The Police Use of Background Information Related to Alleged Victims in Mock Evaluations of Child Sexual Abuse

Alessandro Tadei¹ · Pekka Santtila² · Jan Antfolk¹

Published online: 22 January 2019
© The Author(s) 2019

Abstract

When statistically related to child sexual abuse (CSA), background information can assist decision-making in investigations of CSA allegations. Here, we studied the use of such background information among Finnish police officers. We analyzed their ability to identify and interpret CSA-related and CSA-unrelated background information both when placed in mock scenarios and when presented as separate, individual variables. We also measured the ability to correctly estimate the probability of CSA based on such background information. In the context of mock scenarios, officers were better in discarding CSA-unrelated variables than in identifying CSA-related ones. Within-subject performance across different scenarios was, however, not consistent. When information was presented as separate variables, officers tended to incorrectly consider many CSA-unrelated variables as CSA-related. Officers performed better in recognizing whether actual CSA-related variables increase or decrease the probability of CSA. Finally, officers were inaccurate in identifying variables that are CSA-related only for boys or only for girls. When asked to estimate the CSA probability of mock scenarios, participants were accurate only in assessing low-probability cases, and this was not associated with the ability to identify CSA-related and CSA-unrelated variables. We conclude that police officers would benefit from more training in using background information and from using available decision-making support tools in the context of investigating CSA allegations.

Keywords Child sexual abuse · Police investigation · Background information · Probability estimation · Evidence

Introduction

The decision to press charges in child sexual abuse (CSA) cases depends on the quality and quantity of the available

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11896-019-9312-6>) contains supplementary material, which is available to authorized users.

✉ Alessandro Tadei
atadei@abo.fi

Pekka Santtila
pekka.santtila@nyu.edu

Jan Antfolk
jantfolk@abo.fi

¹ Department of Psychology, Åbo Akademi University, Fabriksgatan 2, 20500 Turku, Finland

² Faculty of Arts and Sciences, New York University (Shanghai), Shanghai, China

evidence. Herman (2010) divided evidence in forensic CSA evaluations into two groups: psychosocial and non-psychosocial. Non-psychosocial evidence, also referred to as strong evidence, includes medical findings, confessions, or videos and photographs. Psychosocial evidence, often considered soft evidence, can, for example, be children's reports and behaviors, or situations that affect the likelihood of the allegation being true (e.g., parental alienation syndrome; Gardner 1998; see also O'Donohue et al. 2016). The presence of non-psychosocial evidence is usually enough to press charges, while psychosocial evidence tends to be considered less persuasive (Herman 2010; Myers 2005). For example, Walsh et al. (2010) have identified four types of evidence that, if present, increase the probability that CSA allegations are prosecuted: a disclosure by the victim, a corroborating witness, a confession by the suspect, or a second report against the suspect. In the absence of stronger evidence, the presence of a corroborating witness was the type of evidence that most

increased the probability that charges would be filed. Herman (2010) demonstrated that only 36% of 894 CSA cases included corroborative non-psychosocial evidence.¹ In cases considered true by mental health and medical professionals, this rose to 54%, suggesting that CSA allegations are often prosecuted based on non-corroborative evidence. It is therefore important that professionals involved in investigations and trials are able to identify and correctly assess the role and importance of psychosocial pieces of evidence in CSA cases.

Statistical Use of Background Information

Traditionally, for psychosocial and non-psychosocial evidence to be considered as important during investigations and trial, they both have to be directly related to the CSA allegations. Videos and medical symptoms, as well as children's statements and non-verbal behaviors, need to be robustly linked to the CSA event to be admitted as evidence in court. There is a third type of information that we here call background information, which can potentially play an important role during an investigation, even if it is not directly linked to the CSA allegation and is not admissible as evidence. Examples of such background information include previous experiences of the alleged victim with other crimes, personal relations with relatives and peers that are not involved in the CSA allegation, or the alleged victim's life style, such as smoking or drinking.

