8-2007

Validation of the Air Force Family Needs Screener

Wendy J. Wyse
University of Tennessee - Knoxville

To the Graduate Council:

I am submitting herewith a dissertation written by Wendy J. Wyse entitled "Validation of the Air Force Family Needs Screener." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Social Work.

Charles Glisson, Major Professor

We have read this dissertation and recommend its acceptance:

William Nugent, Marlys Staudt, Robert T. Ladd

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Wendy J. Wyse entitled “Validation of the Air Force Family Needs Screener.” I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirement for the degree of Doctor of Philosophy, with a major in Social Work.

Charles Glisson
Major Professor

We have read this dissertation
and recommend its acceptance:

William Nugent

Marlys Staudt

Robert T. Ladd

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean
of the Graduate School

(Original signatures are on file with official student records.)

VALIDATION OF THE AIR FORCE FAMILY NEEDS SCREENER

Dissertation Presented for the

Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Wendy J. Wyse

August 2007

DEDICATION

This dissertation is dedicated to my wonderful husband, Todd. I'm so happy I finally found you. I love you, babe.

ACKNOWLEDGEMENTS

There are so many people who have contributed to the completion of this endeavor. I am most grateful to God, who has blessed me beyond measure. I also want to thank my family for their unending love and support. My mother and my father have provided the foundation of love and strength on which I stand today and have helped to shape the woman I am. To Mindy, you have showed me the true meaning of friendship. Thank you for being the most loyal and perseverant person I know.

I am most grateful to my loving husband, Todd. My life is more fulfilled than I ever could have imagined with you in it. Thank you for your encouragement, for your commitment to me and to my career, and for making me Mrs. Wendy Travis.

I am forever indebted to the guidance and mentorship of my friend, Colonel (retired) Jeff Paddock. Without him, I might have never had this opportunity. Likewise, the recent mentorship of Colonel Robert Campbell has been equally pivotal. Thank you both for seeing the “potential” in me, for nurturing it, for listening and for offering your sound advice. You have shown me how I want to mentor others.

Thanks to my peers, Binta and Donna. Binta, you are undoubtedly the strongest woman I know. Donna, thank you for being there when I most needed you. I have the utmost respect and admiration for both of you.

Special thanks go to my wonderful committee. First, to my Chair, Dr. Charles Glisson, thank you for spending so many hours of your time with me. I appreciate your vision and your feedback. Both were absolutely essential to my success. Dr. William Nugent, Dr. Tom Ladd, and Dr. Marlys Staudt, I am grateful for your wisdom, your consultation, your direction and your encouragement.

Thank you Peggy for keeping track of Dr. Glisson and for keeping us organized. Thanks John, for helping me track down paperwork and for helping me find Dr. Nugent. You are both glue that helps hold the college together.

My sincerest thanks go to the Air Force Family Advocacy Program staff members who made this project possible. Specifically, Major Dave Linkh, who supported this project from day one and made sure that I had the resources I needed. To Les Besetsney and Tapan Saha whose skillful manipulation the dataset made my life immeasurably easier. To Meg Walker and Beth Kaiser whose consultation and expertise were vital. Finally, and most importantly, thanks go to those nameless women who bravely completed the Air Force Family Needs Screener. Without them, this project would have never come to fruition.

ABSTRACT

The Air Force Family Needs Screener (FNS) has been utilized as a screening measure for risk of both child maltreatment and intimate partner violence (IPV) within the context of a military primary prevention program called the New Parent Support Program (NPSP) since 1988. However, this measure has only been validated with the initial pilot samples used in its development. The current study sought to assess the reliability, dependability, factor structure and predictive capacity of the FNS using a new sample of mothers. In addition, this study sought to develop shortened versions of the FNS in order to determine if they performed as well as or better than the original version of the measure. Results indicated the original FNS has adequate reliability and dependability, but the cut-score currently used in practice resulted in a fairly low sensitivity rate and corresponding high false negative rate. Since there was no evidence of measurement invariance of a common items model tested across four different population groups, three shortened versions of the FNS made up only of common items to all four groups were also tested. All shortened models appeared to perform as well as the original FNS, suggesting the FNS could be shortened, while retaining its predictive capacity. A shortened model made up of only thirty-six common items was recommended as a revised version of the FNS for implementation with the Air Force. Although not the most parsimonious model, this model retained the highest rate of sensitivity while still improving the acceptability of the measure for staff and respondents alike.

TABLE OF CONTENTS

CHAPTER I - INTRODUCTION 1
Significance of the Problem 1
Rates of Family Violence in the Civilian Population 3
Rates of Family Violence in Military Populations 4
Overview of the Family Violence Screening Debate 6
CHAPTER II – REVIEW OF THE LITERATURE 8
Screening for Family Violence Risk 8
Considerations in Screening 8
Characteristics of Effective Screeners 9
Managing Concerns about Family Violence Risk Screening 17
Risk Factors for Family Violence 20
IPV Risk Factors 21
Child Maltreatment Risk Factors 22
Military-specific Risk Factors 24
Family Violence Risk Screeners 26
Widely-used IPV Measures 27
Other Screening Measures for IPV 28
Widely-used Child Maltreatment Risk Measure 36
Brief Screening Measures for Child Maltreatment 37
Summary of Family Violence Screeners 42
CHAPTER III – CONCURRENT SCREENING FOR IPV AND CHILD MALTREATMENT RISK 43
Support for Concurrent Screening 43
The Air Force Family Needs Screener 47
Background Information on the AF Family Needs Screener 47
Development of the FNS 48
Administration of the FNS 49
Scoring of the FNS 50
Review of FNS research 52
Research Questions for the Current Study 56
CHAPTER IV - METHODS 58
Data Source 58
NPSP Dataset 58
Maltreatment Dataset 59
Design 60
Variables of Interest 62
Family Violence Occurrence 62
Family Violence Risk 63
Respondent Population Groups 63
NPSP Service Utilization 65
Missing Data 65
Samples 66
Data Analyses 70
Reliability and Dependability 70

Confirmatory Factor Analyses....72
Predictive Validity Analyses....74
Development of New Shortened FNS....74
Model Comparison....75
Phase 1 Results....76
Reliability and Dependability of Original Scales....76
Confirmatory Factor Analyses....77
Predictive Capacity of Original FNS....79
Phase 2 Results....81
Exploratory Factor Analysis....81
Reliability and Dependability of Two New Scales....82
Confirmatory Factor Analyses of Common 30 and Common 36 Models....83
Predictive Capacity of the Common 30 and the Common 36 Models....83
Creation of the Shortest Possible Version of the FNS....85
Predictive Validity of Common 12 Model....87
Cross-validation of Prediction Results....88
Summary of Results and Model Comparison....90
CHAPTER VI – DISCUSSION....92
Summary of Major Findings....92
Considerations in Model Comparison....95
Recommendation of Revised FNS....98
REFERENCES....103
APPENDICES....121
VITA....147

LIST OF FIGURES

Figure 1. Sample 1. 125
Figure 2. Sample 2. 126
Figure 3. Sample 3. 127

LIST OF TABLES

Table 1. Theories of Violence Used in the Development of the FNS. 122
Table 2. Current FNS Subscales and Reliabilities of Previous Studies..... 123
Table 3. Respondent Groups Items and Range of Scores. 124
Table 4. Subscales and Items for Common 36 and Common 30 Models. 128
Table 5. Summary of Phases, Analyses, and Samples..... 129
Table 6. Estimates of Reliability and Dependability in Phase 1..... 130
Table 7. Model Fit for Individual Models and Multi-group Comparison. 131
Table 8. Prediction of Subsequent Abuse Controlling for NSPS Services for All Models Using Standardized Scores. 132
Table 9. Cut-score Analyses for All Models..... 133
Table 10. Reliabilities for Common 30 Model Subscales and Total Scale Using Sample 2.1. 134
Table 11. Estimates of Reliability and Dependability for Common 36 and Common 30 Models using Sample 2.2. 135
Table 12. Model Fit for Common 30 and Common 36 Models. 136
Table 13. Prediction of Subsequent Abuse Incident Using Subscales from Common 36 Model. 137
Table 14. Common 12 Model Items. 138
Table 15. Cross-validation of Predictive Validity of Common 12 Model. 139

CHAPTER I - INTRODUCTION

Significance of the Problem

Collectively, child abuse, child neglect, and IPV (also called spouse maltreatment or domestic violence) have been referred to as family violence (Tolan, Gorman-Smith, & Henry, 2006). Although the actual prevalence of family violence varies dramatically depending on the source of the data and the population studied, there is consensus among social work professionals that this problem places a significant proportion of children and adults at risk. Therefore, there is a need for strategies and tools to prevent, assess, and treat it. In 2002, the National Child Abuse and Neglect Data System (NCANDS) reported an estimated 1,400 children died as the result of child abuse or neglect in the United States (U.S. Department of Health and Human Services, 2005). The Center for Disease Control (CDC; 2003) estimated that 1,300 people die each year as the result of intimate partner violence (IPV).

The United States military has recognized the problem of family violence among military members for many years. In 1975, the United States Air Force (AF) became the first branch of the military to develop a child abuse and neglect program. After the eventual inclusion of spouse maltreatment, this program became known as the Family Advocacy Program (FAP; Mollerstrom, Patchner, & Milner, 1992) and was dedicated to the prevention and treatment of family violence among Air Force families. Eventually, the Department of Defense (DoD) approved the first military-wide mandate for the administration of family violence prevention and treatment programs within military communities. The directive defined child and spouse abuse, mandated that each military branch establish a central registry to track all allegations of abuse, and required the

reporting of all such incidents to the respective service's central registry (Department of Defense, 2004). Although DoD directives have helped to bring much needed clarity and continuity to all of the armed services' family violence programs, family violence remains a serious problem for military families. Therefore, the implementation of effective prevention programs is a necessity.

Home visitation programs modeled after Healthy Families of America (HFA) have emerged as one potential child maltreatment prevention program despite mixed evidence regarding its effectiveness (Bilukha et al., 2005; Chaffin & Freidrich, 2004; Duggan, McFarlane et al, 2004; Fraser, Armstrong, Morris, & Dadds, 2000; Olds et al., 1997). In order to identify families who are in need of such prevention services, some home visitation programs have used instruments to screen families for family violence risk. Military family violence prevention programs have important similarities to, as well as distinctions from, civilian prevention programs. Like the civilian sector, the AF has also developed a family violence prevention program based on the home visitation model – the New Parent Support Program (NPSP; Nelson, 1999). And like civilian counterparts, the NPSP utilizes a screening measure in order to assess risk for abuse and prioritize those who are most in need of services. This measure is called the Family Needs Screener (FNS; Kantor & Straus, 1999).

Since the purpose of AF's NPSP is to prevent child abuse and neglect as well as domestic violence (Family Advocacy Program, 2004), the FNS was developed to address all of these facets of family violence. In contrast, most civilian home visitation programs only target child abuse (Guterman, 1999; Fraser et al., 2000). Therefore, screening measures used with these programs often only assess risk for child maltreatment. In fact, a recent meta-analysis found only one study that examined the

effect of home visitation services on the reduction of IPV (Bilukha et al, 2005). Because IPV is also a risk factor for child maltreatment (U.S. Department of Health and Human Services, 2005) and IPV and child maltreatment share many common risk factors (Stith et al., in press; Stith, Smith, Penn, Ward, & Tritt, 2004), the AF's model for family violence prevention is one which might be replicated in the civilian sector.

Rates of Family Violence in the Civilian Population

Reported rates of family violence have varied historically among civilian populations. At least part of this variation is likely due to methodological differences in studies and varied definitions of family violence (Hamberger & Phelan, 2004; Tolan et al, 2006). In one of the first studies of family violence rates, Straus and Gelles (1986) used a longitudinal survey design to compare rates of family violence in a national population. Results indicated that between 1975 and 1986 physical child abuse decreased by 47% and wife beating decreased by 27%, but similarly severe assaults by wives on husbands decreased only 4.3% (Straus & Gelles). More recently, according to the U.S. Department of Health and Human Services (USDHHS), the rate of victimization per 1,000 children in the national population dropped slightly from 13.4 children in 1990 to 12.4 children in 2003 (2005). Children who are the youngest (birth to 3 years of age) had the highest rates of victimization at 16.4 per 1,000 (USDHHS, 2005).

Estimates of rates of intimate partner violence (IPV) are especially varied. For women, rates have been estimated at 47 IPV assaults per 1,000 and for men, 32 assaults per 1,000 (Center for Disease Control [CDC], 2006). However, more conservative estimates originate from the criminal justice system. The Bureau of Justice Statistics (BJS) reported that 5.0 per 1,000 women experienced a violent crime perpetrated by an intimate partner in 2001, a decrease from the rate of 9.8 per 1,000 in

1993 (U.S. Department of Justice [USDOJ], 2003). Rates of violence against men were reported to be 1.6 per 1,000 in 1993 and just 0.9 in 2001 (USDOJ). Clearly, it is likely that the BJS data underestimates the actual prevalence of IPV as it only includes incidents that were reported to the authorities.

Other research has found widely varied IPV prevalence rates among specific groups of people. Hamberger and Phelan (2004) synthesized studies reporting prevalence rates for spouse maltreatment in medical and mental health settings. In the maternal care setting, a review of over 20 studies estimated prevalence rates ranged from 3.2% to 46%. Eleven studies conducted in a primary care setting found rates ranging from 5.5% to 45%. In the mental health setting, ten studies reported spouse maltreatment rates ranging from 8% to 78%, depending on the patients' diagnosis. Finally, of the five studies reviewed from the emergency department setting, rates ranged from 2.2% to 54% depending on the gender of the patient and the patient's presenting problem (Hamberger & Phelan). Hamberger and Phelan stated the wide disparity in prevalence rates can be partly accounted for by (1) varied definitions of abuse; (2) failure to control for potential confounding variables using multivariate statistics or a case-control methodology; and (3) the use of unsystematic or non-standardized approaches to data collection.

Rates of Family Violence in Military Populations

Among the branches of the military, rates of child maltreatment and IPV appear equivalent. A review of six published studies examining rates of abuse in all of the services indicated similar rates of abuse across services (McCarroll et al., 1999). A comparison between civilian and military rates of child maltreatment can easily be obtained due to mandated centralized reporting and investigation of child maltreatment

allegations. Among an Air Force population, trends indicated a slight increase (4.1%) in child abuse reporting from 1987 to 1992, which was comparable to civilian child maltreatment reporting rate increases during a similar time period (Mollerstrom, Patchner & Milner, 1995). Likewise, the substantiation rates of child abuse allegations for an AF population ranged from 46% to 51% from fiscal years 1987 through 1992 (Mollerstrom et al., 1995). An averaged substantiation rate for state child protection programs in 1990 was reported to be 39% (ranging from 15% to 63%), which was very similar to the AF substantiation rate for the same year (Daro & McCurdy, 1991). According to the AF FAP, in 2003, the child substantiation rate per thousand was 7.1 (personal communication, L.K. Besetsny, September 15, 2004) as compared with a substantially higher civilian rate per thousand of 12.4 (U.S. Department of Health and Human Services, 2005).

A study comparing five-year rates of child maltreatment in the Army compared with civilian rates revealed similar findings (McCarroll, Ursano, Fan & Newby, 2004). Overall, civilian rates were consistently double that of Army rates, but a closer examination by type of abuse revealed that rates of neglect in the civilian population accounted for this difference (McCarroll et al., 2004). Therefore, it appears as if the military population may experience a lower rate of child neglect, but may experience similar rates of other types of child maltreatment.

Comparisons between civilian and military populations are more difficult to ascertain when examining IPV. The AF reported a spouse maltreatment substantiation rate ranging between fourteen and sixteen per thousand between the years of 1998 and 2003 (personal communication, L.K. Besetsny, September 15, 2004). There is no civilian counterpart that tracks allegations and substantiation rates for spouse maltreatment

because there is no centralized data collection. As stated above, estimates vary widely and have been reported anywhere from 5 to 47 assaults per 1,000 women (CDC, 2006; USDOJ, 2003). Because of inconsistencies in reporting of IPV and wide disparity in data collection procedures, it is difficult to compare military and civilian rates of IPV.

Overview of the Family Violence Screening Debate

Despite consensus that family violence is an important sociological problem, the past thirty years has produced debate regarding the appropriateness of screening for risk of family violence. Assuming a reliable and accurate screening measure can be developed, proponents have cited the importance of identifying families who are most at risk. There are three reasons this may be important. First, identification of those most at risk would help to ensure that scarce prevention resources are deployed to those who need it the most (Daro & McCurdy, 1994; McCurdy, 1995). Also, Fraser, Armstrong, Morris and Dadds (2000) point out that the ability to identify and engage appropriate target populations has been a constraint of program evaluation efforts. An empirically-validated screening measure may be able to better identify the appropriate target population for services and assist in alleviating this constraint. In addition, screening can help to identify significant risk factors so that prevention services might be appropriately tailored (Avison, Turner & Noh, 1986).

Opponents cite technical, ethical, or social arguments against the use of screeners in the identification of families at risk for family violence. Technical arguments relate to the inherent challenges resulting from screening low base rate events and to the difficulty in developing measures that have been psychometrically validated. Opponents have cited problems associated with the high number of people incorrectly

identified as at risk (termed “false positives”), including the allocation of scarce prevention resources to those not truly at risk or the potential negative consequences of being mislabeled as an “abuser” (Caldwell, Bogat, & Davidson, 1988; Lyons & Doueck, 1996). Ethical and social arguments are often made about the lack of consensus of a reliable definition of family violence (Howing, Wodarski, Kurtz, & Gaudin, 1989; Tolan et al., 2006; Weis, 1989) and inappropriate reliance on reported or confirmed cases of child abuse as the criterion against which prediction is measured (Caldwell et al., 1988).

Despite these concerns, others argue that the potential benefits of screening for family violence can outweigh the potential costs if programmatic contextual factors are considered, the measure is acceptable to both the target population and agency staff, and sufficient attention is paid to the technical challenges of screener development (Daro & McCurdy, 1994; McCurdy, 1995). It seems possible to find a suitable balance so that sparse prevention services resources might be used effectively and the rights of clients are carefully protected. Therefore, in order to determine whether the potential benefits of screening for family violence using the AF FNS outweigh the potential costs, the psychometric properties of the FNS must be thoroughly examined. This study sought to further the psychometric validation of the FNS by building on what is currently known about the FNS and by exploring the predictive validity of the FNS for the first time.

CHAPTER II – REVIEW OF THE LITERATURE

Screening for Family Violence Risk

Family violence screening measures have been developed for use with different populations, in different settings and for different purposes. Regardless of the population screened, the setting, or the purpose of screening, sufficient attention should be paid to the technical challenges of screener development (Daro & McCurdy, 1994; McCurdy, 1995). The next section of this paper will review the key principles of screening for early detection of disease or disorder. Focus will be on the appropriate psychometric and epidemiologic properties of screening instruments. Existing family violence screening measures will then be reviewed with consideration given to these properties.

Considerations in Screening

Since the early 1950's, screening has been defined as "the presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures that can be applied rapidly" (Wilson & Jungner, 1968, p.11). There are two primary purposes of screening: (1) to identify those persons who probably have a particular disease or condition from those who probably do not in order to offer early treatment or to offer a more thorough screening or diagnostic procedure, and (2) to detect risk factors that put a person at higher risk for developing the disease or condition in order to prevent the disease from occurring (Eddy, 1991; Ferrer, 1968). Consistent with Eddy (1991), the term "condition" will be used in this paper as a general term to describe any disease or condition that is the target of screening. The nature of presumptive identification assumes that diagnosis of the condition would follow at some later point in time (Friis & Sellers, 2004). In contrast, diagnosis pertains to the

identification or confirmation of the exact cause of a patient's chief complaints through questions, examinations, and other tests (Sackett & Holland, 1975). This suggests that screening precedes diagnosis, and should not be considered diagnostic by itself (Ferrer, 1968; Wilson & Jungner, 1968).

Pioneers of screening for a condition established principles for early detection (Ferrer, 1968; Wilson & Junger, 1968). The condition to be screened must be an important health problem for which adequate facilities, policy, and treatment services are available. The condition should have a recognizable latent or early symptomatic stage and the natural history of the condition must be adequately understood. Finally, an appropriate test for the condition should be available and the test must be acceptable to the population of interest (Ferrer, 1968; Wilson & Junger, 1968).

Characteristics of Effective Screeners

Effective screeners have specific characteristics. Screening tools must be brief, simple, inexpensive, safe and acceptable to the population of interest (Ferrer, 1968; Friis & Sellers, 2004; Sackett & Holland, 1975). In addition, screening tools should be based on identifiable risk factors, especially those risk factors that have been adequately researched and are modifiable. Finally, screening tools are intended to identify those individuals who may be at risk for a specific condition, but as stated previously, are not intended to be used diagnostically (Wilson & Jungner, 1968). However, the ability of a screening tool to accurately and reliably classify at-risk individuals is the central issue in the screening literature.

There are two bodies of literature that provide important information when evaluating how well a screening measure performs: educational psychology literature and medical/epidemiological literature. Classical measurement theory discusses the

importance of reliability and validity in creating measurement instruments. Medical epidemiology literature highlights the importance of sensitivity and specificity in identification of disease risk. Both of these fields converge when evaluating the effectiveness of screening instruments. The main concepts from each body of literature will be discussed in detail.

Reliability. From classical measurement theory, the psychometric properties of screening instruments must be rigorously investigated. Feldt and Brennan (1989) describe reliability as the "quantification of the consistency and inconsistency in examinee performance" (p.105). The amount of consistency in an individual's score is referred to as "error of measurement" (Feldt & Brennan). Reliability has been quantified in two ways: consistency or agreement within a measurement and consistency of responses across repeated uses of a measure (Feldt & Brennan).

Screening and other measurement instruments can be either norm-referenced or criterion-referenced measures. According to Glaser (1963), the two types of measures differ in the standard used as the reference. Criterion-referenced measures compare one score to an absolute standard (i.e. criterion) or well-defined domain (Brennan, 1980). Norm-referenced measures, however, compare one score to other scores on the measure using a relative rank ordering of all scores, without regard to an absolute standard for comparison (Glaser, 1963). It is important to note that in the context of screening measures, whether or not one individual screens positive for the condition would not logically be based on whether another individual screens positive for the condition. Therefore, when the purpose of the measure is to evaluate the individual's level of risk for having a particular condition, criterion-referenced measures would be more appropriate (Martuza, 1977).

One must consider the inherent differences between criterion-referenced measures and norm-referenced measures when evaluating which estimates of reliability should be used. Classical reliability coefficients (such as Cronbach's α) are not appropriate indices for use with criterion-referenced measures (Popham & Husek, 1969). Instead, other indices, such as Brennan's (1980) dependability indices [e.g. φ or φ (λ)], have been developed for use with criterion-referenced measures. These indices measure the expected agreement in scores across randomly parallel versions of a measurement procedure, either controlling for chance (φ) or given a specified cut score [(φ (λ)]. When establishing the reliability of a set of scores from a screening measure, the most appropriate indicator(s) of reliability must be chosen.

Another important consideration in evaluating the reliability of scores on a screening measure is to understand the stability of the condition that is being screened. In the context of screening measures, it is possible that a person's risk for a particular condition may vary over time, depending on whether the condition is dynamic or static. Therefore, according to Feldt and Brennan (1989), an estimate of the reliability of a set of scores on a screening measure at a particular moment might be preferable to an estimate of a respondent's risk over time. That is, a screening measure may have low test-retest reliability, but responses might still be considered reliable if internal consistency reliability is deemed more important than consistency of measures over time (Cohen & Swerdlik, 2005; Straus & Kantor, 2005).

Validity. The validity of the scores on a screening measure is another important psychometric property that must be evaluated. According to Messick (1989), validity is "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and

actions based on test scores or other modes of assessment" (p.13). Specifically, there are three types of validity: (1) content validity, (2) criterion-related validity, and (3) construct validity. According to Messick (1989), "content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter from the domain about which conclusions are to be drawn" (p.16). For screening instruments, a content-valid measure includes only items that are risk indicators of the condition being screened and adequately addresses all of the subject matter related to that condition.

Criterion-validity is evaluated "by comparing the test scores with one or more external, reference variables considered to provide a direct measure of the characteristic or behavior in question" (Messick, 1989, p. 16). In other words, it speaks to how adequately a test score can be used to infer a person's standing on some criterion of interest (Cohen & Swerdlik, 2005). Criteria can include variables that are assessed prior to, at the same time as, or after the instrument in question is administered. Concurrent validity indicates the extent to which an instrument estimates an individual's present standing on the criterion, predictive validity indicates the extent to which an individual's future level on the criterion is predicted from prior test performance, and postdictive validity indicates the extent to which the instrument estimates an individual's criterion values from the past (Messick, 1989). If the primary purpose of a screening instrument is to identify risk for the future presence of a condition, it is particularly important to establish the predictive validity of the measure (Caldwell et al., 1988).

Finally, construct validity is "evaluated by investigating what qualities a test measures, that is, by determining the degree to which certain explanatory concepts or constructs account for performance on the test" (Messick, 1989, p.16). This is generally based on an integration of evidence that relates to the interpretation or meaning

assigned to scores on a measure (Messick). This type of validity also takes into account the relationships among item scores or between the test and other measure (Messick). In addition, construct validity is strengthened if the interpretation made from scores on a measure is consistent with theoretical rationale (Messick). Many authors have suggested that construct validity is the unifying concept with respect to validity and that it encompasses all of the other types of validity within it (Cohen & Swerdlik, 2005; Mehrens & Lehmann, 1991; Messick, 1989).

Sensitivity and Specificity. A screening measure that is designed to make classification decisions can either correctly or incorrectly classify cases. Classifications made by a screening measure are compared with some independent criterion (Sedlak, 1988). Ideally, this independent criterion would be the reality of a case's true classification. This is rarely ever known, however. Instead, some other criterion must be used as an approximation of reality (Sedlak, 1988). If the purpose of a screening measure is to detect some future occurrence of a condition, then the degree to which the measure accomplishes this is another part of the validity of scores from the measure (Mausner & Kramer, 1985). This can be examined in two different ways: (a) sensitivity – the proportion of individuals identified as at risk for the condition when they actually are at risk for the condition, and (b) specificity - the proportion of individuals identified as not at risk for the condition when they actually are not at risk for the condition (Sedlak, 1988). Ideally, a screening measure would correctly identify all cases, and would therefore have a sensitivity and specificity of 100%; however, this is rarely, if ever, possible in practice (Cochrane & Holland, 1971). Rather, these two characteristics of a screening measure are interdependent (Sedlak, 1988). If a screening measure uses more relaxed criteria for classifications, the sensitivity of the measure will increase, but

the specificity of the measure will generally decrease (Sedlak, 1988). In addition, the lower the sensitivity of a measure, the more cases that are truly at risk will be missed (termed “false negatives”; Sedlak, 1988). The same is true for the relationship between specificity and cases that are incorrectly screened at risk (“false positives”). The lower the specificity of a measure, the more false positives will result (Sedlak, 1988).

