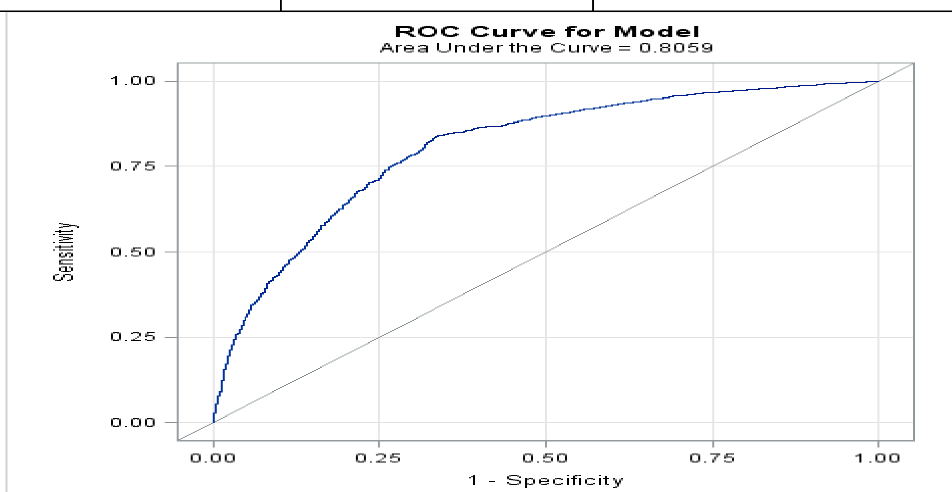


Figure 4.7. Confusion matrix and AUC for altered cut points, where predicted probabilities of 0.1 to 0.40 = a predicted target value of 0 or no maltreatment recurrence and predicted probabilities of 0.60 to 0.99 = a predicted target value of 1 or maltreatment recurrence.

Comparing the Predictive Accuracy of the Neural Network with a Standard Linear Model: Whose Area Under the ROC Curve Is Bigger?

A binary logistic regression model was fitted to the same response and predictor variables used in the neural network to provide a counter point for assessing the relative benefits of using a more complicated model to predict the likelihood of recurrent maltreatment (Beck, King, & Zeng, 2000, 2004; King & Zeng, 2001; Zeng, 1999). As stated earlier, the neural network makes no assumptions about linearity in the predictors but a basic linear regression model (e.g., a binary logistic regression model) does assume linearity in the predictors. That said, a two-level model with random intercepts was used to (a) account for dependency among observations located in the same census tract (with 261 total census tracts), and (b) model any systematic unexplained variation in the probability of recurrent maltreatment that is attributable to differences between census tracts (Gelman, 2006; Gelman & Hill, 2007; Gill & Womack, in press). Hence, in addition to specifying a level one regression model where the probability of maltreatment was regressed on a wide range of individual-level (i.e., child and primary caregiver) risk and protective factors, a level two regression model was specified where the random intercept coefficient was regressed on two census tract-level predictors (i.e., concentrated poverty and residential mobility). The second level regression equation was used to model systematic variation between census tracts while accounting for unexplained variation across census tracts. The second level regression equation thus makes it possible to fit a regression model to individual-level measures (i.e., child and primary caregiver) while (a) accounting for unexplained variation in the probability of recurrent maltreatment that is attributable to differences between community contexts (as measured by residence in a given census tract), and (b) modeling the contextual variation in relationship to key community-level covariates (Gelman, 2006; Gelman & Hill, 2007; Gill & Womack, in press).

The binary logistic regression model with random intercepts assumes a binomial distribution for maltreatment recurrence and a normal distribution for the random intercept with a mean of zero and estimated variance [$\sim N(0, \sigma^2)$]. The estimated variance of the random intercept is referred to as the variance component and can be used in the calculation of the intraclass correlation coefficient (ICC), which estimates the proportion of variance in the probability of maltreatment that is attributable to differences between census tracts (Snijders & Bosker, 1999). Treating census tract membership as a random effect causes the intercepts (each census tract has its own intercept) to vary over children, therefore creating a subject specific effect in which children may differ in their likelihood of being re-reported for maltreatment because of the census tract in which they live. The census tracts within this study are treated as a random sample of all possible census tracts (Goldstein, 1995; Snijders & Bosker, 1999). The model used to estimate the probability of maltreatment recurrence in relationship to individual- and community-level covariates can be represented using matrix notation as follows:

$$\Pr(y_i = 1) = logit^{-1}(\alpha_{j_i} + \boldsymbol{X}_i \beta)$$

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2),$$

where α_{j_i} = the intercept for the i^{th} child in census tract *j*, providing all children living in census tract *j* with a common intercept, $\boldsymbol{X}$ includes all individual-level and community-level explanatory variables, and $N(\mu_\alpha, \sigma_\alpha^2)$ = a common normal distribution for the random intercepts with mean μ_α and standard deviation σ_α^2 (Gelman & Hill, 2007, pp. 263 and 302).

The binary logistic regression model with random intercepts was iteratively fitted to the data across three models. Model 1 is a two-level hierarchical null or intercept-only model with children (Level 1) nested within census tracts (Level 2). The null model was used to estimate the intraclass correlation coefficient (ICC) in order to determine the proportion of unexplained variation in the probability of repeat maltreatment that was attributable to differences between census tracts (Goldstein, 1995; Snijders & Bosker, 1999). Model 2 extends Model 1 by adding all Level 1 (child and primary caregiver) variables. Model 2 was used to estimate the main effects of child- and primary caregiver-level variables on the probability of recurrent maltreatment while controlling for individual-level variables that may account for variation between census tracts (thus taking into account compositional characteristics of the census tract). Model 3 extends Model 2 by adding the two census tract-level variables (i.e., concentrated poverty and residential mobility). Model 3 was used to (a) estimate the effects of census-tract level variables on the probability of recurrent maltreatment, and (b) account for unexplained census tract variations in the probability of recurrent maltreatment (Goldstein, 1995; Snijders & Bosker, 1999). It should be noted that the effect on the probability of maltreatment for a one-unit change in a census tract-level variable applies to all children nested within census tract j. Hence, the effect of a census tract-level predictor varies by census tract and not by individual child (Gill & Womack, in press).

Models 1 through 3 were estimated with PROC GLIMMIX in SAS 9.3; model parameters were estimated with maximum likelihood using Gauss-Hermite quadrature and empirical (i.e., sandwich) standard errors (as opposed to model-based standard errors) were requested. As noted above, the binary logistic regression model with random intercepts was fitted to the same data (and therefore the same response and predictor variables) used in the neural network analysis (*N* = 6,554). While the random intercept or Level 2 variance component ($\tau_{\alpha}^{2} = 0.1559,\ \sigma_{\alpha}^{2} = 0.02945$) was statistically significant [χ^2 = 73.07 (*df* = 1, *N* = 6,554), $p < .001$], the ICC $\left[\frac{\tau_{\alpha}^{2}}{\tau_{\alpha}^{2} + \frac{\pi^2}{3}}\right]$ (Snijders & Bosker, 1999) or the proportion of unexplained variation in the probability of recurrent maltreatment attributable to differences between census tracts was quite low at 0.73%. After entering the individual-level variables in the model, the random intercept variance component ($\tau_{\alpha}^{2} = 0.03100,\ \sigma_{\alpha}^{2} = 0.01986$) remained statistically significant [χ^2 = 3.72 (*df* = 1, *N* = 6,554), $p < .05$], but the proportion of unexplained variation in the probability of recurrent maltreatment attributable to differences between census tracts decreased to an even smaller amount of 0.03%. After entering the census tract-level variables to the model, the random intercept variance component ($\tau_{\alpha}^{2} = 0.003087,$ $\sigma_{\alpha}^{2} = 0.01347$) was no longer statistically significant [χ^2 = 0.05 (*df* = 1, *N* = 6,554), p = 0.4135]. Table 4.5 summarizes the fixed effect parameter estimates and the random intercept estimates. Table 4.6 summarizes and compares the odds ratios for the regression of recurrent maltreatment on the individual-level covariates only (Model 2) and on the individual-level and the community-level covariates (Model 3).

Table 4.5

Regression of Recurrent Maltreatment on the 22 Risk and Protective Factors Using a Binary Logistic Regression Model with Random Intercepts

Parameter	Model 1	Model 2	Model 3
Fixed Effects			
Intercept	0.22*** (0.04)	0.56*** (0.11)	0.64*** (0.11)
Level 1 (Individual Characteristics)			
Child Characteristics			
Ch_Gender		-0.12* (0.05)	-0.13* (0.05)
Ch_Race		0.17* (0.07)	-0.00 (0.07)
Ch_Age		-0.39*** (0.04)	-0.38*** (0.04)
Ch_Birth_Prb		0.25** (0.09)	0.23** (0.09)
Primary Caregiver Characteristics			
Pt_Gender		-0.17 (0.10)	-0.13 (0.10)
Pt_Age		-0.04 (0.03)	-0.04 (0.03)
Pt_Educ		-0.23*** (0.05)	-0.20*** (0.05)
Mom_Hx_Fost_Care		0.12 (0.15)	0.13 (0.15)
First Maltreatment Incident Characteristics			
Substantiation		0.42*** (0.08)	0.41*** (0.08)
Pt_Perp		0.56*** (0.07)	0.54*** (0.06)
Worker-Observed Family Characteristics			
Fam_Protective		-0.20** (0.06)	-0.21** (0.06)

Note. Parameter estimate (logit) listed first followed by the standard error in parentheses.
* p<0.05, ** p<0.01, and *** p<0.001.

Table 4.5

Regression of Recurrent Maltreatment on the 22 Risk and Protective Factors Using a Binary Logistic Regression Model with Random Intercepts (Continued)

Parameter	Model 1	Model 2	Model 3
Level 1 (Individual Characteristics)			
Worker-Observed Perpetrator Characteristics			
Perp_Protective		-0.25*** (0.06)	-0.18** (0.06)
Cross-Sector Service Characteristics			
FCS		-0.84*** (0.09)	-0.84*** (0.09)
FPS		0.72*** (0.16)	0.70*** (0.16)
Ch_MHSA_Pr_02		-0.53*** (0.08)	-0.53*** (0.08)
Pt_MHSA_Pr_02		-1.22*** (0.21)	-1.25***(0.22)
Num_IM_Sp		0.43*** (0.04)	0.38*** (0.04)
Juv_Ct_Pr_02		-1.44*** (0.11)	-1.44*** (0.11)
Spec_Ed_Pr_01		0.52*** (0.12)	0.52*** (0.12)
Spec_Ed_Pr_02		-0.66*** (0.08)	-0.66*** (0.08)
Level 2 (Community Characteristics)			
Comm_Move			0.01 (0.03)
Comm_Pov			0.23*** (0.04)
Random Parameters			
Level 2			
Intercept/ intercept (σ^2_{a0})	0.16*** (0.03)	0.031* (0.02)	0.00 (0.01)
-2log likelihood	8,865.02	7,711.89	7,669.36
N	6,554	6,554	6,554

Note. Parameter estimate (logit) listed first followed by the standard error in parentheses.
* p<0.05, ** p<0.01, and *** p<0.001.

Table 4.6

Odds Ratios from the Regression of Recurrent Maltreatment on Individual-Level Covariates (Model 2) and Individual- and Community-Level Covariates (Model 3)

Parameter	Odds Ratio Individual-Level	*df*	Odds Ratio Individual and Community-Levels	*df*
Fixed Effects				
Level 1 (Individual Characteristics)				
Child Characteristics				
Ch_Gender	0.88*	6,276	0.88*	6,274
Ch_Race	1.18*	6,276	1.00	6,274
Ch_Age	0.68***	6,276	0.69***	6,274
Ch_Birth_Prb	1.28**	6,276	1.26**	6,274
Primary Caregiver Characteristics				
Pt_Gender	0.85	6,276	0.88	6,274
Pt_Age	0.96	6,276	0.96	6,274
Pt_Educ	0.79***	6,276	0.82***	6,274
Mom_Hx_Fost_Care	1.13	6,276	1.14	6,274
First Maltreatment Incident Characteristics				
Substantiation	1.51***	6,276	1.51***	6,274
Pt_Perp	1.74***	6,276	1.72***	6,274
Worker-Observed Family Characteristics				
Fam_Protective	0.82**	6,276	0.81**	6,274

Table 4.6

Odds Ratios from the Regression of Recurrent Maltreatment on Individual-Level Covariates (Model 2) and Individual- and Community-Level Covariates (Model 3) (Continued)

Parameter	Odds Ratio Individual-Level	*df*	Odds Ratio Individual and Community-Levels	*df*
Fixed Effects				
Level 1 (Individual Characteristics)				
Worker-Observed Perpetrator Characteristics				
Perp_Protective	0.78***	6,276	0.84**	6,274
Cross-Sector Service Characteristics				
FCS	0.43***	6,276	0.43***	6,274
FPS	2.05***	6,276	2.00***	6,274
Ch_MHSA_Pr_02	0.59***	6,276	0.59***	6,274
Pt_MHSA_Pr_02	0.30***	6,276	0.29***	6,274
Num_IM_Sp	1.53***	6,276	1.46***	6,274
Juv_Ct_Pr_02	0.24***	6,276	0.24***	6,274
Spec_Ed_Pr_01	1.68***	6,276	1.69***	6,274
Spec_Ed_Pr_01	0.52***	6,276	0.52***	6,274
Level 2 (Community Characteristics)				
Comm_Move	n/a	n/a	1.01	6,274
Comm_Pov	n/a	n/a	1.26***	6,274

Note. *N* = 6,554.

For those variables that were significantly associated with the likelihood of being re-reported for maltreatment, the interpretation of a given fixed effect parameter estimate (e.g., the effect of having a juvenile court petition on the odds of recurrent maltreatment) is conditional on holding both the census tract in which the child was residing as well as the values for all other covariates constant. In this way, the random effects model is said to produce *subject-specific* regression coefficients because the interpretation of a given fixed effect regression coefficient depends on holding each child's random intercept or census tract fixed. Thus, a proper interpretation of the fixed effect regression coefficient compares the odds of recurrent maltreatment for two children who lived in the same census tract with the same covariates, but who differ by one unit in the covariate of interest (Fitzmaurice, Laird, & Ware, 2004; Guo & Zhao, 2000; Hu, Goldberg, Hedeker, Flay, & Pentz, 1998; Hubbard et al., 2010; Larsen & Merlo, 2005).

For key individual-level predictors of interest (these variables are identified through a post-hoc regression tree analysis and are discussed in forthcoming sections of this chapter), the following interpretations of the fixed effect regression coefficients apply. Child age at the first maltreatment report was significantly and negatively associated with the likelihood of being re-reported for maltreatment. For children living in the same census tract, for every one standard unit of increase in the child's age at the time of his/her first maltreatment report (child age was calculated as a *z*-score), the child was 31% less likely to be re-reported for maltreatment, holding all other covariates constant (β = -0.38, *SE* = 0.04, OR = 0.69). The total number of income maintenance spells received was significantly and positively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, for every one standard unit of increase in the number of income maintenance spells received (receipt of income maintenance spells was calculated as a *z*-score) , the child was 46% more likely to be re-reported for maltreatment, holding all other covariates constant (β = 0.38, *SE* = 0.04, OR = 1.46). The primary caregiver's status as the perpetrator of the first maltreatment incident was significantly and positively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, the child whose primary caregiver was found to be the perpetrator of the first maltreatment incident was 72% more likely to be re-reported for maltreatment as compared with the child whose primary caregiver was not found to be the perpetrator of the first maltreatment incident, holding all other covariates constant (β = 0.54, *SE* = 0.06, OR = 1.72).

Receipt of a juvenile court petition on or after the first maltreatment report but before the second report (if a second report occurred) was significantly and negatively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, the child who was issued a juvenile court petition was 76% less likely to be re-reported for maltreatment as compared with the child who was not issued a juvenile court petition, holding all other covariates constant (β = -1.44, *SE* = 0.11, OR = 0.24). Receipt of special education eligibility on or after the first maltreatment report but before the second report (if a second report occurred) was significantly and negatively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, the child who was eligible for special education was 48% less likely to be re-reported for maltreatment as compared with the child who was not eligible for special education, holding all other covariates constant (β = -0.66, *SE* = 0.08, OR = 0.52). Receipt of a family-centered (FCS) service spell on or after the first maltreatment report but before the second report (if a second report occurred) was significantly and negatively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, the child who received an FCS spell was 57% less likely to be re-reported for maltreatment as compared with the child who did not receive an FCS spell, holding all other covariates constant (β = -0.84, *SE* = 0.09, OR = 0.43). The child's receipt of a mental health/substance abuse service on or after the first maltreatment report but before the second report (if a second report occurred) was significantly and negatively associated with the likelihood of being re-reported for child maltreatment. For children living in the same census tract, the child who received a mental health/substance abuse service was 41% less likely to be re-reported for maltreatment as compared with the child who did not receive a mental health/substance abuse service, holding all other covariates constant (β = -0.53, *SE* = 0.08, OR = 0.59).

Interpretation of a statistically significant census-tract level covariate is not as simple as interpreting the fixed effect of an individual-level covariate. While individual-level covariates vary within census tracts, thus facilitating a comparison of the likelihood of repeat maltreatment within the community-based cluster, variables measured at the census tract level do not vary within each community-based cluster. Instead, census tract-level covariates vary between census tracts. Statistically significant census-tract covariates like concentrated poverty explained variation in the likelihood of repeat maltreatment between census tracts (Gill & Womack, in press; Larsen & Merlo, 2005). For all children within a given census tract, for every one standard unit of increase in community-level concentrated poverty (community-level poverty was calculated as a *z*-score), the children in a given census tract were 36% more likely to be re-reported for maltreatment (β = 0.23, *SE* = 0.04, OR = 1.36).

Model 3 was used to calculate the predicted probability of recurrent maltreatment or the probability that y =1 for each observation. This vector of predicted probabilities was used to calculate an ROC curve for the binary logistic regression model; areas under the ROC curve were then compared for the neural network and the binary logistic regression model with random intercepts. As can be seen by comparing Figure 4.5 (shown earlier) with and 4.8 below, the area under the ROC curve for the neural network validation set was larger (0.78) than the area under the ROC curve for the linear model (0.76). Hence, the neural network demonstrated superior predictive validity. Figure 4.9 compares the area under the curve for the neural network model (AUC = 0.78) and the logistic regression model (AUC = 0.76) in comparison with the area under the curve for the two-level hierarchical null or intercept-only model (Model 1) (AUC = 0.64). Comparing the neural network model and the linear model each against the linear intercepts-only model, the researcher can attempt to quantify the relative degree to which the fully specified neural network and the logistic regression models predict the likelihood of recurrent maltreatment above and beyond a model that does not benefit from (a) the information contained in a traditional linear equation, and (b) the information contained within a more complex equation (Beck, King, & Zeng, 2004). By subtracting the AUC of the intercepts only model (AUC = .6401) from the neural network's AUC (.7825) and from the logistic regression's AUC (.7552), the relative difference between each model's improvement over the intercepts-only model can be determined and compared. Specifically, the neural network improved the AUC by .1424 and the logistic regression model improved the AUC by .1151. Comparing these two quantities reveals that the neural network's degree of improvement in predictive accuracy over the baseline model was just under 20% greater than the logistic regression's degree of improvement over the baseline model (where .1424 is just short of being 20% larger than .1151).

A comparison of the confusion matrices, sensitivity, and specificity for the neural network and logistic regression models is an additional way to assess the relative differences in each model's predictive accuracy. Table 4.7 summarizes the proportion of true positives, false positives, true negatives, and false negatives. In short, the sensitivity or the percentage of true positive cases that the logistic regression model correctly identified was higher at .81 compared with .75 for the neural network. However, the specificity or the percentage of true negative cases that the logistic regression model correctly identified was lower at .55 compared with .67 for the neural network. In short, although the logistic regression had a stronger ability to detect high risk cases, the neural network's predictive accuracy struck a better balance between its relative ability to distinguish high risk cases and low risk cases.

A comparison of the relative utility of each model should include not only a thorough assessment of the predictive validity of each model, but also an examination of the degree to which each model accurately represents the true functional form of the relationship between the likelihood of recurrent maltreatment and its predictors. A higher sensitivity value means comparatively little if the logistic regression model contains biased parameter estimates. Additional (and forthcoming) post-hoc analyses provide compelling evidence of nonlinearity in the relationship between recurrent maltreatment and its predictors, to include curvilinear relationships and interaction effects.

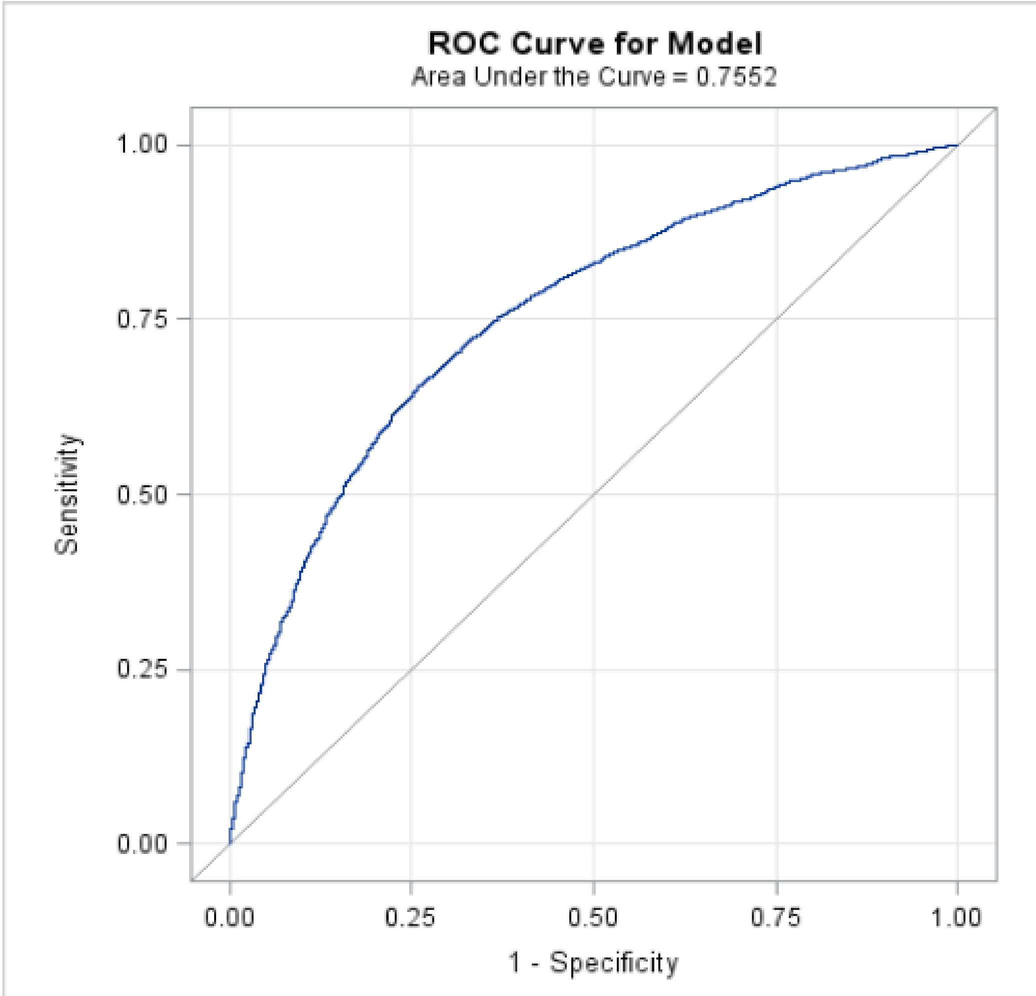


Figure 4.8. Area under the ROC curve for the fully specified binary logistic regression random intercepts model.

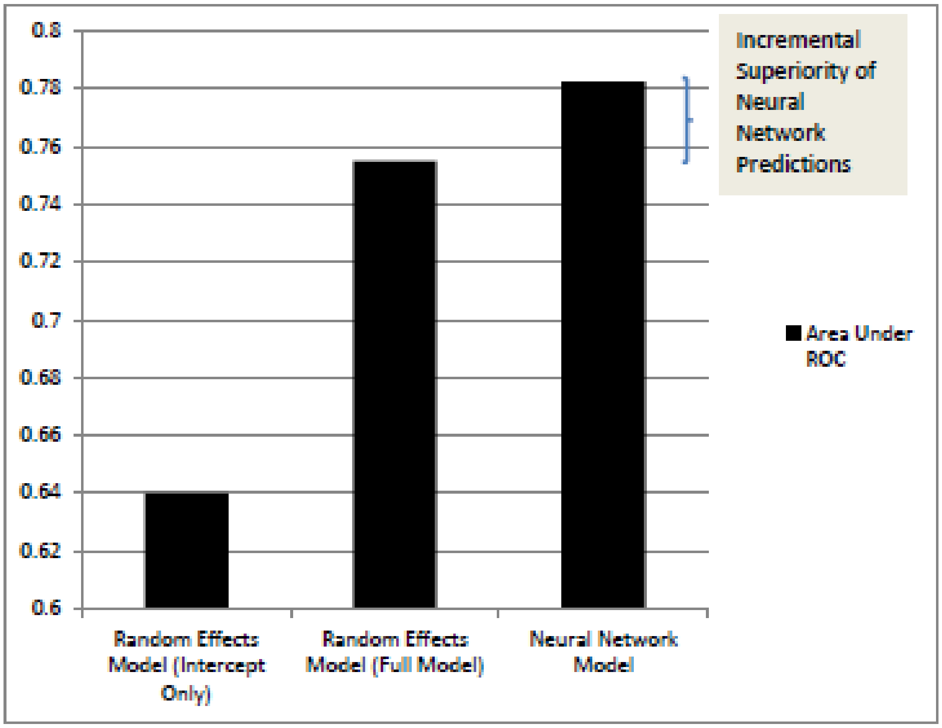


Figure 4.9. Increases in the area under the ROC curve when comparing the predictive accuracy of the neural network with a baseline, intercept-only logistic regression with random effects model and a fully-specified logistic regression with random intercepts model.

Table 4.7

Confusion Matrix for the Binary Logistic Regression Model with Random Intercepts

	Training Set	
Predicted		**Actual**
	Re-Reported (1)	**Not Re-Reported (0)**
Re-Reported (1)	3034 (70.79%) (True Positives)	1252 (29.21%) (False Positives)
Not Re-Reported (0)	734 (32.38%) (False Negatives)	1533 (67.62%) (True Negatives)

Bins, plots, and distributions: Additional methods for comparing the predictive validity of the neural network to a linear model.

An additional way of comparing the predictive validity of the neural network and logistic regression models begins with sorting the predicted probabilities of recurrent maltreatment produced by each model in ascending order; the ordered probabilities are then separated into bins of 0.1 width running from for example, 0.0001 to 0.10, 0.101 to 0.20, 0.201 to 0.30 etc. The average predicted probability of maltreatment recurrence is then calculated for the observations in each bin and the fraction of 1s observed for the response variable, recurrent maltreatment (1 = maltreatment recurrence and 0 = no maltreatment recurrence) is also calculated for the observations in each bin (Beck, King, & Zeng, 2000; King & Zeng, 2001). A model with good predictive accuracy will demonstrate agreement between the average predicted probability of maltreatment for the observations in each bin with the average number of children who actually experienced maltreatment recurrence (where the average level of actual recurrence is measured by the fraction of 1s in each bin) (Beck, King, & Zeng, 2000; King & Zeng, 2001). For example, if a model predicts that children in a given bin have a 30% probability of being re-reported for maltreatment, then approximately 30% of these children should be re-reported for maltreatment. Figures 4.10 and 4.11 plot the fraction of 1s in each bin against the average predicted probability of maltreatment recurrence for each bin for the logistic regression and neural network models respectively. The 45-degree line running through each plot demonstrates perfect agreement between the average number of times maltreatment actually recurred and the average predicted probability of maltreatment for each bin. A model that fits the data will produce a regression line (solid line) that deviates from the 45-degree line (dotted) only by random chance (Beck, King, & Zeng, 2000).

Looking at the plots for the logistic regression and neural network models, it can be seen that both models fit the data well, but the neural network model's fit is closer to the 45-degree dotted line. Moreover, when the neural network model deviates from the dotted line, the deviations are less substantial as compared with the deviations of the logistic regression model from the dotted line. The logistic regression model substantially under-estimated the average probability of recurrence for observations at the low end of the risk spectrum in bins one and two and moderately over-estimated the average probability of recurrence for observations in the middle range of the risk spectrum in bins four, five, and six. The logistic regression model then slightly under-estimated the average probability of maltreatment recurrence at the high end of the risk spectrum for observations in bins eight and nine and slightly over-estimated the average probability of maltreatment recurrence for observations at the highest point of the risk spectrum in bin 10. The neural network model also substantially under-estimated the average probability of maltreatment recurrence at the low end of the risk spectrum but only for observations in bin one. The neural net then slightly over-estimated the average probability of maltreatment recurrence for observations in the middle range of the risk spectrum in bins four, five, and six. Table 4.8 complements Figures 4.10 and 4.11 by providing the average predicted probability of recurrent maltreatment and the average number children who actually experienced recurrent maltreatment for observations in each of the 10 bins.

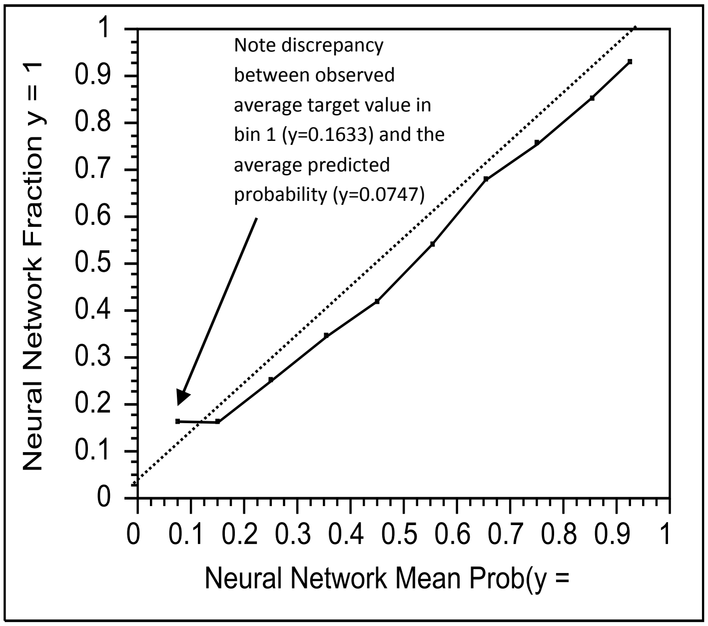


Figure 4.10. Plot of the binned fraction of outcome values where y = 1 by the binned average probability of recurrent maltreatment predicted by the neural network model.

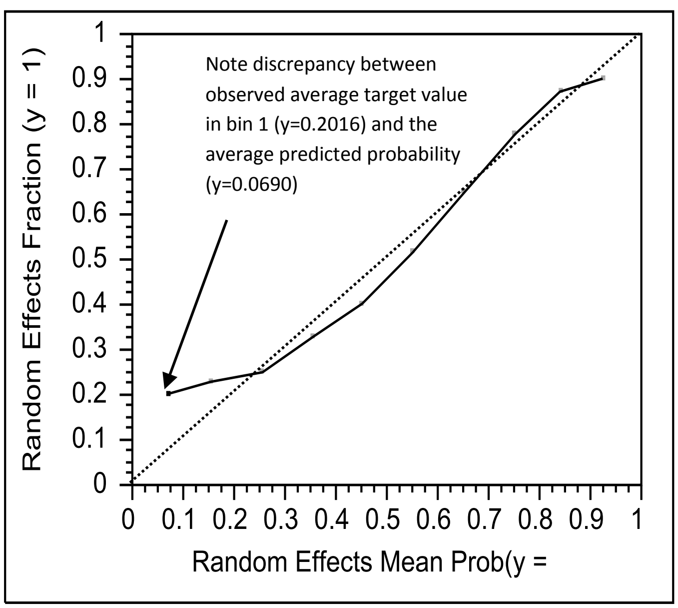


Figure 4.11. Plot of the binned fraction of outcome values where y = 1 by the binned average probability of recurrent maltreatment predicted by the logistic regression with random intercepts model.

Table 4.8

Comparison of the Binned Average Observed Target Values (Fraction of y=1) and Predicted Probabilities (y=1) for Neural Net and Random Effects Models[a]

Bin	Neural Net Model (Observed)	Neural Net Model (Predicted)	Random Effects Model (Observed)	Random Effects Model (Predicted)
1	0.16 (49)	0.07 (49)	0.20 (124)	0.07 (124)
2	0.16 (409)	0.15 (409)	0.23 (284)	0.15 (284)
3	0.25 (586)	0.25 (586)	0.25 (400)	0.26 (400)
4	0.34 (780)	0.36 (780)	0.33 (603)	0.35 (603)
5	0.42 (752)	0.45 (752)	0.40 (853)	0.45 (853)
6	0.54 (695)	0.55 (695)	0.52 (951)	0.55 (951)
7	0.68 (855)	0.65 (855)	0.65 (1001)	0.65 (1001)
8	0.75 (936)	0.75 (936)	0.78 (1255)	0.75 (1255)
9	0.85 (1033)	0.85 (1033)	0.87 (884)	0.84 (884)
10	0.93 (450)	0.92 (450)	0.90 (91)	0.92 (91)

Note. [a] Number of observed values in each bin are placed within parentheses.

A final method that compares the predictive accuracy of the neural network model with the logistic regression model is based on evaluating the overall agreement between the predicted probabilities estimated by each model (Stine, 2011). Evaluating the level of correlation between the two sets of model-predicted probabilities provides insight into just how different the predictions are and by extension the degree to which the neural network model's ability to estimate nonlinear functions of X adds information above and beyond what is obtained by assuming linear functions of X. Upon regressing the neural network-estimated probability of recurrent maltreatment on the logistic regression-estimated probability of recurrent maltreatment, the model was statistically significant [$F = 24{,}477.17$ (Model $df = 1$, Error $df = 6{,}552$), $p < .001$] and an R^2 of 0.79 indicated an excellent fit. The bivariate model accounted for 79% of the variation in the neural network-predicted probability of recurrent maltreatment.