In statistical terms, whether an alleged case of CSA is likely to be true or not can be considered a classification problem. In other words, the allegation needs to be assigned to the right class, that is, a likely true or a likely false allegation. A number of classification techniques are available (e.g., logistic regression, discriminant analysis, K-nearest neighbors, naïve Bayes, classification trees, and others; James et al. 2013), and they have been made accessible to non-statisticians through easy-to-use statistical packages (see, for example, Koul et al. 2017). Seen from a decision-making point of view, background information can be used to provide evidence that can justify prosecution, by estimating the probability of an allegation being true. Using a naïve Bayes classifier, Tadei et al. (2017) identified 42 pieces of background information that predicted the CSA probability with a high level of accuracy for both female and male alleged victims ($AUC = 0.88$ for girls and 0.97 for boys, based on an equal folds cross validation procedure, repeated 100 times). This classifier, called FICSA (Finnish Investigative Instrument of Child Sexual Abuse), was trained on data from more than 11,000 children, either

12 or 15 years old, and 903 background variables (Ellonen et al. 2013). The class variable, CSA, was defined as any sexual act involving a minor and a person 5 years older. More specifically, ten different sexual acts, ranging from a proposal to do anything sexual to actual penetration, were considered for CSA. Based on these definitions, the CSA base-rate in the sample used to calibrate the FICSA was 3% for girls, and 0.7% for boys. To calculate the likelihood of CSA, FICSA needs that part, if not all, of the 42 pieces of background information identified as statistically related to the CSA risk (25 specific for girls, 14 for boys, 3 common) are collected. For example, FICSA might ask if the child has ever tried to smoke, or ever has been robbed. Being a naïve Bayes classifier, FICSA can successfully integrate several pieces of evidence, also when gathered at different times, and the more questions are answered, the more accurate CSA probability estimation will be. Although these pieces of background information required by FICSA do not constitute actual evidence of CSA and would be likely to be non-admissible during a trial in many jurisdictions, the information could still increase the number of CSA allegations that are correctly solved. The authors suggested that knowing the correct baseline probability of a case, when psychosocial or background information related to it is considered, could help investigators prioritize among open cases to optimize the use of investigative resources (Tadei et al. 2018). This could reduce the risk that important pieces of admissible evidence that degrade with time (e.g., physical symptoms, witnesses, child's memories) are invalidated in cases otherwise likely to be prosecuted (Adams 2011; Antfolk et al. n.d.; Heger et al. 2002; Pipe et al. 2004).

When applied to forensic evaluations, it has been demonstrated, however, that not only the creation of a classifier (Tadei et al. 2017; Zawoad and Hasan 2015) but also the interpretation of its results (Fenton and Neil 2012; Koehler 2011; Tadei et al. 2018) can be problematic. When in contrast with their own opinion, expert witnesses and legal professionals tend not to consider the results provided by a statistical tool built on actuarial data (Tadei et al. 2018). This might affect the correctness of the conclusions drawn by forensic professionals, since it has been demonstrated that various cognitive biases might affect decision-making and accuracy also when dealing with statistics (e.g., defense counsel's fallacy, prosecutor's fallacy; Fenton and Neil 2012; Henderson 2002). In sum, background information and classification methods have potential to improve the accuracy of the decisions made by legal professionals, but available studies indicate that more training is needed.

Background Information in CSA Research and Investigations

There is a lack of research on the use police make of background information about alleged CSA victims. The few

¹ As in Herman (2010), we will use the term corroborative evidence to indicate non-psychosocial corroborative evidence only, instead that any type of evidence that might support the allegation.

studies that have considered also background information have generally limited this to include only gender, ethnicity, and/or age of the child (Stroud et al. 2000; Walsh et al. 2010). Interestingly, Stroud et al. (2000) observed how all these three pieces of information, even if not evidence, were statistically related to the probability of a case being referred for criminal prosecution. However, the authors did not specify if it is correct to assume that a particular age range, gender, or ethnicity increases the probability of the alleged abuse being true. In the case this assumption was wrong, the connections between these three variables and CSA could be considered CSA myths. Cromer and Goldsmith (2010) defined CSA myths as false beliefs that may “reflect circumscribed features of perpetrators and victims [and] cause harm through diminishing awareness for CSA and the allocation of resources to prevent CSA and help victims, and/or dissuade victims from disclosing abuse” (p. 619). To our best knowledge, research on CSA myths have not explored if these are common, for example, among police officers. The lack of representativeness of background information is visible in CSA myths research as well. Cromer and Goldsmith (2010) gathered and categorized 119 CSA myths, but, when discussing the beliefs about victim characteristics, they only commented on gender. As far as concerns police officers, they are also subject to cognitive biases and myths as everyone else (Aamodt 2008; Saunders 2012). Thus, we can expect that, because of the absence of research on CSA background information, they can hold false beliefs about the relation between these variables and CSA risk.

