



U.S. Department of Health and Human Services
Assistant Secretary for Planning and Evaluation

**RELIABILITY AND VALIDITY
OF THE NATIONAL INCIDENCE
OF CHILD ABUSE AND
NEGLECT STUDY CONDUCTED
BY WESTAT ASSOCIATES
IN 1988:**

METHODOLOGICAL REVIEW

August 1989

Office of the Assistant Secretary for Planning and Evaluation

The Office of the Assistant Secretary for Planning and Evaluation (ASPE) is the principal advisor to the Secretary of the Department of Health and Human Services (HHS) on policy development issues, and is responsible for major activities in the areas of legislative and budget development, strategic planning, policy research and evaluation, and economic analysis.

ASPE develops or reviews issues from the viewpoint of the Secretary, providing a perspective that is broader in scope than the specific focus of the various operating agencies. ASPE also works closely with the HHS operating divisions. It assists these agencies in developing policies, and planning policy research, evaluation and data collection within broad HHS and administration initiatives. ASPE often serves a coordinating role for crosscutting policy and administrative activities.

ASPE plans and conducts evaluations and research--both in-house and through support of projects by external researchers--of current and proposed programs and topics of particular interest to the Secretary, the Administration and the Congress.

This report was prepared under contract between HHS's Office of Social Services Policy (now the Office of Disability, Aging and Long-Term Care Policy) and Systemetrics. For additional information about this subject, you can visit the ASPE home page at <http://aspe.hhs.gov>. The Project Officer was Karl Ensign.

**RELIABILITY AND VALIDITY OF THE NATIONAL
INCIDENCE OF CHILD ABUSE AND NEGLECT STUDY
CONDUCTED BY WESTAT ASSOCIATES IN 1988:
Methodological Review**

Dr. Deborah Daro
Dr. Elizabeth D. Jones
Karen McCurdy

National Committee for Prevention of Child Abuse

August 1989

Prepared for
Office of the Assistant Secretary for Planning and Evaluation
U.S. Department of Health and Human Services

The opinions and views expressed in this report are those of the authors. They do not necessarily reflect the views of the Department of Health and Human Services, the contractor or any other funding organization.

TABLE OF CONTENTS

| | |
|-----------------------------------------------------------|------|
| OVERVIEW | 1 |
| NATIONAL INCIDENCE STUDY BACKGROUND | 2 |
| KEY FINDINGS AND SUGGESTED CHANGES FOR NIS-3 | 3 |
| SPECIFIC METHODOLOGICAL CONCERNS | 7 |
| Case Duplication..... | 7 |
| Appropriateness of the NIS-2 Sample Counties | 10 |
| Appropriateness of the Sample Non-CPS Agencies | 10 |
| Appropriateness of Sample Sentinels..... | 11 |
| Appropriateness of the Case Weights Utilized..... | 11 |
| Appropriateness of NIS-2 Sample Period | 12 |
| Appropriateness of Weight Trimming..... | 12 |
| SPECIFIC ANALYTIC CONCERNS | 13 |
| CPS Awareness Levels | 13 |
| Case Verification by CPS..... | 17 |
| Subpopulations of Maltreatment | 20 |
| COMPARISONS BETWEEN NIS-1 AND NIS-2 | 25 |
| Are Child Abuse Rates Going Up?..... | 25 |
| CONCLUSIONS | 31 |
| BIBLIOGRAPHY | 33 |
| APPENDICES | |
| APPENDIX A. Consultant Reports | A-1 |
| APPENDIX B. List of Respondents | A-19 |

LIST OF TABLES

| | |
|------------------------------------------------------------------|----|
| TABLE 1: Countable Cases: National Estimates | 10 |
| TABLE 2: CPS Awareness by Demographic Characteristics..... | 15 |
| TABLE 3: CPS Status of Case by Demographic Characteristics | 19 |
| TABLE 4: Types of Abuse by Demographic Characteristics | 22 |
| TABLE 5: Hypotheses Explaining NIS-2 Findings | 27 |

OVERVIEW

In early 1989, the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Department of Health and Human Services, entered into a contract with SysteMetrics and the National Committee for Prevention of Child Abuse (NCPCA) to conduct a methodological review of the Study of National Incidence and Prevalence of Child Abuse and Neglect (NIS-2) completed by Westat Associates in 1986. That study examined the number of child abuse cases recognized by a range of social service, health care and law enforcement professionals in a random sample of counties throughout the United States. The methodology applied in this study paralleled a process undertaken by these same researchers in 1980, thereby making it possible to compare child abuse rates over time. A formal contract to conduct a review of the 1988 study and comparability of its eventual estimates to the figures generated in the 1980 study was awarded by ASPE in March, 1989 to SysteMetrics and NCPCA.

The contract involved three major activities:

- a methodological review of the NIS-2 by two expert sampling statisticians to determine the validity of the procedures undertaken by Westat Associates and the reliability of the estimates made with respect to the incidence of maltreatment;
- informal discussions with several child abuse researchers and policy makers to identify alternative hypotheses for the observed 66% increase in maltreatment between 1980 and 1985; and
- a secondary analysis of the NIS data to determine its utility to address key policy and program issues.

The purpose of this report is to summarize the findings generated by these three activities and to highlight the implications of these findings on further federally-funded national incidence studies.

NATIONAL INCIDENCE STUDY BACKGROUND

The Study of National Incidence and Prevalence of Child Abuse and Neglect (NIS-2) was commissioned by the National Center on Child Abuse and Neglect (NCCAN) in response to a specific Congressional mandate in the Child Abuse Amendments of 1984 (P.L. 98-257). The purpose of this study was to assess the current national incidence of child abuse and neglect and to determine how the severity, frequency, and character of child maltreatment changed since the completion of a similar study in 1980 (NIS-1).

Both the 1986 and 1980 study followed essentially the same design. Data were collected concerning cases of child maltreatment which were recognized and reported to the study by "community professionals" in a national probability sample of 29 counties throughout the United States. These professionals included the local Child Protective Services (CPS) staff in these counties as well as key respondents in a variety of other non-CPS settings such as schools, hospitals, police departments, juvenile probation authorities, day care centers, and mental health agencies. Participating professionals served as "sentinels" by remaining on the lookout during the study's three month data collection period for cases meeting the study's definitions of child maltreatment.

The most recent National Incidence Study employed two definitions of maltreatment. The more restrictive one, which parallels the definitions used in the 1980 study, counted only those cases in which a child had suffered observable harm as a result of abusive or neglectful behavior: Under the second set of definitions, countable cases also included those children who were endangered but not necessarily harmed, as a result of maltreatment. Under the first set of definitions, more than one million children, or 16.3 per 1,000 are estimated to be abused or neglected annually. Applying the second set of definitions, this number increases to 1.6 million, or 25.2 per 1,000.

The over one million estimated child abuse and neglect victims in the 1986 study represent a 66% increase over the number identified in the 1980 study. In reporting this finding, the Westat study team suggested that this increase was probably due more to an increase in the recognition of child maltreatment by community professionals than to any increase in the actual incidence. The research team based this conclusion on two observations: the emphasis in the 1980's on community awareness of the existence of abuse and neglect as well as the need to report suspected maltreatment, and the fact that the greatest increase was in moderate abuse and child sexual abuse, types of maltreatment particularly sensitive to this type of awareness building.

KEY FINDINGS AND SUGGESTED CHANGES FOR NIS-3

This analysis identified a number of cautions to bear in mind when interpreting the NIS study findings or utilizing this data base to addressing critical policy and program concerns. As an incidence study, the consulting statisticians noted the following methodological concerns:

- The inability of NIS-2 to specify the extent of duplication among reported cases, particularly in the sample's large urban counties, introduced an unquantified but potentially significant degree of upward bias into the final estimates, especially in NIS-2.
- The multiple sampling strategies employed at every level of the NIS-2 (i.e. the selection of counties, selection of specific agencies, selection of specific respondents within these agencies, and, in the largest counties, the selection of only a sample of reported cases) further complicated the ability to address the duplication issue and to generate reliable estimates of the total scope of the problem.
- To arrive at national estimates from NIS-2 sampling, the study applied multiple weights to each observation in accordance with the probability of having selected the source who reported it to the study population (i.e. the probability that other professionals working in similar agency settings would identify a similar number of maltreatment cases). This procedure, while necessary, can only be accomplished by making numerous assumptions that the sampling was representative. As discussed below, the statisticians did not feel the extent and direction of the bias introduced by these assumptions was adequately quantified.
- The study contains no specific information on the experience or educational levels of the professionals participating in the study nor on the policies with respect to reporting or staff training operating at the agency level, making it impossible to address in any empirical sense the relationship between professional education or awareness and recognition of maltreatment.

Both reviewers suggested that a very rigorous methodological review be conducted prior to authorizing another national incidence study. It is suggested that an in-depth review by a panel of experts be conducted to determine the most appropriate methods for conducting a national incidence study on such a complex problem as child maltreatment. The panel should consider:

- expanding the types of sentinels utilized in the study to capture a more complete pool of key professionals who have frequent contact with children

- of all ages, such as pediatricians, public health nurses, youth workers, and mental health professionals;
- incorporating non-professional respondents into the sampling frame;
- altering the sample selection process so agencies would be sampled in direct relation to the size of the county in which they are located;
- altering the data collection period.

To assist in the development of NIS-3 it is also recommended that an in-depth review in a limited number of counties be conducted to determine in these few counties the:

- universe of all potential respondent agencies;
- magnitude of duplication rates;
- level of trained and independent knowledge regarding child abuse and neglect reporting requirements and the indicators of maltreatment;
- specific agency-level policies regarding maltreatment reporting, staff training, and reporting responsibilities;
- impacts of state and county policies on the reporting practices of local professionals.

It should be noted, however, that the introduction of any of these changes into subsequent NIS research designs will influence the extent to which one can make direct comparisons with the incidence levels identified in NIS 1 and 2.

Over and above these concerns, lies the broader question of the NIS-2's ability to address emerging policy and program concerns facing the child welfare system. There may well exist an inherent dilemma in trying to address policy questions with a data base developed to look at incidence. For example, it may be useful to shift the emphasis from incidence to incorporate the different aspects of the dynamic processes of CPS agencies which directly affect whether a case is reported or not. As structured, the data base does offer an opportunity to explore the unique characteristics of maltreatment within specific subpopulations. As a sample of over 5,000 child abuse and neglect cases, the data base provides rich information on the types of cases professionals in different settings are observing and the descriptive characteristics of various types of maltreatment. Because of the complex weighting system utilized and the observed differences in the types of cases found in large, medium and small counties, however, the data base is less useful for determining national practice and average performance. Specific changes which would increase the utility of the incidence data to address policy questions include:

- collecting more specific data on an agency's policy with respect to observed cases of maltreatment. For example, many schools require teachers to report suspected cases to the principal or school counselor rather than directly to CPS;
- collecting more information about a professional's level of training and their perceptions regarding the adequacy of this training;

- collecting information which reflects what actually happens to a case once it is reported, including how long it takes to go through the system, which agencies become involved, etc.;
- lengthening the data collection time period to avoid the need for annualized weights;
- collecting explicit information from the sentinels regarding the action they took in response to a suspected child maltreatment incident (e.g., did they report the case?).

Finally, the extent of secondary analysis of any data base is highly correlated to the time it takes to understand the data structure. There were several problems with the NIS-2 data as it was released. First, there are several types of missing data which are not delineated in the code book. For at least six variables, there are values in the raw data which are not coded or represented in the codebook. These variables are: mother's employment (162 cases), father's employment (1477 cases), child's sex (11 cases), family income (142 cases), AFDC (56 cases) father's age (1614 cases) and mother's age (310). Using SPSS-X, the values appear as "system missing". These values could be either missing or not applicable. Preliminary analysis suggested that the majority of these were, in fact, not applicables, but there was not enough information to be certain. For at least 15 other variables, there are values which are coded as blank yet are not adequately explained in the codebook. Further, variables not on a specific form were given the code of "X". These variables needed to be converted to alphanumeric before they could be analyzed in SPSS-X.

Second, the use of different data collection forms, such as the non-CPS, the CPS short form and the CPS long form, further complicated the analyses. The forms did not collect the information in the same manner and they used two different referent points. The non-CPS form bases its questions on the child. The CPS form bases its questions on the adults. Consequently, there is not a direct measure on some very crucial variables. For example, with regard to the child's ethnicity, the non-CPS form collected the ethnicity of the child; the CPS long-form did not ask for the child's ethnicity but asked for the mother/substitute ethnicity. The same is true for household structure. The non-CPS form asked about the relationship of the adults to the child, while the CPS form records information about whether the mother/substitute or father/substitute is in the home. To obtain information on child's ethnicity and household structure for the entire sample, one has to use a combination of variables on the two different forms. These measures are indirect.

Third, some of the evaluative coding was not clearly justified. For example, in Westat's computed variables for type of abuse, it is not clear why cases which were out of scope with respect to time but were on CPS long forms were included in the countable definition. It leads one to suspect that CPS forms were given more credibility than the non-CPS forms. A number of decisions were made and the data went through a number of transformations before it actually made it to the tape. In a few instances, especially with regard to the perpetrator variables, it appeared that even though the

information was collected on both forms, not all the information was included on the data tape.