The logistic regression-predicted probability of recurrent maltreatment were significantly and positively associated with the neural network-predicted probability of recurrent maltreatment [$\beta = 1.0044$ ($t = 156.45$, $SE = 0.01$) $p < .001$]. For every one percent increase in the logistic regression-predicted probability of maltreatment recurrence, the neural network-predicted probability of recurrence increased by 1.0044%. Hence, the overall level of agreement was strong where high predicted probabilities for the logistic regression model corresponded on average with high predicted probabilities for the neural network model and low predicted probabilities for the logistic regression model corresponded on average with low predicted probabilities for the neural network model. However, the correlation between the two sets of predicted probabilities was not perfect: 21% of the variation in the neural network-predicted probability of recurrent maltreatment was not accounted for by the correlation between the two models. Figure 4.12 includes a plot of the neural network predicted-probability of recurrent maltreatment regressed on the logistic regression predicted-probability of recurrent maltreatment. If the two models essentially provided the same information, the correlation would be stronger and the pattern of data points would form a thinner line that hugs the regression line more closely. Instead, the data points formed a relatively thick band around the regression line and the overall pattern of data points was somewhat diffuse.

A predicted probability difference variable was created by subtracting the logistic regression- predicted probabilities of recurrent maltreatment from the neural network-predicted probabilities of recurrent maltreatment. The distribution of the differences in predicted probabilities of maltreatment recurrence was assessed. Figure 4.12 also displays the distribution of the differences in predicted probabilities; it can be seen that a substantial portion of the observations differ in their model-predicted probabilities by ±10% (*n* = 4,331). If the zero point represents no difference between the model predicted-probabilities, then 4,331 observations had a logistic regression-predicted probability that was either 10% greater or 10% less than the neural network-predicted probability.

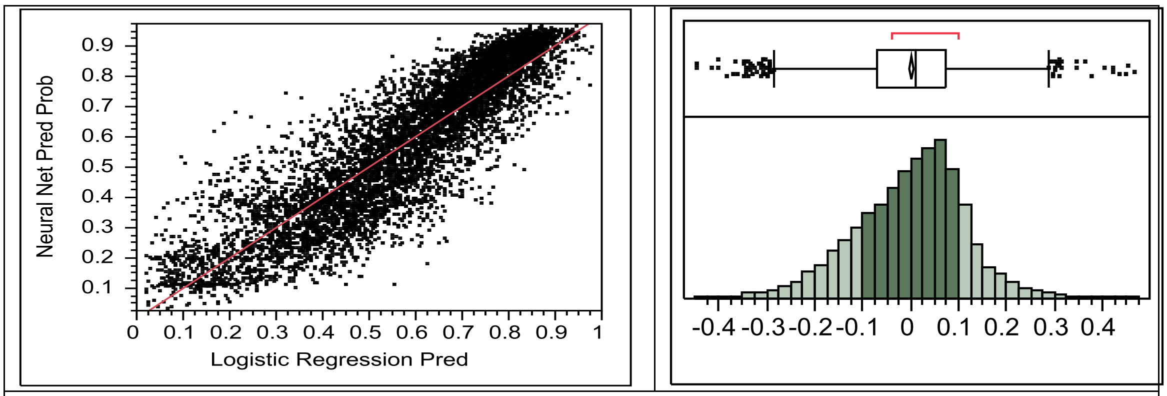


Figure 4.12. Simple regression of the neural network-predicted probability of recurrent maltreatment on the logistic regression-predicted probability of recurrent maltreatment (on the left) and the portion of observations where the difference between the neural network and the logistic regression predicted probabilities of maltreatment differed by ± 10% (on the right). Notice the rather fat and somewhat diffuse band of data points around the regression line indicating the lack of perfect agreement between the neural network and logistic regression-predicted probabilities of recurrent maltreatment.

However, 2,223 observations (34% of the sample) had a difference in the model-predicted probabilities that was larger than 10%; in fact, differences in the model-predicted probabilities ranged as high as 47%. While the models agreed on the predicted probabilities on average, the models did not agree nearly as strongly on the specifics of the individual-level predicted probabilities. Figure 4.13 demonstrates the correspondence between the distribution of the predicted probability differences from 0.10 to 0.47 (n = 1,049) and the location of the more diffuse data points in the plot of the neural network-predicted probability of maltreatment recurrence by the logistic regression-predicted probability of maltreatment recurrence. The observations that are common to both figures are located in the darkened bins of the distribution and the darkened data points in the bivariate plot.

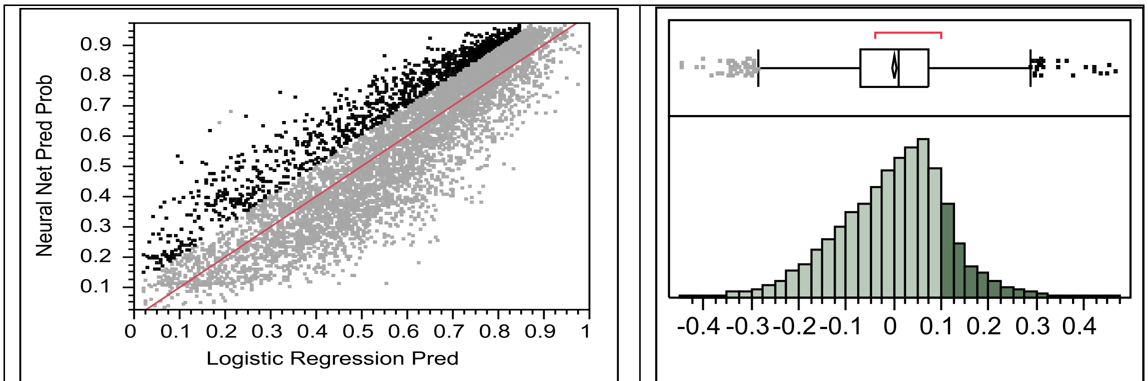


Figure 4.13. Simple regression of the neural network-predicted probability of recurrent maltreatment on the logistic regression-predicted probability of recurrent maltreatment (on the left) and the portion of observations where the difference between the neural network and the logistic regression predicted probabilities of maltreatment differed by 10% to 47% (on the right). The darkened diffuse data points highlighted in the figure on the left correspond with the observations located within the darkened bins in the histogram on the right.

Similarly, Figure 4.14 demonstrates the correspondence between the distribution of the predicted probability differences from -0.10 to -0.45 (n = 1,174) and the location of the more diffuse data points in the plot of the neural network-predicted probability of maltreatment recurrence by the logistic regression-predicted probability of maltreatment recurrence. The observations that are common to both figures are located in the darkened bins of the distribution and the darkened data points in the bivariate plot.

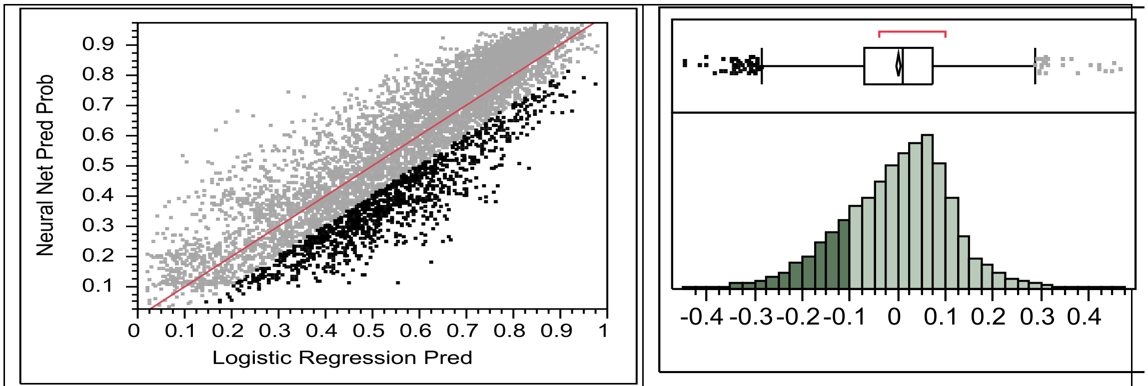


Figure 4.14. Simple regression of the neural network-predicted probability of recurrent maltreatment on the logistic regression-predicted probability of recurrent maltreatment (on the left) and the portion of observations where the difference between the neural network and the logistic regression predicted probabilities of maltreatment differed by -10% to - 45% (on the right). The darkened diffuse data points highlighted in the figure on the left correspond with the observations located within the darkened bins in the histogram on the right.

Putting the Predicted Probabilities into Context: Results from a Post-Hoc Regression Tree

Before proceeding with comparisons between the neural network model and a binary logistic regression model with random intercepts that assumes linearity in the parameters (i.e., network weights) and the predictors, results from the previously described post-hoc regression tree are presented. With the neural network predicted probabilities of recurrence (y = 1) as the response variable and the 22 predictors from the neural network model as the input variables, the regression tree empirically places the estimated probability of recurrence in context. At each recursive binary split, predictors are assessed for their relative ability to explain variance in the predicted probability of recurrence such that the best splitting variable is (a) the one that reduces the sum of squared errors to the greatest extent, and (b) the one that splits the observations into two groups where the average probability of maltreatment recurrence is substantively different (Berk, 2008; Breiman, Friedman, Olshen, & Stone, 1984; Fox, 2000; SAS, 2010; Stine, 2011).

The probability of maltreatment recurrence (Y) is modeled as a constant conditional average (c_m) within each partition of input space, where the mean of Y is dependent upon specific values of successive predictors (X) found within a given partition (R_m) of input space. Hence, the probability of maltreatment recurrence is modeled as follows:

$\hat{c}_m = ave(y_i | x_i \in R_m$ (Hastie, Tibshirani, & Friedman, 2001, p. 269).

The best splitting predictor (*j*) and the best split point (*s*) for a given predictor minimize the sum of squares of difference between the actual and predicted values of the probability of recurrent maltreatment such that the best possible solution is found for the following:

$min_{j,s}\left[min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2\right]$ (Hastie, Tibshirani, & Friedman, 2001, p. 269).

With each successive binary split, the observations are funneled down the tree into internal nodes that ultimately place each observation in a terminal node or leaf. Each leaf provides a context in which the risk of recurrence can be understood in relationship to a conditional average, where all observations in a given leaf have the same average likelihood of recurrence given the sequence of values for a selected subset of predictors. Hence, each leaf organizes the observations into empirically defined risk-based groups where the average probability of recurrence can potentially be altered by addressing the conditions (i.e., the sequence of input values) that define a given risk group. Moreover, due to the sequential nature of the recursive binary splits, each successive predictor variable and corresponding split point is dependent on the previous predictors and split points. Thus, interaction effects that the neural network model may have discovered can be more easily understood. An examination of the successive decision points or splitting rules can provide a great deal of insight into which predictors "take the lead" by defining the splitting process in its early stages and which predictors (and at which values) differentiate groups of observations as they are funneled down the tree before being split into their final nodes.

A regression tree was built using the same sample of observations that was used to build the neural network (N = 6,554). There were no case records with missing values in the data set (the 193 records with missing observations discussed previously were excluded from the neural network analysis and therefore did not have a predicted probability of maltreatment recurrence). Ten-fold cross-validation was used to (a) provide the regression tree with the largest amount of training data possible with which to finding underlying systematic structure, and (b) fine-tune the tree's complexity while estimating the model's generalizability. An additional constraint on model complexity was imposed by implementing a stopping criterion that required each node to include a minimum of 25 case records (Stine, 2011). The regression tree that was produced was large with 198 splits and very good predictive validity with a training R^2 of 0.79 (SSE = 76.21) and a validation R^2 of 0.81 (SSE = 71.63, RMSE = 0.10). Variable importance is measured by SS, which is the degree to which each splitting variable reduces the sum of squared errors; Figure 4.15 below rank orders each predictor by its SS value.

While the regression tree possessed a high degree of predictive validity, with 198 splits, the interpretability of the tree was severely compromised. Hence, a second regression tree was built by limiting the first regression tree to just 20 splits. The second tree had excellent interpretability and retained much of the original proportion of explained variance in the probability of recurrent maltreatment with an R^2 of 0.71 (RMSE

Column Contributions

Term	Number of Splits	SS
Juv_Ct_Pr_02	1	104.75946
Num_IM_Sp	12	79.43408
Spec_Ed_Pr_02	3	26.63146
FCS	4	18.34168
Comm_Pov	25	16.85351
Ch_Age	28	14.13658
Ch_MHSA_Pr_02	8	12.21213
Pt_Perp	8	8.90074
Pt_Educ	22	3.97130
Substantiation	6	2.87137
Perp_Protective	15	2.69416
Fam_Protective	14	2.16947
FPS	1	1.86786
Pt_Age	9	1.28661
Comm_Move	11	1.06032
Ch_Race	9	0.96791
Spec_Ed_Pr_01	1	0.79010
Ch_Gender	15	0.49442
Pt_Gender	2	0.08375
Ch_Birth_Prb	4	0.05729
Pt_MHSA_Pr_02	0	0.00000
Mom_Hx_Fost_Care	0	0.00000

Figure 4.15. Post-hoc 198-split regression tree input variable importance measured by rank ordering in descending order each input's SS statistic (the change in the error sum of squares due to the split).

= 0.13). In fact, an examination of the split history of the first regression tree (with 198 splits) in Figure 4.16 shows that an R^2 in the low 0.70's is reached early in the splitting process at around 20 splits and further increases in R^2 continue at a near glacier-like pace throughout the remaining 178 splits. Additionally, Figure 4.17 summarizes predictor importance for the 20-split tree by rank ordering each predictor by its SS value. The predictors with the highest SS values from the 198-split tree are the same predictors with the highest SS values for the 20-split tree.

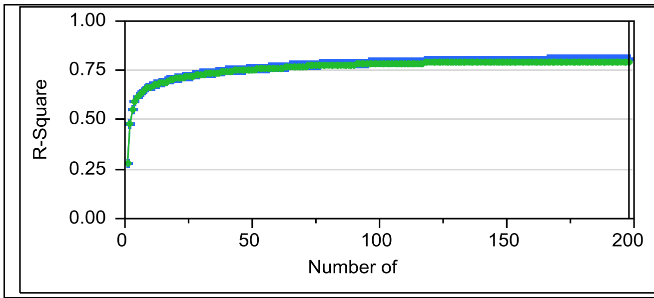


Figure 4.16. Post-hoc regression tree's predictive accuracy as measured by the R^2 plotted by the number of splits (total number of splits = 198).

Column Contributions

	Term	Number of Splits	SS
	Juv_Ct_Pr_02	1	104.75946
XX	Num_IM_Sp	3	77.43598
	Spec_Ed_Pr_02	1	24.93251
	FCS	1	17.45387
	Comm_Pov	2	9.94542
	Ch_Age	4	9.34465
	Ch_MHSA_Pr_02	1	8.67713
	Pt_Perp	3	6.30914
	Substantiation	1	2.10171
	FPS	1	1.86786
	Pt_Educ	2	1.48004
	Pt_Age	0	0.00000
	Comm_Move	0	0.00000
	Ch_Race	0	0.00000
	Perp_Protective	0	0.00000
	Ch_Birth_Prb	0	0.00000
	Fam_Protective	0	0.00000
	Ch_Gender	0	0.00000
	Pt_MHSA_Pr_02	0	0.00000
	Mom_Hx_Fost_Care	0	0.00000
	Pt_Gender	0	0.00000
	Spec_Ed_Pr_01	0	0.00000

Figure 4.17. Post-hoc 20-split regression tree input variable importance measured by rank ordering in descending order each input's SS statistic (the change in error sum of squares due to split).

Hence, the 20-split tree was selected as the model that best represents the prediction of the probability of being re-reported for maltreatment as 21 risk-based groups defined by the sequence of values on a key subset of predictors. Figure 4.18 provides a "small tree" view by replicating the decision structure of the 20-split tree and Table 4.9 describes each of the 21 risk groups by (a) summarizing the sequence of predictors and split points, (b) reporting the average probability of recurrent maltreatment for each group, and (c) providing a count of observations within each group.

An examination of the table summarizing the 21 risk groups provides insight into the embedded nature of the predictors' effects in explaining variation in the probability of being re-reported for maltreatment. Taking the lead are two predictors: (1) a variable that measures involvement in the juvenile justice system through the receipt/no receipt of a first juvenile court petition on or after the first maltreatment report but before the second report (if a second maltreatment report occurred), and (2) a variable that measures the total number of income maintenance spells received before the first maltreatment report up to but not including a second maltreatment report (if a second report occurred). That said, it is important to point out the fact that the values for every continuous predictor have been rescaled as *z*-scores (as described in chapter three). While each continuous predictor could have been converted back to its original unit of analysis to make the regression tree findings more interpretable in a practical sense, keeping the continuous predictors in their *z*-scored format facilitates comparisons between the neural network findings (with probability plots forthcoming) and the regression tree findings. Hence, a score of zero for the number of income maintenance sells does not actually refer to an absence of income maintenance spells, but is in fact the average number of income maintenance spells received by children and their families in the sample.

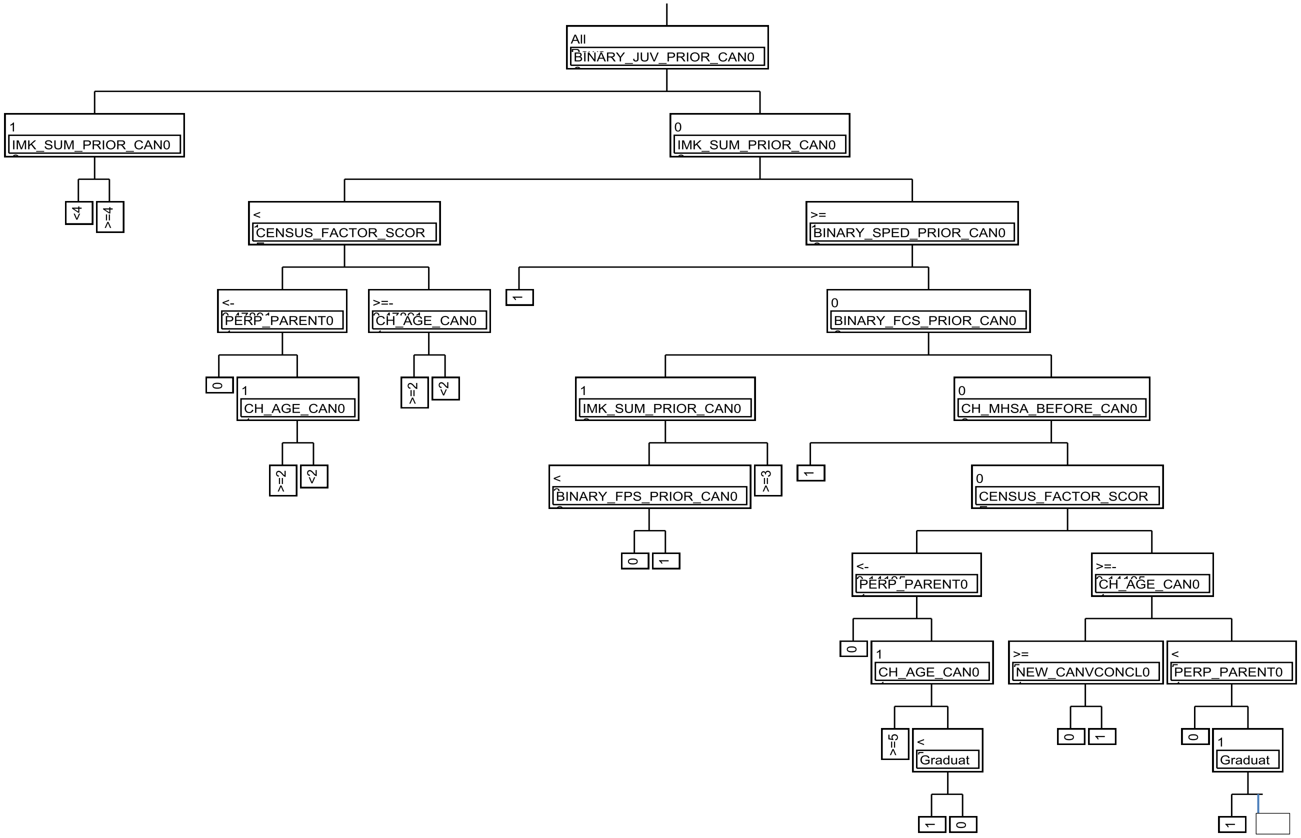


Figure 4.18. Post-hoc regression tree that fits the expected value of the probability of recurrence to homogenous groups given the values of the predictors of maltreatment recurrence.

Table 4.9

Regression Tree-Generated Risk Groups Identified as a Function of Values on Key Predictors

	Mean Probability of Maltreatment Recurrence	Count
1. Yes Juv_Ct & Num_IM_Sp <4	21%	699
2. Yes Juv_Ct & Num_IM_Sp >=4	38%	58
3. No Juv_Ct & Num_IM_Sp <1 & Comm_Pov <-0.47891 & No Pt_Perp	33%	285
4. No Juv_Ct & Num_IM_Sp <1 & Comm_Pov <-0.47891 & Yes Pt_Perp & Ch_Age >=2	42%	805
5. No Juv_Ct & Num_IM_Sp <1 & Comm_Pov <-0.47891 & Yes Pt_Perp & Ch_Age <2	55%	128
6. No Juv_Ct & Num_IM_Sp <1 & Comm_Pov >=-0.47891 & Ch_Age >=2	49%	368
7. No Juv_Ct & Num_IM_Sp <1 & Comm_Pov >=-0.47891 & Ch_Age <2	65%	143
8. No Juv_Ct & Num_IM_Sp >=1 & Yes Spec_Ed	48%	465
9. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & Yes FCS & Num_IM_Sp <3 & No FPS	51%	385
10. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & Yes FCS & Num_IM_Sp <3 & Yes FPS	69%	69
11. Juv_Ct = 0 & Num_IM_Sp >=1 & Spec_Ed = 0 & FCS = 1 & Num_IM_Sp >=3	67%	136
12. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & Yes Ch_MHSA	62%	400
13. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov <-0.14105 & No Pt_Perp	60%	161
14. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov <-0.14105 & Yes Pt_Perp & Ch_Age >=5	69%	333

Table 4.9

Regression Tree-Generated Risk Groups Identified as a Function of Values on Key Predictors

	Mean Probability of Maltreatment Recurrence	Count
15. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov <-0.14105 & Yes Pt_Perp & Ch_Age <5 & Yes Pt_HS_Degree	73%	236
16. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov <-0.14105 & Yes Pt_Perp & Ch_Age <5 & No Pt_HS_Degree	83%	203
17. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov >=-0.14105 & Ch_Age >=5 & No Substantiation	73%	574
18. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov >=-0.14105 & Ch_Age >=5 & Yes Substantiation	88%	109
19. No Juv_Ct & Num_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov >=-0.14105 & Ch_Age <5 & No Pt_Perp	78%	172
20. No Juv_Ct & Num_IM_sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov >=-0.14105 & Ch_Age <5 & Yes Pt_Perp & Yes Pt_HS_Degree	82%	274
21. No Juv_Ct & Mum_IM_Sp >=1 & No Spec_Ed & No FCS & No Ch_MHSA & Comm_Pov >=-0.14105 & Ch_Age <5 & Yes Pt_Perp & No Pt_HS_Degree	88%	551

An examination of the 21 risk groups reveals that values for juvenile court involvement and income maintenance receipt initially define the risk-based trajectories for each of the 21 groups. Interestingly, juvenile court involvement appears to lower the risk of maltreatment recidivism. Despite having a very high split point for the number of income maintenance spells received (less than four spells versus four or more spells), children in risk groups one and two had a low average probability of maltreatment recurrence: 21% and 38% respectively. Specifically, for children involved in the juvenile court and who received less than four income maintenance spells, the average probability of repeat maltreatment was 22%. For children involved in the juvenile court and who received four or more income maintenance spells, the average probability of repeat maltreatment was 38%.

The next set of risk groups, groups three through seven, are defined initially through settings on the two lead predictors: juvenile court involvement and income maintenance receipt. For each group, none of the children were coded as having juvenile court involvement and the split point for the number of income maintenance spells received was much lower in comparison with the split point that defined the first two risk groups. Specifically, children in risk groups three through seven were split into two different groups depending on whether or not the number of income maintenance spells received was less than one or was greater than or equal to one. Despite having a lower split point for poverty measured at the individual level, the average probability of recurrent maltreatment was (with the exception of risk group three) much higher for children without juvenile court involvement as compared with children who received a juvenile court petition. Specifically, (excluding risk group three with an average probability of 33%), children without juvenile court involvement in risk groups four through seven had a 42% minimum average likelihood of recurrence and a 65% maximum average likelihood of recurrence. Beyond a lack of juvenile court involvement and lower values for individual-level poverty as measured through income maintenance receipt, values for three additional predictors defined the risk trajectories of children in groups three through seven: (1) concentrated poverty measured at the census-tract level, (2) the primary caregiver's potential status as the perpetrator of the first maltreatment incident, (3) the child's age at the first maltreatment report.

Children in risk group seven had the highest average probability of maltreatment recurrence at 65% (when comparing the average probabilities of risk groups three through seven) and were defined as being (1) without juvenile court involvement, (2) living in a family that received less than one income maintenance spell, (3) exposed to a community level of concentrated poverty that was greater than or equal to -0.48 (places the child closer to the average level of community poverty and in the direction of higher values for community poverty), and (4) less than 2 years of age.

Risk groups eight through 21 are each defined by the following two conditions: (1) no juvenile court involvement, and (2) receipt of income maintenance support where children are split into two groups depending on whether the family received less than one income maintenance spell as opposed to receiving one or more income maintenance spells. The remaining predictors are concentrated among a fairly small subset: (1) special education eligibility, (2) receipt of family centered services through the child welfare system, (3) a second split point for the number of income maintenance spells received, (4) receipt of family preservation services through the child welfare system, (5) receipt of child mental health/substance abuse services through the community mental health system, (6) exposure to concentrated community-level poverty, (7) the primary caregiver's potential status as the perpetrator of the first maltreatment incident, (8) the child's age at the first maltreatment report, (9) the primary caregiver's educational status (no high school degree versus high school degree plus additional years of education), and (10) the substantiation status of the first maltreatment report. The number of predictors that defined each of the remaining risk groups (groups eight through 21) ranged from a minimum of three (risk group eight) to a maximum of nine (risk groups 15, 16, 20, and 21).

When looking at the average probabilities of repeat maltreatment across risk groups eight through 21, a pattern emerges in relationship to the role that cross-sector services play. In instances where at least one type of service is received -- i.e., special education, family centered services, or child mental health-substance abuse services -- the average probability of recurrent maltreatment is lower; in turn, when no services are received, the average probability of recurrent maltreatment is higher. For example, children in risk group eight had no juvenile court record, received one or more income maintenance spells, and received special education eligibility: the group's average likelihood of repeat maltreatment was 48%. Similar to the first two risk groups, the number of predictors needed to define the composition of the eighth risk-based group is small (three in total) because the reduction in the sum of squared errors accounted for by the three predictors was substantial enough to form a relatively homogenous group of observations. Risk group nine provides another example of this pattern wherein children in this group shared a 51% average likelihood of recurrent maltreatment. Instead of being eligible for special education services, children in risk group nine received a first spell of family centered services, and receipt of this type of service appeared to be able to temper the effects of individual-level poverty as measured by the receipt of income maintenance spells at two different split points.

In contrast, children in risk groups 10, 11, 14, and 16 provide examples of how the receipt of services are (a) only effective in tempering the effects of risk factors like individual-level poverty (as measured by the receipt of income maintenance spells) given the level of poverty experienced; (b) a sort of barrier to preventing a “pile on” of risk factors like individual-level poverty, exposure to community-level poverty, lower levels of child age at the first maltreatment report, the parent’s status as a perpetrator, and the parent’s lack of a high school degree; and (c) only effective in tempering the effects of risk factors like individual-level poverty if the service receipt is limited to certain types of services (i.e., the receipt of family preservations services do not temper the effects of risk factors like individual-level poverty). For example, children in risk group 10 have the same combination of predictors and the same split point values for each predictor as those described for children in risk group nine with the exception of the receipt of family preservation services. Without the receipt of family preservation services and with the receipt of family centered services, children in risk group nine had a 51% average likelihood of maltreatment recurrence. With the receipt of family preservation services and with the receipt of family centered services (and the same split points for individual-level poverty), children in risk group nine had a 69% average likelihood of maltreatment recurrence.

Children in risk group 11 had a nearly identical risk profile as the children in risk group nine with one important difference: the children in risk group 11 experienced a higher level of individual-level poverty as measured by the difference in split points for the number of income maintenance spells received. Specifically, children in risk group nine had no juvenile court involvement, received one or more income maintenance spells, had no special education involvement, received a first family-centered service spell, received less than three income maintenance spells (a second split point for income maintenance), and received no family preservation services. The average likelihood of repeat maltreatment for children in risk group nine was 51%. The children in risk group 11 had a nearly identical risk profile with one exception: at the second split point for the receipt of income maintenance spells, children in risk group 11 received three or more spells as opposed to less than three spells. Despite their receipt of family centered services, the average likelihood of repeat maltreatment increased from 51% to 67%.

Children in risk groups 14 and 16 appear to have experienced a veritable "pile on" of risk factors accompanied by high average probabilities of recurrent maltreatment -- i.e., 69% and 83% respectively -- in conjunction with a lack of service receipt/system oversight. None of the children in risk groups 14 and 16 were juvenile court involved, special education eligible, in receipt of family centered services, and/or in receipt of mental health/substance abuse services. In fact, none of the children in risk groups 14 (65% average probability of maltreatment), 15 (73% average probability of maltreatment), 16 (83% average probability of maltreatment), 17 (73% average probability of maltreatment), 18 (88% average probability of maltreatment), 19 (78% average probability of maltreatment), 20 (82% average probability of maltreatment), and/or 21 (88% average probability of maltreatment) were juvenile court involved, special education eligible, in receipt of family centered services, and/or in receipt of mental health/substance abuse services. The "pile on" of risk factors that drive up the average likelihood of repeat maltreatment include (1) higher levels of household-based poverty as measured by the number of income maintenance spells received, (2) higher levels of exposure to community-based poverty, (3) lower levels of the child's age at his/her first maltreatment report, (4) the parent's status as the perpetrator of the first maltreatment incident, and (5) the parent's lack of a high school degree. Moreover, when the effects of the key risk factors are tempered by the receipt of at least one cross-sector service, all services were delivered on or after the first maltreatment report but before the second report (if a second maltreatment report occurred). Hence, service receipt before the first maltreatment report did not appear to minimize the sum of squared errors and/or temper the effects of key risk factors in relationship to the increased probability of recurrent maltreatment.

A quick examination of Figure 4.17 (shown previously) reveals additional empirical support for the discussion of the predictors that were most influential in explaining variance in the probability of recurrent maltreatment (and consequently minimizing the sum of squared errors). Children were divided into homogenous risk-based groups as a function of the unique values on the predictors featured in the 21 risk groups. Moreover, these same predictors were rank ordered as the most important variables in explaining variation in the probability of maltreatment recurrence while minimizing the sum of squared errors.

In order to better understand the potential contribution that a neural network can make to risk assessment beyond improvements to predictive accuracy, the findings from the regression tree will be used to frame a post-hoc comparison of the ways in which the estimated relationships between the probability of maltreatment recurrence and key predictors differ between the neural network and logistic regression models. As noted earlier, a neural network's contribution to the improvement in risk assessment can first be examined by comparing the neural network's improvement in predictive accuracy as compared with a linear model. Beyond contributions to increasing predictive accuracy, the neural network's open-ended capacity to estimate curvilinear and interaction effects may provide evidence that parameter estimates produced by models that are linear in X_p are biased because the wrong functional form was assumed. Moreover, incorrect assumptions about the ways in which the likelihood of recurrent maltreatment relate to linear functions of X_p have serious implications for treatment planning and the delivery of preventive services within an RNR perspective. As described in earlier chapters, in order for interventions to be effective (i.e., responsive) in reducing the likelihood of future maltreatment, the targets for treatment (i.e., dynamic risk factors) must be accurately conceptualized and operationalized.

The following section evaluates the functional form of the relationship between the average probability of recurrent maltreatment and eight critical X_p as identified by the analysis of the regression tree findings (while holding all other variables in the neural network constant). The first four X_p are the predictors that appear to most prominently increase the risk of recurrent maltreatment: (1) the number of income maintenance spells received, (2) exposure to community-based poverty, (3) the child's age at the first maltreatment report, and (4) the parent's status as the perpetrator of the first maltreatment incident. The next four X_p are the predictors that appear to most prominently decrease the risk of recurrent maltreatment by potentially moderating the effects of the four risk factors that most substantially contribute to an increase in the likelihood of repeat maltreatment: (1) receipt of a juvenile court petition, (2) becoming eligible for special education, (3) receiving a family-centered service spell, and (4) child receipt of a mental health/substance abuse service. For each row of probability plots provided, assumptions of linearity in the eight critical inputs will be assessed by looking for evidence of (a) curvilinearity in the first three continuous risk factors, and (b) interactions among the first four risk factors and the four cross-sector service receipt variables. Moreover, assumptions of linearity as well as the unique contributions of the neural network in more accurately specifying the functional form of X_p will be assessed by comparing probability plots from the neural network analysis with probability plots for the same four risk factors and the same four cross-sector service receipt variables from the binary logistic regression analysis (Beck, King, & Zeng, 2000; King & Zeng, 2001).

Assessing Assumptions of Linearity in X_p: Comparing Probability Plots from the Neural Network and the Binary Logistic Regression Models

Probability and interaction plots for the neural network model.