The Present Study

In the present study, we examined how police use background, non-corroborative, information in evaluations of CSA cases. We also tested if the use of such information was affected by it being contextualized in a scenario, as compared to when considered as individual pieces of information. Finally, we tested if being able to correctly consider all available pieces of valid background information, and correctly exclude information that is not valid, led to more accurate conclusions about the CSA risk probability.

Because of the lack of research about the connection between CSA and background information, we expected police officers not to be able to distinguish a CSA-related piece of information from an unrelated one, both when inserted in scenarios and when presented as individual pieces of information. We expected instead police officers to be able to interpret CSA-related information in the correct direction, that is, either increasing or decreasing the probability of CSA. Based on previous results (Tadei et al. 2018), indicating a low accuracy of experts in estimating CSA probability without the help of decision-support tools (e.g., FICSA; Tadei et al. 2017), we expected a low performance from the participants when

assessing the probability of CSA in the scenarios we presented them, independently of the performance in the CSA-related information selection and interpretation task.

Method

Participants

A total of 135 Finnish police officers participated in the present study. They were recruited through the Police College of Finland. Of the participants, 57% ($n = 77$) completed the online questionnaire in its entirety. We excluded participants only from analyses that required the missing information. Of the entire sample, 75% ($n = 101$) were detectives, the other 25% ($n = 34$) were non-detective officers. On average, participants had worked as police officers for 15 years ($SD = 8$). Of the participants, 62% ($n = 84$) had investigated at least one CSA case, whereas 38% ($n = 51$) had never dealt with a CSA case. Finally, 29% ($n = 39$) of the participants had attended a 1-year training in child interviewing.

Ethical Permission

The Institutional Review Board of the Faculty of Arts, Psychology and Theology of Åbo Akademi University considered the ethical implications of the present study and gave permission to the authors to carry it out. Before starting the questionnaire, participants signed an online consent form where policy about anonymity, possibility to withdraw from the study, and the aims of the research were described.

Measures and Procedures

The study consisted of two separate sections: section one (S1) and section two (S2). In S1, participants were presented with four scenarios (see ESM 1), each of them shortly describing the life situation and habits of four different children: *Minna* (14-year-old girl), *Salla* (8-year-old girl), *Petri* (14-year-old boy), and *Mikko* (8-year-old boy). The four scenarios included all possible combinations of gender (girl/boy) and probability of abuse (high/low). Each scenario contained exactly ten pieces of information (variables). Of these, five pieces were statistically unrelated to CSA and five pieces of information were statistically related to CSA, either increasing or decreasing the CSA risk probability. ESM 1 also reports the Bayes Factors for all the variables related to CSA. All variables were selected from a pool of approximately 1000 variables used to study the life of over 11,000 Finnish children and their experience with crime conducted in 2013 (Ellonen et al. 2013). This same pool of variables was used to train FICSA. We defined a variable as CSA-unrelated when the chi-squared test between the variable and the presence or absence of CSA

experiences was statistically non-significant and the chi-squared test value was close to zero. We considered a variable CSA-related when the chi-squared test between the variable and CSA was statistically significant *and* the variable was used as a feature in FICSA (Tadei et al. 2017).

After each scenario, participants had to report their perceived probability (0–100%) that the child described had been victim of CSA. CSA was defined as the occurrence of one or more experiences ranging from *Receiving a proposal to do anything sexual* to *Sexual penetration* involving a person before age 17, along with at least a 5-year age difference between the victim and the offender (Duodecim 2013). Furthermore, out of the ten variables in each scenario, they had to list the variables they used to estimate the CSA probability and specify, for each of them, if it increased or decreased the CSA risk.

