Psychology, Public Policy, and Law
2013, Vol. 19, No. 2, 179–190

1076-8971/13/$12.00 DOI: 10.1037/a0030312

Do Case Outcomes Change When Investigative Interviewing Practices Change?

Margaret-Ellen Pipe and Yael Orbach
Brooklyn College, CUNY

Michael E. Lamb
University of Cambridge

Craig B. Abbott
Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland

Heather Stewart
Children's Justice Center, Salt Lake City, Utah

To determine whether the introduction of evidence-based practice for interviewing child witnesses was accompanied by changes in the disposition of cases in which sexual abuse was suspected, we compared cases in which the investigative interviews of 3- to 14-year-old alleged victims were conducted either before ($n = 350$) or after ($n = 410$) investigators had been taught to use the NICHD Protocol. Analyses showed that when charges were filed, both pre-Protocol and Protocol interviews were highly (and similarly) likely to yield guilty pleas to one or more counts. However, charges were more likely to be filed when interviewers had been trained to use the Protocol; Protocol interviews were therefore associated with more guilty pleas than non-Protocol interviews. When cases were tried, furthermore, Protocol cases were more likely to yield guilty verdicts. These results point to important implications for policies concerning the investigation of suspected crimes involving young witnesses.

Keywords: structured interview protocols, child abuse, case outcomes, child witnesses

When child sexual abuse is suspected, alleged victims are typically the primary, and often the only, sources of evidence. As a result, decisions regarding both child protection and criminal proceedings depend heavily on the quality of the information obtained from suspected victims during investigative interviews. Recognizing the importance of children's evidence, structured interview protocols for conducting forensic interviews with children are now widely advocated (e.g., Home Office, 2011; Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007; Perona, Bottoms, & Sorenson, 2006). One of these, the NICHD Investigative Interview Protocol, has been subjected to systematic evaluation in the field and has been shown to change both interviewer practices and the "quality" of information elicited from suspected victims (Lamb, Hershkowitz, Orbach, & Esplin, 2008). In the study reported here, we asked whether the introduction of this evidence-based approach to investigative interviewing was accompanied by "downstream effects" on case decisions and outcomes. Previous studies have examined the temporal and procedural aspects of the "case flow" through the criminal justice system (e.g., Cross, Walsh, Simone, & Jones, 2003; Finkelhor, Cross, & Cantor, 2005; Walsh, Lippert, Cross, Maurice, & Davison, 2008), but ours was the first study to focus on the investigative interview as a predictor of such outcomes as the filing of criminal charges, prosecution, and guilty pleas or convictions.

Overview of Case Flow in Child Abuse Cases

Although allegations of child abuse are variously dealt with in different jurisdictions, cases generally follow a predictable path through the child protection and criminal justice systems (Cross et al., 2003; Cross, Whitcomb, & DeVos, 1995; Finkelhor et al., 2005; Giovannoni, 1991). In the jurisdiction where the present study was conducted, all allegations of abuse trigger Child Protective Service (CPS) investigations to determine whether or not the case is "substantiated," with the balance of evidence suggesting that the alleged abuse probably occurred. Informal interviews may be conducted to assess children's safety during these investigations.

When these investigations have not been conducted jointly by CPS workers and police officers, substantiated cases are routinely referred to law enforcement personnel for criminal investigation. Specialist investigative interviewers (mostly police detectives) interview the suspected victims, suspects and witnesses (if available), gather evidence, and decide whether to refer the case for "screening" with District Attorneys (DAs, prosecutors). During the screening process, DAs evaluate the evidence and other information provided by a Multi-Disciplinary Team (MDT) and determine whether or not there is a reasonable likelihood of successful

This article was published Online First November 12, 2012.

Margaret-Ellen Pipe and Yael Orbach, Department of Psychology, Brooklyn College, CUNY; Michael E. Lamb, Department of Psychology, University of Cambridge; Craig B. Abbott, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland; Heather Stewart, Children's Justice Center, Salt Lake City, Utah.

Gail S. Goodman acted as Action Editor for the review of this manuscript.

Correspondence concerning this article should be addressed to Michael E. Lamb, Department of Psychology, University of Cambridge, Free School Lane, Cambridge CB2 3RQ, United Kingdom. E-mail: Mel37@cam.ac.uk

prosecution, in which case they recommend that criminal charges be filed against the suspects. DAs may "decline" cases because, for example, there is insufficient corroborative evidence to support the children's statements, the victims are considered too young to be credible witnesses, the suspects are unknown, or the victims' nonoffending parents refuse to cooperate with the police. After prosecution is initiated, cases may be dismissed by the judiciary or dropped by prosecutors; charges may also be filed later if new evidence emerges, or the victims' status or ability to testify changes.

Once charged, alleged perpetrators must decide whether to plead guilty or go to trial. In plea bargains, which must be approved by judges, the defendants' attorneys negotiate with prosecutors to reduce the severity, and/or to dismiss some of the charges in exchange for guilty pleas to other charges.

Research Examining Predictors of Case Outcomes

Pooling data compiled over a 1-year period (1988–1989) in four urban jurisdictions, Cross and his colleagues (1995) tracked the progress from referral for possible prosecution to disposition of all sex-abuse cases involving 4- to 18-year-old alleged victims. Of cases involving 552 suspects, 60% were accepted for prosecution, 38% were declined, and 2% were assigned to a diversion program. Of those accepted, 5% were dismissed by prosecutors, grand juries or the court, and a few were transferred to other courts or jurisdictions, leaving 54% of the total sample disposed of either by trial (9%) or plea agreement (46%).

Cross and his colleagues (1995, 2003) were among the first to recommend that researchers should take into account all reported cases (including the large proportion of cases that are referred to CPS or law enforcement and are not accepted for screening because they are not substantiated or are referred to other jurisdictions), rather than focus only on cases accepted for prosecution. Cross et al. (2003) thus conducted a meta-analysis involving 24 samples drawn from 21 studies. Most (19, 79%) of the samples only included sexual abuse cases, with the remaining samples (5, 21%) including allegations of both sexual and physical abuse. Only four of the 24 samples included information about the initial referral to law enforcement and/or child protection, and in these four samples, between 40% and 85% of the substantiated cases were referred to the DAs for prosecution. In 14 samples which provided information about the filing of criminal charges, filing rates ranged from 28% to 94%. Data from 18 of the samples showed that 79% of the charged cases were prosecuted. When cases that were dismissed, diverted, or transferred were excluded from the analyses, data from 19 samples showed that 94% of the prosecuted cases led to conviction, 82% on the basis of plea agreements and only 18% at trial (Cross et al., 2003).

Combining all the available data, Cross et al. (2003) estimated that, for every 100 referrals to CPS or police, 66 would be charged, with two suspects diverted to alternative treatment programs, charges against 12 dismissed or transferred, while 52 would be accepted for prosecution. Of the 52 cases accepted for prosecution, 49 suspects would be convicted (43 through plea agreement and 6 through trial), and three would be acquitted at trial. The great majority of cases accepted for prosecution result in conviction, predominantly by plea, and thus the decision to prosecute is the single best predictor of case disposition (Cross et al., 1995, 2003).

It is interesting that there were no significant differences in the proportions of child sexual abuse cases and nonabuse felony cases prosecuted (Cross et al., 1995); in both, the majority of prosecuted cases yielded guilty pleas, with similar conviction rates. More prosecuted cases of sexual abuse (9%) than nonabuse felony cases (3%) went to trial, however.

Other researchers have examined case information and court records in attempts to identify factors associated with the prosecution of child abuse. Child sexual abuse cases are more likely to be prosecuted when the suspects are strangers than when they are parents or familiar individuals (Brewer, Rowe, & Brewer, 1997; Chapman & Smith, 1987; Cross, DeVos, & Whitcomb, 1994; Ménard & Ruback, 2003; Stroud, Martens, & Barker, 2000; Tjaden & Thoennes, 1992); when the cases involve sexual abuse, especially those involving penetration, rather than other types of abuse or less severe abuse (Brewer et al., 1997; Chapman & Smith, 1987; Cross et al., 1994; Sedlak et al., 2005; Tjaden & Thoennes, 1992); when multiple rather than single incidents of abuse are alleged (Chapman & Smith, 1987; Cross et al., 1994; MacMurray, 1989; Tjaden & Thoennes, 1992); when alleged victims are older (Chapman & Smith, 1987; Cross et al., 1994; Stroud et al., 2000; Tjaden & Thoennes, 1992); when the suspects come from minority groups (Stroud et al., 2000; Tjaden & Thoennes, 1992); and when the victims are female (MacMurray, 1989; Sedlak et al., 2005; Stroud et al., 2000; Tjaden & Thoennes, 1992). Prosecution and conviction are more likely in cases involving male suspects, older suspects, and medical evidence (Bradshaw & Marks, 1990; Brewer et al., 1997; Cross et al., 1994; Ménard & Ruback, 2003; Sedlak et al., 2005), suspect statements (Sedlak et al., 2005), shorter intervals between occurrence and reporting (Bradshaw & Marks, 1990; Brewer et al., 1997; Ménard & Ruback, 2003), and the involvement of multiple victims (Brewer et al., 1997; Ménard & Ruback, 2003).