One of the most effective ways of visualizing the functional form of a relationship between the average probability of recurrent maltreatment and any selected predictor is through the examination of a probability plot (Beck, King & Zeng, 2000; Fox, 2000; King & Zeng, 2001). For each plot, the probability of recurrent maltreatment is plotted along the *y*-axis; the *y*-axis displays changes in the average model-predicted probabilities of recurrence across all observations in relationship to (a) the particular value for the input variable featured along the *x*-axis, and (b) constant values for all other variables in the model. Evidence of curvilinearity can be found by examining the degree to which each bivariate regression line is bent or otherwise strays from a straight line. Evidence of interactions can be found by (1) selecting different values for any given predictor along its corresponding *x*-axis, and (2) noting any changes in the functional form or shape of the regression lines for input variables other than the predictor that is being manipulated.

For the figures that follow below, a set of two rows is included per figure. Each row is comprised of four plots where one of the four prominent risk factors is plotted along the *x*-axis (i.e., the number of income maintenance spells received, level of concentrated poverty at the community level, the child's age at the first maltreatment report, and the parent's status as the perpetrator of the first reported maltreatment incident) plus a fifth plot where one of the service receipt/service intervention variables is plotted along the *x*-axis (i.e., juvenile court-involved, special education eligible, receipt of family centered services, and the child's receipt of mental health/substance abuse services). In any given figure, the first row of plots shows the relationship between the average predicted probability of recurrent maltreatment and a selected predictor where (a) the number of income maintenance spells, community-level poverty, and the child's age at the first report were fixed at their average values (~ 0); (b) the parent's status as perpetrator was fixed at a value of 0.9 (i.e., was found to be the perpetrator); and (c) all other variables in the model, including all service receipt/involvement variables were set to -0.9 (i.e., no service receipt/involvement) (Please note: Due to the coding scheme for the neural network described in Chapter 3, response levels for dichotomous variables are represented with -0.9 and 0.9, where -0.9 equals 0 or the "no" category and 0.9 equals 1 or the "yes" category). The second row of plots is the same as the first row with one important exception: the service-related variable plotted along the *x*-axis of the fifth plot has been moved from a value of -0.9 (i.e., no service receipt/involvement) to a value of 0.9 (service receipt/involvement). Each set of two rows per figure highlights the changes in the average predicted probability of recurrent maltreatment as the selected service variable is transitioned from a value of -0.9 to a value of 0.9; moreover, each set of rows highlights the changes in the four prominent risk factors when the selected service variable is transitioned from a value of -0.9 to a value of 0.9. Figure 4.19 evaluates the potential moderating effects of juvenile court involvement, Figure 4.21 evaluates the potential moderating effects of special education eligibility, Figure 4.23 evaluates the potential moderating effects of family-centered service receipt, and Figure 4.25 evaluates the potential moderating effects of the child's receipt of mental health/substance abuse treatment.

In Figures 4.19, 4.21, 4.23, and 4.25, evidence of curvilinearity can be seen by examining the shape of the relationship between each of the first three risk factors and the predicted probability of recurrent maltreatment. Evidence of the interaction effects can be seen by comparing the shape of the curvilinear regression lines for each of the first three risk factors in the top and bottom rows of plots as the service variable in the fifth plot is transitioned from a value of -0.9 to a value of 0.9; changes in the direction or placement of the "peaks and valleys" in the curvilinear regression line indicates an interaction effect. Evidence of an interaction between a given service variable and the one categorical risk factor (i.e., the parent's status as perpetrator) can be seen by looking for changes in the pitch or steepness of the regression line.

To facilitate further exploration of possible interactions between each service variable and each of the four risk factors, interaction plots are provided as well using Figures 4.20, 4.22, 4.24, and 4.26. Each interaction plot (with four plots per figure) places the average predicted probability of recurrent maltreatment on the *y*-axis and one of the four prominent risk factors along the *x*-axis. However, instead of providing one regression line, two lines are provided where changes in the value of the average predicted probability of recurrent maltreatment along the *y*-axis for every one-unit increase in values for the selected risk factor along the *x*-axis is conditioned on a selected value for the service variable. Hence, each regression line represents the relationship between the predicted probability of maltreatment recurrence and a select risk factor when (a) the service variable is set to -0.9 or “no service/system intervention,” and (b) when the service variable is set to 0.9 or “service receipt/system involved.” All variable settings used for the creation of the probability plots were also used to create the interaction plots. Thus, all continuous prominent risk factors were held constant at their average values (~0), the parent’s perpetrator status was set to 0.9 (where the parent is designated as the perpetrator for the first maltreatment incident), and all other predictors not featured in the probability plot were set to

-0.9.

For each selected risk factor featured in the probability plots, the corresponding interaction plot places side by side two regression lines that model changes in the average predicted probability of recurrent maltreatment as a function of a given prominent risk factor when (a) a selected service variable is “turned off,” and (b) the same selected service variable is “turned on.” Evidence of an interaction can be seen when two curvilinear regression lines are dissimilar in their shape such that their ripples or “peaks and valleys” follow a different pattern (thereby producing a different overall shape). Additionally, evidence of an interaction can be seen when two straight regression lines are divergent in their pitch (i.e., they are nonparallel), where one line might be running in a horizontal fashion and the second line might be pitched in a steeper fashion running upward (for a good example, please see the interaction plot for the perpetrator as parent being moderated by receipt of family centered services in Figure 4.24). Finally, evidence of an interaction can be seen when two regression lines (either straight lines or curvilinear lines) cross or intersect each other. As noted by Jaccard (2001), the nonparallel nature of two regression lines or slopes is “indicative of the interaction and the degree of nonparallelness gives some appreciation of the magnitude of the interaction” (p. 54).

One of the advantages in using a neural network to identify a range of potential interaction effects is the ability to identify interactions where neither of the predictors included in the interaction term are required to be first-order polynomial terms; in other words, the predictors involved in the interaction can be higher-order polynomial terms where the probability of recurrent maltreatment is curvilinear in X. Specifically, the sight of two straight nonparallel lines indicates the presence of an interaction where, for example, the log odds of recurrent maltreatment is a linear function of both the independent focal variable (X_1) and the moderator variable (X_2) (Jaccard, 2001). In this case, specifying the interaction for inclusion in a typical logistic regression model (or even a neural network model) can be achieved by creating a product term (i.e., one multiplies X_1 by X_2). However, a product term cannot be used to represent an interaction that occurs when the probability of maltreatment is a nonlinear function of either or both X_1 and X_2 (Jaccard, 2001). Hence, a neural network analysis allows the researcher to keep his/her "options open" by allowing the data to guide the estimation of the target function to include any and all relevant interaction effects as opposed to making *a priori* assumptions about the functional form of the relationship between Y and X_p.

In Figures 4.19, 4.20, 4.21, and 4.22, please note the average probability of recurrent maltreatment given the values of all inputs in the model, located to the far left of each row of probability plots running alongside the *y*-axis. The average likelihood of repeat maltreatment was 90% when (a) the number of income maintenance spells, community-level poverty, and child's age at first maltreatment report were set to their average values (~ 0 for these *z*-scored variables); (b) the parent was identified as the perpetrator of the first maltreatment incident; and (c) all other variables in the model including all service variables were set to -0.9. When the child receive a juvenile court petition and the juvenile court involvement variable was transitioned to a value of 0.9, the average likelihood of recurrent maltreatment dropped to 33%. When the child was determined to be eligible for special education, and the special education eligibility variable was transitioned to a value of 0.9 (holding all other variables constant), the average likelihood of recurrent maltreatment decreased to 46%. When the family received a first spell of family centered services, and the family-centered service variable was transitioned to a value of 0.9 (holding all other variables constant), the average likelihood of recurrent maltreatment decreased to 52%. Finally, when the child received a mental health/substance abuse service and the mental health/substance abuse treatment variable was transitioned to a value of 0.9 (holding all other variables constant), the average likelihood of recurrent maltreatment decreased to 71%.

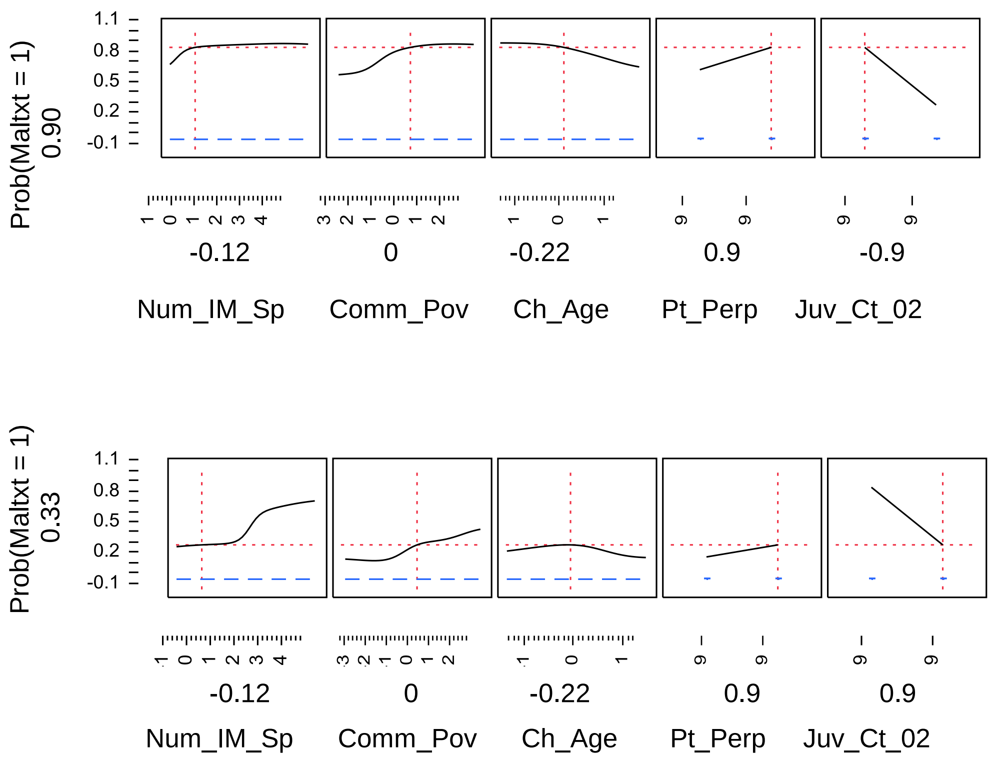


Figure 4.19. Neural network probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no juvenile court involvement in the first row to juvenile court involvement in the second row. Without juvenile court involvement, the average likelihood of repeat maltreatment was 90%; with juvenile court involvement, the average likelihood of repeat maltreatment was 33%.

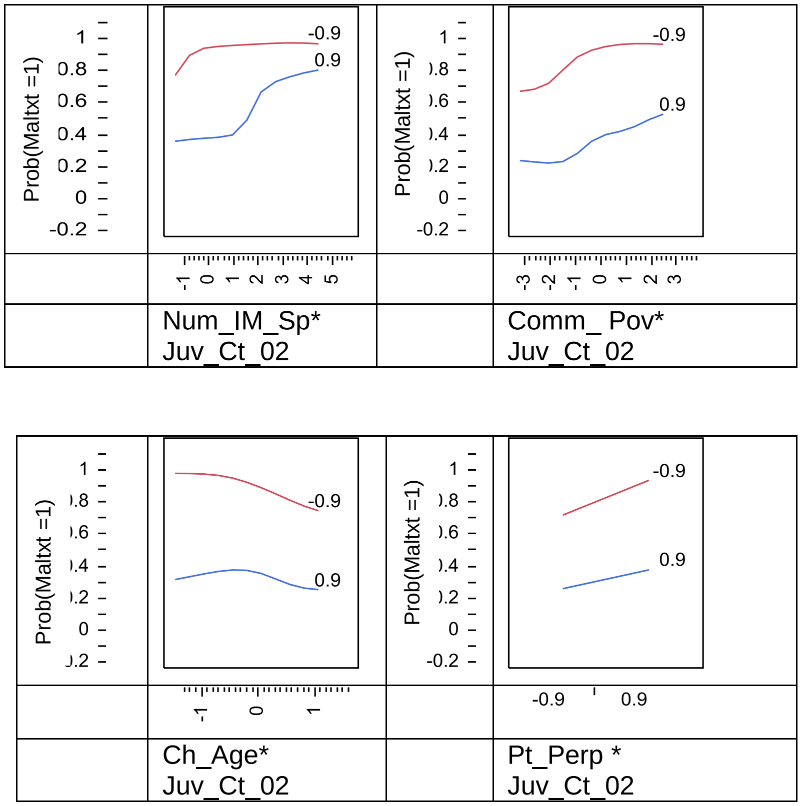


Figure 4.20. Neural network interaction plots where each of the four prominent risk factors was moderated by juvenile court involvement. The average probability of recurrent maltreatment on the *y*-axis is shown in association with each of the four primary risk factors on the *x*-axis. Each plot contains two slopes: (a) one slope for the association between the average probability of recurrent maltreatment and each risk factor when juvenile court involvement was set to "-0.9" ("no" involvement), and (b) the second slope for the association between the average probability of recurrent maltreatment and each risk factor when juvenile court involvement was set to "0.9" ("involvement").

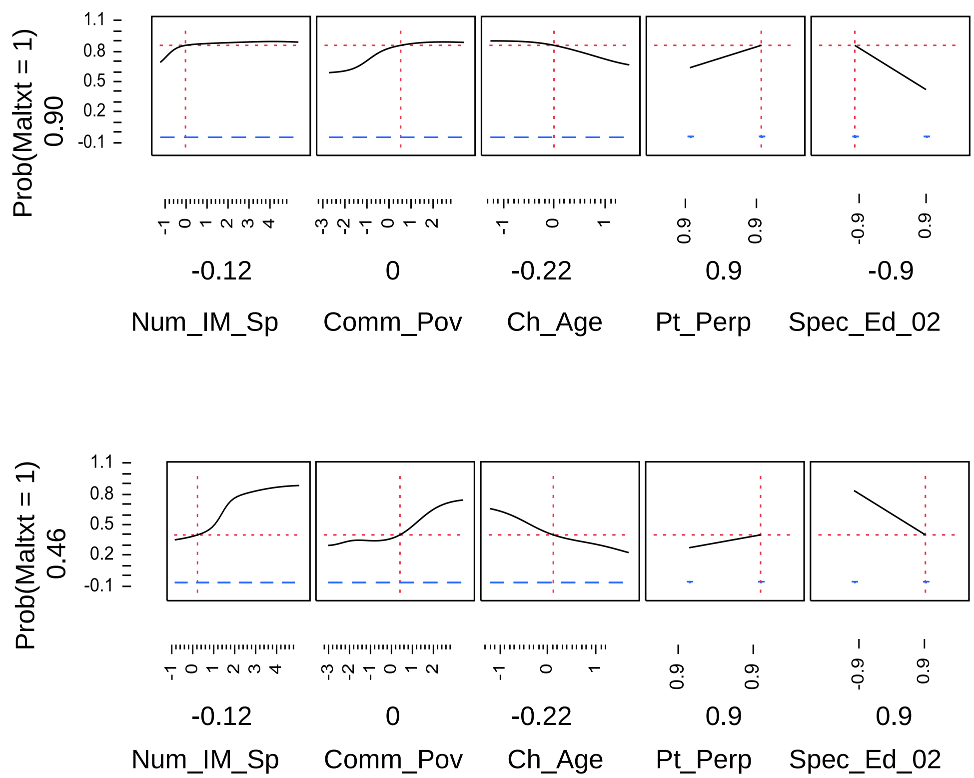


Figure 4.21. Neural network probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no special education eligibility in the first row to special education eligibility in the second row. Without special education involvement, the average likelihood of repeat maltreatment was 90%; with special education eligibility, the average likelihood of repeat maltreatment was 46%.

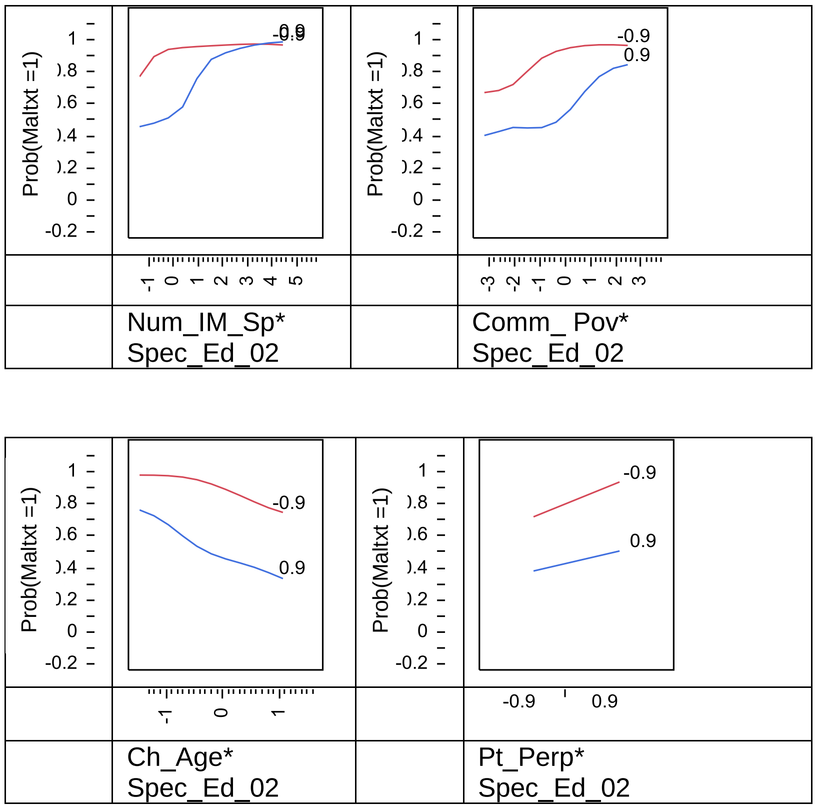


Figure 4.22. Neural network interaction plots where each of the four prominent risk factors was moderated by special education eligibility. The average probability of recurrent maltreatment on the *y*-axis is shown in association with each of the four primary risk factors on the *x*-axis. Each plot contains two slopes: (a) one slope for the association between the average probability of recurrent maltreatment and each risk factor when special education eligibility was set to "-0.9" ("no" eligibility), and (b) the second slope for the association between the average probability of recurrent maltreatment and each risk factor when special education eligibility was set to "0.9" ("eligibility").

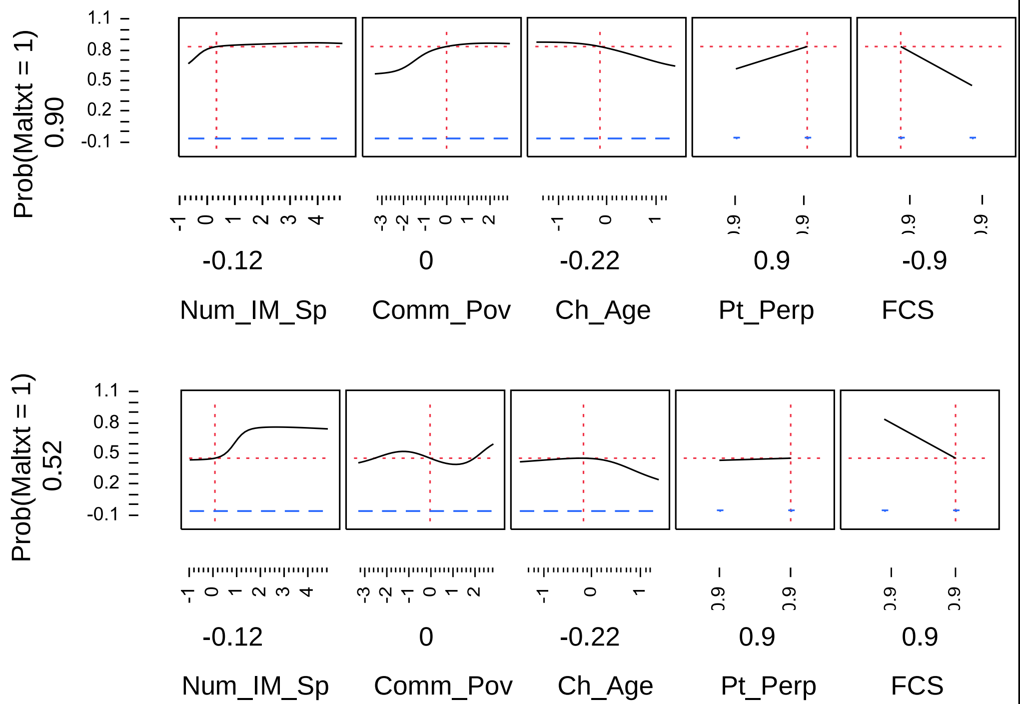


Figure 4.23. Neural network probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no FCS receipt in the first row to FCS receipt in the second row. Without FCS receipt, the average likelihood of repeat maltreatment was 90%; with FCS receipt, the average likelihood of repeat maltreatment was 52%.

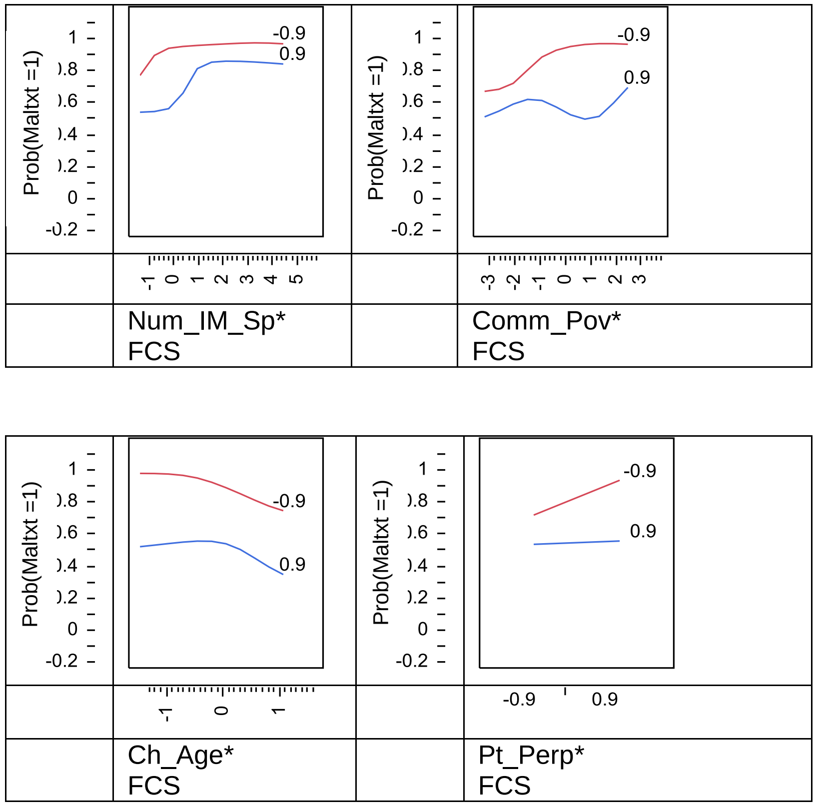


Figure 4.24. Neural network interaction plots where each of the four prominent risk factors was moderated by FCS receipt. The average probability of recurrent maltreatment on the *y*-axis is shown in association with each of the four primary risk factors on the *x*-axis. Each plot contains two slopes: (a) one slope for the association between the average probability of recurrent maltreatment and each risk factor when FCS receipt was set to “-0.9” (“no” receipt), and (b) the second slope for the association between the average probability of recurrent maltreatment and each risk factor when FCS receipt was set to “0.9” (“receipt”).

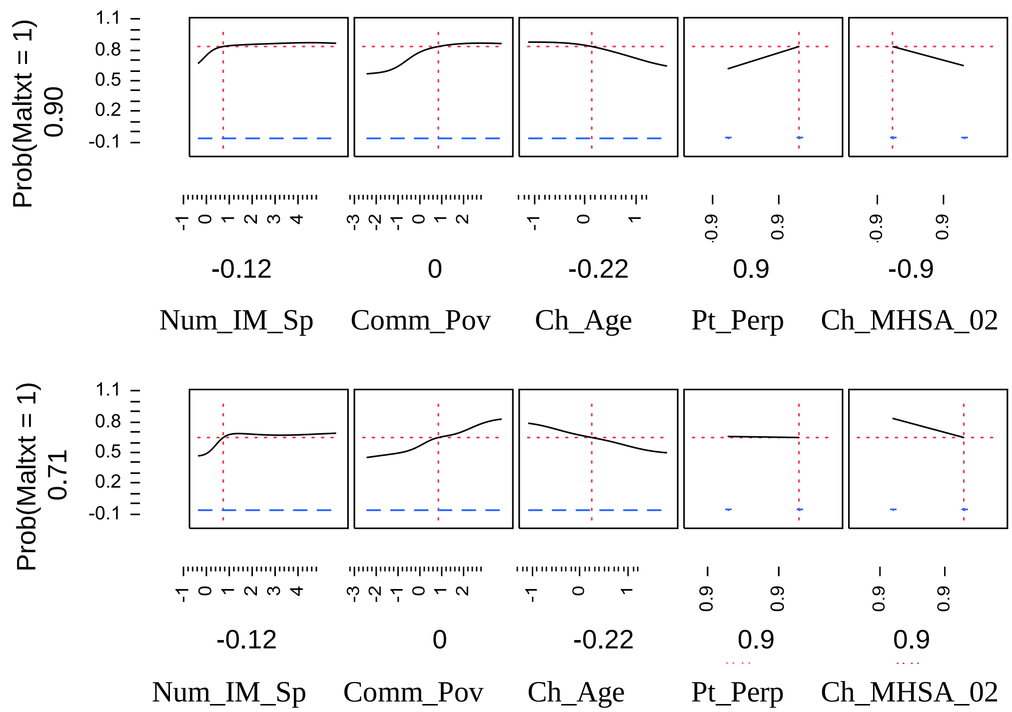


Figure 4.25. Neural network probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no child mental health/substance abuse treatment in the first row to child mental health treatment in the second row. Without child mental health/substance abuse treatment, the average likelihood of repeat maltreatment was 90%; with child mental health/substance abuse treatment, the average likelihood of repeat maltreatment was 71%.

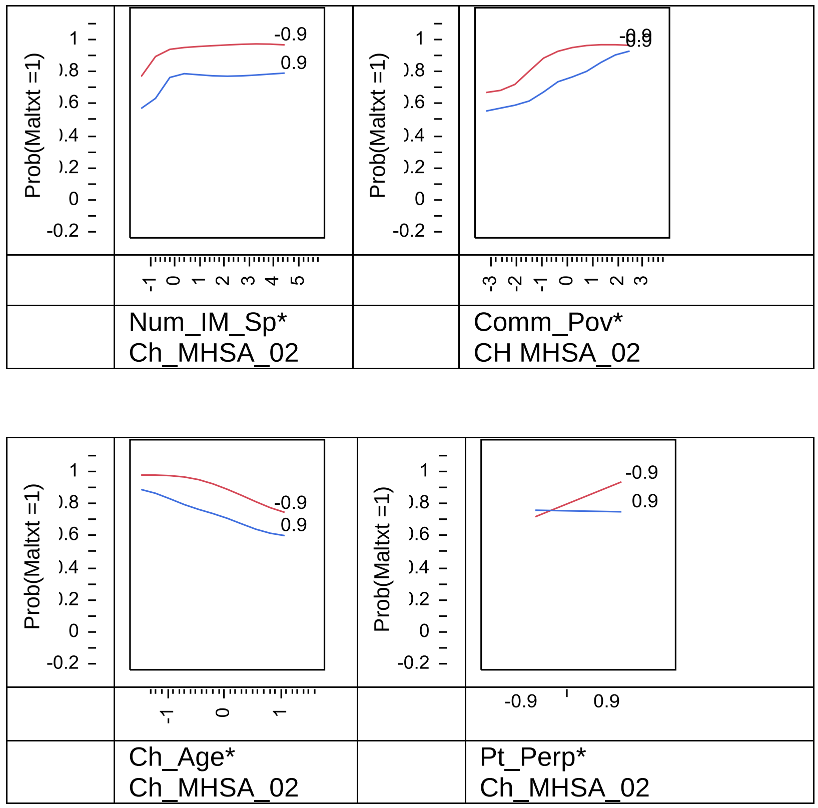


Figure 4.26. Neural network interaction plots where each of the four prominent risk factors was moderated by child receipt of mental health/substance abuse treatment. The average probability of recurrent maltreatment on the *y*-axis is shown in association with each of the four primary risk factors on the *x*-axis. Each plot contains two slopes: (a) one slope for the association between the average probability of recurrent maltreatment and each risk factor when child mental health/substance abuse treatment was set to "-0.9" ("no" treatment), and (b) the second slope for the association between the average probability of recurrent maltreatment and each risk factor when child mental health/substance abuse treatment was set to "0.9" ("treatment").

Probability plots for the logistic regression model.

To provide a counterpoint to the neural network probability plots, the same sets of plots were generated using the predicted probabilities from the logistic regression model and featured in Figures 4.27, 4.28, 4.29, and 4.30. The differences in the ways in which each statistical procedure estimated the functional form of each bivariate relationship (while holding all other predictors constant) are easily detected. In the logistic regression model, parameter estimates were adjusted to improve model fit but in the neural network model, the functional forms of the predictors were adjusted to improve model fit. By comparing the probability plots for each model, comparisons can be made as to how each model estimated the average probability of maltreatment recurrence in relationship to the four prominent risk factors and the four key service variables.

In the neural network model, the average predicted probability of recurrent maltreatment was estimated as a curvilinear function of child age at the first maltreatment report, the number of income maintenance spells received, and community-level poverty. Additionally, in the neural network model, relationships between the average predicted probability of recurrent maltreatment and each of the four prominent risk factors were estimated as being conditional on one or more of the four key service variables. For example, inspection of the interaction plots reveals that juvenile justice involvement moderated the number of income maintenance spells received, community-level poverty, and child age at the first maltreatment report. In each of these cases, the curvilinear regression lines featured in the interaction plot noticeably diverged.

The changes in the average predicted probability of recurrent maltreatment in association with increases in the given risk factor followed different courses depending on the value for juvenile justice involvement. Special education eligibility also moderated the number of income maintenance spells received, community-level poverty, and child age at the first maltreatment report. In each case, an examination of the interaction plots reveals that the curvilinear regression lines were not parallel. The shapes taken by each curvilinear regression line are noticeably different. Receipt of family centered services moderated the number of income maintenance spells received, community-level poverty, and child age at the first maltreatment report; moreover, receipt of family centered services moderated the parent's status as the perpetrator of the first maltreatment incident. Examination of the interaction plots reveals substantial differences in the curvilinear shapes the regression lines take for each risk factor as the setting for the family-centered service variable transitions from -0.9 or "no service" to 0.9 or "service receipt." Notice the degree to which the linear regression lines for the parent's status as perpetrator of the first maltreatment incident diverge. When family centered services were received (i.e., the variable was set to 0.9), the regression line for the average predicted probability of recurrent maltreatment in relationship to the parent's perpetrator status is largely flat and runs in a horizontal fashion. However, when family centered services were not received (i.e., the variable was set to -0.9), the regression line for the average predicted probability of recurrent maltreatment in relationship to the parent's perpetrator status has a considerably steeper slope and indicates a sharp increase in the predicted probability of repeat maltreatment as the parent's status as perpetrator switched from "no" (-0.9) to "yes" (0.9).

Finally, the child's receipt of a mental health/substance abuse service moderated the number of income maintenance spells received, community-level poverty, and the parent's status as the perpetrator of the first maltreatment incident. While the overall shape of the curvilinear regression lines are somewhat similar for the relationship between the average predicted probability of repeat maltreatment and the number of income maintenance spells received, there are some noticeable differences. First, the initial uptick in the predicted probability of recurrent maltreatment begins later for children who received a mental health/substance abuse service. Hence, the uptick in the predicted probability of recurrent maltreatment does not sharply increase until higher values for the number of income maintenance spells received are reached. However, once the uptick in the predicted probability of repeat maltreatment begins for children who received a mental health/substance abuse service, the increase in the probability of a subsequent report is steeper in comparison to the more consistent increase in the probability of a subsequent report for children who did not receive a mental health/substance abuse service.

When looking at the probability plots for the logistic regression model, it is clear that there are no curvilinear relationships and no interaction effects. Hence, all relationships between the probability of maltreatment and each predictor were estimated to be monotonic. Moreover, the slopes appear to be far more subtle in their effects. For example, while the direction of the effect of juvenile court involvement is the same for the logistic regression and the neural network models, the predicted probability of recurrent maltreatment in conjunction with juvenile court involvement (while holding all other predictors constant) decreased by 57% in the neural network model but only by 33% in the logistic regression model. Similar differences in the reduction of the probability of recurrent maltreatment in conjunction with service receipt/service involvement occurred for special education eligibility, receipt of family centered services, and the child’s receipt of a mental health/substance abuse service. In each case, the estimated decrease in the probability of repeat maltreatment was lower for the logistic regression model as compared with the neural network model, which appears to perform better given the advanced capacity to model interaction effects (Beck, King, & Zeng, 2000). Specifically, the decrease in the probability of repeat maltreatment when the child was special education eligible was 44% for the neural network model and 13% for the logistic regression model. Additionally, the decrease in the probability of repeat maltreatment when family centered services were received was 38% for the neural network model and 18% for the logistic regression model. Finally, the decrease in the probability of repeat maltreatment when the child received a mental health/substance abuse service was 19% for the neural network model and 11% for the logistic regression model.

Differences in the strength of the effects of the service-based predictors can also been seen when comparing the relative steepness of the slopes for each predictor across the logistic regression and neural network models. In studies that compared the predicted probability of international conflict across a neural network and a binary logistic regression model, Beck, King, and Zeng (2000, 2004) also found that the effects of explanatory variables as estimated by a logistic regression model were smaller in magnitude as compared with a neural network model. Additionally, changes in the effects of explanatory variables were much stronger for the neural network model when values for the ex ante probability of conflict were moved from low to high.

Please see Figures 4.27, 4.28, 4.29, and 4.30 to examine the probability plots for the logistic regression model.