In S2, participants were presented with a list of 44 variables (see ESM 2 for the list of variables and the Bayes Factors of the ones related to CSA). These variables belonged to one of the following categories: CSA-related for both boys and girls ($n = 1$), related only for boys ($n = 5$), related only for girls ($n = 16$), or unrelated to CSA both for boys and girls ($n = 22$). To define a variable as CSA-related or CSA-unrelated, we used the same approach as in S1. None of the 44 variables overlapped with the ones used in the four scenarios in S1. In S2, the variables were distributed unevenly across the four categories. This is because the number of features used by FICSA is not equal for these categories. FICSA uses 25 variables that are related to CSA only for girls, 14 only for boys, and 3 variables that are related to CSA for both girls and boys. For each variable, participants were asked to specify which of the four categories it belonged to, being informed that a variable could belong to only one of the four categories.

Statistical Analyses

Section 1. To evaluate how accurate participants were in estimating the CSA probability in the four scenarios, we compared the mean of their answers with the probability calculated by FICSA. For more informative descriptive statistics, since the four means were associated with large standard deviations, we decided to use a bootstrapping procedure with 1000 re-samples and a 95% confidence interval. A one-way ANOVA was used to investigate possible differences in the CSA probability estimates both between the scenarios and when compared with the probabilities calculated by FICSA. Descriptive statistics were used to report how well participants performed in excluding CSA-unrelated variables and in retaining CSA-related variables. We also checked whether the CSA-related variables were interpreted in the

correct direction (i.e., increasing or decreasing the CSA probability). Pearson correlation was used to test if a good or bad performance in recognizing CSA-related or CSA-unrelated variables was consistent across all scenarios. Finally, we used Pearson correlation to test if a more correct identification of the CSA-related and CSA-unrelated variables was associated with a more correct estimation of the CSA probability.

Section 2. Descriptive statistics were used to measure how good participants were in choosing the right category for each of the 44 variables listed among CSA-related for both boys and girls, related only for boys, related only for girls, or unrelated to CSA both for boys and for girls.

Results

Accuracy of CSA Probability Estimates

Table 1 shows the average probability estimated by the participants for each scenario, together with standard deviations, 95% confidence intervals, and probability estimated by FICSA.

The descriptive analysis in Table 1 shows how the estimates in all the scenarios are below 50%, indicating a tendency to consider the CSA allegation as unlikely when not supported by corroborative evidence. A one-way ANOVA analysis on the estimates, using the scenarios as factor, showed significant differences between the scenarios ($F(3, 368) = 18.72, p < .001$). However, a *Gabriel's* pairwise post hoc test, chosen because of the small differences in the size of the compared groups (Field 2013), did not indicate differences between scenarios 1 and 2 and between scenarios 3 and 4 as statistically significant ($p = 1.000$), while all the other contrasts were statistically significant ($p < .001$). These results indicate a gender effect, with the scenarios about girls

Table 1 CSA probability estimated by police officers and by FICSA

Scenario	<i>M</i>	<i>SD</i>	BCa 95%CI	FICSA
1 (girl: low)	41.44	24.78	36.76, 46.62	36.90
2 (girl: high)	41.43	24.24	36.66, 46.13	80.80
3 (boy: high)	24.10	20.30	19.95, 28.65	67.60
4 (boy: low)	24.01	18.94	20.20, 27.80	27.80

The column Scenario also gives details about the gender of the child and about the probability calculated by FICSA that could be either low (< 50%) or high (> 50%). *M* and *SD* refer to average probability of CSA estimated by participants. BCa 95%CI = Accelerated 95% confidence intervals. FICSA = Probability estimated by the Finnish Investigative Instrument of Child Sexual Abuse

receiving higher risk estimates than the scenarios about boys. However, the officers do not seem to be sensitive to the differences between scenarios with low and high probabilities of abuse. The comparison between our participants' estimates and the ones from FICSA shows that police officers have been accurate, even if at the limits of the confidence intervals, only in two scenarios out of four. In scenarios 2 and 3, the difference between the two estimates is about 35 points, indicating important misjudgments. We then investigated if the difference between police and FICSA estimates is statistically larger in scenarios 2 and 3 than in scenarios 1 and 4. A one-way ANOVA on the differences between the estimates from our participants and the probabilities calculated by FICSA, using the scenarios as factor, showed significant differences between the scenarios ($F(3, 368) = 58.80, p < .001$). A *Gabriel's* pairwise post hoc test, however, did not identify any statistical difference between scenarios 1 (girl: low) and 4 (boy: low, $p = .13$) and between scenarios 2 (girl: high) and 3 (boy: high, $p = .33$). These results could indicate a tendency among police officers to give low estimates overall when only background information is available, making them accurate for cases with low probability.

