Pergamon

Child Abuse & Neglect 29 (2005) 461–477

Child Abuse & Neglect

Describing maltreatment: Do child protective service reports and research definitions agree?☆

Desmond K. Runyan[a,b,c,*], Christine E. Cox[a,c], Howard Dubowitz[d], Rae R. Newton[e,f], Mukund Upadhyaya[g], Jonathan B. Kotch[h], Rebecca T. Leeb[c], Mark D. Everson[i,c], Elizabeth D. Knight[c]

[a] *Department of Social Medicine, Campus Box 7240, The University of North Carolina School of Medicine, Chapel Hill, NC 27599–7240, USA*
[b] *Department of Pediatrics, The University of North Carolina School of Medicine, Chapel Hill, NC, USA*
[c] *Injury Prevention Research Center, The University of North Carolina, Chapel Hill, NC, USA*
[d] *The Department of Pediatrics, The University of Maryland School of Medicine, Baltimore, MD, USA*
[e] *Department of Sociology, California State University Fullerton, Fullerton, CA, USA*
[f] *Child and Adolescent Services Research Center, San Diego, CA, USA*
[g] *Washington State Department of Social and Health Services, Seattle, WA, USA*
[h] *Department of Maternal and Child Health, The University of North Carolina School of Public Health, Chapel Hill, NC, USA*
[i] *Department of Psychiatry, The University of North Carolina School Medicine, Chapel Hill, NC, USA*

Received 6 August 2002; received in revised form 3 June 2004; accepted 24 June 2004

Abstract

Objective: The National Research Council identified inadequate research definitions for abuse and neglect as barriers to research in child maltreatment. We examine the concordance between child protective services (CPS) classifications of maltreatment type with the determinations of type from two research coding systems. We contrast the two coding systems and the CPS classification, in their ability to predict subsequent difficulties in the psychological functioning of maltreated children at age 8.

Method: The sample included 545 children who were enrolled in Longitudinal Studies of Child Abuse and Neglect (LONGSCAN) with a report of child maltreatment, had data collected at approximately 4 and 8 years of age, and had a lifetime review of CPS records to age 8. CPS Maltreatment reports were coded using LONGSCAN's modification

☆ Research supported by Grants CA-90CA1401, 90CA1433, and 90CA1467 from the Administration For Children, Youth and Families and the Children's Bureau; DHHS & 1 RO1 HD039689 from The National Institute of Child Health and Human Development.

* Corresponding author.

0145-2134/$ – see front matter
doi:10.1016/j.chiabu.2004.06.015

of the Maltreatment Classification System (MMCS) and the Second National Incidence Study maltreatment coding system (NIS-2). The first analyses used reports as the unit of analysis to examine agreement between CPS and research determinations of allegation type. Validation analyses examined outcomes for each type of maltreatment experienced after age 4 under each coding system using the Child Behavior Checklist (CBCL), Trauma Symptom Checklist-Alternative form, and the Vineland Screener as the measures of outcome. Control variables were the CBCL and Battelle Developmental Screener, measured at age 4.

Results: There were a total of 1980 reports of maltreatment for 545 study children although only 1593 CPS reports specified at least one type of maltreatment. There were differences between the type of maltreatment recorded in child protective service records and the conclusions reached by either research classification system. CPS classifications were most discordant with the research systems for emotional abuse and neglect. Nearly 10% of physical and sexual abuse reports, as determined by the MMCS, were classified as neglect by the child protective service agencies. The NIS-2 system and the MMCS had very high Kappa statistics for agreement for physical and sexual abuse. The validity of the research definitions for physical and sexual abuse was demonstrated in models predicting children's functioning at age 8. Prediction of child functioning was significantly but modestly improved in several domains compared to the CPS classifications.

Conclusion: Both research classification systems moderately improved on the prediction of the adverse effects of maltreatment compared to the characterization of a maltreatment exposure as recorded by CPS.

Keywords: Research definitions; Outcomes; Child abuse; Neglect; Emotional maltreatment

Introduction

Systematic inquiry into the impact of child maltreatment on subsequent development has been a clear need since the modern recognition of the problem of child abuse (Kempe, Silverman, Steele, Droegemueller, & Silver, 1962). Among the first studies of outcome was the study by Elmer (1977). However, systematic inquiry into impact and intervention has lagged. In 1993, a US National Academy of Sciences panel systematically reviewed abuse and neglect research and identified a number of inadequacies in the knowledge base of the field (National Research Council, 1993). The panel cited the absence of clear definitions of child maltreatment as a significant deterrent to progress and recommended the development of research definitions as essential for advancement. Prior to the panel's report, attempts to develop standard definitions included the National Incidence Studies (NIS) (National Center on Child Abuse and Neglect, 1988) and the Maltreatment Classification System (MCS) (Barnett, Manly, & Cicchetti, 1993). The NIS effort explicitly attempted to operationalize definitions that could be used to classify reports from child protective service (CPS) agencies and others across multiple legal jurisdictions. The difficulties of defining types of maltreatment and the ambiguousness of measures used in research remain a concern for researchers, practitioners, and policymakers (Haugaard, 2000).

Longitudinal Studies of Child Abuse and Neglect (LONGSCAN) is a consortium of five longitudinal studies of child abuse, begun in 1991, using common measures, definitions, interviewing approaches and data entry systems. The samples were selected using varying criteria, and together roughly represent a continuum of maltreatment experiences (Runyan et al., 1998). LONGSCAN's focus on understanding the impact of abuse and neglect, as described in this issue by English, Bangdiwala, and Runyan (2005), requires reliable information on the exposure of the children to maltreatment in samples from five states, each with significant variation in the legal definitions of maltreatment.

To ensure the most complete and accurate data possible on maltreatment experienced by the children, LONGSCAN has adopted a multiple method, multiple source approach. This includes caregiver report, youth self-report beginning at age 12, and review of official records, including bi-annual review of CPS narrative case records and semi-annual review of state-level Central Registry data.