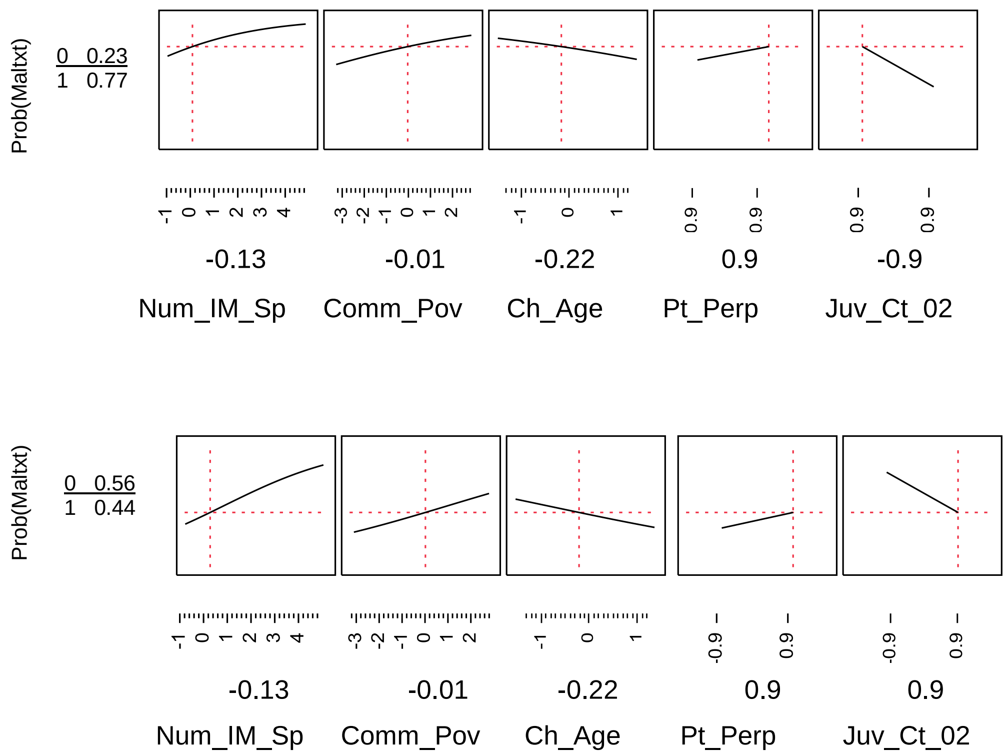


Figure 4.27. Logistic regression probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no juvenile court involvement in the first row to juvenile court involvement in the second row. Without juvenile court involvement, the average likelihood of repeat maltreatment was 77%; with juvenile court involvement, the average likelihood of repeat maltreatment was 44%.

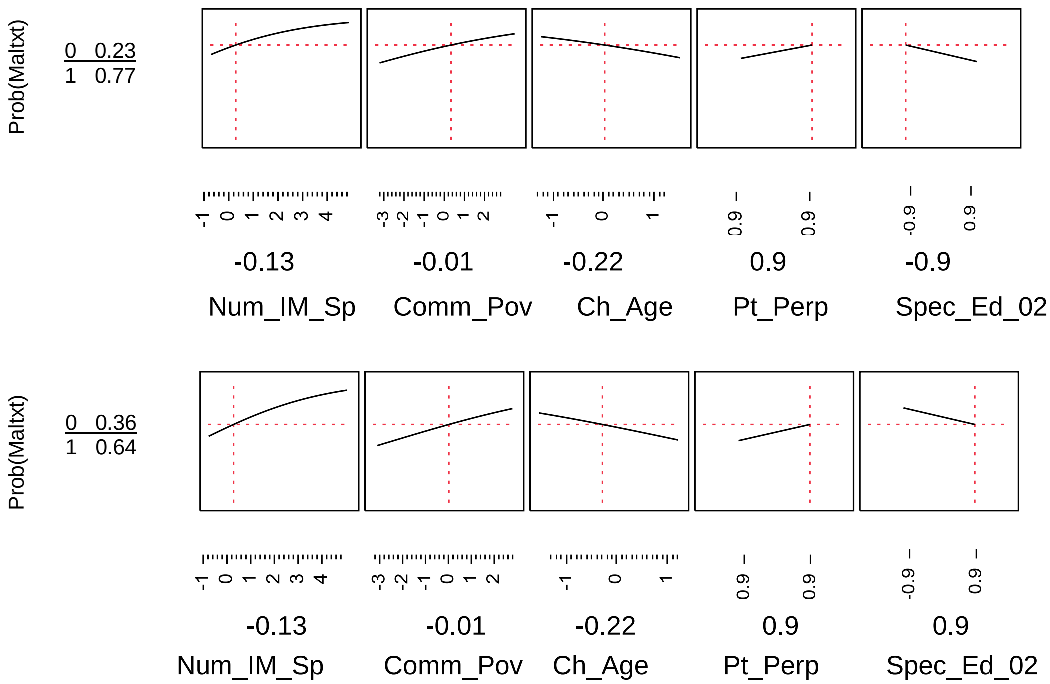


Figure 4.28. Logistic regression probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no special education eligibility in the first row to special education eligibility in the second row. Without special education eligibility, the average likelihood of repeat maltreatment was 77%; with special education eligibility, the average likelihood of repeat maltreatment was 64%.

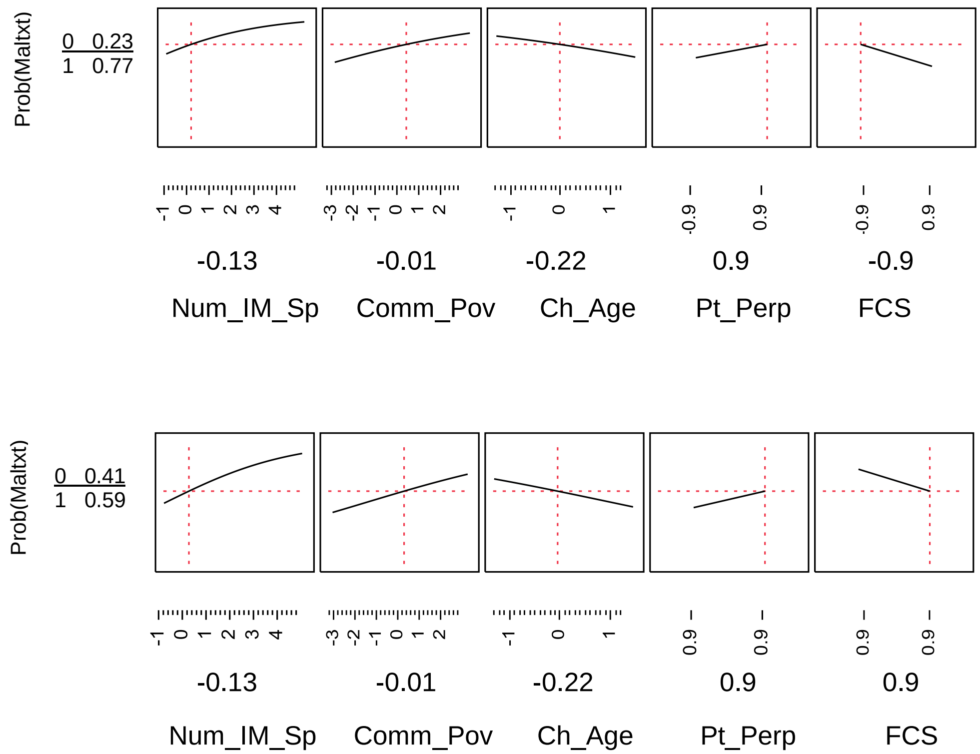


Figure 4.29. Logistic regression probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no FCS receipt in the first row to FCS receipt in the second row. Without FCS receipt, the average likelihood of repeat maltreatment was 77%; with FCS receipt, the average likelihood of repeat maltreatment was 59%.

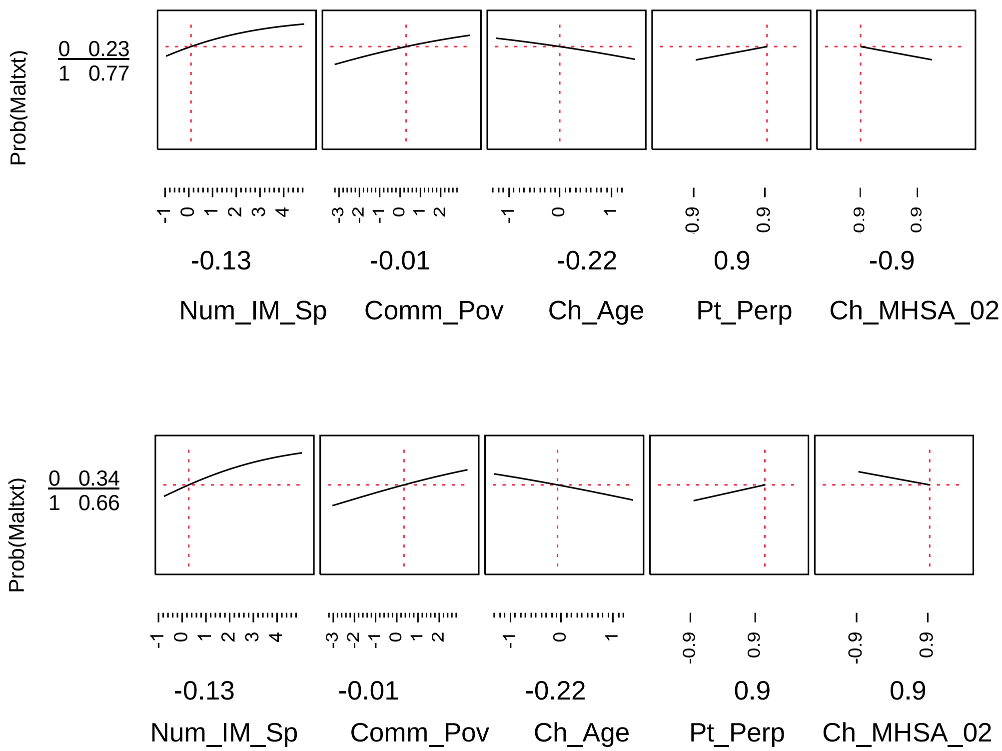


Figure 4.30. Logistic regression probability plots with the average probability of recurrent maltreatment on the *y*-axis and four prominent risk factors on the *x*-axis of the first four respective plots. The fifth plot features the effect of transitioning from no child mental health/substance abuse treatment in the first row to child mental health/substance abuse treatment in the second row. Without child mental health/substance abuse treatment, the average likelihood of repeat maltreatment was 77%; with child mental health/substance abuse treatment, the average likelihood of repeat maltreatment was 66%.

Neural network effects after including the logistic regression-estimated probabilities of repeat maltreatment.

An additional method for evaluating the degree to which a neural network model finds structure in the data above and beyond what would be estimated by a logistic regression model is to include the predicted probabilities of repeat maltreatment as estimated by a logistic regression model as a separate predictor in a neural network analysis (Stine, 2011). Hence, a second neural network analysis was conducted where predicted probabilities of repeat maltreatment for each case record were included as a separate explanatory variable ("Z_RE_PROB") along with all of the originally included explanatory variables. In theory, if there is no structure in the data beyond what would be modeled by a linear combination of the explanatory variables, then the effects of all originally included predictors should be flattened (essentially made null and void) upon including the predicted probabilities of maltreatment estimated by the logistic regression model (Stine, 2011). If the neural network has nothing to add beyond what can be estimated by a linear model, then the slope of the logistic regression-predicted probabilities of maltreatment should be the only slope that demonstrates any real steepness or pitch. All other slopes should resemble horizontal lines. Additionally, there should be no curvilinear effects in the probability plots aside from the very subtle S-shape of the slope for the logistic regression-generated predicted probabilities of maltreatment (Stine, 2011).

As can be seen in Figures 4.31 and 4.32 below, even after introducing the logistic regression-predicted probabilities of maltreatment, the effects of many of the original predictors (to include the majority of the four prominent risk factors and the four key service variables) remain strong, to include curvilinear effects for the number of income maintenance spells received and exposure to community poverty. For the probability plots in Figures 4.31 and 4.32, all continuous variables were held constant at their mean (~0) and all categorical predictors were held constant at -0.9 ("no or "non" response level). However, Figures 4.31 and 4.32 do not adequately represent the interaction effects that the neural network continues to estimate above and beyond the linear effects estimated by the logistic regression model. It is the continued presence of ongoing interaction effects that provides additional evidence of the neural network's utility in discovering structure in the data that would otherwise be missed by a linear model. Figures 4.33 and 4.34 include probability plots for the logistic regression-predicted probabilities of recurrent maltreatment and the four prominent risk factors, with one service variable set to 0.9 ("service received"). In Figure 4.33, the child's receipt of a mental health/substance abuse variable has been turned on and appears to moderate the number of income maintenance spells received, exposure to community poverty, and child age at the first maltreatment report.

Evidence of these interactions can be seen by comparing the shape of the curvilinear regression lines for each of the three risk factors between the probability plots in Figure 4.31 (where the all service variables were turned off) and Figures 4.33 and 4.34 (where the child's receipt of a mental health/substance abuse service and special education eligibility were turned on, respectively). Additionally, evidence of these interactions can be seen by comparing the shapes of the curvilinear regression lines included in each interaction plot in Figures 4.33 and 4.34. For example, when the child's receipt of a mental health/substance abuse service is the moderator variable (please see Figure 4.33), the curvilinear regression lines intersect within the interaction plots for the number of income maintenance spells received and the child's age at the first maltreatment report.

Additionally, the shapes of the curvilinear regression lines in the plot for community poverty are discernibly different. When the child's receipt of a mental health/substance abuse service is set to -0.9 ("no" service receipt), the predicted probability of recurrent maltreatment in relationship to exposure to community poverty holds steady before decreasing slightly and then increasing. In contrast, when the child's receipt of a mental health/substance abuse service is set to 0.9 ("service received"), the predicted probability of recurrent maltreatment in relationship to exposure to community poverty decreases slightly before increasing and then reaching a point where the increase tapers off.

Figure 4.34 includes interaction plots when the child's eligibility for special education services was turned on. The shapes of the curvilinear regression lines in all three interaction plots are discernibly different. Hence, changes in the predicted probability of recurrent maltreatment in relationship to each of the three risk factors are not constant across values for the child's eligibility for special education services.

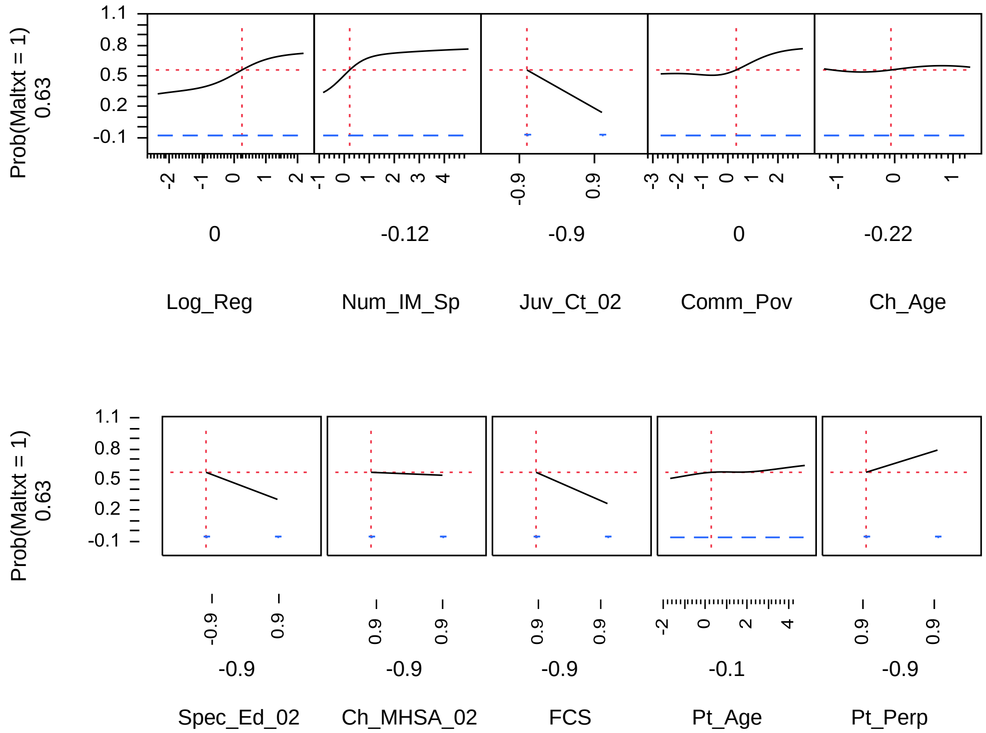


Figure 4.31. Neural network probability plots that include all original predictors plus the predicted probabilities of maltreatment estimated by the binary logistic regression model with random intercepts. Note the continued presence of strong effects across multiple predictors with sharply pitched slopes and curvilinear effects for the number of income maintenance spells received and community poverty (child age at the first maltreatment report and parent age at the child's birth are mildly curvilinear). Hence, the neural network continued to add information about the relationships between the predicted likelihood of recurrent maltreatment and included predictors above and beyond what the logistic regression model estimated.

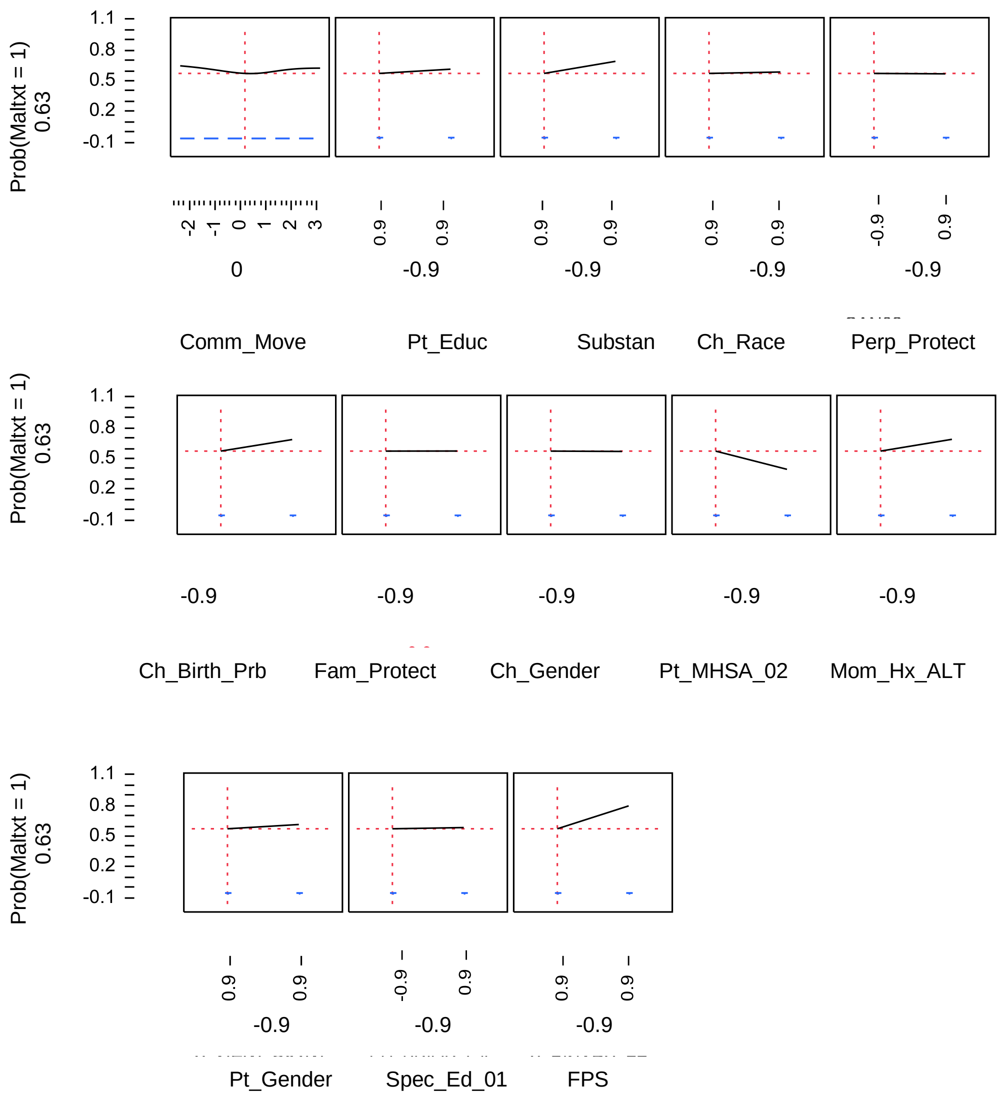


Figure 4.32. Neural network probability plots that include all original predictors plus the predicted probabilities of maltreatment estimated by the binary logistic regression model with random intercepts. Note the continued presence of strong effects across multiple predictors with sharply pitched slopes and curvilinear effects for the number of income maintenance spells received and community poverty (child age at the first maltreatment report and parent age at the child's birth are mildly curvilinear). Hence, the neural network continued to add information about the relationships between the predicted likelihood of recurrent maltreatment and included predictors above and beyond what the logistic regression model estimated.

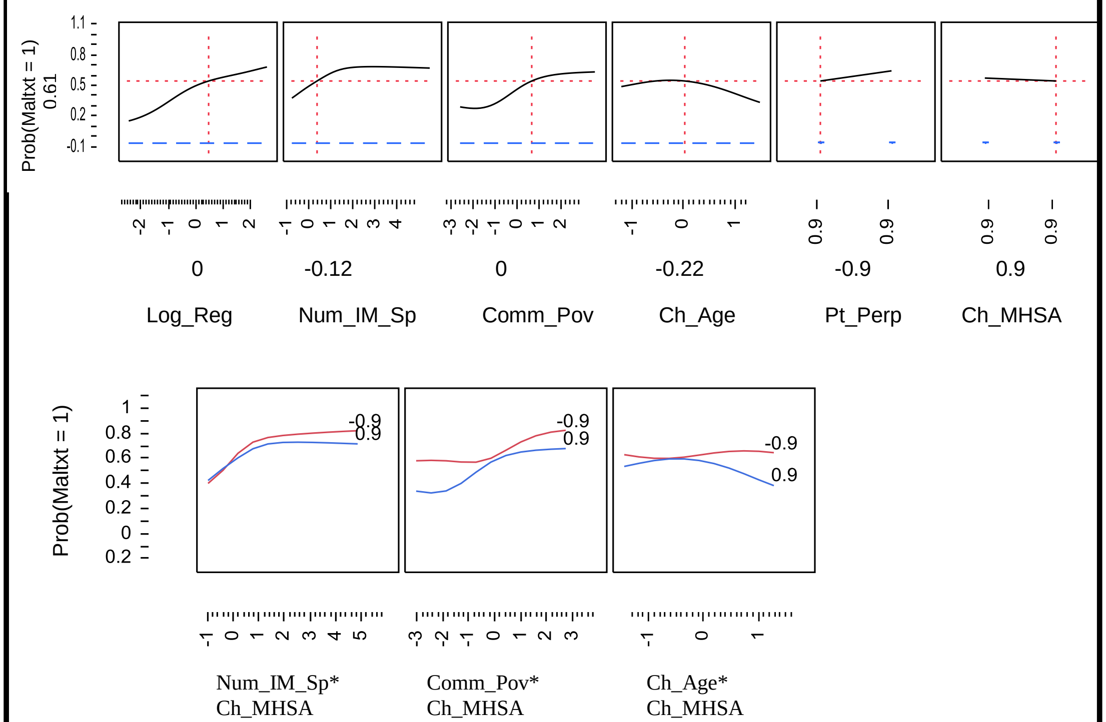


Figure 4.33. Neural network probability plots where the predicted probabilities of maltreatment estimated by the logistic regression model were included as an additional predictor. The first row of plots show the changes in the slopes of the four primary risk factors when the child's receipt of mental health/substance abuse treatment was set to "0.9" for "receipt" of treatment. The second row contains three interaction plots with two slopes for each risk factor: (a) one slope when the child's receipt of mental health/substance abuse treatment was set to "-0.9" or "no receipt" of treatment, and (b) a second slope when the child's receipt of mental health/substance abuse treatment was set to "0.9" or "receipt" of treatment.

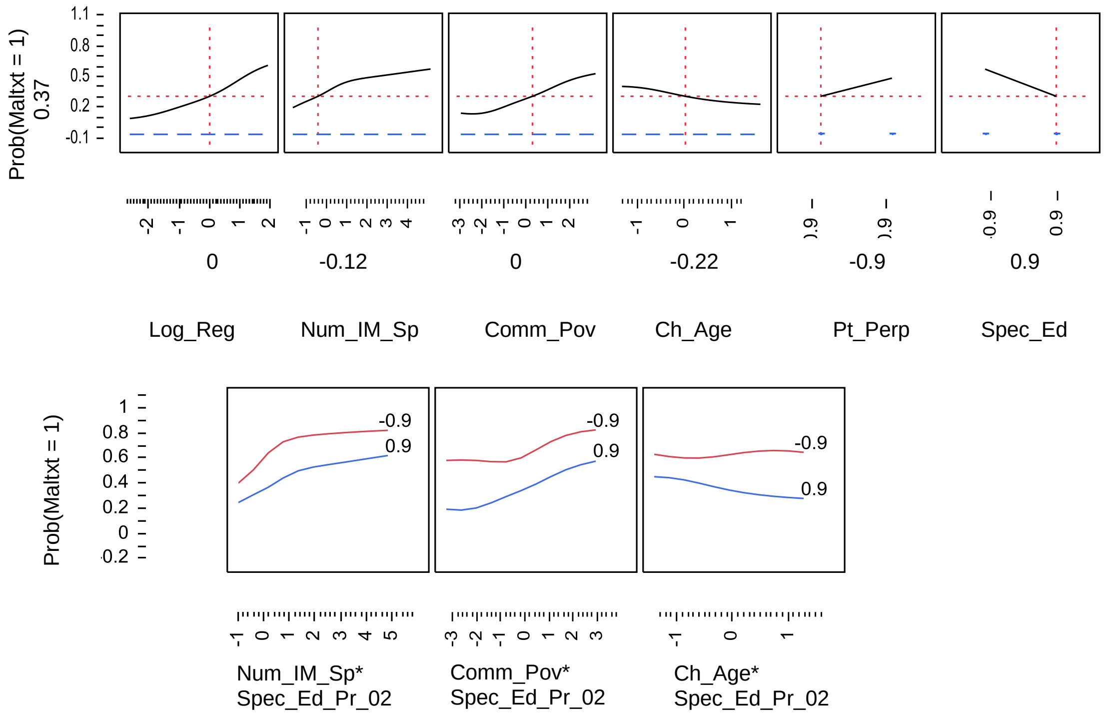


Figure 4.34 Neural network probability plots where the predicted probabilities of maltreatment estimated by the logistic regression model were included as an additional predictor. The first row of plots show the changes in the slopes of the four primary risk factors when the child's special education eligibility was set to "0.9" for "eligibility." The second row contains three interaction plots with two slopes for each risk factor: (a) one slope when the child's special education eligibility status was set to "-0.9" or "not" eligible, and (b) a second slope when the child's special education eligibility status was set to "0.9" or "eligible."

Three-dimensional and two-dimensional plots of the neural network's predicted probability of maltreatment by selected risk factors and service variables.

This final section includes a number of two-dimensional and three-dimensional graphs plotting the neural network-predicted probability of recurrent maltreatment by selected risk factors, service variables, and the observed outcome (i.e., observed values for recurrent maltreatment). The figures that follow provide visual assistance in understanding (a) the degree to which selected risk factors and service variables differentiate children by their likelihood of maltreatment re-report, and (b) the relative complexity in using combinations of predictors to differentiate children who are likely to be re-reported for maltreatment from children who are not likely to be re-reported for maltreatment Figures 4.35 through 4.42 provide three-dimensional views of the probability of recurrent maltreatment in association with two risk factors while holding all other predictors constant. The surface plot contained in each of the Figures 4.35 through 4.42 represents all of the data points (i.e., the plotted values of the predictors for each observation) that satisfy the equation (i.e., the target function) estimated by the neural network model to predict the probability of recurrent maltreatment.

Figure 4.35 provides a three-dimensional view of the relationship between the neural network-predicted probability of maltreatment and (a) the number of income maintenance spells received; and (b) the child's age at the first maltreatment report, while holding all other predictors constant to include holding all service variables constant at their respective "no or non-delivery" levels. All continuous predictors were held constant at their means (~0), the parent's status as the perpetrator was held constant at 0.9 (the parent was the perpetrator), and all remaining categorical predictors were held constant at -0.9 ("no" or "non" categories). Figure 4.35 nicely demonstrates the complexity of the surface that represents the relationship between the predicted probability of maltreatment and two key risk factors. Notice the lack of a smooth surface area; instead, see the ways in which the surface area curves with ripples and bumps that push the surface area into differing levels of elevation.

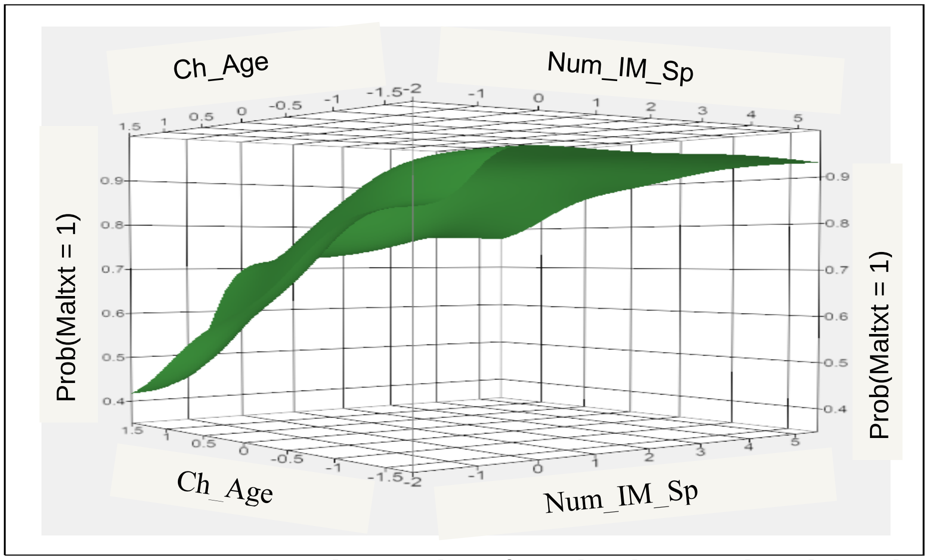


Figure 4.35. Neural network surface plot showing the relationship between the predicted probability of recurrent maltreatment and two key risk factors: number of income maintenance spells received and child age at the first maltreatment report. All service variables were held constant at -0.9 (no service delivery/system intervention). The neural network-predicted surface is far more complex than the logistic regression-predicted surface, which only allows for monotonically increasing or decreasing relationships.

In contrast, Figure 4.36 demonstrates the relatively smooth and even surface that characterizes the same relationship between the probability of recurrent maltreatment and the number of income maintenance spells plus the child's age at the first maltreatment report. The only difference between the surface plots in Figures 4.35 and 4.36 is that the neural network model generated the more complex surface area featured in Figure 4.35

and the logistic regression model generated the less complex surface area featured in Figure 4.36.

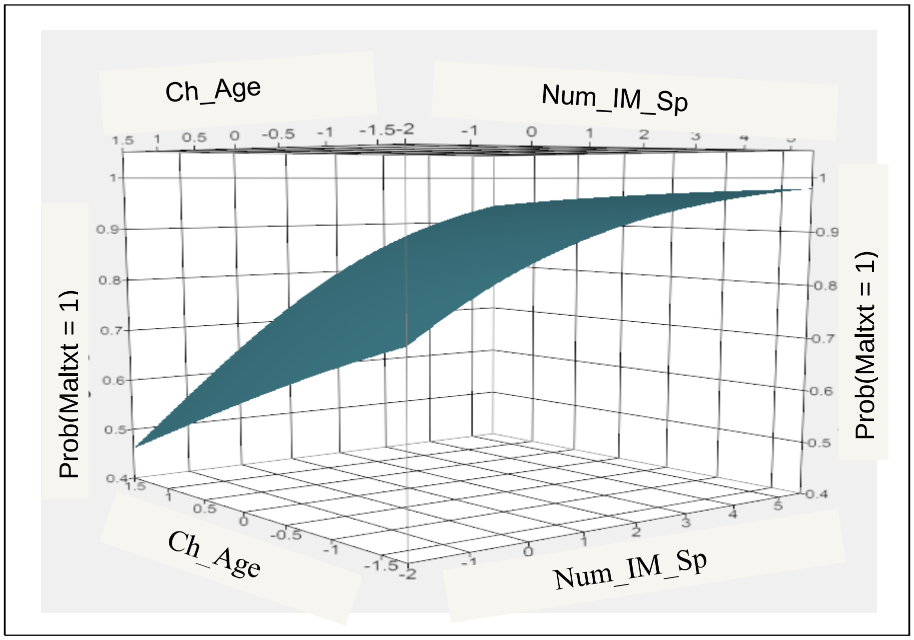


Figure 4.36. Logistic regression surface plot showing the relationship between the predicted probability of recurrent maltreatment and two key risk factors: number of income maintenance spells received and child age at the first maltreatment report. All service variables were held constant at zero (no service delivery/system intervention). The logistic regression-predicted surface resembles a straight piece of 81/2 x 11 paper with a flat as opposed to a curvilinear surface plus a slight S-shaped bend.

Changes in the relationship between the probability of recurrent maltreatment and the number of income maintenance spells plus the child's age at the first maltreatment report can be seen in Figure 4.37 when the juvenile court involvement variable was turned on in the neural network model. The surface area remains complex, but the overall shape of the surface area and the specifics of its curvature have been considerably altered. In contrast, Figure 4.38 shows changes in the relationship between the probability of recurrent maltreatment and the number of income maintenance spells plus the child's age at the first maltreatment report when the juvenile court involvement variable was turned on in the logistic regression model. Notice how little the surface area has changed. Figure 4.39 combines the surface areas for both the neural network model and the logistic regression model into one graph (when the juvenile court involvement variable has been turned on for both models). Notice how different the surface areas are for both models. If a linear model were sufficient for identifying structure in the data, then the surface areas for both models would lie in a parallel fashion to each other.