Selection and Interpretation of CSA-Related and CSA-Unrelated Variables

We next investigated police officers' performance in distinguishing between pieces of background information that have an influence on the CSA probability and the ones that are irrelevant. Table 2 shows, per scenario, the average number of irrelevant variables correctly excluded from the CSA evaluation, the average number of CSA-related variables used to estimate the CSA probability, and the average number of CSA-related variables that were used in the correct direction (i.e., increasing or decreasing the CSA probability). For each of these values, the optimal performance would be 5, as five unrelated and five related CSA variables were included in each scenario.

These results show a better performance in discarding irrelevant variables than in retaining relevant ones. These

Table 2 Average number of CSA-unrelated and CSA-related variables correctly identified and interpreted

Scenario	Unrelated discarded		Related chosen		Related correctly used	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	3.50	1.12	2.08	1.21	1.95	1.25
2	3.24	1.07	1.83	1.29	1.83	1.29
3	3.19	1.19	1.45	1.29	1.20	1.07
4	4.68	0.78	1.47	1.37	1.26	1.13

The maximum performance in each column is 5

numbers partially confirm our hypothesis about the difficulties to distinguish relevant information from irrelevant ones. The tendency not to consider background information as relevant for an accurate CSA evaluation can explain both discarding irrelevant variables and failing to retain relevant variables. The third column indicates that, when a CSA-related variable was included, it was done in the correct direction most of the times (91.7%), hence in line with our hypothesis about the capacity of police officers in understanding if a variable would either increase or decrease the CSA probability.

We studied the correlation between the performances in selecting or excluding the correct variables in the four scenarios. None of the correlations were statistically significant. Moreover, we did not find any statistically significant correlation between the performances in the CSA probability estimation task and the relevant/irrelevant information selection task. The results of the correlations are shown in Table 3.

With “Methods” of the present study, we investigated if police officers were able to separate CSA-related variables from unrelated variables when presented outside the context of a mock scenario. Table 4 shows a crosstab with the categorization done by police in the rows and the correct categorization on the columns.

Police officers correctly identified 49.8% of the unrelated variables (i.e., in more than 50% of the cases they thought an unrelated variable was related to CSA. As participants used almost only two categories, that is, unrelated and related for both, the percentage of correct identification for the unrelated variables would raise to 51.7%, only slightly better than chance level), 74% of the variables related to CSA for both genders, but only 0.8% of the ones related only for girls, and 0.5% of the ones related only for boys. As shown in Table 4, the variables related to CSA only for boys and only for girls have been interpreted, most of the times, as valid for both genders.

Discussion

In the current study, we aimed to investigate how police officers select and interpret background information in CSA mock evaluations, both when part of scenarios and when observed individually. We also investigated if the correct use of background information leads to better estimation of the CSA risk probability.

Table 3 Correlations between CSA probability estimation and relevant/irrelevant information selection task, divided per scenario

Scenario	<i>r</i>	<i>n</i>	<i>p</i>
1	– .12	101	.24
2	.04	94	.71
3	– .17	89	.11
4	– .03	88	.77

Table 4 Crosstab between police answers and correct categorization of CSA-related and unrelated variables

Police answers	Correct answers			
	Unrelated (<i>n</i> = 22)	Rel. only for girls (<i>n</i> = 16)	Rel. only for boys (<i>n</i> = 5)	Rel. for both (<i>n</i> = 1)
Unrelated	49.8	36.3	43.9	23.4
Rel. only for girls	2.4	0.8	0.5	2.6
Rel. only for boys	1.3	0.6	0.5	0.0
Rel. for both	46.5	62.3	55.1	74.0
	100.0	100.0	100.0	100.0