Also related to sensitivity and specificity are the predictive values of a screening measure (the positive and negative predictive values [PPV and NPV]). Specifically, the PPV refers to “the proportion of true positives…among all those who have positive test results” (Mausner & Kramer, 1985, p.220). Therefore, the PPV of a measure is the extent to which that measure can predict the condition for a particular individual. Conversely, the negative predictive value of a screening measure refers to the proportion of true negatives among all of those who have negative test results (Mausner & Kramer, 1985). The positive predictive value (PPV) of a screening measure is calculated using the following formula:

PPV = (Prevalence × Sensitivity) ÷ [(Prevalence × Sensitivity) + (1 - Prevalence)(1 -Specificity)].

Thus, according to Sedlak (1988, p.327), “the positive predictive value of a screening measure … is a function of three factors: (a) its sensitivity; (b) its specificity; and (c) the true prevalence of [the condition]”.

Because sensitivity, specificity, PPV and NPV are all affected by disease prevalence, screening measures developed to identify conditions with a low base rate will generally misidentify a large number of individuals as having the condition when in fact they do not, even with fairly high sensitivity and specificity (Munro, 2004). Therefore, Munro (2004) cautions against over-confidence in the results of instruments used to

screen low base rate events. Since the problem of screening accuracy has been a central issue in the family violence literature for several years (Sedlak, 1988), the characteristics of screening measures described here must be thoroughly investigated, especially if the measure seeks to classify individuals according to future risk for a particular condition.

Establishing Cut-points. If the purpose of the screening measure is to identify whether respondents are either "at risk" or "not at risk" for a condition, another important characteristics of an effective screener is the establishment of a standard or cut-point used to make such determinations. Generally, risk determinations are made for purposes of convenience in order to make decisions about whether or not to provide services or more in-depth assessments.

There are two different methods for setting cut-scores for screening instruments: judgmental methods or empirical methods (Hambleton, 1980). Judgmental methods require decisions to be made about the responses on a screener, usually based on clinical expertise. For example, a particular agency might have clinicians review responses to screeners and determine if a family is "at risk" based on their clinical assessment. This method is relatively arbitrary. The second method is empirically-based. There are generally two different approaches to this method in making decisions about an appropriate cut-score. First, the cut-score may be adjusted based on known group comparisons, depending on the number of false positives and false negatives one is willing to tolerate. Secondly, if the prevalence of a condition in the population is known, the cut-score could be set at a level that is consistent with this population value. Although empirical methods are somewhat less arbitrary, they are only as accurate as the known group or population values used in comparison. Finally, a combination of both

methods may also be used. Clearly, both judgmental and empirical methods of cut-point setting are subject to inaccuracy (Hambleton, 1980), but lead to important decisions made about respondents to a screening measure.

Once a cut-score has been set, there are several important points for consideration in order to validate the cut-score. First, the score that indicates a person is "at risk" for a condition should be the same for everyone in a specified population (Hudson, 1982). For example, are all women and pregnant women at the same level of risk for IPV? If not, different cut-scores may need to be established for women who are pregnant and those who are not, thereby treating pregnant women as a separate population from non-pregnant women. Also, the screening measure must perform in a manner that is generally consistent for a large proportion of the respondents (Hudson, 1982).

Finally, consideration must be given to the use of the cut-score in the clinical or preventative health setting. The cut-score might be set at different point depending on whether the purpose of the cut-score is to assist in the establishment of therapeutic goals or to diagnose a condition (Hudson, 1982). This means that if the purpose of a screening measure is to offer preventative services to a population, the cut-score may be set in such a manner so that a lower number of false negatives are produced as long as the potential negative effects of this misclassification are insubstantial. Clearly, although the establishment of a cut-score for a screening measure is a somewhat arbitrary process, careful consideration of these issues and the inclusion of empirical methods can help to strengthen confidence in this decision.

Thus, reliability, validity, sensitivity, specificity and the establishment of cut-points are all properties of screening measures that result in evidence used to determine how

well the screener is functioning. Although there are no absolute standards for these characteristics, it is apparent that a balance must be reached based upon consideration of the purpose of the screening measure, the population of interest and the programmatic needs of providers. It is reasonable to conclude that a screening measure designed to identify future risk of a condition should demonstrate a high degree of sensitivity and should also be acceptable to the population of interest. In addition, the predictive capacity of the screener should also be established.

Managing Concerns about Family Violence Risk Screening

As discussed previously, opponents to screening for family violence usually state the inherent challenges resulting from screening low base rate events and the problem with misidentification of a large number of false positives (Munro, 2004). Arguments are also made about the lack of consensus of a reliable definition of family violence (Howing et al., 1989) and inappropriate reliance on reported or confirmed cases of child abuse as the criterion against which prediction is measured (Caldwell et al., 1988). There are appropriate ways to manage these concerns, however.

Cochrane and Holland (1971) stated the decision to screen "must ultimately rest on a subjective judgment of the number of false positives and false negatives which would be tolerable to the population and the providers of the screening service" (p.6). Sackett and Holland (1975) argue that since the purpose of screening tests is to separate apparently healthy people into groups of high and low probabilities of risk for a given disease or disorder, the most crucial properties of a screening measure are acceptability to the population of interest and sensitivity of the measure. Other properties such as specificity and precision are of moderate importance for screening measures (Sackett & Holland, 1975). Wilson and Jungner (1968) support this position by asserting

that a screening test is allowed to possess a higher margin of error and may be less valid than a diagnostic test. They go on to say that a fairly high false positive rate may be more acceptable in preliminary screening, but that a high false negative rate may result in missed cases and negative individual outcomes which would be considered less acceptable (Wilson & Jungner, 1968).

For screening families who may be at risk for family violence so that appropriate prevention services might be offered, one might be able to tolerate a higher misclassification rate (into the at-risk category) as long as families who are misclassified are not subjected to any legal, civil or psychological distress as the result of the misclassification (Caldwell et al., 1988). Sedlak (1988) purports that a high false positive rate, in and of itself, does not render a screening measure useless. Instead, the key issue in managing the likely high false positive rate inherent to screening low base rate events is careful consideration of what one does with the results of such a screening. He states that screening measures may be appropriate if they are used to identify individuals, couples or families who warrant further questioning, follow up, or to offer voluntary preventative services (Sedlak, 1988).

When developing and implementing screening measures for use in the prevention setting, there are three important considerations. First, developers of the screening measure should take appropriate steps to avoid the use of derogatory terms such as abuse, violence, injury or harm (Straus & Kantor, 2005). Second, results of screening measures utilized under prevention programs should only be used to provide prevention services and such results should be kept confidential so that the family's privacy is protected to the fullest extent possible. Third, within the context of a prevention program, one must consider the potential disadvantages associated with lowering the

sensitivity of a screening measure. Doing so would likely decrease the number of false positives, but would increase the number of false negatives. If the goal of screening for a prevention program is to offer services to those most at risk, a high number of false negatives would likely lead to missed cases. Such “missed cases” may not receive services and this could result in poor outcomes for some families (Wilson & Jungner, 1968). Consequently, a high false positive rate may be more acceptable to prevention program administrators as compared with a high false negative rate.

This “to screen or not to screen” debate was recently emphasized in a meta-analysis conducted by Guterman (1999). This author reviewed 19 studies that examined the efficacy of home visitation programs and categorized the studies based on the enrollment approaches used by each program. The author defined programs as population-based if they either did not do any screening or if they only employed population or demographically-based screening strategies. Programs were considered screening-based if they actively screened families with checklists or other measures to identify indicators of individual-level risk markers, either alone or in combination with demographic risk indicators. Based on the meta-analysis, the author concluded there was a discernable trend favoring population-based enrollment approaches over those that actively screened for individual-level risk (Gutterman, 1999).

However, a careful review of the results of this study challenges Guterman’s conclusions. The effect sizes for both types of enrollment were very small (ranging from + 3.72 to -.0.07 and + 0.092 to + 0.020, for population-based enrollment and screening-based enrollment respectively depending on the outcome measure used). In addition, the researcher did not conduct any inferential statistics to determine if the differences in mean effect sizes were statistically significant. Because of the small sample size (N=19),

it is likely that the reported mean effect sizes for the two types of program entry would be non-significant. With these important factors in mind, the results of this meta-analysis are inconclusive at best.

If one further analyzes the pros and cons of screening for family violence, one would need to acknowledge that population-based entry into prevention services as defined by Guterman (1999) would likely yield a higher number of false positives than a reasonably valid screening-based entry. Furthermore, even the so-called “false positives” are based on criterion (like substantiated child protective services reports) that are likely underestimated. Therefore, even those who appear to be misidentified as at risk for family violence may in fact still be families who might benefit from an effective prevention program. Admittedly, these conclusions are based on the assumption that one can identify a “reasonably valid” screening measure that would be an improvement over a population-based approach. In order to determine if such a screening measure exists, the next section of this paper will review relevant risk factors for family violence and critically analyze the major screening measures currently in use.

Risk Factors for Family Violence

According to Wilson and Jungner (1968), screening measures must be based on identifiable risk factors which are widely known and modifiable. The purpose of family violence screeners is to detect risk factors that put a person at higher risk so that prevention services can be offered (Eddy, 1991). Other authors have identified the importance of definitional issues in the development of valid screeners for family violence (Caldwell et al., 1988; Howing et al., 1989; Straus & Kantor, 2005; Tolan et al., 2006; Weis, 1989). For at least thirty years, research has produced a plethora of

information regarding risk factors for child abuse, child neglect and IPV. However, much of these data use only bivariate statistics or correlational analyses. In this section, risk factors for child maltreatment and IPV will be summarized. Risk factors for family violence can be characterized using three different types of factors: individual factors, relationship factors or contextual/situational factors (Tolan et al., 2006). These three factors will form the basis for this review of family violence risk factors.

IPV Risk Factors

In a recent review, Riggs and colleagues (2000) summarized risk factors for both victimization and perpetration of IPV, but point out that few of the studies included empirical methods. Individual risk factors for IPV mostly characterize the perpetrator. In a nationally representative household survey, research found that male and female perpetrators who witnessed childhood family violence were more likely to report violent acts (OR ranging from 1.6 to 2.4; Kessler, Molnar, Feurer, & Appelbaum, 2001). In addition, male perpetrators of minor violent incidents were more likely to report depression (OR=1.9), generalized anxiety disorder (OR=2.1), alcohol dependence (OR=1.9) and non-affective psychosis (OR=9.4) prior to the initiation of violence (Kessler et al., 2001) while female perpetrators of minor violence were less likely to report panic disorder (OR=0.6), more likely to report social phobia (OR=1.6) and more likely to report substance dependence (OR=2.1; Kessler et al., 2001). Male perpetrators of severe violent acts were more likely to report dysthymia (OR=7.0), adult antisocial behavior OR = 2.9) and non-affective psychosis (OR=29.0; Kessler et al., 2001). Finding a link between alcohol use and aggressive behavior, Field, Caetano, and Nelson (2004) found that those who reported strong or very strong expectations of aggressive behavior after consuming alcohol were 3.2 times more likely to perpetrate IPV. Similarly, using a large

sample of victims of IPV, Thompson and Kingree (2006) found that women whose partners had been drinking were 1.75 times more likely to be injured when compared with women whose partners were not drinking.

Pan and others (1994) compared a large sample of violent and non-violent men across many different variables. Results are expressed in percentage change in odds for each variable. Compared to non-violent men, violent men who engaged in severe acts of violence were 19% less likely to be older in age, 5% less likely to have a higher income, 128% more likely to have an alcohol problem, 121% more likely to have a drug problem, 183% more likely to experience marital discord, and 74% more likely to experience depressive symptoms.

A recent meta-analysis examining effect sizes from 85 studies for physical abuse perpetration among intimate partners found the largest effect sizes for presence of concurrent emotional abuse (r =.49), forced sex (r =.45), illicit drug use (r =.31), attitudes condoning marital violence (r =.30) and marital satisfaction (r =-.30; Stith et al., 2003). Moderate effect sizes were found for traditional sex-role ideology (r =.29), anger/hostility (r =.26), career or life stress (r =.26), history of partner abuse (r =.24), alcohol use (r =.24), and depression (r =.23; Stith et al., 2003). This review confirms results of previous studies that IPV risk factors span all three factor types. With the exception of family of origin violence and history of partner abuse, the remaining risk factors all appear to be modifiable and would be appropriate risk factors to aid in the development of screening measures for IPV.

Child Maltreatment Risk Factors

Individual risk factors have also been identified as significantly associated with child maltreatment. For victims, younger children had the highest rates of victimization at

16.4 per 1,000 and African-American children (20.4 per 1,000) had nearly twice the rate of victimization as compared to White children (11.0 per 1,000; USDHHS, 2005). Mothers who had a child maltreatment allegation during a child's first four years of life were less likely to report a positive support system (OR=.72), were less educated (OR=.63), were more depressed (OR=1.41), were more likely to report alcohol use (OR=1.77), were more likely to have two or more children in the home (OR=1.52) and were more likely to be receiving income support (OR=1.48; Kotch, Browne, Dufort, Windsor, & Catellier, 1999). Family relationship factors have also been found to be significant. The presence of domestic violence in the family during the first six months of a child's life was found to increase the likelihood of a subsequent child physical abuse incident (OR=3.38), of a psychological abuse incident (OR=2.20), and a neglect incident (OR=2.18) by the child's fifth birthday (McGuigan & Pratt, 2001).

A recent meta-analysis consisting of data from 155 studies and 39 different risk factors revealed large effect sizes between child physical abuse and a parent's perception of the child as a problem (r =.30), parent anger (r =.34), family conflict (r =.39) and family cohesion (r =-.32; Stith et al., in press). For child neglect, large effect sizes were found for six risk factors: child's social competence (r =-.30), parent-child relationship (r =-.48), parent's perception of child as problem (r =.41), parent's level of stress (r =.38), parent's level of anger (r =.35), and parent's self-esteem (r =-.69; Stith et al., in press). This review further validates the findings of previous studies and supports the contention that child maltreatment risk factors span the three risk domains used by Tolan and colleagues (2006). With the exception of individual level factors associated with the victim and a history of antisocial behavior on the part of the perpetrator, the

remaining risk factors appear to be modifiable and would be appropriate indicators for the development of a screening measure for child maltreatment.

Military-specific Risk Factors

There have also been studies conducted specifically on military families or comparisons made between military and civilian families with regard to risk factors for family violence. Demographic factors were found to be associated with a history of IPV in a Navy sample such that women (31% vs. 11%; $\chi^2_{partial}$ =(1, N = 1367) =64.77, p < .001), married participants (38% vs. 34% of other marital statuses; $\chi^2_{partial}$ =(2, N = 1367) =14.21, p < .001) and African American participants (41% vs. 24% of Hispanics and 19% of Whites and other ethnicities; $\chi^2_{partial}$ =(3, N = 1367) =23.59, p < .001) were more likely to have experienced IPV (Merrill, Crouch, Thomsen & Guimond, 2004). Furthermore, Merrill and others also found that alcohol problems were predictive of IPV (OR=3.21). Using logistic regression, Pan and colleagues (1994) found that rates of engaging in the most severe violent behaviors (beating up spouse and threatening partner with gun or knife) were doubled in a military sample as compared with another study's age-matched civilian sample. A review of studies examining military or veteran samples found that alcohol use, depressive symptoms, antisocial traits, family of origin violence, and poor relationship adjustment and satisfaction were associated with the presence of IPV. However, no effect sizes or statistical information were presented in this review. Summarizing this information, it seems as if risk factors associated with IPV in military samples generally do not differ from those associated with civilian samples, but there is some evidence that military men may engage in more severe forms of violence.

A few correlates of child maltreatment among military families have also been reported in the literature; however, many studies report rates or frequencies only

(Dubanoski & McIntosh, 1984; McCarroll et al., 1999; Mollerstrom et al., 1992; Mollerstrom et al., 1995). Merrill and colleagues (2004) found Navy recruits that reported lower socioeconomic status were more likely to be at risk for child physical abuse perpetration (OR=1.80). Furthermore, results of logistic regression analyses indicated that dysphoria (OR=4.88), posttraumatic stress (OR=3.32) and self-dysfunction (OR=7.31) predicted child maltreatment risk. In these analyses, alcohol problems were not significant as a predictor for child maltreatment risk (Merrill et al., 2004). Again, some risk factors appear to be consistent with those found in civilian samples, but more statistically rigorous studies are needed.

Finally, with respect to risk for child maltreatment, there is also some indication that risk factors for military mothers and fathers are different. Schaeffer, Alexander, Bethke and Kretz (2004) examined an Army sample of mothers and fathers who were enrolled in a home visitation program and analyzed a parent's risk for child abuse across measures of depression, family functioning, marital satisfaction, parenting stress, and social support. Results indicated that depression, parenting stress, and family conflict were shared risk factors for both mothers and fathers. Marital dissatisfaction, low social support and low family cohesion were risk factors for mothers only, while low family expressiveness was the only unique risk factor for fathers. Thus, although risk factors are not apparently different among military and civilian parents, there is some indication that risk factors differ for military mothers and fathers.

Military studies also tend to confirm the relationship between child maltreatment and IPV. Army families with a substantiated case of IPV were found to be twice as likely to have a substantiated case of child maltreatment (Rumm et al., 2000). Sixty percent of the child emotional abuse cases in a two-year Army study involved the child witnessing

domestic violence in the home (Jellen, McCarroll, & Thayer, 2001). Merrill and others (2004) found that Navy recruits had a significant positive association between a history of IPV and child physical abuse perpetration risk after controlling for potential confounders ($\chi^2_{partial}$ =(1, N = 1367) =11.30, $p < .001$). Therefore, family violence screening measures developed for use with a military population should include risk factors associated with both IPV and child maltreatment.

Family Violence Risk Screeners

A variety of instruments have been developed to screen for or to measure some construct related to family violence. It is important to note that within the context of screener development, risk factors may be used for three purposes: identification of past abuse, identification of concurrent abuse, or prediction of future risk. Since this paper focuses on the use of screening measures for the prevention of family violence among military families, screening measures used for predictive purposes are of primary interest. However, much of the family violence screener development to date has focused on past or concurrent risk. Therefore, the purpose of the screening measure will be central to this review. Screening measures which met the following criteria were reviewed:

1) Screeners must have been developed as an independent measure for either child maltreatment, IPV, or both;
2) The screener must have been utilized within the context of a family violence prevention program;
3) The screener must have been subjected to at least two or more published research studies with different samples; and,

4) At least two of the research studies must include outcome measures in order to assess the effectiveness of the screener in practice with the target population.

Search results revealed five measures used to screen for IPV and three measures used to screen for child maltreatment. In addition, two measures of IPV and one measure of child maltreatment were identified which are not screening measures per se, but are widely-used as comparison measures in validity studies and considered important measures to review.

Widely-used IPV Measures

Two measures have been developed and are commonly used to quantify past or concurrent levels of family violence: the Conflict Tactics Scale (CTS) and the Index of Spouse Abuse (ISA). These measures are also often used as the basis of comparison to evaluate the validity of IPV screeners.

Conflict Tactics Scale. Straus' (1979) CTS is one of the most widely used self-report instruments to measure the extent to which partners in a dating, cohabitating, or marital relationship engage in physical assault (Straus, 1979). The measure originally consisted of 19 items identifying behaviors that might be taken during a conflict with a partner, scored from 0 (never) to 6 (more than 20 times). Four subscales were determined from these questions: 1) reasoning, 2) psychological aggression, 3) mild physical assault, or 4) severe physical assault. The CTS was later refined into the CTS2 as an improved measure, adding two additional subscales: Sexual Coercion by Partner and Physical Injury Resulting from Assaults.

Some researchers have identified the CTS as the most thoroughly validated measure of IPV (Connelly, Newton, Landsverk, & Aarons, 2000). Specifically, there is evidence that the CTS has a stable factor structure, moderate reliability, strong construct

validity, and moderate concurrent validity (DeVoe & Kantor, 2002). The CTS2 has similar properties: high internal consistency and preliminary adequate evidence of construct validity and discriminate validity (DeVoe & Kantor, 2002). The CTS has been used to measure IPV in many different studies (for example, Houry et al., 2004; Feldhaus et al., 1997; Sherin, Sinacore, Li, Zitter & Shakil, 1998; Eckenrode et al., 2000; Peralta & Fleming, 2003; Connelly et al., 2000).

Index of Spouse Abuse (ISA). The ISA was found to have stable factor structure, high internal consistency reliability, excellent discriminate validity and strong support for construct validity (Hudson & McIntosh, 1981). Furthermore, with the established cut-scores, the ISA correctly classified 90.7% of the sample, with a false positive rate of only 9.3% (Hudson & McIntosh, 1981). The ISA has been used to measure IPV in many different studies (for example, Chen, Rovi, Vega, Jacobs & Johnson, 2005; Coker, Bethea, Smith, Fadden & Brandt, 2002; Feldhaus et al., 1997; Smith, Earp, & DeVellis, 1995). Although this measure is a widely used standard for comparison with other IPV screeners, additional research confirming the initial validation studies could not be located. Although both the CTS and ISA have been used extensively to measure concurrent or past IPV, they have not been developed for use as a screening measure for risk of future family violence.

Other Screening Measures for IPV

Fogarty, Burge, and McCord (2002) identified seven screening measures for IPV used by medical professionals. Two of these measures, the Partner Abuse Inventory (Pan, Ehrensaft, Heyman, O'Leary, & Schwartz, 1997) and the Relationship Chart (Wasson, Jette, Anderson, Nelson, & Kilo, 2000), did not have two or more published studies using different samples with outcome measures and were therefore eliminated

from this review. A thorough search of databases and review of bibliographies identified the following measures which also did not meet the selection criteria: Wife Abuse Inventory (Lewis, 1985), Proximal Antecedents to Violent Episodes (PAVE) Scale (Babcock, Costa, Green & Eckhardt, 2004), the "STaT" measure (Paranjape & Liebschutz, 2003), the Universal Violence Screening Protocol (Heron, Thompson, Jackson, & Kaslow, 2003; Dutton, Mitchell, & Haywood, 1996) and other unnamed or single question screens (Abbott, Johnson, Kozoil-McLain, & Lowenstein, 1995; Siegel, Hill, Henderson, Ernst & Boat, 1999; Berk, He & Sorenson, 2005; Peralta & Fleming, 2003). Therefore, five remaining screening measures met the review criteria, all of which seek to identify past or concurrent IPV in a medical setting. None of the screening measures reviewed were developed to screen individuals or families for future risk of IPV.

Partner Violence Screen (PVS). Feldhaus and colleagues (1997) devised a brief, three-question screening instrument to detect concurrent IPV for use in the emergency department. A positive response to any of the three questions indicates IPV. Researchers compared the PVS to the CTS and the ISA in a large sample of women who presented to 48 randomly selected 4-hour time blocks in two urban emergency departments (Feldhaus et al., 1997). Using the ISA as the standard for comparison, the PVS had a sensitivity of 64.5% and a specificity of 80.3%. Using the CTS as the standard, the PVS had a sensitivity of 71.4% and a specificity of 84.4%. The single question related to past experiences with physical violence was more sensitive and specific than the other questions regarding perceptions of safety, detecting almost as many abused patients as the PVS overall (Feldhaus et al., 1997).

Three additional studies examined the PVS in a medical setting. Holtrop, Fischer, Gray, Barry, Bryant, and Du (2004) evaluated whether use of the PVS increased detection of IPV in a busy pediatric clinic. They compared positive screening results received during a 12-month period of time with the number of referrals for IPV received by the social work department during a previous 12-month period of time. Implementation of the PVS produced 164 referrals to social work services, as compared with only nine referrals during the previous 12 months. Each of these 164 referrals were assessed by social work staff and it was determined that 91.5% of the referrals were actual cases of IPV. Researchers referred to this percentage as the positive predictive value (PPV) of the measure. No additional measures of IPV were included in this study, therefore, the sensitivity and specificity of the PVS could not be calculated. Hence, these results should be taken with caution.

Another study slightly re-worded the first two questions of the PVS and included a third new question related to police involvement in the home (Kozoil-McLain, Coates, & Lowenstein, 2001). This prospective cohort study found that women who screen positive in the emergency department were 46.5 times more likely to report severe physical violence four months later. Sensitivity of the measure to detect physical violence was reported as 80% and specificity was 93%. Sensitivity rates were much lower, however, for verbal aggression and sexual coercion (Kozoil-McLain et al., 2001).

A third cohort study of 215 adult women who presented to an inner-city emergency department provided an opportunity to assess the predictive capacity of the PVS (Houry et al., 2004). Four months after screening, 96 women (44.7%) participated in a telephone interview and rates of IPV were measured using the CTS. Women with a positive PVS were significantly less likely to participate in the follow-up telephone

interview. Despite this, women who screened positive on the PVS were 11.3 times more likely to experience physical violence and 7.3 times more likely to experience verbal aggression compared with those who had a negative PVS screen. Once again, the single question related to physical violence was, by itself, a strong predictor of future violence with positive responses indicating a 10.9 times greater likelihood of physical violence and a 7.06 times greater likelihood of verbal aggression (Houry et al., 2004).

Taken as a whole, these studies provide promising evidence of the predictive capacity of the PVS, especially the single question related to physical abuse. In essence, these data indicate that women who are currently experiencing physical violence are likely going to experience it again in the future. However, it does not provide any information about the risk status of those women who have not yet experienced physical abuse.

Woman Abuse Screening Tool (WAST). The WAST was developed for use by family physicians to aid in the identification and assessment of women who are victims of IPV (Brown, Lent, Brett, Sas & Peterson, 1996). Seven questions make up the WAST, with a choice of three possible ranked responses (two positive and one negative). With a possible range of 0 to 21, developers of the measure did not conduct a cut-score analysis. Instead, they focused on the ability of the first two items to detect IPV (called the short-WAST). The short-WAST was examined in two studies (Brown, Lent, Schmidt, & Sas, 2000; Halpern, Susarla & Dodson, 2005). With a possible range of scores from 0 to 2, a positive response on either of the two questions indicates a positive screen for IPV.

Internal consistency reliabilities for the WAST have been reported to range from .75 to .95 (Brown et al., 1996; Brown, et al., 2000; Fogarty & Brown, 2002). The

measure has also showed evidence of construct and discriminate validity, correlating highly with the Abuse Risk Inventory (ARI; Brown et al., 1996; Brown et al., 2000). However, no published validation studies for the ARI could be identified; therefore, the ARI is not included in this paper and this validation study is thereby weakened as a result.