Reliably coding CPS data poses specific challenges as the form, contents, and completeness of records vary across the four sites contributing data to this analysis. Any systematic approach to coding these data must accommodate varying legal definitions, CPS and judicial practices, and different record keeping systems. As outlined in the paper by English et al. (2005), LONGSCAN has adopted a protocol for coding maltreatment allegations documented in child protective service narrative records using the original CPS classification assigned to the allegation and two alternate coding systems. These include a modified version of the Maltreatment Classification System (MCS) developed by Barnett et al. (1993) and the coding system used by the Second National Incidence Study (NIS-2) (Sedlak, 1986). Each system has its own approach to coding severity, chronicity and type. LONGSCAN's modified version of the MCS, hereafter referred to as the MMCS, added greater specificity to intermediate severity codes, dropped the original chronicity codes, and disaggregated a variety of neglect codes combined in the MCS.

A related research challenge involves the issue of how to assign a maltreatment type code to allegations that often include multiple forms of maltreatment. The multiple possible combinations of physical abuse, sexual abuse, psychological abuse and neglect yield an unwieldy number of coding categories. Research indicating that the effects of maltreatment on child outcomes vary by the type of maltreatment experienced has led researchers to argue that the effects of different types of maltreatment should be examined separately (Manly, Cicchetti, & Barnett, 1994). But on what basis does one develop a hierarchy for deciding what type of maltreatment should be regarded as most important? This question is addressed separately in this issue by Lau et al. (2005).

What are needed are empirical comparisons of different maltreatment coding systems and their ability to predict adverse outcomes among maltreated youth. LONGSCAN, with its relatively large sample of maltreated youth with longitudinal data collected in multiple states, has the ability to compare how reports are classified using the different systems of coding and to examine the predictive ability of the three systems with respect to child outcomes.

In this paper we explore the concordance of the alternate systems in classification of reports and examine their abilities to help understand the impact of maltreatment on children. The two primary objectives of the current study are to (1) compare the concordance of maltreatment determinations made by CPS with the two research coding systems: NIS-2 and the MMCS; and (2) examine validity by comparing their predictions of child outcome at age 8. This approach is thought to be a useful method to clarify the importance of using more precise research definitions to study the impact of maltreatment.

Method

Sample

Data were drawn from the 1354 children enrolled in the *Longitudinal Studies of Child Abuse and Neglect* (LONGSCAN). The origin of the samples has been described elsewhere (Runyan et al., 1998; English et al., 2005). We restricted the sample to include only children who met the following criteria: (1) the child and parent completed age 4 and age 8 interviews; (2) there was a reported allegation of maltreatment

that occurred before the age 8 interview; and (3) their CPS records were searched and reviewed by a LONGSCAN reviewer through the date of the age 8 interview. Children enrolled in LONGSCAN but never reported for child abuse or neglect were excluded. Based on these criteria, the final sample includes 545 children and their primary caregivers. The four data collection sites represented in the final sample are referred to geographically as the East, South, Southwest, and Northwest Sites. All LONGSCAN children from the two west coast sites had been reported for maltreatment by the time of the age 8 interview, whereas a number of children from the Eastern and Southern sites had not been reported and therefore were not included. Thus, 78% of those included live on the West Coast of the United States, while the remaining portion of the sample live on the East Coast. Data from the four sites are pooled and analyzed as a single sample, with site included as a control variable. Analyses were performed at both the report- and the subject-level. A graphic summary of the steps taken to generate the final subject-level sample for this article is in the introductory paper in this issue (English et al., 2005).

Demographic characteristics of participating children and their families at the time of the age 8 interview are shown in Table 1. The majority of the children were of minority ethnic status (i.e., Black, Hispanic, Mixed Race, Asian, and American Indian). The sample was nearly equally split between boys and girls. The range of educational attainment of the participating primary maternal caregivers was broad. On average, caregivers had approximately 12 years of education, and 31.2% of the caregivers had some post-high school education. Participating families had a median income of between $15,000 and $19,999.

Table 1
Sample characteristics at age 4 ($N = 545$)

	%	*N*	Mean (*SD*)	Median
Child gender				
Male	49.2	268		
Female	50.8	277		
Child ethnic status				
Majority status	31.9	174		
Minority status	68.1	371		
Child age (years)			8.2 (.39)	8.1
Caregiver marital status				
Married	33.9	185		
Single/never married	37.1	202		
Separated	9.5	52		
Divorced	17.6	96		
Widowed	1.8	10		
Caregiver education (years)			11.8 (2.31)	12.0
Family income (US$)[a]				$15–19.9K
Family geographic location (“Site”)				
East	12.1	66		
South	10.3	56		
Southwest	39.8	217		
Northwest	37.8	206		

[a] The range for the upper category of income was truncated, so it is not appropriate to report a mean for this variable.

Procedures

The dependent variables were collected during face-to-face interviews with the primary maternal caregiver and the child at both the age 4 and age 8 interviews. The recruitment, measures, and procedures of this study have been reported previously (Runyan et al., 1998) and summarized again in the article by English et al. (2005). Interviewers are trained in a common protocol at the Coordinating Center. Institutional review boards at every site and at the Coordinating Center have approved LONGSCAN protocols. Parents gave informed consent to participate in the study.

Measures

Maltreatment. CPS records are reviewed at least every 2 years at all sites using a project-developed protocol. This protocol permits abstraction of data from each allegation referral and related findings summary in an individual record. The original label of type applied by CPS to each type of maltreatment found in the allegation narrative and in the corresponding summary of findings is coded exactly as reported by CPS. The abstractor then codes the information from the allegation and findings summary narratives using two systems: the system used in the Second National Incidence Study (Sedlak, 1986) and a modification of the Maltreatment Classification System (MCS) developed by Barnett et al. (1993). The MCS has been found to be a valid and reliable system for classifying maltreatment and differentiating among subtypes (e.g., Bolger & Patterson, 2001). As described in English et al. (2005), LONGSCAN's modification of the MCS system (MMCS) includes greater specificity about severity and subtype of maltreatment. Abstractors were trained to achieve 90% congruence with their trainers in both the NIS-2 and MMCS systems before entering the field to code records.