Finally, the use of the sampling weights severely limits the use of the NIS-2 as a public use data set. The primary shortcoming is that, according to Westat, a special software package is needed for any multivariate analysis to obtain the correct standard errors. Unlike the standard social science software packages, the special ones are likely to be difficult to obtain and use without comprehensive documentation. As a result, many analysts will shy away from using the data. Moreover, while it is clear that the weights are necessary to determine incidence, it is unclear that the use of the sampling weights is necessary in others cases. It depends on the question to be answered and the way the model is specified. Statistically, if the weights are a function of the independent variables in the model, then using the unweighted estimates yield consistent estimates of the true regression slope. However, in situations where the weights are a function of the dependent variable, then it is best to use the weights to obtain consistent estimates.¹ Thus, if one is interested in predicting the demographic determinants of abuse, using the weights is not necessary because the types of abuse are not related to the sampling design. In addition, while many national data sets include sampling weights related to their sampling design (e.g., the National Longitudinal Survey of Youth over sampled minorities and disadvantaged whites), most analysts do not use them. Rather, the analysts focus on specifying the correct functional form of their models, a process which precludes the need for sampling weights.

The following suggestions would make the NIS-2 data a more appealing public use data set:

- define all values for all variables using numeric characters including a code for missing, valid skips, invalid skips etc.;
- include all raw data on the data tape in addition to the cleaned or created variables;
- collect all the data on the same data collection instrument; and
- include more documentation in the codebook to give rationale for recoding variables a specific way.

¹ Personal communication with Chris Winship, Sociology Department, Northwestern University.

SPECIFIC METHODOLOGICAL CONCERNS

Shortly following the contract award, NCPA established formal contracts with Dr. Tom Marx, an independent sampling statistician and Dr. Martin Frankel, director of sampling for the National Opinion Research Corporation, University of Chicago to conduct formal reviews of Westat's procedures. The full written reports submitted by the statisticians are found in Appendix A. The purpose of this section is to briefly summarize these comments as they relate to the points outlined above and to incorporate, where appropriate, clarifications made by the Westat team regarding the procedures they followed.

Case Duplication

Of particular concern to the statistical review team was the question of duplication found among the reported cases both between CPS and non-CPS agencies as well as within a CPS agency. As one might imagine, an abused or neglected child might well be identified by a number of professionals over a given period of time. An obvious source of this type of duplication would be a child reported to CPS during the Westat study period by a hospital social worker. Both the social worker and the CPS caseworker would have completed case forms on the child, thereby potentially generated two "incidence" of maltreatment when in fact there was only one. In addition, a child might have been reported to CPS two or more times during the study period. Because the incidence study is designed to estimate the number of Children abused or neglected in a given year rather than the annual number of maltreatment episodes, correcting for this type of "within CPS" duplication is critical.

The Westat study team devoted a good deal of time identifying both types of "duplicate" cases and insuring that the national incidence estimates were based only on "unduplicated" reports. In all counties, duplicate reports were identified and resolved in a uniform manner so that each child was reported on only one data form. Even in the large counties, where sampling was heaviest, Westat unduplicated all children who were reported on more than one data form, whether they were duplicated on more than one CPS form, on more than one non-CPS form or had been reported on both a non-CPS and a CPS data form. However, a number of logistical barriers precluded the absolute identification of all duplicate cases, particularly in the larger counties. These barriers include:

- The sampling of CPS cases within large counties. This strategy meant that in several counties only a small fraction of the reported cases (less than 5%) were entered into the NIS-2 data base. Consequently, it is not known whether a case might have emerged as a duplicate report had the study included a more representative sample of all reported CPS cases.

- The sampling of non-CPS agencies within counties which did not cover similar geographic regions within the county. The Westat study team assumed that the duplication issue was of greater magnitude between CPS and non-CPS agencies than among non-CPS agencies. While this assumption might well be valid, studies conducted on child abuse fatality cases suggest that local hospital, school district and community-based agency personnel are often all involved with chronic abusive or neglectful families. The National Incidence Study offered no systematic way to test the duplication across non-CPS agencies because the relatively small sample of agencies included in the study were rarely located in similar geographic areas. The sample school districts rarely served the same neighborhoods as the sample hospitals, sample day care centers, sample mental health centers, or sample police departments.
- The data collection period. In general, the longer the observation period, the higher the likelihood for duplication either within or across agencies. For example, a recent longitudinal study on the reincidence rates in one Indiana county conducted by Dean Knudsen at Purdue University found that almost one-quarter of all cases reported to CPS in a given calendar year involve children already reported at least once in that year. If the observation period is extended to include both the present and prior calendar year, this percentage increases to over 40%.²

In essence, these conditions resulted in an inability to determine the exact extent of the "between CPS and non-CPS agencies" duplication in the large counties; the "within CPS" duplication in the large counties; and the "among non-CPS agencies" duplication that does not also overlap with CPS in counties where non-CPS agencies were sampled in more than one category (i.e. the medium and large counties). Westat addressed the first of these issues through weighting procedures. Specifically, whenever a child had been reported to CPS by a source that was also participating in the NIS study, the case was assigned a case rate of "1" in the total county estimates (i.e. it was not weighted up to represent anything other than itself). This strategy assumes that any CPS case which could have been reported to the study by a non-CPS participant actually was reported and would have been identified as such had the NIS study sample included all CPS cases. With respect to the second issue, Westat did consider the potential for repeated reports in annualizing the three-month data. The annualization factors utilized in this study were based on NIS-1 information and inherently entail some assumptions about the likelihood of repeat reports. No additional attempts were made to compensate for multiple reports on the same child to CPS in a given year nor to address the issue of "among non-CPS agencies" duplication. Both of these issues, however, were considered by Westat to involve only trivial rates of duplication and, therefore, to have only marginal impacts on the final estimates.

² Because this study is based on the experience of a single county, one should be cautious in interpreting the results as indicative of national practice.

These strategies aside, the consultants concluded that the study team's estimates based upon the sample drawn did suffer from an upward bias, noting that the magnitude of the bias is virtually impossible to define. For example, the application of the NIS-1 annualization figures, while potentially the only estimate of repeat reports available, might well be an underestimate of the problem. If one assumes higher recognition rates among professionals, as the Westat study team did in explaining the observed increase in maltreatment between 1980 and 1986, one also might assume increases in the chances of an abused child being reported repeatedly. By applying the 1980 annualization figure, the Westat study team might have underestimated the rate of duplication and therefore overestimated the actual incidence rate. Similarly, the more intensive sampling in the NIS-2 study, particularly in the large counties, compounded the ability of the study to identify duplications "among non-CPS agencies" at a time of increased education and awareness of child abuse in several professions. As summarized by Dr. Frankel: "The problem with the approach used is that no attempt has been made to quantify the magnitude of the bias. We are certain that the procedure is biased, but we do not know the size of the bias (statistical bias). We cannot assume that the impact of the bias is small."

In response to this conclusion, the Westat study team decided to further examine their assumption that the undetected duplication was trivial. Specifically, the study team analyzed the full set of countable data forms in both the NIS-1 and NIS-2 to see what the national estimates would be if no unduplication had been performed in either study and to examine the degree to which the duplicated estimates were reduced in each study. As summarized in Table 1, the procedures used for unduplication in NIS-1 reduced the total number of maltreated children to 625,063, or 61% of the duplicated total. The NIS-2 procedures reduced the countable cases to 1,025,168, or 66% of the duplicate total. The Westat research team concluded that given the comparability between the two rates, the upward bias introduced into the 1988 study due to a failure to unduplicate all reports was minimal. If the duplicated 1988 totals were reduced by only the 61% rate observed in 1980, the final estimated number of abuse and neglect cases would have been 949,375, only 7% lower than the figure obtained.

This analysis suggest that the issue of duplication, while a serious theoretical concern, might have had limited practical implications on the established national estimates. However, this conclusion is based upon the assumption that minimal changes occurred in the frequency of duplicated reports between NIS-1 and NIS-2. As noted above, this assumption can be questioned given the increased awareness of maltreatment and the professional training which occurred during the 1980's. Further, the slightly higher duplication rate noted in the NIS-2 was achieved with more heavily sampled data, leading one to assume that the difference would have been more pronounced with more comprehensive samples in the larger counties. While it is highly improbable that the failure to account for all duplication explains the increase noted in child abuse and neglect incidence rates between NIS-1 and NIS-2, this issues remains a critical one for future incidence studies.

Appropriateness of the NIS-2 Sample Counties

Both statisticians indicated that the 29 counties included in the National Incidence Study represent both an accurate and representative sample of U.S. counties. While one of the reviewers suggested that the sample counties might be examined for the extent to which they represent other characteristics related to maltreatment (e.g. income distribution, racial composition, etc.), both agreed that the sample was representative along the key dimensions of interest, namely geographic location and population density.

| TABLE 1: Countable Cases: National (Ratio) Estimates | | | |
|-------------------------------------------------------------|------------------------|--------------------------|---------------------------------------|
| | Duplicated Data | Unduplicated Data | |
| | | Forms | “Awareness” Credited Cases |
| CPS | | | |
| NIS-1 | 208,314 | 188,991 | 203,655 |
| NIS-2 | 403,017 | 314,369 | 408,689 |
| NON-CPS | | | |
| NIS-1 | 812,447 | 436,071 | 421,408 |
| NIS-2 | 1,153,335 | 710,799 | 616,479 |
| TOTAL | | | |
| NIS-1 | 1,020,761 | 625,063 | |
| NIS-2 | 1,556,352 | 1,025,168 | |

Appropriateness of the Sample Non-CPS Agencies

As outlined earlier, professionals working in a wide range of community based agencies were utilized in the NIS-2 as "sentinels" to identify cases of maltreatment during the study period. Both of the reviewers expressed reservations regarding the extent to which the non-CPS agencies included in the sample were representative of the potential universe of such agencies in each county. Particular concern regarding this issue was noted in the case of day care providers, mental health agencies and social service agencies, where the total universe of providers was virtually impossible to confirm.. While failing to quantify the universe of potential contributors to a given sample rarely impedes social science research, such knowledge' is more desirable when using a sample to quantify incidence levels than when using a sample to address more descriptive questions such as normative professional practices. The sampling procedures that Westat followed were generally well documented. However, both reviewers cited limitations with the methods used, limitations which might have notable impacts on how one weights the identified cases. This issue is of particular concern in the large counties where a much smaller percentage of the identified universe of agencies were generally selected. For example, the study allowed for a maximum sampling frame of ten schools, 4.5 day care centers, five hospitals, and four mental health agencies regardless of county size. Given the wide variation in child abuse awareness and service levels found among the hospitals, school districts, and community-based agencies within the largest counties in the sample, surveying less

than 5% of the total universe of respondents runs the risk of drawing a pool of respondents unrepresentative of the professional groups being tapped.

Appropriateness of Sample Sentinels

The reviewers expressed similar concerns with respect to the selection of specific respondents within a given agency. For example, social workers were used as the respondents in hospitals rather than emergency room personnel. While it is true that social workers are most likely to be aware of most child abuse cases identified at a hospital, they generally become involved only after other hospital personnel have made the decision to report a case. Given this pattern it is, perhaps, not surprising that the formal reporting rate among cases identified by hospital personnel was among the highest of all professional groups included in the sample (i.e. 66%). By the time a case has come to the attention of a hospital social worker, at least one other professional, be it a nurse or physician, has made the decision to take some action on a given case. The potential bias in this selection of key informants is that the study provides virtually no estimate of the number of cases emergency room staff or other medical personnel observe but fail to report to hospital social workers. Given the behavior identified by other professional groups in the Westat sample, one might assume that this downward bias is rather significant. For example, classroom teachers were used as the only key informants in day care centers and as one of four respondent categories in local school districts. As reported by Westat, these "front line" workers were considerably less likely to formally report an identified case than hospital social workers. Only 16% of the cases identified by day care providers and 24% of the cases identified by school personnel were formally reported.

Appropriateness of the Case Weights Utilized

The National Incidence Study employed a complex, multi-level system of weights in estimating the incidence of abuse and neglect. Each countable case or case which met the study's definition of maltreatment was weighted with respect to a number of properties. A final case weight, derived from the individual weights, was generated to estimate the number of cases with these given properties one might expect to see if a national census of all children were taken. In other words, the national incidence estimates are based on a weighted average of a sample of cases known to a sample of professionals working in a sample of counties. Both of the reviewers raised serious concerns that this multiple weighing procedure adequately compensated for the inherent bias recognized in the many assumptions that were made in generating the sample.

Appropriateness of NIS-2 Sample Period

One of the reviewers questioned the appropriateness of the abbreviated data collection period utilized in NIS-2. Unlike NIS-1 which collected data for a full year, the NIS-2 data collection period covered three months for most sentinels and ten weeks for school personnel. In projecting annual incidence rates based upon this shortened data collection period, the Westat study team used the pace of reporting documented in the NIS-1. This procedure introduces an additional assumption into the study design.

Appropriateness of Weight Trimming

Similarly, the use of weight trimming at the final stages of the analysis may also have altered the national incidence estimates. Under this procedure, the study team trimmed back the total case weights to 2,000 (before annualization) in those instances where the value exceeded 2,000. While this process is well within the range of acceptable statistical adjustments, the process does introduce a downward bias in the final estimates. Given that the average weight applied to cases in this study was 158.3 (before annualization), those cases with "trimmed" weights had a substantial impact on the study's final estimates. However, only 21 cases were "trimmed" in this manner.

SPECIFIC ANALYTIC CONCERNS

In order to explore the utility of the NIS-2 data base in addressing policy and program concerns, a series of analyses were conducted on three principal questions:

- the factors which influence the degree to which CPS is aware of child maltreatment cases recognized by professionals;
- the factors which influence CPS determining if a given report is a founded, indicated or unfounded case of maltreatment; and
- the factors which distinguish different child abuse and neglect subpopulations.

This section of the report summarizes the findings from these analyses and highlights the ways in which the study methodology as opposed to the actual reality of the situation might have influenced the results. Unless otherwise noted, all percentages included in the following tables reflect the distributions for the weighted data.

