FUJP
Open Access Journal

Research Article

Meta-Analysis of the NICHD Investigative Interview Protocol for Child Sexual Abuse Cases

Rohia Nusrat[1], Sumbal Nawaz[2], Sumara Masood Ul Hassan[2]
1.Centre of Counselling and Career Advisory, National University of Science and Technology
2.Department of Behavioral Sciences, National University of Science and Technology
For Correspondence: Sumbal Nawaz. Email: sumbalnawaz@s3h.nust.edu.pk

Abstract

Background. Forensic interviews are pivotal to the investigation of child sexual abuse cases. Following best practice, evidence-based guidelines when conducting such interviews is essential in obtaining a credible and reliable testimony. The investigative interview protocol developed by the National Institute for Child Health and Human Development (NICHD) incorporates these guidelines into a structured interview procedure. An integral part of the NICHD protocol is to promote open-ended, invitational interviewer prompts. The current research was conducted to observe the impact of the NICHD Protocol in encouraging invitational prompts and discouraging suggestive prompts when compared to non-protocol interviews.

Methods. A systematic review and meta-analysis on the impact of the NICHD protocol was conducted, including 11 studies selected following inclusion criteria. The inclusion criteria were determined through the PICO framework. The included literature were experimental studies comparing the NICHD protocol with other structured or unstructured interview methods in interviews conducted with children in suspected cases of child sexual abuse (CSA), with a focus on assessing the quality of investigation using interviewer prompts as a dependent measure. Electronic online databases and Registers/websites including PubMed, PsycINFO, PsycArticles, Web of Science, Wiley Online Library, Cochrane Library, Science Direct, Google Scholar, OSF, Dissertation Abstracts International, and ProQuest Dissertations and These were searched for both peer-reviewed articles and grey literature. The analysis used a random effect model that computed separate weighted mean of the effect sizes shown in independent studies for prompts namely: Directives, Option-posing, Suggestions, and Invitations.

Results. Results of the analysis indicated that overall, interviews with the NICHD protocol had less directives (g= -0.9106), option-posing (-0.9178) and suggestive prompts (g= -0.5516), and more invitational (g= 1.9859) prompts than non-protocol interviews.

Conclusion. These results corroborate with the findings of previous studies. However, these findings should be considered with caution due to the detection of high levels of heterogeneity.

Keywords. Eyewitness testimony, Investigative interviewing, Meta-analyses, Child sexual abuse, NICHD protocol, Interviewer prompt

FUI
Foundation University Islamabad

Introduction

Child sexual abuse (CSA) is a widespread public health issue that causes profound harm to victims, their families, and society (Hailes et al., 2019). Professionals that work with cases pertaining to sexual abuse of children are faced by several difficulties. For the majority of such investigations, there is a scarcity of physical evidence, and the nature of psychological symptoms is not always reliable to differentiate between non-abused and abused children (Kuehnle & Connell, 2012). These are some of the reasons that hamper the accurate detection of sexual abuse in children. Therefore, such hurdles place greater emphasis on the disclosure of the victims for both investigation and treatment purposes. Nevertheless, there are a plethora of variables that can impact the reliability of their reports such as high-stress levels (Chae et al., 2014), attention span (Goodman et al., 2017), and delayed recall (McElvaney, 2015). More particularly, certain social influences such as suggestive interviewing techniques can increase the children's vulnerability to reliably report the event (Otgaar et al., 2018). The understanding of investigators in regards to the validity and reliability of the testaments given by young victims, and the impact the conditions under which an interview is being carried out has on the victim's disclosure has received much attention in the last three decades (Baugerud & Johnson, 2017). The current literature emphasises a scientific approach, resulting in the creation of innovative, theoretically grounded methods. These methods aim to foster the establishment of rapport and support changes in disclosure strategies (Meissner et al., 2017), improve accurate and detailed memory recall (Fisher & Geiselman, 2017), and elevate the evaluation of deception in interview accounts (Vrij, 2018). Literature also recommends the usage of open-ended questions and discourage the usage of suggestive prompts (Lamb et al., 2008). One such structured investigative interview protocol has been developed by National Institute for Child Health and Human Development (NICHD) (Orbach et al., 2000). The NICHD protocol provides explicit guidelines regarding the techniques and prompts that the interviewers can use to obtain an expansive, suggestion-free account from the witnesses (Orbach et al., 2000). This meta-analysis was conducted to answer two questions. First, is the NICHD protocol more effective than other standard interviewing techniques in promoting the use of open-ended, invitational prompts? Second, is the NICHD protocol more effective than other standard interviewing techniques in reducing the use of leading and close-ended prompts (suggestions, option-posing prompts and directive prompts)?

Recommendations Regarding Interviewer Prompts

For forensic interviews that investigate child abuse, there is presently a clear consensus on the best practice guidelines. One of the main purposes of most established interview protocols is to limit prompting by the investigator so that the testimony received is in the interviewee's own words (Lamb et al., 2008). Such accounts are called free narratives and can best be produced by using questions that encourage elaborate responses such as broad, open-ended prompts (Wilson & Powell, 2012). Studies suggest accounts elicited from free recall memory are more credible than when it is provided (accurately or inaccurately) solely in response to the interviewer's prompts (Orbach & Pipe, 2011). This is further corroborated by a meta-analysis conducted by Lavoie et al. (2021) involving 23 studies showed that overall, open-ended questions were moderately effective on the children's descriptive disclosure of a sensitive experience.

Numerous protocols and guidance documents have been curated on these best practice guidelines such as the American Professional Society on the Abuse of Children (APSAC) guidelines (Saywitz et al., 2011), the Corner House Forensic Interview Protocol

(Anderson, 2013) MOGP (Home Office, 1992), Cognitive Interview (Fisher & Geiselman, 1992), the NICHD Protocol (Lamb et al., 2007), the Step-Wise Interview (Yuille et al., 1993), and the Scenario Model (Rispens & van der Sleen, 2017). While these documents have been developed based on similar principle, they vary in the degree to which the emphasise the application of these best practice guidelines (Smethurst, 2023). Literature indicates that even trained interviewers may fail to adhere to the scientifically backed recommendations. Investigators often provide very limited opportunities for children to answer open-ended questions and have a heavy dependency on asking directed or option-posing questions even very early on in the interview (Lamb, 2016). Therefore, research recommends that interviewers should receive proper training with supervision and adhere to investigative interview protocols that are structured or semi-structured in nature (Powell et al., 2010). Compared to other approaches, the NICHD protocol stands out as a highly structured guidance protocol, placing strong emphasis on strict adherence to its guidelines (Myklebust et al., 2023). Moreover, it has undergone extensive empirical research (Anderson et al., 2014).

The NICHD Protocol

A focal theme of the extensively used NICHD investigative interview protocol has been to facilitate the usage of open-ended questions by the investigator and building rapport between the interviewer and the interviewee (Benia et al., 2015). The NICHD protocol divides the forensic interview into multiple phases and provides elaborate guidelines for each phase. In the introductory phase, the investigators introduce themselves and explain to the child about the task at hand, what is expected of them, and the ground rules of the interview. The second phase is called the "Pre-substantive phase". This phase focuses on the rapport development between the interviewer and the interviewee and aims to

provide a supportive and comfortable environment for the child. During this phase, the investigator asks the child to talk about an event unrelated to the one under investigation. Before transitioning to the final phase, the interviewer asks non-suggestive, open-ended questions so that the child can make an allegation or identify the incident under investigation, after which the interview moves toward the free recall phase. If this does not happen, the investigator then carefully proceeds to give more specific prompts until the incident is not identified by the child. In the final phase called "Substantive Phase", interviewer initially gives open-ended invitational prompts so that an in-depth narrative of the incident can be obtained in the child's own words. This may then be followed by "cued invitations" so that the investigator can obtain details about a specific part of the allegation made by the child. In cases where the investigator feels that the information is still incomplete, he/she may cautiously proceed to ask close-ended questions.

Literature indicates that interviewers trained to use the NICHD Protocol and adhere to its guidelines are more likely to use open-ended prompts as compared to those interviewers who do not follow this approach (e.g. Orbach et al., 2000; Yi et al., 2016). Moreover, interviewing child witnesses properly has great relevance in the judicial setting. A study conducted by Pipe et al. (2013) showed that the guilty verdict given to suspects was higher in cases where the interviewer adhered to the guidelines of the NICHD protocol as compared to the cases where untrained investigators conducted the interview. Therefore, good quality case evidence is often linked to good interviewing.

