Development of standardized clinical training cases for diagnosis of sexual abuse using a secure telehealth application

Lori D. Frasiera,*, Ioana Thraenb, Rich Kaplanc, Patricia Goed

a University of Utah School of Medicine, Primary Children’s Center for Safe and Healthy Families, 675 E. 500 So. Suite 300, Salt Lake City, UT 84102, USA
b University of Phoenix, John Sperling School of Business, University of Utah, College of Social Work, Salt Lake City, UT, USA
c University of Minnesota, Children’s Hospitals of Minnesota, Minneapolis, MN, USA
d VisualShare, LLC, Salt Lake City, UT, USA

A R T I C L E   I N F O

Article history:
Received 11 October 2010
Received in revised form 21 June 2011
Accepted 24 June 2011
Available online 8 March 2012

Keywords:
Abuse
Educational technology
WEB education
Telemedicine

A B S T R A C T

Objectives: The training of physicians, nurse examiners, social workers and other health professional on the evidentiary findings of sexual abuse in children is challenging. Our objective was to develop peer reviewed training cases for medical examiners of child sexual abuse, using a secure web based telehealth application (TeleCAM).

Methods: Sixty de-identified cases developed by 2 child abuse pediatricians, were stratified by availability of information (minimal, moderate, comprehensive) for both positive and negative child sexual abuse findings. These cases were narrowed to a set of 30 cases through an expert peer review process using pediatricians with extensive expertise in the evaluation of child sexual abuse. A previously studied secure web-based telehealth application TeleCAM which contains a child abuse workflow, was used to develop, disseminate and review cases. A series of Free Margin agreement statistics are used to select those cases with the highest rates of agreement. A final set of 30 cases are stratified equally by availability of information and for both positive and negative findings, Mantel Haenszel Chi-square was used for trend analysis of the ordered categorical variables.

Results: The highest degrees of inter-rater reliability was found in cases with moderate to comprehensive information. Cases with minimal data had poor kappa agreement indicating that availability of differing levels and types of information contribute to variability in diagnostic findings.

Conclusion: These final cases will be further studied with medical examiners in various settings utilizing TeleCAM as the application for dissemination.

© 2011 Elsevier Ltd. All rights reserved.

Introduction

The evaluation of children and adolescents suspected of having been sexually abused is dependent upon an understanding of many complex issues. The history from the child obtained in a forensically supportable manner has been shown to be the single most important factor in the determination of whether a child has been sexually abused (Berkoff et al., 2008; Heger, Ticson, Guerra, et al., 2002; Kellogg, Menard, & Santos, et al., 2004). The manner in which that history is taken and recorded becomes evidence in legal settings to support or refute allegations of sexual abuse (Berenson et al., 2000; Heger, Ticson, Guerra, et al., 2002). The results of the medical examination may be used by legal authorities to determine if physical evidence is present that confirms the allegation that sexual abuse or assault has occurred. Both clinical experience

* This research has been supported by NIH, NICHD Grant HD 052329 to Ms. Goede is the PI, and owner of VisualShare.

0145-2134/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chiabu.2011.06.006
and research studies have demonstrated that physical evidence is exceedingly rare in children who allege sexual abuse. When prepubertal children with alleged sexual contact are evaluated in the non acute setting more than 95% will have no physical findings that could be discerned as abnormal as compared to non abused children (Berenson et al., 2000; Heger, Ticson, Velasquez, & Bernier, 2002). Additional work suggests that even when acute injuries to the anogenital area are present, those injuries heal quickly and often without physical sequela. The exceptions are when there is severe, penetrating trauma involving extensive tissue injury (McCann, Miyamoto, Boyle, & Rogers, 2007a, 2007b). Medical and legal professionals may place too much emphasis on the physical findings or lack thereof to prove or disprove a sexual act has occurred.

Educators and practitioners in the field of child sexual abuse are often limited by their own experiences and clinical setting. This study attempted to begin the development of a set of standard cases where cross-expert consensus could be gained. Standardized cases can be used as a training tool as well as referral library for examiners. As examiners (pediatricians, nurse practitioners, family medicine physicians, and so forth) are confronted with such practice issues, a set of learning cases that have been vetted by multiple child sexual abuse experts could be very helpful.

Indeterminate and unclear cases tend to confuse new learners and need further study on the confounding issues associated with uncertain outcomes. Guidelines for the interpretation of physical findings that result in a diagnostic confidence that sexual injury has occurred have been published and include the presence of forensic evidence (sperm, semen, DNA, trace evidence) sexually transmitted infections, or pregnancy as diagnostic criteria (Adams et al., 2007). Such evidence can occur in the absence of other physical findings.