Child outcome measures. Our efforts to study alternate Maltreatment Classification Systems require a selection of child developmental outcome variables measured both at age 4 and at age 8. We selected measures that spanned parent report, child report and interviewer assessment. Child behavior was assessed at ages 4 and 8 using the Total Problems, Internalizing, and Externalizing Scales of the Child Behavior Checklist (CBCL) (Achenbach, 1991). The CBCL is a widely used, empirically based measure of children's competencies and behavior problems. The validity and reliability of the 113-item Problem Checklist are well established (Achenbach, 1991). The Internalizing Problem scale combines three subscales, including the Social Withdrawal, Somatic Complaints and Anxiety/Depression scales. The Externalizing Problems scale combines the Delinquent Behavior and Aggressive Behavior subscales. *T* scores less than 60 are considered in the normal range, while *T* scores above 63 are considered in the clinical range, and scores of 61–63 are considered "borderline." We used standardized, or "*T*" scores as they provide an advantage over raw scores in that they are adjusted for child age and gender.

The Socialization and Daily Living Skills subscales of the Vineland Screener were used to assess child adaptive behavior at Age 8. The Vineland Screener (Sparrow, Carter, & Cicchetti, 1993) is an abbreviated form of the parent-report Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1984) that was developed exclusively for research use. The authors report inter-rater reliability coefficients of .98 with lay interviewers, and correlations between the full Vineland and Vineland Screener of the Daily Living Skills and Socialization subscales of $r = .92$ and $r = .93$, respectively (Sparrow et al., 1993). Higher scores reflect higher levels of functioning. While the original Screener items were administered intact, LONGSCAN developed and implemented a more structured set of item-level scoring criteria based upon the scoring

criteria in the full Vineland Survey Form (Sparrow et al., 1984), with additional probe items and behavioral descriptors for each possible score, to enhance standardization of administration and scoring across sites.

The Trauma Symptom Checklist-Alternative Version (TSCC-A) (Briere, 1996) was used to assess child mental health symptoms at age 8. The TSCC-A is an interviewer-administered, child self-report instrument of Post-Traumatic Distress and related psychological symptomatology designed for use with child victims of traumatic events. The alternate version includes five of the original six subscales, including Anxiety, Depression, Anger, Post-Traumatic Stress, and Dissociation. Higher scores reflect greater symptomatology. *T* scores are used in this analysis. *T* scores of 65 or greater are considered clinically significant. The authors describe high internal consistency for the five scales and good convergent, discriminant and construct validity documented in a variety of studies (Briere, 1996).

Control variables. Control variables include five demographic variables: child's gender, child's ethnic status (categorized as minority status versus majority status), child's actual age at the time of the age 8 interview, family income, and LONGSCAN study site; the number of maltreatment reports the child had between the age 4 and the age 8 interviews; and three measures controlling for behavioral, social, and adaptive functioning at age 4. All control variables were used in the regression analyses.

It is assumed that the effects of maltreatment experienced prior to the time of the age 4 interview will already have influenced our measures of child functioning at age 4. The child's *T* score on the CBCL Total Problems scale was used in models predicting age 8 scores on the CBCL Total Problems, Internalizing, and Externalizing scales, and in the models predicting scores on the five TSCC-A scales. For models predicting scores on the Vineland Screener Socialization and Daily Living Skills scales, we used the Personal-Social Skills and Adaptive Behavior domain standard scores from the Battelle Developmental Inventory Screening Test (Battelle Screener) administered at age 4 (Newborg et al., 1988). The Battelle Screener is a standardized, efficient measure of key developmental skills in children taken from the full Battelle Developmental Inventory that uses a structured administration. Trained interviewers administered items to the children, and conducted a face-to-face interview with the caregiver to assess those items that could not be scored based on observation of the child. The authors report test-retest reliability scores of .89 and .96 for the Adaptive and Personal-Social Domain scales respectively when administered to 4-year-old children (Newborg et al., 1988).

Analysis

Examining concordance

To address our first research goal, the three classifications of maltreatment type—the original CPS designation, the NIS-2 codes, and the modified MCS codes (MMCS)—were compared using both Kappa statistics to assess agreement beyond chance, and graphically using the methodology of receiver-operator characteristic curve (ROC) analysis (Fletcher, Fletcher, & Wagner, 1996). We conducted this analysis at the report level using only the reports with a single, valid CPS classification of maltreatment type. A variable representing CPS maltreatment type was coded from a 12-item checklist on the data collection form. The categories for this variable were physical abuse, sexual abuse, neglect, and emotional maltreatment. Items on the checklist that did not fit one of these types were considered invalid. These included "None given," "Dependency," "Caretaker absence/incapacity," "Moral/legal/educational," "Abuse (unspecified)," and

"Don't know." Out of 1980 reports 387 reports with no valid CPS allegation type codes and 167 reports with multiple types of allegations coded by CPS were excluded. An additional 163 reports were dropped because no valid MMCS type could be coded, yielding 1263 as the total N for the concordance analyses for CPS codes by the MMCS codes. For each of the four types of maltreatment, we coded an indicator variable for the NIS-2 and the MMCS coding systems using all available data from the six possible maltreatment type codes that could be recorded for each system.

We cross-tabulated the four-category CPS type variable with each binary variable representing the type of maltreatment allegation coded using the NIS-2 and the MMCS coding systems. Because of its extensive development, level of specification and established validity, the MMCS was treated as the "Gold Standard" or our best estimate of the truth, to which each of the other types of coding was compared.

In the ROC analysis, the sensitivity (classification of reports as abuse when the MMCS classification system determined that they were abuse) and specificity (reports not classified as abuse when there was no abuse coded using the MMCS system) were calculated. We contrasted the sensitivity and false positive rate of both the original CPS classification and the NIS-2 classification to the MMCS classification. In an ROC-type analysis comparing two measures for sensitivity and specificity without consideration of cost or difficulty, the preferred test is the test that is graphically closer to 100% sensitivity and 100% specificity (i.e., a 0% false positive rate). An alternate approach to establishing the relationship between the two alternative coding schemes is to examine agreement using the Kappa statistic. The Kappa statistic will be close to 1 if there is agreement in all of the reports and 0 if none of the reports were classified as the same type by the two coding systems.

We also examined the agreement between CPS type and our research classification of reports based on the MMCS by assigning a single, "predominant type" of maltreatment allegation coded under the MMCS system for each report. For this analysis, an algorithm was used to determine the predominant type for each report. First, the type of maltreatment allegation with the highest severity rating was assigned. If two MMCS types were coded as having equivalent severity, we then selected the type coded with both the highest severity and greatest frequency. If no type tying for greatest severity with another type was coded with greater frequency, type was assigned according to hierarchical ordering of sexual abuse, physical abuse, neglect, and emotional abuse. This hierarchy has been used by other researchers to classify predominant maltreatment subtype (Kinard, 1994; Manly et al., 1994), and mirrors our sense of the convention used in social service agencies.