The NICHD Protocol is considered to be amongst the most extensively researched and empirically validated investigative interview protocols (Fernandes et al., 2024; Herman, 2009). Till date, three bodies of work that synthesise literature regarding the NICHD

protocol exist. Lamb et al. (2007) published a literature review detailing development of the protocol and evaluated its effectiveness to improve the quality of the investigation. Lamb et al. (2008) later published a book, updating the review and summarising both field and laboratory studies that investigated the protocols effectiveness. According to the authors, the NICHD protocol improved the quality of an investigation by encouraging free recall, and improving the questioning techniques and other strategies employed by the interviewer. However, to the best of this researcher's knowledge, limited information could be gathered regarding strategies they had employed to retrieve included studies. Additionally, a study conducted by Benia et al. (2015) has subjected the NICHD protocol to a systematic review and meta-analysis. The study contributed greatly to the field of forensic interviewing as, subjecting the aggregation of literature to a meta-analysis enables the researcher to yield conclusive results that are statistically backed regarding the effectiveness of any intervention. However, as noted by Benia et al. (2015) themselves, only five studies were available that met their inclusion criteria. Furthermore, at the time that this review was conducted, the researchers could identify only one study that directly compared the NICHD Protocol to other structured or semi-structured interview models. Additionally, the review also provided very limited qualitative or quantitative information regarding the quality assessment of the included studies.

Aim and Scope of Current Review

Since the last review conducted in 2015, there have been several research papers that have made direct comparisons of the NICHD Protocol against other protocols (see Erens et al., 2021; Otgaar et al., 2019; Price et al., 2016). Furthermore, current research has also expanded to new geographical and cultural territories which can have implications regarding its universal applicability and efficacy with victims of child abuse belonging to non-western cultural and ethnic backgrounds (see Sumampouw et al., 2019; Yi et al., 2016). Therefore, an expansion in the field of research for the use and effectiveness of NICHD protocol as well as the methodological gaps of previous reviews may justify a revision of the already present literature synthesis.

In light of the above identified gaps, the current research aims to expand the present literature by identifying all experimental researches evaluating the effectiveness of the NICHD Protocol in improving the quality of an interview by comparing the type and number of interviewer prompts/questions used in the NICHD protocol and non-protocol interviews for suspected cases of child sexual abuse. Interviewer prompting is considered a dependant measure of the quality of an investigative interview and it is assumed that the quality improves with greater use of invitational prompts and minimal use of option-posing, directives and suggestive prompts (Lamb et al., 2007). It also aims at estimating the weighted mean of the effect sizes shown in independent studies in order to summarise the results regarding the effectiveness of NICHD protocol for interviewer prompting. Additionally, the current review aims to administer Risk for Bias assessment to evaluate the quality and methodological rigorousness of the included studies.

Method

Search Strategy

A systematic search was conducted to locate the peer-reviewed and non-peer-reviewed articles of interest. Search on various electronic online databases (5) and Registers/websites (3) which included PubMed, PsycINFO, PsycArticles, Web of Science, Wiley Online Library, Cochrane Library, Science Direct, and Google Scholar were performed in late 2021 to the start of 2022. For grey literature, additional databases (3) including OSF, Dissertation

Abstracts International, and ProQuest Dissertations and Theses were searched. The databases were searched using a combination of keywords pertaining to NICHD Investigative Interview Protocol, Child sexual abuse and Interviewer prompts, in line with the study's PICO framework and previous studies conducted in this field. We employed database-specific subject headings for each concept, combining related terms within each concept using the Boolean operator "OR." Subsequently, sets of terms for different concepts outlined above were combined using the Boolean operator "AND." Across all databases, text word searches incorporated adjacency operators (e.g. "NEAR/3") and truncation symbols (e.g. "*") as needed to capture variations in term endings and spellings. No restrictions were applied at the time of search. We also examined the reference lists of retrieved articles and previous reviews for additional references.

Inclusion and Exclusion criteria

The inclusion and exclusion of researches was determined through the PICO framework, which was based on the criteria previously outlined by Benia et al. (2015), with the exception that we also included non-peer-reviewed literature. After conducting the quality assessment, we included studies that (1) were either published in peer-reviewed journal, grey literature, dissertations/thesis or published abstract for which full articles can be retrieved; (2) compared intervention with at least one comparison group that was interviewed using either a standard or any other approach which may include experimental (RCT), quasi-experimental, or pre-post-test design; (3) had interviews conducted with children in suspected cases of child sexual abuse (4) reported interviewer prompt as a dependant measure for the quality of investigation.

The included studies in this meta-analysis must have defined the prompts as follows:

- Invitation: These are open ended statements, questions or utterances that encourage a free recall response from the interviewee. Such prompts may use general cues or cues based on any detail that the child has previously disclosed.
- Directives: A "cued-recall" prompt which directs the interviewee's attention towards information that they have already disclosed and requests further, specific details usually using wh-questions (who, what, when, where, how).
- Option-posing: prompt which directs the interviewee's attention towards information that they have not previously disclosed, requesting the interviewee to approve, disprove, or select an option given by the interviewer.
- Suggestive: The interviewer presents new information during the interview and suggests what answer is expected from the interviewee or give statements or ask questions which involve information that has not been disclosed by the interviewee.

Studies that had participants with the diagnosis of moderate to severe developmental delays or any other psychological disorder based on the diagnostic criteria of DSM-5 or ICD 10 were not included in our study. We also did not include articles focusing on Revised NICHD Investigative Interview Protocol as they are primarily focused on rapport building than interviewer questioning.

Data Screening and Extraction

Studies retrieved databases underwent title and abstract screening via Rayyan software to determine relevance, following removal of duplicates. Two independent screeners used a screening form, and any disagreements were resolved through discussion. Irrelevant articles were excluded based on title and abstract; those

not meeting inclusion criteria were also excluded. If any uncertainty arose, full-text articles were retrieved for further assessment. Percentage agreement and Cohen's Kappa were used to measure inter-rater reliability. The percentage agreement was found to be at 85.8%. Additionally, the value of Cohen's kappa calculated at this stage was .61, showing substantial agreement between the two coders, taking into account any possible agreement by chance between the two raters (Mchugh, 2012) and is considered to be within acceptable range (e.g., Pears & Sutton, 2021).