Regarding the short-WAST, one study failed to report psychometric properties specific to the shortened version of this scale (Brown et al., 2000). Halpern and colleagues (2005) did, however, compare the short-WAST to the PVS in a large cross-sectional study in which women visiting the emergency department with physical injuries were randomly assigned to complete either the short-WAST or the PVS. Injury location (whether the woman's injury was on her head, neck or facial area) was included as an additional IPV marker. Screening results and injury location were compared against self-reports of injury etiology (whether the injury was the result of IPV or some other mechanism). The short-WAST did not perform as well as the PVS in terms of sensitivity and specificity (sensitivity: .58 vs. .92; specificity: .49 vs. .56). Results indicated the short-WAST, alone or in combination with injury location, was not statistically associated with self-reported injury etiology (Halpern et al., 2005). Again, this instrument does not provide any information about the risk status of those who have not yet experienced IPV.

Women's Experience with Battering (WEB) Scale. Smith and colleagues sought to alleviate the perceived limitations of the CTS (the ability to capture only the behavioral features of battering) by developing the Women's Experience with Battering (WEB) Scale (Smith et al., 1995). This measure uses what is termed a "continuous process approach" and quantifies the psychological vulnerability of a respondent created by sustained exposure to physical assault and psychological degradation (Smith et al.,

1995). With a total of 10 items, respondents rate each item on a scale of 1 (Agree Strongly) to 6 (Disagree Strongly) giving a potential range of scores of 10 to 60. A score of greater than 20 is indicative of "battered" (Coker et al., 2002), however, no published analyses of cut-scores could be located for this measure. Development and validation of the WEB Scale using a known-groups approach on a sample of 185 known battered women and 204 known non-battered women revealed high internal consistency reliability ($\alpha = .99$) and good construct validity. In addition, the WEB Scale was highly correlated with the CTS ($r = .71$, $p<.001$) and the ISA ($r = .88$, $p<.001$), providing support for the scale's convergent validity. Correlations with related measures of IPV were moderate, as expected. One interesting note is that the WEB Scale was also significantly negatively correlated with marital satisfaction ($r = -.80$, $p<.001$). Additionally, although the authors suggest that the WEB Scale is not likely to misclassify individuals, they failed to report the sensitivity or specificity of the measure and thus provide no data to support this claim.

The only other known WEB Scale study that provisionally met the criteria established for this review was conducted on a sample of women in the primary care setting of two large medical facilities (Coker et al., 2002). The primary objective of this study was to determine whether IPV screening using a modified ISA and the WEB Scale improved documentation of IPV in the patient's medical records. As compared with the modified ISA, higher WEB scores were associated with a slightly increased likelihood of IPV documentation. However, the authors admit that this modified ISA scale has not been previously validated. Additionally, sensitivity and specificity of the WEB Scale was not reported in this study. The authors did find that 37.5% of the sample scored positive for battering on the WEB Scale, but did not report any other physical or sexual IPV (as

measured by the modified ISA). They concluded that these women were "psychologically battered" (Coker et al., 2002). However, the authors fail to account for the possibility of misclassification by either the WEB Scale, the modified ISA or by both measures. This is especially important given the previous finding that there is a strong negative correlation between marital satisfaction and the WEB Scale. It would be possible that some women who screened positive for battering were in fact experiencing only marital dissatisfaction. Without additional research to address this issue, the validity of the WEB Scale remains in question.

HITS. A short screening mnemonic was developed for use in the family practice setting to identify concurrent IPV based on four common behaviors: hurts, insults, threatens and screams (Sherin et al., 1998). Responses are recorded on a 5-point scale (never, rarely, sometimes, fairly often, frequently) with possible scores ranging from 4 to 20. Cut-score analyses indicated that a score greater than 10 classifies someone as victimized (Sherin et al., 1998). Initial validation studies based on known-groups comparisons revealed good internal consistency ($\alpha = .80$), construct validity, and evidence of concurrent validity as compared with responses on the CTS.

The only other study examined the HITS screen compared with the ISA and WAST in a sample of 202 Hispanic women seen in a family practice setting (Chen et al., 2005). Results revealed acceptable reliabilities ($\alpha = .76$ for the English version and $\alpha = .61$ for the Spanish version) with support for construct validity and concurrent validity. Using a cut-score of 10, the English version of the HITS had a sensitivity of 86% and a specificity of 99%. However, the Spanish version of the scale had much lower sensitivity with this cut-score (22%). The authors recommended a cut-score of 5.5 for the Spanish version of the HITS (Chen et al., 2005). Although preliminary analyses are promising,

additional studies of the HITS scale will be important to further validate these findings. The brevity and simplicity of this measure is particularly appealing. However, this measure does not identify families who may be at risk for future IPV, as the questions are focused on past or current behaviors.

Abuse Assessment Scale (AAS). Developed by the Nursing Research Consortium on Violence and Abuse, the AAS uses five questions to assess remote and recent history (within the past year) of IPV among pregnant women. Women who answer with a positive response to the first three questions on the AAS are considered abused. McFarlane, Parker, Soeken and Bullock (1992) compared the AAS to both the ISA and the CTS in a sample of 691 ethnically diverse pregnant women. Women who screened positive on the three scored questions also scored significantly higher on the ISA, providing support for the construct and criterion validities of the measure. No analyses comparing the AAS to the CTS were reported. The researchers conclude the AAS demonstrated “specific identification of abuse” and summarize the measure is valid and reliable (McFarlane et al., 1992, p.3177). However, they failed to report the reliability, sensitivity or specificity of the AAS in this article.

Ernst, Weiss, Cham and Marquez (2002) sought to expand on the validity of the AAS by evaluating whether the AAS could detect concurrent (ongoing) IPV. The AAS was compared against the Ongoing Abuse Screen (OAS; developed by the researchers) which transformed the AAS questions into present tense forms, and against a single question directly asking about ongoing IPV. Results indicated that with a sensitivity of only 30%, the AAS is not as useful if the purpose of the screening is to determine ongoing IPV for preventative reasons. In the most recent attempt at validation of the AAS, it was compared against the CTS2 to detect physical abuse among pregnant

women. Results were again disappointing with the AAS only correctly identifying 31.7% of both minor and severe physical abuse cases. The authors recommend the AAS is not used as an independent screener for IPV until further modifications and analyses can be conducted. Only two additional studies were identified, but neither provided any outcome measures to compare against the AAS so they are not reviewed here (Norton, Peipert, Zierler, Lima & Hume, 1995; Greenberg, McFarlane, & Watson, 1997).

In summary, IPV screeners currently available have been developed for use in the medical setting for the identification of past or concurrent IPV. There were no measures developed for use in other family violence prevention settings or for the identification of future risk of family violence. In general, of the studies that did include outcome measures, several of these studies failed to report even the most basic psychometric properties of the screening measures. This makes validation of screening measures extremely difficult. Of the screening measures reviewed, the PVS and the HITS reported the highest rates of sensitivity and specificity, although only one study has validated the HITS to date. The other measures have not been adequately validated. Additional studies are needed on the WEB, WAST and AAS before these measures may be considered adequate for the identification of IPV. None of the measures reviewed are validated for the identification of future risk of IPV.

Widely-used Child Maltreatment Risk Measure

Child Abuse Potential Inventory. The Child Abuse Potential Inventory (CAPI) was developed to measure a parent's potential for physical child abuse and has been frequently used to evaluate the validity of child maltreatment screeners. It was originally developed for use by protective services workers and was designed to be used in conjunction with other sources of data collection such as interviews, case histories, or

observations (Milner, 1986). Milner and Wimberley (1980) identified seven factors that appeared descriptive of physical child abusers (distress, rigidity, child with problems, problems from family and others, unhappiness, loneliness, and negative concept of child and self). Based on self-reports, the CAPI measures attitudes and experiences of parenting based on these seven factors. Respondents answer the 160 items with an "agree" or "disagree" response. The measure contains a 77-item abuse scale, an 18-item lie scale and other filler items (Milner, 1986). Scoring is based on a beta-regression weighted procedure that allows a possible range of scores from 0 to 486 with a recommended conservative cut-score of 215 currently used in practice.

The CAPI has been shown to have good internal consistency and high reliability (Burrell, Thompson & Sexton, 1992; Rinehart et al., 2005). Based on the current cut-score and known group comparisons, the correct classification rate was reported as 96.7% for female abusers and 86.6% for women in the general parent population (Caliso & Milner, 1992). While substantial support for construct validity and discriminate validity has accumulated (see Robertson & Milner, 1983; Milner & Gold, 1986; Holden, Willis, & Foltz, 1989; Caliso & Milner, 1992; Rinehart et al., 2005), support for the predictive validity of the CAPI to identify potential for future abuse is much sketchier. Milner, Gold, Ayoub and Jacewitz (1984) reported a moderate, positive relation between abuse scores and subsequent reports of abuse ($\omega^2 = .32$). Likewise, Chaffin and Valle (2003) did find relationship between pre-intervention CAPI scores and future abuse, however, changes in CAPI scores from pre- to post-intervention failed to produce a significant change in the likelihood of future abuse. The CAPI is not suited for use as a screening measure, however, due to the length of the measure and complexity of scoring.

A thorough search of existing databases and review of bibliographies identified several measures used to identify families at risk for child maltreatment. The Brigid Collins Risk Screener (Weberling, Forgays, Crain-Thoreson & Hyman, 2003), the Parenting Risk Assessment (Murry, Baker, & Lewin, 2000), the Maternal Characteristics Scale (Polansky & Gaudin, 1992), and the Dysfunctional Parenting Scale (Larson, Collet, & Hanley, 1987) were identified as screeners that had been implemented in a prevention setting, however, these measures did not meet the criteria of having more than one published study with an outcome measure. This left only three measures that met the review criteria. All of these screeners seek to identify a family's risk for the future occurrence of parenting problems or child maltreatment.

Family Stress Checklist (FSC). Originally named the Carroll-Schmitt Parenting Checklist, the FSC was developed to identify possible indicators of abuse and neglect (Orkow, 1985). Comprised of 10 items, it was originally scored as 0 (no risk), 5 (risk), or 10 (high risk; Orkow, 1985). The initial validation study showed the scale had a sensitivity of 80% and a specificity of 89% (Murphy, Orkow, & Nicola, 1985).

The measure was later called the Kempe FSC and has been used to screen families for Healthy Families of America and Hawaii's Head Start program, home-based child maltreatment prevention programs (Duggan et al., 2000; McGuigan, Katzev, & Pratt, 2003). Acceptance into these prevention programs is based on a two-step process. In step one, medical records are screened for potential demographic risk indicators. Risk indicators can either be present, not present or unknown to the reviewer. An unmarried mother, a mother who received inadequate prenatal care, or a mother who unsuccessfully sought to have an abortion are automatically considered at risk. Additionally, the presence of two or more risk indicators or seven or more unknown risk

indicators are also considered to be a positive risk screen. In step two, the family is interviewed using the FSC. A scoring procedure similar to the original procedure is used with the following possible responses: 0 (no problem), 5 (mild problem), or 10 (severe problem). If either parent receives a total summed score of greater than or equal to 25, the family is considered at risk for child maltreatment and is enrolled into the prevention program.

Korfmacher (2000) conducted a comprehensive review of the psychometric data available on the FSC. He found that scaling of the instrument was inconsistent and the cut-scores varied widely, making study comparisons difficult. He also found a lack of reliability studies, especially inter-rater reliability, and thought this to be especially problematic considering the measure is administered by interview. Korfmacher (2000) also recommended that the FSC is used with other assessment tools, due to a high false positive rate. McGuigan and Pratt (2001) found evidence that mothers' scores on the FSC shortly after birth continued to be a significant predictor of three types of child maltreatment (physical abuse, psychological abuse and neglect) up to the child's fifth birthday. Therefore, although there is evidence of the FSC's predictive validity, it appears as if additional research is still needed to validate initial reliability findings with multiple and single interviewers.

Additionally, the two-step review process is problematic for prevention programs that are seeking to screen a large number of families relatively quickly. Medical records may not be available, depending on the setting of the program and staffing resources may not permit interviewing a large number of families. If the two-step process is altered for other programs, additional reliability and validity studies would need to be conducted. None of the studies reviewed reported sensitivity or specificity rates for this measure.

For these reasons, the FSC may not be a particularly helpful screener for most prevention programs seeking to screen for child maltreatment risk.

Parenting Profile Assessment (PPA). The PPA was developed by Anderson (1987) to assist in the prediction of parenting problems leading to possible physical abuse. The screener is administered through interview with parents and through observation of parenting practices. In addition, the presence or absence of five individual variables is determined (low income, less than high school education, family involvement with police, presence of harsh discipline practices, or moderate to severe life change). The PPA is scored through a weighted process, with scores greater than 21 and the presence of all five individual variables indicative of risk for child abuse (Anderson, 1993).

Internal consistency reliabilities have been reported at .74 and .75 (Anderson, 1993; Anderson, 2000). Sensitivity and specificity of the PPA was reported to be 95.8% and 98.6% for one sample (Anderson, 2000) and 75% and 85.6% for another sample (Anderson, 1993). One potential problem with these data is the standard used for comparison. In the 1993 study, Anderson used self-reports of a nationwide sample of nurses to calculate actual abuse, which has questionable accuracy. In the 2000 study, known groups of abusive and non-abusive mothers were used as the basis for comparison. The latter is a better comparison, but the study failed to report the details of the data collection in the text, and included only a summary. Therefore, a critical analysis of these results is problematic.

Aside from the psychometric issues, there is also one practical problem with using this measure as a screener for family violence. This measure relies on observational data collection. This is not particularly feasible for a prevention program

that seeks to screen a large number of families quickly. Because this measure does not meet the criteria for a brief and simple measure, it does not appear it would be particularly useful.

Michigan Screening Profile of Parenting (MSPP). Helfer, Hoffmeister, and Schneider (1978) developed the MSPP as an instrument to identify parents who are more likely to experience parenting problems that may lead to child abuse or neglect. The self-administered measure consists of five sections: Emotional Needs Met, Relationship with Parents, Expectations of Children, Dealing with Others and Coping. Respondents answer the four sections which contain one to 30 questions on a seven-point scale for a total of 50 items.

The MSPP was initially determined to have a sensitivity and specificity of 83% and 77%, respectively (Anderson, 1993) with internal consistency reliabilities ranging from .62 to .85 (as reported in Gaines, Sandgrund, Green & Power, 1978). Subsequent validation of the measure has shown evidence of content validity (Gaines et al., 1978; Spinetta, 1978). Several studies used known-groups analyses to provide some evidence of this measure's ability to discriminate among mothers who have abused their children, mothers who have neglected their children, and non-abusive or non-neglectful mothers (Gaines et al., 1978; Spinetta, 1978). However, even with the inclusion of several other related variables, the MSPP was only able to account for 12% of the variance in differences among abusers, neglecters, and non-abusers (Gaines et al., 1978). Furthermore, validation of the MSPP has virtually halted since the late 1970's, with no attempts to replicate sensitivity or specificity, or to establish the predictive capacity of the measure.

In summary, two of the child maltreatment screeners reviewed require observational or interview techniques as part of the screening process. This procedure hinders the ability of these measures to screen a large number of families for risk of child maltreatment in an efficient manner. Although these measures may have promising preliminary psychometric properties, additional studies are needed to further validate these findings and address the limitations as discussed previously. The MSPP also shows promise as an effective screener for child maltreatment risk, but further validation of the measure is required, especially in terms of the ability of the measure to predict risk.

Summary of Family Violence Screeners

Although a few of the existing IPV measures are brief, simple, and have sensitivity rates above 80%, they do not address the need to predict future occurrence of violence. Their complexity and reliance on observation or interview data collection hinders their efficiency in the prevention setting. Moreover, none of the measures reviewed thus far screen for IPV at the same time as child maltreatment. There is growing evidence that concurrent screening for both types of violence may be important for family violence prevention. The military has pioneered concurrent screening for IPV and child maltreatment for entry into a home-based, family violence prevention program. The military model for screening for family violence is theoretically, administratively and empirically supported in the literature. The next section of this paper will review the available support for the concurrent screening of both IPV and child maltreatment and will examine the AF Family Needs Screener in detail.

CHAPTER III – CONCURRENT SCREENING FOR IPV AND CHILD MALTREATMENT RISK

Support for Concurrent Screening

There is evidence to support the concurrent screening for IPV and child maltreatment within the military setting. There are four factors that provide support for this position: 1) administrative and programmatic structure of the military, 2) theoretical justification, 3) established co-occurrence of these two types of abuse, and 4) reduced treatment effect of home visitation programs when co-occurrence is identified. Each of these four factors will be discussed in detail.

First, according to current military directives, the AF Family Advocacy Program (FAP) is mandated to address both types of family violence. From an administrative standpoint, allegations of child abuse and IPV are managed in the same manner within the FAP. Regardless of the type of maltreatment, allegations are logged into the AF Central Registry, family members are assessed and interviewed, and allegations are determined by a committee who reviews the incident information. Because both types of maltreatment are considered important sociological problems for military families, screening for IPV and child maltreatment simultaneously makes programmatic sense.

The second reason for concurrent screening for both types of abuse is supported by theory. Early theoretical development on the causes of family violence identified 15 theories which were organized into three broad categories: intraindividual theory, sociocultural theory and social-psychological theory (Gelles & Straus, 1979). Intraindividual theories incorporate the role of individual traits and their impact on family violence. Examples of individual traits linked to family violence include: alcohol and drug

abuse, poor self-esteem, the presence of psychological disorders such as depression or post-traumatic stress disorder, and self-dysfunction (Bugental & Happaney, 2004; Kotch et al., 1999; Merrill, Crouch, Thomsen, & Guimond, 2004; and Pan et al., 1994). All of these factors have been shown to have a positive correlation with family violence.

The second category of theory, sociocultural theories, generally focuses on the influence of social class, education and income level on family violence. Research has found younger, unpartnered, less educated and lower income African American mothers appeared to be more at risk for child abuse (see Combs-Orme, Martin, Fox, & Faver, 2000; Combs-Orme, Cain, & Wilson, 2004; Kotch et al., 1999; Murphey & Braner, 2000). Sociocultural theories also hypothesize the role of the abused woman's loss of power and control by a male-dominated culture as the cause for family violence (see Bugental & Happaney, 2004).

Finally, social-psychological theories are the last grouping of theories related to family violence. Here, the environment and social variables shape one's experience with family violence. Social learning theory is one example of a social-psychological approach to understanding family violence. Bandera (1973) hypothesized that human behavior is learned from observing others through modeling of that behavior. This theoretical approach assumes that risk of family violence is related to previous exposure to family violence and exposure to aggressive or violent behaviors. More recently, all three of these theoretical categories have been combined. Gelles and Straus' work has been adapted and an ecological approach is commonly used in research to explain the relationship between individual, family, community and societal factors related to family violence (Little & Kantor, 2002). Hence, it is clear that both IPV and child maltreatment share common theoretical bases.

The third justification for the concurrent screening of IPV and child maltreatment in the military setting relates to empirical support for high rates of co-occurrence between these two types of abuse. One longitudinal study of families enrolled in an early intervention program found that domestic violence occurred in 38% of the families with confirmed child maltreatment, and that domestic violence preceded child maltreatment in 78% of these cases (McGuigan & Pratt, 2001). Another study found that among female caretakers of children receiving child abuse forensic interviews at a child advocacy center, 67% of the caretakers reported a history of emotional abuse, 64% reported physical abuse, and 47% reported sexual abuse (Pulido & Gupta, 2002). Families with a history of domestic violence in the first six months of an infant's life were three times as likely to have a confirmed incident of child physical maltreatment by the time that child was age five (McGuigan & Pratt, 2001).

Although the longitudinal study cited above provides an indication that IPV may precede child maltreatment, much of the research in this area cannot determine causality. Certainly, it is apparent that IPV and child maltreatment share many common risk factors. However, one cannot rule out that for certain populations, child maltreatment may lead to IPV. In any case, the co-occurrence of the two provides support for concurrent screening of both.

Finally, there is recent evidence that the effectiveness of home visitation services in the prevention of child maltreatment may be limited if there is domestic violence in the home (Eckenrode et al., 2000). Researchers conducted a fifteen-year follow-up study of a randomized control trial and found a statistically significant interaction effect for home visitation services and the presence of domestic violence on the prevention of child maltreatment. Specifically, the positive treatment effect of home visitation services

decreased as the number of domestic violence incidences in the home increased (Eckenrode et al., 2000). The authors suggest that modifications are needed in order to strengthen the impact of home visitation programs when domestic violence is present. Given this finding, it would be important for home visitation programs to screen for risk of IPV along with the risk of child maltreatment so that prevention services might be tailored depending on the family context.

To sum, there are several factors that support simultaneous screening risk of child maltreatment and IPV. Current military policies and administrative processes, current theoretical approaches, evidence of interdependence, and evidence of a significant interaction effect between home visitation services, domestic violence, and the prevention of child maltreatment provide theoretical and practical support for screening for family maltreatment as a whole. First, there are practical reasons to treat families in a holistic manner, identifying both IPV and child maltreatment risk simultaneously. Scare resources require efficiency in screening techniques. If a brief screening measure can consistently and accurately identify families at risk for either type of violence, many hours of staff time spent interviewing families, administering and scoring measures, or reviewing case records might be saved.

Secondly, there is recent evidence that the efficacy of home-visitation programs on the prevention of child maltreatment may be attenuated if IPV is present in the home. From a clinical perspective, families experiencing multiple types of violence must be accurately identified so that the most effective treatment services can be provided. Therefore, the development of an accurate screener that simultaneously screens for both IPV and child abuse would be a significant contribution to the field.

The military provides one prevention model that incorporates concomitant screening for both IPV and child maltreatment. The rationale for this model will be reviewed and a critical analysis of the military's family violence screening measure will be conducted. The only known measure developed to screen for risk of both child maltreatment and IPV is the AF Family Needs Screener. Information pertaining to the development of this screener will be reviewed next and research studies involving the screener will be critically analyzed.

The Air Force Family Needs Screener

Background Information on the AF Family Needs Screener

As discussed previously, the Department of Defense (DOD) has recognized the importance of prevention programs such as home visitation programs and has supported the implementation of such programs. The NPSP provides home-based nursing and social work interventions to pregnant parents or parents with young children who are at risk of family violence. The AF Family Needs Screener (FNS) was developed with three uses in mind: (1) to assist the NPSP staff in making classification decisions about the allocation of services based on family needs, (2) to provide a means to better assess, plan, and conduct clinical interventions for the NPSP families, and (3) to provide a more systematic means of assessing family well-being at program entry (Kantor & Straus, 1999). The FNS was not developed to comprehensively assess all of the characteristics associated with multiple types of maltreatment, but was developed to capture those major characteristics of families that are particularly associated with physical assaults on parents or young children, with the assumption that many of these characteristics are associated with other types of maltreatment as well (Kantor & Straus).

Development of the FNS

Kantor and Straus (1999) developed and tested the FNS for use with mothers. They cite two primary reasons for excluding fathers as respondents. Mothers were chosen because they are most often the primary recipient of NPSP services and are the primary caretakers of children in most families (Kantor & Straus). In addition, the authors state that mothers were found to be more accessible and more willing to complete assessment instruments as compared to fathers (Kantor & Straus). Currently, fathers may complete the FNS, but their responses are only used in the assessment of the family in order to develop a plan for treatment services and not to make classification decisions.

The FNS was developed using an integrated framework of 15 different theories of family violence, organized under three broad subheadings: intraindividual theories, social-psychological theories, and socio-cultural theories (Bersani & Chen, 1988; Kantor & Straus, 1999; Gelles & Straus, 1979). Table 1 (all tables and figures are located in the Appendix) summarizes each of these theories and the risk markers identified for that theory which were included in the FNS. Based on the items included in the FNS, three of the fifteen mentioned theories do not appear to be represented by the FNS items. These are: symbolic interaction theory, functional theory of violence, and general systems theory. Although systems theory may provide an overarching explanation for the relationship between risk markers and different types of violence, it does not explain a specific risk marker, per se. The other twelve theories do appear to be represented among the FNS items.

Based on these theories, the FNS is divided into ten different subscales: demographic characteristics, stress, relationship discord, support, substance abuse,

violence approval, family of origin violence and neglect, self-esteem, depression and prior family violence (Kantor & Straus, 1999). The FNS is comprised of 58 items (53 scored items including plus five additional non-scored items; see Table 2). The FNS subscales most strongly supported by theory are: demographic characteristics, family of origin abuse and neglect, stress, relationship discord, violence approval, depression and prior family violence. Based on a review of Gelles and Straus' (1979) summary, the theoretical basis for substance abuse, support and self-esteem as risk markers for family violence are somewhat weaker. However, substance abuse, support, and self-esteem are all supported empirically by research as risk factors for family violence (as discussed in Chapter 2).

Administration of the FNS

The FNS is administered to pregnant women or women with children age three or younger who wish to receive NPSP services. Many women are screened during obstetrical orientation, without regard to any other risk factors. Other women are referred to the NPSP and complete a FNS if they wish to receive NPSP services. Two categorizations are made for families who complete the FNS: Low needs or high needs. Low needs families are offered one home visit by a nurse and are provided with referral services as indicated. High needs families are assigned a nurse as a case manager and are provided with intensive, in-home nursing and social work services. The nature of the services provided is dependent upon the individualized needs of the family and the services plan that is developed by the case manager and approved by the family. The overall assumption that drives the allocation of services is that the score on the FNS, coupled with an assessment by a clinical staff member, would determine the risk level of the family and thereby guide the implementation of services.

There are essentially three ways a family is categorized as high needs. First, based on the mother's responses, if a family scores at or higher than the established cut-score of 9, the family is categorized as high needs. Second, even if a mother scores below the cut-point, a family may be clinically assessed by NPSP staff as high needs. Third, a family will be categorized as high needs if the mother responds positively to one of the five automatic high needs questions which will be discussed in the next section.

Scoring of the FNS

Cut-score analyses were conducted by Kantor and Straus (1999) to determine the most appropriate score with which to make classification decisions. An examination of the distribution of the scores from one sample indicated the 75th percentile was at a cut-score of 10 (77.8% precisely). A score of 9 included 71.7% of the cumulative frequency of the sample, and the researchers decided to err on the side of being more inclusive rather than less inclusive. Therefore, the cut-score of 9 was tested on another sample and the final decision to set the cut point at 9 was supported in this second sample (Kantor & Straus). This is the cut-score currently used in practice.