CPS Awareness Levels

One of the most frequently cited findings from both of the NIS studies is the relatively small percentage of cases known to professionals which were actually investigated by child protective service agencies. Overall, Westat reported that only 46% of the cases identified in the present study were known to and investigated by local child protective service agencies. This figure is comparable to the rate of reporting noted in the initial national incidence study (NIS-1). In that study, only 33% of the cases identified by professionals had been formally reported.

In defense of professionals not reporting all known cases, many practitioners feel they can better protect the child by not reporting known or suspected cases. These workers cite the inflexibility in certain child protective service procedures and the poor follow-through during the investigative and treatment planning process as resulting in increased client frustration, anger at the system, and a sense of personal betrayal by the professional from whom they had originally sought assistance (Alfaro, 1984). More recent research suggest that such reasons, while continuing to be cited, may not be the primary reasons for not reporting. Zellman (1990) found in her survey of 912 professionals that the most frequently endorsed reason for failing to report was a lack of sufficient evidence that abuse or neglect had occurred. Also important was the fact that the observed act was not serious enough to report, that they suspected the abuse had already been reported by another source, and that the situation had resolved itself.

In addition to concern over the significant number of cases not being reported to protective services, it has long been suspected that professionals are influenced by a client's race, income or marital status in determining when to report. In a secondary

analysis of the NIS-1 data, Newberger (1983) found that a disproportionate number of unreported cases were victims of emotional abuse, in families of higher income, whose mothers were alleged to be responsible for the injuries and who were white. The severity of harm resulting from the maltreatment was found to be a significant discriminating factor between reported and unreported cases only when income level was excluded from the analyses, suggesting that class and race, not severity, define who does and who does not get reported.

In order to investigate the potential reasons behind the significant number of cases known to professionals but not known or investigated by CPS observed in the NIS-2 data, NCPA conducted a number of crosstabulations between this variable and certain demographic characteristics identified by ASPE.³ These characteristics included:

- child's age;
- child's sex;
- child's race;
- age of mother;
- employment status of mother;
- household income;
- household AFDC status;
- family composition;
- number of children in the family;
- type of abuse; and
- county size in which the incident occurred.

The results of these crosstabulations are presented in Table 2. As this table indicates, the proportion of countable cases included in this analysis which were known to and investigated by CPS is 42.5%, slightly less than the number reported by Westat.⁴

Contrary to the Newberger finding, race and income do not distinguish between cases known to CPS and cases known only to professionals. While two-parent families were slightly less likely to be known to CPS than single parent families (39% versus 42%), the overall pattern in the crosstabulations does not suggest extensive screening by professionals based on questions of race, income or family composition. The one variable which might suggest a particular bias on the part of CPS or professionals is the high proportion of mothers not in the labor force who were known to CPS and the very low proportion of mothers looking for work not known to CPS. This pattern may suggest a tendency among professionals to have greater faith in the parenting abilities of mothers actively seeking employment. On the other hand, this variable may be highly

³ Characteristics with over 40% missing data were excluded from the analyses. These include age of father (64%), employment status of father (42%), age of perpetrator (67%) and type of perpetrator (58%). In addition, coding difficulties with the "source of report" data complicated the use of this variable in this analysis.

⁴ The NCPA definition of countable cases excludes the 103 cases which were out of scope with respect to time but were on the CPS long forms. We also excluded 6 cases where the child was not a victim based on codes on both the role indicated and role alleged variables.

correlated with a number of other factors, such as the child's age, which influence CPS behavior. For example, if mothers not in the labor force are more likely to have younger children, the observed pattern might well reflect the tendency of CPS to be more aware of younger children rather than any overt opinions of parenting capabilities influenced by maternal employment status.

| TABLE 2: CPS Awareness by Demographic Characteristics (Weighted Data)^a | | |
|----------------------------------------------------------------------------------------------|--------------------------------------|--------------------------------------|
| | Aware (%) N=646,131 (42%) | Unaware (%) 874,099 (58%) |
| CHILD'S AGE | | |
| 0-2 years | 46 | 54 |
| 3-5 years | 60 | 40 |
| 6-9 years | 47 | 50 |
| 10-12 years | 43 | 56 |
| 12+ years | 30 | 70 |
| Unknown | 95 | 5 |
| CHILD'S SEX | | |
| Male | 44 | 52 |
| Female | 40 | 60 |
| Unknown | 88 | 12 |
| RACE | | |
| White | 42 | 58 |
| Black | 40 | 60 |
| Other | 43 | 57 |
| Unknown | 78 | 22 |
| AGE OF MOTHER | | |
| 12-19 years | 51 | 49 |
| 20-25 years | 57 | 43 |
| 26-34 years | 55 | 45 |
| 35-70 years | 40 | 60 |
| Unknown | 28 | 72 |
| EMPLOYMENT STATUS MOTHER | | |
| Employed Fulltime | 41 | 59 |
| Employed Parttime | 45 | 55 |
| Looking for Work | 25 | 75 |
| Not in Labor Force | 56 | 44 |
| Unknown | 32 | 68 |
| INCOME | | |
| Under \$15,000 | 44 | 56 |
| \$15,000 plus | 43 | 57 |
| Unknown | 38 | 62 |
| AFDC | | |
| Yes | 47 | 53 |
| No | 50 | 50 |
| Unknown | 28 | 72 |
| FAMILY COMPOSITION | | |
| Two Parent | 39 | 61 |
| Female Head | 42 | 58 |
| Male Head | 42 | 58 |
| Unknown | 48 | 52 |

| TABLE 2 (continued) | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------|--------------------------------------------|
| | Aware (%) N=646,131 (42%) | Unaware (%) 874,099 (58%) |
| NUMBER OF CHILDREN | | |
| 1 Child | 37 | 63 |
| 2 Children | 40 | 60 |
| 3 Children | 44 | 56 |
| 4+ Children | 47 | 53 |
| Unknown | 47 | 53 |
| TYPE OF ABUSE | | |
| Physical Abuse | 53 | 47 |
| Sexual Abuse | 50 | 50 |
| Emotional Maltreatment | 39 | 61 |
| Physical Neglect | 51 | 49 |
| Educational Neglect | 13 | 87 |
| Other Maltreatment | 41 | 59 |
| COUNTY SIZE | | |
| Large SMSA | 32 | 68 |
| Other SMSA | 45 | 55 |
| Non SMSA | 54 | 45 |
| a. Demographic Variables with more than 40% missing data are excluded from this table. These are Age of Father, Employment of Father, Age of Perpetrator, Type of Perpetrator. | | |

Demographic factors found to influence whether or not a case is reported and investigated by CPS include the victim's age, maternal age, number of children in a family, and type of abuse. Not surprisingly, child protective services are more likely to be aware of maltreatment involving younger children than teenagers, a pattern indicative of the greater importance placed on protecting young infants, toddlers and early school-aged children. The very high percentage of children 3 to 5 known to CPS (60%) was somewhat surprising. This pattern might have been generated by the high percentage of this age group which were involved in physical abuse and child sexual abuse. Given the fact that younger children were more likely to be known to CPS than older children, it was not surprising to find younger mothers more likely known to CPS than older mothers.

As anticipated, public investigative energies tend to focus on maltreatment forms which produce more solid physical evidence of harm to the child. As a result, a lower proportion of educational neglect, emotional maltreatment and sexual abuse cases are reported to and investigated by CPS. While a professional may suspect a child is being mistreated in these manners, proving these suspicions, particularly in the absence of a formal disclosure by the victim or admission of guilt by a perpetrator, can be extremely problematic. Further, many child protective service agencies are faced with the need to prioritize because existing resources are insufficient to allow for a thorough investigation of all reports (Wells, Fluke, Downing and Brown, 1989). As a result, first priority is given to those types of maltreatment perceived as most severe and offering the clearest grounds for further judicial action.

As reported in Table 2, levels of CPS awareness differed rather notably for counties of different size. CPS is less likely to be aware of cases in large SMSAs and

more likely to be aware of cases in rural areas than those in medium size SMSAs. Again, this pattern is not surprising and may well reflect the difficulty CPS workers in the most urban counties face in balancing increased reports with stable or decreasing revenues. Because of the structure of the NIS data, it is not possible to determine if the absence of a case on the CPS listing is a function of the professional choosing not to report the case or CPS choosing not to investigate the case. In any event, it is consistent with current studies of CPS practice to conclude that the factors which would increase the likelihood of a CPS worker not reporting a case (e.g. feeling they can handle the case better without reporting, uncertainty over how CPS will respond) and factors which would result in more rigorous screening of reports prior to investigation may be more acute in urban communities.

For many of the variables included on Table 2, notable differences in the likelihood of CPS awareness of a case were found among the missing data categories. This pattern underscores the difficulty in using the NIS data to interpret questions of policy. If the data base was complete, the patterns observed above might not have been supported. While a certain degree of missing data is inevitable in any research project, variables with missing data on a quarter or more of the cases or individual cases which cannot be fully documented pose significant difficulties in drawing reliable policy and practice conclusions.

Case Verification by CPS

A frequently debated issue in the field is the acceptable level of unfounded cases. Because no child abuse reporting system in the country requires that professionals or individuals be absolutely certain abuse or neglect has occurred before filing a formal report, it is expected that some percentage of these reports will be found, following an investigation, not to involve abuse or neglect. In certain jurisdictions, the substantiation rate has dropped dramatically in recent years promoting some to argue that child protective services are being asked to function as an all purpose social service agency rather than as a specialized' unit dealing only with abuse and neglect (Besharov, 1986).

While individual jurisdictions have experienced significant variation in the percentage of substantiated cases, the national substantiation rate for child abuse reports has been remarkably consistent over the past ten years, hovering around 50% (AAPC, 1988). Further, the term unsubstantiated does not always imply the absence of maltreatment. As Finkelhor (1990) has argued, cases may be termed unsubstantiated for such diverse reasons as an investigation never occurred, the case is currently an active CPS case, or CPS had no services to offer the family. Even if the substantiation rate is as poor as some argue (i.e. 30%), this percentage would still compare favorably with the confirmation rate experienced by other emergency response systems. For example only about one-third of all calls to fire departments involve an actual fire and reviews of police time studies suggest beat patrol officers spend less than 25% of their time dealing with violent crimes (Daro, 1988).

Very little empirical work has been done across jurisdictions regarding the characteristics of cases more or less likely to be substantiated. Barriers to this type of research include the different criteria used across states to determine whether a report does or does not constitute maltreatment and the practice in many jurisdictions of purging, all information on a case once it has been classified as unfounded or unsubstantiated. While the NIS-2 data is also hampered by the different standards for substantiation employed among the sample CPS agencies, the data do offer a unique opportunity to analyze these decisions in light of certain descriptive characteristics.

To identify characteristics associated with whether or not CPS substantiates a case, NCPCA conducted crosstabulations of whether a case was substantiated or unfounded by all but one of the descriptive characteristics used to examine CPS awareness. This analysis did not explore the relationship between case verification and type of abuse due to the specific sample utilized for the dependent variable. This analysis is based on all cases reported to CPS. Thus this sample differs from the sample used to investigate CPS awareness because it includes cases which the Westat study team deemed not countable. Because the non-countable cases were not assigned a maltreatment type in Westat's final structuring of the data, this variable is omitted from the present analysis.

As the results in Table 3 show, CPS founded or indicated 54% of the cases. These two categories were combined for purposes of this analysis because a county level examination of the data found that several of the counties exclusively used either the "founded" or the "indicated" category. This suggest that local CPS policy rather than any objective interpretation of these terms governed how respondents classified the cases they were documenting.

As noted in the previous analysis, race and income factors do not appear to play a significant role in determining whether a case will be substantiated. Only marginal differences were noted in the proportion of lower income and AFDC recipients with substantiated reports as compared to the proportion substantiated for those with annual incomes over \$15,000 and not receiving public assistance. The one economic variable which did seem to distinguish between substantiated and unsubstantiated cases was maternal employment. Mothers who were employed part-time or who considered themselves not in the labor force were more likely to be substantiated cases than were mothers who worked full-time.