Regression analyses

Finally, we examined the validity of the different coding systems by comparing the ability of each to predict children's functioning at age 8 using hierarchical regression analyses. These analyses were done at the subject-level. A total of 10 measures of functioning at age 8 were predicted with identical regression models for each type of maltreatment (i.e., physical abuse, sexual abuse, neglect, and emotional maltreatment experienced after age 4). The maltreatment variable in each model was systematically varied using the indicators of maltreatment type from each of the three coding systems. Outcomes examined included the Child Behavior Checklist Total Problem, Internalizing and Externalizing scores, the Vineland Screener Daily Living and Socialization scores, and the Anger, Depression, Anxiety, Dissociation, and PTSD scores from the Trauma Symptom Checklist.

Our procedure for coding the indicators of maltreatment types for each child was based on examining all reports occurring during the period between the age 4 interview and the age 8 interview. Approximately

half of the subjects ($n = 272$) had a total of 632 reports during this period for which one or more of the three classification systems coded a valid allegation type. A variable was created for each type of maltreatment (physical abuse, sexual abuse, neglect, and emotional maltreatment) for each coding system. These 12 variables were entered into otherwise identical regression models to predict outcomes in order to see whether one coding system was stronger than its alternatives. That is, for each type of maltreatment, the CPS variable for that type was entered into the model, then was replaced by the MMCS variable, then by the NIS-2 variable.

Variables were entered in blocks as follows:

Block 1: Demographic controls: Child gender, Child ethnic status, and actual age at the time of the age 8 interview.
Block 2: Control variables: Number of reports alleging the type of maltreatment of interest between age 4 and age 8; and either the age 4 CBCL Total Problems *T* score, or the Battelle Personal-Social Skills domain standard score, or the Battelle Adaptive Behavior domain standard score, depending on the outcome variable.
Block 3: Indicator representing the presence of the type of maltreatment of interest from each maltreatment coding system, entered in separate models.
Block 4: Income.
Block 5: Site.

To gauge the usefulness of the alternate classifications to describe the phenomena being experienced by the children, we assumed that harm would result from child abuse or neglect. We fitted competing regression models, alternating the variable entered as block 3, to predict outcomes on major measures of child well being at age 8. The competing models were inspected to see if one of the systems of coding emerged as more predictive of adverse outcomes.

Results

Concordance between CPS and research definitions

Agreement between the MMCS codes and CPS recorded allegations based on the predominant type coded from MMCS, for each reported allegation of child abuse or neglect in the social service records is presented in Table 2. Table 2 includes the number of reports and row and column percentages for each cell. The column percentages represent the percentage of agreement (or disagreement) of the CPS recorded allegations with the predominant type assigned using the MMCS codes. For example, agreement between the MMCS predominant allegation type of physical abuse and CPS recorded allegations for physical abuse is 82%. Similarly, agreement for sexual abuse, neglect and emotional abuse is 90%, 85% and 37%, respectively. The column percentages that lie off the main diagonal of Table 2 demonstrate the nature of disagreements between LONGSCAN's assigned MMCS codes and recorded CPS allegations. In column 1, 9% of reports assigned a MMCS code of physical abuse had been originally recorded by CPS as an allegation of neglect. Similarly, 9% of children assigned an MMCS code of sexual abuse upon record review originally had been recorded by CPS as neglect. The MMCS system increased the total number of reports classified as neglect, from the total coded as neglect by CPS, by re-classifying 36 reports of physical abuse, 32 reports of sexual abuse, and 11 cases of emotional abuse as neglect. The total number

Table 2
Type of maltreatment allegation designated by child protective services by predominant type of narrative-based allegation coded by LONGSCAN[a] for 1263 maltreatment reports from CPS records[b]

Coding scheme	Modified Maltreatment Classification System Codes[c]				
	Physical abuse	Sexual abuse	Neglect	Emotional	Total
Child protective service codes					
Physical	222	1	36	16	275
Row (%)	81	.4	13	6	
Column (%)	82	1	5	16	
Sexual	8	101	32	2	143
Row (%)	6	71	22	1	
Column (%)	3	90	4	2	
Neglect	25	10	663	35	733
Row (%)	3	1	91	5	
Column (%)	9	9	85	36	
Emotional	4	0	11	36	51
Row (%)	8	–	22	71	
Column (%)	2	–	1	37	
Other[d]	12	0	40	9	61
Row (%)	20	–	65	15	
Column (%)	4	–	5	9	
Total	271	112	782	98	1263

[a] Predominant type of maltreatment alleged when multiple types were coded was determined using the following algorithm: type with highest severity (excluding emotional maltreatment, as severity is coded using a different scheme), then type with highest severity and frequency when two types were coded with equal severity, then if no type was coded more frequently than any other, type was assigned in the following hierarchy: sexual abuse, physical abuse, neglect, and then emotional abuse.

[b] Row percents represent the percentage of the total number of reports classified as that type by CPS that were also classified as that type by the MMCS allegation codes, using the predominant type algorithm. Column percents represent the percentage of the total number of reports classified as that type by the MMCS allegation codes, using the predominant type algorithm, that were also classified as that type by CPS.

[c] Of 163 reports excluded from this table because LONGSCAN coders were unable to code the report as one of these four types using MMCS, CPS classified 22 (14%) as physical abuse, 29 (18%) as sexual abuse, 87 (53%) as neglect, 2 (1%) as emotional maltreatment, and 23 (14%) as some other type.

[d] Other type categories used by CPS included Dependency, Caretaker absence/incapacity, Moral/legal/educational, and unspecified abuse.

of cases assigned a code for emotional abuse in the MMCS increased to 98 from the 51 assigned by CPS. However, only 36 cases were coded by both CPS and the MMCS as emotional abuse. Overall, both CPS and the MMCS classified reports as the same type of maltreatment in 81% of 1263 cases, and 19% of cases were classified differently by CPS and by the MMCS coding system. NIS-2 and CPS similarly agreed for 81% of cases (not shown in Table 2).