The scoring of the FNS is somewhat complex, but is simplified by the use of computerized scoring in the field. Appendix A shows the FNS in its entirety, with a sample score sheet shown in Appendix B. Omitting the five non-scored items, responses for the other 53 items are dichotomized as either not at risk (0) or at risk (1) on the scoring sheet. For example, question #21 reads, "This is a very stressful time for me." Respondents choose from one of four possible answers: (1) strongly disagree, (2) disagree, (3) agree, or (4) strongly agree. For this question, respondents who agree or strongly agree with this question are considered "at risk" and would receive one point towards the total score on the FNS. Respondents who disagreed or strongly disagreed

to this item would receive a zero toward the total score on the FNS. All items positively scored as “at risk” are added to produce a composite score, ranging from 0 to 53. The scoring sheet at Appendix B indicates which responses would warrant an “at risk” score (or score of 1) for each item.

Based on the cut-score analyses as discussed previously, a composite score of 9 or greater indicates the family is at “high risk” for family violence. In an effort to minimize any potentially negative stigmatization for respondents, terminology was changed from “high risk” to “high needs” (Kantor & Straus, 1999). In addition, five items are deemed “automatic high needs” items. A positive response on any of the following five items will automatically categorize a family as “high needs” regardless of the total composite score: feeling out of control, uncontrolled anger, suicidal thoughts, or a past history of child or partner abuse (Kantor & Straus).

An additional complication in scoring is the omission of some items depending on the respondent’s pregnancy or partner status. There are 43 common items (k) that are completed by every respondent (7 common demographic items plus an additional 36 common items on the other subscales). Table 3 details how many additional items are completed if a respondent is pregnant and/or has a partner. Hence, the range of possible scores on the FNS is different depending on the status of the respondent. Only those respondents who are both pregnant and have a partner complete all 53 scored items. The cut-score analyses conducted on the FNS (Kantor & Straus, 1999) did not take into account the variability in the number of items completed depending on whether the respondent is pregnant or has a partner. Therefore, it is unknown whether the cut-score should be different for a respondent who is not pregnant or who does not have a partner. This would be an important contribution to future research.

Review of FNS research

There are only two studies that have been conducted evaluating the psychometric properties of the FNS. Unfortunately, both studies utilize the same two samples. Kantor and Straus (1999) conducted the initial analyses and Pittman and Taylor (2002) attempted to replicate the initial analyses. These studies will be summarized and evaluated in this section.

Kantor and Straus (1999) conducted reliability analyses, with comparisons across two different samples in two pilot studies. Validity analyses sought to determine whether the FNS worked in theoretically consistent ways and to establish whether the instrument could distinguish between groups that are known to differ (Kantor & Straus). Cut-score analyses were also conducted, but they did not take into account different populations of responders. As mentioned previously, all of the current data on the FNS has been collected and validated for mothers only.

Results indicated the FNS was quickly and easily administered and positively received by both staff and families (Kantor & Straus, 1999). In addition, the FNS demonstrated excellent internal consistency reliability (see Table 2; $\alpha = .91$), good construct validity and adequate concurrent validity based on known groups comparison and moderate correlations with the CAPI. On the basis of known group comparisons and with the clinical cut point of 9, the FNS produced a sensitivity of 74%, a specificity of 75% and 89%, a false positive rate of 11% and 25% and a false negative rate of 26% (Kantor & Straus). The false positive rate and specificity of the measure was different in the two pilot studies. The second pilot study had a higher false positive rate (25%) but also included a larger sample size (569 vs. 159) and higher risk mothers. Thus, it seems as if the FNS produces more false positives when the sample includes more families

who have a higher risk for maltreatment. This is an important distinction since it is precisely these cases (those who may be at high risk for maltreatment) that are of utmost interest in screening families for risk of maltreatment.

A reanalysis conducted by Pittman and Taylor (2002) utilized the same sample as Kantor and Straus (1999) to replicate the initial findings. Internal consistency reliability coefficients were highly similar to Kantor's for the first pilot group and precisely the same for the second pilot study. Pittman's findings on subscale reliability generally replicated the initial study and suggested acceptable-to-good reliability for seven of the nine subscales and good-to-excellent reliability for the total scale (see Table 2). The demographics subscale had poor reliability, but researchers argued that these items, as a subscale, are of little value to begin with. Pittman was unable to replicate Kantor's findings perfectly in the factor analysis and raised questions about the originally conceptualized factor structure, especially for items on the social support, stress, self-esteem and depression domains. Pittman and Taylor also questioned the imbalance of the number of items in each of Kantor's conceptual domains. They stated that a positive response on a domain with 5 to 8 items may lead that respondent to a high needs classification since that set of items is overrepresented. Likewise, the same is true for a low needs classification. The domains with the most number of items are: Support, Family of origin violence, Stress and Relationship Discord. It may be important to investigate if mothers who responded positively to these domains were more likely to receive a high needs classification as compared with the other domains.

Pittman and Taylor (2002) also compared demographic characteristics on screener subscales and found that younger mothers reported more problems on several of the subscales. To the surprise of the researchers, pregnancy status appeared to have

the opposite effect on risk status. Non-pregnant mothers reported more relationship distress, more prior family violence, and a greater likelihood of having a history being reported for a child or spouse abuse case. However, this finding may be the result of a sampling bias. Most bases sample pregnant women as part of an obstetrical orientation, which includes a very heterogeneous sample of mothers (primary prevention referral source). Mothers who are not pregnant are generally either self-referred or referred to the FAP by a medical provider or another concerned person who feels the respondent may benefit from NPSP services (secondary prevention referral source). Thus, these non-pregnant mothers may experience more risk factors because they are actually at higher risk.

Pittman and Taylor (2002) were also unable to replicate the cut-score analyses exactly. However, with a cut-score of 9, they found the measure had a sensitivity of 72.5, a specificity of 76.8, a false positive rate of 23.2, and a false negative rate of 27.5, using Kantor's definition of maltreatment. Analyses of individual screener items in the prediction of maltreatment revealed that only a small number of items consistently contributed to accurate classifications of the maltreating group. These findings should be interpreted with caution, however, because the methods the researchers employed in these analyses did not take into account the inter-correlations between the items. Based on all of these findings, the researchers recommended a modified FNS consisting of 24 items which they believed performed as well as or better in predicting risk reliably and consistently (Pittman & Taylor).

The reader will recall that the FNS was designed in a manner such that only women who are pregnant and have a partner answer all 53 items. Women who fit the other three profiles of respondents as outlined in Table 3 complete a fewer number of

items. Therefore, some respondents did not answer items that were not applicable to them. Both Kantor and Straus (1999) and Pittman and Taylor (2003) decided to treat non-applicable items as data missing at random. Hence, they employed listwise deletion of these cases or in some analyses, they omitted variables altogether. As a result of these decisions, only women who were pregnant and who had a partner were included in all of the analyses. It appears as if other potentially at risk groups would have a decreased likelihood of meeting the established cut-score of 9 simply because they complete a fewer number of items. For example, a single mother who is not currently pregnant might “appear” at lower risk compared with a pregnant woman with a partner, simply because there are fewer opportunities for her to meet the cut-score on the FNS. It would be important to investigate whether the FNS performs differently for different populations of respondents.

Although the studies completed thus far on the FNS are notable, there are a number of ways these findings could be strengthened. First, it would be important to establish that the psychometric properties are constant across a new sample of mothers. Second, the FNS is designed to assess four different populations of respondents, but these populations have not been taken into account during analyses. Future research should account for these different respondent populations in the study design and analyses. Third, there is speculation that because some of the domains on the FNS are weighted more heavily with a greater number of items than others, classification decisions may be biased toward these domains. This needs to be empirically investigated. Fourth, all of the analyses thus far have been completed from a cross-sectional, known groups approach. Kantor and Straus (1999) recognized that establishing the predictive validity of the FNS would be an important contribution to

future research. Therefore, it would be important to establish the sensitivity, specificity, false positive and false negative rates, and the predictive validity of the FNS by examining actual maltreatment outcomes over time.

Besides the two studies reviewed above, a comprehensive review of multiple databases revealed no additional studies using the FNS. In addition, personal communication with USAF FAP administrative personnel confirmed there are no other unpublished studies for this measure. As discussed, there are significant gaps that are apparent as the result of this review. Therefore, although the FNS is a promising measure for the concurrent screening of IPV and child maltreatment, additional studies are warranted.

Research Questions for the Current Study

This dissertation sought to address many of these gaps in knowledge related to the FNS. First, this study utilized a new sample of mothers and expectant mothers in order to validate the initial findings on the FNS on a new sample. Secondly, the four respondent populations were considered in order to determine if the FNS works consistently for each group. Third, this study specifically investigated the predictive capacity of the FNS on future occurrence of family violence. Therefore, the following research questions were examined:

Research Question 1: What is the reliability coefficient for the original FNS scale as currently used in practice?

Research Question 2: What is the dependability of the original FNS using the cut-score currently used in practice?

Research Question 3: Does the factor structure of the FNS differ for the four respondent populations (e.g. pregnant women with partners, pregnant women without

partners, non-pregnant mothers with partners and non-pregnant mothers without partners)?

Research Question 4: To what degree does the original FNS predict the occurrence of family violence as currently used in practice?

Research Question 5: To what degree do the common items among the scales predict the occurrence of family violence?

Research Question 6: To what degree can the predictive capacity of the FNS improve?

Research Question 7: To what degree can the predictive capacity of the FNS improve if alternative scoring procedures are used (e.g. continuous scoring or alterations in cut-score)?

CHAPTER IV - METHODS

This study utilized a population-based retrospective, non-experimental research design. Mothers and expectant mothers who are either active duty military personnel or the wives of active duty military personnel were the target population for this study. The secondary data used for this study were collected at 79 different AF bases worldwide from October of 2002 through March of 2006 and are contained in the Air Force System of Records (FASOR). All of the mothers and expectant mothers in the sample completed the FNS either as a part of an initial obstetrical orientation program or because they expressed interest in receiving New Parent Support Program (NPSP) services through the AF FAP. All of the women in the sample were either pregnant or had a child under the age of 3 at the time they completed the FNS.

Data Source

Two different datasets contained within FASOR were used to derive the samples for this study. The NPSP dataset contains information on all mothers who voluntarily completed the FNS as part of their interest in the NPSP. The maltreatment dataset consists of detailed information regarding any known allegations of both child and spouse abuse involving an active duty member or family member. Data collection procedures for each dataset will now be explained in more detail.

NPSP Dataset

Women interested in receiving services through the AF's NPSP completed the FNS as an initial screening measure, usually in the NPSP clinic office or in obstetrical orientation classes. During the time frame of interest, 57,263 women were logged into the NPSP dataset. A FAP staff member scored the measure using a computerized

scoring program. Item responses and total scores were uploaded to the central NPSP database. Only mothers' FNS scores were included in the NPSP database; data on fathers were excluded. For each case, the NPSP database contains the date of FNS completion, raw responses to each of the 53 FNS items, the total FNS score, high or low needs classification, positive responses to any of the five automatic high needs items, referral source, and numerous demographic characteristics.

In addition to information specific to the FNS, the NPSP dataset also includes clinical information on families who accepted and were provided NPSP services. These data include the date services were initiated and terminated, the number of visits completed, and the reason for case closure. Since the NPSP is a voluntary program, this additional program information is lacking on families who refused NPSP services or who were determined ineligible for services.

Maltreatment Dataset

As discussed previously, DOD directives define child and spouse abuse, mandate that each military branch establish a central registry to track all allegations of abuse, and require the reporting of all such incidents to the respective service's central registry (Department of Defense, 2004). The Air Force's central registry contains information on allegations of family violence reported to an AF FAP program. During the time frame of interest, 33,957 cases of family violence were reported to the AF FAP and logged into the AF central registry.

Allegations of family violence may be received from a variety of sources including neighbors, school officials, daycare workers, medical providers, or commanders. When an incident of alleged family violence occurs, information regarding the incident is provided to the FAP. If the allegation meets initial criteria, a case is opened in the FAP

and the family is requested to attend an intake interview to discuss the allegation with a clinician. Following this interview, a multidisciplinary committee reviews the case and determines if it meets AF criteria for abuse or neglect. Information included in the AF central registry was collected during the intake interview or was provided by other agency staff such as child protection workers, medical professionals or law enforcement authorities. The following information is logged into the maltreatment database: incident referral date, incident status determination and date, whether the victim was a spouse or a child, severity of incident at intake, type of maltreatment, demographic information of the military member, military member's role in the incident, and offender's relationship to the victim. Other information regarding allegations of family violence are included in this dataset, but will not be used in this study.

All personal identifiers were removed from both of the datasets and a random case identification number was assigned to each case by FAP data specialists before the data were released to the researcher. Data received for this study did not contain any personal identifying information. Since the secondary data used for this study did not contain any personal identifying information, a certificate of exemption from review by the full Institutional Review Board for Research Involving Human Subjects was granted by the University of Tennessee Knoxville.

Design

This study was conducted in two phases. The first phase sought to examine the psychometric properties and predictive capacity of the original FNS, while the second phase of the study used these results to conceptualize a new version of the FNS and to determine whether it performed better under comparative analyses.

In order to examine the current functioning of the FNS, the following psychometric properties of the measure were examined: the factor structure using confirmatory factor analysis (CFA), reliability of the overall scale and subscales, and dependability of cut-score classifications. To complete Phase 1, the predictive capacity of the FNS was examined using a case-control design (Schlesselman, 1982). The case-control design is a research methodology that examines outcomes or effects retrospectively to identify potential predictors or causes. Schlesselman (1982) defines the cases in a case-control study as the individuals with a particular condition. Alternatively, controls are defined as those individuals in whom the condition is absent.

In the current study, cases were identified as those in the source population who had a substantiated family violence incident following their completion of the FNS. This group is referred to as the "Maltreatment group". Controls were defined as those in the target population who completed the FNS, but whose family did not have a known allegation of family violence either before or after their completion of the FNS. Families who had any allegation of family violence, even if it was unsubstantiated, were excluded from the control group. Specific criteria for the selection of controls will be discussed later in this chapter. Based on the scoring procedures outlined by Kantor and Straus (1999), the exposure or predictor variables of interest were the items and total score on the FNS that were thought to predict risk for family violence.

Some authors have suggested that because the case-control design is non-experimental, use of the terminology "control group" may be misleading (Geffner, Rosenbaum & Hughes, 1988). These authors recommend the term "comparison group" instead. Therefore, this paper adopts the term "comparison group" instead of "control group" due to the non-experimental nature of this study.

Phase 2 used the results of the factor and reliability analyses to propose a new version of the FNS. Alterations in scoring procedures and cut-score changes were examined to determine if the predictive capacity of any new measures could be improved over the original FNS. All analyses from Phase 1 were repeated with the new versions of the measure and results were compared with the original FNS.

Variables of Interest

Family Violence Occurrence

Family violence was defined as any incident of child physical, emotional, or sexual abuse, child neglect, or intimate partner violence. Incidents of elder abuse were excluded in this study, as the AF currently does not track allegations of elder abuse. Furthermore, incidents of family violence must have been entered into the AF central registry as a substantiated incident of family violence. This study was not able to account for family violence incidents which may have been undetected by the AF FAP.

The presence or absence of a substantiated family violence allegation was the outcome variable for this study. Within the Maltreatment dataset, each allegation of family violence received one of three possible case determinations: (1) Substantiated – the information available regarding the allegation suggested the incident met the established criteria for family violence; (2) Unsubstantiated, Unresolved – the information available regarding the allegation was insufficient to determine whether or not the incident met criteria for family violence; and (3) Unsubstantiated, Did Not Occur –the information available regarding the allegation was sufficient in order to determine the allegation did not meet the established criteria for family violence. Allegations that were either unsubstantiated as unresolved or unsubstantiated as did not occur were coded as

family violence absent (0). Substantiated allegations of family violence were coded as family violence present (1).

Family Violence Risk

The principle focus of this study was the performance of the FNS as a measure for the risk of family violence. The reader will recall that the FNS was developed with three uses in mind: (1) to assist the NPSP staff in making classification decisions on the allocation of services based on family needs, (2) to provide a means to better assess, plan, and conduct clinical interventions for the NPSP families, and (3) to provide a more systematic means of assessing family well-being at program entry (Kantor & Straus, 1999). The FNS was not developed to comprehensively assess all of the characteristics associated with multiple types of maltreatment, but was developed to capture those major characteristics of families that are particularly associated with physical assaults on parents or young children, with the assumption that many of these characteristics are associated with other types of maltreatment as well (Kantor & Straus). The development, administration, and scoring of the FNS were described in detail in Chapter Three. Additionally, Appendices A and B detail the items on the FNS and the scoring procedures. Hence, the total scores on existing or revised versions of the FNS served as the predictor variables for the study.

Respondent Population Groups

The reader may recall from Chapter 3 that mothers who complete the FNS answer different questions depending on whether or not they are currently pregnant or have a partner. Hence, the range of possible scores on the FNS is different depending on the status of the respondent. For example, only those respondents who are both pregnant and have a partner complete all 53 scored items. Because of this, respondents

were categorized into one of four respondent populations: (1) women who are *not* pregnant and who do *not* have a partner (NoPgNoPt); (2) women who *are* pregnant but who *do not* have a partner (PgNoPt); (3) women who *are not* pregnant but who *do* have a partner (NoPgPt) and (4) women who *are* pregnant and who *do* have a partner (PgPt).

Unfortunately, data were not collected on pregnancy or partner status outside of the items on the FNS. Additionally, the FNS does not define what is meant by "partner" and this is left open to the respondent's interpretation. Therefore, respondents were categorized into the four groups according to patterns of responses on specific items on the FNS that related to pregnancy or partner status.

Specifically, pregnancy status was determined by examining whether or not responses were given to one or two pregnancy-related questions (items #14 and #15). Item #14 ("My partner is very supportive of this pregnancy") was only included for those respondents who indicated they had a partner, while item #15 ("This is an unplanned pregnancy") was used to determine pregnancy status for all respondents. Respondents were categorized as pregnant if they responded to either item #14 or item #15.

Partner status was determined by examining whether responses were given to five partner-related questions (item #17 – "My partner treats me well", item #18 – "My partner and I have a very good relationship", item #19 – "I wish my partner and I got along better", item #20 – "I have thought seriously about ending my relationship with my partner", and item #28 – "My partner sometimes drinks five or more drinks of alcohol at a time, but mostly on weekends"). Respondents were categorized as having a partner if they responded to one or more of these five questions.

Based on responses to these pregnancy and partner-related items, a new variable called "Respondent Population Group" was created. Sample sizes for the four

groups varied widely. The NoPgNoPt group had the smallest sample size (N = 136), while the PgPt group had the largest sample size (N = 47,531). This finding is somewhat understandable given that the NPSP recruits heavily in obstetrical orientation programs and tends to focus services on prevention at the earliest point in pregnancy.

NPSP Service Utilization

For those cases that accepted and received NPSP services, the dosage of the service was recorded in the NPSP dataset as a count variable. The number of home or office visits represents the amount of NPSP services received by the family. A higher number of visits indicated more home or office visitation services were provided. Unfortunately, there was no information contained within the NPSP database that would provide information about quality or the effectiveness of services provided, nor does the database indicate specific therapeutic approaches that were used with clients. Therefore, this variable only indicates a rough estimate of the dosage of services, but not the extent, quality or effectiveness of these services.

Missing Data

Several decisions were made regarding the handling of missing data. A vigilant examination of the nature of missing data for specific cases was conducted. Creation of respondent populations enabled the researcher to identify some missing data that were likely conditional (i.e. data that were not applicable to the respondent) and distinguish it from data that were missing at random (MAR) or data missing completely at random (MCAR). Finally, when possible, an expectation maximization (EM) algorithm was used to impute missing data on quantitative variables. Researchers have found that EM

provides parameter estimates that come very close to population values, even with substantial missing data (Choi, Golder, Gillmore, & Morrison, 2005; Pigott, 2001).

Samples

In order to answer the research questions presented for this study, different samples were derived for different analyses. There were, in essence, three different samples used for the analyses. Each of these samples will now be described in detail.

Both Sample 1 and Sample 2 were derived from the 57,263 women who were entered in the NPSP dataset during the time frame of interest. Because not all women were supposed to answer all of the questions on the FNS, cases were examined with respect to missing data only on the FNS items that were common to all four respondent populations (36 items). Originally, only cases that did not contain any missing data on these 36 items (N = 45,794) were going to be used in the analyses. However, this would have eliminated 21% of the original sample. Therefore, cases were included in the sample if they contained less than 10 percent of missing data on the common 36 items (N = 51,468 or 89.8% of the original sample). Consideration was given to lowering the amount of tolerated missing data to less than 20 percent, but it was determined that only approximately 1,000 additional cases would result (91.6% of the original sample). Therefore, only cases with less than 10 percent missing data on the common items were retained. Next, any remaining missing data that was quantitative in nature was imputed using EM.

An examination of those cases missing *less* than 10 percent of the common items on the FNS (cases that were selected to be included in the analyses) as compared to those cases missing *more* than 10 percent of common items (cases that were not

selected to be included) did reveal some interesting differences. Compared with those who were selected, cases that were not selected were more likely to state the pregnancy was unplanned (67.9% vs. 40.9%). Those not selected were also more likely to wish that they and their partner got along better compared with those who were selected (39.5% vs. 16.4%). Finally, those cases not selected for the study were more likely to be single as compared with those selected (30.2% vs. 11.7%). Although 90% of the sample was retained, caution must be taken when interpreting results of the data analyses since the cases selected for the study appear to be biased towards the married mother who has planned her pregnancy and has a more positive relationship with her spouse.

Sample 1 was derived by dividing the cases missing less than 10% of the common items into the four respondent population groups based on the responses to the pregnancy and partner related questions as described in the previous section (see Figure 1). The following sample sizes resulted: NoPgNoPt (n = 136), PgNoPt (n = 729), NoPgPt (n = 3072), and PgPt (n = 47,531). The pregnant and partnered group is substantially larger than the other three population groups. Again, as discussed previously, this finding is not surprising considering the NPSP recruits heavily in obstetrical orientation programs and targets pregnant mothers for early intervention. Sample 1 was used when analyses focused specifically on differences between the four respondent populations.

Sample 2 was derived by randomly dividing the cases missing less than 10% of the common items into two approximately equal groups. One half of Sample 2 (Sample 2.1; N = 25896) will be used in Phase 1 analyses. The other half of Sample 2 (Sample 2.2; N = 25574) will be used for analyses during Phase 2 to cross-validate findings during Phase 1 and to conduct new analyses (see Figure 2). Samples 2.1 and 2.2 were

compared across three different demographic characteristics. Results revealed no significant differences with respect to rank of the military sponsor (X^2 [25] = 28.31, p = .294) and with respect to mother's ethnicity (X^2 [9] = 8.01, p = .533). However, Samples 2.1 and 2.2 did have significant differences with respect to mother's age (X^2 [3] = 13.02, p = .005). Sample 2.2 had more mothers who were age 18 or younger (52.3% vs. 47.7%), more mothers who were between the ages of 30 and 40 (50.7% vs. 49.3%), and more mothers who were 41 years old and older (52.2% vs. 47.8%).

Finally, Sample 3 was derived in two steps for the predictive validity portion of the analyses. First, cases in the NPSP dataset that had a substantiated incident of family violence after completing the FNS were selected as part of the *maltreatment group* (n = 1127) of Sample 3. None of the families included in the maltreatment group had more than one family violence incident and none of the families completed more than one FNS screener. Cases that were missing more than 10% of the scored items on the FNS were not included in this portion of the analyses. Therefore, 1,075 cases (95.39% of the original sample) were included in Sample 3 as the maltreatment group.

Second, consistent with the case-control design, a *comparison group* for Sample 3 was also selected from the NPSP dataset. Cases that had completed the FNS and did not have a known allegation of family violence were eligible for inclusion into the comparison group. Cases were excluded from the comparison group if there was any allegation of family violence, even if the allegation was unsubstantiated. This reduced potential non-differential misclassification bias for the comparison group (Rothman & Greenland, 1989). To further refine selection of cases to the comparison group, an approximately equal number of comparison group cases were matched with maltreatment group cases.

Matching. Geffner and colleagues (1988) point out that case-control designs may result in significant bias unless groups are carefully matched on a number of relevant criteria. Variables that are appropriate for matching are those factors that are believed to have an effect on the outcome of interest independently of other exposure variables (Schlesselman, 1982). Conversely, however, one must also be careful not to overmatch, which may reduce the validity or statistical efficacy of the comparison (Schlesselman, 1982). Matching that either does not improve or reduces the efficiency of a case-control comparison is considered unnecessary (Schlesselman, 1982). In addition, statistical analyses must take the matching into consideration or there is risk of a reduction in validity of the study (Schlesselman, 1982).

A confounding variable is defined as the extraneous factors responsible for the differences in the frequency of the outcome variable between the case group and the comparison group (Rothman & Greenland, 1998). The principle characteristic of a confounding variable is that it is predictive of the frequency of the outcome variable, especially within the comparison group (Rothman & Greenland). Furthermore, a potential confounder should be associated with both the exposure and the outcome of the study (Rothman & Greenland).

In light of this, potential confounding variables were identified for this study. The date that the FNS was completion is a potential confounder. If any differences between the groups on the outcome variable were actually due to a shorter period of time between completion of the FNS and a subsequent incident of family violence, this would bias findings. Likewise, demographic characteristics like rank (which equates to socioeconomic status in the military population) and ethnicity are also potential confounders. Matching on these characteristics ensures any differences between the

groups are not attributable to these demographic characteristics. Finally, because prevention services were offered to families who were screened at risk on the FNS, NPSP service utilization is another important confounding variable. One would want to ensure that a lower frequency of the outcome variable in the comparison group was not due to the prevention services provided to these families.

This study matched the comparison group with the maltreatment group according to the date of FNS completion, the rank of sponsor, and the ethnicity of FNS responder. Because NPSP service utilization may have an effect on whether or not a family experienced family violence, this variable was controlled for statistically in the data analyses and was not matched. Other demographic characteristics, such as education level, age, or relationship status, may also be considered appropriate confounders. However, since all of these demographic characteristics are counted as scored items on the FNS, they could not be included as matched factors. Out of 51,468 eligible cases in the NPSP dataset, 1,064 cases were successfully matched with the maltreatment group, for a total sample size of 2,139 cases in Sample 3 (see Figure 3; 1,064 comparison group cases plus 1,075 maltreatment group cases). Sample 3 was used to conduct the predictive validity analyses for both phases.

Data Analyses

Reliability and Dependability

Sample 1 was used for all reliability and dependability analyses for the four respondent population groups. Sample 2.1 was used to estimate the reliability and dependability for the 36 items that are common to all respondents (the Common 36 Model; see Table 4 for specific subscales and items which make up this model).