Joint law-enforcement and CPS investigations are pursued in many jurisdictions, often in the context of Children's Advocacy Centers (CACs) or Child Abuse Assessment Centers (CAACs). Cases assessed in CACs (or CAACs) are more likely to be prosecuted (Jensen, Jacobson, Unrau, & Robinson, 1996; Steele, Norris, & Komula, 1994; Tjaden & Anhalt, 1994; Walsh et al., 2008) although the reasons are not clear. Joa and Edelson (2004) found that 76% of the cases in which children were seen at CAACs led to the filing of charges, compared with 39% of the cases in which the children were not seen at CAACs. In addition, more counts were charged overall, more counts were charged against biological fathers and stepfathers, and more defendants pleaded or were found guilty following CAAC involvement. Prosecutions were also more common when cases involving children aged 4 to 6 years or over 12 years were assessed at CAACs rather than elsewhere (Joa & Edelson, 2004).

Jones, Cross, Walsh, and Simone (2005) concluded that adherence to seven widely recommended investigative practices (the use of MDTs, videotaping of all forensic interviews, specialized training for forensic interviewers and forensic medical examiners, victim advocacy programs, improved access to victims' mental health treatment, and the establishment of CACs) was associated with a higher likelihood of both CPS substantiation and the filing of charges against suspects. As Jones and colleagues (2005) ob-

served, however, little research has been conducted on the effectiveness of specific programs. In particular, research has focused on investigative procedures, rather than the effect of investigative interviewing practices, as in the present study.

Whether or not they are valid, most allegations of sexual abuse cannot be pursued in the criminal justice system because the allegations themselves are unclear and inadequately detailed, and because corroborative evidence is often lacking (Goodman, 2006; Poole & Lamb, 1998). This makes research on the quality of investigative interviewing especially important. Accordingly, we asked in the present study whether the introduction of an evidence-based approach to interviewing child witnesses, the NICHD Investigative Interview Protocol, was accompanied by changes in case outcomes, in light of prior research showing that introduction of the NICHD Protocol significantly improves interviewing practices and the quality of the information elicited from alleged victims (see Lamb et al., 2008, for review).

The NICHD Interview Protocol

Structured protocols for forensic interviews with children are widely advocated in the USA, Israel, Canada, the UK and Europe (e.g., Cyr & Lamb, 2009; Hershkowitz, Horowitz, & Lamb, 2005; Home Office, 2011; Lamb et al., 2008; Perona et al., 2006; Poole & Lamb, 1998), but to date only the NICHD Investigative Interview Protocol has been systematically evaluated in the field (Cyr & Lamb, 2009; Lamb et al., 2009; Lamb, Orbach, Hershkowitz, Horowitz, & Abbott, 2007; Lamb, Orbach, Hershkowitz, et al., 2007; Lamb et al., 2008; Orbach et al., 2000; Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001). This Protocol has been adopted as the standard protocol for forensic interviews of child witnesses in a number of centers in the US and Canada, is explicitly recommended to forensic investigators in the UK (Home Office, 2011), and has been used in all interviews of suspected child abuse victims, witnesses, and suspects in Israel since 1996 (Hershkowitz et al., 2005).

The NICHD Protocol trains interviewers to first provide children with opportunities to practice recalling experienced events and to maximize the amount of information elicited from free recall memory. The strategies and techniques employed in the Protocol were developed in accordance with widespread evidence that free recall memory prompts are most likely to elicit accurate information, whereas prompts that depend on recognition processes are associated with more erroneous responses (Perona et al., 2006; Poole & Lindsay, 1998; Wood & Garven, 2000; see Lamb et al., 2008, for review).

Extensive research has demonstrated that the NICHD Protocol improves the quality of information obtained from alleged victims, including very young children (see Lamb et al., 2008, for a review). In Protocol interviews, absolutely and proportionally more information is elicited using open-ended prompts, whereas significantly fewer and proportionally fewer details are elicited using directive, option-posing, and suggestive utterances than in non-Protocol interviews.

These findings are important in view of experimental and field research demonstrating that the way memories are accessed is a crucial determinant of their accuracy and that information elicited using recall processes is more accurate than information elicited using recognition processes (see above).

The Present Study

We expected that improvements in the quality of investigative interviews might affect the progress of cases through the criminal justice system and thus asked whether the introduction of a best-practice interview, the NICHD Protocol, was associated with changed case outcomes within the justice system. Because use of the NICHD Protocol should reduce ambiguities in children's accounts and increase the amount of central information they report, especially in free recall, we predicted that introduction of the Protocol would be associated with a higher likelihood that investigations would lead to arrests and the filing of criminal charges. Conversely, we expected that improved interviewing practices might decrease the number of cases declined by prosecutors during the screening process. When charges were filed, we expected that more would result in conviction, either through plea negotiations or at trial.

Method

To determine whether the progress of child sexual abuse cases through the criminal justice system changed following the introduction of the NICHD Protocol for conducting interviews, we compared case outcomes and dispositions before (1994 to mid-September 1997) and after (mid-September 1997 to 2000) the introduction of the NICHD Protocol. Only cases involving suspicions of sexual abuse were included in the study. The same detectives, prosecutors and judges handled the cases from 1994 to 2000 with no changes in leadership or formal policy during the study period (other than introduction of the Protocol following engagement with the NICHD researchers), thereby minimizing the potential confounding effects of these variables. A particular advantage of this before-and-after design was that the same interviewers conducted interviews in the two periods. Nonetheless, because interviews in the two conditions were conducted at different time periods (before and after training of interviewers on the NICHD Protocol), we cannot discount the possibility of potential confounds, including increased interviewing experience. In this retrospective study, it was not feasible to employ an experimental design with random assignment of cases to the two interview conditions which, as noted in the discussion, would have introduced a different confound (different interviewers in different conditions because it is not possible for the same interviewers to be both trained and untrained). However, the proportions of cases in which suspects were charged were, for successive pre-Protocol years, 40% (1994/95), 36.4% (1996), and 42.2% (first part of 1997). Following introduction of the Protocol, the percentages of cases in which charges were filed were 49.1% (second part of 1997), 42.5%, (1998) and 47.6% (for 1999/2000). There was some variation across years, both prior to and following introduction of the Protocol, therefore, but no clear trend prior to introduction of the Protocol, suggesting that greater experience alone could not account for differences between the two time periods studied.

All the forensic interviews included in the present study were conducted at the Salt Lake County Children's Justice Centre (CJC) by 16 police detectives who specialized in conducting forensic interviews with children. In September 1997, they were trained to conduct forensic interviews with children using the NICHD Protocol in a 5-day workshop that included information about child

development, especially in the preschool years, as well as opportunities to be monitored and given feedback during practice interviews, and was complemented by feedback on field interviews (see Sternberg, Lamb, Orbach, et al., 2001, for more information). Prior to the introduction of the Protocol, there was no formal or systematic training in forensic interview practices; individual investigators had attended one or more of the following training sessions: 1-hr break-out sessions at conferences, 2-day training workshops focusing on the investigative process (not exclusively on victim interviewing), 2-day training based on the Kempe Center Model of interviewing (The Kempe Center for the Prevention & Treatment of Child Abuse & Neglect, University of Colorado, Denver School of Medicine, Department of Pediatrics, Aurora, CO), the Attorney General's Task Force on Satanic Abuse training conference, and local conferences. In many cases, the primary training occurred "on the job," with guidance by other interviewers, a practice widespread at the time.

The Sample, Case Characteristics, and Outcomes

The sample comprised investigative interviews with 871 suspected victims of sexual abuse between the ages of 2.8 and 13.97 years. All suspected victims of sexual abuse referred to Salt Lake County Police Departments were considered for inclusion if the interviews had been conducted by one of the 16 experienced police officers from the two police departments who were trained in 1997 to use the Protocol. Initially, 1,029 interviews were identified for inclusion. Of these, 133 interviews were dropped because the alleged perpetrator was implicated in at least one other case; 25 additional interviews were dropped for lack of information about police outcome. Of the remaining 871 interviews, 388 cases were initiated in the pre-Protocol period and 483 in the Protocol period. The number of interviews conducted by each police officer ranged from 3 to 134 ($M = 54.4$, $SD = 40.5$). Most interviews were the first formal forensic interviews of the alleged victims, although second forensic interviews were sometimes performed.

Based on all available sources of information, including CPS and police reports, CJC intake forms, and the CJC database, two trained researchers compiled the archive used in the study. Most of the information was objectively recorded and required no judgment or rating. Queries and ambiguities were discussed with an additional trained researcher.