Somewhat counterintuitive was the finding that two parent households were more likely than single parent households to involve substantiated maltreatment and that reports involving older children were more likely to be substantiated than ones involving younger children. Given the findings of the previous analysis and the characteristics typically ascribed to CPS caseloads, one might have anticipated higher substantiation rates for reports involving single parent families and infants. One interpretation of these findings is that screening along these dimensions might indeed take place prior to the decision to file a formal report. Professionals or individuals may require a higher

standard of proof before formally reporting a case involving two parents or older children. As a result, cases with these characteristics which do find their way to protective services may include a higher than average percentage of more serious or more provable acts of maltreatment.

| TABLE 3: CPS Status of Case by Demographic Characteristics (Weighted Data) | | |
|---------------------------------------------------------------------------------------|--------------------------------------------------|----------------------------------------|
| | Founded/Indicated (%) N=838,108 (54%) | Unfounded (%) 704,314 (46%) |
| CHILD'S AGE | | |
| 0-2 years | 46 | 54 |
| 3-5 years | 50 | 50 |
| 6-9 years | 56 | 44 |
| 10-12 years | 57 | 43 |
| 12+ years | 60 | 40 |
| Unknown | 50 | 50 |
| CHILD'S SEX | | |
| Male | 51 | 49 |
| Female | 57 | 43 |
| Unknown | 56 | 44 |
| RACE | | |
| White | 53 | 47 |
| Black | 58 | 42 |
| Other | 56 | 44 |
| Unknown | 52 | 48 |
| AGE OF MOTHER | | |
| 12-19 years | 33 | 67 |
| 20-25 years | 47 | 53 |
| 26-34 years | 55 | 45 |
| 35-70 years | 63 | 37 |
| Unknown | 53 | 47 |
| EMPLOYMENT STATUS MOTHER | | |
| Employed Fulltime | 45 | 55 |
| Employed Parttime | 58 | 42 |
| Looking for Work | 51 | 49 |
| Not in Labor Force | 58 | 42 |
| Unknown | 57 | 43 |
| INCOME | | |
| Under \$15,000 | 54 | 46 |
| \$15,000 plus | 50 | 50 |
| Unknown | 61 | 39 |
| AFDC | | |
| Yes | 53 | 47 |
| No | 52 | 48 |
| Unknown | 63 | 37 |
| FAMILY COMPOSITION | | |
| Two Parent | 56 | 44 |
| Female Head | 51 | 49 |
| Male Head | 49 | 51 |
| Unknown | 56 | 44 |

| TABLE 3 (continued) | | |
|----------------------------|--------------------------------------------------------|----------------------------------------------|
| | Founded/Indicated (%) N=838,108 (54%) | Unfounded (%) 704,314 (46%) |
| NUMBER OF CHILDREN | | |
| 1 Child | 50 | 50 |
| 2 Children | 53 | 47 |
| 3 Children | 57 | 43 |
| 4+ Children | 58 | 42 |
| Unknown | 52 | 48 |
| COUNTY SIZE | | |
| Large SMSA | 57 | 43 |
| Other SMSA | 57 | 43 |
| Non SMSA | 43 | 54 |

Similarly, the importance of county size in determining whether a case is substantiated also illustrates an interesting counterpoint to the previous analysis. While cases identified in the largest counties were least likely to have been noted among CPS caseloads, the largest counties demonstrated the highest substantiation rate. In contrast, the rural counties which had demonstrated the highest CPS recognition rate, recorded the lowest substantiation rate. One interpretation of this pattern is that professionals in the largest counties both within and outside CPS do more rigorous screening of cases, reporting only the most serious or best documented maltreatment incidents. In smaller counties, professionals may be more willing to report and CPS more willing to investigate a wider range of maltreatment charges. As a result, fewer of these cases are substantiated or accepted for service. Again, the absence of any specific information on agency policies within the NIS-2 data base precludes further empirical exploration of this issue.

Subpopulation of Maltreatment

The third analytical question NCPA examined was what demographic characteristics distinguish different types of maltreatment. Child maltreatment, as a summative term, incorporates a wide range of behaviors. Parents who beat their children, the father who sexually molests his daughter, and the single parent who fails to ensure that her children attend school or receive adequate medical care are guilty, in the eyes of the law, of the same infraction --- child maltreatment. From a public policy perspective, child maltreatment is the generic problem comprising a variety of different, but theoretically similar, behaviors. Mistreatment of children or the failure to care for children is the central legal and policy issue; precisely how parents or caretakers choose to mistreat their children is of secondary concern. For purposes of clinical practice, however, quite the reverse is true.

As more is known about the diversity within the maltreatment population, unique subpopulations are being singled out for specific programmatic or legislative attention (Daro, 1988). Four major types of maltreatment are consistently cited in the literature -- physical abuse, physical neglect, emotional maltreatment and sexual abuse. In distinguishing among these four types, researchers have drawn on such diverse

variables as the characteristics of the perpetrator, the characteristics of the victim, and the underlying personal and environmental factors which led to the maltreatment.

The NIS-2 offers a limited opportunity to explore demographic differences among families experiencing different types maltreatment. For purposes of this analysis, acts of maltreatment are divided into six categories: physical abuse, sexual abuse, emotional maltreatment, physical neglect, educational neglect, and other maltreatment. The NCPA categorization system differs from the one employed by Westat in two ways. First, it includes only the first form of abuse, rather than all three forms combined. This provides a more straightforward description of the children involved. In our classification the case would be assigned whatever type was listed by the study source as the first type of abuse. Second, the types of abuse are collapsed into different categories. For example, we combined emotional abuse and neglect into emotional maltreatment. Other maltreatment includes general abuse, general neglect and other maltreatment.

NCPA crosstabulated the types of abuse by the demographic characteristics used in the two previous analyses. These results are presented in Table 4. The most frequent form of maltreatment was physical neglect (32% of the sample), followed by physical abuse (21% of the sample), educational neglect (17% of the sample), emotional maltreatment (16% of the sample), sexual abuse (8% of the sample), and other maltreatment (7% of-the sample). This distribution pattern is slightly different from the distribution suggested by studies of reported cases of maltreatment. In 1986, almost 55% of all reports involved child neglect, 38% involved physical abuse, 16% involved sexual abuse, 8% involved emotional maltreatment, and 8% involved other, unspecified types of maltreatment (AAPC, 1988). The higher proportion of sexual abuse cases and the lower proportion of emotional maltreatment in the two samples most likely reflect the differences in professional practice. Professionals may well be more likely to observe, but not report, incidences of emotional maltreatment and less likely to observe but more likely to report cases of sexual abuse.

As expected, the NIS-2 data support the notion of unique subpopulations of maltreatment. On most of the dimensions tested, the characteristics of the victims and their families differed depending upon the first type of maltreatment indicated for the case. Emotional maltreatment and educational neglect were more common among children 10 years of age or older than under 10. In contrast, young children were far more likely to be victims of child neglect. As would be expected, maternal age followed this same distribution, with younger moms more likely being noted in cases involving physical neglect and older moms being noted in cases involving emotional maltreatment and educational neglect.

Perhaps one of the most surprising distributions noted in Table 4 was the age distribution for victims of sexual abuse. The age category with the highest proportion of sexual abuse was 3 to 5 year olds, considerably younger than the average age of victims (9.19) noted among reports of sexual abuse (AAPC, 1988). However, a growing number of clinical studies, while often based on more limited samples, suggest that the onset of sexual abuse may occur when victims are much younger than implied in the

reporting statistics. One major sexual abuse treatment center has reported that over 25% of their victims are five years of age or younger (Summit,1983).

| TABLE 4: Types of Abuse by Demographic Characteristics (Weighted Data) | | | | | | |
|-----------------------------------------------------------------------------------|----------------------------------------------|-----------------------------------------|-------------------------------------------------|----------------------------------------------|----------------------------------------|---------------------------------------|
| | Types of Abuse (%) | | | | | |
| | Physical Abuse N=317560 (21%) | Sexual Abuse 115047 (8%) | Emotional Maltreat. 266467 (16%) | Physical Neglect 488585 (32%) | Educ. Negl. 25591 (17%) | Other Mal. 106661 (7%) |
| CHILD'S AGE | | | | | | |
| 0-2 years | 19 | 3 | 7 | 61 | -- | 11 |
| 3-5 years | 23 | 12 | 10 | 45 | 5 | 5 |
| 6-9 years | 20 | 9 | 14 | 34 | 15 | 8 |
| 10-12 years | 20 | 6 | 22 | 25 | 17 | 11 |
| 12+ years | 22 | 8 | 18 | 20 | 27 | 5 |
| Unknown | 8 | -- | 18 | 65 | -- | 9 |
| CHILD'S SEX | | | | | | |
| Male | 22 | 4 | 15 | 33 | 19 | 7 |
| Female | 20 | 12 | 15 | 31 | 14 | 8 |
| Unknown | 12 | -- | 62 | 26 | -- | -- |
| RACE | | | | | | |
| White | 20 | 8 | 17 | 30 | 18 | 8 |
| Black | 19 | 5 | 14 | 38 | 16 | 8 |
| Other | 27 | 9 | 14 | 34 | 13 | 5 |
| Unknown | 20 | 16 | 21 | 31 | 11 | 2 |
| AGE OF MOTHER | | | | | | |
| 12-19 years | 19 | 4 | 8 | 57 | -- | 13 |
| 20-25 years | 25 | 5 | 7 | 51 | 5 | 8 |
| 26-34 years | 22 | 11 | 20 | 37 | 7 | 5 |
| 35-70 years | 19 | 10 | 19 | 24 | 19 | 9 |
| Unknown | 20 | 5 | 13 | 28 | 28 | 7 |
| EMPLOYMENT STATUS MOTHER | | | | | | |
| Employed Fulltime | 25 | 10 | 18 | 26 | 15 | 7 |
| Employed Parttime | 22 | 9 | 12 | 31 | 22 | 4 |
| Looking for Work | 21 | 6 | 14 | 27 | 15 | 17 |
| Not in Labor Force | 19 | 7 | 17 | 38 | 14 | 5 |
| Unknown | 19 | 8 | 12 | 33 | 23 | 6 |
| INCOME | | | | | | |
| Under \$15,000 | 16 | 7 | 15 | 37 | 17 | 7 |
| \$15,000 plus | 31 | 10 | 18 | 21 | 15 | 6 |
| Unknown | 21 | 7 | 15 | 31 | 18 | 9 |
| AFDC | | | | | | |
| Yes | 15 | 5 | 14 | 43 | 12 | 12 |
| No | 26 | 10 | 18 | 29 | 12 | 6 |
| Unknown | 20 | 7 | 14 | 27 | 28 | 5 |
| FAMILY COMPOSITION | | | | | | |
| Two Parent | 24 | 10 | 20 | 27 | 14 | 6 |
| Female Head | 20 | 6 | 12 | 36 | 17 | 10 |
| Male Head | 19 | 4 | 8 | 35 | 17 | 17 |
| Unknown | 17 | 8 | 14 | 36 | 22 | 4 |
| NUMBER OF CHILDREN | | | | | | |
| 1 Child | 26 | 8 | 13 | 29 | 18 | 6 |
| 2 Children | 22 | 9 | 16 | 33 | 14 | 6 |
| 3 Children | 20 | 8 | 19 | 33 | 12 | 8 |
| 4+ Children | 18 | 7 | 16 | 37 | 11 | 11 |
| Unknown | 17 | 5 | 10 | 24 | 42 | 2 |
| COUNTY SIZE | | | | | | |
| Large SMSA | 21 | 8 | 13 | 33 | 20 | 6 |
| Other SMSA | 20 | 8 | 18 | 29 | 17 | 8 |
| Non SMSA | 21 | 7 | 15 | 38 | 12 | 8 |

The only type of abuse showing a different distribution pattern by sex was child sexual abuse, where the victims were three times more likely to be female than male. Again, this pattern is consistent with the majority of clinical research and official reporting data (AAPC, 1988). However, like with age, extensive interviews with men incarcerated for child sexual abuse reveal a surprisingly high childhood victimization pattern involving males, as do interviews with random and nonrandom samples of adult males (Gebhard, Gagnon, Pomroy and Christenson, 1965; Finkelhor, 1979; Groth, 1983; Abel, Becker, Cunningham-Rathnor, Renlean, Kaplan and Raid, 1984; Finkelhor, 1984). These studies suggest that professionals may be disinclined to suspect sexual abuse with young boys, thereby accounting for the relatively low percentage of male victims reported to the Westat study team.

The only notable difference involving the victim's race were the over-representation of blacks in cases of physical neglect. The high proportion of blacks living below the poverty line and the well documented correlation between poverty and child neglect most likely account for this distribution (Pelton, 1981). Consistent with this pattern was the higher proportion of unemployed mothers reported for child neglect. These women most likely are parenting relatively young children (a situation correlated with a higher frequency of child neglect) or may be enrolled in a workfare/job placement program associated with AFDC (a variable also associated with a higher incidence of child neglect).

Significant differences were noted in the income for families involved in various types of maltreatment. While the over-representation of lower income families among cases of physical neglect was anticipated, the large proportion of families with annual incomes over \$15,000 noted among those cases involving physical abuse was somewhat surprising. To a certain extent, this pattern underscores the universality of the child abuse problem, with income not being as strong a predictor of family violence as some would contend.

The typical family composition does differ by type of maltreatment. On balance, physical abuse and sexual abuse cases include a higher proportion of two parent families and single parent families headed by males while single parent families headed by women are overrepresented among cases involving physical neglect. Similarly, families with greater number of children are more common among cases involving physical neglect, while only children are overrepresented among cases of physical abuse and sexual abuse. Again, these patterns are not surprising and are consistent with the characteristics of these subpopulations reported by others (Finkelhor, 1986, Daro, 1988, AAPC, 1988, and Gelles and Straus, 1989).

The NIS-2 methodology, namely the use of professionals to identify cases of maltreatment, may have its strongest impact with respect to the descriptions of various subpopulations. In interpreting the characteristics most likely associated with different forms of maltreatment, it is important to bear in mind that not all families or all children had an equal likelihood to be observed by the sentinels selected for this study. For example, young children had significantly less probability than school-aged children of

being included in the study. Unless the child was in a day care center or was taken to a hospital emergency room with an injury serious enough to have warranted referring the child to a hospital social worker or had been formally reported to CPS, he or she would have had no way of being included in this study. In short, this data base may seriously underreport the range of maltreatment involving children under 5. A more accurate assessment of this population might have been possible had the study included a sample of pediatricians or professionals working in well-baby clinics.

This issue aside, the relatively modest differences observed among the various subpopulations suggest that the types of demographics included in the NIS-2 are generally poor predictors of maltreatment. With the possible exception of physical neglect, maltreatment does indeed cut across all income groups and family structures. If the NIS data is to be considered a useful empirical base for building credible models to distinguish among different types of maltreatment it may be necessary to include a broader range of variables than is currently available. Based on the experiences of others who have pursued this type of research, it appears essential to capture, in greater detail, perpetrator characteristics and the underlying personal and environmental factors which contribute to elevated risks for maltreatment.