Estimates of internal consistency reliability (Cronbach's alpha) were obtained for each of these scales using the alpha reliability procedure in the SPSS statistical software package. Measures with higher alpha reliability have higher degrees of internal consistency. Estimates of dependability were calculated using phi or phi of lambda for a particular cut score. Dependability indices indicate the expected agreement in scores across randomly parallel versions of a measurement procedure, either controlling for chance (φ) or given a specified cut score [(φ (λ)]. Hence, greater values on these indicators suggest greater agreement in scores. For the original measure, lambda was defined as 9 (which is the cut-score currently used in practice).

Missing Data on Demographic Questions. Estimates of reliability and dependability require that all conceptualized subscales are included in these calculations. Missing demographic data were a significant problem with respect to all groups and models. These data could not be imputed as this would bias the reliability estimates upward. Thus, the sample sizes used for reliability and dependability analyses were subjected to listwise deletion of missing data. This significantly reduced the sample sizes for the four respondent population groups in Sample 1.

For estimates including the demographic subscales, 62.2% of the sample was retained for the PgPt group, 58.8% of the sample was retained for the NoPgPt group, 22.2% of the sample was retained for the PgNoPt group, and 24.3% of the sample was retained for the NoPgNoPt group. Estimates of reliability and dependability for each of the scales were also calculated omitting the demographic items. Although these reliabilities do not reflect the internal consistency of the scales as currently used in practice, these samples retained a much higher percentage of cases in the analyses. Eighty-seven percent of the sample was retained in the PgPt group, 82.6% of the

sample was retained for the NoPgPt group, 58% of the sample was retained for the PgNoPt group, and 72.8% of the sample was retained for the NoPgNoPt group. Finally, Sample 2.1 also had significant missing data, especially on the demographic items. Analyses including the demographic items retained 65.2% of the sample; while 90.6% of the sample was retained when the demographic items were omitted.

Confirmatory Factor Analyses

Confirmatory factor analysis (CFA) using structural equation modeling (SEM) was used to determine how well the data fit the original FNS measurement model and whether the common items on the FNS functioned similarly for each of the four respondent populations. The first task was to determine how well each respondent population group's set of questions fit the data for that particular respondent population. Each of the four respondent population groups answer a different set of items on the FNS (see Table 3 for the specific number of items completed by each group) so each of the four respondent populations has a unique measurement model. Therefore, using Sample 1, each of the four models' factor structure was examined separately.

In addition to testing the four population specific models, the factor structure of the items common to all four groups, the Common 36 Model, was also tested as a potential shortened version of the FNS. Sample 2.1 was used for this analysis, which will be replicated in Phase 2 with Sample 2.2.

For the last part of the CFA, the Common 36 Model was tested across all four respondent populations using the four respondent groups in Sample 1 to determine whether any measurement model invariance exists. If no measurement model invariance existed, then the Common 36 Model could be used in subsequent analyses as the preferred model on the basis of parsimony.

For the CFA, the demographics subscale was handled differently than previous analyses. Since the seven demographic questions on the FNS do not comprise a single latent construct, they cannot be included in the CFA. Bollen and Long (1993) point out that structural equation models must make substantive sense. Since the seven demographic questions do not make substantive sense as a single construct, they were eliminated from the CFA altogether.

When evaluating model fit using SEM, the most popular methods involve use of the χ^2 goodness-of-fit statistic and several fit indexes (Hu & Bentler,1995). Researchers have cautioned against over-reliance on the χ^2 goodness-of-fit statistic for assessing model fit, especially with increased statistical power associated with a large sample (Bollen & Long, 1993; Hu & Bentler, 1995). Because the χ^2 goodness-of-fit statistic is influenced by sample size, trivial differences in the model may result in rejection of a good model. Furthermore, Bollen and Long (1993) recommend the use of several different fit indices, and state that reliance on any one index is to be avoided. Based on these guidelines, three measures of model fit were selected.

Browne and Cudeck (1993) recommended the use of the root mean square error of approximation (RMSEA) as a measure of the discrepancy per degree of freedom. RMSEA values of 0.05 or less indicate a close fit of the model in relation to the degrees of freedom and values of between 0.05 and 0.08 indicate a moderate fit (Browne and Cudeck (1993). Hu and Bentler (1999) recommended reporting the discrepancy between the observed correlations and the model-reproduced correlations. The standardized root mean square residual (SRMR) represents the average value across all standardized residuals (Byrne, 2001). With a potential range of 0 to 1, SRMR value close to .08

indicating a well-fitting model (Hu & Bentler, 1999). Finally, Hu and Benter (1999) recommended the use of the comparative fit index (CFI) with a revised cut-off criterion of close to .95. Finally, the issue of parsimony was also evaluated.

Predictive Validity Analyses

The final part of Phase 1 examined the predictive capacity of the original FNS. Binary logistic regression was conducted on Sample 3 to determine whether the FNS scores were predictive of the occurrence of family violence. The predictive validity of the original FNS was evaluated by calculating a total score using the FNS scoring procedures as outlined by Kantor and Straus (1999). This total score was then entered in a regression equation as the predictor variable. The presence or absence of substantiated family violence was used as the outcome variable. In addition, sensitivity, specificity, positive predictive value (PPV), negative predictive value, NPV, false positive rate (FPR), false negative rate (FNR) and overall correct classifications were calculated and examined based on different cut-score classifications.

Development of New Shortened FNS

Data from the above analyses were reviewed and exploratory factor analysis (EFA) was conducted using the Sample 2.1 to determine if a shortened version of the FNS could be developed without a loss of reliability or validity. Based on the results of the factor analyses, poorly functioning items or subscales were deleted and the remaining items were re-organized into new subscales. Reliability and dependability analyses were then repeated using Sample 2.2 (N = 25574). In addition, the use of a new scoring procedure was also examined. Raw scores on all items were summed and a total score was included in the regression equation as a predictor variable. Just as in Phase 1, Sample 3 was used for these analyses.

Model Comparison

The original FNS, the Common 36 Model and the newly modified FNS were then compared according to the results of the CFA, the estimates of reliability and dependability, and the predictive validity analyses. In order to compare the predictive capacity of the models, the scored totals for each of the models were converted to z-scores. Binary logistic regression was repeated using the summed z-scores as the predictor variable and the presence or absence of abuse as the outcome variable. This procedure allowed for comparison of the regression coefficients and odds ratios across different scales and scoring procedures. In addition, the overall correct classification and sensitivity of each model were also used as the basis for comparison.

CHAPTER V - RESULTS

Table 5 summarizes the two phases of this study, detailing the analyses conducted in each phase and sample utilized for that analysis. With respect to Phase 1, results of reliability analyses, dependability analyses, factor analyses and predictive validity analyses are presented for the original FNS, the four respondent population groups and for the 36 items common to all four respondent population groups (the Common 36 Model; see Table 4 for specific subscales and items included in this model). In addition, results for the test of measurement invariance across the four respondent population groups are presented. Phase 2 results are described next including the results of the exploratory factor analyses and rationale for the development of shortened versions of the FNS. Results of reliability analyses, dependability analyses, factor analyses and predictive validity analyses are presented for the shortened models. Finally, all of the models are compared with respect to predictive capacity.

Phase 1 Results

Reliability and Dependability of Original Scales

Research questions #1 and #2 relate to the reliability and dependability of the original FNS as currently used in practice. Estimates of internal consistency and dependability were calculated for the Sample 1 respondent population groups using the items conceptualized by Kantor and Straus (1999) for each group and for the Common 36 Model using Sample 2.1. All scored demographics questions were included in these calculations. When demographic items were included, the PgPt respondent group model was identical to the original FNS model, since the PgPt group completes all of the scored items on the FNS. As shown in Table 6, all estimates of reliability were fairly low

for each of the models tested when the demographic subscale was included. Reliabilities ranged from .620 to .783, with the original FNS (also the PgPt group) having a reliability estimate of .725. Reliability for the Common 36 Model using Sample 2.1 (α = .733) was very similar to the original FNS when the demographics subscale was included.

Estimates of the dependability of classification decisions while controlling for chance (φ) and with a cut-score of 9 (φ[λ=9]) for each model including the demographic items are also summarized in Table 6. Estimates of the consistency in classification decisions across randomly parallel versions of the measurement instrument while controlling for chance were markedly low for all of the models. Phi of lambda using a cut score of 9 showed improvement in classification decisions across all models.

Because significant numbers of cases were lost when the demographics subscale was included in these estimates, these analyses were repeated omitting the demographics questions but using the same samples. Table 6 displays these results. All of the models showed marked improvement across estimates of Cronbach's Alpha, Phi and Phi of lambda. Omitting demographics, the models' estimates of reliability ranged from .897 to .929. Estimates of dependability ranged from .853 to .909 when controlling for chance, and from .993 to .994 for the cut-score of 9. The original FNS had a reliability of .924 when the demographics questions were omitted. These data suggest that, as expected, the demographics subscale appears to lower the reliability and dependability of these models.

Confirmatory Factor Analyses

Respondent Populations. Next, this study sought to answer research question #3 regarding the factor structure of the original FNS. Because each of the four respondent population groups answers a slightly different number of questions on the FNS, the four

respondent population group models were tested separately (demographic questions were omitted for all four groups as described in the previous chapter). The NoPgNoPt model included 36 items; the PgNoPt model included 39 items; the NoPgPt model included 41 items; and the PgPt model included all 44 items on the FNS (again, excluding demographic questions).

Two different estimation procedures were considered for these analyses: Asymptotically Distribution Free (ADF) estimation and Maximum Likelihood (ML) estimation using expectation maximization (EM). Although ADF was developed as an alternative estimator that does not assume multivariate normality of the measured variables, studies have shown that ADF only works well when sample size are extremely large (>5,000 cases; Gold, Bentler, & Kim, 2003; Hu & Bentler, 1995). However, Gold, Bentler, and Kim found that non-normality corrected EM outperformed ADF regardless of the missing-data mechanism in a Monte Carlo study. Therefore, although some of the models' sample sizes are large enough to warrant the use of ADF, it appears as if EM implementation of ML still outperforms ADF (Gold, Bentler & Kim, 2003). Thus, EM implementation of ML was the estimation procedure of choice for this study.

Table 7 displays the fit indices for each of the four population group models. Only one fit index suggested that all four models had close fit (SRMR ranged from .061 to .083). One fit index (RMSEA ranged from .063 to .082) suggested that two models had close fit (PgPt and NoPgPt) while two models did not have close fit (PgNoPt and NoPgNoPt). A third fit index suggested all of the models lacked good fit (CFI ranging from .771 to .844). The NoPgPt model fit the data the best compared with the other three models (X^2 = 9908.620@743, $p < .000$, CFI = .844, SRMR = .061, RMSEA = .063).

Common 36 Model. The factor structure of the Common 36 Model was examined next. The reader will recall that the Common 36 Model is comprised of the 36 items that are common to all four of the respondent population groups. As shown in Table 7, the Common 36 Model fit the data slightly better compared with the four population-specific models (X^2 = 51227.046@557, p < .000, CFI = .874, SRMR = .049, RMSEA = .059). However, this model also reflected mixed results with respect to the fit indices. Both the SRMR and RMSEA suggested close fit to the data, while the CFI suggested the model lacked good fit.

Multi-Group Comparison. In order to test for measurement invariance across the four population groups, the Common 36 Model was examined across the four population groups' datasets. Results are shown in Table 7. There was no evidence of measurement invariance across population groups for the Common 36 Model. Measurement weights, structural covariances, structural residuals, and measurement residuals all appeared to be working similarly across all of the population groups.

Predictive Capacity of Original FNS

In order to answer research question #4 regarding the predictive capacity of the original FNS, abuse status (presence or absence) was regressed on the total FNS score using the original scoring procedures as outlined by Kantor and Straus (1999) and described in detail in Chapter 3. In order to determine the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall correct classification (OCC) of the FNS, a cut-score of 9 (which is the cut-score currently used in practice) was used to classify respondents as "at risk" or "not at risk". Then, the regression equation was re-calculated using this cut-score as the basis for "at-risk" and "not at-risk" classifications. Results are summarized in Table 8.

Using the total FNS score converted to a standardized z-score, the overall model was statistically significant (X^2 [1] = 196.92, N = 2139, p < .000). Specifically, for every one standard deviation increase in scores on the FNS, the likelihood of having had a substantiated abuse incident nearly doubled (Wald [1] = 165.69, p < .000, Odds Ratio = 1.95). In order to examine whether NPSP service dosage had an effect on the occurrence of abuse, abuse status was regressed on the total FNS score and NPSP service dosage. The overall model was statistically significant (X^2 [2] = 198.15, N = 2139, p < .000). More specifically, for every one standard deviation increase in FNS scores, mothers were nearly twice as likely to have had a subsequent abuse incident when controlling for NPSP service utilization (Wald [1] = 150.26, p < .000, Odds Ratio = 1.93). However, there was not a significant relationship between the outcome variable and the dosage of home visitation services when controlling for the total FNS score (Wald [1] = 1.21, p = .271, Odds Ratio = 1.01).

Using a cut-score of 9, the overall model was also statistically significant (data not shown; X^2 [1] = 126.49, N = 2139, p < .000). Specifically, as mothers were classified as “at risk” (had a total FNS score of 9 or greater), they were more than two and a half times as likely to have had a substantiated abuse incident (see Table 9; Wald [1] = 121.523, p < .000, Odds Ratio = 2.79). Based on a cut-score of 9, the FNS correctly classified 61.6% of respondents. Sensitivity of the measure was relatively low (48.9% of the respondents who subsequently experienced a family violence incident were correctly classified as at risk for a family violence incident). Seventy-four percent of the cases were correctly classified as not at risk for a subsequent incident of family violence.

In order to determine if other cut-scores would perform more favorably, cut-scores of 4 through 10 were also examined. Results were compared across all indicators

and are shown in Table 9. Overall correct classifications (OCC) were at the highest rate with a cut-score of 7 and the odds ratio remained fairly stable across all cut-scores examined. With a cut-score of 6, the OCC rate was slightly lower that the highest rate (60.7% compared with 62.2% using a cut-score of 7), but the sensitivity was increased from 63.4% to 70.6%.

Phase 2 Results

Exploratory Factor Analysis

This phase of analyses sought to identify any potential shortened versions of the original FNS that performed as well as or better than the original version. Using Sample 2.1, the first step to identifying a potentially shortened version of the FNS was to examine the factor structure using exploratory factor analysis (EFA). Using Maximum Likelihood estimation procedure (with Varimax rotation), results revealed there were seven components with eigenvalues greater than 1. Based on a seven factor model, the EFA was repeated on the same sample, extracting seven factors from the data. Item #39 ("My income is often inadequate for basic needs…") cross-loaded on four factors and item #55 ("I feel sad quite often") loaded equally on two factors. These items were excluded from the new model as they did not support only one factor.

Using the remaining items, seven subscales were identified (see Table 4). Four of the subscales were identical to the original FNS subscales: Substance Abuse, Violence Approval, Family of Origin Abuse, and Prior Family Violence. Three of the subscales were altered. Stress/Coping was now made up of items #21 and #22 (previously from the Stress subscale), item #23 (previously from the Relationship Discord subscale) and items #24 and #25 (previously from the Support subscale). Items

from the original Self-esteem subscale remained intact except that one additional item was added (item #54 from the Depression subscale). Therefore, this subscale was re-named “Self-Esteem/Depression). Finally, the new Interpersonal Support subscale was created from items previously found on the Support subscale (items #45-51) and the Depression subscale (items #52 and #53).

Reliabilities were computed for each new subscale separately and are shown in Table 10. Results indicated that the Prior Family Violence subscale had very low alpha reliability (α = .477). Therefore, this subscale (which was comprised of only two items) was deleted. Additionally, the Stress/Coping subscale would have an improved alpha reliability from .760 to .778 if item #24 was deleted. Likewise, the Violence Approval subscale saw an improvement of alpha reliability from .731 to .754 if item #29 was deleted. These two additional items were deleted leaving 30 items remaining for the new model. This model, named the Common 30 Model, was subsequently examined and compared with the Common 36 Model and the original FNS.

Reliability and Dependability of Two New Scales

Using Sample 2.2, estimates of internal consistency were calculated for the new Common 30 Model and for the Common 36 Model. Results are shown in Table 11 and indicate that both models have nearly identical alpha reliabilities (α = .918 and .920, respectively). The six subscales that comprised the Common 30 Model had reliability estimates ranging from α = .757 to .880. The nine subscales that comprise the Common 36 Model had reliability estimates ranging from α = .474 to .880. Prior Family Violence had the lowest alpha reliability of all of the subscales at α = .474. This subscale is comprised of only two dichotomous (yes or no response) items, thus it is not surprising that this subscale has a fairly low estimate of internal consistency. Estimates of

dependability were also very similar for both models (φ = .906 for the Common 30 Model and φ = .906 for the Common 36 Model). Hence, it appears as if both shortened models have very similar estimates of internal consistency reliability and dependability.

Confirmatory Factor Analyses of Common 30 and Common 36 Models

The factor structure of the Common 30 and Common 36 Models were examined using Sample 2.2. The reader will recall that the Common 36 Model's factor structure was already tested using Sample 2.1 in Phase 1 analyses. Therefore, for the Common 36 Model, Phase 2 analyses served as cross-validation of these initial findings. As shown in Table 12, once again the Common 30 Model and the Common 36 Model fit the data with very similar results. The fit indices of both models reflected mixed results. Both the SRMR and RMSEA suggested close fit to the data (SRMR = .0483 for the Common 30 Model and .0492 for the Common 36 Model; RMSEA = .067 and .059, respectively) while the CFI for both models suggested they lacked good fit (CFI = .871 and .874, respectively). Results of the cross-validation of the Common 36 Model revealed that the data fit the model in a nearly identical manner using Sample 2.1 and Sample 2.2 as shown in Table 12.

Predictive Capacity of the Common 30 and the Common 36 Models

In order to answer research question #5, #6 and #7, Sample 3 was employed to evaluate the capacity of the Common 30 and Common 36 models to predict the occurrence of family violence. Two different scoring procedures were used to score each model. First, the original scoring procedures as outlined by Kantor and Straus (1999) were utilized. With this scoring method, total scores on the Common 30 Model ranged from 0-30 while total scores on the Common 36 Model ranged from 0-36. To examine an alternate scoring procedure, raw responses (scored as 1, 2, 3, or 4) on each item were

summed to form a total score. With this method, total scores on the Common 30 Model ranged from 0-120 while total scores on the Common 36 Model ranged from 0-148.

In order to convert the total scores to a comparable metric, the scores for each model using the two scoring methods were converted to z-scores using the SPSS statistical software package. This standardized total score was then entered into the regression equation. In essence, there were four models tested in this section of the analyses: 1) The Common 30 Model using original scoring, 2) the Common 30 Model using raw sum scoring, 3) the Common 36 Model using original scoring, and 4) the Common 36 Model using raw sum scoring.

With respect to the Common 30 Model using original scoring, the overall model was significant (see Table 8; $\underline{X}^2$ [1] = 167.27, $\underline{N}$ = 2139, $\underline{p}$ < .000). More specifically, for every one standard deviation increase in scores on the Common 30 Model using original scoring, mothers were nearly twice as likely to experience a subsequent abuse incident (Wald [1] = 142.84, $\underline{p}$ < .000, Odds Ratio = 1.84). With respect to the Common 30 Model using raw sum scoring, the overall model was also significant ($\underline{X}^2$ [1] = 115.34, $\underline{N}$ = 2139, $\underline{p}$ < .000). Again, for every one standard deviation increase in scores on the Common 30 Model using raw sum scoring, mothers were 1.6 times as likely to experience a subsequent incident of abuse (Wald [1] = 107.21, $\underline{p}$ < .000, Odds Ratio = 1.65).

With respect to the Common 36 Model using original scoring, the overall model was significant ($\underline{X}^2$ [1] = 182.59, $\underline{N}$ = 2139, $\underline{p}$ < .000). More specifically, for every one standard deviation increase in scores on the Common 36 Model using original scoring, mothers were nearly twice as likely to experience a subsequent abuse incident (Wald [1] = 152.96, $\underline{p}$ < .000, Odds Ratio = 1.91). Finally, with respect to the Common 36 Model using raw sum scoring, the overall model was again significant ($\underline{X}^2$ [1] = 122.16, $\underline{N}$ =

2139, p < .000). More specifically, for every one standard deviation increase in scores on the Common 36 Model using raw sum scoring, mothers were 1.69 times more likely to experience a subsequent abuse incident (Wald [1] = 112.60, p < .000, Odds Ratio = 1.69).

Controlling for NPSP Services. In order to control for a potential confounding variable, each of the models were included in the regression equation again along with the NPSP services variable. Results were similar for most of the models and are shown in Table 8. NPSP service dosage did have a statistically significant relationship with abuse status when controlling for the total score on the Common 36 Model using raw sum scoring (Wald [2] = 3.90, p < .048, Odds Ratio = 1.02). NPSP service dosage was not significantly related to a subsequent abuse incident when controlling for each of the other models tested. Therefore, it appears as if changes in scoring procedures from original scoring to raw sum scoring did not improve the predictive capacity of the models, even after controlling for NPSP service dosage.

Cut-score analyses. Cut-score analyses were also conducted on the Common 30 and Common 36 Models using original scoring. Results of the cut-score analyses are shown in Table 9. Overall correct classifications were at the highest level using a cut-score of 4 for the Common 30 Model and a cut score of 5 for the Common 36 Model. However, these cut scores produced a moderate sensitivity rate of 59.5% and 55.9%, respectively. Using a cut-score of 3 for both models, the OCC rate was slightly lower than the highest rate (60.4% for both models), but the sensitivity of the models were increased to 69.8% and 73.6%, respectively.

Creation of the Shortest Possible Version of the FNS

Since results suggest that shortening the original FNS from 53 items to 36 items and then again to 30 items resulted in little change in the reliability, dependability or predictive capacity of the measure, the decision was made to explore how brief the measure could become without diminishing its predictive ability. In order to explore this, the subscales of the Common 36 Model were examined in order to determine which of the subscales predicted a subsequent abuse incident. Sample 3 was utilized and the nine subscales that make up the Common 36 Model were regressed on abuse status. Findings are presented in Table 13 and suggest that Prior Family Violence was, by far, the best predictor of a subsequent abuse incident (Wald [1] = 41.23, $p < .000$, Odds Ratio = 4.219). Therefore, even though the two items that make up the Prior Family Violence subscale (item #56 – Have you or your partner been involved in a suspected or verified case of child abuse or neglect?; and item #57 – Have you or your partner been involved in a suspected or verified case of spouse abuse?) do not perform well as a single construct (as evidenced by low alpha reliability estimate of .474), these items do contribute significantly to prediction of a future abuse incident. Therefore, these items were included as potential predictors of abuse.

The next step was to regress the thirty items that make up the Common 30 Model *plus* items #56 and #57 in a backwards step-wise regression with Sample 3 data in order to determine which items contributed the most towards the prediction of a subsequent incident of family violence. Results suggested that twelve items were most predictive (see Table 14 for a description of these items). As expected, items #56 and #57 remained in the regression equation, along with ten other items.

The reader may recall from Chapter 3 that the original FNS was developed with five items that were designated “automatic high needs items”. The content of these five

items was such that a positive response to any of these items would automatically categorize a respondent as “at risk”, regardless of the overall score received on the measure. The backwards step-wise regression resulted in the inclusion of all five of these “automatic high needs items” in the prediction model. This finding may provide evidence of construct validity for these presumed “high risk” items.

Predictive Validity of Common 12 Model

The twelve items that resulted from the backwards step-wise regression were used to form the Common 12 Model. Because of the low number of items included in this scale, subscales were not created. Sample 3 was randomly split into two approximately equal halves. Sample 3.1 was used for to conduct the backwards step-wise regression analysis. Like previous analyses, items comprising the Common 12 Model were summed using original scoring and the total score was converted to a z-score. The summed z-scores were then regressed on abuse status using binary logistic regression. Table 8 displays the results showing the overall model was yet again significant (X^2 [1] = 91.59, N = 1035, p < .000). More specifically, for every one unit increase in scores on the Common 12 Model using original scoring, mothers were nearly twice as likely to have experienced a subsequent incident of abuse (Wald [1] = 73.30, p < .000, Odds Ratio = 1.95). When controlling for NPSP services, the overall model was again significant (X^2 [1] = 91.67, N = 1035, p < .000) and there was virtually no change to the prediction model (Wald [2] = 70.80, p < .000, Odds Ratio = 1.96). NPSP Services dosage was not statistically related to abuse status when controlling for the total score on the Common 12 Model.

The estimate of internal consistency reliability for the Common 12 Model (α = .790) was similar to the original FNS including demographics (α = .729) but was much

lower than the original FNS when the demographics subscale was omitted (α = .924). This result would be expected since estimates of reliability are sensitive to the number of items included in the calculation. Phi of the Common 12 Model (φ = .755) was lower than the other models tested (φ = .853 to .909), but not considerably lower.

Results of the cut-score analyses for the Common 12 Model revealed that with a cut-score of 1, the Common 12 Model correctly classified 61.9% of the cases, with a FNR of only 38.1%. With a cut-score of 1, the Common 12 Model had a consistency of classification decisions across randomly parallel version of the measure of 98.15% (φ[λ = 1] = .9815), which was very similar to that of the other models tested. Overall correct classifications were at the highest rate with a cut-score of 2 (61.4%). However, sensitivity dropped to 39.5% with this cut-score.

Cross-validation of Prediction Results

The final step in the development of the Common 12 Model was to attempt to cross-validate the prediction findings on Sample 3.2. Dummy coding was used to assign cases to one of two groups: the original sample (Sample 3.1; 0) and the replication sample (Sample 3.2; 1). Next, a five step binary logistic regression was used with the presence or absence of a subsequent incident of abuse as the outcome variable.

In Step 1, the dummy variable was entered into the regression equation in order to determine if there were significant differences with respect to each group and the outcome variable. In order for the cross-validation to be successful, there would be no significant relationship between the dummy variable and the outcome variable. Next, the total score on the Common 12 Model was added to the equation as a predictor of the outcome. In order for the cross-validation to be successful, there would be a significant relationship between the total score on the Common 12 Model and the outcome

variable. The third step involved the creation of an interaction term (Dummy variable X the Common 12 total score). This step is the crux of the cross-validation effort and provides information about whether the effect of the total score on the outcome variable is different depending on the group assignment. Fourth, each of the variables that were not retained in the backwards stepwise regression analysis (20 items) were added to the total score on the 12 items that made up the Common 12 Model and this new total score (made up of 13 items) was entered into the regression equation. For example, item 21 was omitted from the Common 12 Model, but in Step 4 this item was scored and added to the total score on the Common 12 Model. This procedure provides information about whether any of the items that were omitted from the Common 12 Model actually contribute to the prediction of the outcome variable when the item is added to the Common 12 items. Finally, twenty interaction terms were created (Dummy variable X each of the new 13 item scales) and added into the regression equation. This step provides information about whether the each of the 13 item scales predicted the outcome variable differently depending on group assignment.