Case Characteristics included the interview condition (pre-Protocol, Protocol), the number of alleged incidents of abuse (single, multiple), abuse severity (exposure, touch, penetration), location of incident/s (in home, out of home, unknown/unclear), victim's age at interview (years), victim's gender (male, female), victim's ethnicity (Caucasian, Hispanic, other), victim's disability (mental, physical, none), suspect's familiarity with the victim (immediate family coresiding, other family, familiar—not related, unfamiliar), suspect's age (juvenile, adult), suspect's ethnicity (Caucasian, Hispanic, other), suspect's gender (male, female), number of suspects (single, multiple), number of victims (single, multiple), availability of corroborative evidence (medical, material, witnesses), referral date (date the case was referred to the CJC by CPS or law enforcement personnel), and interview date (date child was interviewed by CJC personnel).

Case outcomes represented all decision-making points between the initial referral for investigation and court disposition:

- *Substantiated*—determinations by CPS during initial case evaluation that abuse had probably occurred.
- *Transferred*—decisions made by police officers to refer cases to other police departments because the possible crimes were believed to have occurred outside their jurisdictions.
- *Exceptionally cleared*—cases dismissed by police investigators and not submitted for screening with the DAs for reasons beyond the investigators' control (e.g., known suspects could not be located or had died).
- *Unfounded*—assessments, made by police officers following the forensic interviews, that there were no credible bases for the allegations, implying they may have been untrue. Such cases were dismissed by police investigators.
- *Not cleared*—decisions made by police officers not to submit cases for screening after interviewing the children because there was insufficient corroborative evidence, the victims were considered too young to be credible witnesses, the abusers were unknown, or the victims' parents refused to cooperate with the police, rendering successful prosecution highly unlikely. Such cases were closed by police investigators.
- *Submitted for screening*—decisions by police investigators to submit cases for screening with DAs to determine whether criminal charges should be filed.
- *Screened/declined*—decisions by DAs not to file criminal charges because prospects of successful prosecution appeared low after the interviews, other evidence, and information from the MDTs had been considered.
- *Arrested/charged (filed)*—decisions by DAs to file criminal charges because the likelihood of successful prosecution appeared high.
- *Arrested/Charged (not filed)*—cases initially approved by DAs (arrested/charged), in which charges are eventually not filed for criminal proceedings because additional information was later obtained.
- *Diverted*—cases diverted by DAs after prosecution had been initiated (i.e., charges had been filed) to an alternative intervention, such as a treatment program.
- *Plea Agreements*—agreements reached between the DAs and suspects regarding the disposition of criminal charges, typically involving reductions in the level/seriousness of the charges filed (e.g., from felony charges to misdemeanor charges) or in the number of counts charged.
- *Submitted for trial*—in the absence of plea agreements, cases were submitted to trial by criminal courts or by juvenile courts (when the suspects were under 18 years at the time of the alleged incidents).
- *Case disposition*—statements specifying one of three types of court decisions—upholding, reducing, or dismissing criminal charges following plea agreements or trials:

  - *Case Dismissed*—Dismissal of all charges against the defendant.
  - *Pled Guilty*—Usually to reduced charges, following negotiations during a plea agreement.
  - *Found Guilty*—on some or all counts following a trial.

Results

The sample of 871 cases was reduced to 760 for the purposes of the analyses: 64 cases (27 pre-Protocol period, 37 Protocol period) were exceptionally cleared (forced inaccessibility to suspects or victims) and 47 (11 pre-Protocol, 36 Protocol) were transferred to other jurisdictions. These cases are not considered further because the outcomes were determined by jurisdictional circumstances, not the evidence. Further, the following variables had to be excluded from all analyses because information was missing for too many cases: location of incident, victim's disability, suspect's ethnicity, number of incidents, and availability of corroborative evidence.

Of the 364 cases in which charges were filed (see Table 1), 86.0% (n = 313) involved felony-only charges, 7.4% (n = 27) involved misdemeanor-only charges, and 6.6% (n = 24) involved both felony and misdemeanor charges. The total number of counts per suspect ranged from one to eight (M = 1.92, SD = 1.15), with 45% (n = 165) of the cases having one count, 32% (n = 116) having two counts, and 23% (n = 83) having three or more counts. A logistic GEE (Generalized Estimating Equation) with interview condition as the dependent variable revealed that differences in the mean number of total, felony, and misdemeanor counts per case in the pre-Protocol and Protocol periods were not significantly different.

Overview of the Analyses

We were primarily interested in identifying the case characteristics that predicted (i) whether criminal charges would be filed and prosecution sought, expecting that interview condition would have a significant effect on the likelihood of charges being filed, with a higher proportion of Protocol than pre-Protocol cases submitted by the DAs for prosecution, and (ii) whether convictions, acquittals, or dismissals of charges were more likely in the Protocol than the pre-Protocol condition.

The dependent measures had characteristics that required special analytic methods to avoid inflated Type I errors and low power: they were dummy-coded binomial variables (e.g., suspect arrested with charges filed = 1, other outcome = 0), they were clustered (an alleged victim was interviewed by only one detective, and each detective interviewed multiple victims) so were potentially correlated, and the number of interviews conducted by each interviewer varied, resulting in unbalanced clusters. Binomial logistic regression (Agresti, 2007; Hartmann, Pelzel, & Abbott, 2011) using the method of Generalized Estimating Equations (GEE; Hanley, Negassa, Edwardes, & Forrester, 2003; Hedeker & Gibbons, 2006; Zeger, Liang & Albert, 1988) was thus chosen to control Type I error and maximize power while controlling for the clustering of cases. GEE was applied to adjust for correlated dependent measures that resulted from differences in outcome due to interviewers; that is, differences in cluster means. For dummy-coded binomial outcomes, the cluster mean is the proportion of cases coded 1. Heterogeneity of the cluster proportions was tested using the Pearson chi-square test with N-1 degrees of freedom (N = number of interviewers). A statistically significant chi-squared test indicated that GEE was required rather than standard logistic methods (Snijders & Bosker, 1999). We conducted both simple (one IV in the equation) and multiple (all IVs in the equation) logistic GEEs to examine the total and unique association (i.e., controlling for the other IVs), respectively, between each IV and the dependent variables. IVs were entered into analyses simultaneously. All IVs were categorical and had two or more levels. One level of each variable served as the comparison group for that variable, as shown in Tables 2 and 3. For both simple and multiple regressions, the effects of each variable were interpreted with respect to the reference category.

First, we examined the association between interview condition and other case characteristics. For these analyses, interview condition (pre-Protocol, Protocol) was treated as the dependent variable and other case characteristics as predictor variables. Second, we examined the association between interview condition and the likelihood that suspects were arrested and charged after controlling for other case characteristics. Third, we examined the characteristics of cases in which charges had been filed with respect to a range of dispositions (e.g., guilty plea, plea reduced, dismissal of all charges, found guilty, and acquitted).

Table 1
Outcomes by Interview Condition

	Pre-Protocol		Protocol		Total	
Case outcome (variable)	*n*	%	*n*	%	*n*	%
Not submitted for screening						
Unfounded	42	12.0	41	10.0	83	10.9
Not cleared	49	14.0	49	12.0	98	12.9
Submitted for screening						
Screened declined	98	28.0	72	17.6	170	22.4
Arrest/charges filed	147	42.0	217	52.9	364	47.9
Arrest/charges not filed	10	2.9	26	6.3	36	4.7
Pending	2	0.6	3	0.7	5	0.7
Inactive	2	0.6	2	0.5	4	0.5
Total	350	100	410	100	760	100

Note. Outcomes in this table were the result of decisions by police officers or district attorneys during the investigative and screening phases.

Association Between Interview Condition and Case-Related Variables

Table 2 shows the number and percentage of cases that were in the Protocol or pre-Protocol periods overall, and for each category (level) of case variables. For example, for Severity of Sexual Abuse, 53.9% of the 760 cases were Protocol interview cases, and, conversely, 46.1% were pre-Protocol cases. Of the 257 cases involving allegations of penetration, 135 (52.5%) were Protocol interview cases, and conversely, 122 (47.5%) were pre-Protocol cases. Also shown are the odds ratios (ORs) quantifying the magnitude of the between-category (for "reference" categories in the Table) or between-condition (for nonreference categories) differences.