The author developed a 52 item coding sheet that helped identify important information pertaining to participant characteristics (interviewee age, gender, nature of reported abuse, interviewer training duration and delivery agent), study design (type of experiment, and sample sizes in experimental and comparison group) and outcome measure (means and standard deviations of all categories of interviewer prompts for both experimental and comparison group). The full text articles were initially coded for their identifying information and eligibility criteria after which only the eligible articles were coded for study design, participant, intervention group and outcome characteristics.
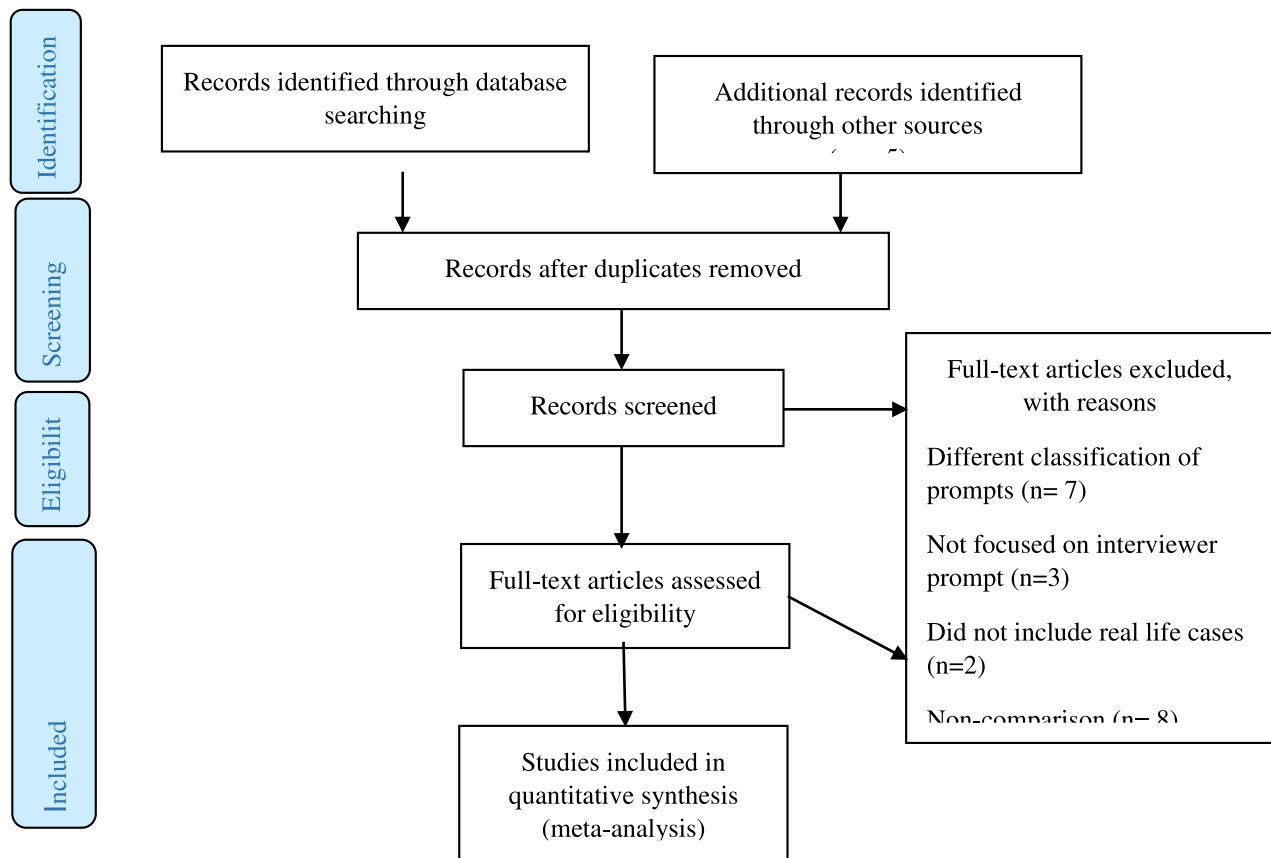

Table 1. *PRISMA Flow diagram for included studies*

Assessment of Risk of Bias

In order to assess the quality and methodological rigorousness of the included individual studies, Risk of Bias was calculated with the help of a modified version of ROBINS-1. It is considered a well-established tool that is both valid and reliable in determining the quality of non-randomised studies (Thode et al., 2021). In comparison to its contemporaries, ROBINS-1 uses an absolute scale approach (Sterne et al., 2016) meaning that rather than assessing whether bias favours a specific outcome (positive or negative), it concentrates on quantifying the overall magnitude or extent of bias without specifying its direction. It also places a study on a common, comparable metric as that of an RCT (Schünemann et al., 2019), enabling in a more consistent and reliable quality assessment approach across different research designs. This was used as the eligible studies selected after the full-text screening process were identified to be non-RCTs. We added an additional domain to titled "other sources of bias" relating to possible biases associated with article funding source, allegiance of the author and authors' possible conflict of interests. Studies were coded as either "low", "moderate", "high" or "unclear" for biases risk on each domain based on the criteria outlined by ROBINS-I tool.

Data Analysis Strategy

The Meta-Analysis was conducted using the random effects model as it accounts the random, study level differences of the included studies pertaining to their research designs, settings, sample size, data collection methodologies etc. This model also aids in a more balanced distribution of weights assigned to each study in order to ensure that the study's significance is not overly impacted by its sample size (Bornstein et al., 2010). Using this model is also consistent with past literature that suggest that variability in researches looking at child maltreatment and exposure may stem from variances in methodological approaches and operational definitions commonly employed in this research domain (Azzopardi et al., 2019).

Statistical analyses in were conducted in R using the Meta-Analysis packages "metafor", "meta", and "dmetar". This included summary estimates of effect sizes, assessment of heterogeneity, sensitivity analysis, and publication bias analysis. This study provides summary estimates of effect sizes using Hedge's g (using means and standard deviations; see Table 1) to quantify the difference between the intervention and comparison group in the included studies as it is preferred for studies with small sample size. Effect sizes were calculated at 95% confidence interval. Forest plots were also constructed to provide a visual summary and display study level effect sizes.

Additionally, variability between the outcome of the included studies, heterogeneity (variability), was computed through: Q, I^2 and Tau statistics. Assessing heterogeneity helps us gauge how diverse the data set of the included studies is and hence, how reliable the pooled summary measure might be. Q statistics assesses for the presence of heterogeneity, I^2 quantifies the variation because of heterogeneity and tau statistics provides information regarding the variability contributed by the true effect sizes of research, not just due to chance. Higher values across these statistics indicate higher levels of heterogeneity. Guidelines provided by Higgins and Green (2011) were used to report I^2 statistics and are as follows: 0%= variability because of chance, 25%=low variability, 50%=moderate variability, and 75%=high variability. In case heterogeneity was detected (p < .05), analysis was performed twice, with and without outlying studies so that no relevant data was excluded from the analysis. This helped reduce the chances of introducing any additional bias. Sensitivity analyses (procedures to assess the impact of independent variable on dependent variables by removing studies from the meta-analysis in a stepwise approach) was

used to identify such studies (e.g., Tawfik et al., 2019).

In addition to this, an assessment of publication bias was carried out through the help of funnel plots, Egger's Test and P-curve (e.g., Horry et al., 2021). These assessments operate on the assumption that bias may arise due to the tendency to publish only statistically significant results. The funnel plot is a visual tool in which the studies are represented by dots across and if these dots assemble in an inverted funnel shape, it indicates absence of publication bias (Peters et al., 2008). Egger's test is a statistical method that assess the presence of publication bias by examining the relationship between the effect sizes and their standard errors. If the p value is less than $p < 0.05$, it suggests the presence of publication bias (Egger et al., 1997). P-curve analysis is another visual and statistical tool that looks at the distribution of p-values of the included studies. If p-values tend to concentrate around 0.05 (the common cut-off for statistical significance), it could suggest the presence of publication bias or selective reporting (Simonsohn et al., 2014a).

Protocol Registration

The protocol has been registered on Open Science Framework and can be accessed through the DOI: 10.17605/OSF.IO/YN5W6.

Results

Study Characteristics

An overview of the characteristics of included studies indicated that 66.3% of the participants (interviewees) were females, whereas males constituted 29.4% of the participants and 4.3% did not specify. The overall age range of the sample ranged between 3-17 years with the mean age being 9.23 years (ranging from 8.4-9.8 years). Fondling was the most reported form of abuse at 40.6%, followed by reports of penetration a t 15% and indecent exposure at 3.3 % respectively. The interviews in the included studies were conducted by either police officers (64.46%) or social workers (43%). Additionally, variation in the reported duration of the training conducted with the interviewers ranged from 2 to 5 days. In regard to comparison/control group interview techniques, majority of the studies used a standard interviewing technique. These were generally defined as a typical method of investigation that the interviewers would use in their everyday practice that were not based on any clearly identifiable guidelines or structured protocols. Four of the included studies compared the NICHD protocol to a semi-structured or structured approach such as Memorandum of Good Practices (Lamb et al., 2009; Price et al.,2016), Tree House Method (Erens et al., 2021) and Dutch Scenario Model (Otgaar et al.,2019).

Risk of Bias

ROBINS-1 tool with an additional factor of "other sources" of bias was used to test for the methodological rigorousness of the included studies in our meta-analysis, showed low risk of bias. Any study without any major flaw which scored "low" in at least six domains was considered to have overall low levels of risk of bias.

Meta Analyses of Interviewer Utterances

Our study included eleven data sets from eleven studies, for interviewers prompts as dependant measure of interviewer quality were coded for both experimental and comparison groups. Since four prompt types are targeted in our research, we divided our systematic review and meta-analysis accordingly.