Table 3 presents the sensitivity, specificity and predictive values for the NIS-2 and CPS codes predicting the MMCS code. It demonstrates the high level of agreement between NIS-2 and the MMCS for all types of maltreatment. For physical abuse, agreement was particularly strong with a positive predictive value of 99% for the MMCS code of physical abuse when the NIS-2 code was physical abuse for a report. The

Table 3
Sensitivity, specificity, positive predictive value and Kappa statistics for child protective service and National Incidence Study-2 Coding of Abuse Allegations using the modified Maltreatment Classification System as the Gold Standard

Type	Sensitivity (%)	Specificity (%)	Positive predictive value[a] (%)	Kappa[b]
CPS physical abuse	74	95	81	
NIS-2 physical abuse	98	100	99	.981
			(.975 substantiated cases)	
CPS sexual abuse	87	95	65	
NIS-2 sexual abuse	92	100	97	.961
			(.971 substantiated cases)	
CPS neglect	82	76	83	
NIS-2 neglect	84	93	94	.743
			(.785 substantiated cases)	
CPS emotional abuse	16	99	87	.224
NIS-2 emotional abuse	72	96	85	.722
			(.207 substantiated cases)	

[a] Positive predictive value is the probability that the type of maltreatment coded by the CPS system or by the NIS-2 system will be the same type as coded by the MMCS system.

[b] Kappa statistics were calculated for all 1920 records of maltreatment coded by NIS-2 and MMCS. CPS comparisons were limited to 1426 records with a single valid CPS code.

Kappa statistic for this comparison was .981. The agreement between CPS and the MMCS system was much lower, with a sensitivity of 74%, a specificity of 95% and a positive predictive value of 81%.

There was a similarly high level of agreement between the NIS-2 and MMCS codes for sexual abuse with 97% positive predictive value for the NIS-2 code predicting the MMCS code (Kappa = .961). CPS codes for sexual abuse had only a 65% positive predictive value with much greater misclassification.

The agreement between the NIS-2 and the MMCS codes for neglect was not as strong as for physical abuse, with a Kappa of .743 and a positive predictive value of 94% for a NIS-2 code of neglect predicting that the MMCS code would also be neglect. When the analysis was restricted to substantiated cases only, the Kappa between the NIS-2 and MMCS codes climbed to .785. There was greater disagreement between the CPS classification for a report of neglect and the assigned MMCS code. In particular, the CPS code of neglect had poorer specificity. A large number of cases recorded as neglect by CPS were "false positives" in that the MMCS coding reassigned 24% to other forms of maltreatment.

The type of abuse with the least agreement was emotional abuse. CPS recorded an allegation in just 53 reports of maltreatment. The MMCS codes for emotional maltreatment added another 239 reports that met research criteria for this type of maltreatment. The sensitivity was just 16% for a CPS report identifying emotional maltreatment as the type and the Kappa statistic for CPS and MMCS was low at .224 for all reports and even lower, at .207 for substantiated reports. In contrast, the agreement between the NIS-2 and MMCS codes was good with a Kappa of .722 although the positive predictive value for the NIS-2 code of emotional abuse predicting an MMCS classification of emotional abuse was just 85%.

The ROC curve analysis illustrates that for all types of maltreatment, there is relatively close agreement between the NIS-2 and MMCS codes and less agreement between the CPS and MMCS codes. In particular, a CPS code of emotional maltreatment is very insensitive although quite specific. The NIS-2 code for emotional maltreatment is more sensitive and somewhat less specific.

Regression analyses predicting child outcomes

Hierarchical regression models were used to assess the relative ability of the three different coding systems for maltreatment reported after the age 4 interview to predict child functioning at age 8. Ten outcome variables and four predictor variables, each measured by each of the three alternative coding systems, resulted in a total of 120 separate regression equations. We do not present the considerable output from this exploratory analysis in its entirety. (Complete tables for all outcomes are available from the corresponding author.) In this presentation, we focus on three specific questions. First, in the third hierarchical block, does entry of the maltreatment code for type of maltreatment into the regression model result in a significant increase in overall variance explained (R^2 change)? Second, is the unstandardized regression coefficient for the maltreatment type code statistically significant in the final regression model containing all control variables? And finally, are there equations in which there is a statistically significant amount of the variance explained by one of the coding systems that is not matched by the other two coding systems?

Table 4 presents the unstandardized regression coefficients and block 3 R^2 change values for all models that contained at least one statistically significant outcome. There were no significant findings for the TSCC-A scales, and these are not included in Table 4. Table 4 reveals four main findings. First, the results for MMCS and NIS-2 coding systems are very similar. Second, using the determination of maltreatment type by CPS, the entry of the maltreatment code in the third hierarchical block significantly increased the amount of variance explained in some of the outcomes for three forms of maltreatment: emotional maltreatment predicting CBCL Internalizing scores, physical abuse predicting CBCL Total Problems scores, and Neglect predicting Vineland Socialization. The amount of variance explained also increased significantly when the MMCS and the NIS-2 systems codes for sexual abuse were entered in the third block in all three models predicting CBCL broadband scores. While statistically significant, use of either the MMCS or NIS-2 research codes of physical abuse modestly contributed to the variance in Vineland Daily Living scores.

Third, sexual abuse accounts for over half of all significant outcomes (7 out of 13), counted as either a significant change in R^2 or a significant parameter estimate, and physical abuse and emotional maltreatment each account for almost a fourth (3 out of 13). Sexual abuse explains CBCL Internalizing scores for all three systems and CBCL Externalizing and Total scores only when coded by MMCS or NIS-2. Physical abuse explained a significant amount of the variance in CBCL Total Problem scores when coded by any of the three systems, and Vineland Daily Living scores only when coded by one of the research systems. Emotional maltreatment, when coded by MMCS, explains a significant amount of variance in CBCL Total Problems scores. When coded by CPS, emotional maltreatment explains a significant amount of the variance in CBCL Internalizing scores. Neglect only predicts Vineland Socialization scores, and only when coded by CPS.