The proportions of Protocol cases among the interviewers (i.e., cluster means) were significantly heterogeneous $\chi^2(15) = 190.51$, $p < .001$. Logistic GEEs, therefore, using pre-Protocol cases as the reference group, were then conducted to determine whether there were group differences with respect to case characteristics (see Table 2). They revealed statistically significant differences with respect to victim age, victim gender, and victim ethnicity: Compared to the youngest alleged victims, 5- to 6-year-olds and 10- to 13-year-olds were 2.1 (p = .001) and 1.45 (p = .017) times more

Table 2
Associations With Interview Condition: Results of Simple and Multiple Logistic Generalized Estimating Equations (GEE)

Variable	*N*	Pre-Protocol *n* (row %)	Protocol *n* (row %)	Estimate (SE)	Wald χ^2 (*df* = 1)	Unadjusted[a] OR (95% CI)	*p*	Adjusted[b] OR (95% CI)	*p*
Severity of sexual abuse									
Exposure	39	14 (35.9)	25 (64.1)	.479 (.517)	.856	1.614 (.586, 4.447)	.355	1.879 (.601, 5.880)	.278
Touch	430	187 (43.5)	243 (56.5)	.161 (.223)	.520	1.174 (.759, 1.818)	.471	1.372 (.858, 2.194)	.187
Penetration (Reference)	257	122 (47.5)	135 (52.5)	.101 (.419)	.058	1.107[c] (.486, 2.518)	.809	.371[d] (.105, 1.313)	—
Not reported	34	27 (79.4)	7 (20.6)	−1.451 (.407)	12.695	.234 (.105 .521)	.001	.187 (.100, .350)	.001
Victim's age (years)									
2.8 to 4 (Reference)	135	75 (55.6)	60 (44.4)	−.223 (.379)	.346	.800[c] (.380, 1.683)	.556	.371[d] (.105, 1.313)	—
5 to 6	164	62 (37.8)	102 (62.2)	.721 (.210)	11.751	2.056 (1.362, 3.106)	.001	2.199 (1.465, 3.302)	.001
7 to 9	215	99 (46.0)	116 (54.0)	.382 (.226)	2.863	1.465 (.941, 2.279)	.091	1.447 (.926, 2.262	.105
10 to 13	246	114 (46.3)	132 (53.7)	.370 (.155)	5.705	1.447 (1.069, 1.960)	.017	1.583 (1.057, 2.368)	.026
Victim's gender									
Male	214	89 (41.6)	125 (58.4)	.252 (.138)	3.319	1.286 (.981, 1.686)	.068	1.362 (1.043, 1.778)	.023
Female (Reference)	546	261 (47.8)	285 (52.2)	.088 (.320)	.076	1.092[c] (.584, 2.043)	.783	.371[d] (.105, 1.313)	—
Victim's ethnicity									
Caucasian	609	273 (44.8)	336 (55.2)	.758 (.346)	4.788	2.133 (1.082, 4.205)	.029	1.789 (.925, 3.457)	.084
Hispanic	80	36 (45.0)	44 (55.0)	.751 (.297)	6.378	2.119 (1.183, 3.794)	.012	2.234 (1.122, 4.450)	.022
Other race (Reference)	41	26 (63.4)	15 (36.6)	−.550 (.458)	1.445	.577[c] (.235, 1.415)	.229	.371[d] (.105, 1.313)	—
Not reported	30	15 (50.0)	15 (50.0)	.550 (.708)	.604	1.733 (.433, 6.937)	.437	1.406 (.460, 4.295)	.549
Suspect's gender									
Male	616	261 (42.4)	355 (57.6)	−.060 (.261)	.053	.942 (.564, 1.571)	.818	1.010 (.597, 1.709)	.970
Female (Reference)	44	18 (40.9)	26 (59.1)	.368 (.312)	1.388	1.444[c] (.783, 2.663)	.239	.371[d] (.105, 1.313)	—
Not reported	100	71 (71.0)	29 (29.0)	−1.263 (.438)	8.324	.283 (.120, .667)	.004	.221 (.079, .622)	.004
Suspect's age									
Juvenile	280	113 (40.4)	167 (59.6)	.269 (.146)	3.410	1.308 (.984, 1.740)	.065	1.289 (.875, 1.900)	.199
Adult (Reference)	411	193 (47.0)	218 (53.0)	.122 (.338)	.130	1.130[c] (.583, 2.190)	.718	.371[d] (.105, 1.313)	—
Not Reported	69	44 (63.8)	25 (36.2)	−.687 (.303)	5.154	.503 (.278, .910)	.023	.306 (.147, .638)	.002
Victim/suspect familiarity									
Immediate family (Reference)	279	127 (45.5)	152 (54.5)	.180 (.281)	.408	1.197[c] (.690, 2.077)	.523	.371[d] (.105, 1.313)	—
Other family	137	69 (50.4)	68 (49.6)	−.194 (.221)	.773	.823 (.534, 1.270)	.379	.772 (.493, 1.210)	.260
Familiar, not related	300	135 (45.0)	165 (55.0)	.021 (.255)	.007	1.021 (.620, 1.683)	.934	1.123 (.637, 1.978)	.688
Unfamiliar	19	14 (73.7)	5 (26.3)	−1.209 (.484)	6.234	.298 (.115, .771)	.013	.705 (.220, 2.260)	.557
Not reported	25	5 (20.0)	20 (80.0)	1.207 (.672)	3.222	3.342 (.895, 12.479)	.073	57.150 (6.094, 535.968)	.001

Note. For each variable, the category counts sum to $N = 760$ (pre-Protocol = 350, 46.1%; Protocol = 410, 53.9%).
[a] Unadjusted Odds Ratios from simple regression with one categorical IV (all levels) in analysis. [b] Adjusted Odds Ratios from multiple regression with all categorical IVs in analysis. [c] Reference category is the comparison group within an IV. The OR of a Reference category is the odds of Protocol cases (# of Protocol cases ÷ # of pre-Protocol cases) for that category. The OR of a non-Reference category is the odds ratio (odds of Protocol cases for category ÷ odds of Protocol cases for Reference category). Expected OR = 1 when H_0 is true. [d] Intercept value for adjusted results: Estimate (SE) = −.991 (.644), Wald = 2.363, $p = .124$; When multiple categorical IV's are analyzed jointly, the combined Reference categories characterize the comparison group.

likely to be Protocol than non-Protocol cases, respectively, and these effects were amplified when the adjusted ORs were examined. The adjusted ORs also indicated that, controlling for all other IVs, boys were 1.36 times more likely than girls to be interviewed after the Protocol was introduced ($p = .023$). Further, in comparison with alleged victims from other ethnic groups (primarily African American and Pacific Islander combined), Protocol cases were 2.13 ($p = .029$) and 2.12 ($p = .012$) times more likely than non-Protocol cases to involve Caucasian and Hispanic complainants, respectively. When controlling for all other IVs, the OR for the Caucasian group weakened (adjusted OR = 1.78, $p = .084$) while the OR for the Hispanic group strengthened (adjusted OR = 2.23, $p = .022$). The vast majority of cases involved Caucasian or Hispanic complainants, however, and these proportions did not differ by interview condition (see Table 2).

Predicting Filed Charges Versus Other Outcomes

As shown in Table 1, charges were filed in 364 (or 47.9%) of the 760 cases: 42% of the pre-Protocol-period cases, and 52.9% of the Protocol-period cases. A total of 181 cases (23.8% of the 760) were not submitted for screening, and a further 170 (29.4% of the 579 submitted for screening) were declined by DAs, with no charges thus filed (37.8% of the pre-Protocol cases and 22.5% of the Protocol cases submitted for screening). Other outcomes were very infrequent (see Table 1).

The proportions of cases with filed charges were significantly heterogeneous among the interviewers, $\chi^2(15) = 31.48$, $p = .008$. Logistic GEEs, therefore, were conducted to examine the case characteristics, including interview condition, associated with filing of charges. The results are presented in Table 3 where we focus on the adjusted ORs to control for the significant associations between interview condition and other case-related characteristics. Statistically significant adjusted ORs revealed that six case characteristics were associated with the filing of charges: interview condition, severity of sexual abuse, victim's age, suspect's gender, suspect's age, and victim/suspect familiarity.

Charges were significantly more likely to be filed in Protocol cases (52.9%) than in pre-Protocol cases (42%). The estimated

Table 3
Associations With Charges Filed: Results of Simple and Multiple Logistic Generalized Estimating Equations (GEE)