All the values are percentages, calculated by column

Selection and Interpretation of Background Information

When the background variables were contextualized in a scenario, police officers performed better in discarding the CSA-unrelated variables that in taking in consideration the related ones. This can be explained by the fact that participants overall tended to use a small number of variables (on average, 3 out of 10 per scenario) to estimate the CSA probability. On the other hand, police officers performed very well in understanding if a variable, when correctly identified as CSA-related, would have increased or decreased the CSA probability. When evaluating the police officers' performance in understanding the correct direction of the chosen CSA-related variables (91.7% overall), we must keep in mind the different baselines of variables that increase and decrease the CSA probability in each scenario. In the four scenarios, the distribution of the CSA-related variables was uneven. In scenarios 1 and 2, four variables increased the CSA probability and only one decreased it. In scenarios 3 and 4, three variables increased and two decreased it the CSA probability. However, considering that if officers had defined all CSA-related variables as increasing the CSA probability, this would have led to 80 and 60% accuracy, the 91.7% accuracy still shows that participants performed well.

When the variables were not contextualized in scenarios, police officers performed well in identifying variables related to CSA for both genders, but worse when the variable was unrelated to CSA, or related only for boys or only for girls. In particular, in more than 50% of the times, they considered CSA-unrelated variables as related. This result goes in the opposite direction compared to when the variables were inside scenarios. This might be due to the fact that, as demonstrated by the low CSA estimates, participants did not consider the CSA risk in the scenarios very likely and so tended to disregard a higher number of variables. When the variables were not part of a scenario, instead, police officers were not influenced by their own opinions about the CSA having occurred or not. Another possible explanation is that the single CSA-

related variable presented individually, compared to the ones in the scenarios, simply was more well known to the police officers. The reversed could be the case for CSA-unrelated variables, with more well-known variables being presented in the scenarios and less known presented separately. Finally, participants almost never considered the possibility that a variable could be related to CSA only for a specific gender. This result indicates the need for a much better training about gender-specific CSA risk and protective factors (i.e., variables that respectively increase or decrease CSA risk).

Estimation of CSA Probability

The results show that police officers, when asked to evaluate the CSA probability in scenarios containing only background information, tend to provide low values overall. On one hand, this conservativeness is understandable because of the short length of the scenarios and the complete absence of corroborative evidence. On the other hand, if the conservativeness translates to real allegations, high probability cases risk not to be identified as such. In all the scenarios, on average, they estimated a probability lower than 50%, meaning that they considered it was unlikely that the abuse had actually taken place. As a consequence of consistently estimating low probabilities, police officers were more accurate in evaluating low-probability scenarios. Furthermore, participants consistently estimated the CSA risk for a boy to be 15% less probable than when a comparable scenario was about a girl. The difference between the CSA prevalence for boys and girls (in Finland 0.7 and 3%, respectively; Tadei et al. 2017) might explain the difference between genders in the police officers' estimates. An alternative explanation could be the presence of a gender bias making police officers consider girls at a higher risk than boys in a comparable situation. Unfortunately, there is a lack of scientific literature about the perception of CSA risk for boys and girls and its accuracy. However, we might hypothesize that, since it is commonly known that most people who commit CSA are males (Faller 1996; McLeod 2015), oversimplistic reasoning could lead to the assumption that male

sex offenders would abuse girls in almost the totality of cases. Another reason why girls are perceived as at much higher risk than boys could be because the distribution of sexual assaults among adults indicates that women are more likely to be victimized (Breiding et al. 2015; Tjaden and Thoennes 2000). People, therefore, could tend to transpose the same prevalence to the children. These are possible explanations of the gender bias we identified in the present study, but they should be scientifically investigated to even be considered as plausible.

Correlation coefficients close to zero and absence of statistical significance for the association between ability of estimating the CSA probability and selection and interpretation of the CSA-related variables indicate a difficulty from police officers to evaluate the correct impact of each risk and protective factor on the CSA probability and to compute an overall probability for each scenario. This result is in line with what has been observed in a recent study by Tadei et al. (2018), which showed that expert psychologists from the Finnish CSA investigative units might benefit from using decision-making support tools when estimating CSA probability in cases that present also background information.