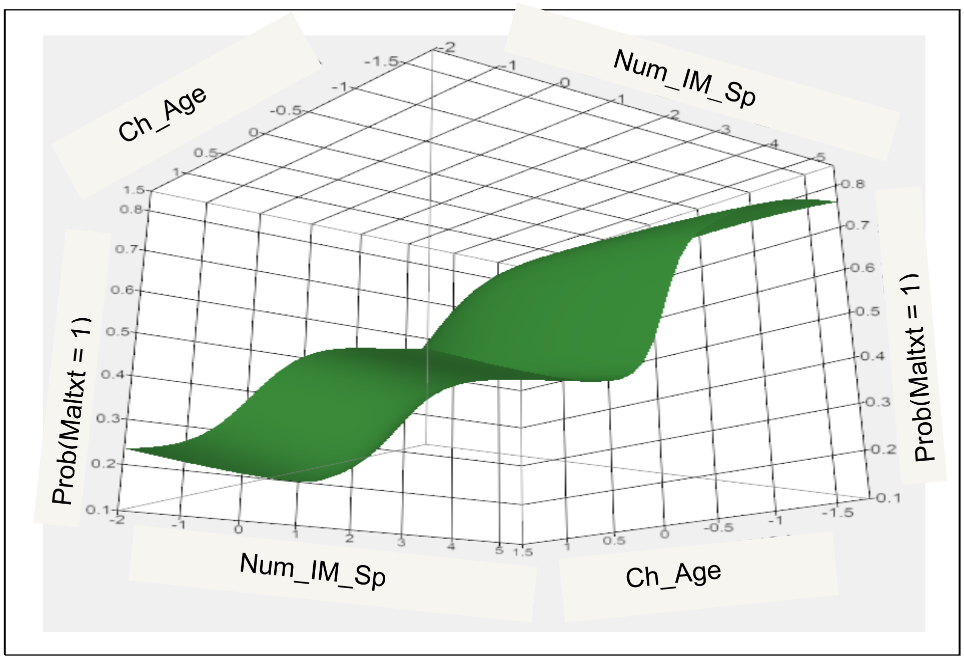


Figure 4.37. Neural network surface plot showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with juvenile court involvement set to "0.9" ("yes" for involvement).

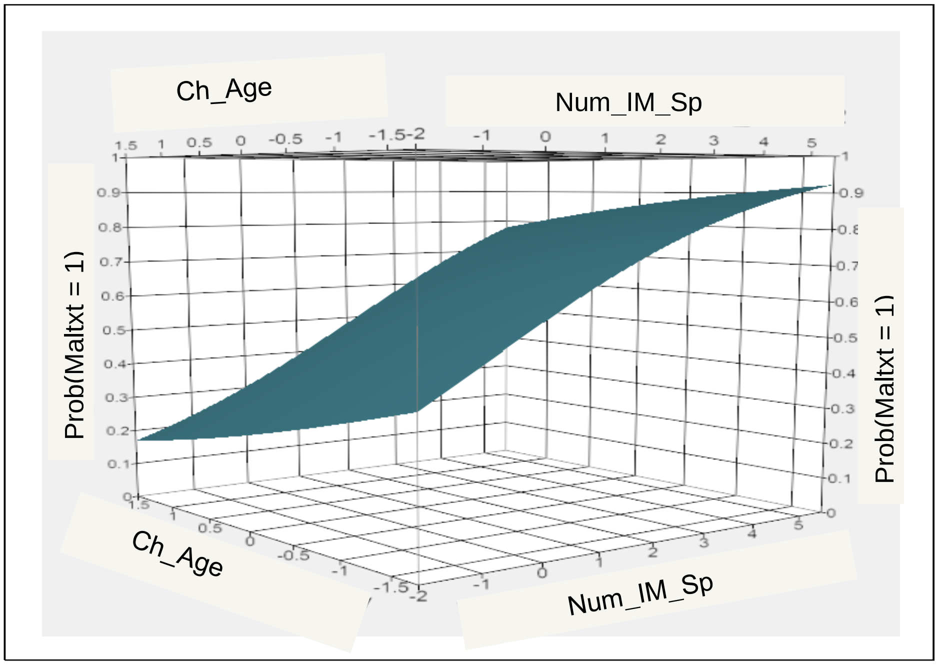


Figure 4.38. Logistic regression surface plot showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with juvenile court involvement set to "1" ("yes" for involvement).

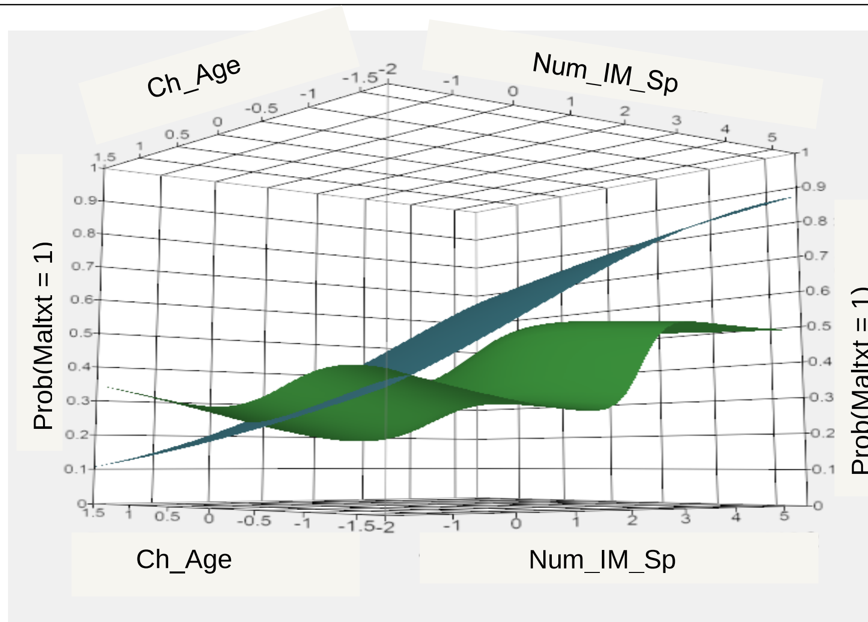


Figure 4.39. Comparison of the logistic regression (top) and neural network (bottom) surface plots showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with juvenile court involvement set to “0.9” or “1” (“yes” for involvement).

The remaining three-dimensional figures show changes in the relationship between the probability of recurrent maltreatment and the number of income maintenance spells plus the child’s age at the first maltreatment report when different service variables were turned on in the neural network model. Specifically, in Figure 4.40, the special education eligibility variable was the only service variable that was turned on, while in Figure 4.41, the family centered service variable was the only service-based variable that was turned on. Finally, in Figure 4.42 the child receipt of a mental health/substance abuse service was the only service-based variable that was turned on. In each case, the surface area remains complex, but the overall shape of the surface area and the specifics of its curvature are discernibly different as various service-based variables were turned on and different service-based variables were turned off.

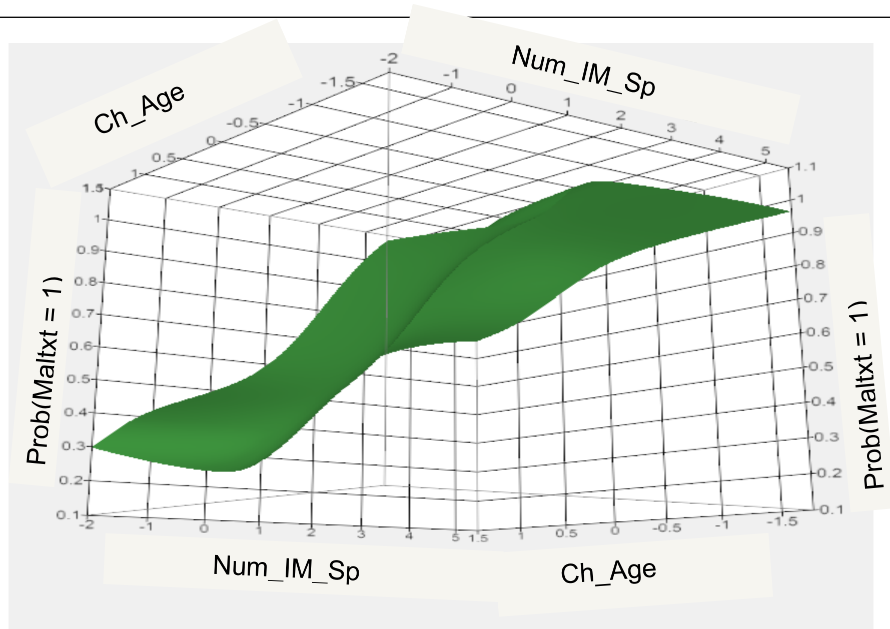


Figure 4.40. Neural network surface plot showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with special education eligibility set to “0.9” (“yes” for involvement).

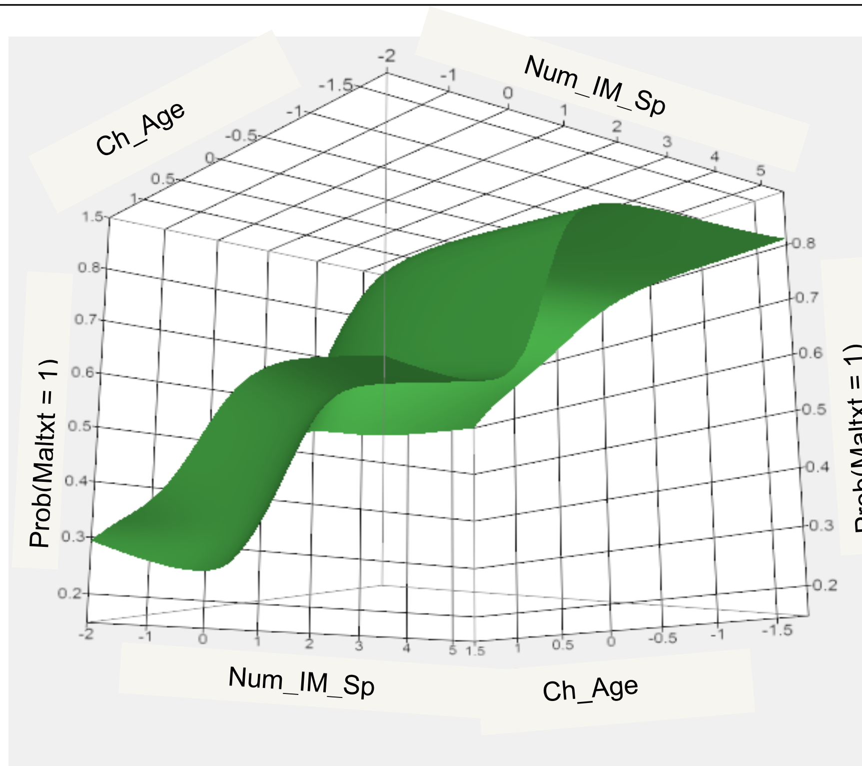


Figure 4.41. Neural network surface plot showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with FCS receipt set to “0.9” (“yes” for receipt).

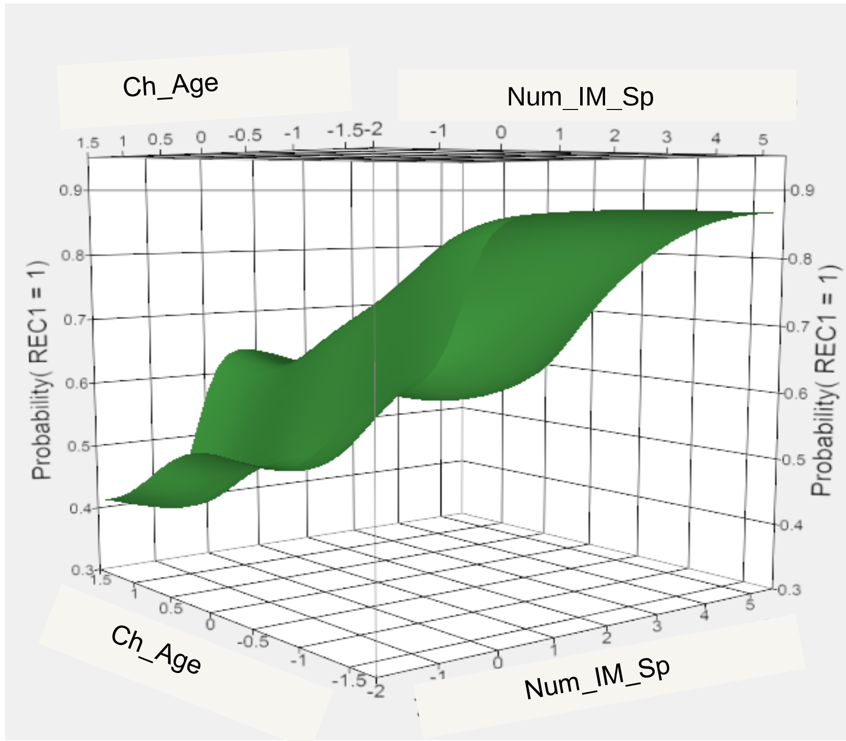


Figure 4.42. Neural network surface plot showing the relationship between the predicted probability of recurrent maltreatment and the number of income maintenance spells received in addition to child age at the first maltreatment report with child mental health/substance abuse treatment set to “0.9” (“yes” for treatment).

Figures 4.43 and 4.44 provide two-dimensional views of the relationship between the neural network-predicted probability of recurrent maltreatment and a service-based variable by values for the observed recurrent maltreatment variable. For example, Figure 4.43 plots the probability of recurrent maltreatment by the receipt of family centered services (FCS) by the observed values (i.e., the actual values) for the recurrent maltreatment outcome. The two boxplots towards the left of the figure and above the "0" setting for FCS receipt visually display the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that did not receive FCS. The two boxplots towards the right of the figure and above the "1" setting for FCS receipt visually display the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that received FCS. For the two boxplots on the left, the boxplot with the solid line displays the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that did not receive FCS and who were actually re-reported for maltreatment. The boxplot with the broken line displays the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that did not receive FCS and who were not actually re-reported for maltreatment. Likewise, for the two boxplots on the right, the boxplot with the solid line displays the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that received FCS and who were actually re-reported for maltreatment. The boxplot with the broken line displays the range of predicted probabilities of recurrent maltreatment estimated by the neural network for children in families that received FCS and who were not actually re-reported for maltreatment.

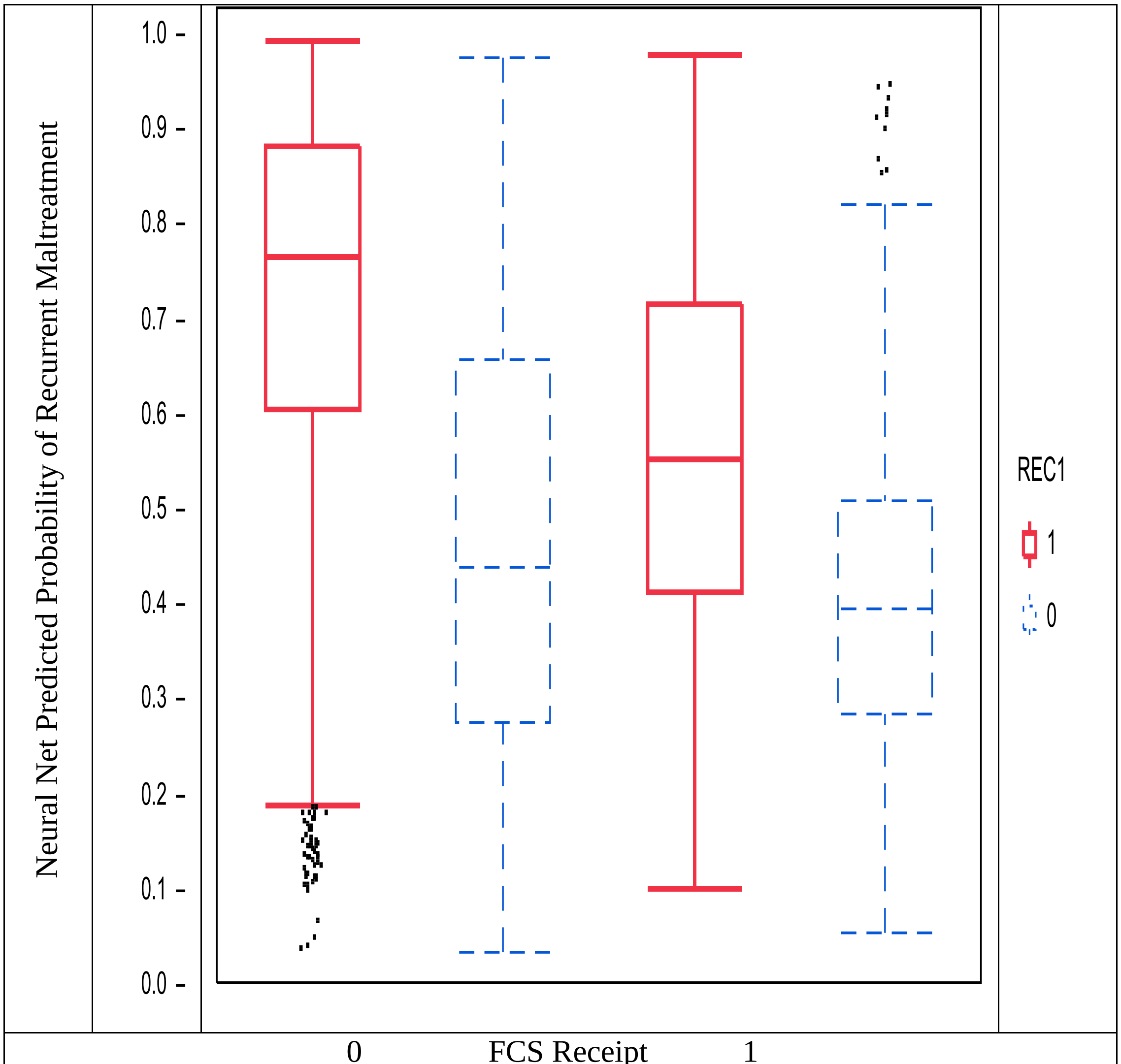


Figure 4.43. Neural network predicted probability of recurrent maltreatment vs. family centered service receipt by observed values for recurrent maltreatment.

Figure 4.43 facilitates an exploration of between and within group differences. In terms of examining between group differences, one looks for the degree to which the boxplots separate children who received FCS and who did not receive FCS into groups that are characterized by substantively different estimated probabilities of recurrent maltreatment. If FCS has a strong effect on the likelihood of recurrent maltreatment, then the boxplots for children who did and did not receive FCS should be positioned in discernibly different places along the range of predicted probabilities of maltreatment. If the FCS receipt is truly decisive in partitioning children into groups that are highly likely to be re-reported and highly unlikely to be re-reported, then there should be almost no overlap between the positioning of the boxplots along the *y*-axis. As can be seen in Figure 4.43, FCS receipt exerts some influence in that children falling within the interquartile range and within families that received FCS tend to have a lower range of estimated probabilities of recurrent maltreatment as compared with children falling within the interquartile range and within families that did not receive FCS. However, the differentiation between the predicted probabilities of recurrent maltreatment as a function of FCS receipt is far from stark. There is considerable overlap between the range of predicted probabilities of recurrent maltreatment for children who did and did not receive FCS, both in terms of overlap between children falling inside and outside of the interquartile range. Rather than exerting a strong direct influence or main effect on the predicted probability of recurrent maltreatment, the influence of FCS might be better captured in relationship to its ability to moderate the influence of prominent risk factors.

In terms of examining within group differences, one looks for the degree to which the boxplots separate children within each category of FCS receipt by the observed outcomes for maltreatment. Specifically, one evaluates the magnitude of differentiation or degree of separateness between the range of predicted probabilities of recurrent maltreatment for children who were actually re-reported versus the children who were not actually re-reported. If the neural network was accurate in estimating the predicted probabilities of recurrent maltreatment as a function of the predictors, and in this case as a function of FCS receipt, then the range of estimated probabilities of recurrent maltreatment for children who were actually re-reported should be substantively higher than the range of estimated probabilities for children who were not actually re-reported for maltreatment. As can be seen in Figure 4.43, children falling within the interquartile range and who were re-reported for maltreatment tend to have a lower range of estimated probabilities of recurrent maltreatment as compared with children falling within the interquartile range and who were not re-reported for maltreatment (this was true for children who did and did not receive FCS). However, the differentiation between the predicted probabilities of recurrent maltreatment for children who were and were not actually re-reported for maltreatment is not absolute. There is considerable overlap between the range of predicted probabilities of recurrent maltreatment for children who did and did not receive FCS, both in terms of overlap between children falling inside and outside of the interquartile range. Overall, in terms of between group and within group differentiation, FCS receipt appears to exert a moderate level of influence in being able to differentiate children who will and will not be re-reported for maltreatment. Generally speaking, children who received FCS fell within a reduced range of estimated probabilities of recurrent maltreatment.

As a point of reference, Figure 4.44 plots the probability of recurrent maltreatment by the juvenile court involvement by the observed values (i.e., the actual values) for the recurrent maltreatment outcome. Overall, in terms of between group and within group differentiation, juvenile court involvement appears to exert an even stronger level of influence in being able to differentiate children who will and will not be re-reported for maltreatment. Generally speaking, children who were juvenile court involved fell within a reduced range of estimated probabilities of recurrent maltreatment.

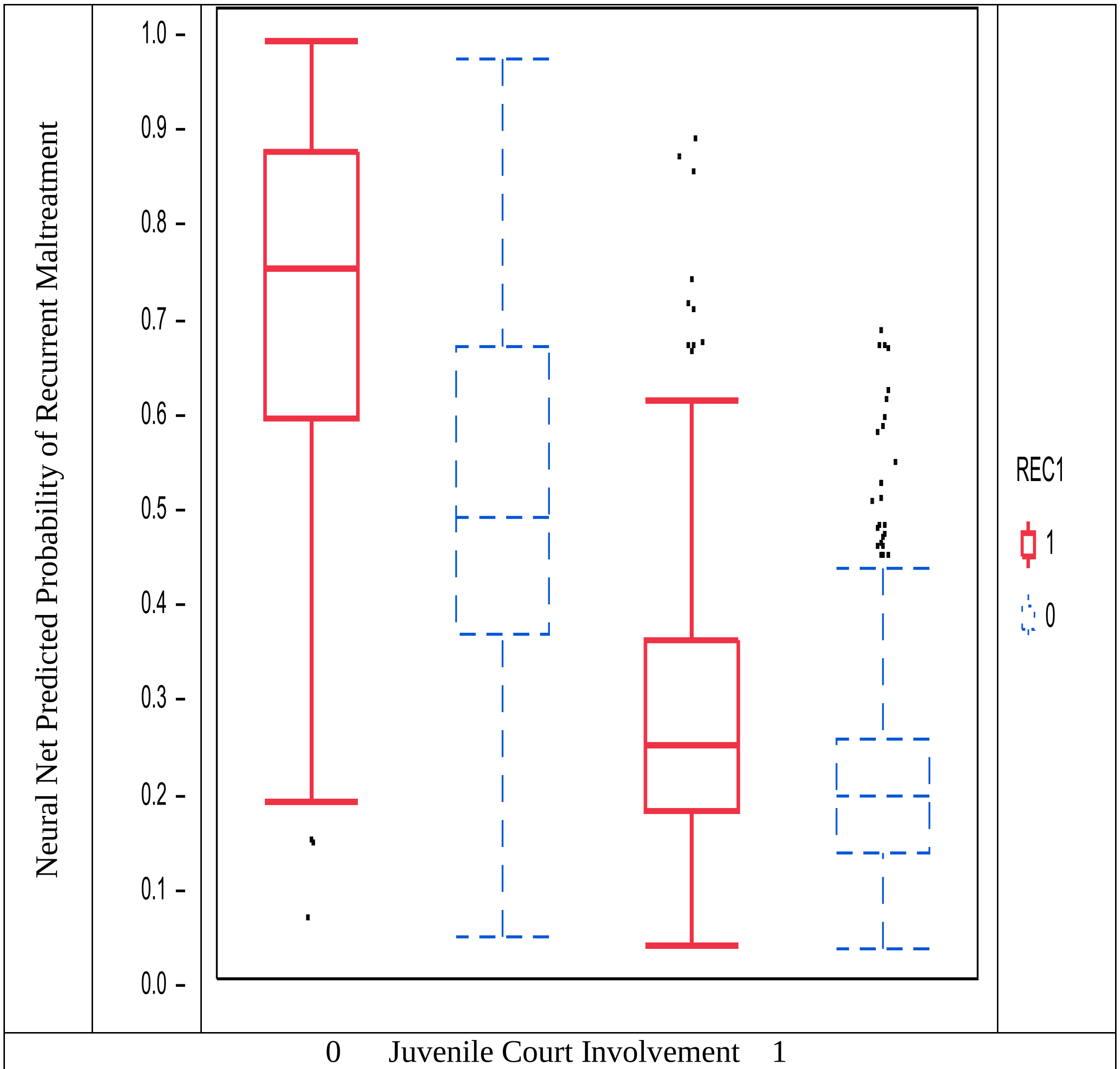


Figure 4.44. Neural network predicted probability of recurrent maltreatment vs. issued juvenile court petition by observed values for recurrent maltreatment.

Figures 4.45 through 4.47 provide two-dimensional views of the relationship between the neural network-predicted probability of recurrent maltreatment and a selected continuous risk factor by FCS receipt and by values for the observed recurrent maltreatment variable. In all cases, it is easy to see that the relationships between the probability of maltreatment recurrence and the selected risk factor (i.e., the child's age at the first maltreatment report, the number of income maintenance spells received, and the exposure to community-level poverty) by FCS receipt are curvilinear. Classification of children into “likely” and “unlikely” to be re-reported groups as a function of the values for selected risk factors and for values of FCS receipt can be evaluated by examining the curvilinear regression lines between groups and within groups. Between groups, one evaluates the degree to which the range of estimated probabilities of recurrent maltreatment differ for values of X (i.e., the selected risk factor) given values for FCS receipt. Within groups, one evaluates the degree to which the range of estimated probabilities of recurrent maltreatment differ for children who were actually re-reported versus children who were not actually re-reported (for each category of FCS receipt). For example, between groups, the curvilinear regression lines for the predicted probabilities of recurrent maltreatment by child age in Figure 4.45 are not only discernibly different for values of FCS receipt (i.e., evidence of an interaction effect), but they also estimate a different range of probabilities of recurrent maltreatment given values of child age. Specifically, children who received FCS services have on average a lower estimated likelihood of recurrent maltreatment across values of child age in comparison with children who did not receive FCS services. Within groups, the curvilinear regression lines for the predicted probabilities of recurrent maltreatment by child age also estimate a different range of probabilities of recurrent maltreatment given values of child age (and within categories of FCS receipt). Differences in the predicted probabilities of recurrent maltreatment by observed values for maltreatment occurrence can be assessed by comparing the solid curvilinear regression line (for children who were actually re-reported for recurrent maltreatment) with the dotted curvilinear regression line (for children who were not actually re-reported for maltreatment). Specifically, children who were actually re-reported for maltreatment have on average a higher estimated likelihood of recurrent maltreatment across values of child age in comparison with children who were not actually re-reported for maltreatment (for both categories of FCS receipt).

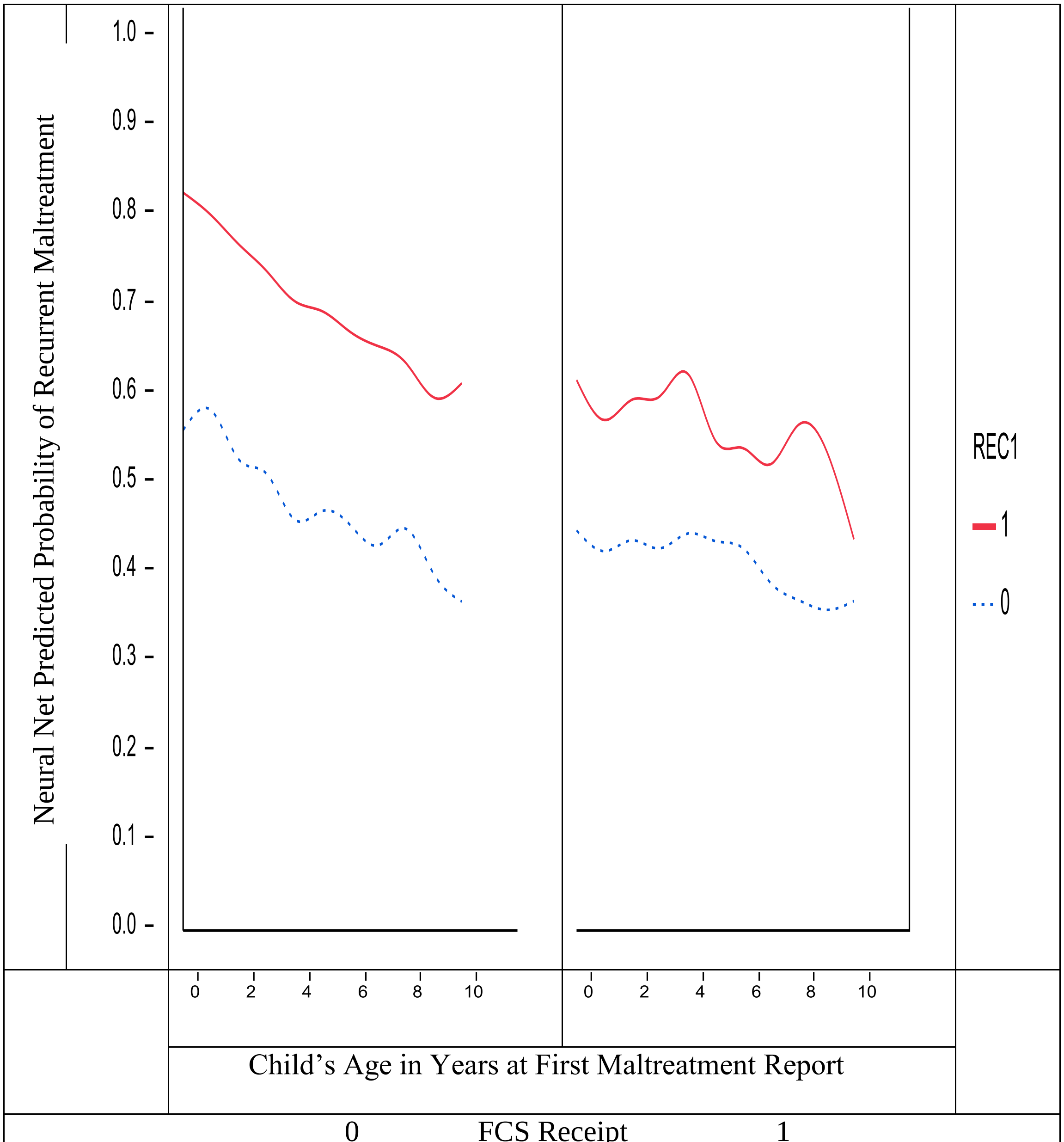


Figure 4.45. Neural network predicted probability of recurrent maltreatment vs. child age at first report by family centered service receipt and observed values for recurrent maltreatment.

Although the overall ability to separate children who received FCS into a lower range of estimated probabilities of recurrent maltreatment holds for the number of income maintenance spells received and exposure to community poverty, there appears to be a threshold effect. Specifically, reductions in the estimated probabilities of recurrent maltreatment as conditioned on the receipt of FCS also appear to be conditioned on values for the number of income maintenance spells received and community-level poverty. In looking at Figure 4.46, there is a discernible decrease the in the range of estimated probabilities of recurrent maltreatment for children who received FCS as compared with children who did not receive FCS up until the number of income maintenance spells received reaches three to four spells. At three to four income maintenance spells received, the estimated probabilities of maltreatment increase markedly and the distance separating the curvilinear regression lines for children with and without FCS receipt shrinks. Hence, differences in the estimated probabilities of recurrent maltreatment between groups (i.e., children who received or did not receive FCS) appear to be dependent upon the number of income maintenance spells received. In terms of within group differences, the curvilinear linear regression lines collide for children who received FCS and who were (a) actually re-reported for recurrent maltreatment or (b) not actually re-reported for maltreatment. Hence, the estimated probabilities of recurrent maltreatment were the same for children who received FCS and who were (a) actually re-reported for maltreatment or (b) not actually re-reported for maltreatment when the number of income maintenance spells received reached six to seven.