Variable	*N*	Charges Filed *n* (row %)	Estimate (SE)	Wald χ^2 *df* = 1	Unadjusted[a] OR (95% CI)	*p*	Adjusted[b] OR (95% CI)	*p*
Interview condition								
Pre-Protocol (Reference)	350	147 (42.0)	−.323 (.164)	3.855	.724[c] (.525, .999)	.050	.049[d] (.013, .185)	—
Protocol	410	217 (52.9)	.440 (.154)	8.130	1.553 (1.147, 2.101)	.004	1.458 (1.057, 2.012)	.022
Severity of sexual abuse								
Exposure	39	16 (41.0)	−.781 (.229)	11.673	.458 (.292, .717)	.001	.314 (.192, .511)	.001
Touch	430	192 (44.7)	−.633 (.135)	22.128	.531 (.408, .691)	.001	.557 (.418, .741)	.001
Penetration (Reference)	257	155 (60.3)	.418 (.134)	9.721	1.520[c] (1.168, 1.977)	.002	.049[d] (.013, .185)	—
Not reported	34	1 (2.9)	−3.915 (.958)	16.712	.020 (.003, .130)	.001	.022 (.003, .163)	.001
Victim's age (years)								
2.8 to 4 (Reference)	135	42 (31.1)	−.795 (.196)	16.378	.452[c] (.307, .664)	.001	.049[d] (.013, .185)	—
5 to 6	164	79 (48.2)	.722 (.182)	15.695	2.058 (1.440, 2.941)	.001	1.961 (1.354, 2.840)	.001
7 to 9	215	118 (54.9)	.991 (.181)	29.882	2.694 (1.888, 3.843)	.001	2.708 (1.812, 4.046)	.001
10 to 13	246	125 (50.8)	.827 (.160)	26.744	2.287 (1.672, 3.130)	.001	2.383 (1.596, 3.560)	.001
Victim's gender								
Male	214	104 (48.6)	.039 (.154)	.065	1.040 (.768, 1.408)	.800	.984 (.717, 1.352)	.923
Female (Reference)	546	260 (47.6)	−.095 (.142)	.450	.909[c] (.688, 1.201)	.502	.049[d] (.013, .185)	—
Victim ethnicity								
Caucasian	609	299 (49.1)	.309 (.266)	1.342	1.362 (.808, 2.296)	.247	.963 (.534, 1.739)	.901
Hispanic	80	34 (42.5)	.043 (.350)	.015	1.043 (.525, 2.072)	.903	.739 (.308, 1.768)	.496
Other (Reference)	41	17 (41.5)	−.345 (.234)	2.164	.708[c] (.447, 1.121)	.141	.049[d] (.013, .185)	—
Not reported	30	14 (46.7)	.211 (.505)	.175	1.235 (.459, 3.324)	.676	1.013 (.285, 3.604)	.985
Suspect's gender								
Male	616	317 (51.5)	2.113 (.570)	13.740	8.270 (2.706, 25.270)	.001	9.765 (3.073, 31.026)	.001
Female (Reference)	44	5 (11.4)	−2.054 (.512)	16.102	.128[c] (.047, .350)	.001	.049[d] (.013, .185)	—
Not reported	100	42 (42.0)	1.731 (.558)	9.623	5.648 (1.892, 16.865)	.002	14.837 (4.937, 45.592)	.001
Suspect's age								
Juvenile	280	165 (58.9)	.620 (.203)	9.338	1.860 (1.249, 2.768)	.002	1.933 (1.371, 2.725)	.001
Adult (Reference)	411	179 (43.6)	−.259 (.136)	3.617	.772[c] (.591, 1.008)	.057	.049[d] (.013, .185)	—
Not reported	69	20 (29.0)	−.637 (.249)	6.550	.529 (.325, .861)	.010	.562 (.359, .879)	.012
Victim/suspect familiarity								
Immediate family (Reference)	279	123 (44.1)	−.238 (.170)	1.959	.788[c] (.565, 1.100)	.162	.049[d] (.013, .185)	—
Other family	137	77 (56.2)	.487 (.207)	5.557	1.628 (1.086, 2.440)	.018	1.664 (1.079, 2.567)	.021
Familiar not related	300	156 (52.0)	.318 (.203)	2.439	1.374 (.922, 2.047)	.118	1.280 (.833, 1.967)	.260
Unfamiliar	19	3 (15.8)	−1.436 (.815)	3.105	.238 (.048, 1.175)	.078	.244 (.057, 1.042)	.057
Not reported	25	5 (20.0)	−1.149 (.426)	7.265	.317 (.138, .731)	.007	.362 (.145, 903)	.029

Note. For each variable, the category counts sum to $N = 760$ (Charges filed = 364, 47.9%).
[a] Unadjusted Odds Ratios from simple regression with one categorical IV (all levels) in analysis. [b] Adjusted Odds Ratios from multiple regression with all categorical IVs in analysis. [c] Reference category is the comparison group within an IV. The OR of a Reference category is the odds of charges filed (# cases with charges filed ÷ # cases with charges not filed) for that category. The OR of a non-Reference category is the odds ratio (odds of charges filed for category ÷ odds of charges filed for Reference category). Expected OR = 1 when H_0 is true. [d] Intercept value for adjusted results: Estimate (SE) = −3.012 (.674), Wald = 19.938, $p = .000$; When multiple categorical IVs are analyzed jointly, the combined Reference categories characterize the comparison group.

odds that charges would be filed in Protocol cases were 1.46 times (or 46%) higher than in pre-Protocol cases, when the effects of other case characteristics were controlled. With respect to the severity of the sexual abuse, 60.3% of the cases involving penetration compared with 41% and 44.7% of the cases involving exposure and touch, respectively, resulted in charges being filed with significantly reduced odds for touch and exposure (.31 and .56 times as likely than with penetration). Cases involving touch were 1.78 times more likely than those involving exposure to have charges filed (95% CI: 1.0, 3.14; $p = .049$). The youngest (2.8- to 4-year-old) alleged victims were least likely (31.1%) to have charges filed: For 5- to 6-year-olds, the percentage was 48.2%, 1.96 times as likely as with the youngest children; for 7- to 9-year-olds, 54.9%, 2.7 times as likely; and for 10- to 13-year-olds, 50.8%, 2.4 times as likely. In addition, in comparison to 5- to 6-year-olds, charges were 1.38 times more likely to be filed for 7- to 9-year olds (95% CI: 1.04, 1.83; $p = .025$) and 1.22 times more likely to be filed for 10- to 13-year olds (95% CI: .98, 1.51; $p = .083$).

Nearly half (44.1%) of the cases in which suspects were immediate family members resulted in charges being filed. Extended-(other) family suspects were 1.66 times more likely to be charged (56.2%) than immediate-family suspects, while unfamiliar suspects (of whom there were only 19) were .244 times as likely to be charged (15.8%) as immediate-family members (see Table 3), perhaps because some unfamiliar suspects were not identified. Unfamiliar suspects were also .146 times as likely to be charged as extended family suspects (95% CI: .038, .561; $p = .005$) and .19 times as likely to be charged as familiar but unrelated suspects (95% CI: .051, .707; $p = .013$). Extended family members were 1.30 times more likely to be charged than familiar but unrelated suspects (95% CI: 1.0, 1.69; $p = .05$).

There were few female suspects (n = 44, 5.8% of 760), and a small percentage of them (11.4%) were charged; males were 9.76 times more likely to be charged than females. Notably, suspect gender was not reported for more cases (n = 100) than there were female suspects, raising a question about the reliability of the gender difference. However, the estimated adjusted odds for males was still 1.61 times greater than the odds for females and gender-not-reported suspects combined (95% CI: .955, 2.05; p = .074), suggesting that the gender difference was reliable.

Juvenile suspects (58.9%) were 1.93 times more likely to be charged than adult suspects (43.6%). Post hoc analyses showed that, compared to cases with adult suspects, cases with juvenile suspects were less likely to involve the oldest (10- to 13-year-olds) alleged victims (22.9% vs. 39.7%) or immediate family members (30.0% vs. 45.3%), and were more likely to involve male alleged victims (37.1% vs. 21.2%). Cases involving juvenile suspects were also more likely to involve penetration (37.9% vs. 31.1%) and less likely to involve touch (52.5% vs. 59.9%) than were cases involving older suspects.

Final Dispositions

As shown in Table 4, 97% (n = 353) of the 364 cases with charges filed reached disposition, 2.7% (n = 10) remained active, and information were missing for .3% (n = 1). The majority of cases, 59% (n = 215), concluded with a single disposition: Of these cases, 42.3% (n = 91) of the suspects pled guilty as charged, 32.6% (n = 70) pled guilty to reduced charges, and 16.3% (n = 35) had all counts dismissed. Of the 138 cases in which there were mixed dispositions, 71% (n = 98) of the suspects pled guilty as charged to at least one count with the remaining counts dismissed, 16% (n = 22) pled guilty to at least one reduced charge with the remaining counts dismissed, and 3% (n = 4) pled guilty to at least one count as charged as well as to reduced charges. Overall, of the 364 suspects charged, 84.3% (n = 307) pled guilty, or were found guilty at trial to one or more counts.

To assess differences in disposition related to interview condition, the possible combinations of disposition were sorted into one of seven categories: still active or no information, diverted, all charges dismissed, trial–not guilty, trial–guilty, pled guilty as charged, and plea reduced (see Table 4). A Fisher's exact test revealed a statistically significant association between trial–not guilty and interview condition, but there were no other significant associations between interview condition and disposition (ps = .13 to 1.00). For the 23 cases that went to trial, guilty verdicts were significantly more likely in Protocol than in pre-Protocol cases (Fisher's exact test, p = .045) with only 1 of the 11 Protocol cases not resulting in conviction, in contrast with six of the 12 pre-Protocol cases (see Table 4). There was no association between interview condition and dismissal, or between dismissal and other case characteristics.