Finally, the MMCS and the NIS-2 codes of sexual abuse were better predictors of CBCL Internalizing and CBCL Total Problems outcomes than CPS coding in models though the effect size was larger when coded by MMCS. For Vineland Daily Living scores, physical abuse was a significant predictor when coded by MMCS and NIS-2, and the effects were stronger when it was coded by NIS-2. CPS classifications of type were better predictors of outcomes than both the MMCS and NIS-2 codes in three cases. Physical abuse was a better predictor of CBCL Total Problem scores, neglect was a better predictor of Vineland Socialization scores, and emotional maltreatment was a better predictor of CBCL Internalizing scores than when these types were coded by the MMCS or the NIS-2.

Table 4
Block 3 R^2 change, final model β values, and post hoc comparisons

	CPS		MMCS		NIS-2	
	R^2 change	β	R^2 change	β	R^2 change	β
CBCL Internalizing						
Physical abuse	.002	−2.088	.001	−1.184	.001	−1.247
Sexual abuse	.006	4.645*	.022***	8.289***	.017***	7.314***
Neglect	.003	1.476	.001	.665	.003	1.316
Emotional maltreatment	.014**	7.108***	.004	2.511	.001	1.414
CBCL Externalizing						
Physical abuse	.001	−2.046	.002	−2.085	.003	−2.261
Sexual abuse	.003	3.339	.030***	10.293***	.020***	8.615***
Neglect	.001	1.554	.000	−.275	.000	.000
Emotional maltreatment	.001	.677	.002	2.573	.003	2.815
CBCL Total Problems						
Physical abuse	.005*	−3.612**	.003	−2.365	.004	−2.605
Sexual abuse	.004	3.075	.028***	9.861***	.020***	8.469***
Neglect	.001	1.570	.000	.395	.001	.979
Emotional maltreatment	.003	2.895	.004	3.109*	.002	2.547
Vineland Daily Living						
Physical abuse	.005	3.960	.010*	6.846*	.012**	7.451**
Sexual abuse	.001	−2.918	.000	1.742	.000	−.466
Neglect	.000	.000	.000	.000	.000	.687
Emotional maltreatment	.003	4.428	.001	−1.470	.004	−3.728
Vineland Socialization						
Physical abuse	.000	−.329	.001	1.817	.002	2.417
Sexual abuse	.001	−2.918	.006	−6.697	.005	−5.916
Neglect	.007*	−4.810**	.001	−1.751	.002	−2.508
Emotional maltreatment	.001	1.776	.000	−.618	.009*	−5.268

Note. There were no significant findings for the TSCC scales, and these have been excluded.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Discussion

Despite the recognition of the need for clearer definitions of maltreatment among researchers in response to the call by the National Research Council (1993), consensus on the definitions of maltreatment continues to challenge researchers. Maltreatment definitions and classification schemes vary over time, across cultures and geographic boundaries, and by professional discipline (Cicchetti & Manly, 2001). Even within the same discipline there is debate about such issues as whether maltreatment should be defined based on the behavior of the perpetrator, the experience of the child, or by some combination of these (Barnett et al., 1993). A unified approach to defining maltreatment also appears unlikely because of differences across disciplines in the way that definitions are utilized (Barnett et al., 1993; Cicchetti & Manly, 2001; Zuravin, 1991). For example, maltreatment definitions based on research criteria for

the purpose of understanding child outcomes may not be useful to CPS in carrying out their work of determining appropriate responses to individual reports.

Though the lack of uniform research definitions of maltreatment remains an issue, we found moderate levels of agreement between the original child protective services classification of reports and two different research classification systems upon reviewing CPS records. Overall, the concordance between the original child protective service classification and the research classification using the MMCS is in the 80–90% range for physical abuse, sexual abuse and neglect. We found much greater disagreement among emotional maltreatment reports with potentially significant misclassification. The ROC curve analysis and the Kappa statistics confirm close agreement between the two alternative research definition coding systems. For every type of maltreatment, the sensitivity and specificity of the NIS-2 codes were very close to the MMCS codes. Although the high level of agreement may have been influenced by the concurrent coding of the record for both systems by the same reviewer, there was much less agreement for neglect and emotional abuse.

Although there were relatively moderate differences in the classification of maltreatment types between coding systems, we found evidence that a coding system with replicable coding criteria beyond that used by CPS is important in order to describe accurately children's maltreatment experiences. The research definitions improved prediction in a sample already at high risk because of early life experiences. For maltreatment reports occurring after age 4 in a sample of children, many of whom had been reported before age 4, we found that sexual abuse and physical abuse, using research definitions, were significant but weak predictors of adverse effects. In contrast, the original CPS classifications demonstrated little or no impact on outcomes.

Interestingly, the CPS classification of neglect was more powerful in predicting socialization, as measured with the Vineland Screener, than either the MMCS or the NIS-2. Although the models were not significant, the CPS classification of neglect was also more powerful than either the NIS-2 or the MMCS neglect code in predicting Externalizing or Total Problems scores. This result may represent the misclassification of other types of maltreatment in CPS reports as neglect. Leiter and Johnsen (1997), for example, found that among their sample of 657 substantiated reports of neglect found in the North Carolina central registry, 133 of those reports were for alleged abuse. In contrast, they found only six reports of neglect that were classified as abuse upon substantiation. One possible reason for this is that since neglect is more difficult to prove than abuse, and hence criminal prosecution is less likely, caregivers may be given the option of stipulating to the less serious charge of neglect and agreeing to intervention by social service agencies when the allegation was for physical or sexual abuse.

These exploratory analyses comparing the predictive ability of the original CPS classifications and two research definition systems for subsequent child function represent a novel addition to the current discussion about the need to improve research definitions. Taking advantage of child functional assessments at age 4 and 8 in a cohort of high-risk children, we were able to control for child functioning before the exposure to new maltreatment. Of note, these longitudinal analyses are potentially limited because exposure to maltreatment prior to age 4 may have a latent effect not already accounted for in the baseline assessment at age 4. Further, some forms of maltreatment, such as neglect or emotional abuse, may have their most damaging effects at younger or older ages, and thus our attempt to account for adverse influences after age 4 may be too late or too early. Nevertheless, we found significant adverse effects of physical and sexual abuse that were unrelated to the particular sample (study site), ethnicity, or socio-economic status.