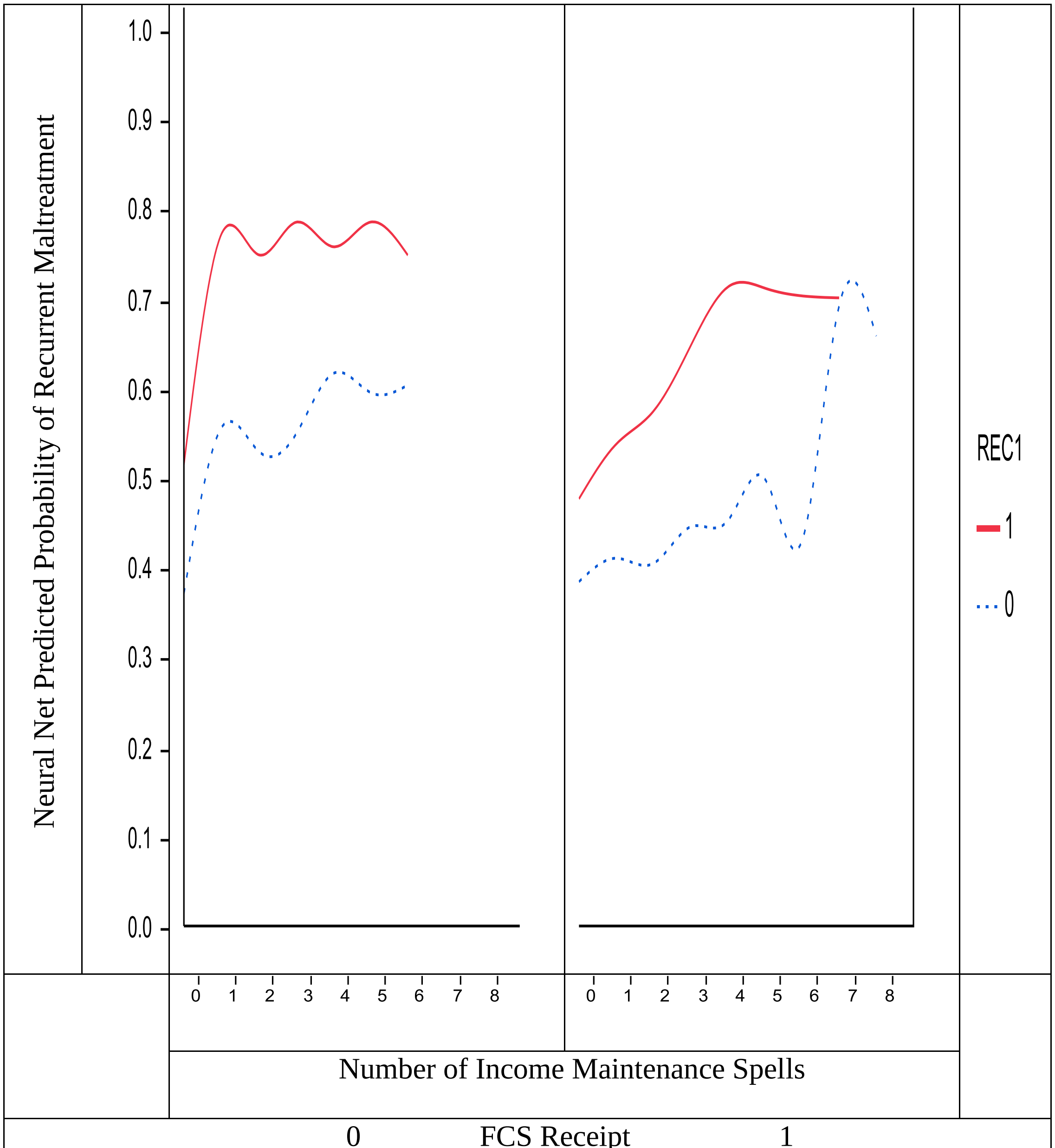


Figure 4.46. Neural network predicted probability of recurrent maltreatment vs. number of income maintenance spells received by family centered service receipt and observed values for recurrent maltreatment.

While less pronounced, threshold effects can be seen in Figure 4.47 where the probabilities of recurrent maltreatment are plotted by community poverty and values for FCS receipt and the recurrent maltreatment outcome. Discernible differences in the

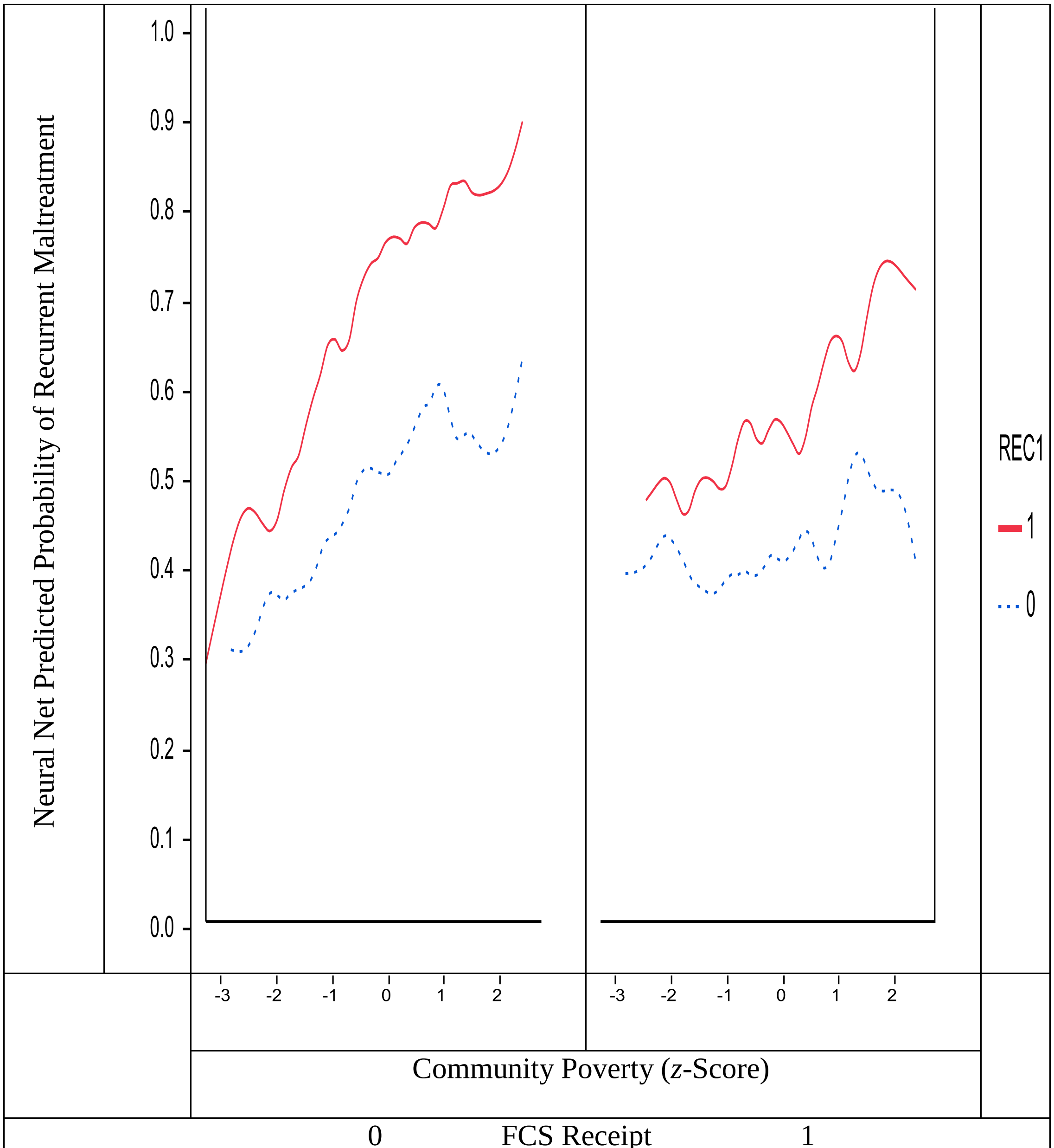


Figure 4.47. Neural network predicted probability of recurrent maltreatment vs. community poverty by family centered service receipt and observed values for recurrent maltreatment.

range of probabilities of recurrent maltreatment estimated for increasing values of community-poverty by FCS receipt as well as observed values for recurrent maltreatment wax and wane depending upon the values for community poverty. The distance between a pair of curvilinear regression lines (e.g., comparing the pairs of regression lines within each category of FCS receipt) appears to be greatest when community-poverty is set at mid-range values as opposed to extremely low or extremely high values.

Chapter 5: Discussion

Summary of Neural Network and Post-Hoc Analytical Findings: Implications for Differential Response Practice

Neural network modeling was selected as the analytic method for this dissertation study in an effort to improve the predictive accuracy of risk assessment for recurrent maltreatment. Moreover, measures of predictive accuracy were compared for the neural network model with a standard linear model in order to determine if the added complexity of the neural network was warranted. The neural network model proved to be more accurate as measured by an increased area under the ROC curve that was larger by .02 units. As noted by Beck, King, and Zeng (2004), it is difficult to numerically categorize the magnitude of a reported difference in the area under the ROC curve and to subsequently describe in standard units what degree of difference constitutes a small, moderate, or large improvement in predictive accuracy. Objectively speaking, the neural network outperformed the linear model in terms of predictive accuracy. Moreover, based upon a thorough examination of the extant literature to date, the area under the ROC curve reported for the neural network model in this dissertation study represents the highest level of predictive accuracy reported for any risk assessment model. Future studies containing risk assessment models should include multiple measures of predictive accuracy such as the area under the ROC curve, a misclassification rate, a confusion matrix, sensitivity, and specificity to facilitate a more thorough and transparent comparison of the predictive utility of various approaches to risk assessment.

Beyond predictive accuracy, one of the most compelling findings in this dissertation study is evidence of nonlinearity, both in terms of curvilinear relationships and interaction effects. Evidence of nonlinearity is critical for the re-evaluation of the findings across the 19 key studies of recurrence because none of these studies specified the risk of recurrent maltreatment as being associated with curvilinear terms and/or interactions where individual poverty, community poverty, child age, and/or parent status as the perpetrator were moderated by service system involvement. Improperly specified regression models produce biased parameter estimates and biased standard errors (Cohen, Cohen, West, & Aiken, 2003). The implications of this study's findings are far reaching and include evidence that calls into question the standard practice of assuming that the relationship between the risk of recurrent maltreatment and a rich selection of predictors is linear. Rather than expend more effort on finding a wider array of predictors to include in a linear model, future studies should focus on the functional form of the relationship between recurrent maltreatment and a more limited number of properly specified predictors that effectively differentiate high-risk cases from low-risk cases.

Post-hoc analyses revealed that the findings from a neural network analysis can be meaningfully deconstructed. Probability plots can be enormously helpful in visualizing the relationships between the likelihood of recurrence and selected predictors to include understanding the ways in which changes in the likelihood of recurrent maltreatment are dependent on the values of selected predictors as well as the interactions between particular predictors. A regression tree can be used to zero-in on the predictors that explain the largest proportion of variance in the likelihood of recurrence, and findings from a regression tree analysis can be used to create a set of empirically-supported risk groups. These risk groups can be used to differentiate children who are at a relatively higher risk of re-report from children who are at a relatively lower risk of re-report as a function of the values for a select number of risk factors. These risk groups are meaningful because they place children in context by describing the average probability of recurrence for each group of children who share the same values for a limited number of predictors. This approach to risk assessment aligns with suggestions that approaches to risk assessment may be more clinically meaningful if workers understood the likelihood of future maltreatment as being higher or lower on average for different groups of children who are defined as relatively similar to some and different from others in relationship to characteristics that are familiar to child welfare workers (Baird & Wagner, 2000; Shlonsky & Wagner, 2005).

In terms of daily practice, risk groups serve as an empirically-based context in which differential response workers can engage families around the specific ways that the risk factors operate. The limited number of risk factors that collectively determine the average likelihood of recurrent maltreatment can inform the context in which the specific mechanisms for change are identified. For example, a family may be classified very quickly as falling within a particular risk group on the basis of their history of income maintenance receipt, the age of the child at the first maltreatment report, and the primary caretaker's identification as the perpetrator of the first maltreatment event. These characteristics function as a starting point by providing the worker with an average likelihood of recurrent maltreatment and a circumscribed set of factors that (a) may be modifiable, or (b) may lead to a set of modifiable characteristics. Poverty is modifiable while child age and the parent's status as the perpetrator for the index event are not; however, an assessment of the family should include a targeted discussion of the ways in which raising a child of a particular age in an environment characterized by a specific level of material deprivation influences the caretaker's actions. Zeroing in on the influential risk factors can help to uncover the behavioral pathways that are modifiable and amenable to intervention. Concentrating assessment and engagement efforts around a set of risk factors that are empirically linked to variation in the likelihood of recurrent maltreatment places a much-needed ground floor in the treatment planning process. By inserting a ground floor in the development of service plans, workers would be required to train what is typically a limited set of resources on the assessment of risk factors that are most likely to lead to a substantial reduction in the risk of recurrence.

The overall contribution of this neural network analysis and subsequent post-hoc analyses is the potential -- aided by further and repeated testing in the future -- to locate the combination of risk factors that give the best possible “initial read” on a family’s likelihood of recurrent maltreatment. Predictive accuracy can be increased in two ways: (a) by estimating a mathematical process that correctly classifies children into risk groups as a function of the provided predictors, and (b) by estimating the true target function that relates recurrent maltreatment to its predictors without assuming monotonically increasing or decreasing effects. Investing in the science that identifies the best combination of predictors for determining that initial read means cutting down on the guess work that occurs when workers have to evaluate the potential effects of an innumerable number of risk factors. An accurate initial read reduces the likelihood of recurrence by more effectively training limited resources on those risk factors that matter most. Differential response workers can spend more time engaging families around a discussion that is designed to identify the mechanisms linking a more general set of risk factors to a more personalized set of interventions that promote child safety and overall family functioning.

Overview: Comparing the Neural Network Findings to Inconsistent Findings in the Extant Literature

As noted in earlier chapters, there is a lack of consistency among the findings produced by the 19 key studies summarized on pages 57 - 71. These studies represent the state of the art in using administrative data to identify the factors that are most influential in increasing or decreasing the risk of recurrent maltreatment. This lack of consistency in findings is particularly noteworthy in relationship to the four prominent risk factors identified by this dissertation study, as well as in relationship to the four service-based variables that appear to moderate the four prominent risk factors.

In comparing the findings from this dissertation study to findings generated by the collection of 19 key studies, several important points need to be addressed. First, none of the 19 key studies has provided a clear theoretical framework that explains how the various risk factors relate to each other and how proposed relationships among risk factors can be used to tailor the delivery of preventive services given a family's unique constellation of dynamic risk factors. Despite the important calls to continue to address the need to engage in cross-sector service system collaboration and coordination (see, e.g., Bai, Wells, & Hillemeier, 2009; Green, Rockhill, & Burns, 2008; Romanelli et al., 2009; Jonson-Reid, 2011; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Kolko, Herschell, Costello, & Kolko, 2009; Smith & Mogro-Wilson, 2008), none of the 19 studies has specifically described how to integrate risk assessment findings and/or scores with service-planning activities for the purpose of matching the treatment needs of children and their primary caregivers with a set of generally and specifically responsive services. Hence, this dissertation study takes a first step in addressing this gap by using a unique constellation of statistical techniques to identify four prominent risk factors and the four system-based responses that can potentially be used to lower a family's risk of returning to the child welfare system for ongoing reports of maltreatment.

Second, in line with the first point, one of the difficulties in figuring out how best to apply risk assessment findings to improvements in service delivery practice is the lack of a clear conceptualization of what a service actually represents. Traditionally, risk assessment was designed to predict the likelihood of recurrence in lieu of any intervention (Cash, 2001; Doueck, English, DePanfilis, & Moote, 1993; Fuller, Wells, & Cotton, 2001); hence, including service delivery in a risk assessment study alters the family's original risk level. That said, if one is interested in understanding how service system responses can be altered to improve child and family outcomes such as reducing the proportion of families that are re-reported for maltreatment, then it makes perfect sense to include measures of cross-system service delivery in risk assessment studies (Camasso & Jagannathan, 2000; Wald & Woolverton, 1990). Ultimately, including measures of service delivery in risk assessment studies follows an edict made by Fluke (2008) in his seminal commentary on the state of preventive service delivery in child welfare systems: "The reduction of reentry is most likely to be achieved by attending to how a CPS agency intervenes with children and families" (p. 750). Given the importance of including measures of service delivery as dynamic factors that can be modified to decrease a family's risk of re-report, it is equally as important to define what the delivery of a service actually represents. Specifically in the extant literature to date, it is unclear if service delivery is being used as a proxy for a condition that serves as a risk factor (e.g., using income maintenance receipt as a proxy for poverty, which is associated with an increased risk of recurrent maltreatment) or if service delivery is being used to represent a system response. Moreover, if service delivery is being used to represent a system response, it is unclear if researchers conceptualize the system response as (a) representing the system's assessment or "read" on the relative risk of recurrence imposed by each family, (b) representing the system's attempt to reduce the family's relative level of risk, and/or (c) representing a set of activities that are theoretically capable of altering any of the child or primary caregiver risk factors. This lack of clarity makes it very difficult to move the field of risk assessment from what is a rather large collection of individual findings to a conceptual plan of how to use specific cross-sector services to reduce the likelihood of recurrent maltreatment by modifying the most powerful and proximal risk factors.

Improving the delivery of preventive services in relationship to specific attempts to modify key proximal risk factors will require that child welfare workers engage families in structured conversations around the ways in which each family personally experiences the prominent risk factors. For example, in addition to collecting information about the child's age at the initial maltreatment report and the family's exposure to poverty as measured by the number of income maintenance spells received to date, the worker will need to engage the family around questions that target how having a child at a particular age within a household that is constrained by a specific level of material deprivation challenges the primary caregiver's coping skills and the child's growth and development. In sum, if risk assessment is going to be helpful in matching a family's set of dynamic risk factors to a responsive service plan, then greater attention must be paid to (a) clearly delineating the specific constructs that are represented by each risk factor included in risk assessment studies, and (b) testing hypothesized relationships among risk factors with a specific emphasis on identifying key moderating effects. Finally, additional attention needs to be paid to identifying data collection techniques that will facilitate a greater understanding of the ways in which families experience the effects of primary risk factors to include (a) underlying mechanisms that explain how a key risk factor increases the likelihood of maltreatment, and (b) cross-factor interactions that occur when risk factors potentially moderate each other.

Third, risk assessment studies typically do not report measures of their models' predictive accuracy (for rare exceptions, please see Jonson-Reid, Drake, & Kohl, 2009; Marshall & English, 1999). Generally speaking, model fit is represented through the presentation of a statistic (e.g., a Wald chi-square value) and a *p*-value that indicates the model had at least one statistically significant effect, thus allowing the rejection of the global null hypothesis. Results are typically confined to the reporting of partial regression coefficients when, as stated previously, most studies assume that the likelihood of recurrent maltreatment is a linear function of all predictors included in the model. None of the risk assessment studies has included higher-order polynomial terms and few have included interaction terms. When interaction terms have been included, they have been limited to interactions between maltreatment type by substantiation status (Connell, Bergeron, Katz, Saunders, & Tebes, 2007), post-investigation services by substantiation status (Connell, Bergeron, Katz, Saunders, & Tebes, 2007), substantiation by type of child welfare service received (Drake, Jonson-Reid, & Sapokaite, 2006), victim disposition (i.e., report was substantiated or indicated) by post-investigation services (Fluke, Shusterman, Hollinshead, & Yuan, 2008), maltreatment type by child age (Jonson-Reid, 2002), child race/ethnicity by child age (Jonson-Reid, 2002), child race/ethnicity by maltreatment type (Jonson-Reid, 2002), alternative response by foster care (Ortiz, Shusterman, & Fluke, 2008), and victim disposition by foster care (Ortiz, Shusterman, & Fluke, 2008). Only one study tested for an interaction between two dynamic risk factors: family stress by social support deficits (where social support deficits were measured in relationship to relationship to informal helping systems)

(DePanfilis & Zuravin, 1999b). In short, none of the interaction terms included in various risk assessment models have tested for the moderation of dynamic child- or primary- caregiver risk factors by service system interventions.

Without measures of predictive accuracy such as confusion matrices or the area under ROC curves, it is impossible to compare the relative predictive validity of various risk assessment models. Hence, comparisons of various risk assessment models are limited to an evaluation of the reported partial regression coefficients. As noted previously, there is a striking lack of consistency among the findings produced by the 19 key studies that largely define the state of the art in risk assessment for recurrent maltreatment. This lack of consistency may be indicative of biased parameter estimates, biased standard errors, and biased significance tests due to incorrect model specification where the risk of recurrent maltreatment has been assumed to be a linear function of predictors such as child age at the first maltreatment report as well as exposure to poverty (Cohen, Cohen, West, & Aiken, 2003). Studies that model the risk of recurrent maltreatment as a linear function of child age, for example, assume that the risk of recurrent maltreatment is a monotonically decreasing function of child age, where a one-unit increase in the child's age is associated with a constant magnitude of decrease in the likelihood of recurrent maltreatment. In other words, it is assumed that the decrease in the likelihood of recurrent maltreatment is constant across all values of the child's age at the first maltreatment report (Cohen, Cohen, West, & Aiken, 2003). However, this dissertation study provided evidence of curvilinearity in child age (among other prominent risk factors), which means that an accurate specification of the relationship between the risk of recurrent maltreatment and the child's age at the first maltreatment report should assume that an increase or decrease in the likelihood of recurrent maltreatment changes as the value of the child's age changes. Moreover, this dissertation study provided evidence of conditional or interaction effects where changes in the likelihood of maltreatment in association with child's age are not only dependent on the value of the child's age, but the value of service-based moderating variables as well. Hence, the effects of individual prominent risk factors such as child age are not purely additive (Cohen, Cohen, West, & Aiken, 2003).

In the sections that follow, inconsistencies in reported findings for the four prominent risk factors and four service-based moderators are discussed. It should be noted that in some cases, the number of studies containing certain predictors such as child mental health services received after a first maltreatment report is quite limited. The data used for this dissertation study are unusual in that administrative records were merged for children and primary caregivers from among a uniquely diverse array of cross-sector service delivery systems. Additionally, this dissertation study was unique in that it included a full complement of risk and protective factors at the child, primary caregiver, family, maltreatment incident, cross-sector service, and community levels that are not typically included in the same model (i.e., the configuration of risk and protective factors included in the neural network models represents to the best degree possible the inclusion of all potentially relevant predictors). Hence, cross-study comparisons can be limited by the degree to which other studies included the same predictors found in this study's neural network model. Nonetheless, as the evidence of discrepancies among findings in the extant literature mount, there is certainly a basis for questioning the degree to which the likelihood of recurrent maltreatment is actually a linear function of its predictors.

Taking Stock of the Extant Literature: Examining the Inconsistencies

Child age at the time of the first maltreatment report.

Child age at the first maltreatment report is generally found to have a significant and negative association with the likelihood of recurrent maltreatment, where younger children are more vulnerable to the risk of recurrent maltreatment as compared to older children (see e.g., Bae, Solomon, & Gelles, 2007, 2009; English, Marshall, Brummel, & Orme, 1999; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fryer & Miyoshi, 1994). That said, findings across key studies do not demonstrate support for a monotonically decreasing relationship between child age at the first maltreatment report and the likelihood of repeat maltreatment where the decrease in the likelihood of a re-report is constant across all values for child age. Among the studies that measured child age at the first maltreatment report as an ordinal-level variable (see e.g., Connell, Bergeron, Katz, Saunders, & Tebes, 2007; English, Marshall, Brummel, & Orme, 1999; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fluke, Yuan, & Edwards, 1999; Fryer & Miyoshi, 1994; Fuller, Wells, & Cotton, 2001; Jonson-Reid, 2002; Lipien & Forthofer, 2004; Ortiz, Shuterman, & Fluke, 2008) multiple studies reported findings where the assumed increase or decrease in the likelihood of repeat maltreatment for a particular age group in relationship to a reference category was not statistically significant. For example, Fuller, Wells, and Cotton (2001) looked for the potential presence of differences in the likelihood of recurrent maltreatment for children who were (a) 0 to 2 years of age or (b) 3 to 5 years of age in relationship to children who were 6 to 18 years of age. For children who were 3 to 5 years of age, there was no statistically significant difference in the likelihood of recurrent maltreatment as compared with children who were 6 to 18 years of age. Similarly, in a study conducted by Connell, Bergeron, Katz, Saunders, and Tebes (2007) there was no statistically significant difference in the likelihood of recurrent maltreatment for children who were 1 to 5 years of age in comparison with children who were under 1 year of age. Additionally, children who were 1 year of age did not experience a significantly different likelihood of recurrent maltreatment as compared with children who were under 1 year of age (Fluke, Shusterman, Hollinshead, & Yuan, 2008), and children who were 7 to 10 years of age did not experience a significantly different likelihood of recurrent maltreatment in comparison with children who were 1 to 6 years of age (Jonson-Reid, 2002).

In an analysis of case-level data provided by 10 states within the National Child Abuse and Neglect Data System, Fluke, Yuan and Edwards (1999) compared the likelihood of recurrent maltreatment for children divided into the following four age groups: (a) 0 to 2 years of age, (b) 3 to 5 years of age, (c) 6 to 11 years of age, and (d) 12 to 17 years of age. Rather than evidence of a significantly decreasing likelihood of repeat maltreatment as the child's age increased across age-based groups, the authors noted the following:

> From state to state, some categories of age had statistically different rates of recurrence in relation to each other. The analysis of age categories was not successful in confirming the findings of Fryer and Miyoshi (1994) from Colorado that very young children were more likely to recur, with the one exception of Vermont. However for 9 of 10 states, the oldest age group, the 12 to 17 category, had a lower rate of recurrence compared to the other age categories. (p. 641)

An examination of the survival function with respect to child age provided for the state of Louisiana for data from years 1994 through 1995 provided evidence that the survival curves for children in the 0 to 2 year category and the 3 to 5 year category were merged (i.e. the curves were lying on top of each other), and the distance between the survival curves for children in these two age-based groups and the survival curve for children in the 6 to 11 year category was not substantial.

Among the studies that measured child age at the first maltreatment report as a continuous variable (see e.g., Bae, Solomon, & Gelles, 2007, 2009; Drake, Jonson-Reid, & Sapokaite, 2006; Marshall & English, 1999; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010), three studies reported a change in the direction of the partial regression coefficient that measured the rate of change in the likelihood of repeat maltreatment in relationship to an increase in child age. Specifically, Drake, Jonson-Reid, and Sapokaite (2006) ran two models that assessed the likelihood of a second maltreatment report in relationship to a wide array of predictors to include the child's age at the first maltreatment report. The first model tracked children who were 0 to 11 years of age at the time of the first report and the second model tracked children who were 4 to 11 years of age at the time of the first report. The partial regression coefficient for child age from the first model provided evidence of a statistically significant negative association between child age and the likelihood of a maltreatment re-report (HR = 0.97). However, the partial regression coefficient for child age from the second model provided evidence of a statistically significant positive association between child age and the likelihood of a maltreatment re-report (HR = 1.05). A similar switch in the direction of the slope for child age was detected in findings from a follow-up study that used the same data to identify and compare the predictors of a first, second, third, and fourth re-report for maltreatment (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). When predicting the likelihood of a first re-report (i.e., a second report for maltreatment), child age at the time of the first report was significantly and negatively associated with the likelihood of repeat maltreatment (HR = 0.97). However, child age was then shown as being significantly and positively associated with the likelihood of a second re-report (HR = 1.02), a third re-report (HR = 1.03), and a fourth re-report (HR = 1.03).

Finally, English, Marshall, Brummel, and Orme (1999) reported a decrease in the proportion of children with a second maltreatment referral for three ordinal-level age groups, where 31.08% of children who were 0 to 5 years of age at the first report were re-referred for maltreatment, 28.77% of children who were 6 to 11 years of age at the first report were re-referred for maltreatment, and 21.85% of children who were 12 to 17 years of age at the first report were re-referred for maltreatment. However, in a follow-up study using the same data, Marshall and English (1999) reported a significant and positive association between child age at the first report when measured as a continuous variable and the likelihood of being re-referred for maltreatment (HR = 1.06).

Exposure to poverty at the individual/household and community levels.

In general, the receipt of income maintenance (IM) vis-à-vis AFDC/TANF spells can be used as a proxy for individual (i.e., child exposure) or household-level poverty. As conceptualized by Needell, Cuccaro-Alamin, Brookhart, and Lee (1999), exposure to poverty may increase the probability of child maltreatment by influencing the ways in which primary caregivers interact with their children. The mechanism that links exposure to poverty with parenting behaviors appears to be variation in the ways in which parents psychologically experience the stress of acute and chronic material deprivation. Moreover, exposure to the stressful constraints imposed by poverty may decrease a primary caregiver's effectiveness in negotiating the hardships of material deprivation as well as the inevitable incidence of new challenges related to parenting. Broadening the conceptualization of stress as the mechanism that links poverty to changes in parenting behaviors that can lead to child maltreatment, Coulton, Crampton, Irwin, Spilsbury, and

Korbin (2007) place this "parenting behavioral pathway" into a neighborhood context. Specifically, variation in child maltreatment that can be attributed to differences between neighborhoods is linked to differences in the availability and quality of social resources as well as differences in the mediating effects of social processes such as social cohesion (i.e., the degree of mutual trust) and informal social control (i.e., the degree to which community members experience a shared expectation of action around a specific common goal). In short, variation in child maltreatment between neighborhoods can be explained by the ways in which parents in neighborhoods with different levels of social resources as well as social cohesion and informal social control (i.e., collective efficacy) experience the constraints of poverty.

In a study that examined the degree to which exposure to poverty might increase the likelihood of ongoing child welfare involvement for children who had at least one child maltreatment report, Jonson-Reid, Drake, and Kohl (2009) compared two groups of child welfare-involved children: (1) children in the CAN/AFDC group who had also received at least one IM spell, and (2) children in the CAN Only group who had no known IM receipt. In comparison with children in the CAN Only group, children in the CAN/AFDC group had a significantly higher proportion of primary caregivers with (a) a disability (CAN/AFDC = 6.5% and CAN Only = 1.1%), (b) a history of mental health treatment (CAN/AFDC = 11.4% and CAN Only = 1.6%), and (c) a substance abuse problem (CAN/AFDC = 13.3% and CAN Only = 2.6%). Additionally, 63.8% of the children in the CAN/AFDC group had a recurrent report while 33.3% of the children in the CAN Only group experienced a recurrent report.

Among children with a first report for neglect, having a parent with a history of mental health treatment (in comparison to children whose parents did not have a history of mental health treatment) increased the likelihood of a re-report by 71% and having a parent with a substance abuse problem increased the likelihood of a re-report by 69%. That said, children without a history of poverty were 51% less likely to experience a recurrent report for maltreatment. Among children with a first report for physical abuse or a first report for sexual abuse, having a parent with a history of mental health treatment increased the likelihood of a re-report by 105%, while children without a history of poverty were 60% less likely to experience a recurrent report for maltreatment (Jonson-Reid, Drake, & Kohl, 2009). All told, the findings from this study appear to provide some support for the hypothesized increase in the risk of maltreatment in relationship to exposure to poverty as well as parenting characteristics that are associated with variation in parenting behavior and the stress that accompanies financial duress.

Among the small but important number of studies that have been able to assess the risk of recurrent maltreatment in relationship to IM receipt (among a number of other risk factors) by merging administrative records from child welfare and IM systems (see e.g., Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Drake, & Kohl, 2009; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Needell, Cuccaro-Alamin, Brookhart, & Lee, 1999), the findings are mixed as to how the timing and duration of poverty are related to the likelihood of recurrent maltreatment. One of the difficulties encountered in comparing findings was the lack of continuous measures used to capture a child's exposure to household-level poverty. Rather than tracking the total number of IM spells received within a specific timeframe, studies typically measured exposure to poverty with dichotomous measures. The only study that employed official child welfare and IM administrative records as well as a continuous measure of IM receipt was conducted by Needell, Cuccaro-Alamin, Brookhart and Lee (1999).

Needell et al. (1999) identified 63,768 children in 10 California counties who received a first-known IM spell between 1990 and 1995 and followed the children by tracking any subsequent reports of maltreatment between the date of their first known IM spell and 1995. Overall, 16.65% of the children were subsequently reported for maltreatment; moreover, the overall proportion of reported children increased to 27% when the follow-up period was extended to five years for the cohort of interest, which first received IM support in 1990. The child's age at the time of the initial AFDC receipt was significantly and negatively associated with the likelihood of being reported for maltreatment, but only for children who were 3 to 5 years of age in comparison to children who were less than 1 year of age. Although the number of months of AFDC receipt was significantly and positively associated with the likelihood of being reported for maltreatment (OR = 1.02), the number of breaks in AFDC receipt was also significantly and positively associated with the likelihood of receiving a maltreatment report (OR = 1.21). Hence, the effect of poverty as measured by AFDC receipt does not appear to have been monotonically increasing if measures of receipt and breaks in receipt could both produce an increase in the odds of a maltreatment report occurring.

In theory, if social scientists use AFDC receipt to measure exposure to poverty (as opposed to the receipt of a service that is meant to provide financial support), then the remission of or breaks in AFDC receipt should measure a reduced exposure to or remission in exposure to poverty (as opposed to the cessation of a service that is meant to provide financial support). Hence, if the relationship between poverty and the probability of being reported for maltreatment is truly a monotonically increasing association, then the remission of poverty as measured by a break in AFDC receipt should be negatively associated with the probability of being reported for maltreatment. Additional discrepancies in findings related to the relationship between AFDC receipt and the likelihood of repeat maltreatment are described below.

Although not measured on a continuous scale, several other studies have included AFDC receipt as a measure of exposure to poverty. For example, Drake, Jonson-Reid, and Sapokaite (2006) measured exposure to poverty through IM receipt by assessing the extent to which a permanent cessation of AFDC benefits influenced the likelihood of receiving a second maltreatment report. The cessation of IM support was measured as a dichotomous variable that could have occurred before the first maltreatment report or after the first maltreatment report. Cox regression models were run separately for children who were 0 to 11 years of age at the time of the first maltreatment report (Model #1) and 4 to 11 years of age at the time of the first maltreatment report (Model #2). For both models, families experiencing a permanent cessation of IM benefits were significantly less likely to be re-reported for maltreatment in comparison with families that did not experience a permanent cessation of IM benefits. Additionally, for both models, families that experienced a permanent cessation of IM benefits after the first maltreatment report experienced a more substantial decrease in the likelihood of being re-reported for maltreatment. Specifically, families in Model #1 that experienced a permanent cessation of IM benefits *before* the first maltreatment report were 12.2% less likely to be re-reported for maltreatment (in comparison to families that did not experience a permanent cessation of IM benefits before the first maltreatment report); families that experienced a permanent cessation of IM benefits *after* the first maltreatment report were 31.6% less likely to be re-reported for maltreatment (in comparison to families that did not experience a permanent cessation of IM benefits after the first maltreatment report). Families in Model #2 that experienced a permanent cessation of IM benefits *before* the first maltreatment report were 30.6% less likely to be re-reported for maltreatment (in comparison to families that did not experience a permanent cessation of IM benefits before the first maltreatment report); families that experienced a permanent cessation of IM benefits *after* the first maltreatment report were 44% less likely to be re-reported for maltreatment (in comparison to families that did not experience a permanent cessation of IM benefits after the first maltreatment report).