Discussion

The primary objective of the present study was to determine whether the introduction of the evidence-based NICHD Protocol was associated with a change in the outcomes of cases in which child sexual abuse was suspected. The results showed significant differences between cases conducted prior to and after the introduction of the Protocol at two crucial decision points: The filing of charges by prosecutors and the final judicial disposition, through either plea negotiation or trial. Specifically, charges were more likely to be filed following the introduction of the Protocol. Once charges were filed, both pre-Protocol and Protocol interviews were highly (and similarly) likely to yield guilty pleas to one or more counts. As in earlier research on the outcomes of child sexual abuse cases (see Cross et al., 2003, for review), a minority of cases in the present sample proceeded to trial, but when they did, Protocol interviews were associated with a significantly higher rate of conviction. Unfortunately, we do not have information about the trials themselves or about the fact-finders' decision-making processes.

The fact that prosecutors were more likely to file charges following the introduction of the Protocol is most significant. Prior to introduction of the Protocol, charges were filed in 42% of investigated cases, whereas 52.9% of the investigated cases led to the filing of charges after the Protocol was introduced. After the Protocol was introduced, conversely, proportionally fewer cases were dropped by police investigators prior to screening (i.e., "not cleared"), and thus not submitted to the prosecutors for screening, or "screened but declined" by prosecutors. As in previous studies (e.g., Cross et al., 1995, 2003; MacMurray, 1989), the initial screening seemed to be crucial. Many nonprosecuted cases were declined during the screening process, and proportionally fewer

Table 4
Disposition by Interview Condition

Disposition	Pre-Protocol		Protocol		Total			
	n	% [a]	n	% [a]	N	%[a]	$\chi^2(1)$	p[c]
Still active/No information	7	4.76	4	1.84	11	3.02	FET[b]	.13
Diverted	2	1.36	2	.92	4	1.10	FET[b]	1.00
All charges dismissed	13	8.84	22	10.13	35	9.62	.17	.72
Trial	12	8.16	11	5.07	23	6.32	1.42	.27
Not guilty	6	4.08	1	.46	7	1.92	FET[b]	.02
Guilty	6	4.08	10	4.61	16	4.40	.06	1.00
Plea agreement	113	76.87	178	82.03	291	79.95	1.44	.29
Pled guilty	77	52.38	122	56.22	199	54.67	.52	.52
Plea reduced	36	24.49	56	25.81	92	25.27	.08	.81
Total	147	100.00	217	100.00	364	100.00		

[a] Column %. [b] Fisher's Exact Test. [c] p-values are exact.

cases were declined in the Protocol (17.6%) than in the pre-Protocol (28%) period. As a result, proportionately more Protocol than pre-Protocol cases were prosecuted.

Use of the Protocol may have been associated with more charges than pre-Protocol interviews because Protocol interviews have been shown to involve fewer option-posing or inappropriate questions and more of the recall prompts and techniques that elicit higher quality and more compelling information than non-Protocol interviews (e.g., Cyr & Lamb, 2009; Hershkowitz, 2001, 2002; Hershkowitz, Lamb, & Orbach, 2008; Lamb, Sternberg, Orbach, Esplin & Mitchell, 2002; Lamb et al., 2003, 2008, 2009; Lamb, Orbach, Hershkowitz, et al., 2007; Orbach et al., 2000; Sternberg, Lamb, Orbach et al., 2001). Improvements in the quality of victim statements are likely to have accounted for increases in the proportion of cases in which charges are filed, although other factors may have been important, too. For example, Darvish, Hershkowitz, Lamb, and Orbach (2008) found that forensic interviews conducted using the Protocol yielded more investigative leads (i.e., information suggesting new avenues for investigation) as well as more central, stronger, and more verifiable leads than those produced in non-Protocol interviews, and the availability of such leads may have facilitated more thorough and productive investigations of the alleged crimes.

In addition to interview condition, several other variables were significantly related to the filing of charges. Cases involving the youngest alleged victims (between the ages of 2.8 and 4 years) were the least likely to yield criminal charges, regardless of interview condition. Age differences in filing rates have also been reported by other researchers (e.g., Cross et al., 1994, 1995; Stroud et al., 2000; Tjaden & Thoennes, 1992), however, and there are several reasons why cases involving the youngest children were less likely to be prosecuted. Although young children can remember and provide coherent accounts of their experiences, even after long delays (Fivush, 1997; Fivush & Shukat, 1995; Lamb, Sternberg, Orbach, Esplin et al., 2002; Pipe, Gee, Wilson, & Egerton, 1999), they typically retrieve significantly less information than older children do and interviewers often need to provide more prompts to elicit the information than when interviewing older children (Hershkowitz, Lamb, Orbach, Katz, & Horowitz, 2012; Lamb et al., 2003). Previous field research has also shown that a surprisingly large proportion (more than half) of young (3- to 6-year-old) suspected victims of abuse do not make allegations when interviewed (Hershkowitz, Horowitz, & Lamb, 2007; Pipe et al., 2007). Hershkowitz, Horowitz, and Lamb (2007) suggested that younger children are disproportionately likely to misunderstand the purpose and focus of the investigative interview or the abuse itself, thus failing to report experiences that they could discuss if they recognized the investigators' interest. Younger children may also be more reluctant to disclose and talk about abuse (Pipe et al.,2007; Pipe & Goodman, 1991). Thus both cognitive and motivational factors might have affected the quality of the information children provided in their interviews, even when interviewed under the optimal conditions promoted by the Protocol. Less complete accounts provided by the younger children in the present study may have meant they did not provide sufficient evidence to convince prosecutors that a conviction could be obtained at trial.

As noted above, children may be motivated to withhold information or deny that they were abused because they wish to protect familiar suspects, especially caregivers on whom they are dependent, and/or when they have been pressured to remain silent (Hershkowitz, Horowitz et al., 2007; Paine & Hansen, 2002; Pipe et al., 2007; Yuille, Tymofievich, & Marxsen, 1995). Hershkowitz, Orbach, and colleagues (2007), for example, found that children who were suspected victims of parental abuse provided proportionally fewer informative responses and more uninformative responses (e.g., omission, "don't know," "don't want to talk," "don't remember") than children who were suspected victims of suspects who were not their parents. They also provided fewer details per response than did counterparts believed to have been abused by individuals other than their parents. Similarly, Hershkowitz, Horowitz, and colleagues (2007) reported that almost half of the children whose parents were divorced failed to disclose sexual abuse when this was suspected and that very young children more readily made allegations against familiar nonfamily members but were less likely to make allegations against parents and stepparents. Clearly, it is problematic that some child victims remain vulnerable because they are reluctant to discuss abuse and/or lack the linguistic or cognitive abilities to provide detailed accounts of abuse that might lead to protective action being taken. Approaches to the investigation of abuse in these cases, and the development of interviewing techniques sensitive to the needs of these children, are clearly needed (see Hershkowitz, Orbach et al., 2007, for further discussion).

Juvenile suspects were significantly more likely to be charged (58.9% of the 280 cases involving juvenile suspects led to charges) than were adult suspects (43.6% of the 411 cases involving adults suspects: see also Tjaden & Thoennes, 1992), regardless of interview condition. It is interesting that Protocol-era interviews that went to trial were also more likely to involve juvenile suspects (45.5%, five cases) than were pre-Protocol interviews (16.7%, two cases), although given the small numbers of cases this difference cannot be considered reliable.

Two other variables were associated with the filing of charges regardless of interview condition: abuse severity and gender of suspect. Cases involving the most severe abuse (penetration) were more likely to have charges filed, as in previous studies (Brewer et al., 1997; Chapman & Smith, 1987; Cross et al., 1994, 1995; Sedlak et al., 2005; Stroud et al., 2000; Tjaden & Thoennes, 1992). Once charges were filed, however, there were no differences in disposition or in the likelihood that the cases would be tried. Male suspects were much more likely to be charged than female suspects, although cases involving female suspects were relatively rare (5.8%). The gender of the alleged victims had no effect on case outcome, whereas, as in previous studies, charges were more likely when female victims were involved (e.g., Cross et al., 1994, 1995; Stroud et al., 2000; Tjaden & Thoennes, 1992).

As in previous studies (e.g., Cross et al., 1995; Stroud et al., 2000), only a small percentage of cases (8.2% and 5.1% for pre-Protocol and Protocol cases, respectively) with filed charges went to trial, with rates at the low end of those reported in Cross et al.'s (2003) meta-analysis. Plea agreements speeded up case disposition, circumvented the greater uncertainty of trials for both prosecution and defense, and, of course, ensured that child victims did not have to appear in court to be cross-examined. Although few cases went to trial, the high rate of convictions in trials following introduction of the Protocol is nonetheless striking. Whereas previous studies have reported trial conviction rates between 50% and

75% (see Cross et al., 1995), trial conviction rate for Protocol cases in the present study amounted to 91%. Only one of the 11 Protocol cases (9%) tried yielded an acquittal in contrast to six of the 12 pre-Protocol cases (50%). Cross et al. (2003) point out that although only a small number of cases go to trial, they are important both because of their potential impact on children who must testify (see, e.g., Goodman, Batterman-Faunce, & Kenney, 1992), and on future decisions by prosecutors and defendants (and their attorneys), potentially influencing the much larger category of plea negotiations.