COMPARISONS BETWEEN NIS-1 AND NIS-2

The problems raised with respect to the weights and the differences in approach between the NIS-1 and the NIS-2 call into question the ability to generate valid statements regarding trends in child abuse levels between the two time periods. Of particular concern is the use of a 3-month data collection period in the NIS-2 study compared to a one year data collection period in the NIS-1 study; the differences in the representation of large counties in the two studies; and the duplication issue outlined earlier. While the Westat study team took great care to address all of these issues and to adhere to comparable methods in constructing the estimates used in both studies, the methodological issues noted by the statisticians in the NIS-2 approach precluded a definitive answer on this question. As has been discussed earlier, both the pool of sentinels utilized and the effects of weight trimming might have biased the estimate downwards, while the duplication issue might have had the opposite effect. While these factors may lead to inaccurate estimates of child maltreatment, they are not of utmost concern in explaining the observed differences between NIS-1 and NIS-2 since presumably the bias would have been roughly the same magnitude in both studies. Rather than resulting from methodological differences, it is more likely that the observed 66% increase in child abuse and neglect between 1980 and 1985 is an absolute increase resulting from a variety of factors, including professional awareness, environmental and economic changes and personal behavior.

As stipulated in our original research design, NCPA staff informally polled a number of child welfare administrators, advocates and academic researchers regarding their opinions regarding the NIS-2 conclusion. Specifically, respondents were asked three questions relative to this issue:

- do you believe the incidence of child abuse increased markedly between 1980 and 1985?
- if so, do you perceive the increase to be solely a function of increased professional awareness?
- what alternative explanations can you suggest for the Westat finding?

Respondents also were asked how they have used the NIS-2 findings and how valid they considered the overall methodology to be in accurately measuring child abuse levels.

Appendix B lists the individuals contacted as part of this effort.

Are Child Abuse Rates Going Up?

Respondents noted that both NIS-1 and NIS-2 offer only an indication of the number of child abuse cases observed by professionals on an annual basis, not the true

incidence of maltreatment. The issue under debate, therefore, is whether professionals identified more maltreatment cases in 1985 because they were more observant or because they come in contact with more actual cases. If one believes the latter explanation to be true, it does not necessarily follow that the total incidence of maltreatment has increased. Similar number of cases might have existed in 1980 but, for whatever reason, were not brought to the attention of professionals.

Only one-third of the respondents supported the Westat study team's interpretation of the findings, with the remaining two-thirds posturing alternative explanations for the documented 66% increase. All of those interviewed agreed that professional awareness and pressure to observe and report child abuse has increased in recent years, particularly with respect to child sexual abuse. Such pressure most certainly accounts for increased recognition of the problem. As one respondent noted, it is "implausible that three times as many Americans are having sex with their children as ten years ago." With respect to other forms of maltreatment, however, the majority of the respondents viewed increased awareness as accounting for only a fraction of the observed increase.

Factors identified as contributing to a real increase in the number of cases observed ranged from changes in the broader social-economic sphere, such as increased child poverty, homelessness and societal violence, to changes in family structure and parental behavior, such as divorce, teenage parenting and substance abuse. Interestingly, those respondents who spend the majority of their time conducting basic research or who work for the Federal Government were more likely than other respondents to support the Westat conclusion. Those more likely to disagree included child welfare administrators, child advocates and researchers who, in addition to research, spend a significant portion of their time in clinical settings. One reason for this division may be the influence reality plays in shaping one's perception of a given phenomenon. Rising reporting rates and a real increase in the number of serious physical and sexual abuse cases seen by local child protective service caseworkers seem indicative of a problem on the rise, not a problem under control.

Table 5 summarizes the arguments for and against the major interpretations of the 66% difference in the number of child abuse cases noted by professionals in 1980 and 1985. As this table indicates, compelling arguments can be made for each conclusion. Certainly, there has been an increase in professional recognition of child abuse, particularly child sexual abuse and within rural areas. A number of states require social workers, educators, and health care professionals to obtain special training on the identification and response to child abuse as a condition for certification or licensing. Beyond the formal training professionals receive on this topic is a general public which has less tolerance for all forms of maltreatment than in the past. To the extent professionals observe a child being beaten, belittled or neglected, they do appear more likely to labeled such behavior as child abuse than they might have in 1980.

On the other hand, the conditions in which a growing number of children live in this country are fertile grounds for mistreatment. The rate of child poverty, the growing

number of children living in shelters for the homeless, and the drug epidemic has led many thoughtful observers to conclude that the risk for child abuse is higher today than at any time in recent memory. This increased risk coupled with a decrease in the range of supportive services for those families with the most limited resources give credence to the belief that serious child abuse may be on the rise.

TABLE 5: Hypotheses Explaining NIS-2 Findings

| Hypothesis | Supporting Arguments | Non-Supporting Arguments | Testing Methods | Explanatory Power |
|-----------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>1. The increase in the recognition of child abuse noted between the NIS-1 and NIS-2 is a statistical artifact.</p> | <ul style="list-style-type: none"> - Greater number of large urban counties in NIS-2, suggests that a more complete accounting for the scope of the problem in these areas could be obtained than was feasible under NIS-1. - A broader representative of sentinels in the NIS-2 resulted in a larger pool of respondents with ongoing contact with very high populations (i.e. those seeking mental health services) - The policy of adjusting the number of maltreatment cases observed by each sentinel to reflect the number documented by their more "cooperative" colleagues (i.e. exit evaluation adjustments) may inflate the NIS-2 estimates - Because Westat had conducted a similar study in 1980, the NIS-2 was implemented with a greater degree of expertise and efficiency, thereby reducing the likelihood of selecting less than optimal sample agencies or sentinels | <ul style="list-style-type: none"> - While methodological changes were made between NIS-1 and NIS2, the two studies are essentially comparable, with both having randomly selected counties, agencies and respondents. - Statistical adjustments were made to account for the different reporting periods used and for the heavier sampling of cases which occurred in the NIS-2. | <ul style="list-style-type: none"> - Calculate the incidence rate for 1985 without the exit evaluation adjustments - Determine the percentage of the incidence rate in both studies accounted for by the largest urban counties and then compare the abuse incidence rates generated by the remaining counties | <ul style="list-style-type: none"> - Little evidence exists to suggest the entire observed 66% increase was due to methodological changes in the two studies. However, it is quite possible that the exit evaluation adjustments and the larger proportion of urban counties in the NIS-2 sample somewhat inflated the 1980-1985 difference. |

TABLE 5 (continued)

| Hypothesis | Supporting Arguments | Non-Supporting Arguments | Testing Methods | Explanatory Power |
|----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>2. Since 1980, professional recognition, not the actual incidence, of child maltreatment has increased.</p> | <ul style="list-style-type: none"> - Mandated training of many key social and health care professionals with respect to child maltreatment suggest sentinels were better equipped to recognize maltreatment in 1985 than in 1980 - Increase coverage of child abuse particularly sexual abuse and child abuse fatalities in both the visual and print media has heightened sensitivity to these concerns. - Fear of possible prosecution for failing to protect a child may make professionals more vigilant in considering child abuse as a possible cause of any observed physical or emotional condition - The increase number of working mothers has resulted in professionals having access to more children of preschool age, and therefore able to identify abuse cases which in the past might have been known only to family members or friends | <ul style="list-style-type: none"> - While professionals became more aware of child sexual abuse during the early 1980's, training with respect to physical abuse and neglect has been in place since the 1960's - As child abuse awareness increases, parents or caretakers become more cautious in exposing children with obvious injuries to professional scrutiny. - Despite increased training, professionals still have a tendency not to get involved in suspected maltreatment cases as evidenced by the continued low rate at which they formally report the cases they identify (i.e. if increased child abuse training and awareness accounted for the full 66% increase, one might have expected a higher percentage of cases would have been formally reported since such procedures are frequently discussed and encouraged during such training sessions.) | <ul style="list-style-type: none"> - Compare the number of cases observed by sentinels in agencies with aggressive child abuse training and reporting policies to those working in agencies with little or no emphasis on such training - Compare the recognition rate of sentinels in communities with extensive child abuse prevention efforts and public awareness campaigns to those working in communities lacking such services and campaigns - Compare the recognition rate of sentinels who have completed formal training programs on child abuse to those without such training | <ul style="list-style-type: none"> - Increased professional awareness is a more valid explanation for the observed increase in sexual abuse cases than for the increases observed in other categories of maltreatment. If one assumes the increase in the sexual abuse was fully accounted for by increased awareness, that would still leave a 55% increase in other forms of maltreatment between the two reporting periods. |

TABLE 5 (continued)

| Hypothesis | Supporting Arguments | Non-Supporting Arguments | Testing Methods | Explanatory Power |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>3. The incidence of child abuse increased between 1980 and 1985 due to a variety of factors including:</p> <ul style="list-style-type: none"> - Increases in the rate of child poverty - High rates of teen parenting - Increases in the rate of substance abuse (particularly crack) - Increases in homelessness - Increases in divorce and custody disputes - Increase in general public awareness and commitment to preventing maltreatment - Increases in other forms of family and societal violence (e.g. domestic violence, murder, rape, etc. - Decline in social services and child welfare resources - Decline in the accessibility of health care services, particularly birth control and abortions | <ul style="list-style-type: none"> - Nationwide, living conditions for children most at risk have deteriorated as evidenced by increases in child poverty, teen parenting, homelessness and drug use, particularly crack leading to higher rates of physical abuse and neglect - Levels of serious injury to young children have increased or remained high throughout the 1980's (e.g. child abuse fatalities, infant mortality, accidental deaths, SIDS, etc.) - The rise of a very active anti-abortion/pro-life movement has inhibited the ability of poor women to seek abortions and the ability of teenagers to seek birth control resulting in a larger number of unwanted births - Increases in societal violence including domestic violence, murders rapes, and gang-related violence suggest an environment increasingly hostile to the nurturing of children - An uncertain economic future of those employed in manufacturing and un-skilled jobs increase personal stress levels, increases which have been associated with higher rates of physical abuse | <ul style="list-style-type: none"> - Overall employment rates and the nation's economy are actually brighter today than in 1980. - A significant number of hospitals, school districts and community-based agencies are involved in child abuse prevention efforts, offering parents, particularly new parents, assistance in meeting their children's needs. - While drug usage has reached epidemic proportions in many of the nation's urban centers, this behavior has experienced its largest growth since 1985. - The number of teen parents, while remaining high, has not been increasing in recent years. | <ul style="list-style-type: none"> - Determine, to the extent possible, the degree to which these risk factors played a role in the abuse and neglect cases observed by professionals--the stronger the relationship the more likely that increases in the frequency of these conditions would result in increased maltreatment rates. | <ul style="list-style-type: none"> - Recognizing that the "true" incidence of maltreatment is not captured by either of the Westat studies, the increases noted in the risk factors cited under this hypothesis project a less than ideal environment for children, particularly poor children. While it remains a theoretical argument as to whether the total level of maltreatment in this country increased between 1980 and 1985, it is plausible that such factors as child poverty, homelessness, and parental drug use combined to produce a larger number of serious physical abuse and neglect cases than existed in 1980. |

| TABLE 5 (continued) | | | | |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------|-----------------|-------------------|
| Hypothesis | Supporting Arguments | Non-Supporting Arguments | Testing Methods | Explanatory Power |
| | <ul style="list-style-type: none"> - Changing family structure results in fewer adults to share routine child rearing responsibilities (e.g. single-parent households) and children being exposed to a revolving number of non-family adults such as boyfriends and child care providers. - Decreases in Federal support for social and health care services coupled with declining or static state revenues for child welfare services have resulted in fewer service resources for high-risk families and children. | | | |

Little evidence exist to suggest that the 66% increase is merely an artifact of the methodology employed by Westat. While differences do exist in the sampling methods and statistical procedures employed in the two studies, it is not clear that these differences would account for such a dramatic increase in the number of observed cases.

Unfortunately, there is not a method to determine, empirically, which arguments represent the "correct" interpretation of the Westat data. Only two of the respondents said that they use these data as a method for measuring change in the scope of the problem. As reported above, serious questions exist as to the validity of any NIS-1 and NIS-2 comparisons. The use of a different balance of sample counties, agencies and sentinels, coupled with the absence of any data regarding the training of these professionals, make it extremely difficult to "prove" professional awareness accounted for the entire 66% difference. Further, as constructed, the data base includes no information relative to the alternative explanations for this increase posed by our respondents. No data are available regarding the rates of poverty, domestic violence, homelessness, or substance abuse in the sample counties nor are there any indicators of the levels of social and health care services in the communities surveyed. In short, because the NIS-1 and NIS-2 do not offer definitive measures for the scope of the child abuse problem in 1980 and 1985, it remains a theoretical debate as to whether overall maltreatment rates have increased or simply moved up in the Westat pyramid design (i.e. cases previously known only to perpetrators or family members are now known to professionals).

CONCLUSIONS

The need for a National Incidence Study is self-evident. Both practitioners and policy makers want to know how much child abuse there is in this country and whether or not it is on the rise. In the absence of a National Incidence Study figure, the public and the professional community would be forced to rely upon reported rates of child abuse, taken from a system which suffers from many identifiable downward biases. For example, we know that the reporting rates between given states or counties will vary depending upon: (a) explicit state policy and definitions of maltreatment; (b) the extent to which funding for the local child welfare system is sufficient to allow for comprehensive investigations of all reports; and (c) the extent to which all potential abusers have an equal likelihood to be formally reported for maltreatment.

Given the need for a National Incidence Study, the question then becomes how best to accomplish this task. Presently the sole alternative methodology relies upon professional judgments to move beyond the estimated small percentage of actual cases which are formally reported. While everyone is in agreement that this approach does not identify the total universe of maltreatment cases, it is believed to represent a more comprehensive and potentially less biased estimate than that suggested by current reporting rates. Short of observing parental behavior in a random sample of American households, the method may indeed be the best available technology for moving the field closer to a more accurate estimate of the scope of the maltreatment problem. Despite the problems raised by the reviewers, both of the National Incidence Studies utilized justifiable methods and generated estimates, which if one accepts the majority of the assumptions made in the course of developing the study design, are as credible as anything currently available.