Physicians and other medical providers may be unaware of the rarity of physical evidence in children and adolescents. Myths and misconceptions about the physical injuries resulting from sexual abuse in children persist despite evidence to the contrary. This may be due to lack of specialized training in undergraduate and graduate medical/nursing education as well as difficulty disseminating consensus data on physical evidence in child sexual abuse. Newly trained examiners are often expected to provide accurate assessment of findings with little additional oversight or mentoring. Research indicates that less experienced examiners are overly influenced by the history alone, and have the expectation that a child must have findings (Paradise, Winter, Finkel, Berenson, & Beiser, 1999; Sinal et al., 1997). Investigators and prosecutors may exert undue influence over an isolated medical provider to find evidence. Adherence to peer reviewed research standards should be the standard of practice. Significant legal and social consequences result when examination findings are misinterpreted because legal authorities rely heavily on such physical evidence when pressing charges and prosecuting cases.

A case- or problem-based format or didactic lecture is a common method or training. Training modules should include complete clinical data combined with physical examination and laboratory/forensic information. Expert consensus is one method in determining which cases meet the standards for evidence. Experts in child abuse have been shown to have higher inter-rater reliability statistics when evaluating cases; however, this is not universal (Paradise et al., 1999; Sinal et al., 1997). Therefore the development of training cases that have a high consensus among experts is critical. Paralleling the need for training is the rapidly emerging use of telehealth and telemedicine applications for peer review, case consultation, and oversight (Kellogg, Lamb, & Lukefahr, 2000). Linkage of geographically disparate medical clinics, with child abuse pediatricians has been challenging. Combining case based training with a secure telehealth application offers tremendous advantages in both areas.

Recently the National Children’s Alliance mandated photodocumentation of genital findings by examiners as one criteria for accreditation of Children’s Advocacy Centers who perform or refer children for medical exams (“National Children’s Alliance Standards for Accredited Members,” 2008). Photodocumentation is required to be submitted to a peer review process that can be documented during accreditation. Web based training cases, which have themselves undergone extensive review by experts have never been available for such purposes. The medical history and forensic interview information from a child victim also should contain contextual and historical details that provide reliability to this information.

Examiners should be alert to the fact that details regarding an event of child sexual abuse are disturbing and emotionally charged. The concept that “children never lie,” has been put forth in many settings as a reason to believe a child. Additionally, the interpretation of the child’s statements (i.e., he put “it” inside of me) compels examiners and investigators alike to put their own frame of reference into the assessment. Although there is research and clinical evidence to the contrary, the idea that the child’s medical examination will be able to prove or disprove the child’s statement is pervasive. Also, many medical, dermatological, and anatomic conditions have been described which “resemble” traumatic sequelae, or are confusing to an examiner (Adams & Horton, 1989; Frasier, 2011; Jenny, Kirby, & Fuquay, 1989). The inexperienced examiner may interpret those findings as “abnormal” merely due to the fact that they have not encountered such findings previously. External review becomes important, especially if performed by examiners with extensive knowledge in the field.

This study is designed to develop a standardized set of cases that ultimately can be used as a training set utilizing a secure web based application. The set of cases may be used to train new examiners or provide refresher cases for experienced examiners. They can be uploaded into a digital library or placed on a secure web page for examiners to review. Aggregating images for purposes of diagnosis of child sexual abuse across domains (e.g., radiology, photodocumentation, and pathology), workflow (e.g., Child Advocacy Centers, physician offices, inpatient hospital care) time, and geographies (rural vs. urban or primary vs. tertiary centers) can greatly enhance clinical collaboration, diagnostic efficiencies, continuity of care, and
ultimately patient care outcomes. The research was conducted as part of a National Institutes of Health grant to develop and test a web-based system for child abuse medical assessment.