Results of the cross-validation are presented in Table 15. Overall, the results reflect the successful cross-validation of the Common 12 Model for the prediction of a subsequent abuse incident. As expected, the overall model for Step 1 was not statistically significant (X^2 [1] = 1.28, p = .257) which indicates there were no differences between the two groups with respect to the outcome variable. The overall model for Step 2 was statistically significant (X^2 [1] = 176.38, p < .000). This suggests that, once again, the Common 12 Model was statistically associated with the outcome variable when controlling for group assignment. In Step 3, the creation of the interaction term ensures that any potential interaction effect between group assignment and total score on the

Common 12 Model is accounted for. Although the overall model was statistically significant (X^2 [3] = 177.70, p < .000), the interaction between group assignment and total score on the model was not statistically significant (Wald [3] = .027, p = .868, Odds Ratio = 1.01). Results of the fourth step suggested that none of the 20 non-selected items, when added back into the Common 12 Model, improved the prediction of abuse of abuse. Additionally, there was no statistically significant interaction effect across group membership when the 20 non-selected items were added to the Common 12 Model. Hence, these results support the cross-validation of the predictive capacity of the Common 12 Model.

Summary of Results and Model Comparison

Using different cut-scores and scoring procedures for each model, this study compared the predictive capacity of the original FNS with several shortened versions of the screener: the Common 36 Model, the Common 30 Model and the Common 12 Model. All models displayed good internal consistency and dependability, and all displayed similar mixed results in terms of the CFA analyses. Results of reliability and dependability analyses for the Common 12 Model were somewhat lower than for the other models tested. This result is expected due to the fewer number of total items on the Common 12 Model, as estimates of internal consistency are sensitive to the total number of items included on the scale.

Comparison of the predictive capacity of the models revealed that all models had very similar predictive capacity. Optimal cut-scores for each of the four models were determined using cut-score analyses. According to Table 8, the original FNS had the highest overall correct classification rate of 62.2% using a cut-score of 7. This cut-score improved sensitivity of the measure to 63.4%, with a 39.1% FPR and a 36.6% FNR. This

cut-score would be the recommended cut-score for those wishing to obtain the highest possible correct classifications. However, lowering the cut-score to 6 increased the sensitivity of the FNS to 70.6%, with a FPR of 49.3% and a FNR of 29.4%. Use of this cut-score would allow for capture of more mothers who are truly at risk, so that prevention services may be offered, but would also capture a higher percentage of mothers who are not truly at risk for a subsequent incident of abuse.

The highest OCC rates for the other models were: 61.9% on the Common 36 Model with a cut-score of 5, 61.4% on the Common 30 Model with a cut-score of 4, and 61.4% on the Common 12 Model with a cut-score of 2. It appears as if the best balance between sensitivity and specificity for each of the models was obtained with a cut-score of 7 for the original FNS, a cut-score of 4 for the Common 36 Model, a cut-score of 4 for the Common 30 Model, and a cut-score of 1 for the Common 12 Model.

Of these four models and cut-scores, the Common 36 Model had the highest sensitivity rate of 64.2%. However, the Common 12 Model produced the closest balance between sensitivity and specificity as compared with the other models. Although the Common 12 Model lagged behind the other models with respect to reliability and dependability, results of the predictive validity analyses revealed it could predict subsequent abuse as well as the other models.

CHAPTER VI – DISCUSSION

This dissertation sought to address some of the technical issues identified by both sides of the screening debate with respect to one measure of family violence risk – the Air Force Family Needs Screener. After reviewing several considerations for the screening of family violence and discussing characteristics of effective screening measures, several measures that screen for risk for child maltreatment or spouse abuse were reviewed. Next, the psychometric properties of the AF Family Needs Screener (FNS) were examined in detail. This study sought to further the validation of the FNS by testing the measure's reliability, dependability and factor structure on a new sample of mothers and by reporting the predictive capacity of the FNS with data collected retrospectively. Finally, this study sought to develop a shortened version of the FNS that would retain the characteristics of effective screening measures and perform as well as, or better than, the original FNS.

Summary of Major Findings

This dissertation addressed three primary gaps in knowledge related to the FNS. First, a new sample of respondents was utilized so in order to replicate the initial findings related to the psychometric properties of the FNS. This was particularly important since the FNS has been widely used with a military population for many years and had only been tested on one sample of respondents. Results revealed that estimates of internal consistency reliability for the FNS as currently used in practice (including the demographic subscales) were somewhat lower than obtained in the pilot study and replication of the pilot study (Kantor & Straus, 1999; Pittman & Taylor, 2001). However, when the demographics subscale was omitted, estimates of internal consistency

reliability were almost identical to the pilot study. Not surprisingly, it appears as if the demographic questions included on the FNS are detrimental to the factor structure of the measure. Estimates of dependability were not reported in previous studies, but the FNS was found to have good dependability using the cut-score currently used in practice.

The second gap in knowledge addressed by this study related to the specification of the four population respondent groups in the analyses. This study was the first to examine questions on the FNS specific to four respondent population groups. Previous studies treated items that were not applicable to all respondents as missing data and subjected this data to listwise deletion (Kantor & Straus, 1999; Pittman & Taylor, 2002). The current study created four respondent population groups and tested each group's unique measurement model separately based on the items specific to each group in order to determine whether these questions contributed to the overall factor structure of the FNS, and whether the FNS performed consistently for each of these four respondent population groups. Results suggested the population-specific items did not significantly contribute to the factor structure of each model tested. In addition, there was no evidence of measurement invariance across the four population groups for the Common 36 Model. These results provided empirical evidence to support the shortening of the FNS to include only items common to all groups and provided evidence that these common items perform consistently for all four groups.

This finding is important because the original FNS was designed with some inherent complexity that had not yet been considered empirically. Screening measures must be simple, inexpensive, safe, and acceptable to the population of interest (Ferrer, 1968; Friis & Sellers, 2004; Sackett & Holland, 1975). Moreover, screening measures must perform in a manner that is generally consistent for a large proportion of the

respondents (Hudson, 1982). Thus, the use of only common items for all four population groups improves the simplicity of the measure and the results provide evidence that the measure performs consistently across the four groups.

The third gap in knowledge addressed by this study involved obtaining empirical evidence about the predictive capacity of the FNS. As stated by Caldwell and colleagues (1988), if the primary purpose of a screening instrument is to identify risk for the future presence of a condition, it is particularly important to establish the predictive validity of the measure. One of the three uses of the FNS as stated by Kantor and Straus (1999) is to assist staff in making risk classification decisions so that prevention services are allocated first to those who are most at risk. NPSP home-visitation services are offered to those families who are classified as “at-risk” on the FNS. Therefore, inaccurate classification decisions may result in families who truly are at risk having been overlooked for prevention services and may result in an increased potential for negative individual outcomes (Wilson & Junger, 1968). Although the effectiveness of NPSP prevention services has not yet been examined in the literature, the NPSP seeks to decrease the risk of family violence. Misclassification of risk status could be potentially costly to some families as well as to the Air Force. This study was the first to examine the predictive validity of the FNS.

Results of the predictive analyses for the FNS as currently used in practice revealed that for every one unit increase in the FNS, women were over two and a half times as likely to have experienced a subsequent incident of family violence. However, results of cut-score analyses for the original FNS using a cut-score of 9 were somewhat disappointing. Using this cut-score, the FNS correctly identified slightly less than half of those who eventually did have a subsequent incident of abuse, leaving more than half of

the cases who did have a subsequent incident of abuse undetected. Kantor and Straus (1999) reported a much higher rate of correct classification of at risk respondents using the same cut score in their pilot study, but their study utilized known groups as the outcome variable. Using the report of a subsequent abuse incident as the outcome variable, the current study reported a much lower rate of correct classifications of those truly at risk for family violence.

Opponents have cited the inappropriate reliance on reported or confirmed cases of abuse as the criterion against which prediction is measured as a weakness in family violence screening (Caldwell, et al., 1988). If one assumes that reported or confirmed cases of abuse are underestimated, then use of this criterion as the outcome in this study only underestimates the number of cases categorized as true positives. Hence, this would explain why the use of such criterion would lower the rate of true positives as compared with the use of known-groups as the criterion for classification. It does not seem appropriate to cite the reliance on reported or confirmed cases of abuse as a weakness in the family screening debate. Rather, one must recognize that in the absence of absolute knowledge of reality, reported or confirmed cases of abuse simply have a tendency to underestimate true positive rate.

Considerations in Model Comparison

The comparison of models and the selection of a preferred model is a particularly challenging enterprise. It requires consideration of the properties of effective screeners along with careful assessment of programmatic contextual factors. There are no absolute standards to guide this decision-making process. Therefore, one must balance the purpose of the screening measure, the needs of the population of interest, and

programmatic needs in order to prioritize the unique characteristics of each model under consideration.

There are three properties of effective screening measures that are particularly important, especially with respect to screening in primary prevention settings. First, the measure must be acceptable to the population of interest and to agency staff (Sackett & Holland, 1975). Especially in primary prevention programs, screeners should be brief and simple (Ferrer, 1968; Friss & Sellers, 2004; Sackett & Holland, 1975). Second, screeners used in a primary prevention setting are intended to identify individuals who may be at risk for a particular condition. Therefore, the measure must correctly classify a high rate of respondents who are actually at risk for the condition. Thirdly, if the primary purpose of a screening instrument is to identify risk for the future presence of a condition, it is particularly important to establish the predictive validity of the measure (Caldwell, et al.1988).

In order to select the best model and cut-score for use with a particular prevention program, one must balance empirical results obtained with practical, programmatic needs. The potential benefits of screening for family violence can outweigh the potential costs if such programmatic contextual factors are considered, the measure is acceptable to both the target population and agency staff, and sufficient attention is paid to the technical challenges of screener development (Daro & McCurdy, 1994; McCurdy, 1995).

One approach to model selection involves selecting the model and cut-score that most equally balances rates of correct classification for true positives with rates of correct classification for true negatives. Using this approach, the overall correct classification rate could be selected for each model at a particular cut-score so that this

rate is maximized. Based on the results presented in this study, the original FNS using a cut-score of 7 resulted in the highest overall correct classification rate of any of the models tested. However, this approach does not take into account acceptability to the population, nor does it necessarily identify a higher proportion of cases that are truly at risk.

Another approach would be to select the most parsimonious model that predicts the outcome variable as well as or better than the original scale. Therefore, if two models predict in a fairly similar manner, the briefest and simplest model would be selected. Using this approach, based on the results of this study, the Common 12 Model using a cut-score of 1 would be the selected model. This approach considers acceptability and predictive capacity, but use of this approach may still result in an unacceptable rate of correct classification of true positives.

Combining the two previous approaches and taking into account programmatic contextual factors creates yet another approach. If prevention programs wish to capture as many women as reasonably possible who are true positives, then lowering the cut-score beyond the level that balances the overall correct classification rate may be required. This decision would result in a higher rate of true positives. However, in doing this, there is an associated increase in the false positive rate which may create an overburdening of prevention staff. In this example, instead of balancing the true positive rate with true negative rate (as in the first approach), the challenge becomes balancing the true positive rate with the false positive rate so as to not overburden staff resources. In this example, as long as prevention staff resources are available, the decision to capture a higher rate of true positives may be desirable given this specific programmatic context.

Sedlak (1988) purports that a high false positive rate, in and of itself, does not render a screening measure useless. Instead, the key issue in managing the likely high false positive rate inherent to screening low base rate events is careful consideration of what one does with the results of such a screening. Screening measures with higher false positive rates may be appropriate if they are used to identify individuals, couples or families who warrant further questioning, follow up, or to offer voluntary preventative services (Sedlak, 1988). Hence, the selection of a preferred model must ultimately take into account those psychometric properties that are most important with consideration of the context in which the measure is being utilized.

Recommendation of Revised FNS

Because the FNS is utilized in the NPSP to identify risk of future family violence, I believe that the most important characteristic of a revised version of the FNS is its predictive validity. Since the FNS is used to identify those individuals whom will be offered prevention program services, the measure's ability to correctly classify those who are truly at risk for the condition is also of utmost importance. Within the context of a primary prevention program, I would recommend lowering the cut score of a model so that the rate of correct classifications of true positives are increased, even if this results in a higher rate of false positives. I agree with Sedlak's (1988) assertion that more false positives can be tolerated within the context of a primary prevention program. Finally, acceptability to the population of interest is an important property for consideration, but it is not to be considered over and above the two characteristics previously mentioned.

Based on this prioritization, I would recommend the Common 36 Model with a cut-score of 3 for implementation for the NPSP. The Common 36 Model performed consistently across all four respondent populations, and had good reliability,

dependability, and adequate factor structure. With a cut-score of 3, for every one unit increase on the Common 36 Model using original scoring, women were about two and a half times as likely to experience a subsequent incident of family violence. The cut-score of 3 for this model correctly classified nearly 74% of respondents who were at risk. This cut-score provides a balance between capturing a high rate of true positives, while remaining mindful of the potential to overburden staff if the cut-score is lowered any further. Of course, with a cut-score of 3, the Common 36 Model does produce a fairly high percentage of false positive classifications. However, for reasons stated previously, this false positive rate can be tolerated within the context of this primary prevention program.

Both the Common 30 Model with a cut-score of 3 and the Common 12 Model with a cut-score of 1 had comparable predictive validity, but these models produced lower sensitivity rates at these cut-scores. Although these two models have fewer items than the Common 36 Model, the purpose of the FNS is to correctly classify respondents as at risk for family violence. Therefore, choosing a model with higher sensitivity is preferred over a model with simply fewer items. The Common 36 Model does simplify the original FNS by reducing the number of items from 53 down to 36 items, all of which are common items. The elimination of items that are only completed by respondents who are pregnant or who have a partner creates a more parsimonious version of the FNS. This would further contribute to improved simplicity of the measure since all respondents would answer all items on the measure. This will no doubt increase the acceptability of the measure for respondents and staff alike, in addition to saving time and resources in the administration and scoring of the measure. Although the Common 36 Model is not

the most parsimonious model examined, the correct classification and predictive capacity of the model are more important that merely choosing the shortest model.

Limitations of the study must be considered when interpreting the findings. Because this research design is based on a convenience sample from two existing databases, findings are only generalizable to active duty military mothers or wives of active duty military fathers. Findings cannot be generalized to other populations. Future studies should examine the utility of the FNS for use with different populations including fathers and non-military affiliated mothers.

Respondent population groups were created based on inferences about the pattern of responses and omissions of responses to specific questions on the FNS. It is difficult, if not impossible, to know precisely why a respondent failed to answer a specific question. Although the inclusion of more than one question in order to determine pregnancy status or partner status helps to alleviate this issue to some extent, there is no way to be absolutely certain these inferences were made correctly. And since there were no other data available to verify these inferences, the possibility of misclassification of respondents cannot be overlooked. Because it is possible that these classifications were made incorrectly, results specific to each respondent population group should be considered with caution. Results did suggest that the Common 36 Model worked consistently across the four respondent population groups, but future research should replicate this analysis with better defined population groups.

New Parent Support Program service dosage was included in this study as a potential confounding variable. Results suggested that, for the most part, when the dosage of NPSP services variable was included in the regression equation, there was no relationship between NPSP services and the outcome variable, when controlling for the

total score on the measure. Caution should be used when interpreting this finding as information regarding the efficacy of NPSP services. NPSP services were examined in this study in order to control for a potential confounding variable, not as a test of service effectiveness. Additionally, the NPSP services dosage variable is merely a count of the number of visits each mother received. Hence, future research would need to address this research question specifically and would need to more clearly define the type and extent of the services provided in order to evaluate its effectiveness.

The efficacy of the NPSP is especially important for future research. All conclusions regarding the need to classify individuals as at risk for family violence are based on the assumption that family violence prevention programs can be effective at reducing the risk of violence. The efficacy of home visitation programs similar to the NPSP has been mixed (Bilukha et al., 2005; Chaffin & Friedrich, 2004; Duggan, McFarlane et al, 2004; Fraser, Armstrong, Morris, & Dadds, 2000; Olds et al., 1997). In addition, there have been no known studies examining the efficacy of the NPSP specifically. Therefore, this is most certainly an important direction for future research.

Finally, this study took into account some potential confounding variables and attempted to control for their possible effect on the outcome of interest. The use of matching and inclusion of NPSP service dosage into the regression model were two such attempts. However, due to the non-experimental nature of this study, it is possible that there were unknown confounding variables which have affected the results of this study. For example, although a very high percentage of the original sample was retained in order to comprise Sample 1, it appears as if these cases included in the study were more likely to be married women who planned their pregnancy and who have a more positive relationship with their spouses. It would be important to specifically examine

women who were single parents who had unplanned pregnancies in order to cross-validate findings to this population.

Current theoretical approaches to family violence, evidence of interdependence of spouse abuse and child maltreatment, and evidence of a significant interaction effect between home visitation services, domestic violence, and the prevention of child maltreatment provide theoretical and practical support for screening for family maltreatment as a whole. The AF FNS is the only known measure developed to screen for risk of both child maltreatment and IPV concurrently. Results suggest that the FNS is a brief screening measure which can consistently classify respondents as at risk for the subsequent occurrence of family violence. However, the original FNS with a cut-score of 9 misclassified a significant number of respondents as not at risk even though they did experience a subsequent family violence incident. In addition, the complicated design of the FNS, with inclusion of population group-specific items, adds to the complexity of the measure and only increases the probability that items would be completed incorrectly.

Results of this study suggest that the Common 36 Model is one shortened version of the FNS that is able to predict the subsequent occurrence of family violence as well as the original measure, and will likely be more acceptable to respondents than the longer, more complex original version of the FNS. Therefore, the Common 36 Model shows promise as family violence screening measure that may be practical for use in a variety of social service settings that provide prevention or treatment services for both child maltreatment and spouse abuse.

REFERENCES

Abbott, J., Johnson, R., Kozoil-McLain, J., & Lowenstein, S. R. (1995). Domestic violence against women: Incidence and prevalence in an emergency department population. *JAMA*, *273(22),* 1763-1767.

Anderson, C. (1987). Assessing parenting potential for child abuse risk. *Pediatric Nursing*, *13(5),* 323-328.

Anderson, C. (1993). The Parenting Profile Assessment: Screening for child abuse. *Applied Nursing Research*, *6(1)*, 31-39.

Anderson, C. (2000). Revisiting the Parenting Profile Assessment to screen for child abuse. *Journal of Nursing Scholarship*, *32(1),* 53.

Avison, W. R., Turner, R. J., & Noh, S. (1986). Screening for problem parenting: Preliminary evidence on a promising instrument. *Child Abuse & Neglect*, *10*, 157-170.

Babcock, J. C., Costa, D. M., Green, C. E., & Eckhardt, C. I. (2004). What situations induce intimate partner violence? A reliability and validity study of the Proximal Antecedents to Violent Episodes (PAVE) Scale. *Journal of Family Psychology*, *18(3)*, 433-442.

Bandera, A. (1973). Aggression: A social learning analysis. Englewood Cliffs, NJ: Prentice-Hill.

Berk, R. A., He, Y., & Sorenson, S. B. (2005). Developing a practical forecasting screener for domestic violence incidents. *Evaluation Review*, *29(4),* 358-383.

Bersani, C. A., & Chen, H. (1988). *Sociological perspectives in family violence*. In V. B. Van Hasselt, R. L. Morrison, A. S. Bellack, & M. Hersen (Eds.), Handbook of family violence (pp. 57-86). New York: Plenum.

Bilukha, O., Hahn, R. A., Crosby, A., Fullilove, M. T., Liberman, A., Moscicki, E. et al. (2005). The effectiveness of early childhood home visitation in preventing violence: A systematic review. *American Journal of Preventative Medicine*, *28 (2S1),* 11-39.

Bollen, K.A. & Long, J.S. (Eds.). (1993). *Testing Structural Equation Models*. Newbury Park, CA: Sage.

Brennan, R. L. (1980). *Application of generalizability theory*. In R. A. Berk (Ed.), Criterion-referenced measurement: The state of the art (pp. 186-232). Baltimore: Johns Hopkins University Press.

Browne, M.W. & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K.A. Bollen & J.S. Long (Eds.), Testing Structural Equation Models (pp.136-162). Newbury Park, CA: Sage.

Brown, J. B., Lent, B., Brett, P. J., Sas, G., & Pederson, L. L. (1996). Development of the Woman Abuse Screening Tool for use in family practice. *Family Medicine, 28(6),* 422-428.

Brown, J. B., Lent, B., Schmidt, G., & Sas, G. (2000). Application of the Woman Abuse Screening Tool (WAST) and WAST-short in the family practice setting. *Journal of Family Practice*, *49(10),* 896-903.

Bugental, D. B., & Happaney, K. (2004). Predicting infant maltreatment in low-income families: The interactive effects of maternal attributions and child status at birth. *Developmental Psychology*, *40(2),* 234-243.

Burrell, B., Thompson, B., & Sexton, D. (1992). The measurement integrity of data collected using the Child Abuse Potential Inventory. *Educational and Psychological Measurement*, *52*, 993-1001.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications and Programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Caldwell, R. A., Bogat, G. A., & Davidson, W. S. (1988). The assessment of child abuse potential and the prevention of child abuse and neglect: A policy analysis. American *Journal of Community Psychology*, *16(5),* 609-624.

Caliso, J. A., & Milner, J. S. (1992). Childhood history of abuse and child abuse screening. *Child Abuse & Neglect*, *16,* 647-659.

Centers for Disease Control and Prevention (2003). *Costs of intimate partner violence against women in the United States*. Atlanta (GA): CDC, National Center for Injury Prevention and Control; Retrieved February 15, 2006 from www.cdc.gov/ncipc/pub-res/ipv_cost /ipv.htm.

Center for Disease Control and Prevention (2006). *Intimate partner violence: Fact sheet*. (2005, September 28). Retrieved Feb 13, 2006 from http://cdc.gov/ncipc/ factsheets/ipvfacts. htm.

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

Chaffin, M., & Valle, L. A. (2003). Dynamic prediction characteristics of the Child Abuse Potential Inventory. *Child Abuse & Neglect*, *27*, 463-481.

Chen, P. H., Rovi, S., Vega, M., Jacobs, A., & Johnson, M. S. (2005). Screening for domestic violence in a predominately Hispanic clinical setting. *Family Practice*, *22(6)*, 617-623.

Choi, Y., Golder, S., Gillmore, M.R., & Morrison, D.M. (2005). Analysis with Missing Data in Social Work Research. *Journal of Social Service Research, 31(3)*, 23-48.

Cochrane, A. L., & Holland, W. W. (1971). Validation of screening procedures. *British Medical Bulletin*, *27(1)*, 3-8.

Cohen, R. J. & Swerdlik, M. E. (2005). Psychological testing and assessment: An introduction to test and measurement (6th ed.). New York: McGraw-Hill.

Coker, A. L., Bethea, L., Smith, P. H., Fadden, M. K., & Brandt, H. M. (2002). Missed opportunities: Intimate partner violence in family practice settings. *Preventative Medicine*, *34*, 445-454.

Combs-Orme, T., Cain, D. S., & Wilson, E. E. (2004). Do maternal concerns at delivery predict parenting stress during infancy? *Child Abuse & Neglect*, *28(4)*, 377-392.

Combs-Orme, T., Martin, L., Fox, G. L., & Faver, C. A. (2000). Risk for child maltreatment: New mothers' concerns and screening test results. *Children and Youth Services Review*, *22(7),* 517-537.

Connelly, C. D., Newton, R. R., Landsverk, J., & Aarons, G. A. (2000). Assessment of intimate partner violence among high-risk postpartum mothers: Concordance of clinical measures*. Women & Health*, *31(1),* 21- 37.

Department of Defense (2004). *Family advocacy program* (Directive 6400.1). Washington, D.C.: Author.

Daro, D. & McCurdy K. (1991). Current trends in child abuse reporting and fatalities: The results of the 1990 annual fifty state survey. Chicago, IL: National Committee for Prevention of Child Abuse.

Daro, D. & McCurdy, K. (1994). Preventing child abuse and neglect: Programmatic interventions. *Child Welfare*, *73(5),* 405-430.

DeVoe, E. R., & Kantor, G. K. (2002). Measurement issues in child maltreatment and family violence prevention programs. *Trauma, Violence, & Abuse*, *3(1)*, 15-39.

Dubanoski, R. A., & McIntosh, S. A. (1984). Child abuse and neglect in military and civilian families. *Child Abuse & Neglect*, *8(1),* 55-67.

Duggan, A., McFarlane, E., Fuddy, L., Burrell, L., Higman, S. M., Winsham, A. et al. (2004). Randomized trial of a statewide home visiting program: Impact in preventing child abuse and neglect. Child Abuse & *Neglect*, *28*, 597-622.

Duggan, A., Windham, A., McFarlane, E., Fuddy, L., Rohde, C., Buchbinder, S. et al. (2000). Hawaii's healthy start program of home visiting for at-risk families: Evaluation of family identification, family engagement, and service delivery. *Pediatrics, 105(1)*, 250-259.

Dutton, M. A., Mitchell, B., & Hatwood, Y. (1996). The emergency department as a violence prevention center. *JAMA*, *51(3)*, 92-95+117.

Eckenrode, J., Ganzel, B., Henderson, C. R., Smith, E., Olds, D. L., Powers, J. et al. (2000). Preventing child abuse and neglect with a program of home nursing visitation: The limiting effects of domestic violence. *JAMA*, *284(11)*, 1385-1391.

Eddy, D. M. (1991). *How to think about screening*. In D. M. Eddy (Ed.), Common screening tests (pp.1-21). Philidelphia, PA: American College of Physicians.

Ernst, A. A., Weiss, S. J., Cham, E., & Marquez, M. (2002). Comparison of three instruments for assessing ongoing intimate partner violence. *Medical Science Monitor*, *8(3),* 197-201.

Family Advocacy Program (2004). *Family advocacy program commanders guide*. Retrieved January 31, 2005, from http://www.dod.mil/fapmip/module1/mod1_2.htm.

Feldhaus, K. M., Kozoil-McLain, J., Amsbury, H. L., Norton, I.M., Lowenstein, S. R., & Abbott, J. T. (1997). Accuracy of 3 brief screening questions for detecting partner violence in the emergency depertment. *JAMA*, *277(17),* 1357-1361.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R.L. Linn (Ed.) Educational measurement (pp.105-146). New York: Macmillan.