However, the vast number of assumptions underlying the development of national estimates using the present methodology and the inability to quantify the direction and size of the total bias resulting from each of these assumptions requires a tremendous leap of faith in accepting the final incidence figures as accurate. One needs to accept, as given, that valid estimates have been made for the level of duplication in the sample; the representativeness and comprehensiveness of the sample agencies and sentinels; and the appropriateness of the case weights, annualization rates and weight trimming. As suggested by the statisticians, it seems prudent to move cautiously in designing subsequent National Incidence Studies. Simple replication of the existing design is not viewed as the most useful course of action. Determining reliable estimates rests on the ability to establish a methodology based upon the fewest number of assumptions and, for those assumptions one does make, conducting the most stringent sensitivity tests possible so as to quantify the magnitude and direction of any potential error introduced by these assumptions. As Dr. Frankel noted: "When the purpose of a survey is the estimation of a critical total, I believe that extensive justification should be provided for any shift from unbiased to biased estimation procedures." Each of the weighing systems employed in the NIS-2 involved biased estimation procedures, procedures which were not in the view of the statisticians sufficiently justified, largely

because the information needed to justify them is unavailable. We strongly recommend addressing this concern in designing any subsequent national incidence studies.

Further, future national incidence studies need to be designed with a careful eye toward the development of a data base which can address critical child welfare policy and program issues. To accomplish this task, additional data may need to be collected as part of the NIS. Specific changes which would increase the utility of the incidence data to address policy questions include information on agency policies, professional training and behavior with respect to reporting and various local systemic and environmental factors which influence maltreatment levels.

BIBLIOGRAPHY

- Abel, G. Becker, J. Cunningham-Rathner, J., Renlean, J., Kaplan, M. and Reid, J. (1984). The Treatment of Child Molesters (memo). New York: SBC-TM.
- Alfaro, J. (1984). "Summary of findings and issues: Survey of impediments to mandated reporting of suspected child abuse and neglect". Report to the Mayor's Task Force on Child Abuse and Neglect, City of New York.
- American Association for Protecting Children. (1988). Highlights of Official Child Neglect and Abuse Reporting - 1986. Denver, CO: American Humane Association.
- Bersharov, D. (1986). "Unfounded allegations -- a new child abuse problem." The Public Interest, 83(Spring), 18-33.
- Daro, D. (1988). Confronting Child Abuse. New York: Free Press.
- Finkelhor, D. (1980). "Is child abuse overreported?" Public Welfare (Winter) 22-29.
- Finkelhor, D. (1986). A Sourcebook on Child Sexual Abuse. Beverly Hills, CA: Sage.
- Finkelhor, D. (1984). Child Sexual Abuse: New Theory and Research. New York: Free Press.
- Finkelhor, D. (1979). Sexually Victimized Children. New York: Free Press.
- Gebhard P., Gagnon, J., Pomroy, W. & Christenson, C. (1965). Sex Offenders: An Analysis of Types. New York: Harper & Row.
- Gelles, R. and Straus, M. (1989). Intimate Violence. New York: Simon and Schuster.
- Newberger, E. The Helping Hand Strikes Again. Testimony given before the Subcommittee on Family and Human Services, Committee on Labor and Human Resources, U.S. Senate, April 11, 1983.
- Pelton, L. (1981). Social Context of Child Abuse and Neglect. New York: Human Services Press.
- Summit, R. (1983). "The child sexual abuse accommodation syndrome." Child Abuse and Neglect 7, 177-193.

Wells, S., Fluke, J., Downing, J., and Brown, C. (1989). Screening Child Protective Services: Executive Summary. Washington D.C.: Center on Children and the Law, American Bar Association.

Zellman, G. (1990). "Child abuse reporting and failure to report among mandated reporters." Journal of Interpersonal Violence 5:1 (March). 3-22.

APPENDIX A. CONSULTANT REPORTS

Martin Frankel & Associates, Inc.
14 Patricia Lane
Cos Cob, Connecticut 06807

May 25, 1989

Deborah Daro, DSW
Director, Center on Child Abuse Prevention Research
National Committee of Prevention of Child Abuse
332 S. Michigan Avenue, Suite 1600
Chicago, Illinois 60604-4357

Dear Deborah:

The purpose of this letter is to provide a written report which assesses the sampling, weighting and statistical methods employed in conducting the NIS-2 Survey (Study of National Incidence and Prevalence of Child Abuse and Neglect). As per your specifications, the following questions are addressed:

Did the sampling techniques employed produce an accurate and representative sample of counties?

Did the sampling techniques employed produce an accurate and representative sample of social service, criminal justice and health care professionals?

Are cases weighted in a manner which provides for credible national estimates?

Were comparisons with the 1980 data based on the accurate and appropriate use of various statistical procedures?

Were the projections that were made on the two sets of definitions used in NIS-2 valid?

What, if any, changes would you suggest in the sampling or methodology if the study were to be repeated in several years?

In providing my answers to these questions I have provided an overall answer of YES, NO or UNCERTAIN first. This overall answer is followed by an explanation.

Did the sampling techniques employed produce an accurate and representative sample of counties?

YES

The term accuracy is generally used to in conjunction with an estimate produced by a sample. The term representative is generally used, in a non-technical sense, to describe a sample that is selected in such a way that it has the potential for producing accurate and reliable estimates.

It is clear from the NIS-2 documentation that sample of counties was selected in such a way that a probability sample was produced. As such, the sample has the potential for use in the production of unbiased estimates with known¹ reliability.

Did the sampling techniques employed produce an accurate and representative sample of social service, criminal justice and health care professionals?

UNCERTAIN

Within selected counties the selection of social service, criminal justice and health care professionals involved a two stage sampling process. In the first stage of within county sampling (second stage, overall), sampling frames of agencies was constructed and a probability selection was carried out from this frame. In the second stage of within county sampling (third stage, overall) sampling frames of individuals were constructed and these were sampled by probability methods.

However, the probability or non-probability nature of the sample of professions selected by this process rests on a number of critical assumptions. First, is the assumption that all appropriate health care professions within the county are included in the frames that were prepared.

Insufficient information is provided in the technical documentation for the to assess the degree to which these assumptions are fully satisfied. I have particular questions about the frame of child care centers, as well as the within agency frames in the case of hospitals, schools, and child care centers.

Specifically, I would like to know if any procedures were used to assure the completeness of the frame of child care centers. I would also like to know what procedures were used to assure that the person level frames within selected agencies were complete and up to date. Pending the resolution of these questions, I am not in a position to assess the probability nature of the sample of social service, criminal justice and health care professionals.

¹ Strictly speaking, the use of a single primary selection (PS) per stratum does not allow for the estimation of unbiased estimates of sampling reliability. However, it is generally accepted that with a single PS per stratum it is possible to produce conservative estimates of sampling reliability.

Are cases weighted in a manner which provides for credible national estimates?

UNCERTAIN, and very probably NO

The most serious questions that I have concerning the estimates produced by the NIS-2 are related to the weighting of cases. In particular, I question the validity of the assumptions that were used in dealing with the potential problems of duplication of cases within CPS agencies and within non-CPS agencies and professionals, the duplication of cases between the various CPS and non-CPS providers and the duplication of cases over time (annualization).

On page 6-23 of the Report on Data Processing & Analysis: Study of National Incidence and Prevalence of Child Abuse and Neglect: 1988 the following paragraph appears:

"The attempt for obtain unbiased estimates of the extent for duplication in a population by using the incidence of duplicates in a sample can only rarely produce reliable estimates (footnote L.Kish (1965). Survey Sampling. New York John Wiley and Sons. See remark 11.2.1), and can often lead to negative estimates of population counts. For this reason, that approach was not used with the NIS-2 database. Instead, a more robust approach (albeit a biased one) was used."

I am in agreement with this paragraph that the approach used was biased. This is also pointed out by Kish on page 390 "Note that eliminating only the duplicate selections actually found in the sample would not correct the selection bias."

I am in disagreement with the implication that the procedure adopted was more robust.

The problem with the approach used is that no attempt has been made to quantify the magnitude of the bias. We are certain that the procedure is biased, but we do not know the size of the bias. We can not assume that the impact of the bias is small. The impact on the final estimates might be 5%, or it might be 75% or more. We just do not know.

In addition to the potentially serious bias introduced by ignoring the problem of "hidden" duplication in the population, I have a number of questions about the potential bias introduced by the "Reliability Adjustment" (p.6-20), the "Exit Evaluation Adjustment" (p. 6-21) and "Weight Trimming" (p. 6-28). Not enough detail is provided in order fully understand these adjustments or speculate on their potential bias impact. When the purpose of a survey is the estimation of a critical total, I believe that extensive justification should be provided for an shift from unbiased to biased estimation procedures.

Were comparisons with the 1980 data based on the accurate and appropriate use of various statistical procedures?

UNCERTAIN

Insufficient detail about the 1980 study was provided in order to answer this question. It appears that significant changes were made in the sample design and weighting procedures from HIS-1 to NIS-2. To the extent that the bias of the weighting procedures was different in the 1988 from 1980, this would impact the accuracy and appropriateness of NIS-1 to NIS-2 comparisons. In summary, it appears to be impossible to definitively attribute differences between results in the NIS-1 and the NIS-2 to either real change or method effect. It might be either or it might be both.

Were the projections that were made on the two sets of definitions used in NIS-2 valid?

UNCERTAIN

My, uncertainty about the validity of the two sets of definitions use in NIS-2 is linked to my uncertainty about the sampling of professionals and the weighting that was used in projection. The sampling and weighting were proper, then it is certainly possible to develop two valid sets of projections based on two sets of definitions, within the same sample survey.

What, if any, chances would you suggest in the sampling or methodology if the study were to be repeated in several years?

I would recommend that an extensive study be carried out in order to provide empirical evidence about the amount and nature of report duplication in three dimensions: within agency, across persons in the same agency, across agencies, across persons in different agencies and across time. This study should focus on large and medium sized counties, where extensive sampling is employed.

This empirical information is required in order to allow for the design of an NIS that will provide scientifically credible estimates.

Sincerely yours,

/s/

Martin R. Frankel, Ph.D.

MARX SOCIAL SCIENCE RESEARCH, INC.

196 Appleton Street
Cambridge, Massachusetts 02138
(617) 876-0962

April 17, 1989

Deborah Daro, DSW
Director
Center on Child Abuse Prevention Research
National Committee for Prevention of Child Abuse
322 S. Michigan Avenue, Suite 950
Chicago, IL 60604-4357

Dear Deborah:

This letter contains my responses to the six questions you asked the statisticians to address in Task 2. My responses are based on the materials you mailed me: Study Findings, Final Report: Appendices, Report on Data Processing & Analyses, Report on Data Collection, Report on County Sample Selection Process and Public Use Tape Documentation Manual.

Did the sampling techniques employed produce an accurate and representative sample of counties?

Probably yes.

All U.S. counties with at least 2,800 children in school were stratified by geographic region and degree of urbanization. Within each cell formed by these strata, counties were listed sequentially from northeast to southwest. Each was assigned a size equal to the number of children in school within the county. By dividing the total number of children in school in all counties by 27, a sampling interval was chosen that would yield a final cluster sample of 27 counties. Using a random number to initiate the selection process, a systematic sample was drawn by using the sample interval to choose counties proportional to size within each cell.

For the purpose of sampling, contiguous counties with less than 2,800 were aggregated until their combined size totalled at least 2,800 children. One of the set of combined counties was chosen at random. It turned out to be composed of two, individual counties. The 27 combined counties and 2 combined counties were designated sampling units) of the study.

My one suggestion on sampling would be to statistically test how representative of the universe of counties within its cell each selected sample of counties is. This could be done not only for the two strata dimensions of geography and population density, but

also with respect to other characteristics possibly related to child maltreatment such as income distribution and racial/ethnic composition.

Did the sampling techniques employed produce an accurate and representative sample of social service, criminal justice and health care professionals?

Yes, but I have some reservations.

The selection of non-CPS agencies is detailed in Chapter II of the Report on Data Collection.

Juvenile Probation Departments, County Sheriff/State Police and County Public Health Departments were censused within each PSU. They are therefore representative.

Schools, day care centers, hospitals, municipal police departments and social services/mental health agencies were all sampled within PSU. Whether they are likely to be representative depends on how the sampling was done.

Schools were stratified by grade spans (K-5, 6-9, 7-12), and further by size (over or under 1,000 students) and/or by race (over or under 50% white students) if there were a sufficient number of schools to form one or both of these additional strata. A random sample of ten schools was chosen from the cells formed by these strata so that the number of sampled schools in each cell was roughly proportional to the total number of schools in each cell.

Unless a goal of NIS-2 was to compare small to large schools, I wonder why schools were not randomly sampled proportional to size within grade and (possibly) race categories. Given that the student enrollment for each school in a PSU was readily available, this selection method might have sharpened incidence estimates.

The national universe of day care centers available for sampling came from a list purchased from a market research company. Was this list comprehensive? In the text, Westat discusses the high rate of turnover for day care centers without mention of any steps to verify how complete the list was.

Within each PSU, day care centers were ordered by descending enrollment. A systematic sample of twenty day care centers was then selected with probability proportional to size (enrollment) after random choice of a starting number. As high refusal rates were anticipated, day care centers were oversampled in the ratio of three backup centers for each primary (targeted) center.

Short-stay, general or children's general hospitals with 4,000 or more annual admissions were eligible for inclusion. In PSU's with 6 or less hospitals, all were selected. Otherwise, hospitals were sampled.

Children's hospitals were always selected. Public hospitals were selected with greater probability than private hospitals and hospitals with fewer than 4,000 admissions a year were only used where necessary. Hospitals were stratified into four size categories.