Methods

TeleCAM, a secure web based system, was used to collect and test cases used for peer reviewed diagnostic assessment in this study as part of a Phase 2 Small Business Innovative Research Grant funding by the NIH, and NICHD. TeleCAM is a telehealth information system that provides a standards-based technological infrastructure for the collection, storage, and display of digital images via a secured web technology (Lazadikou, 2010). This application has previously demonstrated excellent usability among remote centers engaged in medical child abuse assessments (Thraen, Frasier, Cochella, Yaffe, & Goede, 2008). IRB approval was obtained July 14, 2009. Two expert reviewers originated and submitted 60 cases (30 each) from a practice pool collected over 20 years. Both expert reviewers (LF and RK) are board certified in pediatrics and child abuse pediatrics with over 20 years experience each in evaluating child abuse, and participating in peer review processes. Sixty cases were selected for convenience and selected with excellent photodocumentation and/or clinical history. The dichotomous outcomes of “negative” or “positive” in regards to child sexual abuse were varied by ordered sets of information. Sets contained minimal (historical, or interview information only, no images or laboratory information), moderate (reported historical or interview information combined with digital images of physical findings) or comprehensive information (reported historical, or interview information, digital images of physical findings, and additional medical findings such as laboratory, pathology, and radiological findings). All cases were de-identified to meet HIPAA standards and are used for teaching purposes only.

The cases consisted of actual and composite cases that were representative examples of the expert reviewer experience in the field of child sexual abuse evaluation (sexually abused children). Case collection was based on a standardized case workflow consisting of the following fields: Demographics, History, Child Development and Behavior, Vitals, Physical Exam, Genital Exam, Anal Exam and Clinical images (Fig. 1). Cases were uploaded to a secure web-based research database. The database for case collection was MySequel (“MySQL database,” 2010) running on standard Apache (Dulovic et al., 2010) web configuration. The authentication was a role-based access control so that each expert reviewer was authenticated to create, upload and edit the candidate cases. The project team described a coding scheme for candidate case entry and image upload specific to the two submitting expert reviewers. Once cases were uploaded (n = 60, 30 cases from each expert), each expert reviewed and determined that each other’s cases were qualified and met the information stratification criteria. Once one hundred percent agreement was achieved between the submitting experts. The cases were then randomized for a second round of reviews by 6 national expert/peer reviewers.

Six expert peer reviewers (4 female and 2 male) from across the US (CA, MO, NH, NV, NY, TX) were recruited from a panel of child abuse experts belonging to the Ray E. Helfer Society. Six reviewers were chosen as a convenience sample. None of the original reviewers declined to participate in the study. Expanding to additional reviewers and examiners could be second phase of this project. The reviewers were recruited based upon their extensive experience in the field and background as reviewers of cases. All of the expert reviewers have extensive expertise in child abuse evaluation. The Ray E. Helfer Society is an honorary society of physicians who are nominated for membership based upon their standing and participation in the field of abused and neglected children. The reviewers were all board certified in General Pediatrics. At the time of the study, the Child Abuse Pediatrics Board examination had not been administered, and so this was not a criterion for recruitment.

Each of the 6 expert/peer reviewers reviewed 60 cases each for a total of 360 case outcome determinations. Each reviewer was assigned a unique login and password that was auto generated and sent via with a URL link that redirected them to the secure research web site. Authenticated reviewers had “read only” access to anonymous and randomized cases. Each expert/peer reviewer had access to the following fields: Demographics, History, Child Development and Behavior, Vitals, Physical Exam, Genital Exam, Anal Exam and Clinical images and lab results. Each reviewer was required to submit a dichotomous response of positive or negative findings for child sexual abuse with an optional field for additional comments. The reasons for the dichotomous options in the training cases, which may not reflect clinical practice, were to eliminate the indeterminate category where consensus was very difficult to reach. In reality, most investigative agencies want to know if the exam reflects sexual abuse or not. In clinical practice when a finding by Adams et al. (2007) is indeterminate, most referral agencies see that as a negative examination. This study aimed to develop a set of training cases where the outcomes were clear based on the expert consensus. In the future, an additional set of indeterminate cases could be developed as a second training tool that can illustrate the varying degrees of interpretation and determinations and the issues associated with the lack of agreement. Both child and adolescent cases were selected. Although some examiners may be limited to either population, and state laws differ in age of consent for adolescents, the authors in this study chose a broad range of cases that, in their experience, could be seen in a variety of clinical settings. Conviction rates or confessions were not included in the case information. Such information is very difficult to obtain. Additionally the cases were meant to replicate a clinical assessment of a case, not a legal one. Confession and conviction data may be helpful at a later point in a similar study.