Table 2

Descriptive statistics for prompts included in meta-analyses

Study	Sample size (n)		Invitation (M, SD)		Suggestions (M, SD)		Directive (M, SD)		Option posing (M, SD)	
	NICHD	Control	NICHD	Control	NICHD	Control	NICHD	Control	NICHD	Control
Cyr & Lamb., 2009	83	83	20.7 (12.40)	6.60 (5.60)	2.90 (3.20)	5.40 (3.60)	13.20 (11.90)	27 (17.10)	9.60 (9.30)	23.10 (17.60)
Cyr et al., 2012	45	45	0.37 (0.17)	0.08 (0.047)	0.05 (0.04)	0.07 (0.05)	0.27 (0.11)	0.37 (0.10)	0.20 (0.10)	0.32 (0.10)
Dion & Cyr, 2008	17	17	19.50 (7.20)	6.20 (4.50)	2.50 (1.40)	6.20 (3.70)	12.60 (7.60)	21.10 (15.30)	9.30 (6)	14.10 (9.70)
Erens et al., 2021	38	30	35 (17)	11 (12)	6 (9)	25 (33)	57 (26)	57 (22)	113 (74)	118 (57)
Orbach et al., 2000	55	50	14.53 (9.90)	4.30 (3.22)	4.62 (5.35)	9.10 (7.59)	26.24 (22.85)	44.96 (29.50)	9.78 (8.06)	28.60 (18.15)
Lamb et al., 2008	50	50	22.72 (11.31)	6.40 (4.27)	3.58 (3.78)	8.24 (7.63)	21.38 (15.45)	41.88 (27.12)	14.44 (12.85)	28 (22.14)
Otgaar et al., 2019	50	49	0.34 (0.16)	0.06 (0.04)	0.06 (0.06)	0.02 (0.02)	0.43 (0.12)	0.76 (0.11)	0.18 (0.08)	0.16 (0.10)
Price et al., 2016	48	46	0.56 (0.19)	0.01 (0.02)	0.03 (0.03)	0.11 (0.09)	0.16 (0.08)	0.41 (0.15)	0.13 (0.09)	0.34 (0.13)
Sternberg et al., 2001	50	50	15 (7.44)	5.56 (3.57)	3.12 (3.27)	6.64 (4.13)	17.54 (11.07)	26.36 (14.46)	12.30 (6.94)	21.56 (12.29)
Sumampouw et al., 2019	50	24	0.34 (0.16)	0.04 (0.05)	0.06 (0.06)	0.03 (0.05)	0.43 (0.12)	0.44 (0.13)	0.18 (0.08)	0.49 (0.14)
Yi et al., 2016	36	36	0.11 (0.11)	0.04 (0.05)	0.19 (0.1)	0.25 (0.09)	0.37 (0.08)	0.39 (0.13)	0.35 (0.26)	0.31 (0.12)

Invitational Prompts

All eleven studies included in our review reported that interviews conducted using the NICHD protocol contained more open-ended invitational prompts as compared to non-protocol interviews. Furthermore, these results in all of the included studies were statistically significant. Six studies examined and reported age-related differences regarding the interviewer prompts (Cyr & Lamb, 2009; Lamb et al., 2008; Orbach et al., 2000; Price et al., 2016; Sternberg et al., 2001) and details relevant to experienced abuse (Cyr & Lamb, 2009; Erens et al., 2021; Lamb et al., 2008; Orbach et al., 2000; Price et al., 2016). Overall, these studies reported that the majority of abuse-relevant details were elicited in response to invitational prompts. Moreover, results of these studies indicated that more invitations were given by the interviewers in the NICHD protocol group regardless of the participant's age as compared to non-protocol group. Also, older children reported more details relevant to the case being investigated as compared to younger children (the cut-off age being 6-7 years for this group) in response to invitational prompts. Results of the meta-analysis showed a combined effect of g= 1.9859 (95% CI=1.47; 2.5), p<.0001. This indicates that there is a large difference between the number of open-ended prompts asked by interviewers in the NICHD protocol and the comparison group, and the protocol group contained significantly more invitational prompts. The effect sizes amongst the studies showed significant heterogeneity (tau^2 =0.5172, p<0.0001; I^2= 84.5, 95% CI= 73.9, 90.8). Due to high heterogeneity, sensitivity analysis was conducted in which study carried out by Price et al. (2016) was identified as an outlier. The removal of this outlier showed a slight reduction in g (Hedge's g= 1.7815; CI=1.4607, 2.1023; p < 0.0001), however, the difference between the two groups was still significantly large. Moreover, this reduced the level of heterogeneity (tau^2 = 0.1446, p= 0.0016) so that 66.2% of heterogeneity could be contributed to between study differences. Egger's test and P-curve analysis did not indicate presence of publication bias.
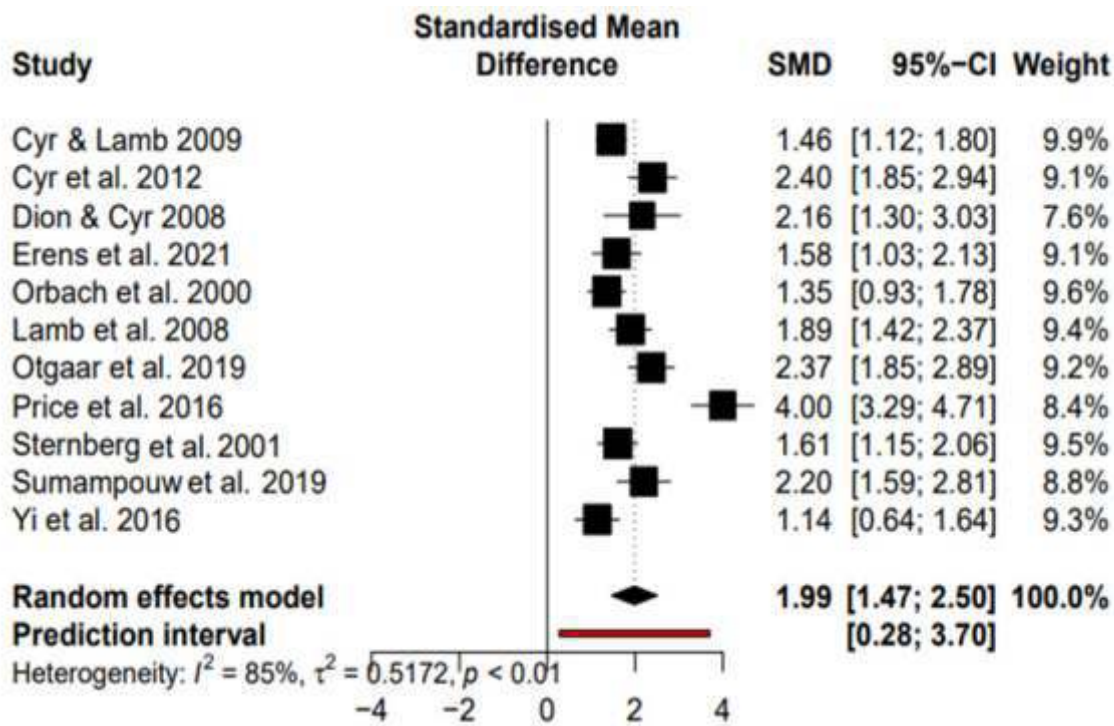

Study	Standardised Mean Difference	SMD	95%-CI	Weight
Cyr & Lamb 2009		1.46	[1.12; 1.80]	9.9%
Cyr et al. 2012		2.40	[1.85; 2.94]	9.1%
Dion & Cyr 2008		2.16	[1.30; 3.03]	7.6%
Erens et al. 2021		1.58	[1.03; 2.13]	9.1%
Orbach et al. 2000		1.35	[0.93; 1.78]	9.6%
Lamb et al. 2008		1.89	[1.42; 2.37]	9.4%
Otgaar et al. 2019		2.37	[1.85; 2.89]	9.2%
Price et al. 2016		4.00	[3.29; 4.71]	8.4%
Sternberg et al. 2001		1.61	[1.15; 2.06]	9.5%
Sumampouw et al. 2019		2.20	[1.59; 2.81]	8.8%
Yi et al. 2016		1.14	[0.64; 1.64]	9.3%
Random effects model		1.99	[1.47; 2.50]	100.0%
Prediction interval			[0.28; 3.70]	