In a follow-up study using the same data, Jonson-Reid, Emery, Drake, and Stahlschmidt (2010) measured exposure to poverty through IM receipt by dichotomously identifying which families received AFDC benefits before the first maltreatment report. Families that received IM support before the first maltreatment report were 23% less likely to be re-reported for maltreatment (in comparison with families that did not receive IM support before the first maltreatment report). This relationship did not hold, however, when looking at the relationship between IM support before the first maltreatment report and the likelihood of being reported to child welfare for a third report, a fourth report, and a fifth report. In models predicting the likelihood receiving a third, fourth, and fifth report, IM receipt before the first report was not significantly associated with the likelihood of recurrent maltreatment.

Exposure to poverty was also measured through IM receipt by dichotomously identifying which families received AFDC benefits after the first maltreatment report. Families that received IM support after the first maltreatment report were 12% more likely to be re-reported for maltreatment (in comparison with families that did not receive IM support after the first maltreatment report). The direction of this relationship switched in models assessing the likelihood of being reported to child welfare for a third report and a fourth report. In each case, AFDC support was measured as whether or not the family received IM benefits after the maltreatment report that directly preceded the outcome of interest. For example, for the model assessing the likelihood of being reported to child welfare for a third time in relationship to AFDC benefits and all other predictors, AFDC receipt was measured as having occurred if the benefits were received after the second report but before the third report (if a third report occurred). Families that received AFDC benefits after the second report were 14% less likely to be reported to child welfare for a third time, and families that received AFDC benefits after the third report were 13% less likely to be reported to child welfare for a fourth time. There were no significant differences in the likelihood of being reported to child welfare for a fifth time for families that received AFDC benefits after the fourth maltreatment report.

A third study that used data from the National Child Abuse and Neglect Data System included a dichotomous measure of poverty based upon caseworker responses to indicators of financial hardship contained within the NCANDS public assistance and financial problems section (Connell, Bergeron, Katz, Saunders, & Tebes, 2007). A family was coded in the study as having experienced poverty if the NCANDS public assistance item and/or financial problems item provided information documenting (a) the family's receipt of public benefits such as AFDC support or Medicaid, and/or (b) the child's removal from the home in cases where the primary caregiver was unable to provide sufficient financial resources so as to meet a basic standard of care for the child. In comparison with families that were not living in poverty, families that were living in poverty were 226% more likely to be re-reported for maltreatment. A final study that used child welfare records to assess the likelihood of recurrent maltreatment in relationship to a variety of child, primary caregiver, and household characteristics used a five-item scale to measure survival stress as a combination of a lack of resources for basic needs, a lack of shelter, housing in poor repair, over-crowded housing, and lack of or poor use of health care (DePanfilis & Zuravin, 1999b). Survival stress was not significantly associated with the likelihood of a re-report for maltreatment.

Similar to household-level measures of poverty, findings related to community or neighborhood measures of poverty have also proven to be inconsistent when used as a predictor in relationship to the likelihood of recurrent maltreatment. Very few studies have been able to merge administrative child welfare records with community-level measures of poverty such as U.S. Census variables that capture dimensions of poverty within census tracts (see e.g., Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Drake, & Kohl, 2009; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Way, Chung, Jonson-Reid, & Drake, 2001). Median household income as reported in 1990 U.S. Census data has been used as a proxy for community-level poverty where a family's neighborhood has been defined as occurring within a census tract. Median household income has been shown to have a statistically significant and negative association with the likelihood of being reported to child welfare for the second time, with decreases in the likelihood of repeat maltreatment that range from 0.5% to 3% for every $1,000 increase in median household income (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Drake, & Kohl, 2009; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010; Way, Chung, Jonson-Reid, & Drake, 2001). That said, median household income has also been reported as having no significant association with the likelihood of a second report of maltreatment depending upon regression model specification criteria such as limitations to the child's age at the first maltreatment report (i.e., restricted to 4 to 11 years of age) (Drake, Jonson-Reid, & Sapokaite, 2006) and type of maltreatment at the first report (i.e., where median household income was not statistically significant for first reports of physical abuse and for sexual abuse) (Jonson-Reid, Drake, & Kohl, 2009). Moreover, median household income has not been reported as having a significant association with the likelihood of repeat maltreatment when the outcome of interest was the third, fourth, and/or fifth maltreatment report (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010).

Additional studies have linked 1990 U.S. Census data to individual-level measures of risk and maltreatment based on survey responses (e.g., responses to the Child Abuse Potential Inventory or the Conflict Tactics Scale) as opposed to official child welfare records (Coulton, Korbin, & Su, 1999; Kim, 2004; Molnar, Buka, Brennan, Holton, & Earls, 2003). These multilevel studies have generally reported a very low proportion of variance in reported measures of maltreatment as being attributable to differences between neighborhoods and have varied in significant results for community-level measures of poverty after taking into account child, primary caregiver, and family risk factors. For example, Coulton, Korbin, and Su (1999) and Molnar, Buka, Brennan, Holton, and Earls (2003) found that approximately 2% of the variation in primary caregiver-reported measures of child maltreatment was attributable to between neighborhood differences, but only Coulton et al. found that neighborhood-level poverty was significantly associated with differences in maltreatment variation between neighborhoods (where impoverishment was a factor score that combined percent of households with children that were female headed, percent of poor persons, percent of residents unemployed, percent of vacant housing units, percent of 1980-1990 population, and percent of residents classified as Black). Beyond differences in the ways in which child maltreatment outcomes and individual-level risk factors are measured, differences in the ways in which neighborhoods are operationalized will affect findings regarding relationships between variation in individual-level measures of maltreatment and neighborhood-level measures of poverty (Aron et al., 2010; Coulton, Crampton, Irwin, Spilsbury, & Korbin, 2007; Coulton, Korbin, and Su; 1999; Freisthler, Merritt, & LaScala, 2006).

The primary caregiver's status as perpetrator of the first maltreatment incident. Of the 19 key studies of recurrent maltreatment, only one study included the primary caregiver's perpetrator status as a risk factor. In fact, Way, Chung, Jonson-Reid, and Drake (2001) focused specifically on the ways in which perpetrator status, perpetrator characteristics (i.e., gender and race/ethnicity), case substantiation status, maltreatment type, and median household income were related to the likelihood of a second maltreatment report. Findings from the study revealed that large proportions of perpetrators both with and without substantiated initial reports for maltreatment were subsequently re-reported for child maltreatment. For example, 42.4% of all perpetrators in the study were re-reported for maltreatment, where 52% of the perpetrators with a substantiated initial report for neglect were re-reported for maltreatment, and 44% of the perpetrators with an unsubstantiated initial report for neglect were re-reported for maltreatment. Re-report rates for perpetrators with substantiated and unsubstantiated initial reports for physical abuse were about the same. The re-report rate for perpetrators with an unsubstantiated case of sexual abuse was actually higher than the re-report rate for perpetrators with a substantiated case of sexual abuse. Specifically, 34% of perpetrators with an unsubstantiated initial report for sexual abuse were re-reported for maltreatment, and 25% of perpetrators with a substantiated initial report for sexual abuse were re-reported for maltreatment (Way, Chung, Jonson-Reid, & Drake, 2001).

As noted by the authors, "perpetrator recidivism is of particular interest to practitioners and researchers. This is because child welfare interventions are generally geared to produce changes in the behavior of the perpetrator/adult caregiver rather than the child" (Way, Chung, Jonson-Reid, & Drake, 2001, p. 1094). Given a perpetrator's ongoing proximity to the child who was victimized when the perpetrator is also the primary caregiver, it stands to reason that including perpetrator status and perpetrator characteristics in risk assessment models makes sense. While substantiation has typically been included as a risk factor in risk assessment models, Way, Chung, Jonson-Reid, and Drake (2001) provide evidence that a significant effect of substantiation status becomes less important in light of the substantial re-report rates for perpetrators of substantiated and unsubstantiated incidents across maltreatment types. Moreover, substantiation status has been found to vary in its effects on the likelihood of a report regardless of maltreatment type. For example, some studies reported a significant increase in the likelihood of a subsequent maltreatment report for a substantiated index event in comparison with an unsubstantiated index event (Drake, Jonson-Reid, & Sapokaite, 2006; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010); conversely, other studies reported a significant decrease in the likelihood of a subsequent maltreatment report for a substantiated index event in comparison with an unsubstantiated index event (Connell, Bergeron, Katz, Saunders, & Tebes, 2007; Marshall & English, 1999; Ortiz, Shusterman, & Fluke, 2008). Finally, in two of the studies where the likelihood of a subsequent report was shown as decreasing when the index event was substantiated, the likelihood of a subsequent maltreatment report was also shown as increasing if the child had a previous history (i.e., before the index event in the study) of substantiated or indicated reports (Connell, Bergeron, Katz, Saunders, & Tebes, 2007; Ortiz, Shusterman, & Fluke, 2008).

Cross-sector service delivery.

Despite the relatively large number of studies that have included child welfare post-investigation services as predictors of recurrent maltreatment (see, e.g., Bae, Solomon, & Gelles, 2007, 2009; dePanfilis & Zuravin, 1999a, 2001, 2002; Drake, Jonson-Reid, & Sapokaite, 2006; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fluke, Yuan, & Edwards, 1999; Lipien & Forthofer, 2004), there is no generally accepted consensus regarding the construct that service delivery represents. Options for the potential effects of post-investigation services include (a) a representation of the family's relative risk of recurrent maltreatment measured in relationship to the receipt or absence of services, or (b) a representation of the family's relative risk of recurrent maltreatment measured in relationship to the intensity of the intervention that was delivered (i.e., where no services represents the lowest level of risk, family centered services represents a moderate level of risk, and family preservation services and foster care placement represent much higher levels of risk) (dePanfilis & Zuravin, 1999a, 2001; Drake, Jonson-Reid, & Sapokaite, 2006; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). In fact, this conceptualization of service delivery can be extended to other cross-sector types of service receipt such as the parental and child receipt of mental health services, child receipt of special education eligibility, and child involvement in the juvenile justice system (Drake, Jonson-Reid, & Sapokaite, 2006). Additional options for the potential effects of post-investigation services include (a) a potential failure to meet a family's needs as evidenced by a lack of service receipt or service receipt that is not adequate for meeting a family's diverse array of needs (Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fluke, Yuan, & Edwards, 1999; Hélie & Bouchard, 2010; Jonson-Reid, 2002), (b) a potential success in decreasing the likelihood of future maltreatment in relationship to presumably specific but unnamed treatment objectives as well as family access and motivation to receive treatment (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010), and/or (c) the system's intrusion into and ultimately surveillance over families currently engaged in service receipt (dePanfilis & Zuravin, 1999a, 2001; Fluke, Shusterman, Hollinshead, & Yuan, 2008; Fluke, Yuan, & Edwards, 1999; Hélie & Bouchard, 2010).

Given the lack of consistency in the effects of service receipt for child welfare-involved families, it is not surprising that several different and at times overlapping perspectives have emerged in order to describe the observed effects of services on the likelihood of repeat maltreatment. That said, greater conceptual clarity is needed in order to advance the accuracy and the clinical applicability of risk assessment findings in daily child welfare practice. One of the most important questions that needs to be addressed is whether or not services should be conceptualized as strictly a measure of the family's relative risk of being re-reported for maltreatment or if services should be conceptualized as system responses that are capable of influencing the likelihood of a re-report by moderating specific risk factors. Service influence on the likelihood of recurrence has only been modeled through interactions with substantiation or victim disposition status. These conceptualizations have not been successful in identifying the potential mechanisms through which service delivery might be influencing the risk of re-report. Moreover, statistically significant interactions with substantiation or victim status have raised more questions than they have answered; and they, thus, have failed to clarify concerns regarding the effects of services in general or the effects of services for specific subpopulations. For example, as noted by Fluke, Shusterman, Hollinshead, & Yuan, 2008), victims who received services had a lower risk of being re-reported as compared with victims who did not receive services. However, non-victims who received services had a greater risk of being re-reported as compared to non-victims who did not receive services. While this may suggest a protective effect for children who were found to be victims in relationship to their first maltreatment report, this finding does not provide any insight as to how the service is effective in reducing the risk of a re-report or why victims would enjoy greater benefits as opposed to non-victims.

In terms of the subpopulations that receive services, several studies have examined the likelihood of receiving post-investigation services in relationship to an array of child, primary caregiver, family, and maltreatment incident characteristics. For example, Jonson-Reid, Drake, and Kohl (2009) used logistic regression models to identify the characteristics that influenced the likelihood of receiving in-home or foster care services within 45 days of a first maltreatment report. The authors used a 45-day timeline until the start of services to identify cases that would be considered higher risk or more serious as compared to cases where services were either not delivered or services began after the 45 day cut-off. Separate regression models were run for children with an initial report for neglect, physical abuse, or sexual abuse. Children with a first report for neglect were significantly more likely to receive services within 45 days of the first maltreatment report if they were children of color (OR = 1.37), if they had a parent with a substance abuse problem (OR = 1.48), and/or if a mandated reporter was responsible for making the initial report (OR = 2.53). Children with a first report for physical abuse were significantly more likely to receive services within 45 days of the first maltreatment report if they were children of color (OR = 1.43), if they had a developmental or learning disability (OR = 1.65), if they had a parent with a substance abuse problem (OR = 2.15), and/or if a mandated reporter was responsible for making the initial report (OR = 3.08). Children with a first report for sexual abuse were significantly more likely to receive services within 45 days of the first maltreatment report if they were older at the time of the first report (OR = 1.09), if they had a developmental or learning disability (OR = 1.69), if a mandated reporter was responsible for making the initial report (OR = 1.85), and/or if the first maltreatment incident was severe enough to cause the child physical harm (OR = 2.49). Conversely, children with a first report for sexual abuse were significantly less likely to receive services within 45 days of the first maltreatment report if they came from a household that had never received income maintenance benefits (OR = 0.51).

In a separate study, DePanfilis and Zuravin (2001) assessed the likelihood of receiving post-investigation services in relationship to the mother's race/ethnicity and age, the presence of a parental substance abuse problem, the number of children in the household, the existence of a prior substantiated report, and the maltreatment type at the initial report. Among families with a substantiated report for maltreatment, none of the predictors listed above were statistically significant in increasing the likelihood of receiving post-investigation services; in fact, two of the predictors listed above were statistically significant in decreasing the likelihood of receiving post-investigation services. Specifically, families with a previous substantiated report were 22% less likely to receive post-investigation services as compared with families that did not have a prior substantiated report. Additionally, families with a substantiated index report for neglect were 20% less likely to receive post-investigation services as compared with families with a substantiated index report for physical abuse. Based on the findings of the two studies described above, it is doubtful that service receipt is a strong proxy for the family's overall risk of being re-reported for maltreatment. It is more likely that service receipt is a better representation of a system response that may take into account select risk factors that might or might not have been included in the regression model. Based on the low to moderate *c* characteristics (i.e., to represent the area under the ROC curve) (*c* = .66 for an initial report for neglect, *c* = .71 for an initial report for physical abuse, *c* = .72 for an initial report for sexual abuse) for models predicting the likelihood of receiving post-investigation services within 45 days of the first report, the relationship between the array of risk factors included as predictors and the service receipt outcome is far from a perfect fit (Jonson-Reid, Drake, and Kohl, 2009). Thus, instead of treating service receipt as a proxy for the combination of risk factors a system should be responding to you, it makes more sense to evaluate the effects of service receipt on the likelihood of recurrent maltreatment in relationship to key risk factors.

In terms of the increases or decreases in the likelihood of recurrence that have been attributed to service receipt, the findings are diverse. Using administrative child welfare data from Maryland, DePanfilis and Zuravin (1999a) found that families that did not receive post-investigation services, in comparison with families that received post-investigation services, were significantly less likely to have a subsequent substantiated report both while the case will still open and during a two-year follow-up period after the case had closed (thus controlling for a potential surveillance effect). Using a subset of the same data, DePanfilis and Zuravin (2001) found again that a statistically significant decrease in the likelihood of a subsequent substantiated report occurred for families that did not receive post-investigation services as compared with families that received post-investigation services. Specifically, just 4% of families that did not receive services experienced a recurrence in five years, while 26% of families that received services experienced recurrence within five years. However, in a third study using the same data, DePanfilis and Zuravin (2002) reported that families documented as having attended the services identified in their treatment plan were 32% less likely to experience recurrent maltreatment.

Using administrative child welfare data from 10 Florida counties, Bae, Solomon, and Gelles (2007) found that the receipt of court-ordered service did not significantly increase or decrease the likelihood of a substantiated or an unsubstantiated re-report. That said, the length of CPS involvement was significantly associated with an increase in the likelihood of both a substantiated re-report (HR=1.01) and an unsubstantiated re-report (HR=1.01). Moreover, upon disaggregating the sample by type of maltreatment for subsequent substantiated reports, a receipt of court-ordered services as opposed to no court-ordered services was significantly and positively associated with the likelihood of a subsequent substantiated report for sexual abuse (HR=1.52), but the receipt of court-ordered services was not significantly associated with the likelihood of a subsequent substantiated report for neglect and/or physical abuse. Conversely, the length of CPS involvement was significantly and positively associated with the likelihood of a subsequent substantiated report for neglect (HR=1.01) and for physical abuse (HR=1.01), but the length of CPS involvement was not significantly associated with the likelihood of a subsequent substantiated report for sexual abuse. Using a subset of the same data (7 counties as opposed to 10 counties), Bau, Solomon, and Gelles (2009) compared the likelihood of having multiple subsequent substantiated reports versus no subsequent substantiated reports for (a) children who were placed in foster care in comparison with children who received court-ordered permanency, and (b) children who received general CPS services in comparison with children who received court-ordered permanency. Both foster care placement (OR=1.36) and the receipt of general CPS services (OR=1.46) were significantly associated with an increase in the likelihood of having multiple subsequent substantiated reports for maltreatment. Additionally, the receipt of general CPS services (OR=1.38) was significantly associated with an increase in the likelihood of having multiple subsequent substantiated reports for maltreatment versus one subsequent substantiated report for maltreatment.

Analyzing data from the National Child Abuse and Neglect Data System (NCANDS), Connell, Bergeron, Katz, Saunders, and Tebes (2007), used a Cox regression model to assess the contributions of a range of child, family, maltreatment incident, and CPS intervention variables in explaining the likelihood of experiencing a re-report for maltreatment. The delivery of post-investigation services was not significantly associated with the likelihood of a re-report. However, other studies using NCANDS data have found that the delivery of post-investigation services is significantly and positively associated with the likelihood of repeat maltreatment. For example, Fluke, Shusterman, Hollinshead, and Yuan (2008) studied the effect of post-investigation services when receipt of services occurred if a family received any number of a type of services to include family preservation or family support services within 90 days of the index event's disposition. No post-investigation services was used as the reference group. The delivery of post-investigation services was significantly associated with an increase in the likelihood of an unsubstantiated re-report (HR=1.35) and a substantiated re-report (HR=1.74). Foster care placement was also significantly associated with an increase in the likelihood of an unsubstantiated re-report (HR=2.19) and a substantiated re-report (HR=4.24). That said, the direction of the effect for foster care placement changed for a separate study using NCANDS data, where placement of the child in foster care was significantly associated with a decrease in the likelihood of an unsubstantiated re-report (HR=0.93) (Ortiz, Shusterman, & Fluke, 2008).

Additional discrepancies in findings can be found by comparing the effects of post-investigation service delivery for studies using child welfare administrative data from Florida (Lipien & Forthofer, 2004), California (Jonson-Reid, 2002), and Missouri (Drake, Jonson-Reid, & Sapokaite, 2006; Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). Using no service delivery as the reference group, Lipien and Forthofer reported a statistically significant and positive association between the delivery of short-term services (OR=1.22) and the likelihood of a subsequent substantiated report for maltreatment; additionally, the authors reported a statistically significant and positive association between the delivery of in-home services (OR=1.70) and the likelihood of a subsequent substantiated report for maltreatment. However, placement of the child in relative foster care was significantly and negatively associated (OR=0.81) with the likelihood of a subsequent substantiated report.

Conversely, studies using administrative child welfare data from California and Missouri provide evidence that the delivery of post-investigation services can be associated with a significant decrease in the likelihood of a re-report. Using open for services as the reference group, Jonson-Reid's (2002) study of the likelihood of a maltreatment re-report for children in 10 California counties provided evidence of a statistically significant increase in the risk of a re-report for children with investigated cases who did not receive post-investigation services (OR=1.22). Using no service delivered as the reference group for each type of service delivery that was tested, Drake, Jonson-Reid, and Sapokaite (2006) studied the relationship between the likelihood of a re-report and the delivery of family centered services (FCS), family preservation or intensive in-home services (FPS), and foster care placement. Two Cox regression models were run where Model #1 included children who were 0-11 years of age at the time of the first maltreatment report, and Model #2 included children who were 4-11 years of age at the time of the first maltreatment report. For both models, receipt of FCS was significantly and negatively associated with the likelihood of a re-report (Model #1 HR=0.72, Model #2 HR= 0.56), FPS or FPS and FCS was significantly and positively associated with the likelihood of a re-report (Model #1 HR=1.44, Model #2 HR=1.85), and placement in foster care was significantly and positively associated with the likelihood of a re-report (Model #1 HR=2.46, Model #2 HR=4.54). In a follow-up study using the same data, the direction of the effects of FPS and FCS and foster care placement switched. Using no service delivered for each type of service delivery that was tested, Jonson-Reid, Emery, Drake, and Stahlschmidt (2010) reported that FCS was significantly associated with a decrease in the likelihood of a re-report (HR=0.50), FPS and FCS was significantly associated with a decrease in the likelihood of a re-report (HR=0.74), and foster care placement was also significantly associated with a decrease in the likelihood of a re-report (HR=0.82).

In addition to providing inconsistent findings regarding the statistical significance, magnitude of effects, and direction of effects for child welfare post-investigation service delivery, findings regarding the delivery of child mental health and special education services are contradictory. Drake, Jonson-Reid, and Sapokaite (2006) found that the child's receipt of a mental health/substance abuse service prior to or within one year of the first report for maltreatment was significantly associated with an increase in the likelihood of a re-report (HR=2.06). Additionally, special education eligibility for an emotional disturbance prior to or within one year of the first report for maltreatment was significantly associated with an increase in the likelihood of a re-report (HR=1.49). Finally, juvenile court involvement prior to or within one year of the first maltreatment report was significantly associated with a decrease in the likelihood of a re-report (HR=0.61). That said, a follow-up study using the same data found that the direction of effects differed for the delivery of child mental health services and special education involvement (Jonson-Reid, Emery, Drake, & Stahlschmidt, 2010). Specifically, child eligibility for special education before the first maltreatment report was significantly associated with an increase in the likelihood of a re-report (HR=1.08), while eligibility for special education after the first maltreatment report was significantly associated with a decrease in the likelihood of a re-report (HR=0.43). Additionally, child receipt of a mental health service after the first maltreatment report was significantly associated with a decrease in the likelihood of a re-report (HR=0.65). The effect of juvenile court involvement was not tested.

Inconsistent Findings in the Extant Literature: What about Door B?

Based upon a thorough examination of key recurrent maltreatment risk assessment studies, a foundation of contradictory findings was identified. As noted earlier, none of the studies tested for curvilinear effects and a very limited number of studies tested for a restricted range of interaction effects. Hence, the typical risk assessment model assumes that the likelihood of recurrent maltreatment is a linear combination of a range of child, parent, family, maltreatment incident, and service delivery predictors. Given the lack of consistency in findings for the four prominent risk factors and the four moderating service variables identified in this dissertation study, one has to ask the question: What about door B - - i.e., what if the likelihood of recurrent maltreatment is not a linear combination of selected predictors?

As noted by Beck, King, and Zeng (2000), a neural network is the ideal statistical method of choice when the structure in the data is best captured by a model capable of estimating “complex, nonlinear, and contingent relationships” (p.22). Inconsistent findings are likely to result if highly context-dependent relationships are estimated by a linear model that assumes that the effects of the predictors are the same over all observations and therefore averages the effects of the variables over all of the observations. In cases where included interaction terms allow the effects to vary across the observations, “the degree of variation represented is still quite limited” (Beck, King, & Zeng, 2000, p. 23).

It should be noted that the 19 key studies assessed in this dissertation study are of extreme importance to the field of recurrent maltreatment risk assessment research. These studies have set the tone for scholarship that attempts to identify the predictors of recurrent maltreatment, and they have made continued advancements possible by creating a foundation from which to learn. Hence, any notation of inconsistency among the findings reported by these 19 key studies is in no way a criticism of the rigor or legitimacy of the scholarship that has been conducted. Rather, notations of inconsistent findings are used to define an empirical reason for deciding to open Door B by applying random forest and neural network analyses in order to explore the possibility that linear models are missing important structure in the data.

Implications for Policy and Practice

In addition to limitations imposed by questionable levels of predictive accuracy, the literature has not generally provided a clear explication of how risk assessment findings can be used to inform child welfare policy and practice (Shlonsky & Wagner, 2005). While the predictive accuracy of the neural network created for this dissertation study is far from perfect, it does appear to represent the highest reported level of predictive accuracy achieved to date. Moreover, in additional neural network analyses not described in this dissertation but that were conducted as part of the overall dissertation study, classification accuracy as measured by the area under the ROC curve increased, moving from a low of 0.79 to a high of 0.94 for different types of re-reported maltreatment as defined by first and second re-reports. Hence, despite a level of predictive accuracy that still leaves sizeable room for misclassification error, the methodology used in this dissertation study was able to produce a classification model that beats the predictive accuracy of a linear model and that uses administrative data - - a source of data that is relatively inexpensive to produce and easy for child welfare workers to access. Moreover, none of the measures used in this dissertation study were based on abstracting data from case records and/or collecting data from families during structured clinical interviews or observation studies. While the predictive accuracy of the neural network model would likely increase upon including such measures, it is important to note that the basic measures found in administrative data can be used to predict the likelihood of a child returning for a re-report with a reasonably good level of accuracy.

The post-hoc analyses conducted for this dissertation study identified a very small subset of features that explained a little over 70% of the variation in the predicted probability of recurrent maltreatment as estimated by the neural network model. Hence, these eight features - - i.e., four risk factors and four service based moderators - - link the assessment of risk with the practice of creating a potential plan for preventive intervention.

Two of the four prominent risk factors, number of income maintenance spells received and exposure to community poverty, are dynamic to the extent that child welfare professionals can work with families in order to identify the specific ways in which poverty increases the likelihood of recurrent maltreatment through a stress-based pathway that alters parental behavior. Beyond coordinating with social welfare professionals who manage the dissemination of income maintenance support, child welfare professionals need to be able to identify the specific poverty-based targets for treatment that each child welfare disseminated intervention will address. Moreover, child age at the first maltreatment report and the parent's status as perpetrator of the first maltreatment incident may not be dynamic, but they do strongly influence the context in which poverty increases the likelihood of maltreatment. Hence, the identification of specific poverty-based targets for treatment should take into account how the likelihood of recurrent maltreatment is influenced in relationship to changes in the child's age as well as the additional supports or interventions primary care givers who are also the perpetrators will need to cope with the stress of caring for their children.

All four of the service-based variables represent dynamic factors to the extent that service system contact appears to moderate the risk of recurrence. Hence, it is important to understand the mechanisms that cause an apparent decrease in risk for one group of children and the lack of such a decrease in risk for the other group of children. As noted by Jonson-Reid (2011), system contact is not the equivalent of understanding exactly which services a child/family received and the characteristics of service delivery to include dose, duration, and quality. While there are many questions that remain unanswered regarding the specific ways in which system contact influences the risk of recurrent maltreatment, to include basic knowledge of just what a system contact is comprised of, it does appear that juvenile court involvement, special education eligibility, FCS receipt, and child mental health/substance abuse service receipt moderate the risk of repeat maltreatment. Children who had contact with at least one of these systems had a lower average likelihood of maltreatment in comparison with children who had no contact with any of the four systems. One of the benefits of including system contact within a risk assessment model is the ability to develop a better understanding of how risk of recurrence changes within the context of system-based experiences a child/family can have.

An increased understanding of how system contact influences other predictors provides greater insight into the ways in which service system contacts could be altered to decrease the risk of recurrent maltreatment. Altering service system contacts could include a variety of approaches such as (a) improving access to and awareness of the need to seek system intervention, (b) improving utilization patterns to promote service receipt for those who need and would benefit from system contact, (c) improving the general and specific responsivity of system contact in relationship to the modification of specific dynamic risk factors, and/or (d) improving surveillance methodology to better understand who has system contact at what points in time and for which reasons. Of course, all of the approaches to system improvement assume that treatment need as defined within an RNR perspective has been adequately conceptualized; specifically, the ability to alter system contact in relationship to treatment need means that researchers understand which modifiable factors raise or lower the risk of repeat maltreatment.

A second step to improving the effects of system contact involves the use of policy and practice to promote the matching of system-based interventions to modifiable risk factors. An effective treatment matching process will only work to the extent that intervention components are capable of altering targeted dynamic risk factors in a fairly prescribed manner. Hence, a successful implementation of a treatment matching procedure depends on the ability to conceptualize how interventions will trigger the mechanisms of change that alter targeted dynamic risk factors for the overall purpose of decreasing the risk of repeat maltreatment. Nonlinear analyses of merged administrative data sets can provide researchers with the tools to develop and test conceptual frameworks for change based on how children and families with specific characteristics interface with a range of public sector service systems. Administrative data can be used to develop such a conceptual framework by providing information regarding who comes to the attention of a service system, the individual and event level characteristics the system responds to, the documented actions the system takes, and case based outcomes (Jonson-Reid & Drake, 2008).

Ultimately, neural networks can be incredibly helpful when attempting to account for the myriad of contingent relationships among the predictors of the likelihood of recurrent maltreatment. Studies regarding the relationships between child maltreatment and key areas for intervention such as a child's need for special education and mental health services paint a picture of a very complex set of interactions where no one risk factor is likely to operate in the same way across all individuals (see, e.g., Jonson-Reid, Drake, Kim, Porterfield, & Han, 2004; Lee & Jonson-Reid, 2009; Leslie et al., 2005). Due to constraints in the types of services that child welfare agencies typically can offer as well as the community referrals that would feasibly result in adequate access to affordable care, indiscriminately calling for improved collaboration and coordination across various systems is not likely to result in structural improvements to the overall delivery of preventive services. Instead, calls for coordination and collaboration should be located within specific approaches to identify and test a limited number of the best available interventions designed to modify key risk factors within various sectors of care (e.g., how best to modify the well-mapped effects of poverty for a family with a two year old child whose mother was the perpetrator of the index event and who is slated to receive FCS).

Future Directions

Like any study, the analyses include in this dissertation are affected by several limitations. *First*, the generalizability of the findings may be limited by the degree to which the sample of children and their primary caregivers can be used to represent the child welfare and cross-sector service delivery experiences of other child welfare-involved children. Specifically, the children in this study lived in a Midwestern metropolitan area, and were largely poor; moreover, information regarding child and primary caregiver mental health treatment was limited to records provided by the department of mental health to include participation in Medicaid and non-Medicaid programs. Data regarding participation in insured and out-of-pocket mental health treatment located in the private sector was not available. Data regarding the child's participation in special education services and the child's juvenile court involvement were limited to one city/county area.

Second, the generalizability of the estimated target function may be limited by the degree to which the cross-validation method accurately estimated the performance of the target function on new data (presumably generated by the same underlying mechanism characterizing the training and validation samples). *Third*, findings from this dissertation study were based upon analyses of child welfare records in a state that employs a differential response to child welfare. That said, the records contained no specific information about the tract to which children were assigned. Generalizability to other populations of children involved in the differential response (DR) systems could be limited to the extent that any DR system is comprised of children and families with potentially unique characteristics. Hence, generalizability is not so much influenced solely by the features of the DR system in and of itself, but by the context in which any given DR system operates to include the particular geographic region in which the DR system is located, the characteristics of the child welfare and community-based organizations, and the characteristics of the families reported to and served by the DR system. The best means of addressing these limitations of generalizability is to engage in future studies that use larger samples of administrative records for children living in a more diverse range of urban and rural areas. Moreover, future studies should include administrative records for children who had system contact with public and private sector organizations throughout multiple counties or regions in various states as opposed to being limited to a more narrow range of service delivery systems that are located in a circumscribed geographic area.

Fourth, measures of system contact should be expanded to include more information regarding the characteristics of service delivery such as dosage and duration. Measures of system contact would also be more informative if various measures of treatment need could be included, such as standardized test scores or data from structured clinical interviews. The administrative data used for this study provided the opportunity to create a wide and rich array of variables that describe basic characteristics across the child, primary care giver, family, perpetrator, maltreatment incident, community, and cross-sector service delivery systems. However, the administrative records did not allow for a very deep exploration of the mechanisms that might be responsible for variation in the likelihood of future maltreatment as a result of parent-child interactions, parental responses to risk factors such as poverty, and child and parent treatment need relative to mental health and special education issues.

Fifth, the child welfare administrative records used for this study were very limited in terms of the degree to which key worker-observed family and perpetrator characteristics varied. Administrative data used for future studies should include measures that workers routinely implement as part of an assessment process for all families. Clinical observations are likely to be characterized by a "hit or miss" approach, wherein the quantity and quality of data may vary considerably across workers. In order to increase the reliability and validity of the key predictors, researchers need to work with child welfare administrators to identify the best possible range of measures across electronic and paper-based records that can be used to better understand how the risk of repeat maltreatment varies in relationship to modifiable factors. *Sixth*, the administrative records used for this study included little, if any, time varying data that could be used to assess for changes in parent and child characteristics in relationship to system contact and the likelihood of repeat maltreatment. Future studies should focus on expanding the nature and quality of predictors by identifying the measures that could be used to track variation in modifiable risk factors across time.

Seventh, and finally, future studies should consistently report measures of model predictive accuracy such as the area under the ROC curve, misclassification rates, sensitivity, specificity, and confusion matrices. Consistency in findings should be taken into account in order to better gage the degree to which each additional study is helping researchers to develop and test an increasingly specific set of hypotheses that can be used to improve preventive service delivery.

References

Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural networks*. Thousand Oaks, CA: SAGE Publications.

Allen, A. O. (1990). *Probability, statistics, and queueing theory with computer science applications*. San Diego, CA: Academic Press.