Limitations of the Present Study

One of the strengths of the current study was that the same detectives conducted the pre-Protocol and Protocol interviews, thereby unconfounding the effects of the interviewer and of the Protocol. The potential disadvantage of this design is, of course, that the outcome data were necessarily collected in different (albeit adjacent) time periods, before and after introduction of the interview Protocol. A between-subjects design in which the same detectives were randomly assigned to conduct either Protocol or pre-Protocol interviews was not feasible. Random assignment of different detectives to standard interviewing and Protocol conditions might have been possible had the jurisdiction been larger, and had this not been a retrospective study.

There was continuity of personnel over the period of data collection; the detectives, supervisors, prosecutors and judges who handled these cases throughout the study period were the same and there were no changes in leadership or declared policy other than adoption of the Protocol. Nonetheless, there may have been a general "drift" that influenced outcomes, and of course the participating interviewers were necessarily more experienced later in the study period. Comparison of case outcomes on a year-by-year basis prior to (i.e., 1994 to mid-1997) and after (mid-1997 to 2000) introduction of the Protocol failed to reveal any systematic trends attributable to time-related variables other than the introduction of the Protocol interview. It is also possible that awareness of the Protocol's reported strengths, rather than the quality of the investigations themselves, changed the practices of DAs and defense attorneys in ways that increased the likelihood of conviction.

A second limitation (attributable to the retrospective nature of the study) is that there were some missing data with respect to case (victim and suspect) characteristics. Pre-Protocol cases were more likely to have missing data because once the Protocol was introduced, case characteristics were often tabulated as part of ongoing research.

Third, our study focused on one county in the state of Utah. Utah has unique demographic characteristics, with relatively high proportions of Caucasians and people belonging to the Church of Jesus Christ of Latter-day Saints (Mormons). Salt Lake County is not as homogeneous in these respects as other parts of the state, but approximately 55% of the population identify as Mormons. Previous studies have found regional differences in prosecution rates, with higher rates in rural and urban than in suburban areas (e.g., Ménard & Ruback, 2003), as well as differences across ethnic groups, with higher rates of prosecution for ethnic minority suspects (e.g., Stroud et al., 2000; Tjaden & Thoennes, 1992), but there is no obvious reason for suspecting that the findings might have been affected by demographic characteristics of the population studied. The interview Protocol has now been proven effective in studies conducted in several different countries, with a wide range of suspected victims (see Lamb et al., 2008, for review). Further studies, including replication of the present study by researchers from another research group, are necessary to establish whether improved interview practices similarly are associated with differences in outcomes in other investigative and judicial systems. In such research, it would also be important to ascertain how closely interviewers adhered to the Protocol after being trained to use it.

Investigations of child sexual abuse do not readily lend themselves to tightly controlled experimental studies. Extracting information for research purposes that has been recorded as part of an investigation is time consuming and frustrating. Nonetheless, we were able to obtain necessary information about a very large number of cases, and our results complement those obtained in similarly large studies by other researchers (e.g., Cheit & Goldschmidt, 1997; Cross, Jones, Walsh, Simone, & Kolko, 2007; Davis & Wells, 1996; Stroud et al., 2000). However, replication of the current study in different jurisdictions, particularly in a prospective design involving multiple baselines, would overcome the limitations noted here.

Conclusions

The quality of forensic interviewing practices is of utmost importance if the rights of both child victims and innocent suspects are to be protected. When child abuse is suspected, children's verbal allegations often constitute the only available evidence, and our research group has been developing and evaluating an evidence-based forensic interview Protocol for nearly two decades. Evidence that charges are more likely to be filed and that suspects are more likely to be convicted at trial provides strong endorsement indeed of a best-practice approach to interviewing, with potentially important implications for policy and practice regarding child victims.

Because so few cases go to trial, as Cross et al. (1995) concluded, ". . . it seems more productive to focus improvement efforts on the pretrial phase of prosecution. For example, enhancing the quality of investigations may have a substantially larger impact on prosecution than rulings that affect the admissibility of evidence, even if we take into account the "ripple" effect that new precedents have on the whole system" (p. 1439). The findings reported here show that when interviews are conducted in accordance with best practices, there may be an associated change in case outcomes.

While the findings reported here provide very strong support for the use of the NICHD Protocol, and complement the large number of studies evaluating its use in the field (Lamb et al., 2008; Pipe et el., 2007; Orbach et al., 2000; Orbach & Lamb, 2007), there is of course need for further research, especially to replicate the very striking association between the NICHD Protocol and conviction rates. In addition, approaches to interviewing suspected victims of within-family abuse and very young children need to be considered further in light of the distinctive effects noted above. It is nevertheless clear that improvements in the quality of investigative interviewing are associated with increases in the probability that substantiated allegations of sexual abuse are addressed appropriately in the criminal justice system.

References

Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). Hoboken, NJ: Wiley. doi:10.1002/0470114754

Bradshaw, T. L., & Marks, A. E. (1990). Beyond a reasonable doubt: Factors that influence the legal disposition of child sexual abuse cases. *Crime & Delinquency, 36,* 276–285. doi:10.1177/0011128790036002006

Brewer, K. D., Rowe, D. M., & Brewer, D. D. (1997). Factors related to prosecution of child sexual abuse cases. *Journal of Child Sexual Abuse, 6,* 91–111. doi:10.1300/J070v06n01_07

Chapman, J. R., & Smith, B. E. (1987). *Child sexual abuse: An analysis of case processing*. Washington, DC: American Bar Association.

Cheit, R. E., & Goldschmidt, E. (1997). Child molesters in the criminal justice system: A comprehensive case-flow analysis of the Rhode Island docket (1985–1993). *New England Journal on Criminal and Civil Confinement, 23,* 267–301.

Cross, T. P., DeVos, E., & Whitcomb, D. (1994). Prosecution of child sexual abuse: Which cases are accepted? *Child Abuse & Neglect, 18,* 663–677. doi:10.1016/0145-2134(94)90016-7

Cross, T. P., Jones, L. M., Walsh, W. A., Simone, M. K., & Kolko, D. (2007). Child forensic interviewing in Children's Advocacy Centers: Empirical data on a practice model. *Child Abuse & Neglect, 31,* 1031–1052. doi:10.1016/j.chiabu.2007.04.007

Cross, T. P., Walsh, W. A., Simone, M., & Jones, L. M. (2003). Prosecution of child abuse: A meta-analysis of rates of criminal justice decisions. *Trauma, Violence, & Abuse, 4,* 323–340. doi:10.1177/1524838003256561

Cross, T. P., Whitcomb, D., & DeVos, E. (1995). Criminal justice outcomes of prosecution of child sexual abuse: A case flow analysis. *Child Abuse & Neglect, 19,* 1431–1442. doi:10.1016/0145-2134(95)00106-2

Cyr, M., & Lamb, M. E. (2009). Assessing the effectiveness of the NICHD Investigative Interview Protocol when interviewing French-speaking alleged victims of child sexual abuse in Quebec. *Child Abuse and Neglect, 33,* 257–268. doi:10.1016/j.chiabu.2008.04.002

Darvish, T., Hershkowitz, I., Lamb, M. E., & Orbach, Y. (2008, March). *The effect of the NICHD interview protocol on the elicitation of investigative leads in child sexual abuse investigations*. Paper presented in the conference of the American Psychology and Law Society (AP-LS), Jacksonville, FL.

Davis, N. S., & Wells, S. J. (1996). *Justice system processing of child abuse and neglect cases: Final report, Volume II*. Washington, DC: American Bar Association and National Institute of Justice.

Finkelhor, D., Cross, T. P., & Cantor, E. N. (2005). The justice system for juvenile victims: A comprehensive model of case flow. *Trauma, Violence, & Abuse, 6,* 83–102. doi:10.1177/1524838005275090

Fivush, R. (1997). Event memory in early childhood. In C. Nelson (Ed.), *The development of memory in childhood* (pp. 139–161). Hove, UK: Psychology Press.

Fivush, R., & Shukat, J. R. (1995). Content, consistency, and coherence of early autobiographical recall. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (Vol. 1, pp. 5–23). Thousand Oaks, CA: Sage.

Giovannoni, J. M. (1991). Unsubstantiated reports: Perspectives of child protection workers. *Children & Youth Services, 15,* 51–62. doi:10.1300/J024v15n02_03

Goodman, G. S. (2006). Children's eyewitness memory: A modern history and contemporary commentary. *Journal of Social Issues, 62,* 811–832. doi:10.1111/j.1540-4560.2006.00488.x

Goodman, G. S., Batterman-Faunce, J. M., & Kenney, R. (1992). Optimizing children's testimony: Research and social policy issues concerning allegations of child sexual abuse. In D. Cicchetti & S. Toth (Eds.), *Child abuse, child development, and social policy* (pp. 139–166). Norwood, NJ: Ablex.