Limitations

To measure how correct participants were in estimating CSA probability, we compared their values with the ones we obtained by running FICSA on the four scenarios and that we considered correct. Even if it has been demonstrated that FICSA has excellent diagnostic validity ($AUC = 0.88$ for girls and 0.97 for boys; Tadei et al. 2017), there is still the possibility for a scenario to be among the false positives or false negatives. In this case, the police officers' estimations might be right and the FICSA's ones, that here we considered correct, might be wrong. However, we limited this risk basing our analyses on four scenarios rather than only one. Furthermore, the risk of wrong estimates by FICSA is also reduced by our results being in line with previous studies.

Earlier, we illustrated the possibility that the complete absence of information about a possible CSA event might have driven the participants towards lower estimates. This is supported by the fact that a CSA allegation where only background information is present would rarely be considered worthy of an investigation. However, a previous study (Tadei et al. 2018) demonstrated that the presence of information related to the alleged CSA event does not always influence the probability estimated using background information only. In any case, we could not add this type of evidence to our scenarios since FICSA would not take it in consideration, i.e., the probability provided by FICSA would be wrong, hence, we could not evaluate the police officers' performance.

Conclusions

With this study, we demonstrated that police officers partially use the available background information when evaluating CSA allegations. However, the relevance given to these pieces of information seems to be dependent on the police officers' own opinion about the CSA probability. Hence, we highlight the importance of better research and training on the nature and use of background information in CSA cases, with a special focus on the gender-specific risk and protective factors. We also suggest a consistent use, during the entire investigation process, of decision-making support tools based on actuarial data to help understand correctly the probabilistic value of each piece of evidence. Finally, we identified the presence of a possible gender bias that would make the officers estimate a much higher CSA risk in the scenarios involving girls. This bias, and the reasons behind it, necessitate a deeper exploration.

Acknowledgements Open access funding provided by Åbo Akademi University (ABO).

Compliance with ethical standards

The Institutional Review Board of the Faculty of Arts, Psychology and Theology of Åbo Akademi University considered the ethical implications of the present study and gave permission to the authors to carry it out. Before starting the questionnaire, participants signed an online consent form where policy about anonymity, possibility to withdraw from the study, and the aims of the research were described.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Aamodt MG (2008) Reducing misconceptions and false beliefs in police and criminal psychology. *Crim Justice Behav* 35(10):1231–1240. <https://doi.org/10.1177/0093854808321527>
- Adams JA (2011) Medical evaluation of suspected child sexual abuse: 2011 update. *Journal of Child Sexual Abuse* 20(5):588–605. <https://doi.org/10.1080/10538712.2011.606107>
- Antfolk, J., Rönholm, M., Salo, B., Rantamäki, S., & Santtila, P. (n.d.). Children's suggestibility in interviews – a meta-analysis of the effects of child's age, post-event delay, event-type, and question-type
- Breiding MJ, Smith SG, Basile KC, Walters ML, Chen J, Merrick MT (2015) Prevalence and characteristics of sexual violence, stalking, and intimate partner violence victimization—National Intimate Partner and Sexual Violence Survey, United States, 2011. *Am J Public Health* 105(4):e11–e12. <https://doi.org/10.2105/AJPH.2015.302634>