Our analyses used both a limited age of exposure and a specific age at which to examine outcomes in a group that had all been reported to CPS previously. Additional studies examining exposure at earlier ages and examining outcomes at other ages are needed before conclusions can be drawn about the real importance of improving definitions in order to advance the understanding of child abuse and neglect. These analyses reflect the evidence of maltreatment as documented in CPS records. It remains to be seen how well these records describe the lives of LONGSCAN children. Any single data source is limited in its ability to fully capture children's maltreatment experiences. As they age, LONGSCAN subjects will be asked directly about their maltreatment experiences, as children's subjective experience of maltreatment likely influences its effects on functioning. However, even child report may be limited as a source of data about child experiences. Their memories of early childhood maltreatment may not be accurate. In addition, children may have different explanatory models that may influence their perception of what constitutes maltreatment or their attributions of the causes of maltreatment. Systematic efforts to inquire about maltreatment from both the child's and the parent's perspective may be needed to understand the nature of maltreatment that has occurred.

Comparing the codes assigned by CPS with the research codes defined by very specific criteria is an important contribution to understanding the differences in the labels assigned to children's experiences by CPS workers and by researchers. This study has shown the potential usefulness of standardized research coding schemes, with strict, detailed criteria for assigning codes, for more accurately describing maltreatment reports and for generating statistics that are comparable across geographic boundaries and over time. Considering the recent discussions about the apparent decline in physical and sexual abuse nationally since 1993, we caution that the misclassification of maltreatment type by social services may be systematic and the patterns of misclassification may be changing over time. CPS classifications may be influenced by other considerations such as whether a conclusion of abuse creates different obligations on an agency than a finding of neglect. For example, a pattern of reluctance to report families to police by labeling cases neglect may be one plausible explanation for the apparent decline in child physical and sexual abuse nationally. Additional work is needed to improve classifications of type of maltreatment and to examine patterns of classification over time in the same jurisdiction to understand how structural or policy factors influence the classification of maltreatment reports.

References

Achenbach, T. M. (1991). *Manual for Child Behavior Checklist/4–18 and 1991 Profile*. Burlington, VT: University of Vermont, Department of Psychiatry.

Barnett, D., Manly, J. T., & Cicchetti, D. (1993). Defining child maltreatment: The interface between policy and research. In D. Cicchetti & S. L. Toth (Eds.), *Advances in applied developmental psychology: Child abuse, child development and social policy* (pp. 7–73). Norwood, NJ: Ablex Publishing Corp.

Bolger, K. E., & Patterson, C. J. (2001). Developmental pathways from child maltreatment to peer rejection. *Child Development, 72*(2), 549–568.

Briere, J. (1996). *Trauma Symptom Checklist for children: Professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

Cicchetti, D., & Manly, J. T. (2001). Editorial: Operationalizing child maltreatment: Developmental processes and outcomes. *Development and Psychopathology, 13*, 755–757.

Elmer, E. (1977). A follow-up study of traumatized children. *Pediatrics, 60*, 185–187.

English, D. J., Bangdiwala, S. I., & Runyan, D. K. (2005). The dimensions of maltreatment: Introduction. *Child Abuse & Neglect, 29*(5), 441–460.

Fletcher, R., Fletcher, S., & Wagner, E. (1996). *Clinical epidemiology, the essentials* (3rd ed.). Baltimore: Williams & Wilkins.
Haugaard, J. (2000). The challenge of defining child sexual abuse. *The American Psychologist*, *55*, 1036–1039.
Lau, A. S., Leeb, R. T., English, D., Graham, J. C., Briggs, E. C., Brody, K. E., & Marshall, J. M. (2005). What's in a name? A comparison of methods for classifying predominant type of maltreatment. *Child Abuse & Neglect*, *29*(5), 533–551.
Leiter, J., & Johnsen, M. C. (1997). Child maltreatment and school performance declines: An event-history analysis. *American Educational Research Journal*, *34*(3), 563–589.
Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemueller, W., & Silver, H. R. (1962). The battered child syndrome. *Journal of the American Medical Association*, *181*, 17–24.
Kinard, E. M. (1994). Methodological issues and practical problems in conducting research on maltreated children. *Child Abuse & Neglect*, *18*, 645–656.
Kinard, E. M. (2001). Characteristics of maltreatment experience and academic functioning among maltreated children. *Violence and Victims*, *16*(3), 323–337.
Manly, J., Cicchetti, D., & Barnett, D. (1994). The impact of subtype, frequency, chronicity, and severity of child's maltreatment on social competence and behavior problems. *Development and Psychopathology*, *6*, 121–143.
National Center on Child Abuse and Neglect. (1988). *Study findings: Study of national incidence and prevalence of child abuse and neglect: 1988*. Washington, DC: US Department of Health and Human Services, Office of Human Development Services.
National Research Council. (1993). *Understanding child abuse and neglect*. Washington, DC: National Academy Press.
Newborg, J., Stock, J. R., Wnek, L., Guidubaldi, J., Svincki, J., Dickson, J., & Markley, A. (1988). *Battelle Developmental Inventory with recalibrated technical data and norms: Screening test examiner's manual* (2nd ed.). Allen, TX: DLM, Inc.
Runyan, D. K., Curtis, P., Hunter, W. M., Black, M. M., Kotch, J. B., Bangdiwala, S., Dubowitz, H., English, D., Everson, M., & Landsverk, J. (1998). LONGSCAN: A Consortium for Longitudinal Studies of Maltreatment and the Life Course of Children. *Aggression and Violent Behavior*, *3*(3), 275–285.
Sedlak, A. J. (1986). *Second National Incidence Study of child abuse and neglect (NIS-2)*. Washington, DC: US Department of Health and Human Services.
Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales: Interview edition, survey form manual*. Circle Pines, MN: American Guidance Service.
Sparrow, S. S., Carter, A. S., & Cicchetti, D. V. (1993). *Vineland screener: overview, reliability, validity, administration, and scoring*. New Haven, CT: Yale University Child Study Center.
Zuravin, S. J. (1991). Research definitions of child physical abuse and neglect: Current problems. In R. Starr & D. Wolfe (Eds.), *The effects of child abuse and neglect: Issues and research* (pp. 100–128). New York: Guilford Press.