No further details are provided on how hospitals were sampled after they had been typed and sized. I assume there was random selection within the cells formed by the type and size strata.

I also assume that the census of children's hospitals and oversampling of public hospitals was an attempt to concentrate sentinels in institution types that were less prevalent or in settings where rates of child maltreatment cases seen were thought to be higher. The latter consideration would increase the number of countable cases reported to CPS from institutions with higher rates of such cases, thus improving the precision of incidence estimates.

A similar sampling strategy was followed with police departments. The larger the municipality served by the police department was, the higher was the police department's probability of selection into the sample.

All departments were taken in PSU's with 5 or fewer departments. In other PSU's, between 4 and 7 departments were sampled depending, among other things, on the number of police departments and the population of the PSU's. The area served by each department was assigned to one of six population strata. Probability of selection varied by stratum size from certainty in the stratum for cities of 500,000 or more to lower probabilities for strata with successively smaller population ranges. This method of sampling may be expected to generate larger numbers of countable maltreatment cases and more precise estimates of incidence than, for example, selecting departments proportional to size.

All known social services and mental health agencies were considered for the sampling frame except for government-administered social services agencies. To be included in the sampling frame the agency also had to be a nonresidential provider of counselling, therapy and/or emotional support to families or children in the general population of the PSU's, and to know its clients well enough to complete the non-CPS study data form.

Westat compiled a sampling frame from yellow pages listings of "social services," special directories of community services, and the knowledge of CPS staff and key participants at non-CPS study agencies.

From the sampling frame they took all agencies from PSU's with 4 or less agencies, and sampled 4 or 5 agencies randomly from the remaining PSU's.

I have two concerns about the sampling of social services and mental health agencies. Why was size of agency in terms of number of clients or families served not

considered in selecting agencies? Why were only 3 to 5 agencies selected regardless of the number of agencies in the PSU? (To take an extreme instance, 5 of 9 agencies were selected in Kern, CA and 5 of 235 agencies were selected in Los Angeles.) Ignoring size in sampling agencies coupled with the small sample of agencies from areas of high population density lead to imprecise incidence estimates of the countable maltreatment cases reported to CPS from social services and mental health agencies, especially from the PSU's that contain large cities.

Are cases weighted in a manner which provides for credible, national estimates?

Probably yes. My confidence in the work would rise if Westat answered questions I have about how some of the statistical formulas used were derived, filled in omitted steps in argument and exposition, and demonstrated that the weighting methodology was executed without major errors.

Chapter 6 of the Report on Data Processing & Analysis presents the weighting and estimation methodology. The statisticians provided a methodology so that each countable case could be assigned a weight for each of a number of properties it might have. A final case weight, derived from the individual weights, represents an estimate of the number of cases with these properties one might expect to see if a national census of maltreated children were taken.

I believe that the statisticians considered all the properties of cases that needed weighting. However, in reviewing Chapter 6, I was not always sure what they had done. I can't tell whether all of the quantities derived in the tables of Chapter 6 were calculated from the appropriate formula. There are entries in both the text and tables that are wrong. Clearly, the quality of Chapter 6 is inferior to the quality of other parts of NIS-2 I have reviewed. These problems may cast doubt on the validity of the final incidence figures.

Because of the length and complexity of chapter 6, I am not going to try to summarize and review each aspect of the work. Instead, I will state my questions and concerns with the corresponding page references.

Non-response adjustments for agencies were made across PSU's, within agency type and by non-response classes within agency. Is this sufficient? If the causes of non-response are related to countable cases encountered by an agency, then further adjustment would be required to avoid bias. There are two problems: non-response of an agency and non-response to particular items or questions on a report form. (See Statistical Analysis with Missing Data, Roderick J.A. Little & Donald B. Rubin, John Wiley, 1987. Also, see Multiple Imputation for Nonresponse in Surveys, Donald B. Rubin, John Wiley, 1987.)

I am not sure what the quantity n_h on page 6-3 represents. Section 6.2.1.1 says that 110 hospitals were originally selected, and on page 6-4 n_h is defined as "the number originally selected." The selected column in Table 6-1 totals 104. This disagrees

with 110, the number originally selected. It also disagrees with 105, the number obtained by adding to 110 the 4 replacement hospitals and subtracting from 114 the 9 out-of-scope hospitals. If n_h is indeed the number of hospitals originally selected, I am puzzled by the formula for P_g on page 6-3 whose denominator in the numerator of the entire expression is $n_h + a_h + c_h$. This seems to double count the last two terms which are already present in n_h . I would have thought that the proportion out-of-scope in a stratum would have been $(a_h + b_h) / n_h$.

I would appreciate written clarification on the definition of quantities as well as the argument that led to this formula and the formula for hospital weights on page 6-4 which uses P_g . These formulas or variants of them are used repeatedly in Chapter 6 to assign weights to the various agencies.

In table 6-1 my calculation of the out-of-scope adjustment for hospital strata 5,6 and 7 in PSU 25 yields 0.80, not 0.93 as reported. Four out-of-scope adjustments of 0.81 reported for PSU 5, strata 5 and 6, and PSU 6, strata 3 and 5, should be 0.80 and not 0.81. And why is stratum 6 in PSU 6 not combined with stratum 5 as it is in PSU 5?

Table 6-2 also confuses me. For small hospitals (stratum 8), table 6-1 shows 1 hospital that refused and was not replaced. Whether the adjustment factor is calculated as $10/9$ or as $11/10$, the result is not 1.016999 as in the table. Perhaps size measures were used to calculate the adjustment factors. How were the adjustment factors for non-response in Table 6-2 calculated?

For schools on pages 6-7 and 6-8, the questions on the formulas for P_g and for W_h , the agency weight for stratum h in group g , carry over from my queries about hospitals.

Additionally, I have some problems with Table 6-3 which derives school agency weights. First, PSU 18, stratum 1 shows 85 schools sampled out of 8 in the frame. This is a typo, and should have been 5 schools.

Page 6-7 states that replacement selections were made for 17 schools. The total shown in Table 6-3 is 15.

For the purpose of calculating the estimated proportion out-of-scope, the strata that were combined sometimes varied by PSU. For example, in PSU 3, strata 4 and 5 are combined to yield an estimate for both strata of 0.2981. In PSU 17, strata 5 and 6 are combined to produce a joint out-of-scope estimate of 0.2241 while stratum 4 gets an out-of-scope estimate of 0. PSU 25, stratum 5 gets an out-of-scope estimate of 0.5 while its stratum 4 gets an out-of-scope estimate of 0. Can Westat supply a rationale for the apparent inconsistencies?

On page 6-8 I do not see where the formula for day care center weights W_{ij} comes from. Also the total sample size in table 6-5 is 124. Above the table it is claimed that "Of the 141 day care centers recruited, 16 ... refused" leaving 125 in the sample.

Also, in this table I am not sure how the adjustment factor was calculated. Was the cumulative size of 311 centers in a region-urbanization category divided by the cumulative size of all cooperating centers in that category?

For social services and mental health agencies the weight formula given on page 6-14 and the weights calculated in table 6-6 have no out-of-scope adjustment. Yet, on page 6-14 Westat declares that "of the 218 selections, 110 were out-of-scope." Why did Westat not make out-of-scope adjustments here when they had previously?

Additionally, refusals were not, as with other agencies, handled with a separate adjustment that, I presume, would be used as a multiplier of the weight. Rather, as with out-of-scope agencies they were included in both the numerator and denominator of the single weight calculated for these agencies. Given the absence of stratification for social services and mental health agencies, I don't think this approach biased the weights.

For the sampling of municipal police departments, out-of-scopes were included in both the frame count and sample count on page 6-17. Again, I don't understand why Westat departed from previous practice with these agencies.

Incidentally, in this portion of the chapter the tables get out of sequence. Table 6-7 precedes table 6-6 and table 6-9 precedes table 6-8.

On page 6-23 the assumption under which duplication weights were derived for reports of countable cases is given. This is "that no more duplication occurred in the population than was seen in the sample. That is, if a case was found to be duplicated, it was assumed that there were no more duplicates in the population. When no duplicates were found in the sample for a given case, it was assumed to be entirely unduplicated in the population."

If I understand this assumption correctly as applying to the entire PSU, it must lead to an upward bias in the estimate of incidence. Suppose, for example, we find an unduplicated case in a hospital. Can we assume that the case has not been recorded in any other non-sampled agency of whatever type? Might there be a correlation with severity of maltreatment and duplication so that the assumption causes a bias that varies with severity? Perhaps cases with moderate harm to the child are more apt to be duplicated since cases where the injury is severe or fatal are not as likely to be transferred from one agency to another. And what distinction, if any, is made between the duplication of a single maltreatment and two maltreatments that occur within the three-month sampling window?

On pages 6-21 and 6-22 "Exit Evaluation Adjustment" is described. Each participant was rated on a 0 to 5 scale on the level and quality of his/her participation. Each group weight was inflated by the ratio of the total number of participants to the number of participants with favorable (3 or higher) ratings. We are not told how this group weight was derived. I am also uncertain whether individual participants were separately weighted or whether a single weight for the group was derived. And I would

like to know more about the criteria used to rate each participant on the 0 to 5 scale. If the phenomenon of partial participation was at all widespread it seems important to assess it as accurately as possible. Otherwise large errors and/or biases would degrade incidence estimates.

On page 6-25 a formula for weighting intra-agency duplicates from the non-CPS sector is given. I do not understand how that formula was derived.

Similarly, I do not understand why the exponent $(1 + m/n)$ is needed in the formula on 6-26. 'm' is the number of CPS short form reports, and 'n' is the number of CPS long form reports.

Westat describes the process of trimming its final weights to reduce sampling error on pages 6-28 through 6-29. On page 6-29 they claim that "the increase in precision outweigh[ed] the addition of bias." The reduction in the mean weight was 3% from 163.3 to 158.3, and the reduction in the coefficient of variation of the weights was 17%.

I am not an expert on trimming, and have only seen it applied to data in both tails of a distribution. But I wonder whether bias could not have been controlled better if some of the low-end weights had been raised so that the mean remained stable. Another possibility might have been to apply some compressive function such as a log to the weights. I am not saying that I disagree with what the Westat statisticians did, but that I am uncomfortable with a technique that invariably produces a bias in one direction.

On page 6-36 I am not familiar with the jackknifed variance estimation of Rust and Kalton. Would Westat be willing to send me a copy of their article? The reference is K. Rust & G. Kalton (1987), "Strategies for Collapsing Strata for Variance Estimation", J. Official Statist., 3(1), 69-81.

Were the comparisons with the 1980 data based on the accurate and appropriate, use of various, statistical procedures?

Yes.

Appendix E of Final Report: Appendices describes how the over time analyses were performed. The standard techniques of t-tests and Analysis of Variance were used to derive estimates. However, two modifications were introduced to compensate for the multiple and complex dependencies introduced by multi-stage cluster sampling so that these procedures produced statistics described as "t-like" and "F-like."

The first technique, that of jackknifing estimates of variance, I mentioned at the end of the last section. As I understand it, in NIS-2, 28 estimates of the mean are made by dropping one PSU at a time, reweighting up the three other PSU's in its foursome of

demographically similar PSU's, and leaving the remaining 24 PSU's with their original weights. A comparable technique was used for NIS-1 PSU's. The variance of these estimates around the estimate of the mean made from the full sample can then be computed. This number is a superior estimate of the true, but unknown, variance since it does not assume that maltreatment cases came from a simple random sample. Rather, the jackknifed variance estimate will reflect whatever dependencies arise from the cluster sampling design even though we may not be aware of them.

The second modification was a reduction in the degrees of freedom associated with a test statistic. Westat denotes the t-like and F-like statistics by t' and F' . A degree of freedom represents one independent measurement of the value of a variable. Clearly, the degrees of freedom can not be based on the case weights as each observation has been weighted up by one to several orders of magnitude to generate national estimates.

What Westat reports having done on page E-7 is to take as the maximum number of degrees of freedom available the number of independent jackknife replicate estimates used in generating the [variance] estimate (21 for the NIS-2 estimates)," and, I believe, 13 for the NIS-1 variance estimates.

Where the F' statistic is used, as in determining whether there is age by year interaction in the incidence of sexual misuse or physical abuse between studies, further reduction in degrees of freedom was made. Age was collapsed into 6 levels for the purpose of comparing NIS-1 and NIS-2 incidence. This would, I believe, imply $6*((13-1) + (21-1)) = 192$ degrees of freedom for the error sum of squares estimate in an ANOVA. Westat essentially argues on page E-7 that because the samples in different cells are not independent and the variance within cells may not be constant, the true degrees of freedom would lie between $(13-1) + (21-1) = 32$ and 192. Their actual quote is that the "degrees of freedom for the SS within is between a number somewhat less than q and $m*q$." I think they must have meant 'a number between q and $m*q$.'

In the example just given I wonder if the degrees of freedom might be specified more exactly. It might entail working with the ratio of probabilities of drawing the obtained sample under random sampling to drawing it under the multi-stage cluster sampling actually used. Furthermore, since the constancy of variance within cells can and should be statistically tested any time an ANOVA or t-test is conducted, should degree of freedom reductions be made unless test results indicate non-constancy of within-cell variance?

In reporting final results Westat based degrees of freedom on weights with footnotes indicating when degrees of freedom based on observations over all cells, or observations in a single cell would have made significance levels marginal or non-significant. In my opinion, the body of the report should have used degrees of freedom based on observations over all cells since false positives may result in misdirected effort and expenditure.

Would Westat be willing to clarify a question I have on notation? On page E-7 'q' is defined as "the number of independent replicate estimates available for each cell." On the next page, E-8, 'q' is defined as "the number of levels for the factor under consideration." I had thought that 'm', defined on page E-7 as "[the number of] independently sampled cells" was equivalent to the definition of 'q' on page E-8.