The 6 peer reviewers were blinded to the outcomes of the cases and were then asked to select a determination for normal or abnormal findings. Agreement analysis was conducted and the 30 cases with the highest degree of inter-rater reliability were selected from the original 60 cases and will be used as a gold standard set of cases for future review by field examiners. The flow of the review process is displayed in Fig. 2.
Results

Percentage agreements were calculated for individual cases and then used to calculated kappa agreement by category. A free marginal kappa was selected to determine agreement within categories of types of information (minimal, moderate, comprehensive) and finding outcomes normal (negative findings for sexually abused children) or abnormal (positive findings for sexually abused children) findings. A free margin kappa was chosen because reviewers were not forced into specific number of cases per category. Kappa Agreement results of good agreement is defined to be .61 to .80 and very good to be
from .81 to 1. The 60 original cases were split into samples of 10 cases each between positive and negative findings and across the 3 information categories (minimum, moderate, and comprehensive information) Kappa results for each outcome are described below in Table 1. This table demonstrates the varying degree of agreement by category of information and final determination as assessed by the 6 expert reviewers and in contrast to the 2 submitting experts. These differences were anticipated by the research design which requires a vetting process from an original 60 training cases agreed to by 2 expert reviewers, to 30 training case outcomes agreed to by an additional 6 expert reviewers, and to a final set of 24 training case outcomes agreed to through field examination.

Kappa agreements per category were either good or very good among the expert reviewers. A hypothesis test for 2 independent proportions for agreement between abnormal and normal findings indicate that the differences were not significant (p = .6385). Given that the original set of cases had full agreement between the 2 gold standard expert reviewers, a Chi-square goodness of fit test was conducted to test between the expected rate of agreement and the observed agreement rate by all 8 reviewers. An overall p value of .04 was obtained indicating that there are significant differences in agreement on outcomes (normal vs. abnormal) across the 8 expert reviewers. Further evaluation indicated significant differences in outcome with minimal information type at a p value of .002. No significant differences were found with moderate and comprehensive information. Sensitivity and specificity tests were also conducted by types of information and are found in Table 2. As the information type moved toward comprehensive information, sensitivity and specificity both approached 1. Qualitative information regarding both the cases and the use of the telehealth application was elicited from the reviewers. Cases were submitted in a manner that mirrored how actual clinical scenarios present themselves. Narrative content provides qualitative information that is extremely helpful when analyzing case materials. Some comments reflected state law on age of consent for sexual activity. In one state for example, a case may not be considered statutory rape based upon the ages of

Fig. 2. Flow of review process.

<table>
<thead>
<tr>
<th>Information</th>
<th>Normal (negative) findings</th>
<th>Abnormal (positive) findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal—interview info only</td>
<td>K = .66</td>
<td>K = .69</td>
</tr>
<tr>
<td>Moderate—interview info plus documented physical findings</td>
<td>K = .83</td>
<td>K = .95</td>
</tr>
<tr>
<td>Comprehensive—interview info, documented physical findings, and medical results</td>
<td>K = .95</td>
<td>K = .81</td>
</tr>
</tbody>
</table>
Table 2
Sensitivity and specificity results stratified by types of information.

<table>
<thead>
<tr>
<th>Information type</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>.753</td>
<td>.759</td>
</tr>
<tr>
<td>Moderate</td>
<td>.908</td>
<td>.986</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>.987</td>
<td>.951</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information type</th>
<th>True positives</th>
<th>False positives (type I error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>.763</td>
<td>.240</td>
</tr>
<tr>
<td>Moderate</td>
<td>.987</td>
<td>.013</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>.95</td>
<td>.048</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information type</th>
<th>True negatives</th>
<th>False negatives (type II error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>.750</td>
<td>.247</td>
</tr>
<tr>
<td>Moderate</td>
<td>.900</td>
<td>.092</td>
</tr>
<tr>
<td>Comprehensive</td>
<td>.987</td>
<td>.013</td>
</tr>
</tbody>
</table>

Table 3
Examples of narrative comments provided by the expert reviewers.

- **Comprehensive negative case:**
  1) I have been unable to enlarge the images. I do not know if this is a problem on my end. I have been able to enlarge them for the other cases. I have not been able to use the slideshow feature on any cases.
  2) Impressive skin lesions. No wonder she is grouchy. I would have gotten a rapid strep test also as one can have both EBV and strep.
  **Comprehensive positive case:**
  1) If the urine is really hers and there is confirmation of sperm in her urine, this would be very concerning. However the child still needs to be sent for interviewing.
  2) Leaving the info that she has not had a period in two months and not mentioning the medical care, including a pregnancy test is unsatisfying, but not necessary to form an opinion about the scenario.
  **Moderate negative case:**
  1) The normal exam of course does not mean that nothing happened, but it also does not mean something did.
  2) The child needs a forensic interview. The behaviors are very concerning as is the history of possibility of being exposed to a sexual offender. There is not enough information to reach a conclusion one way or the other.
  **Moderate positive case:**
  1) The photo with the Foley catheter show complete clefts at 3 and 9, which could be from healed lacerations. Has this girl had other sexual partners? In the absence of acute injuries, sperm or pregnancy, this history and physical exam is indeterminate for abuse.
  2) I am not clear on the description of the anal fining being past the anal verge.
  **Minimal negative case:**
  1) Difference in age makes this unlawful intercourse with a minor.
  2) Good detail from adolescent. Is the written results of the anal exam automatically generated? The history says he will not let anyone look at his butt.
  **Minimal positive case:**
  1) There was not enough information. If the pregnancy tests and STD tests are negative and the boyfriend is not an adult I would call this negative.
  2) Based solely on the events described by the child, this does appear to have been the cause of the inappropriate touching.