Résumé

Objectif: Le Conseil National des Recherches a jugé que les recherches définissent mal les mauvais traitements et la négligence de sorte que ces définitions posent des difficultés pour la recherche. Les auteurs ont examiné la concordance entre, d'une part, la façon dont les services de protection de l'enfance classent les diverses formes de maltraitance et, d'autre part, deux méthodes de classement utilisées dans des recherches. Ils soulignent le contraste entre les deux systèmes de classements de la recherche et le système utilisé par les services de protection de l'enfance, à savoir leur habileté de prédire des problèmes psychologiques d'enfants maltraités âgés de 8 ans.
Méthode: L'échantillon comprenait 545 enfants enrôlés dans une étude dite LONGSCAN (études longitudinales sur les mauvais traitements et la négligence) suite à un incident de maltraitance, et pour qui on a recueilli des données à l'âge de 4 et 8 ans. Ces enfants ont aussi fait l'objet d'un suivi dans les dossiers de la protection de l'enfance de la naissance jusqu'à l'âge de 8 ans. Les signalements aux autorités de la protection infantile ont été codifiés en se servant des modifications que LONGSCAN a apportées au Système de Classification des Mauvais Traitements (dit MMCS en anglais) ainsi que le système de codification du

Second National Incidence Study (NIS-2)—soit la deuxième étude sur l'incidence nationale des mauvais traitements. Les premières analyses ont utilisé les signalements en tant qu'unité d'analyse pour examiner la concordance dans les allégations entre les services de protection de l'enfance et la recherche. Pour valider les constats, les chercheurs ont analysé les retombées pour chaque type de mauvais traitements vécus après l'âge de 4 ans. Pour ce faire, ils ont utilisé le Child Behavior Checklist (CBCL), le Trauma Symptom Checklist (version alternative) et le Vineland Screener. Pour contrôler les variables, ils ont administré le CBCL et le Battelle Developmental Screener à l'âge de 4 ans.

Résultats: Au total, on compte 1.960 signalements de mauvais traitements affectant les 545 enfants qui ont fait l'objet de l'étude, bien que les 1.593 dossiers de la protection de l'enfance n'aient précisé qu'un type de maltraitance. On a noté des différences entre les types de mauvais traitements relevés dans les dossiers de la protection de l'enfance et les conclusions tirées des deux systèmes de codification de la recherche. C'est entre la classification des services de protection de l'enfance et les systèmes de codification de la recherche qu'on a remarqué le plus haut niveau d'incohérence, et ceci par rapport à la négligence et les mauvais traitements affectifs. Près de 10% des signalements d'agressions physiques et sexuelles que le MMCS a précisés, ont été classés en tant que néligence par les services de protection de l'enfance. Le système NIS = 2 et le MMCS démontrent une statistique Kappa très élevée en ce qui a trait à la cohérence pour les agressions physiques et sexuelles. On note les modèles conçus pour prédire le fonctionnement des enfants à l'âge de 8 ans se servent de définitions valides pour les agressions physiques et sexuelles. La capacité de prédire le fonctionnement des enfants s'améliore de façon modérée dans quelques domaines, comparée aux classements des autorités de la protection de l'enfance.

Conclusions: Les deux systèmes de classement utilisés dans la recherche améliorent de façon modérée la capacité de prédire les conséquences néfastes des mauvais traitements, comparés à la façon dont les autorités de la protection de l'enfance conçoivent les mauvais traitements vécus.

Resumen

Objetivo: El Consejo de Investigación Nacional identificó que algunas definiciones que la investigación da el maltrato y la negligencia infantil son inadecuadas y dificultan el desarrollo de la investigación. Se examinó la concordancia entre las diferentes clasificaciones de maltrato de los Servicios de Protección Infantil y las definiciones de dos sistemas de codificación de la investigación. Se contrastaron los dos sistemas de codificación con la clasificación de los Servicios de Protección Infantil (CPS) en su habilidad para predecir las dificultades posteriores en el funcionamiento psicológico de los niños maltratados a la edad de ocho años.

Método: La muestra se compuso de 545 niños que participaron en el estudio del Longitudinal Studies of Child Abuse and Neglect (LONGSCAN) que contaban con una notificación de maltrato infantil. Los datos fueron recogidos a una edad aproximada de entre 4 y 8 años, y la revisión a lo largo del tiempo de los informes de los CPS fue a la edad de 8 años. Los informes de maltrato del CPS fueron codificados con una modificación del sistema del LONGSCAN utilizado para clasificar el maltrato (MMCS) y el sistema de codificación del Segundo Estudio Nacional de Incidencia (NIS-2). Los primeros análisis examinaron el acuerdo entre el CPS y las determinaciones de la investigación según el tipo de maltrato. Los análisis de validación examinaron las consecuencias de cada tipo de maltrato experimentado a partir de la edad de 4 años según cada sistema de codificación. Se utilizaron el Child Behavior Checklist (CBCL), el formulario del Trauma Symptom Checklist-Alternative, y el Vineland Screener como instrumentos para

medir las consecuencias del maltrato. Las variables de control fueron el CBCL y el Battelle Developmental Screener, medido a la edad de 4 años.

Resultados: Hubo un total de 1,980 notificaciones de maltrato en los 545 niños estudiados aunque solamente 15,493 informes de los Servicios de Protección Infantil (CPS) identificaron al menos una forma de maltrato. Hubo diferencias entre el tipo de maltrato recogido en los expedientes de los CPS, y las conclusiones a las que se llegaba con cada método de clasificación. Las clasificaciones de los CPS fueron las más discordantes con los sistemas de investigación del maltrato y negligencia emocional. Cerca de un 10% de los informes de maltrato físico y abuso sexual, según determinaba la MMCS, fueron clasificados como negligencia en los CPS. En los casos de maltrato físico y abuso sexual, se obtuvieron coeficientes altos (índices Kappa) de acuerdo entre los sistemas NIS-2 y MMCS. Se demostró la validez de las definiciones de investigación del maltrato físico y el abuso sexual en los modelos que predecían el funcionamiento del niño a la edad de 8 años. La predicción del funcionamiento infantil fue significativa pero modestamente mejorada en diferentes ámbitos de las clasificaciones de los CPS.

Conclusiones: Ambas clasificaciones de la investigación mejoraron moderadamente la predicción de efectos adversos del maltrato en comparación con la caracterización de la exposición al maltrato según fue recogida por los CPS.