Heterogeneity: I^2 = 85%, τ^2 = 0.5172, p < 0.01

Figure 1. *Meta-Analysis of Invitations*

Suggestive Prompts

All eleven studies reported that the difference between protocol and non-protocol interviews for suggestive prompts were statistically significant. Nine studies reported that investigators in the NICHD protocol group asked fewer suggestive questions as compared to non-protocol group. On the other hand, studies conducted by Otgaar et al. (2019) and Sumampouw et al. (2019) did not support these results. The outcome of meta-analysis indicated the presence of a moderate effect size as indicated by Hedge's g= -0.5516 (95% CI=-0.99; -0.103), p= 0.0207. This indicates that there is moderate difference between the number of suggestive prompts asked by interviewers in the NICHD protocol and the comparison group, and the non-protocol group contained significantly more suggestive prompts. Results showed significant heterogeneity (tau^2=0.3913, p<0.0001; I^2= 88.0, 95% CI= 80.6; 92.6). Sensitivity analysis indicated results from studies conducted by Otgaar et al. (2019) and Sumampouw et al. (2019) as outliers. The removal of these outliers showed an increase in g (Hedge's g=-0.8075, 95%CI=-0.9818; -0.6332, p< 0.0001). This indicates that, in exception of the two excluded studies, the overall results indicate that there are significantly more suggestive prompts in non-protocol interviews and that the difference between the two group is large. This reduced heterogeneity (tau^2 = 0.0287, 95%CI=0.0; 0.161) so that 0% of heterogeneity (I^2= 0.0, 95%CI= 0.0; 64.8) could be contributed to between study differences. Funnel plot and Egger's test indicated absence of asymmetry, and P-curve analysis did not indicate presence of publication bias.
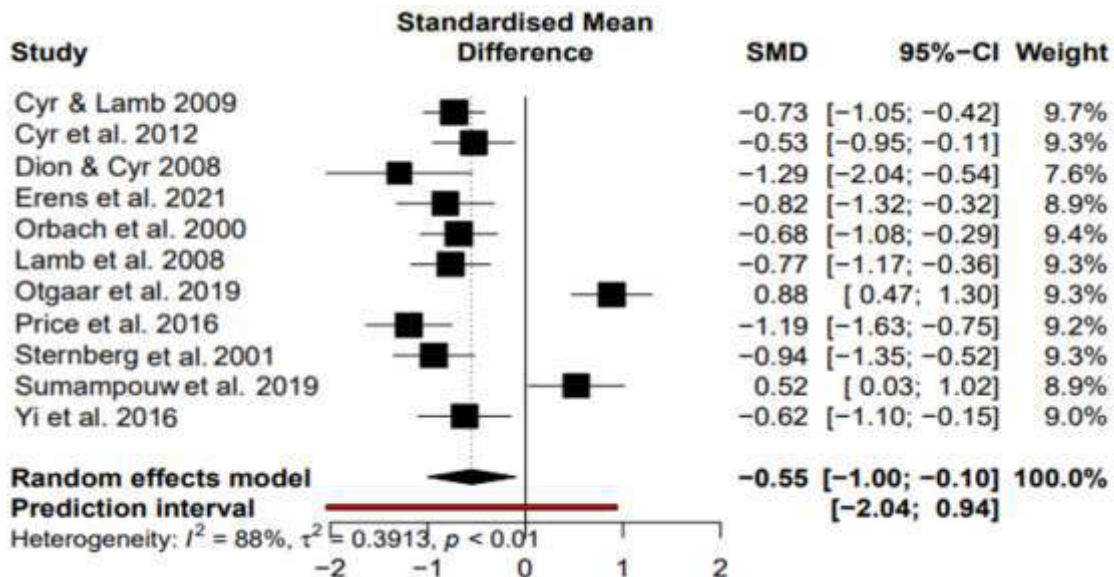

| Study	Standardised Mean Difference	SMD	95%–CI	Weight
Cyr & Lamb 2009		−0.73	[−1.05; −0.42]	9.7%
Cyr et al. 2012		−0.53	[−0.95; −0.11]	9.3%
Dion & Cyr 2008		−1.29	[−2.04; −0.54]	7.6%
Erens et al. 2021		−0.82	[−1.32; −0.32]	8.9%
Orbach et al. 2000		−0.68	[−1.08; −0.29]	9.4%
Lamb et al. 2008		−0.77	[−1.17; −0.36]	9.3%
Otgaar et al. 2019		0.88	[0.47; 1.30]	9.3%
Price et al. 2016		−1.19	[−1.63; −0.75]	9.2%
Sternberg et al. 2001		−0.94	[−1.35; −0.52]	9.3%
Sumampouw et al. 2019		0.52	[0.03; 1.02]	8.9%
Yi et al. 2016		−0.62	[−1.10; −0.15]	9.0%
Random effects model		−0.55	[−1.00; −0.10]	100.0%
Prediction interval			[−2.04; 0.94]	
Heterogeneity: I^2 = 88%, τ^2 = 0.3913, p < 0.01				

Figure 2. *Meta-Analysis of Suggestive*

Directive Prompts

Eight of the included studies reported that non-protocol interviews contained more directive questions as compared to protocol interviews. All of these studies, except for one conducted by Dion and Cyr (2008) reported the results as being statistically significant. On the other hand, three studies conducted by Erens et al. (2021), Sumampouw et al. (2019) and Yi et al. (2016) reported very little difference or non-treatment effect between their participant groups. The combined effect size for Directives was g= -0.9106 (95% CI= -1.48,-0.34), p=.0051. This

indicates that there is a large difference between the number of directive prompts asked by interviewers in the NICHD protocol and the comparison group, and the non-protocol group contained significantly more directive prompts. The effect sizes amongst the studies showed significant heterogeneity (tau^2=0.657, p<0.01; I^2= 90.2 ,95% CI= 84.5, 93.8). Results from Otgaar et al. (2019) and Price et al. (2016) were identified as outliers through sensitivity analysis. The removal of these outliers showed a decrease in g (Hedge's g= -0.58; 95% CI= -0.88, -0.29; P= 0.0017). This reduction indicates that the difference between the groups was influenced by outliers and post their removal, we observe a decrease in the difference from large to moderate levels, albeit it still being statistically significant. Moreover, this reduced the level of heterogeneity (tau^2 = 0.1044, p=0.0034). Egger's test indicated lack of asymmetry and P-curve analysis did not indicate presence of publication bias.
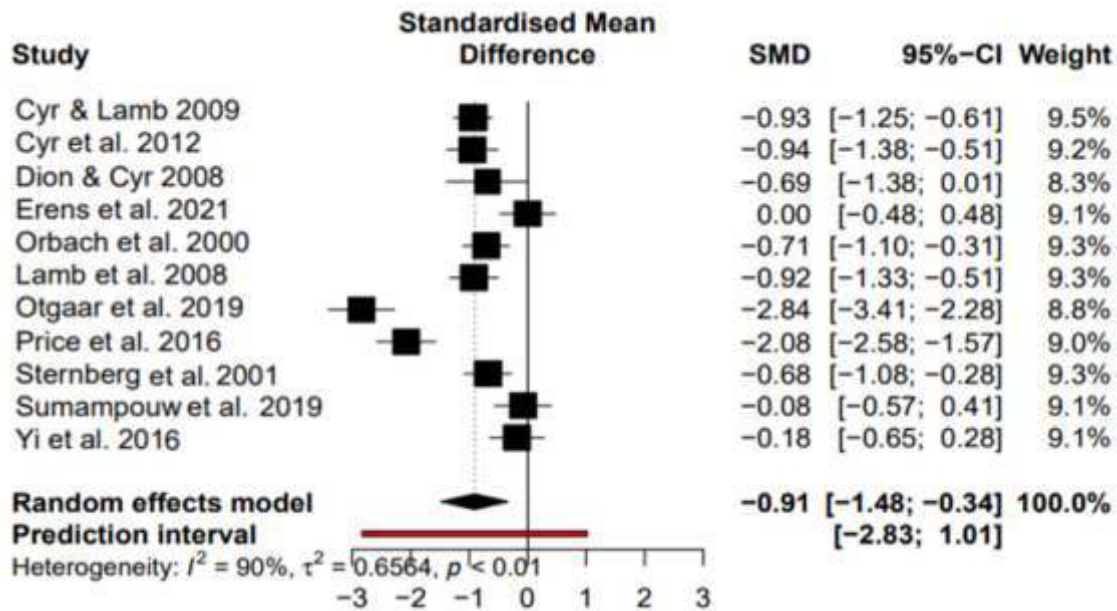
Figure 3. *Meta-Analysis of Directives*