The victim and the actor. In another state, this may clearly be illegal. The reviewers were also quick to point out typographical errors, photographic issues, inability to use the program effectively, and apparent conflicts in the data or missing information. Table 3 is an example of a few of those comments stratified by amount of information. Those comments often were technical (software) or clinical or both. These issues actually reflect common problems with cases that experts are asked to review, and are reflective of actual practice. It is common for an expert to be asked to review a case that either has missing or poorly transmitted information.

**Discussion**

This study, using a secure web based telehealth application (TeleCAM) which supports text and image information accessibility, demonstrates that high quality information must be available in order to develop training cases that can be used to enhance diagnostic accuracy in sexual abuse medical evaluations. High quality data consists of detailed historical information and interpretable photodocumentation (sufficient lighting, composition, magnification, and exposure) that clearly demonstrates the findings noted on the examination. In some cases the addition of comprehensive data (laboratory, forensic) is important when other information is not available. Less experienced examiners overcall physical findings in favor of the positive historical information (bias of history). Our study demonstrates that in the presence of excellent historical information but subsequent apparent conflicting data (i.e., a child’s history of penetration, and a normal genital examination) expert examiners will agree. Despite the normal findings, a history may support a positive finding of sexual abuse. A history in isolation (minimal information) no matter how detailed, results in a lower agreement among experts. One limitation of this study is the lack of power calculations. Also, the panel of experts who reviewed the cases was “forced” into either positive or negative determinations. Examiners in clinical practice often refer to examinations and situations as inconclusive.
or indeterminate based upon the information provided. In reality, law enforcement and child protective authorities will interpret inconclusive cases as negative. The geographic dispersion of the experts may not represent the entire country, and the expert reviewers were selected on the authors’ personal knowledge of the experts experience and direct experience reviewing cases with this peer group of experts. This may have biased the results; however, there are relatively few expert examiners who met the criteria for expert knowledge and experience thus making the pool for selecting examiners small. Additional experts with different experiences may have altered the results of the study. Further study of why and how experts agree is warranted.

Conclusion

Medical literature and clinical experience demonstrate that examiners and reviewers of sexual abuse medical examinations have various degrees of experience, levels of specialization and biases that may influence their opinions regarding physical evidence. Total agreement among experts in any field remains illusory and probably unobtainable. However the degree of agreement can be maximized by the quality and amount of information that is available. Many attempts over the years have been made to improve examiners knowledge of physical findings, both normal and abnormal in child sexual abuse cases. This study demonstrates a method of evaluating inter-rater reliability among experts in the field using a standardized set of cases supported by a web based information technology. Training of medical examiners from professional schools (medical and nursing) is sporadic and limited.

Most physicians receive little training in residency. Nurses and nurse practitioners are often trained through continuing education programs or on-the-job training. The greatest future potential to standardize training and expertise is the new Child Abuse Pediatrics (CAP) fellowship recently approved by the Accreditation Council for Graduate Medical Education (ACGME). The ACGME has set specific content and practice criteria for specialty fellows training in CAP fellowships. The CAP Subboard certifying examinations will reflect the content and practice of a broad based CAP practice. Other primary care physician and nursing organizations will benefit from more standardized training among the physician leaders in this field. The next phase of this study will use the selected 30 cases for a second validation process with field examiners resulting in a final set of 24 gold standard training cases. TeleCAM as an application for web based, workflow oriented, peer review and expert oversight is also demonstrated as secure and accessible support technology for the field of child sexual abuse evaluation.

Acknowledgements

We wish to thank the expert reviewers for their participation in the study: Joyce Adams, Nancy Kellogg, Neha Meta, Vince Palusci, Timothy Kutz, and Wendy Gladstone.

References