Allison, P. D. (2002). *Missing data* (Series/Number 07-136). Thousand Oaks, CA: SAGE Publications.

Andrews, D.A. (2006). Enhancing adherence to Risk-Need-Responsivity: Making quality a matter of policy. *Criminology & Public Policy, 5*, 595-602.

Andrews, D.A., & Bonta, J.L. (1998). *The psychology of criminal conduct*. Cincinnati, OH: Anderson.

Andrews, D.A., & Bonta, J.L. (2006). *The psychology of criminal conduct (4th Ed.)*. Newark, NJ: Lexis/Nexis/Matthew Bender.

Andrews, D. A., Bonta, J. L., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*, 19-52.

Andrews, D.A., Bonta, J.L., & Wormith, S.J. (2006). The recent past and near future of risk and and/or need assessment. *Crime & Delinquency, 52*, 7-27.

Andrews, D. A., & Dowden, C. (2007). The risk-need-responsivity model of assessment and human service in prevention and corrections: Crime-prevention jurisprudence. *Canadian Journal of Criminology and Criminal Justice, 49*, 439-464.

Aron, S. B., McCrowell, J., Moon, A., Yamano, R., Roark, D. A., Simmons, M., Tatanashvili, Z., & Drake, B. (2010). Analyzing the relationship between poverty and child maltreatment: Investigating the relative performance of four levels of geographic aggregation. *Social Work Research, 14*(1), 169-179.

Austin, J. (2006). How much risk can we take? The misuse of risk assessment in corrections. *Federal Probation, 70*(2), 58-63.

Bae, H., Solomon, P.L., & Gelles, R.J. (2007). Abuse type and substantiation status varying by recurrence. *Children and Youth Services Review, 29*, 856-869.

Bae, H., Solomon, P.L., Gelles, R.J. (2009). Multiple child maltreatment recurrence relative to single recurrence and no recurrence. *Children and Youth Services Review, 31*, 617-624.

Baird, C., & Wagner, D. (2000). The relative validity of actuarial- and consensus-based risk assessment systems. *Children and Youth Services Review, 22*(11/12), 839-871.

Baird, C., Wagner, D., Healy, T., & Johnson, K. (1999). Risk assessment in child protective services: Consensus and actuarial model reliability. *Child Welfare, 78*(6), 723-748.

Barth, R.P. (2009). Preventing child abuse and neglect with parent training: Evidence and opportunities. *The Future of Children, 19*(2), 95-118.

Beck, N., King, G., & Zeng, L. (2000). Improving quantitative studies of international conflict: A conjecture. *American Political Science Review, 94*(1), 21-35.

Beck, N., King, G., & Zeng, L. (2004). Theory and evidence in international conflict: A response to de Marchi, Gelpi, & Grynaviski. *American Political Science Review, 98*(2), 379-389.

Bia, G., Devroye, L., & Lugosi, G. (2008). Consistency of random forests and other averaging classifiers. *Journal of Machine Learning Research, 9*, 2015-2033.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Cambridge, UK: Oxford University Press.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* New York City, NY: Springer Science and Business Media, LLC.

Bolger, K.E., & Patterson, C.J. (2001). Pathways from child maltreatment to internalizing problems: Perceptions of control as mediators and moderators. *Development and Psychopathology, 13*, 913-940.

Bolger, K.E., Patterson, C.J., & Kupersmidt, J.B. (1998). Peer relationships and self-esteem among children who have been maltreated. *Child Development*, *69*, 1171-1197.

Bolker, B. M. (2008). *Ecological models of data in R*. Princeton, NJ: Princeton University Press.

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123-140.

Brieman, L. (2001a). Random forests. *Machine Learning*, *45*(1), 5-32.

Breiman, L. (2001b). Statistical modeling: The two cultures. *Statistical Science, 16*(3), 199-231.

Camasso, M.J., & Jagannathan, R. (2000). Modeling the reliability and predictive validity of risk assessment in child protective services. *Children and Youth Services Review, 22*, 873-896.

Cash, S.J. (2001). Risk assessment in child welfare: The art and science. *Children and Youth Services Review, 23*(11), 811-830.

Chaffin, M., & Friedrich, B. (2004). Evidence-based treatments in child abuse and neglect. *Children and Youth Services Review, 26*, 1097-1113.

Chen, F. W., & Blackwell, W. J. (2009). *Neural networks in atmospheric remote* *sensing.* Lexington, MA: Massachusetts Institute of Technology.

Cheng, B., & Titterington, D.M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science, 9*(1), 2-54.

Christenson, B., Curran, S., DeCook, K., Maloney, S., & Merkel-Holguin, L. (2008). The intersection between differential response and family involvement approaches. *Protecting Children, 23*(1 & 2), 88-95.

Cicchetti, D., & Toth, S.L. (2005). Child maltreatment. *Annual Review of Clinical Psychology, 1*, 409-438.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/ correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Conley, A. (2007). Differential response: A critical examination of a secondary prevention model. *Children and Youth Services Review, 29*, 1454-1468.

Conley, A., & Duerr Berrick, J. (2008). Implementation of differential response in ethnically diverse neighborhoods. *Protecting Children, 23*(1 & 2), 30-38.

Connell, C.M. Bergeron, N., Katz, K.H., Saunders, L., & Tebes, J.K. (2007). Re-referral to child protective services: The influence of child, family, and case characteristics on risk status. *Child Abuse & Neglect, 31*, 573-588.

Connolly, M. (2005). Differential responses in child care protection: Innovative approaches in family-centered practice. *Protecting Children, 20*(2 & 3), 8-20.

Coulton, C. J., Crampton, D. S., Irwin, M., Spilsbury, J. C., & Korbin, J. E. (2007). How neighborhoods influence child maltreatment: A review of the literature and alternative pathways. *Child Abuse & Neglect, 31*, 1117-1142.

Coulton, C. J., Korbin, J. E., & Su, M. (1999). Neighborhoods and child maltreatment: A multilevel study. *Child Abuse & Neglect, 23*, 1019-1040.

Cox, I., Gaudard, M. A., Ransey, P. J., Stephens, M. L., & Wright, L. T. (2010). *Visual Six Sigma: Making data analysis lean*. Hoboken, NJ: John Wiley & Sons, Inc.

Cullen, F.T. (2005). The twelve people who saved rehabilitation: How the science of criminology made a difference. *Criminology, 43*(1), 1-42.

Cutler, D. R., Edwards, T., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology, 88*(11), 2783-2792.

DePanfilis, D., & Zuravin, S.J. (1999a). Epidemiology of child maltreatment recurrences. *Social Service Review,73*, 218-239.

DePanfilis, D., & Zuravin, S.J. (1999b). Predicting child maltreatment recurrences during treatment. *Child Abuse & Neglect, 23*, 729-743.

DePanfilis, D., & Zuravin, S.J. (2001). Assessing risk to determine the need for services. *Children and Youth Services Review, 23*(1), 3-20.

DePanfilis, D., & Zuravin, S.J. (2002). The effect of services on the recurrence of child maltreatment. *Child Abuse & Neglect, 26*, 187-205.

Diaz-Uriarte, R. & Alvarez de Andres, S. (2006). Gene selection and classification microarray data using random forests. *BMC Bioinformatics, 7*, 3.

Dodge, K., Pettit, G., & Bates, J.E. (1997). How the experience of early physical abuse leads children to become chronically aggressive. In D. Cicchetti & S.L. Toth (Eds.), *Rochester symposium on developmental psychopathology, Vol. 8: Trauma: Perspectives on theory, research, and intervention* (pp. 263-288). Rochester, NY: University of Rochester Press.

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*(5), 441-452.

Drake, B., & Jonson-Reid, M. (2000). Substantiation and early decision points in child welfare: A conceptual reconsideration. *Child Maltreatment*, *5*(3), 227-235.

Drake, B., & Jonson-Reid, M. (2008). *Social work research: From conceptualization to dissemination*. Boston, MA: Allyn & Bacon.

Drake, B., Jonson-Reid, M., & Sapokaite, L. (2006). Rereporting of child maltreatment: Does participation in other public sector services moderate the likelihood of a second maltreatment report? *Child Abuse & Neglect, 30*, 1201-1226.

Drake, B., Jonson-Reid, M., Way, I., & Chung, S. (2003). Substantiation and recidivism. *Child Maltreatment*, *8*(4), 248-260.

Dunteman, G. H. (1989). *Principal components analysis* (). Thousand Oaks, CA: SAGE Publications.

Elkan, C. (2011). *Text mining and topic models*. Davis, CA: University of California at Davis. Retrieved from the University of California at Davis: http://cseweb.ucsd.edu/users/elkan/250Bwinter2011/topicmodels.pdf

Elkan, C. (2012). *Maximum likelihood, logistic regression, and stochastic gradient training*. Davis, CA: University of California at Davis. Retrieved from the University of California at Davis: http://cseweb.ucsd.edu/~elkan/250B/logreg.pdf

English, D.J. (1998). The extent and consequences of child maltreatment. *The Future of Children, 8*(1), 39-53.

English, D.J., & Graham, J.C. (2000). An examination of relationships between children's protective services social worker assessment of risk and independent LONGSCAN measures of risk constructs. *Children and Youth Services Review,* *22*, 897-933.

English, D.J., Graham J.C., Litrownik, A.J., Everson, M., & Bangdiwala, S.I. (2005). Defining maltreatment chronicity: Are there differences in child outcomes? *Child Abuse & Neglect*, *29*, 575-595.

English, D.J., Marshall, D.B., Brummel, S., & Orme, M. (1999). Characteristics of repeated referrals to child protective services in Washington State. *Child Maltreatment, 4*(4), 297-307.

English, D.J., Upadhyaya, M.P., Litrownik, A.J., Marshall, J.E., Runyan, D.K., Graham, J.C., & Dubowitz, H. (2005). Maltreatment's wake: The relationship of maltreatment dimensions to child outcomes. *Child Abuse & Neglect, 29*, 597-619.

English, D.J., Wingard, T., Marshall, D., Orme, M., & Orme, A. (2000). Alternative response to child protective services: Emerging Issues and Concerns. *Child Abuse & Neglect, 24*, 375-388.

Éthier, L.S., Lemelin, J.-P., & Lacharité, C. (2004). A longitudinal study of the effects of chronic maltreatment on children's behavioral and emotional problems. *Child Abuse & Neglect, 28,* 1265-1278.

Evans, J. S., & Cushman, S. A. (2009). Gradient modeling of conifer species using random forests. *Landscape Ecology*, *24*, 673-683.

Faraggi, D., LeBlanc, M., & Crowley, J. (2001). Understanding neural networks using regression trees: An application to multiple myeloma survival data. *Statistics in Medicine, 20,* 2965-2976.

Faraggi, D., & Simon, R. (1995a). Maximum likelihood neural network prediction models. *Biometrical Journal*, *37*, 713-725.

Faraggi, D., & Simon, R. (1995b). The maximum likelihood neural network as a statistical classification model. *Journal of Statistical Planning and Inference, 46,* 93-104.

Ferguson, J. L. (2002). Putting the "what works" research into practice: An organizational perspective. *Criminal Justice and Behavior, 29*, 472-492.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2004). *Applied longitudinal analysis.* Hoboken, NJ: John Wiley & Sons.

Fluke, J.D. (2008). Child protective services rereporting and recurrence—Context and considerations regarding research. *Child Abuse & Neglect, 32*, 749-751.

Fluke, J.D. (2009). Allegory of the cave: On the theme of substantiation. *Child Maltreatment, 14*(1), 69-72.

Fluke, J.D., Shusterman, G.R., Hollinshead, D.M., & Yuan, Y.-Y.T. (2008). Longitudinal analysis of repeated child abuse reporting and victimization: Multistate analysis of associated factors. *Child Maltreatment, 13*(1), 76-88.

Fluke, J.D., Yuan, Y.-Y.T., & Edwards, M. (1999). Recurrence of maltreatment: An application of the National Child Abuse and Neglect Data System (NCANDS). *Child Abuse & Neglect, 23*, 633-650.

Foster, E.M., Prinz, R.J., Sanders, M.R., & Shapiro, C.J. (2008). The costs of a public health infrastructure for delivering parenting and family support. *Children and Youth Services Review, 30*, 493-501.

Fox, J. (2000). *Multiple and generalized nonparametric regression* (07-131). Thousand Oaks, CA: SAGE Publications.

Freisthler, B., Merritt, D. H., & LaScala, E. A. (2006). Understanding the ecology of child maltreatment: A review of the literature and directions for future research. *Child Maltreatment, 11*(3), 263-280.

Fromm, S. (2001). *Total estimated cost of child abuse and neglect in the United States: Statistical evidence.* Retrieved from Prevent Child Abuse America: http://www.preventchildabuse.org/index.shtml

Fryer, G.E., & Miyoshi, T.J. (1994). A survival analysis of the revictimization of children: The case of Colorado. *Child Abuse & Neglect, 18,* 1063-1071.

Fuller, T.L., Wells, S.J., & Cotton, E.E. (2001). Predictors of maltreatment recurrence at two milestones in the life of a case. *Children and Youth Services Review*, *23*(1), 49-78.

Gambrill, E. (2008a). Decision making in child welfare: Constraints and potentials. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 175-193). Oxford: Oxford University Press.

Gambrill, E. (2008b). Providing more effective, ethical services: The philosophy and process of evidence-based (-informed) practice. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 51-65). Oxford: Oxford University Press.

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

Gambrill, E., & Shlonsky, A. (2001). The need for comprehensive risk management systems in child welfare. *Children and Youth Services Review, 23*(1), 79-107.

Garson, G.D. (1991). A comparison of neural network and expert systems of algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review, 9*(3), 399-434.

Garson, G. D. (1998). *Neural networks: An introductory guide for social scientists*. Thousand Oaks, CA: SAGE Publications.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432-435.

Gelman, A., & Hill, J. (2007). *Analytic methods for social research: Data analysis using regression and multilevel/hierarchical models*. New York City, NY: Cambridge University Press.

Gendreau, P. (1996). Offender rehabilitation: What we know and what needs to be done. *Criminal Justice and Behavior, 23*(1), 144-161.

Genuer, R., Poggi, J. M., & Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters, 31*(14), 2225-2236.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London, England: Arnold.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Gill, J. (2006). *Essential mathematics for political and social research*. New York City, NY: Cambridge University Press.

Gill, J., & Womack, A. (in press). The multilevel model framework. In B. Marx, M. Scott, & J. Simonoff (Eds.), T*he SAGE handbook of multilevel modeling*. Thousand Oaks, CA: SAGE Publications, Inc. Retrieved from http://artsci.wustl.edu/~jgill/papers/mlmchap4.pdf

Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *Crime & Delinquency, 52*(1), 178-200.

Gotwalt, C. M. (2011). *JMP 9 Neural platform numerics*. Cary, NC: SAS Institute, Inc.

Graham, J.C., English, D.J., Litrownik, A.J., Thompson, R., Briggs, E.C., & Bangdiwala, S.I. (2010). Maltreatment chronicity defined with reference to development: Extension of the social adaptation outcomes findings to peer relations. *Journal of*

Family Violence, 25(3), 311-324.

Green, B. L., Rockhill, A., & Burrus, S. (2008). The role of interagency collaboration for substance-abusing families involved with child welfare. *Child Welfare, 87*(1), 29-61.

Gujarati, D.N. (2003). *Basic econometrics* (4th ed). Boston, MA: McGraw Hill.

Guo, S., & Zhao, H. (2000). Multilevel modeling for binary data. *Annual Review of Sociology, 26*, 441-462.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed. Revised). Chicago, IL: The University of Chicago Press.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction.* New York City, NY: Springer.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hélie, S., & Bouchard, C. (2010). Recurrent reporting of child maltreatment: State of knowledge and avenues for research. *Children and Youth Services Review, 32*, 416-422.

Holder, T. (2000). The basis for ongoing CPS intervention: CPS staff focused survey summary of findings. In T.D. Morton & W. Holder (Eds.), *Issues and strategies for assessment approaches to child maltreatment* (pp. 14-25). Retrieved from the National Resource Center on Child Maltreatment: http://www.nrccps.org/PDF/IssuesandStrategiesforAssessment.pdf

Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. *American Journal of Epidemiology, 147*(7), 694-703.

Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., Bruckner, T., & Satariano, W. A. (2010). To GEE or not to GEE. Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Epidemiology, 21*(4), 467-474.

Huebner, R.A. (2005). *Program evaluation of the multiple response system: Kentucky*. Retrieved from the National Quality Improvement Center on Differential Response in Child Protective Services: http://www.differentialresponseqic.org/resources/studies/program-evaluation-of-the.html

Ireland, T.O., Smith, C.A., & Thornberry, T.P. (2002). Developmental issues in the impact of child maltreatment on later delinquency and drug use. *Criminology, 40*, 359-396.

Jaccard, J. (2001). *Interaction effects in logistic regression* (Series/Number 07-135). Thousand Oaks, CA: SAGE Publications Inc.

Jagannathan, R., & Camasso, M.J. (1996). Risk assessment in child protective services: A canonical analysis of the case management function. *Child Abuse & Neglect, 20*, 599-612.

Johnson, C., Sutton, E.S., & Thompson, D.M. (2005). Child welfare reform in Minnesota. *Protecting Children, 20*(2 & 3), 55-60.

Jones, P.R., Schwartz, D.R., Schwartz, I.M., Obradovic, Z., & Jupin, J. (2006). Risk classification and juvenile dispositions: What is the state of the art? *Temple Law Review, 79*(2), 461-498.

Jonson-Reid, M. (2002). After a child abuse report: Early adolescents and the child welfare system. *Journal of Early Adolescence, 22*(1), 24-48.

Jonson-Reid, M. (2011). Disentangling system contact and services: A key pathway to evidence-based children's policy. *Children and Youth Services Review, 33*, 598-604.

Jonson-Reid, M., & Drake, B. (2008). Multisector longitudinal administrative databases: An indispensable tool for evidence-based policy for maltreated children and their families. *Child Maltreatment, 13*(4), 392-399.

Jonson-Reid, M., Drake, B., Chung, S., & Way, I. (2003). Cross-type recidivism among child maltreatment victims and perpetrators. *Child Abuse & Neglect, 27*, 899-917.

Jonson-Reid, M., Drake, B., Kim, J., Porterfield, S., & Han, L. (2004). A prospective analysis of the relationship between reported child maltreatment and special education eligibility among poor children. *Child Maltreatment, 9*(4), 382-394.

Jonson-Reid, M., Drake, B., & Kohl, P. (2009). Is the overrepresentation of the poor in child welfare caseloads due to bias or need? *Children and Youth Services Review, 31*, 422-427.

Jonson-Reid, M., Emery, C. R., Drake, B., & Stahlschmidt, M. J. (2010). Understanding chronically reported families. *Child Maltreatment*, *15*(4), 271-281.

Kaplan, C., & Merkel-Holguin, L. (2008). Another look at the national study on differential response in child welfare. *Protecting Children, 23*(1 & 2), 5-21.

Kim, J. S., (2004). Neighborhood effects on the etiology of child maltreatment: A multilevel study. *Dissertation Abstracts International, 65*(08), 3157B (UMI No. 3143286).

King, G., & Zeng, L. (2001). Research note: Improving forecasts of state failure. *World Politics, 53*(July), 623-658.

Kirk, R.S. (2008). Development and field testing of a family assessment scale for use in child welfare practice settings utilizing differential response. *Protecting Children, 23*(1 & 2), 71-87.

Knoke, D., & Trocmé, N. (2005). Reviewing the evidence on assessing risk for child abuse and neglect. *Brief Treatment and Crisis Intervention*, *5*(3), 310-327.

Kohl, P.L., Jonson-Reid, M., & Drake, B. (2009). Time to leave substantiation behind: Findings from a national probability study. *Child Maltreatment*, *14*(1), 17-26.

Koutsoukos, A. D., Rubinstein, L. V., Faraggi, D., Simon, R. M., Kalyandrug, S., Weinstein, J. N., Kohn, K. W., & Paull, K. D. (1994). Discrimination techniques applied to the NCI *in vitro* anti-tumor drug screen: Predicting biochemical mechanism of action. *Statistics in Medicine, 13*, 719- 730.

Larsen, K., & Merlo, J. (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology, 161*(1), 81-88.

Lee, M. Y., & Jonson-Reid, M. (2009). Needs and outcomes for low income youth in special education: Variations by emotional disturbance diagnosis and child welfare contact. *Children and Youth Services Review, 31*, 722-731.

Leslie, L. K., Gordon, J. N., Lambros, K., Premji, K., Peoples, J., & Gist, K. (2005). Addressing the developmental and mental health needs of young children in foster care. *Developmental and Behavioral Pediatrics, 26*(2), 140-151.

Lin, Y., & Jeon, Y. (2006). Random forests and adaptive nearest neighbors. *Journal of the American Statistical Association, 101*(474), 578-590.

Lipien, L., & Forthofer, M.S. (2004). An event history analysis of recurrent maltreatment reports in Florida. *Child Abuse & Neglect*, *28*, 947-966.

Loman, L.A. (2006). *Families frequently encountered by child protection services: A* *report on chronic child abuse and neglect*. Retrieved from the Institute of Applied Research: http://www.iarstl.org/papers.htm

Loman, L.A., & Siegel, G.L. (2004a). *An evaluation of the Minnesota SDM Family Risk Assessment: Final report*. Retrieved from the Institute of Applied Research: http://www.iarstl.org/papers.htm

Loman, L.A., & Siegel, G.L. (2004b). *Differential response in Missouri after five years: Final report*. Retrieved from the Institute of Applied Research: http://www.iarstl.org/papers.htm

Loman, L.A., & Siegel, G.L. (2004c). *Minnesota alternative response evaluation: Final report*. Retrieved from the Institute of Applied Research: http://www.iarstl.org/papers.htm

Loman, L.A., & Siegel, G.L. (2005). Alternative response in Minnesota: Findings of the program evaluation. *Protecting Children, 20*(2 & 3), 78-92.

Luongo, G. (2007). Re-thinking child welfare training models to achieve evidence-based practices. *Administration in Social Work, 31*(2), 87-96.

Maher, E.J., Jackson, L.J., Pecora, P.J., Schultz, D.J., Chandra, A., Barnes-Proby, D.S. (2009). Overcoming challenges to implementing and evaluating evidence-based interventions in child welfare: A matter of necessity. *Children and Youth Services Review, 31*, 555-562.

Manly, J.T., Cicchetti, D., & Barnett, D. (1994). The impact of subtype, frequency, chronicity, and severity of child maltreatment on social competence and behavior problems. *Development and Psychopathology, 6*, 121-143.

Marshall, D.B., & English, D.J. (1999). Survival analysis of risk factors for recidivism in child abuse and neglect. *Child Maltreatment, 4*(4), 287-296.

Marshall, D.B., & English, D. J. (2000). Neural network modeling of risk assessment in child protective services. *Psychological Methods, 5*(1), 102-124.

Merkel-Holguin, L. (2005). Differential response: A common sense reform in child welfare. *Protecting Children, 20*(2 & 3), 2-3.

Merkel-Holguin, L., Kaplan, C., & Kwak, A. (2006, November). *National study on differential response in child welfare*. Denver, CO: American Humane Association and Child Welfare League of America.

Mitchell, T. M. (1997). *Machine learning*. New York City, NY: WCB McGraw Hill.

Molnar, B. E., Buka, S. L., Brennan, R. T., Holton, J. K., & Earls, F. (2003). A multi-level study of neighborhoods and parent-to-child physical aggression: Results from the Project on Human Development in Chicago Neighborhoods. *Child Maltreatment, 8*, 84-97.

National Quality Improvement Center on Differential Response in Child Protective Services. (2009). *Differential response in child protective services: A literature review* (CFDA Number: 93:670). Retrieved from http://www.differentialresponseqic.org/assets/docs/qic-dr-lit-review-sept-09.pdf

Needell, B., Cuccaro-Alamin, S., Brookhart, A., & Lee, S. (1999). Transitions from AFDC to child welfare in California. *Children and Youth Services Review, 21*(9/10), 815-841.

Ortiz, M.J., Shusterman, G.R., & Fluke, J.D. (2008). Outcomes for children with allegations of neglect who receive alternative response and traditional investigations: Findings from NCANDS. *Protecting Children, 23*(1 & 2), 57-70.

Paik, H. (2000). Comments on neural networks. *Sociological Methods & Research, 28*(4), 425-453.

Pampel, F. C. (2000). *Logistic regression: A primer* (Series/Number 07-132). Thousand Oaks, CA: SAGE Publications.

Pandya, A. S., & Macy, R. B. (1996). *Pattern recognition with neural networks in C++*. Boca Raton, FL: CRC Press.

Romanelli, L. H., Landsverk, J., Levitt, J. M., Leslie, L. K., Hurley, M. M., Bellonci, C… Webb, M. B. (2009). Best practices for mental health in child welfare: Screening, assessment, and treatment guidelines. *Child Welfare, 88*(1), 163-188.

Rosenberg, D.A., & Krugman, R.D. (1991). Epidemiology and outcome of child abuse. *Annual Review of Medicine, 42*, 217-224.

Ryan, S., Wiles, D., Cash, S., & Siebert, C. (2005). Risk assessments: Empirically supported or values driven? *Children and Youth Services Review, 27*, 213-225.

Rycus, J.S., & Hughes, R.C. (2008). Assessing risk throughout the life of a child welfare case. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 201-213). Oxford: Oxford University Press.

Sall, J. (2009). *Predicting with bootstrap forest and gradient boosting*. Cary, NC: SAS Institute, Inc.

Sampson, R. J. (1997). Collective regulation of adolescent misbehavior: Validation results from eighty Chicago neighborhoods. *Journal of Adolescent Research, 12*(2), 227-244.

SAS Institute. (2010). *JMP version 9: Modeling and multivariate methods*. Cary, NC: SAS Institute, Inc.

SAS Institute. (2009). *SAS/STAT 9.3 userguide. What's new in SAS/STAT 9.3.* Cary, NC: SAS Institute, Inc.

Sawyer, R., & Lohrbach, S. (2005a). Differential response in child protection: Selecting a pathway. *Protecting Children, 20*(2 & 3), 44-53.

Scarcella, C.A., Bess, R., Zielewski, E.H., Warner, L., & Geen, R. (2004). *The cost of protecting vulnerable children IV: How child welfare funding fared during the recession*. Retrieved from the Urban Institute: http://www.urban.org/UploadedPDF/411115_VulnerableChildrenIV.pdf

Schene, P.A. (1998). Past, present, and future roles of child protective services. *The Future of Children 8*(1), 23-38.

Schene, P. (2005). The emergence of differential response. *Protecting Children, 20*(2 & 3), 4-7.

Schwartz, I.M., Jones, P.R., Schwartz, D.R., & Obradovic, Z. (2008). Improving social work practice through the use of technology and advanced research methods. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 214-229). Oxford: Oxford University Press.

Schwartz, D.R., Kaufman, A.B., & Schwartz, I.M. (2004). Computational intelligence techniques for risk assessment and decision support. *Children and Youth Services Review, 26*, 1081-1095.

Scott, K.L., Wolfe, D.A., & Wekerle, C. (2003). Maltreatment and trauma: Tracking the connections in adolescence. *Child and Adolescent Psychiatric Clinics, 12*, 211-230.

Shalizi, C. (2009). *Classification and regression trees* (36-350, Data Mining). Pittsburg, PA: Carnegie Mellon University. Retrieved from Carnegie Mellon University: http://www.stat.cmu.edu/~cshalizi/350/lectures/22/lecture-22.pdf

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379-423. Retrieved from Bell Labs: http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Children and Youth Services Review, 27,* 409-427.

Shusterman, G.R., Fluke, J.D., Hollinshead, D.M., & Yuan, Y.-Y. T. (2005). Alternative responses to child maltreatment: Findings from NCANDS. *Protecting Children, 20*(2 & 3), 32-42.

Siegel, G.L., & Loman, L.A. (2000). *The Missouri family assessment and response demonstration impact evaluation: Digest of findings and conclusions*. Retrieved from the Institute of Applied Research: http://www.iarstl.org/papers.htm

Smith, M. (1993). *Neural networks for statistical modeling*. New York City, NY: Van Nostrand Reinhold.

Smith, B. D., & Mogro-Wilson, C. (2008). Inter-agency collaboration: Policy and practice in child welfare and substance abuse treatment. *Administration in Social Work, 32*(2), 5-24.

Smith, C.A., & Thornberry, T. (1995). The relationship between child maltreatment and adolescent involvement in delinquency. *Criminology, 33*, 451-481.

Smyth, P. (2007). *Machine learning: Linear classifiers and boosting.* Irvine, CA: University of California at Irvine. Retrieved from the University of California at Irvine: www.ics.uci.edu/~smyth/courses/cs271/topic12_learning_part2.ppt

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: SAGE Publications Inc.

Stine, R. A. (2011). *Data mining*. Philadelphia, PA: Statistical Horizons.

Strobl, C., Malley, J., & Tutz, G. (2009). *An introduction to recursive partitioning*

(Technical Rreport Number 55). Munich, Germany: Institut Fur Statistik, Ludwig Maximilians-Universitat Munchen.

Taxman, F. (2006). Assessment with flair: Offender accountability in supervision plans. *Federal Probation, 70*(2), 2-7.

Thanner, M. H., & Taxman, F. S. (2003). Responsivity: The value of providing intensive services to high-risk offenders. *Journal of Substance Abuse Treatment, 24*, 137-147.

U.S. Department of Health and Human Services. (2003a). *National study of child protective services systems and reform efforts: A summary report*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2003b). *National study of child protective services systems and reform efforts: Findings on local CPS practices*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2003c). *National study of child protective services systems and reform efforts: Review of state CPS policy*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (2009). *Child maltreatment 2007*. Washington, DC: U.S. Government Printing Office.

Wald, M.S., & Woolverton, M. (1990). Risk assessment: The emperor's new clothes? *Child Welfare, LXIX*, 483-511.

Waldfogel, J. (1998). Rethinking the paradigm for child protection. *The Future of Children, 8*(1), 104-119.

Waldfogel, J. (2000a). Protecting children in the 21st century. *Family Law Quarterly, 34*(3), 311-328.

Waldfogel, J. (2000b). Reforming child protective services. *Child Welfare, 79*(1), 43-57.

Waldfogel, J. (2008). The future of child protection revisited. In D. Lindsey & A. Shlonsky (Eds.), *Child welfare research: Advances for practice and policy* (pp. 235-241). Oxford: Oxford University Press.

Way, I., Chung, S., Jonson-Reid, M., & Drake, B. (2001). Maltreatment perpetrators: A 54-month analysis of recidivism. *Child Abuse & Neglect, 25*, 1093, 1108.

Yuan, Y.-Y. T. (2005). Potential policy implications of alternative response. Protecting *Children, 20*(2 & 3), 22-31.

Zeng, L. (1999). Prediction and classification with neural network models. *Sociological Methods & Research, 27*(4), 499-524.

Footnotes

[1] Unless otherwise noted, the terms re-report, re-referral, repeat, and recurrent are used interchangeably to denote the re-occurrence of maltreatment following an initial event of maltreatment. Both the initial occurrence of maltreatment and the subsequent occurrences of maltreatment are measured through the existence of a hotline report of maltreatment that is accepted by and assigned to a response by the CPS worker. Thus, initial and subsequent events of maltreatment represented in this dissertation proposal have not necessarily been substantiated. It is common practice within the extant literature on DR to measure initial and subsequent events of maltreatment via hotline reports that are assigned to a response within the DR system (i.e., the case is assigned to the investigation or assessment track) because cases sent to the assessment track cannot be substantiated (Ortiz, Shusterman, & Fluke, 2008). Furthermore, studies have identified similar predictors for subsequent substantiated reports and unsubstantiated reports of maltreatment (Bae, Solomon, & Gelles, 2007; English, Marshall, Brummel, & Orme, 1999; Fluke, Shusterman, Hollinshead, & Yuan, 2008) as well as a similar risk of recidivism for substantiated and unsubstantiated cases at the index event (Drake, Jonson-Reid, Way, & Chung, 2003; Kohl, Jonson-Reid, & Drake, 2009).

[2] Although JMP Pro 9 assumes a multinomial distribution for any categorical variable (this assumes that categorical response levels are mutually exclusive and therefore not ordered), for a categorical variable with just two response levels, the assumption of a multinomial distribution reduces to an assumption of a binomial distribution (the binomial distribution is a special case of the multinomial distribution, where the number of categorical response levels equals two).

[3] It is not possible to obtain factor scores when using PROC FACTOR in SAS 9.3 unless the data are entered as individual variables as opposed to a correlation matrix. Due to the dichotomous nature of the worker-observed family and perpetrator characteristics, assuming that a linear association exists between any two of the variables, as measured by a Pearson correlation, is inappropriate. Hence, rather than enter the dichotomous variables as singled indicators to be combined by SAS into a Pearson correlation matrix, the %POLYCHOR SAS macro was used to create a tetrachoric correlation matrix for each set of worker-observed family and perpetrator characteristics. The tetrachoric correlation matrix was then entered as the data for the execution of each respective principle component analysis. As noted in Chapter 3, the objective of the principle component analyses with the family and perpetrator characteristics was not so much conducted for the purpose of obtaining a set of factor scores, but rather was conducted for the purpose of identifying an empirical basis for combining specific family and perpetrator characteristics.

[4]Despite the elimination of the variable that measured the receipt of a first FCS spell within 45 days of the first maltreatment report, the receipt of the first FCS spell without the 45 day constraint (receipt on or after the first report but before the second report) remained in the neural network model.

[5]As per the extant literature, a binary logistic regression model (with random intercepts) was used as a counterpoint against which the neural network model's predictive accuracy and flexibility in approximating a wide range of functional forms could be compared (see e.g., Beck, King & Zeng, 2000, 2004; King & Zeng, 2001; Zeng, 1999). Moreover, the selection of a binary logistic model allowed the researcher to specify a two-level hierarchical model that assessed for subject specific effects by

modeling variation in the likelihood of recurrent maltreatment in relationship to between census tract differences.