Were the projections that were made on the two sets of definitions used in NIS-2 valid?

If I may replace the word "projections" with the word "estimates," my answer is yes.

Page 2-6 of Study Findings refers to the two sets of definitions used for determining countability in NIS-2 as "original" and "revised." The original definitions were identical to those of NIS-1 "concerning both the perpetrator of the acts/omissions and the degree of harm to the child." The revised definitions expanded the degree of harm to include endangerment and relaxed criteria as to what constitutes perpetration.

Comparisons between years were made using the more restrictive original definitions. Estimates confined to NIS-2 cases were made using the expanded, revised definitions. Within NIS-2, the sample of cases meeting the original definitions was a subset of the sample of cases-meeting the expanded definitions.

From my reading of Chapter 6 of the Report on Data Processing & Analysis, I do not think separate weights were (or needed to be) calculated for the cases in the "original" and "revised" NIS-2 samples. However, incidence estimates for NIS-1 were re-estimated by the same method by which NIS-2 estimates were made. Westat found that the NIS-2 method led to more precise incidence estimates (page 6-31, section 6.7).

They proceed, on page 6-31, to describe a difference in the way population totals were incorporated into estimation of incidence rates as opposed to estimation of total cases: "Estimates of incidence rates were found to be more precise when these were based entirely on the sample (i.e., used estimates of the U.S. population totals based on the sample of study counties)... However, estimates of totals were more accurate when these incorporated more reliable population size figures for the period in question using census data."

How this was implemented is not evident to me from the formulas for incident rate, r , and estimated total, y' , on page 6-32. In the formula for incidence rate P_h "denotes the population of persons under 18 in PSU h ." The formula for total number is $y' = rP$, where "P denotes the population of persons under 18 in the U.S." Does the second part of the quoted passage in the preceding paragraph merely mean that the incidence rate was multiplied by the census population of the U.S.?

What if any changes would you suggest in the sampling or methodology if the study was to be repeated in several years?

If the primary focus of the study is with changes in incidence, I would select the same 29 counties that were used in NIS-2. I would also determine whether the population densities (urbanization) of those counties was representative of the entire nation.

From the 28 PSU's formed by these counties, I would randomly select one or two small PSU's with the goal of surfacing a higher proportion of the child maltreatment incidents that occurred within the PSU. On page 7-2 of the "Summary and Conclusions" chapter of Study Findings Westat lists numerous sources of knowledge about child maltreatment that were not tapped by NIS-2. These include "private schools, private physicians, medical clinics not affiliated with hospitals or health departments, clinical social workers or mental health professionals in private practice ... neighbors, relatives and the children themselves." Westat concludes: "Thus, the estimates provided by this study should be regarded as minimum estimates of the numbers of abused and neglected children." I would have added that the study estimates must be underestimates unless we believe that all of these listed sources taken together would add a negligible number of new countable cases to the pool.

In a subsequent study, I urge that statistical sampling of these sources be taken within the one or two PSU's selected for comprehensive coverage. Field interviews and investigations might enable us to uncover a substantial proportion of countable cases that are now, presumably, undetected.

In a subsequent study, I would also try to obtain more evidence on the cause(s) of the increase between 1980 and 1986 in the incidence of moderate sexual and physical abuse that widened with increasing age of the child. See Section 5.2.3 of Study Findings. This is an important question, for if the observed increase in incidence is due, in substantial part, to an increase in maltreatment of older children, then there has been a dramatic increase in violence towards older children in the space of 6 years.

In Chapter 7, "Summary and Conclusions," of Study Findings, Westat examines three alternative explanations to the increase cited in the previous paragraph: study methodology, increased reporting, and increased recognition. Westat essentially rejects methodology and reporting as causes of the increase in incidence by appealing to quantitative evidence from the study. I would like to see these arguments presented more fully with the quantitative evidence that supports them.

Having rejected the hypotheses of methodology and reporting, Westat accepts the hypothesis of recognition. Their argument is, basically, that if violence toward children had increased over the six year period, we would expect it to manifest rather uniformly across all categories of severity, across emotional as well as sexual and physical abuse, and across all ages.

Two additional arguments might be made to support Westat's position. The first is based on a study conducted by Richard J. Gelles and Murray A. Straus. They found that "in a telephone survey of two parent families with children over three years of age ... [there was] a decrease in the self-reported incidence of physical abuse by parents between 1975 and 1985." (See Study Findings, page xi.) Suppose, that during that decade, abusing parents felt themselves to be at greater risk of detection and punishment by the state. I would then expect a decrease in self-reported incidence as well as an increase in incidence by community professionals due to increased recognition and increased reporting.

The second argument concerns the magnitude of the increase. Incidence of physical abuse had increased by 58% and sexual abuse more than tripled between 1980 and 1986. (See Study Findings, page xxi.) I would only find changes of that magnitude credible if the level of stress and of violence in the society at large had climbed dramatically over the six year period.

One way of testing the "recognition" hypothesis would be to survey the media over the six-year period. On television, in newspapers, in magazines, in the professional journals, is the amount of space devoted to child maltreatment increasing? Of the total space devoted to child maltreatment in a particular medium, has the proportion devoted to sexual or physical abuse of older children been increasing? During the 1980-86 period, have child protection laws and regulations been expanded to cover older children? Do community professionals feel their attention on child abuse, particularly older children, has been increasing over the last five or ten years?

Sincerely yours,

/s/

Thomas J. Marx, Ed.D.

MARX SOCIAL SCIENCE RESEARCH, INC.

196 Appleton Street
Cambridge, Massachusetts 02138
(617) 876-0962

June 1, 1989

Deborah Daro, DSW
Director
Center on Child Abuse Prevention Research
National Committee for Prevention of Child Abuse
322 S. Michigan Avenue, Suite 950
Chicago, IL 60604-4357

Dear Deborah:

As a statistician I appreciate Karl Ensign's desire "to pin down Marty Frankel and Tom Marx as much as possible concerning the effects of bias on the measurement of abuse and neglect." If I could have delivered to you the type of quantitative bias estimates that Karl depicts in his graph, broken out by source of bias, I would have.

Unfortunately, after I completed my review of NIS-2, I felt that I did not possess sufficient information to make quantitative estimates of the effects of bias by source. These information lacks fell into three categories: What did Westat do? What were the means, variances etc. of quantities that would permit an estimate of bias? Did the incomplete and careless work in the exposition, explanation and presentation of results, formulas and procedures carry over to data collection, data cleaning and data processing?

In the rest of this letter, I am going to respond to the points Karl raised in his letter to you of 10 May. Then I will recommend how I think HHS ought to fund NIS-3.

Karl's Letter

Slope A in Karl's graph is possible except I don't know to what extent cities were underrepresented in NIS-1, what the difference in incidence rates between cities and non-cities was in either NIS-1 or NIS-2, or by how much duplication in NIS-2 biased incidence estimates upwards. (I must admit that our long, conference telephone conversation on the duplication question didn't clear it up for me.)

In calculating annual rates, Westat adjusted both for seasonality and multiple, different episodes of abuse within the year. The duplication adjustment was applied independently. For this reason I don't think annualization interacts with duplication bias, if it exists.

Data trimming does introduce a small bias in the mean. Westat found a 3% downward bias in the trimmed mean. They traded this downward bias for a 17% reduction in the coefficient of variation. This implies a 19% reduction in the standard deviation. Although the SD may be biased from events that took place before trimming, the SD is not biased by trimming. It is the SD of the weighted and trimmed data. The reason Westat trimmed was to achieve a large reduction in the SD thereby increasing the power of statistical tests of significance.

Karl thought that the reduction in variance (or equivalently SD) affected the slope of the line connecting the Abuse/Neglect incidence rate estimates of NIS-1 and NIS-2. This is not so. The variance reduction makes it more likely that an incidence rate change between NIS-1 and NIS-2 is detected by a t' or F' test.

Karl advocates that "one should pay attention to the effects of data weighting." I couldn't agree more. With case weights differing by several orders of magnitude and an extremely complex, multi-stage, weighting procedure, the potential for huge, unpredictable biases exists. The questions I had posed on weighting in my report to you were not answered in our conference call. (I think a better way to have answered these questions would have been through a letter to you, Frankel and me from Westat's statistician.) Additionally, how do we know that the complex weighting scheme, even if correct, was correctly implemented by the Westat computer programs?

My concern about the undersampling of social services and mental health agencies in large PSUs cited by Karl leads not to increases in bias, but to increases in the error or uncertainty of incidence estimates. In other words, the undersampling in large PSUs inflates variance, thus decreasing power. (To provide a quantitative estimate of the inflation, more information would be required than I found in the Westat reports.)

Recommendations for NIS-3

NIS-3 should be preceded by a methodological study that covers study design, sampling, data collection, data cleaning and statistical analyses. The methodological study should begin with an audit of NIS-1 and NIS-2.

What's worked from the first two surveys should be kept. What hasn't worked should be improved. If the audit uncovered errors that had a major impact on NIS-1 or NIS-2 estimates and statistical tests, the data should be reanalyzed.

One thought I had on the duplication issue was that this attribute of the universe of cases may be captured by random sampling just as any other attribute (such as age of victim) can. As long as cases are randomly selected from the universe, the duplication rate in the sample will estimate the duplication in the universe.

If the sample size is chosen to attain some specified power, additional cases could be sampled to make up the loss in sample size occasioned by duplicates. If an

initial sample of size n had a duplication rate of r , then a second stage sample of size $rn/(1-r)$ would increase the sample size enough to approximate the desired power. If a two-stage sample draw were infeasible, r might be guessed and a single sample of size $n/(1-r)$ selected.

Once the methodological study was completed and approved, NIS-3 proper could begin. In funding the methodological study and NIS-3, I recommend that HHS be sure the contractor(s) have satisfactory fees and ample deadlines. In particular, if funds are limited, I think HHS should reduce the scope of the study rather than asking the contractor to complete the full scope as best it can. Without knowing anything about why Westat produced some inferior work, I would guess that it was deadline pressure, money pressure, or both.

Sincerely yours,

/s/

Tom Marx, Ed.D.

APPENDIX B. LIST OF RESPONDENTS

Jose D. Alfaro, Director
Personnel Training & Research
The Children's Aid Society
150 East 45th Street
New York, NY 10017

Dr. Richard Barth
School of Social Welfare
University of California
120 Haviland Hall
Berkeley, CA 94720

Linda Blick
National Resource Center on Child
Sexual Abuse
The Chesapeake Institute
11141 Georgia Avenue, S-310
Wheaton, MD 20902

Foster Centola
Illinois Dept. of Children and Family
Services
406 West Monroe
Springfield, IL 62701

Dr. David Finkelhor
Family Research Laboratory
University of New Hampshire
Durham, NH 03824

Dr. James Garbarino, President
Erikson Institute
25 West Chicago Avenue
Chicago, IL 60610

Dr. Richard Gelles, Dean
College of Arts & Sciences
University of Rhode Island
Kingston, RI 02881

Dr. David Gil, DSW
Heller Graduate School
Brandeis University
Waltham, MA 02154

Dorothy V. Harris
11154 Wood Elves Way
Columbia, MD 21044

John Holtkamp, Program Mngr.
Child Protective Services
Iowa DHS
Hoover State Office Building, 5th Fl.
Des Moines, IA 50310

Helaine Hornby
National Child Welfare Resource Center
for Mgmt. & Admin.
University of Southern Maine
246 Deering Avenue
Portland, ME 04102

Beverly Jones
APWA
810 First Street, NE, S-500
Washington, DC 20002

Dr. Richard Krugman, Director
C. Henry Kempe National Center for the
Prevention and Treatment of Child
Abuse and Neglect
1205 Oneida Street
Denver, CO 80220

Penny Maza
Administration of Children, Youth and
Families
Div. of HHS
929 L Street, NW
Washington, DC 20001

Dr. Eli Newberger, Director
Family Development Study
Children's Hospital Medical Center
300 Longwood Avenue
Boston, MA 02115

Mr. Edward Van Dusen
Dept. of Health & Welfare
Division of Family & Children Svcs.
Statehouse
Boise, ID 83720

Betsey Rosebaum
National Association of Public Child
Welfare Administrators
APWA
810 First Street, NE, S-500
Washington, DC 20002

Patricia Schene, Director
American Association for Protecting
Children
9725 E. Hampden Avenue
Denver, CO 80231

Dr. Murray A. Straus
Professor of Sociology
University of New Hampshire
Durham, NH 03824

Susan Webber, Director
National Center on Child Abuse &
Neglect
P.O. Box 1182
Washington, DC 20013

Susan Wells
Director of Child Abuse Research
National Legal Resource Center for
Child Advocacy and Protection
ABA
1800 M Street, NW
Washington, DC 20036

Mr. Richard Winters
Dept. of Social & Health Services
Div. of Children & Family Svcs.
MS, OB-41
Olympia, WA 98504

To obtain a printed copy of this report, send the full report title and your mailing information to:

U.S. Department of Health and Human Services
Office of Disability, Aging and Long-Term Care Policy
Room 424E, H.H. Humphrey Building
200 Independence Avenue, S.W.
Washington, D.C. 20201
FAX: 202-401-7733
Email: webmaster.DALTCP@hhs.gov

RETURN TO:

Office of Disability, Aging and Long-Term Care Policy (DALTCP) Home
[\[http://aspe.hhs.gov/office_specific/daltcp.cfm\]](http://aspe.hhs.gov/office_specific/daltcp.cfm)

Assistant Secretary for Planning and Evaluation (ASPE) Home
[\[http://aspe.hhs.gov\]](http://aspe.hhs.gov)

U.S. Department of Health and Human Services Home
[\[http://www.hhs.gov\]](http://www.hhs.gov)