Option Posing Prompts

Eight of the included studies reported that non-protocol interviews contained more option- posing questions as compared to protocol interviews. All of these studies, except for one conducted by Dion and Cyr (2008) reported statistically significant results. The greatest difference can be observed in the study conducted by Sumampouw et al. (2019). On the other hand, three studies conducted by Erens et al. (2021), Otgaar et al. (2019) and Yi et al. (2016) reported very little difference or non-treatment effect between their experimental and comparison groups. The mean of effect sizes was g= -0.9178 (95% CI:-1.54; -0.29, p= 0.0081. This indicates that there is a large difference between the number of option-posing prompts asked by interviewers in the NICHD protocol and the comparison group, and the comparison group contained significantly more option-posing prompts. High variability amongst studies was indicated (tau^2=0.7937, p<0.0001, I^2= 91.6. 95%CI= 87.0; 94.6). Results from Otgaar et al. (2019), Sumampouw et al. (2019), and Yi et al. (2016) were identified as outliers. The removal of these outliers showed a slight increase in g (Hedge's g=-0.9687 ; CI=-1.41, -0.53; p= 0.0013), still indicating large effect. This indicates that the outliers did not heavily influence the difference between

protocol and non-protocol interviews for option-posing prompts. Moreover, this led to a reduction in the level of heterogeneity (tau^2= 0.2298, 95%CI= 0.06, 1.13) so that 78.8% of heterogeneity could be contributed to between study differences. Egger's test and P-curve analysis did not indicate presence of publication bias.
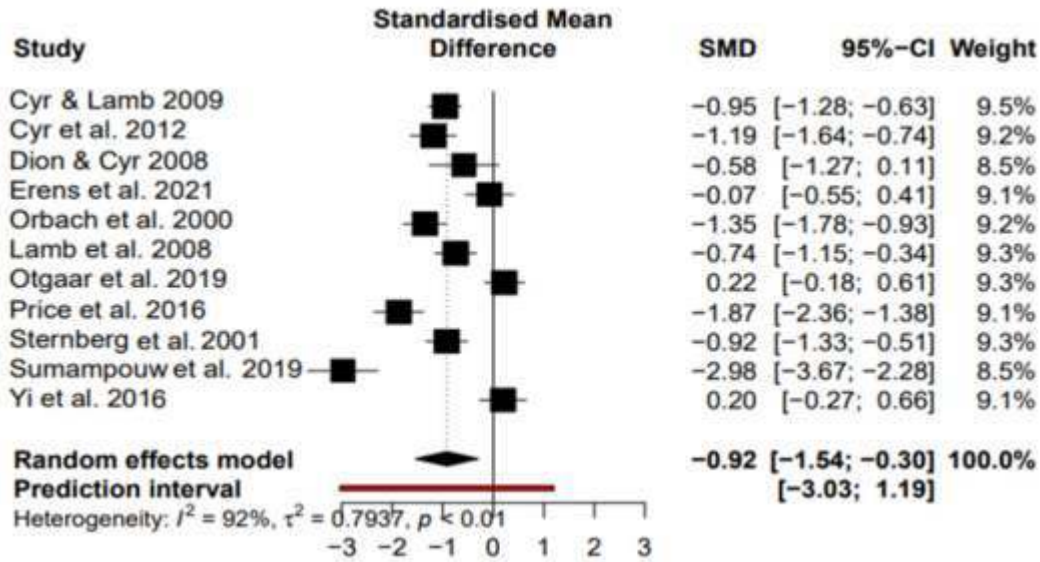

Figure 4. *Meta-Analysis of Option Posing*

Discussion

The purpose of this research was to synthesise and amalgamate the effect sizes regarding the impact of the NICHD protocol in comparison to other interviewing techniques in promoting the usage of open ended questions and reducing the usage of close-ended questions. This meta-analyses showed that interviewers who used the NICHD protocol asked more invitational/open ended questions and less directive, option-posing and suggestive questions in comparison to non-protocol interviews. However, the effect sizes driven from the results of each included individual study, for each prompt type show large effects and corroborates with the trends indicated by combined effect sizes.

For suggestive prompts, two out of eleven studies (Otgaar et al., 2019; Sumampouw et al., 2019) indicate that protocol interviews yield more suggestive prompts as compared to non-protocol interviews. These results are not supported by previous literature. Upon further inspection to understand this heterogeneity, the following points were identified. Otgaar et al. (2019)

proposed that this disparity may be due to the "philosophical underpinnings" of the comparison protocol (The Scenario model), encouraging greater use of directive prompts in comparison to other kinds of prompts (including suggestive line of questioning). Additionally, the Scenario protocol in also built around "best practice guidelines" that discourage the use of suggestive questioning. Sumampouw et al. (2019) posits that the standard practices of Indonesian police officers place a heavier reliance on using directive and option-posing questions to gain disclosure. This may limit the opportunity of asking suggestive questions. The two authors suggest that, while the proportionally low number of suggestive prompts in the comparison groups is a promising feature of those techniques, it cannot be considered as an indicator of them being good quality practices. This is because they place heavier reliance on directive and option-posing prompts, both of which are not in line with best practice guidelines

Overall, our results corroborate with the findings of previous reviews conducted by Benia

et al. (2015) and Lamb et al. (2007). These studies report on NICHD protocol's effectiveness in using questioning techniques that are in line with the well-accepted, best practice guidelines in the field of forensic interviewing. This is an important finding as, in addition to other factors, previous literature has established a strong relationship between the interviewers asking open-ended question and its facilitation with children reporting detailed, accurate descriptions of sensitive experiences (Lavoie et al., 2021). More closed questions reduce the likelihood of obtaining elaborate accounts (Erens et al., 2021). Similar results have often been replicated in laboratory settings, placing importance of obtaining "free recall narratives" elicited by invitational prompts (e.g., Goodman et al., 1991; Hutcheson et al., 1995). Additionally, prioritizing open-ended questioning is also linked with increase in supportive comments given to the child witness; thus improving rapport, encouraging the child to be more cooperative and increasing chances of accurate disclosure (Hershkowitz et al., 2006). As a result, the relative quantity of each type of question and prompt used can be an indicator to assess the overall quality of forensic interviews investigating child abuse. Therefore, based on our results we can conclude with some certainty that the NICHD protocol produces better quality investigative interviews as compared to other standardized and non-standardized protocols.

The current study was successful in bridging certain literature gaps. Firstly, the review synthesized the results of the available literature by subjecting them to statistical analysis, which was done in only one more research conducted by Benia et al. (2015). Therefore, this study provides corroboration to the present consensus regarding the impact of the NICHD protocol. Secondly, while the total number of studies may still be considered low evidence base, this meta-analysis included 6 additional studies in comparison to the previous meta-analysis. Furthermore, these additional articles provided diversity in the data pool in terms of geographical locations in which the studies were conducted (Erens et al., 2021; Sumampouw et al., 2019; Yi et al., 2016) and comparison interview protocols (Erens et al., 2021; Lamb et al., 2008; Otgaar et al., 2019; Price et al., 2016). An important finding was that invitational prompts were unanimously greater in number in all protocol interviews as compared to non-protocol regardless of geographical or comparison protocol. This diversity, particularly in the context of comparison interview protocols does to an extent fill the gap identified in the previous meta-analytical review. Lastly, the current review assessed the quality of the evidence included by conducting risk of bias along with publication bias assessment. This has helped us establish that the results of this meta- analysis may need to be considered with caution as several studies indicated possible risk of bias in various fields, particularly in the area related to "other source of bias" due to the involvement of the developers of the protocol in a given study. Additionally, conducting the publication bias analysis, particularly p-curve analysis, was useful as it showed p-hacking was not evident in evidence base. negating the likelihood of selective reporting in published literature (Gadbury & Allison, 2012).

Limitations

While attempts were made to ensure the rigorousness of this review, certain limitations in the current review are present. The major limitation of this study is that we found substantial heterogeneity within outcome variables even in post-outlier removal results except for suggestive prompts. This poses challenges for the interpretation of our results. Summary effect sizes may not be entirely reliable as significant heterogeneity implies that the NICHD protocol may not be effective in a consistent manner across its application in various situations. This may impact the results of publication bias analysis, particularly of p-curve analysis as heterogeneity leads to the overestimation of true effect sizes (Van Aert et al., 2016). Moreover, due to the inclusion of small number of studies that met our

inclusion criteria, a large evidence base could not be generated to gather extensive information regarding the effectiveness of the NICHD protocol in facilitating the use of prompts in accordance with the best practice guidelines. While several other article that studied the NICHD protocol were identified, they did not focus on evaluating the protocol in comparison to a control group but focused on evaluating certain components of testimonies given by children.