Ferrer, H. P. (1968). *Screening for health: Theory and practice*. London: Butterworths.

Field, C. A., Caetano, R., & Nelson, S. (2004). Alcohol and violence related cognitive risk factors associated with the perpetration of intimate partner violence. *Journal of Family Violence*, *19(4)*, 249-253.

Fogarty, C. T., & Brown, J. B. (2002). Screening for abuse in Spanish-speaking women. *Journal of the American Board of Family Practitioners*, *15,* 101-111.

Fogarty, C. T., Burge, S., & McCord, E. C. (2002). Communicating with patients about intimate partner violence: Screening and interviewing approaches. *Family Medicine*, *34(5),* 369-375.

Fraser, J. A., Armstrong, K. L., Morris, J. P., & Dadds, M. R. (2000). Home visiting intervention for vulnerable families with newborns: Follow-up results of a randomized controlled trial. *Child Abuse & Neglect*, *24(11)*, 1399-1429.

Friis, R. H., & Sellers, T. A. (2004). *Epidemiology for public health practice* (3rd ed.) Sudbury, MA: Jones and Bartlett.

Gaines, R., Sandgrund, A., Green, A. H., & Power, E. (1978). Etiological factors in child maltreatment: A multivariate study of abusing, neglecting, and normal mothers. *Journal of Abnormal Psychology, 87(5),* 531-540.

Geffner, R., Rosenbaum, A. & Hughes, H. (1988). Research Issues Concerning Family Violence. In V. B. Van Hasselt, R. L. Morrison, A. S. Bellack, & M. Herson (Eds.), *Handbook of Family Violence.* New York: Plenum Press.

Gelles, R., & Straus, M. A. (1979). *Determinants of violence in the family: Toward a theoretical integration*. In W.A. Burr, R. Hill, F.I. Nye, & I.L. Reiss (Eds.), Contemporary theories about the family (pp. 549-581). New York: Macmillan.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist, 18*, 519-521.

Gold, M.S., Bentler, P.M., & Kim, K.H. (2003). A Comparison of Maximum-Likelihood and Asymptotically Distribution-Free Methods of Treating Incomplete Nonnormal Data. *Structural Equation Modeling, 10(1),* 47-79.

Greenberg, E. M., McFarlane, J., & Watson, M. G. (1997). Vaginal bleeding and abuse: Assessing pregnant women in the emergency department. *American Journal of Maternal/Child Nursing*, *22(4),* 182-186.

Guterman, N. B. (1999). Enrollment strategies in early home visitation to prevent physical child abuse and neglect and the "Universal versus targeted" debate: A meta-analysis of population-based and screening-based programs. *Child Abuse & Neglect, 23(9),* 863-890.

Halpern, L. R., Susarla, S. M., & Dodson, T. B. (2005). Injury location and screening questionnaires as markers for intimate partner violence. *Journal of Oral and Maxillofacial Surgeons, 63*, 1255-1261.

Hamberger, L. K., & Phelan, M. B. (2004). Domestic violence screening and intervention in medical and mental healthcare settings. New York: Springer.

Hambleton, R. K. (1980). *Test score validity and standard-setting methods*. In R.A. Berk (Ed.), Criterion-referenced measurement: The state of the art (pp. 80-123). Baltimore: Johns Hopkins University Press.

Helfer, R., Hoffmeister, J. & Schneider, D. (1978). *MSPP: A manual for use of the Michigan Screening Profile of Parenting*. Boulder, CO: Test Analysis and Development Corporation.

Heron, S. L., Thompson, M. P., Jackson, E., & Kaslow, N. J. (2003). Do responses to an intimate partner violence screen predict scores on a comprehensive measure of intimate partner violence in low-income black women? *Annals of Emergency Medicine*, *42(2),* 483-491.

Holden, E. W., Willis, D. J., & Foltz, L. (1989). Child abuse potential and parenting stress: Relationships in maltreating parents. *Journal of Consulting and Clinical Psychology*, *1(1),* 64-67.

Holtrop, T. G., Fischer, H., Gray, S. M., Barry, K., Bryant, T., & Du, W. (2004). Screening for domestic violence in a general pediatric clinic: Be prepared! *Pediatrics, 114*, 1253-1257.

Houry, D, Feldhaus, K., Peery, B., Abbott, J., al-Bataa-de-Montero, S., & Levine, S. (2004). A positive domestic violence screen predicts future domestic violence. *Journal of Interpersonal Violence, 19(9)*, 955-966.

Howing, P. T., Wodarski, J. S., Kurtz, P. D., & Gaudin, J. M. (1989). Methodological issues in child maltreatment research. *Social Work Research & Abstracts, 25(3)*, 3-5.

Hu, L. & Bentler, P.M. (1995). Evaluating Model Fit. In R.H. Hoyle (Ed.), Structural Equation Modeling: Concepts, Issues and Applications (pp.76-99). Thousand Oaks, CA: Sage.

Hu, L. & Bentler, P.M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling, 6(1),* 1-55.

Hudson, W. W. (1982). The clinical measurement package: A field manual. Homewood, IL: Dorsey.

Hudson, W. W., & McIntosh, S. R. (1981). The assessment of spouse abuse: Two quantifiable dimensions. *Journal of Marriage and the Family, 43(4)*, 873-885+888.

Jellen, L. K., McCarroll, J. E., & Thayer, L. E. (2001). Child emotional maltreatment: A 2-year study of US Army cases. *Child Abuse & Neglect, 25(5)*, 623-639.

Kantor, G. K., & Straus, M. A. (1999). *Report on the USAF Family Needs Screener*. Durham: University of New Hampshire.

Kessler, R. C., Molnar, B. E., Feurer, I. D., & Appelbaum, M. (2001). Patterns and mental health predictors of domestic violence in the United States: Results from the National Comorbidity Survey. *International Journal of Law and Psychiatry, 24*, 487-508.

Korfmacher, J. (2000). The Kempe Family Stress Inventory: A review. Child Abuse & Neglect, 24(1), 129-140.

Kotch, J. B., Browne, D. C., Dufort, V., Winsor, J., & Catellier, D. (1999). Predicting child maltreatment in the first four years of life from characteristics assessed in the neonatal period. *Child Abuse & Neglect, 23(4)*, 305-319.

Kozoil-McLain, J., Coates, C. J., & Lowenstein, S. R. (2001). Predictive validity of a screen for partner violence against women. *American Journal of Preventative Medicine, 21(2)*, 93-100.

Larson, C. P., Collet, J., & Hanley, J. A. (1987). The predictive accuracy of prenatal and postpartum high risk identification. *Canadian Journal of Public Health, 78*, 188-192.

Lewis, B. Y. (1985). The Wife Abuse Inventory: A screening device for the identification of abused women. *Social Work, 30(1)*, 32-35.

Little, L., & Kantor, G. K. (2002). Using Ecological Theory to understand intimate partner violence and child maltreatment. *Journal of Community Health Nursing, 19(3)*, 133-145.

Lyons, P., & Doueck, H. J. (1996). Risk assessment for child protective services: A review of the empirical literature on instrument performance. *Social Work Research, 20(3)*, 143-156.

Martuza, V. R. (1977). Applying norm-referenced and criterion-referenced measurement in education. Boston: Allyn and Bacon.

Mausner, J. S. & Kramer, S. (1985). *Epidemiology: An Introductory Text.* Philadelphia: Saunders.

McCarroll, J. E, Newby, J. H., Thayer, L. E., Ursano, R. J., Norwood, A. E., & Fullerton, C. S. (1999). Trends in child maltreatment in the US Army, 1975-1997. *Child Abuse & Neglect, 23(9)*, 855-861.

McCarroll, J. E., Ursano, R. J., Fan, Z., & Newby, J. H. (2004). Comparison of U.S. Army and civilian substantiated reports of child maltreatment. *Child Maltreatment, 9(1)*, 103-110.

McCurdy, K. (1995). Risk assessment in child abuse prevention programs. *Social Work Research, 19(2),* 77-88.

McGuigan, W. M., & Pratt, C. C. (2001). The predictive impact of domestic violence on three types of child maltreatment. *Child Abuse & Neglect, 25*, 869-883.

McGuigan, W. M., Katzev, A. R., & Pratt, C. C. (2003). Multi-level determinants of retention in a home-visiting child abuse prevention program. *Child Abuse & Neglect, 27*, 363-380.

McFarlane, J., Parker, B., Soeken, K., & Bullock, L. (1992). Assessing for abuse during pregnancy: Severity and frequency of injuries and associated entry into prenatal care. *JAMA, 267(23),* 3176-3178.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th edition). Ft. Worth, TX: Holt, Rinehart, and Winston.

Merrill, L. L., Crouch, J. L., Thomsen, C. J., & Guimond, J. M. (2004). Risk for intimate partner violence and child physical abuse: Psychosocial characteristics of multi-risk male and female Navy recruits. *Child Maltreatment, 9(1),* 18-29.

Messick, S. (1989). *Validity*. In R. L. Linn (Ed.) Educational measurement (pp.13-104). New York: Macmillan.

Milner, J. S. (1986). *The Child Abuse Potential Inventory: Manual* (2nd ed.). Webster, NC: Psytec.

Milner, J. S., & Gold, R. G. (1986). Screening spouse abusers for child abuse potential. *Journal of Clinical Psychology, 42(1),* 169-172.

Milner, J. S., Gold, R. G., Ayoub, C., & Jacewitz, M. M. (1984). Predictive validity of the Child Abuse Potential Inventory. *Journal of Consulting and Clinical Psychology, 52(5)*, 879-884.

Milner, J. S., & Wimberley, R. C. (1980). Prediction and explanation of child abuse. *Journal of Clinical Psychology, 36*, 875-884.

Mollerstrom, W. W., Patchner, M. A., & Milner, J. S. (1992). Family violence in the Air Force: A look at offenders and the role of the Family Advocacy Program. *Military Medicine, 157(7)*, 371-374.

Mollerstrom, W. W., Patchner, M. A., & Milner J. S. (1995). Child maltreatment: The United States Air Force's response. *Child Abuse & Neglect, 19(3)*, 325-334.

Munro, E. (2004). A simpler way to understand the results of risk assessment instruments. *Children and Youth Services Review, 26*, 877-887.

Murphey, D. A., & Braner, M. (2000). Linking child maltreatment retrospectively to birth and home visit records: An initial examination. *Child Welfare, 76(6),* 711-728.

Murphy, S., Orkow, B., & Nicola, R. M. (1985). Prenatal prediction of child abuse and neglect: A prospective study. *Child Abuse & Neglect, 9*, 225-235.

Murry, S. K., Baker, A. W., & Lewin, L. (2000). Screening families with young children for child maltreatment potential. *Pediatric Nursing, 26(1)*, 47-54.

Nelson, J. P. (1999). *Development and evolution of the Family Advocacy Program in the Department of Defense.* In J.G. Daley (Ed.), Social work practice in the military (pp. 51-66). New York: Haworth Press.

Norton, L. B., Peipert, J. F., Zierler, S., Lima, B., & Hume, L. (1995). Battering in pregnancy: An assessment of two screening methods. *Obstetrics and Gynecology, 85(3)*, 321-325.

Olds, D., Eckenrode, J., Henderson, C. R., Kitzman, H., Poers, J., Cole, R. et al. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect. *JAMA, 278 (8)*, 637-643.

Orkow, B. (1985). Implementation of a family stress checklist. *Child Abuse & Neglect, 9*, 405-410.

Pan, H. S., Ehrensaft, M. K., Heyman, R. E., O'Leary, K. D., & Schwartz, R. (1997). Evaluating domestic partner abuse in a family practice clinic. *Family Medicine, 29(7),* 492-495.

Pan, H. S., Neidig, P. H., & O'Leary, K. D. (1994). Predicting mild and severe husband-to-wife physical aggression. *Journal of Consulting and Clinical Psychology, 62(5)*, 975-981.

Paranjape, A., & Liebschuz, J. (2003). "STaT": A three-question screen for intimate partner violence. *Journal of Women's Health & Gender-based Medicine, 12(3),* 233-239.

Peralta, R. L., & Fleming, M. F. (2003). Screening for intimate partner violence in a primary care setting: The validity of "feeling safe at home" and prevalence results. *Journal of the American Board of Family Practice, 16*, 525-532.

Pigott, T.D. (2001). A Review of Methods for Missing Data. *Educational Research and Evaluation, 7(4),* 353-383.

Pittman, J. F., & Taylor, L. (2002). *Reanalysis of the USAF FAP Family Needs Screener*. Auburn, AL: Auburn University.

Polansky, N. A., & Gaudin, Jr., J. M. (1992). The Maternal Characteristics Scale: A cross validation. *Child Welfare, 71(3),* 271-280.

Popham, W. J., & Husek, T. R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6,* 1-9.

Pulido, M. L., & Gupta, D. (2002). Protecting the child and the family: Integrating domestic violence screening into a child advocacy center. *Violence against Women, 8(8),* 917-933.

Riggs, D. S., Caulfield, M. B., & Street, A. E. (2000). Risk for domestic violence victimization: Factors associated with perpetration and victimization. *Journal of Clinical Psychology, 56(10)*, 1289-1316.

Rinehart, D. J., Becker, M. A., Buckley, P. R., Dailey, K., Reichardt, C. S., Graeber, C. et al. (2005). The relationship between mothers' child abuse potential and current mental health symptoms: Implications for screening and referral. *Journal of Behavioral Health Sciences & Research, 32(2),* 155-166.

Robertson, K. R., & Milner, J. S. (1983). Construct validity of the Child Abuse Potential Inventory. *Journal of Clinical Psychology, 38(3)*, 426-429.

Rothman, K.J. & Greenland, S. (1998). *Modern Epidemiology* (2nd ed.). Philidelphia: Lippincott Williams & Wilkins.

Rumm, P. D., Cummings, P., Krauss, M. R., Bell, M. A., & Rivara, F. P. (2000). Identified spouse abuse as a risk factor for child abuse. *Child Abuse & Neglect, 24(11)*, 1375-1381.

Sackett, D. L., & Holland. W. W. (1975). Contoversy in the detection of disease. *Lancet, 2(7930)*, 357-359.

Schaeffer, C.M., Alexander, P.C., Bethke, K., & Kretz, L.S. (2005). Predictors of child abuse potential among military parents: Comparing mothers and fathers. *Journal of Family Violence, 20(2),* 123-129.

Schlesselman, J. J. (1982). *Case-Control Studies: Design, Conduct, Analysis*. New York: Oxford University Press.

Sedlak, A. J. (1988). Prevention of Wife Abuse. In V. B. Van Hasselt, R. L. Morrison, A. S. Bellack, & M. Herson (Eds.), *Handbook of Family Violence.* New York: Plenum Press.

Sherin, K. M., Sinacore, J. M., Li, X., Zitter, R. E., & Shakil, A. (1998). HITS: A short domestic violence screening tool for use in a family practice setting. *Family Medicine*, *30(7),* 508-512.

Siegel, R. M., Hill, T. D., Henderson, V. A., Ernst, H. M., & Boat, B. W. (1999). Screening for domestic violence in the community pediatric setting. *Pediatrics*, *104(4),* 874-877.

Spinetta, J. J. (1978). Parental personality factors in child abuse. *Journal of Consulting and Clinical Psychology, 46(6)*, 1409-1414.

Smith, P. H., Earp, J. A., & DeVillis, R. (1995). Measuring battering: Development of the Women's Experience with Battering (WEB) Scale. *Women's Health, 1(4)*, 273-288.

Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M. et al. (in press). Risk factors in child maltreatment: A meta-analytic review of the literature. *Journal of Aggression and Violent Behavior.*

Stith, S. M., Smith, D. B., Penn, C. E., Ward, D. B., & Tritt, D. (2004). Intimate partner physical abuse perpetration and victimization risk factors: A meta-analytic review. *Journal of Aggression and Violent Behavior, 10(1)*, 65-98.

Straus, M. A. (1979). Measuring intrafamily conflict and violence: The Conflict Tactics (CTS) Scale. *Journal of Marriage and the Family, 41,* 75-88.

Straus, M. A., & Gelles, R. J. (1986). Societal change and change in family violence from 1975 to 1985 as revealed by two national surveys. *Journal of Marriage and the Family, 48(3)*, 465-479.

Straus, M. A., & Kantor, G. K. (2005). Definition and measurement of neglectful behavior: Some principles and guidelines. *Child Abuse & Neglect, 29*, 19-29.

Thompson, M. P., & Kingree, J. B. (2006). The roles of victim and perpetrator alcohol use in intimate partner violence outcomes. *Journal of Interpersonal Violence, 21(2)*, 163-177.

Tolan, P., Gorman-Smith, D., & Henry, D. (2006). Family violence. *Annual Review of Psychology, 57*, 557-583.

U.S. Department of Health and Human Services. (2005). *Child maltreatment 2003: Reports from the states to the National Center on Child Abuse and Neglect*. Washington, DC: Government printing office.

U.S. Department of Justice (2003). *Bureau of justice statistics crime data brief: Intimate partner violence, 1993-2001*. Retrieved April 2, 2006, from http://www.cdc.gov/ncipc/pub-res/ipv_cost/ipv.htm.

Wasson, J. H., Jette, A. M., Anderson, J., Johnson, D. J., Nelson, E. C., & Kilo, C. M. (2000). Routine, single-item screening to identify abusive relationships in women. Journal of *Family Practice, 49(11),* 1017-1022.

Weberling, L. C., Forgays, D. K., Crain-Thoreson, C., & Hyman, I. (2003). Prenatal child abuse risk assessment: A preliminary validation study. *Child Welfare, 132(3)*, 319-334.

Weis, J.G. (1989). *Family violence research methodology and design*. In L. Ohlin & M. Tonry (Eds.), Family Violence (pp.117-162). Chicago: University of Chicago Press.

Wilson, J. M. G., & Jungner, G. (1968). *Principles and practice of screening for disease.* World Health Organization, Geneva.

APPENDICES

Appendix A. Tables and Figures.

Table 1. Theories of Violence Used in the Development of the FNS.

Theoretical subtype	Theory	Risk Marker(s) included in FNS
Intraindividual Theories	Psychopathology	Depression
	Alcohol and Drugs	Substance abuse
Social-Psychological Theories	Frustration-Aggression	Violence approval, Family of origin violence
	Social Learning Theory	Violence approval, Family of origin violence
	Self-Attitude Theory	Self-esteem
	Clockwork Orange Theory	Stress
	Symbolic Interaction	None
	Exchange Theory	Relationshp Discord
	Attribution Theory	Relationshp Discord
Sociocultural Theories	Functional Theory of Violence	None
	Culture-of-Violence Theory	Socioeconomic Status
	Structural Theory of Violence	Socioeconomic Status, Stress
	General Systems Theory	No specific risk markers - many causes/roots
	Conflict Theory	Stress
	Intrafamily Resource Theory	Socioeconomic Status, Support, Relationship Discord

Table 2. Current FNS Subscales and Reliabilities of Previous Studies.

Domain	Total Items	Scorable Items	Kantor's Alphas[a]	Pittman's Alphas[a]
Demographics	14	9	.35	.45
Stress	5	5	.69	.69
Relationship Discord	5	5	.85	.85
Support	10	10	.81	.81
Substance Abuse	3	3	.68	.68
Violence Approval	4	4	not reported	not reported
Fam of Origin Violence	6	6	.79	.79
Self-Esteem	5	5	.80	.80
Depression	4	4	.75	.75
Prior Fam Violence	2	2	.32	.31
Total:	58	53	.91	.91

Note: [a] Using the Pilot 2 sample in both studies (N=658).

Table 3. Respondent Groups Items and Range of Scores.

Respondent Group	# items completed (excluding demographics)	Range	# items completed (including demographics)	Range
NoPgNoPt	36 items = (k)	0-36	k + 7	0-43
PgNoPt	39 items = (k + 3)	0-39	k + 10	0-46
NoPgPt	41 items = (k + 5)	0-41	k + 14	0-50
PgPt	44 items (k + 8)	0-44	k + 17	0-53

Note: k is the number of common items answered by all respondents.

Three questions are pregnant specific; Five questions are partner specific; Two demographics questions are partner specific; Seven demographics questions are answered by all respondents.

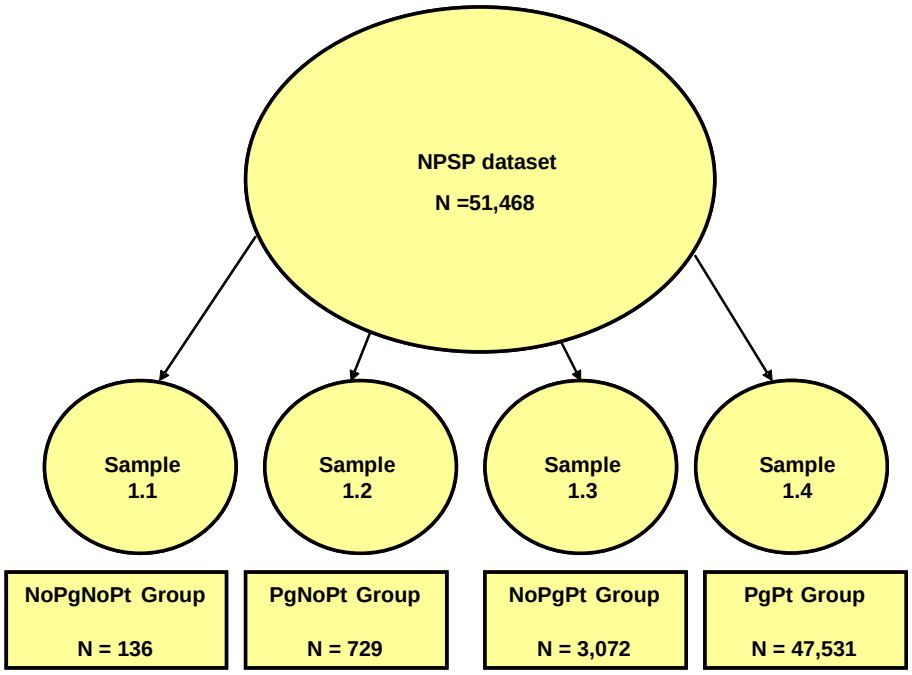

Figure 1. Sample 1.

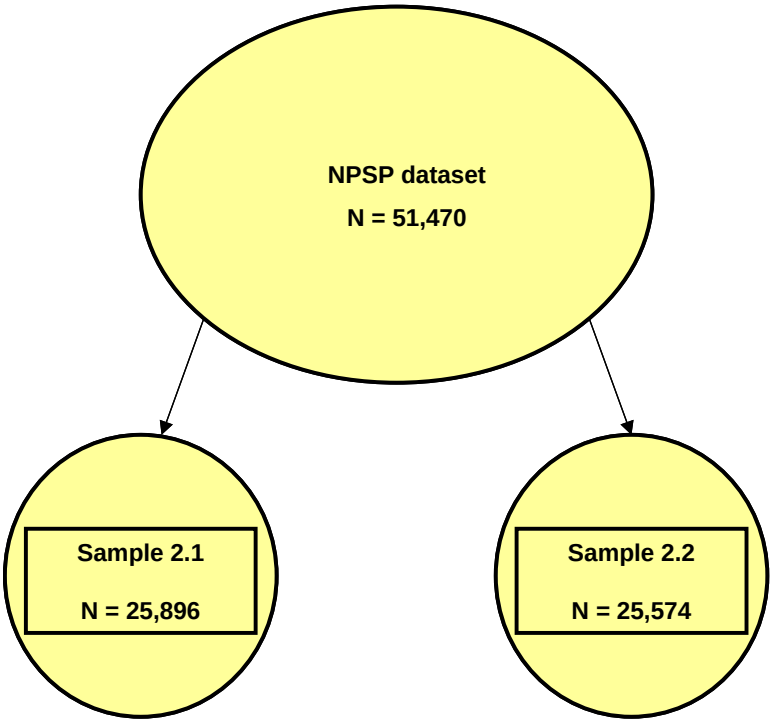

Figure 2. Sample 2.

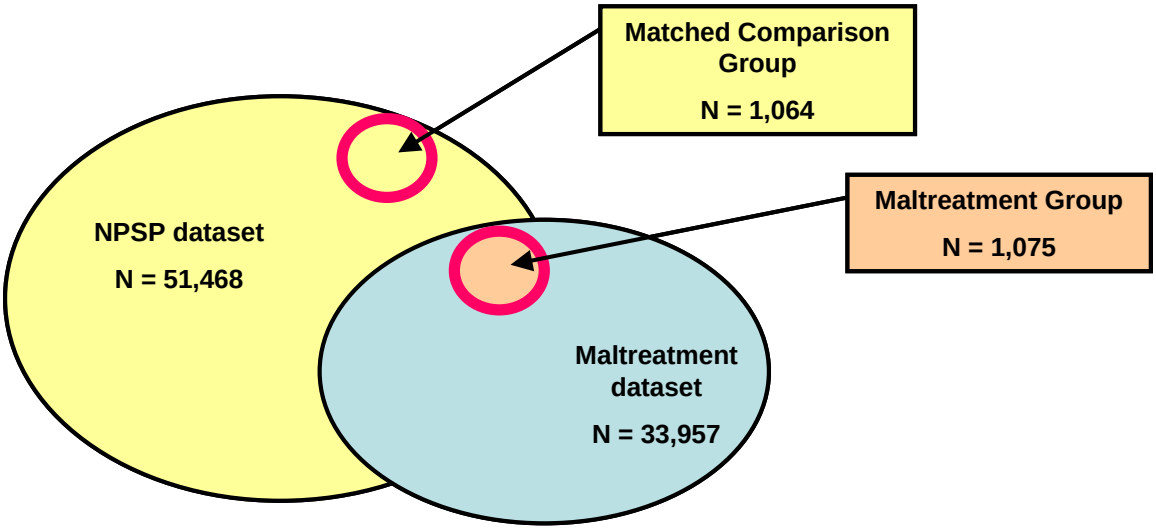

Figure 3. Sample 3.

Table 4. Subscales and Items for Common 36 and Common 30 Models.

Common 36 Model		Common 30 Model	
Subscale	Items Included	Subscale	Items Included
Stress	21,22	Stress	21-23, 25
Support	24, 25, 39, 45-51	Interpersonal Support	45-53
Relationship Discord	23	Substance Abuse	26,27
Substance Abuse	26,27	Violence Approval	30-32
Violence Approval	29-32	Family of Origin Violence	33-38
Family of Origin Violence	33-38	Self Esteem/Depression	40-44, 54
Self-Esteem	40-44		
Depression	52-55		
Prior Family Violence	56,57		

Table 5. Summary of Phases, Analyses, and Samples.