Hanley, J. A., Negassa, A., Edwardes, M. D., & Forrester, J. E. (2003). Statistical analysis of correlated data using generalized estimating equations: An orientation. *American Journal of Epidemiology, 157,* 364–375. doi:10.1093/aje/kwf215

Hartmann, D. P., Pelzel, K. E., & Abbott, C. B. (2011). Design, measurement, and analysis in developmental research. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (6th ed., pp. 109–197). New York, NY: Psychology Press.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.

Hershkowitz, I. (2001). Children's responses to open-ended utterances in investigative interviews. *Legal and Criminological Psychology, 6,* 49–63. doi:10.1348/135532501168190

Hershkowitz, I. (2002). The role of facilitative prompts in interviews of alleged sex and abuse victims. *Legal and Criminological Psychology, 7,* 63–71. doi:10.1348/135532502168388

Hershkowitz, I., Horowitz, D., & Lamb, M. E. (2005). Trends in children's disclosure of abuse in Israel: A national study. *Child Abuse & Neglect, 29,* 1203–1214. doi:10.1016/j.chiabu.2005.04.008

Hershkowitz, I., Horowitz, D., & Lamb, M. E. (2007). Individual and family variables associated with disclosure and non-disclosure of child abuse in Israel. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A.-C. Cederborg (Eds.). *Child sexual abuse: Disclosure, delay and denial* (pp. 63–76). Mahwah NJ: Erlbaum.

Hershkowitz, I., Lamb, M. E., & Orbach, Y. (2008, March). *The effects of the NICHD investigative interview protocol on assessment of credibility in child sexual abuse allegations*. Paper presented in the conference of the American Psychology and Law Society (AP-LS), Jacksonville, FL.

Hershkowitz, I., Lamb, M. E., Orbach, Y., Katz, C., & Horowitz, D. (2012). The development of communicative and narrative skills among preschoolers: Lessons from forensic interviews about child abuse. *Child Development, 83,* 611–622. PMid:22181976.

Hershkowitz, I., Orbach, Y., Lamb. M. E., Sternberg, K. J., Pipe, M.-E., & Horowitz, D. (2007). Suspected victims of abuse who do not make allegations: An analysis of their interactions with forensic interviewers. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A.-C. Cederborg (Eds.), *Child sexual abuse: Disclosure, delay and denial* (pp. 97–114). Mahwah, NJ: Erlbaum.

Home Office. (2011). *Achieving best evidence in criminal proceedings: Guidance for vulnerable and intimidated witnesses, including children*. London, UK: Home Office Communications Directorate.

Jensen, J. M., Jacobson, M., Unrau, Y., & Robinson, R. L. (1996). Intervention for victims of child sexual abuse: An evaluation of the Children's Advocacy Model. *Child & Adolescent Social Work Journal, 13,* 139–156.

Joa, D., & Edelson, M. G. (2004). Legal outcomes for children who have been sexually abused: The impact of child abuse assessment center evaluations. *Child Maltreatment, 9,* 263–276. doi:10.1177/1077559504267004PMid:15245679

Jones, L. M., Cross, T. P., Walsh, W. A., & Simone, M. (2005). Criminal investigation of child abuse: The research behind "best practices". *Trauma, Violence, & Abuse, 6,* 254–268. doi:10.1177/1524838005277440

Lamb, M. E., Hershkowitz, I., Orbach, Y., & Esplin, P. W. (2008). *Tell me what happened: Structured investigative interviews of child victims and witnesses*. Chichester, UK: Wiley.

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). Structured forensic interview protocols improve the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse & Neglect 31,* 1201–1231.

Lamb, M. E., Orbach, Y., Hershkowitz, I., Horowitz, D., & Abbott, C. B. (2007). Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? *Applied Cognitive Psychology, 20,* 1–14. doi:10.1002/acp.1318

Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., . . . Bowler, L. (2009). Use of a structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology, 23,* 449–467. doi:10.1002/acp.1489

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., & Mitchell, S. (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with alleged abused children? *Applied Developmental Science, 6,* 35–41. doi:10.1207/S1532480XADS0601_04

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology, 71,* 926–934. doi:10.1037/0022-006X.71.5.926PMid:14516241

MacMurray, B. K. (1989). Criminal determination for child sexual abuse. *Journal of Interpersonal Violence, 4,* 233–245. doi:10.1177/088626089004002007

Ménard, K. S., & Ruback, R. B. (2003). Prevalence and processing of child sexual abuse: A multi-data-set analysis of urban and rural counties. *Law and Human Behavior, 27,* 385–402. doi:10.1023/A:1024084916626

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse and Neglect, 24,* 733–752. doi:10.1016/S01452134(00)00137-X

Orbach, Y., & Lamb, M. E. (2007). Young children's references to temporal attributes of allegedly experienced events in the course of forensic interviews. *Child Development, 78,* 1100–1120. doi:10.1111/j.1467-8624.2007.01055.x

Paine, M. L., & Hansen, D. J. (2002). Factors influencing children to self-disclose sexual abuse. *Clinical Psychology Review, 22,* 271–295. doi:10.1016/S0272-7358(01)00091-5

Perona, A. L., Bottoms, B. L., & Sorenson, E. (2006). Research-based guidelines for child forensic interviews. *Journal of Aggression, Maltreatment, & Trauma, 12,* 81–130. doi:10.1300/J146v12n03_04

Pipe, M.-E., Gee, S., Wilson, J. C., & Egerton, J. M. (1999). Children's recall one and two years after the event. *Developmental Psychology, 35,* 781–789. doi:10.1037/0012-1649.35.3.781

Pipe, M.-E., & Goodman, G. S. (1991). Elements of secrecy: Implications for children's testimony. *Behavioral Sciences & the Law, 9,* 33–41. doi:10.1002/bsl.2370090105

Pipe, M.-E., Sternberg, K., Lamb, M. E., Orbach, Y., Stewart, H., & Esplin, P. (2007). Factors associated with non-disclosures of suspected abuse during forensic interviews. In M.-E. Pipe, M. E. Lamb, Y. Orbach, & A-C Cederborg (Eds.), *Child sexual abuse: Disclosure, delay and denial* (pp. 77–96). Mahwah NJ: Erlbaum.

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association. doi:10.1037/10301-000

Poole, D. A., & Lindsay, D. S. (1998). Assessing the accuracy of young children's reports: Lessons from the investigation of child sexual abuse. *Applied and Preventive Psychology, 7,* 1–26. doi:10.1016/S0962-1849(98)80019-X

Sedlak, A. J., Doueck, H. J., Lyons, P., Wells, S. J., Schultz, D., & Gragg, F. (2005). Child maltreatment and the justice system. *Research on Social Work Practice, 15,* 389–403. doi:10.1177/1049731505277947

Snijders, T. A., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage.

Steele, P. D., Norris, M., & Komula, K. (1994). *Evaluation of Children's Safehouse of Albuquerque*. Albuquerque, NM: The Youth Resource and Analysis Center, Institute for Social Research, University of New Mexico.

Sternberg, K. J., Lamb, M. E., Orbach, Y., Esplin, P. W., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free recall prompts in the course of forensic interviews. *Journal of Applied Psychology, 86,* 997–1005. doi:10.1037/0021-9010.86.5.997

Stroud, D., Martens, S. L., & Barker, J. (2000). Criminal investigation of child sexual abuse: A comparison of cases referred to the prosecutor to those not referred. *Child Abuse & Neglect, 24,* 689–700. doi:10.1016/S0145-2134(00)00131-9

Tjaden, P. G., & Anhalt, J. (1994). *The impact of joint law enforcement–child protective services investigations in child maltreatment cases*. Final report for Grant number 90-CA-1446, submitted to the National Center on Child Abuse and Neglect. Denver, CO: Center for Policy Research.

Tjaden, P. G., & Thoennes, N. (1992). Predictions of legal intervention in child maltreatment cases. *Child Abuse & Neglect, 16,* 807–821. doi:10.1016/0145-2134(92)90083-4

Walsh, W. A., Lippert, T., Cross, T. P., Maurice, D. M., & Davison, K., S. (2008). How long to prosecute child sexual abuse for a community using a children's advocacy center and two comparison communities? *Child Maltreatment, 13,* 3–13. doi:10.1177/1077559507307839

Wood, J. M., & Garven, S. (2000). How sexual abuse interviews go astray: Implications for prosecutors, police, and child protective services. *Child Maltreatment, 5,* 109–118. doi:10.1177/1077559500005002003

Yuille, J. C., Tymofievich, M., & Marxsen, D. (1995). The nature of allegations of child sexual abuse. In T. Ney (Ed.), *True and false allegations of child sexual abuse: Assessment and case management* (pp. 21–46). New York, NY: Brunner/Mazel.

Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equations approach. *Biometrics, 44,* 1049–1060. doi:10.2307/2531734

Received March 25, 2012
Revision received July 20, 2012
Accepted August 16, 2012 ■