Implications and Future Directions

The results of this review have implications for policymakers, child protective services, social services, police departments, and forensic and clinical psychologists amongst other fields that are involved in the investigation of child abuse cases. It is important to note evidence (e.g. Lavoie et al., 2021) supports that the use of open ended questions positively relates to the disclosure of sensitive information by child witnesses. However, in regards to future researches, there are certain limitations that should be addressed. Moderator analysis should be conducted to identify possible factors that may impact the consistency of the application of the NICHD interview protocol in improving quality of investigation. It is also important to study the effectiveness of invitational prompts for different victim population be that pertaining age (pre-school children who may not have developed conversational skills or may have limited vocabulary to answer open ended questions), personality characteristics (e.g. children who are reluctant to trust and open up), and geographical region (e.g. how children from different culture interact with a person of authority) etc. In expansion of this, it is important to study how invitational questions can be tweaked and what flexibility protocols such as that established by the NICHD allow for such tweaks so that the practices can be best suited to the witness in question.

In regards to Pakistan, to the best of our knowledge, there is presently no evidence backed structured guidelines/protocol that has been implemented by policy makers at national level

regarding forensic investigation child abuse cases (Malik, 2012). Moreover, there is a scarcity in indigenous literature that can potentially encourage and guide such decision making. Considering child abuse is a persistent and rising concern in Pakistan (Mehnaz, 2018), it is important that the gap pertaining to the areas of detection, intervention and investigation of child abuse be catered to. Our results indicate the NICHD protocol's effective in improving interview quality is also generalizable to non-western, Asian countries and hence, can possibly be adapted within the context of Pakistan.

Declarations
Data sharing. The article is based on already published work, so data sharing is not applicable for this article. However, extracted data is provided in Table 2 of the article.
Funding. No funding was received for this review.
Human and animal rights. This article does not include data collection by any of the authors. However, ethical review was done by the departmental ethical committee.
Conflict of interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

References
Anderson, G. D., Anderson, J. N., & Gilgun, J. F. (2014). The influence of narrative practice techniques on child behaviors in forensic interviews. *Journal of Child Sexual Abuse*, 23(6), 615-634. htt ps://doi. org/10.1080/10538712.20 14.93288
Anderson, J. (2013). The Corner House forensic interview protocol: An evolution in practice for almost 25 years. *APSAC Advisor*, 4, 2-7. https: //shorturl.at/hjqU6
Azzopardi, C., Eirich, R., Rash, C. L., MacDonald, S., & Madigan, S. (2019). A meta-analysis of the prevalence of child sexual abuse

disclosure in forensic settings . *Child Abuse & Neglect* , *93*, 291-304. https://doi.org/10.1016/j.chiabu.2018.11.020

Baugerud, G. A., & Johnson, M. S. (2017). The NICHD Protocol: Guide to Follow Recommended Investigative Interview Practices at the Barnahus ?. *Collaborating against Child Abuse: Exploring the Nordic Barnahus model*, 121-143. https://doi.org/10.10 07/978-3-319-58388-4 _6

Benia, L. R., Hauck-Filho, N., Dillenburg, M., & Stein, L. M. (2015). The NICHD investigative interview protocol: A meta-analytic review. *Journal of Child Sexual Abuse*, 24, 259–279. https://doi.org/10.1080/105 38712.2015.1006749

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97-111.https://doi.org/ 10.1002/jrsm.12

Chae, Y., Goodman, G. S., Larson, R. P., Augusti, E. M., Alley, D., VanMeenen, K. M., ... & Coulter, K. P. (2014). Children's memory and suggestibility about a distressing event: The role of children's and parents' attachment . *Journal of Experimental Child Psychology* , *123*, 90-111. https://doi.org/10.1016/j.jecp .2014.01.005

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 &* *Neglect*, *33*(5), 257-268. https://doi. org/10.1016/j.chiabu.2008.04.002

Cyr, M., Dion, J., McDuff, P., & Trotier-Sylvain, K. (2012). Transfer of skills in the context of non-suggestive investigative interviews: Impact of structured interview protocol and feedback. *Applied Cognitive Psychology*, 26(4),516-524.https://doi .org/10.1002/acp. 2822

Dion, J., & Cyr, M. (2008). The use of the NICHD protocol to enhance the quantity of details obtained from children with low verbal abilities in investigative interviews: A pilot study. *Journal of child sexual abuse*, *17*(2), 144-162. https://doi.org/ 10.1080/10538710801916564

Egger, M., Zellweger-Zahner, T., Schneider, M., Junker, C., Lengeler, C., & Antes, G. (1997). Language bias in randomized controlled trials published in English and German. *The Lancet*, *350*, 326–329.https://doi.org/10.1016/ S0140-6736(97)02419-7

Erens, B., Otgaar, H., de Ruiter, C., van Bragt, D., & Hershkowitz, I. (2021). The NICHD interview protocol used by Dutch child protection workers: Effects on interview style, children's reported information and susceptibility to suggestion. *Applied Cognitive Psychology*, 36(1),7-18. https://doi.org/10.1002/acp.3893

Fernandes, D., Gomes, J. P., Albuquerque, P. B., & Matos, M. (2024). Forensic interview techniques in child sexual abuse cases: a scoping review. *Trauma, Violence, & Abuse*, 25(2), 1382-1396. https://doi .org/10.1177/15248380231177317

Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The*

cognitive interview. Charles Thomas Publishers. https://psycnet.apa.org/record/1992-98595-000

Fisher, R. P., & Geiselman, R.E. (2017). Investigative interviewing. In V. Hasselt & M. Bourke (Eds.), *Handbook of Behavioral Criminology*. https://doi.org/10.1007/978-3-319-61625-4_2 6

Gadbury, G. L., & Allison, D. B. (2012). Inappropriate fiddling with statistical analyses to obtain a desirable p-value: tests to detect its presence in published literature. https://doi.org/10.1371/journal.pone.0046363

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B. M., & Rudy, L. (1991). Children's testimony about a stressful event: Improving children's reports. *Journal of Narrative and life history*, *1*(1), 69-99. https://doi.org/10.1075/jnlh.1.1. 05chi

Goodman, G. S., Jones, O., & McLeod, C. (2017). Is there consensus about children's memory and suggestibility?. *Journal of interpersonal violence*, *32*(6), 926-939. https://doi.org/10.1177/0886260516657358

Hailes, H. P., Yu, R., Danese, A., & Fazel, S. (2019). Long-term outcomes of childhood sexual abuse: an umbrella review. *The Lancet Psychiatry*, *6*(10), 830-839. https://doi.org/10.1016/S2215-0366(19) 30 286-X

Herman, S. (2009). Forensic child sexual abuse evaluations: Accuracy, ethics, and admissibility. In K. Kuehnle & M. Connell (Eds.), *The evaluation of child sexual abuse allegations: A comprehensive guide to assessment* (pp. 247–266). Hoboken, NJ: Wiley http://dx.doi.org/10.13140/2.1.3235.4082

Hershkowitz, I., Orbach, Y., Lamb, M. E., Sternberg, K. J., & Horowitz, D. (2006). Dynamics of forensic interviews with suspected abuse victims who do not disclose abuse. *Child Abuse & Neglect*, *30*(7),753-769.https://doi.org/10.1016/j.chiabu.2005.10.016

Higgins, JPT., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions*. The Cochrane Collaboration. https://handbook-5-1.cochrane.org/

Home Office. (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. HM Stationery Office. https://shorturl .at/bqCI1

Horry, R., Hughes, C., Sharma, A., Gabbert, F., & Hope, L. (2021). A meta-analytic review of the Self-Administered Interview©: Quantity and accuracy of details reported on initial and subsequent retrieval attempts. *Applied Cognitive Psychology*, *35*(2), 428-444 https://doi.org/10.1002/acp.3753