Phase	Analyses	Model Tested	Sample
I	Reliability/Dependability	Respondent Group Models	Sample 1
		Common 36 Model	Sample 2.1
	CFA	Respondent Group Models	Sample 1
		Common 36 Model	Sample 2.1
		Measurement Invariance	Sample 1
	Predictive Validity Analyses	Original FNS	Sample 3
II	EFA	Development of Common 30	Sample 2.1
	Reliability/Dependability	Common 36 Model	Sample 2.2
		Common 30 Model	Sample 2.2
	CFA	Common 36 Model	Sample 2.2
		Common 30 Model	Sample 2.2
	Predictive Validity Analyses	Common 36 Model	Sample 3
		Common 30 Model	Sample 3
		Common 12 Model	Sample 3.1
		Cross-validation of Common 12 Model	Sample 3.2

Table 6. Estimates of Reliability and Dependability in Phase 1.

	With Demographics[c]					Without Demographics[c]				
Model	N	n(i)	Phi	Phi(L=9)	Alpha	N	n(i)	Phi	Phi(L=9)	Alpha
Common36[a]	16891	43	.180	.915	.733	23471	36	.905	.993	.920
PgPt[b]	29776	53	.147	.922	.725	41341	44	.909	.994	.924
NoPgPt	1805	50	.195	.932	.783	2537	41	.909	.994	.929
PgNoPt	162	46	.173	.945	.620	423	39	.853	.993	.897
NoPgNoPt	33	43	.229	.940	.778	99	36	.875	.993	.904

[a] Using Sample 2 (Half 1)

[b] PgPt Model is identical to Original FNS

[c] Missing data not imputed

Table 7. Model Fit for Individual Models and Multi-group Comparison.

Model	N	Chi Sq	df	CFI	SRMR	RMSEA
Common	25896	51227.05***	557	.874	.0490	.059
PgPt	47531	177055.42***	866	.805	.0668	.065
NoPgPt	3072	9908.62***	743	.844	.0608	.063
PgNoPt	729	3199.24***	667	.771	.0681	.072
NoPgNoP	136	1059.98***	557	.780	.0834	.082
Common	25574	45704.74***	390	.871	.0483	.067
Mulit-Group Comparison		Chi Sq	df	CFI	SRMR	RMSEA
Unconstrained		103061.62***	2228	.871	.0489	.030
Measurement weights		104747.22***	2333	.869	.0499	.029
Structural covariances		105396.40***	2429	.869	.0500	.029
Structural residuals		106067.84***	2444	.868	.0500	.029
Measurement residuals		122732.59***	2555	.847	.0502	.030

Note: *** $p < .000$

Table 8. Prediction of Subsequent Abuse Controlling for NSPS Services for All Models Using Standardized Scores.

Step	Variables Entered	Chi-square	df	B[a]	Wald	p	Exp(B)[a]
1	Original FNS	196.92***	1	.669	165.60***	.000	1.95
2	Original FNS	198.15***	2	.655	150.26***	.000	1.93
	NPSP Services			.013	1.21	.271	1.01
1	Common 36 Original Scoring	182.59***	1	.646	152.96***	.000	1.91
2	Common 36 Original Scoring	184.24***	2	.630	138.23***	.000	1.88
	NPSP Services			.015	1.61	.204	1.02
1	Common 36 Raw Sum Scoring	122.16***	1	.525	112.60***	.000	1.69
2	Common 36 Raw Sum Scoring	126.22***	2	.504	99.66***	.000	1.66
	NPSP Services			.024	3.90*	.048	1.02
1	Common 30 Original Scoring	167.27***	1	.612	142.84***	.000	1.84
2	Common 30 Original Scoring	169.54***	2	.594	128.20***	.000	1.81
	NPSP Services			.018	2.21	.137	1.02
1	Common 30 Raw Sum Scoring	115.34***	1	.502	107.21***	.000	1.65
2	Common 30 Raw Sum Scoring	119.30***	2	.482	94.69***	.000	1.62
	NPSP Services			.023	3.81	.051	1.02
1	Common 12 Original Scoring	91.59***	1	.667	73.30***	.000	1.95
2	Common 12 Original Scoring	91.67***	2	.672	70.80***	.000	1.96
	NPSP Services			-.004	.076	.783	.996

Note: [a] Standardized Regression Coefficient and Odds Ratio calculated using z-scores.

Note: *** p < .000, * p<.05

Table 9. Cut-score Analyses for All Models.

Original FNS								
Cut Score	Sensitivity	Specificity	FNR	FPR	PPV	NPV	OCC	Exp(B)
4	0.854	0.321	0.146	0.679	0.560	0.685	0.589	2.77
5	0.775	0.422	0.225	0.578	0.575	0.650	0.599	2.51
6	0.706	0.507	0.294	0.493	0.591	0.630	0.607	2.47
7	0.634	0.609	0.366	0.391	0.621	0.622	0.622	2.70
8	0.557	0.682	0.443	0.318	0.639	0.604	0.619	2.70
9	0.489	0.744	0.511	0.256	0.659	0.591	0.616	2.79
10	0.434	0.794	0.566	0.206	0.681	0.582	0.613	2.96
Common 36 Model								
Cut Score	Sensitivity	Specificity	FNR	FPR	PPV	NPV	OCC	Exp(B)
1	.9340	.1569	.0660	.8430	.5281	.7017	.547	2.63
2	.8326	.3280	.1674	.6720	.5559	.6597	.582	2.43
3	.7358	.4699	.2642	.5301	.5838	.6378	.604	2.47
4	.6419	.5883	.3581	.4117	.6117	.6192	.615	2.56
5	.5591	.6786	.4409	.3214	.6373	.6037	.619	2.68
Common 30 Model								
Cut Score	Sensitivity	Specificity	FNR	FPR	PPV	NPV	OCC	Exp(B)
1	.9116	.1786	.0884	.8214	.5286	.6667	.547	2.24
2	.8121	.3675	.1879	.6325	.5647	.6594	.591	2.51
3	.6986	.5094	.3014	.4906	.5899	.6259	.604	2.41
4	.5953	.6335	.4047	.3665	.6214	.6078	.614	2.54
5	.4958	.7143	.5042	.2857	.6368	.5837	.604	2.46
Common 12 Model								
Cut Score	Sensitivity	Specificity	FNR	FPR	PPV	NPV	OCC	Exp(B)
1	.6189	.5970	.3811	.4030	.5977	.6181	.608	2.41
2	.3949	.8270	.6051	.7130	.6884	.5855	.614	3.21
3	.2515	.9335	.7485	.0665	.7853	.5631	59.8	4.71

Table 10. Reliabilities for Common 30 Model Subscales and Total Scale Using Sample 2.1.

Factors	Initial Items	Alpha
Stress Coping	21,22,23,24,25	.760
Sub Abuse	26,27	.890
Violence Approval	29-32	.731
Fam of Origin Violence	33-38	.834
Self Esteem/Depression	40-44, 54	.833
Interpersonal Support	45-53	.868
Prior FV	56,57	.477
Scale Alpha		.916

Table 11. Estimates of Reliability and Dependability for Common 36 and Common 30 Models using Sample 2.2.

Subscales	Common 36 Model	Common 30 Model
Stress	.711	
Support	.833	
Relation. Discord	N/A	
Sub Abuse	.880	.880
FoO Violence	.837	.837
Viol Approval	.728	.757
Self Esteem	.823	
Depression	.759	
Prior FV	.470	
Stress/Coping		.780
Interpersonal Support		.866
Self Esteem/Depression		.833
Overall Alpha	.920	.918
Phi	.905	.907

Table 12. Model Fit for Common 30 and Common 36 Models.

Model	N	Chi Sq	df	CFI	SRMR	RMSEA
Common 36 (Half #1)	25896	51227.05***	557	.874	.0490	.059
Common 36 (Half #2)	25574	50411.11***	557	.874	.0492	.059
Common 30 (Half #2)	25574	45704.74***	390	.871	.0483	.067

Note: *** p < .000

Table 13. Prediction of Subsequent Abuse Incident Using Subscales from Common 36 Model.

	Chi-square	df		
Common 36 Model	265.58***	9		
Subscale	B[a]	Wald	p	Exp(B)[a]
Stress	.139	3.48	.062	1.15
Relationship Discord	.675	17.21***	.000	1.96
Substance Abuse	-.146	.374	.541	.864
Support	.134	18.95***	.000	1.14
Violence Approval	-.055	.726	.394	.957
Family of Origin Violence	.173	28.94***	.000	1.19
Self-Esteem	-.102	1.73	.188	.903
Depression	.097	1.39	.238	1.10
Prior Family Violence	1.44	41.23***	.000	4.22

Note: [a] Standardized Regression Coefficient and Odds Ratio calculated using z-scores; *** $p < .000$

Table 14. Common 12 Model Items.

Item #	Question
22	At times I feel out of control, like I am losing it.
23	Uncontrolled anger can be a problem in my family.
25	I feel very isolated
26	I sometimes drink enough to feel really high or drunk.
27	I sometimes drink five or more drinks of alcohol at a time, but mostly on weekends.
34	When I was a teenager, I was hit a lot by my mother or father.
35	When I was growing up, I saw my mother or father hit or throw something at their partner.
43	I feel I do not have much to be proud of.
46	There is someone I can talk to openly about anything.
54	There are times when I feel life is not worth living.
56	Have you or your partner been involved in a suspected or verified case of child abuse or neglect?
57	Have you or your partner been involved in a suspected or verified case of spouse abuse?

Table 15. Cross-validation of Predictive Validity of Common 12 Model.

Step	Variables Entered	Chi-square	df	B[a]	Wald	p	Exp(B)[a]
1	Dummy Variable	1.28	1	.101	1.28	.257	1.11
2	Dummy Variable	177.67***	2	.075	.647	.421	1.08
	Common 12 Model			.433	139.61***	.000	1.54
3	Dummy Variable	177.70***	3	.064	.302	.583	1.07
	Common 12 Model			.427	66.13***	.000	1.53
	Dummy*Common12			.012	.027	.868	1.01
1	Dummy Variable	184.95***	3	.063	.265	.606	1.07
	Common12 + Item 21			.351	76.24***	.000	1.42
	Dummy*Comm12+21			.006	.011	.916	1.01
1	Dummy Variable	183.50***	3	.044	.152	.697	1.05
	Common12 + Item 30			.383	68.55***	.000	1.47
	Dummy*Comm12+30			.038	.322	.570	1.04
1	Dummy Variable	191.79***	3	.058	.263	.608	1.06
	Common12 + Item 31			.414	73.14***	.000	1.51
	Dummy*Comm12+31			.023	.108	.742	1.02
1	Dummy Variable	181.70***	3	.043	.141	.707	1.04
	Common12 + Item 32			.379	68.83***	.000	1.46
	Dummy*Comm12+32			.037	.324	.569	1.04
1	Dummy Variable	188.49***	3	.049	.178	.673	1.05
	Common12 + Item 33			.353	73.27***	.000	1.42
	Dummy*Comm12+33			.018	.094	.759	1.02
1	Dummy Variable	198.61***	3	.038	.111	.739	1.04
	Common12 + Item 36			.385	73.91***	.000	1.47
	Dummy*Comm12+36			.028	.194	.660	1.03
1	Dummy Variable	191.44***	3	.033	.080	.778	1.03
	Common12 + Item 37			.348	73.361	.000	1.42
	Dummy*Comm12+37			.032	.293	.588	1.03

Table 15 (continued).

Step	Variables Entered	Chi-square	df	B[a]	Wald	p	Exp(B)[a]
1	Dummy Variable	194.40***	3	.039	.117	.733	1.04
	Common12 + Item 38			.379	72.79***	.000	1.46
	Dummy*Comm12+38			.025	0.162	.687	1.03
1	Dummy Variable	192.25***	3	.053	.225	.635	1.06
	Common12 + Item 40			.411	72.41***	.000	1.51
	Dummy*Comm12+40			.024	.127	.721	1.02
1	Dummy Variable	193.80***	3	.049	.187	.666	1.05
	Common12 + Item 41			.410	72.94***	.000	1.51
	Dummy*Comm12+41			.029	.176	.674	1.03
1	Dummy Variable	181.96***	3	.053	.218	.641	1.05
	Common12 + Item 42			.377	69.44***	.000	1.46
	Dummy*Comm12+42			.012	.036	.848	1.01
1	Dummy Variable	186.42***	3	.042	.143	.705	1.04
	Common12 + Item 44			.385	69.14***	.000	1.47
	Dummy*Comm12+44			.036	.285	.593	1.04
1	Dummy Variable	195.22***	3	0.041	.127	.721	1.04
	Common12 + Item 45			.402	72.92***	.000	1.49
	Dummy*Comm12+45			.028	.175	.676	1.03
1	Dummy Variable	202.03***	3	.063	.310	.578	1.06
	Common12 + Item 47			.409	76.92***	.000	1.51
	Dummy*Comm12+47			.013	.039	.844	1.01
1	Dummy Variable	190.95***	3	.076	.443	.506	1.08
	Common12 + Item 48			.388	74.86***	.000	1.47
	Dummy*Comm12+48			.008	.016	.901	1.01
1	Dummy Variable	196.29***	3	.061	.261	.610	1.06
	Common12 + Item 49			.380	76.35***	.000	1.46
	Dummy*Comm12+49			.012	.039	.843	1.01

Table 15 (continued).

Step	Variables Entered	Chi-square	df	B[a]	Wald	p	Exp(B)[a]
1	Dummy Variable	198.18***	3	.026	.049	.825	1.03
	Common12 + Item 50			.376	73.49***	.000	1.46
	Dummy*Comm12+50			.036	.326	.568	1.04
1	Dummy Variable	198.26***	3	.060	.279	.597	1.06
	Common12 + Item 51			.411	74.95***	.000	1.51
	Dummy*Comm12+51			.016	.055	.814	1.02
1	Dummy Variable	191.09***	3	.054	.228	.633	1.06
	Common12 + Item 52			.378	72.61	.000	1.46
	Dummy*Comm12+52			.020	.097	.756	1.02
1	Dummy Variable	189.91***	3	.054	.231	.630	1.06
	Common12 + Item 53			.404	71.45***	.000	1.50
	Dummy*Comm12+53			.023	.118	.731	1.02

Note:*** p < .000

Appendix B. The Air Force Family Needs Screener.

DATE ___/___/___ NPSP ID#: _ _ _ _ _ _ _ _ _ _
BASE _____

U.S.A.F. Family Advocacy New Parent Support Program
Family Needs Screener

1. What is your military status? (**PLEASE CIRCLE**)
❶ Active Duty Member
❷ Family Member, Spouse
❸ Retired Military
❹ Family Member, Daughter
❺ Other (**SPECIFY**):_____

2. What is the sponsor's military status? (**PLEASE CIRCLE**)
❶ Active Duty
❷ Retired Military
❸ Other (**SPECIFY**): _____

3. What is your marital status? (**PLEASE CIRCLE**)
❶ Single
❷ Married
❸ Divorced
❹ Separated
❺ Widowed

4. What is your current living situation? Are you: (**PLEASE CIRCLE**)
❶ Living together with your partner/spouse
❷ Living alone (or with children only)
❸ Living with your parents (or other adults)
❹ Other living situation (**SPECIFY**):

5. How long have you been living together: _____Years _____Months _____Not Applicable

6. Are you currently pregnant or in the process of adoption? (**PLEASE CIRCLE**)
❶ Yes
❷ No (**GO TO QUESTION 7**)

(a) No. of Weeks Pregnant_____

7. Did you have or adopt a baby over the last 12 months? (**PLEASE CIRCLE**)
❶ Yes
❷ No

➡ GO TO NEXT PAGE

DATE ___/___/___ NPSP ID#: __ __ __ __ __ __ __ __

8. How many children are living with you? (**SPECIFY**): _____

9. Do you have any children living with you who are from a prior relationship? (either yours or your partner's) (**PLEASE CIRCLE**)

❶ Yes
❷ No

10. What is your age? _____

11. What is your partner's age?_____ (**SKIP IF NOT APPLICABLE**)

Ethnic Group

12.Which of these ethnic groups do you and your partner consider yourself? (**PLEASE CIRCLE**)

1.YOU		2.YOUR PARTNER	
❶	Pacific Islander	❶	Pacific Islander
❷	Asian	❷	Asian
❸	Native Amer. Or Alaskan Native	❸	Native Amer. Or Alaskan Native
❹	White but not Latino	❹	White but not Latino
❺	Black but not Hispanic	❺	Black but not Hispanic
❻	Latino or Hispanic	❻	Latino or Hispanic
❼	Multi-racial	❼	Multi-racial
❽	Some other group (**SPECIFY**):_____	❽	Some other group (**SPECIFY**):_____

Education

13.What is the last year of school that you and your partner completed? (**PLEASE CIRCLE**)

1.YOU		2.YOUR PARTNER	
❶	7th Grade or Less	❶	7th Grade or Less
❷	8th Grade	❷	8th Grade
❸	Some High School/GED	❸	Some High School/GED
❹	High School Graduate	❹	High School Graduate
❺	Some College	❺	Some College
❻	College Graduate	❻	College Graduate
❼	Post-B.A. Training	❼	Post-B.A. Training
❽	Advanced Degree	❽	Advanced Degree

➡ GO TO NEXT PAGE

DATE ___/___/___ NPSP ID#: _ _ _ _ _ _ _ _ _

INSTRUCTIONS: FOR EACH QUESTION, PLEASE READ THE FOLLOWING STATEMENTS AND CIRCLE THE BEST RESPONSE

GO TO QUESTION 17 IF YOU ARE NOT CURRENTLY PREGNANT

	Strongly Disagree	Disagree	Agree	Strongly Agree
14. My partner is very supportive of this pregnancy.	❶	❷	❸	❹
15. This is an unplanned pregnancy.	❶	❷	❸	❹
16. This is not a good time for me to have a baby.	❶	❷	❸	❹

GO TO QUESTION 21 IF YOU ARE NOT CURRENTLY IN A RELATIONSHIP

17. My partner treats me well.	❶	❷	❸	❹
18. My partner and I have a very good relationship.	❶	❷	❸	❹
19. I wish my partner and I got along better.	❶	❷	❸	❹
20. I have thought seriously about ending my relationship with my partner.	❶	❷	❸	❹
21. This is a very stressful time for me.	❶	❷	❸	❹
22. At times I feel out of control, like I'm losing it.	❶	❷	❸	❹
23. Uncontrolled anger can be a problem in my family.	❶	❷	❸	❹
24. I only have a few friends/family to help with the baby (my children).	❶	❷	❸	❹
25. I feel very isolated.	❶	❷	❸	❹
26. I sometimes drink enough to feel really high or drunk.	❶	❷	❸	❹
27. I sometimes drink five or more drinks of alcohol at a time, but mostly on weekends.	❶	❷	❸	❹

GO TO QUESTION 29 IF YOU ARE NOT CURRENTLY IN A RELATIONSHIP

28. My partner sometimes drinks five or more drinks at a time, but mostly on weekends.	❶	❷	❸	❹
29. It is sometimes necessary to discipline a child with a good, hard spanking.	❶	❷	❸	❹
30. I can think of a situation when I would approve of a wife slapping a husband's face.	❶	❷	❸	❹
31. I can think of a situation when I would approve of a husband slapping a wife's face.	❶	❷	❸	❹
32. It is sometimes necessary for parents to slap a teen who talks back or is getting into trouble.	❶	❷	❸	❹
33. When I was a child I was spanked or hit a lot by my mother or father.	❶	❷	❸	❹

➡ GO TO NEXT PAGE

DATE ___/___/___ NPSP ID#: __ __ __ __ __ __ __ __

INSTRUCTIONS: FOR EACH QUESTION, PLEASE READ THE FOLLOWING STATEMENTS AND CIRCLE THE BEST RESPONSE

	Strongly Disagree	Disagree	Agree	Strongly Agree
34. When I was a teenager, I was hit a lot by my mother or father.	❶	❷	❸	❹
35. When I was growing up, I saw my mother or father hit or throw something at their partner.	❶	❷	❸	❹
36. My parents helped me when I had problems.	❶	❷	❸	❹
37. I have unhappy memories of my childhood.	❶	❷	❸	❹
38. My parents did not comfort me when I was upset.	❶	❷	❸	❹
39. My income is often inadequate for basic needs (rent, food, clothing, transportation, etc.).	❶	❷	❸	❹
40. I feel that I have a number of good qualities.	❶	❷	❸	❹
41. I feel that I am a person of worth, at least on an equal basis with others.	❶	❷	❸	❹
42. I frequently feel as if I am not as good as others.	❶	❷	❸	❹
43. I feel I do not have much to be proud of.	❶	❷	❸	❹
44. All in all, I am inclined to feel that I am a failure.	❶	❷	❸	❹
45. Someone I'm close to makes me feel confident in myself.	❶	❷	❸	❹
46. There is someone I can talk to openly about anything.	❶	❷	❸	❹
47. There is someone I can talk to about problems in my relationship.	❶	❷	❸	❹
48. I have someone to borrow money from in an emergency.	❶	❷	❸	❹
49. I have someone to take care of my child/children for several hours if needed.	❶	❷	❸	❹
50. I have someone who helps me around the house.	❶	❷	❸	❹
51. I have someone I can count on in times of need.	❶	❷	❸	❹
52. I usually wake up feeling pretty good.	❶	❷	❸	❹
53. I think good things will happen to me in the future.	❶	❷	❸	❹
54. There are times when I feel life is not worth living.	❶	❷	❸	❹
55. I feel sad quite often.	❶	❷	❸	❹

	YES	NO
56. Have you or your partner been involved in a suspected or verified case of child abuse or neglect?	❶	❷
57. Have you or your partner been involved in a suspected or verified case of spouse abuse?	❶	❷

END OF QUESTIONNAIRE *THANK YOU*

Page 4 of 4

Appendix C. The FNS Scoring Sheet.

Date: _____ Case Id: _____

U.S. Air Force Family Advocacy New Parent Support Program Family Needs Screener Scoring Sheet

Instructions for scoring the USAF Family Advocacy New Parent Support Program Family Needs Screener. Numbered Items on the Family Needs Screener instrument correspond to those in the "ITEM" column below. If the person answered the questionnaire item with the responses found next to the items below place a "1" in the "Score" column. For example, if a person answered question 1 below with either "1" (Active Duty) or "4" (Dependent Daughter), enter 1 on the same line under the "Person's Score" column. If the person answered with any other option, place a 0 in the "Score" column. For items that say "Do not Score", or which were omitted by a client, then place a dash as needed. Add the 1's in the "Score" column to obtain the total needs score.

Item	Grading Criteria	Answer	Score
1	1-4=1		
2	Do not score		
3	1,3,4,5=1	2	
4	2=1		
5	Less than 1 year=1		
6	Do not score		
7	Do not score		
8	Do not score		
9	"yes"=1		
10	Less than 20=1		
11	Less than 20=1		
12	Do not score		
13.1	(you) 1,2,3=1		
13.2	(partner) 1,2,3=1		
14	1,2=1		
15	3,4=1		
16	3,4=1		
17	1,2=1		
18	1,2=1		
19	3,4=1		
20	3,4=1		
21	3,4=1		
22	3,4=1 (HIGH NEED)		
23	3,4=1 (HIGH NEED)		
24	3,4=1		
25	3,4=1		
26	3,4=1		
27	3,4=1		
28	3,4=1		

Item	Grading Criteria	Answer	Score
29	3,4=1		
30	3,4=1		
31	3,4=1		
32	3,4=1		
33	3,4=1		
34	3,4=1		
35	3,4=1		
36	1,2=1		
37	3,4=1		
38	3,4=1		
39	3,4=1		
40	1,2=1		
41	1,2=1		
42	3,4=1		
43	3,4=1		
44	3,4=1		
45	1,2=1		
46	1,2=1		
47	1,2=1		
48	1,2=1		
49	1,2=1		
50	1,2=1		
51	1,2=1		
52	1,2=1		
53	1,2=1		
54	3,4=1 (HIGH NEED)		
55	3,4=1		
56	1=1 (HIGH NEED)		
57	1=1 (HIGH NEED)		

TOTAL NEEDS SCORE

Needs Subscales

	Category	Total
A.	Demographics (1-13.2)	0
B.	Stress (14-16; 21, 22)	0
C.	Relationship Discord (17-20, 23)	0
D.	Support (24, 25; 39, 45-51)	0
E.	Substance Abuse (26-28)	0

	Category	Total
F.	Violence Approval (29-32)	0
G.	Family of origin Violence and Neglect (33-38)	0
H.	Self Esteem (40-44)	0
I.	Depression (52-55)	0
J.	Prior Family Violence (56,57)	0

VITA

Major Wendy J. Wyse graduated from Grand Valley State University in Allendale, Michigan in 1992 with a Bachelors Degree in Health Sciences. Major Wyse transferred to George Williams College at Aurora University, Aurora, Illinois, in 1993 where she worked as a Graduate Assistant for Residence Life while working towards her Masters Degree in Social Work. She obtained her MSW in 1996 having completed her thesis on the effects of a school-based prenatal curriculum on incidences of low birth weight deliveries. Major Wyse was hired as a full-time medical social worker at Provena Mercy Center Hospital where she worked for two years before joining the United States Air Force. Major Wyse received her commission in August of 1998 to the Biomedical Sciences Corp. She was awarded the Hoyt D. Vandenburg Award for Academic Excellence during Commissioned Officer Training. Following training, Major Wyse was assigned to Kelly Air Force Base in San Antonio, Texas where she was appointed the Chief of the Family Advocacy Program. After transferring to Lackland Air Force Base the following year, she held several different positions including the Assistant Director of Family Advocacy, Medical Social Worker for the Pediatric and Pediatric Intensive Care Units, and Assistant Director of Alcohol and Drug Addition and Treatment Program. In 2001, Major Wyse was assigned to Travis Air Force Base, California, where she continued working as Chief, Family Advocacy Program. From February to July of 2003, Major Wyse was deployed to the United Arab Emirates in support of Operation Iraqi Freedom. At an austere location, Major Wyse provided mental health clinical services to a population of 1000 Air Force, Air National Guard, and Reserve personnel. Following her deployment, she was one of two officers selected to obtain a doctorate degree in social work. Major Wyse began Ph.D. studies at the University of Tennessee, Knoxville,

in August of 2004 and was the co-author for two publications while completing her degree requirements. Major Wyse married Todd Travis in April of 2007 and the couple will re-locate to Warner Robins, Georgia in June of 2007 where Wendy will assume duties as the Mental Health Flight Commander at Robins Air Force Base, Georgia.