- Cromer LD, Goldsmith RE (2010) Child sexual abuse myths: attitudes, beliefs, and individual differences. *J Child Sex Abuse* 19(6):618–647. <https://doi.org/10.1080/10538712.2010.522493>
- Duodecim (the Finnish Medical Society). (2013). Current care guidelines: investigating suspicions of child sexual abuse. Retrieved from <http://www.kaypahoito.fi/web/kh/suositukses/suositus;jsessionid=7E3A7E56839CC04A143BED0DA7F6BD97?id=hoi34040>
- Ellonen, N., Fagerlund, M., Kääriäinen, J., Peltola, M., & Sariola, H. (2013). Child Victim Survey 2013 [dataset]. Version 1.0 (2015-03-16). Finnish Social Science Data Archive. Retrieved from <http://urn.fi/urn:nbn:fi:fsd:T-FSD2943>
- Faller KC (1996) A clinical sample of women who have sexually abused children. *Journal of Child Sexual Abuse* 4(3):13–30. https://doi.org/10.1300/J070v04n03_02
- Fenton N, Neil M (2012) Risk assessment and decision analysis with Bayesian networks. Taylor and Francis Group, Boca Raton, FL
- Field A (2013) Discovering statistics using IBM SPSS statistics. SAGE, London
- Gardner RA (1998) The parental alienation syndrome: a guide for mental health and legal professionals, 2nd edn. Creative therapeutics, Cresskill
- Heger A, Ticson L, Velasquez O, Bernier R (2002) Children referred for possible sexual abuse: medical findings in 2384 children. *Child Abuse Negl* 26(6–7):645–659. [https://doi.org/10.1016/S0145-2134\(02\)00339-3](https://doi.org/10.1016/S0145-2134(02)00339-3)
- Henderson J (2002) The use of DNA statistics in criminal trials. *Forensic Sci Int* 128(3):183–186. [https://doi.org/10.1016/S0379-0738\(02\)00198-6](https://doi.org/10.1016/S0379-0738(02)00198-6)
- Herman S (2010) The role of corroborative evidence in child sexual abuse evaluations. *J Investig Psychol Offender Profiling* 7(3):189–212. <https://doi.org/10.1002/jip.122>
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning. An introduction to statistical learning. Springer, New York. <https://doi.org/10.1007/978-1-4614-7138-7>
- Koehler JJ (2011) Misconceptions about statistics and statistical evidence. In: *Handbook of trial consulting*. Springer, Boston, MA, pp 121–133. <https://doi.org/10.1017/CBO9781107415324.004>
- Koul A, Becchio C, Cavallo A (2017) PredPsych: a toolbox for predictive machine learning-based approach in experimental psychology research. *Behav Res Methods* 50:1–16. <https://doi.org/10.3758/s13428-017-0987-2>
- McLeod DA (2015) Female offenders in child sexual abuse cases: a national picture. *J Child Sex Abuse* 24(1):97–114. <https://doi.org/10.1080/10538712.2015.978925>
- Myers JE (2005) Myers on evidence in child, domestic, and elder abuse cases. Aspen Publishers, New York
- O'Donohue W, Benuto LT, Bennett N (2016) Examining the validity of parental alienation syndrome. *J Child Custody* 13(2–3):113–125. <https://doi.org/10.1080/15379418.2016.1217758>
- Pipe ME, Sutherland R, Webster N, Jones C, La Rooy D (2004) Do early interviews affect children's long-term event recall? *Appl Cogn Psychol* 18(7):823–839. <https://doi.org/10.1002/acp.1053>
- Saunders CL (2012) The truth, the half-truth, and nothing like the truth. *Br J Criminol* 52(6):1152–1171. <https://doi.org/10.1093/bjc/azs036>
- Stroud DD, Martens SL, Barker J (2000) Criminal investigation of child sexual abuse: a comparison of cases referred to the prosecutor to those not referred. *Child Abuse Negl* 24(5):689–700
- Tadei A, Santtila P, & Antfolk J (2018) Usability and Effectiveness of the Finnish Investigative Instrument of Child Sexual Abuse. <https://doi.org/10.31234/osf.io/vehk8>
- Tadei A, Pensar J, Corander J, Finnälä K, Santtila P, Antfolk J (2017) A Bayesian decision-support tool for child sexual abuse assessment and investigation. *Sexual Abuse: A Journal of Research and Treatment*. <https://doi.org/10.1177/1079063217732791>
- Tjaden P, Thoennes N (2000) Prevalence and consequences of male to female and female to male intimate partner violence as measured by the national violence against women survey. *Violence Against Women* 6(2):142–161 803973233
- Walsh WA, Jones LM, Cross TP, Lippert T (2010) Prosecuting child sexual abuse: the importance of evidence type. *Crime Delinq* 56(3):436–454. <https://doi.org/10.1177/0011128708320484>
- Zawoad, S., & Hasan, R. (2015). Digital forensics in the age of big data: challenges, approaches, and opportunities. In *2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems* (pp. 1320–1325). <https://doi.org/10.1109/HPCC-CSS-ICSS.2015.305>