Hutcheson, G. D., Baxter, J. S., Telfer, K., & Warden, D. (1995). Child witness statement quality. *Law and Human Behavior*, *19*(6), 631-648. https://doi.org/10.1007/BF01499378

Kuehnle, K., & Connell, M. (2012). Child sexual abuse evaluations. In R. K. Otto & I. B. Weiner (Eds.), *Handbook of psychology* (Vol. 11, 2nd ed., pp. 579–613). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/978 1118133880.hop211024

Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: Finding water, leading horses, but can we get them to drink ?. *American*

Psychologist, *71*(8), 710. https:// doi. org/10.1037/amp 0000 039

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. http://ndl.ethernet.edu.et/bitstream/12 3456789/14675/1/125.pdf

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P. W., & Horowitz, D. (2007). A structured forensic interview protocol improves 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. https://doi. org/10.1016/j.chiabu.2007. 03.021

Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J. A. N., 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: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(4), 449-467. https://doi.org/10.1002/acp.14 89

Lavoie, J., Wyman, J., Crossman, A. M., & Talwar, V. (2021). Meta-analysis of the effects of two interviewing practices on children's disclosures of sensitive information: Rapport practices and question type. *Child Abuse & Neglect*, *113*, 104930. https://doi.org/10.1016/j.chiabu.2021. 104930

Malik, M. A. (2012). A Working Paper on Tackling Child Sexual Abuse The country need a specialized Child Sexual Abuse Unit within the police department. https://cgss.com.pk/ paper /pdf/Tackling-Child-Sex-Abuse.pdf

McElvaney, R. (2015). Disclosure of child sexual abuse: Delays, non-disclosure and partial disclosure. What the research tells us and implications for practice. *Child Abuse Review*, *24*(3), 159-169. https://doi.org/10.1002/car. 2280

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia medica*, *22*(3), 276-282. https://hrcak. srce.hr/8939 5

Mehnaz, A. (2018). Child abuse in Pakistan-Current perspective. *National Journal of Health Sciences*, 3(4), 114-117.https://ojs. njhsciences.com/ index .php/njhs/ article/view/67

Meissner, C. A., Surmon-Böhr, F., Oleszkiewicz, S., & Alison, L. J. (20 17). Developing an evidence-based perspective on interrogation: A review of the US government's high-value detainee interrogation group research program. *Psychology, Public Policy, and Law*, *23* (4),438.https://doi. org/10.1037/law 0000136

Myklebust, T., La Rooy, D. J., & Peixtoto, C. E. (2023). The National Institute of Child Health and Human Development Protocol. *Interviewing and interrogation: a review of research and practice since World War II*, 367-387. https://repositorio. ucp.pt/bitstrea m/10400.14/42994/1/88019939.pdf

Orbach, Y., & Pipe, M. E. (2011). Investigating substantive issues. *Children's testimony: A handbook of psychological research and forensic practice*, 147-164. https://doi.org/10.1002/97811199984 95.ch8

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 & Neglect*, 24, 733–752. https://doi.org/10.1016 /S0145-2134 (00)00137 -X

Otgaar, H., Howe, M. L., Merckelbach, H., & Muris, P. (2018). Who is the better eyewitness? Sometimes adults but at other times children. *Current Directions in Psychological Science*, 27, 378–385.https://doi.org/10.11 77/0963 721418770998

Otgaar, H., La Rooy, D., Horselenberg, R., Hershkowitz, I., de Ruiter, C., Blezer, L., ... & Kollau, R. (2019). Assessing the quality of child investigative interviewing in the Netherlands. *Applied Cognitive Psychology*, *33*(5), 889-897. https:// doi.org/10.1002/acp.3521

Pears, S., & Sutton, S. (2021). Effectiveness of Acceptance and Commitment Therapy (ACT) interventions for promoting physical activity: a systematic review and meta-analysis. *Health psychology review*, *15*(1), 159-184. https://doi. org/10.1080/17437199.2020.1727759

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, *61*(10), 991–996. https://doi.org/10.1016/j.jclinepi. 2007 .11.010

Pipe, M. E., Orbach, Y., Lamb, M. E., Abbott, C. B., & Stewart, H. (2013). Do case outcomes change when investigative interviewing practices change? *Psychology, Public Policy,*

and Law, 19, 179–190. https://doi. org/10.1037/a0030 312

Powell, M. B., Wright, R., & Clark, S. (2010). Improving the competency of police officers in conducting investigative interviews with children. *Police Practice & Research: An International Journal, 11*(3), 211–226. https://doi. org/10.1080/1561426090283070

Price, E. A., Ahern, E. C., & Lamb, M. E. (2016). Rapport-building in investigative interviews of alleged child sexual abuse victims. *Applied Cognitive Psychology, 30*(5), 743-749. https://doi.org/10.1002/acp.32 49

Rispens, I., & van der Sleen, J. (2017). Interviewing witnesses in the Netherlands. In *International developments and practices in investigative interviewing and interrogation* (pp. 210-223). Routledge. https://doi.org/10.4324/97 81315769684

Saywitz, K. J., Lyon, T. D., & Goodman, G. S. (2011). Interviewing children. U: Myers, JEB (ur.) *The APSAC Handbook on Child Maltreatment, 3d ed.*,(str. 337-360) Los Angeles. https://www.nationalcac.org/wp-content/uploads/2016/10/Interviewin g-childen.pdf

Schünemann, H. J., Cuello, C., Akl, E. A., Mustafa, R. A., Meerpohl, J. J., Thayer, K., Morgan, R. L., Gartlehner, G., Kunz, R., Katikireddi, S. V., Sterne, J., Higgins, J. P., Guyatt, G., & GRADE Working Group (2019). GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. *Journal of Clinical Epidemiology, 111*, 105–114. https:

//doi.org/10.1016/j.jclinepi. 2018.01. 012

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534–547. https://doi.org/10.1037/a0033 242

Smethurst, A. (2023). *Scaffolding Children's Communication in Investigative Interviews* (Doctoral dissertation, Teesside University). https://shorturl.at/aklR3

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*(5), 997. https://doi.org/10.1037/0021-9010. 86.5.997

Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., ... & Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *bmj*, 355. https://doi.org/10. 1136 /bmj.i4919

Sumampouw, N. E., Otgaar, H., La Rooy, D., & De Ruiter, C. (2019). The quality of forensic child interviewing in child sexual abuse cases in Indonesia. *Journal of Police and Criminal Psychology, 35*(2), 170-181. https:// doi.org/10.1007/s11896-019-0 9342-5

Tawfik, G. M., Dila, K. A. S., Mohamed, M. Y. F., Tam, D. N. H., Kien, N. D., Ahmed, A. M., & Huy, N. T. (2019). A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Tropical Medicine and Health, 47*, 1-9.

https://doi.org /10.1186/s41182-019-0 165-6

Thode, R., Solanki, G. S., Aggarwal, A., Belekar, V., Lakkakula, U., & Goyal, R. (2021). PND54 A Comparison of NOS and ROBINS-I Tools for Quality Assessment of Observational Studies. *Value in Health, 24*, S169. https://doi.org/10. 1016/j.jval.2021. 04.838

van Aert, R. C., Wicherts, J. M., & van Assen, M. A. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*(5), 713-729. https://doi.org/10.1177/174569 16166 50874

Vrij, A. (2018). Verbal lie detection tools from an applied perspective. In P. J. Rosenfeld (Ed.), *Detecting Concealed Information and Deception: Recent Developments* (pp. 297–327). Elsevier Academic Press. https:// doi.org/10.1016 /B978-0-12-81272 9-2.00013-6

Wilson, C., & Powell, M. (2012). *A guide to interviewing children: Essential skills for counsellors, police lawyers and social workers*. Routledge. https://doi.org/10.4324 /9780203390 023

Yi, M., Jo, E., & Lamb, M. E. (2016). Effects of the NICHD protocol training on child investigative interview quality in Korean police officers. *Journal of Police and Criminal Psychology*, 31, 155–163. https://doi.org/10.1007/s11 896-015-9170-9

Yuille, J. C., Hunter, R., Joffe, R., & Zaparnuik, J. (1993). Interviewing children in sexual abuse cases. In G. S. Goodman & B. L. Bottoms (Eds.), *Child Victims, child witnesses:*

Understanding and improving testimony (pp. 95-115). Guildford Press. https://psycnet.apa.org/record/1993-97464-005