



Is Current Child Forensic Practice Unethical?

What Research Says about the Reliability of Our Case Decisions

Mark D. Everson PhD

Presented at San Diego International Conference
on Child and Family Maltreatment

January 26, 2012



Agenda

- The challenge
- The argument
- The rebuttal
- The new worry

Key References

- Herman, S. (2005). *Improving decision making in forensic child sexual abuse evaluations*. Law and Human Behavior, 29(1), 87-120.
- Herman, S. (2009). *Forensic child sexual abuse evaluations: Accuracy, ethics and admissibility*. Chapter in Kuehnle, K. & Connell, M. (Eds.). The evaluation of child sexual abuse allegations: A comprehensive guide to assessment and testimony, p. 247-266, Hoboken, NJ: Wiley.



The Money Quote

“Because of the high risks of harm to the innocent and vulnerable, ***ethical*** CSA evaluators should not employ ***faulty diagnostic techniques*** known to have very high false positive error rates, ***even if*** this makes it impossible to substantiate some true, but uncorroborated allegations of sexual abuse....



Quote, continued

“The “technique” referred to is the process of making ***inferences*** about the validity of abuse allegations on the basis of ***psychosocial evidence***, including children’s reports in investigative interviews.”

Herman, 2009, p.258



Herman's Proposed Reforms

1. Substantiation decision only if hard definitive evidence
2. Dismissal of all psychosocial evidence, including behavior symptoms and child's disclosure statement
3. Universal adaptation of NICHD protocol
4. Shifting primary investigative responsibility in CSA cases from CPS and MH to LE



Herman's Argument



Hard Evidence

Not defined – Described as “corroborative,”
“clear and convincing”

Sufficient to make a judgment, definitive

Examples include:

- Perpetrator confession
- Diagnostic medical evidence
- Photographs or videos of the abuse
- Eye witness



Psychosocial Evidence

Not defined – Described as requiring subjective interpretation

Examples include:

- Content or quality of child's verbal report
- Behavior symptoms or changes
- Psychosocial context of allegations
- Psychosocial history of various parties



Common Beliefs Among Evaluators

- Hard, definitive evidence is uncommon.
- Child's verbal statement is the central evidence in most cases of CSA.

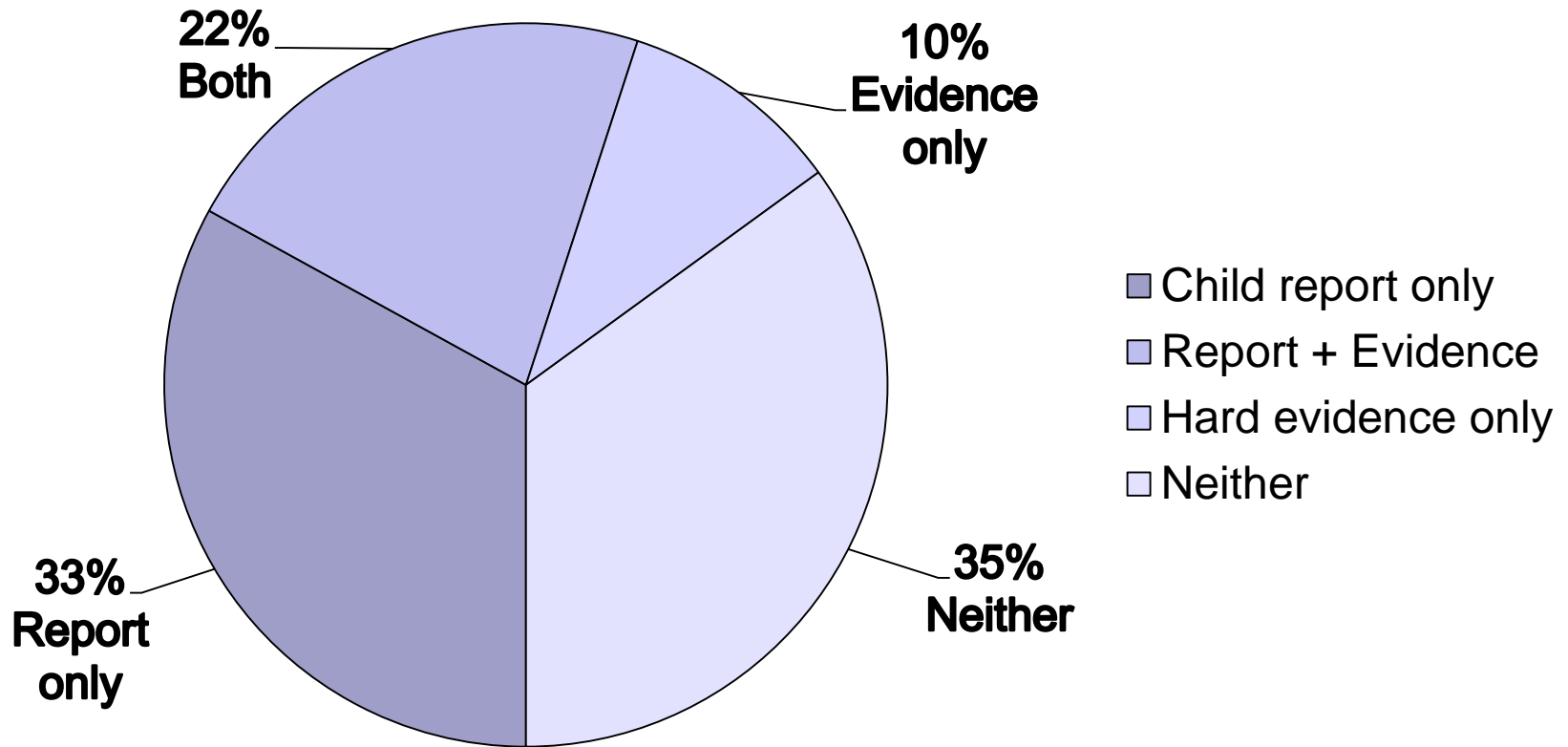


Are these beliefs true?

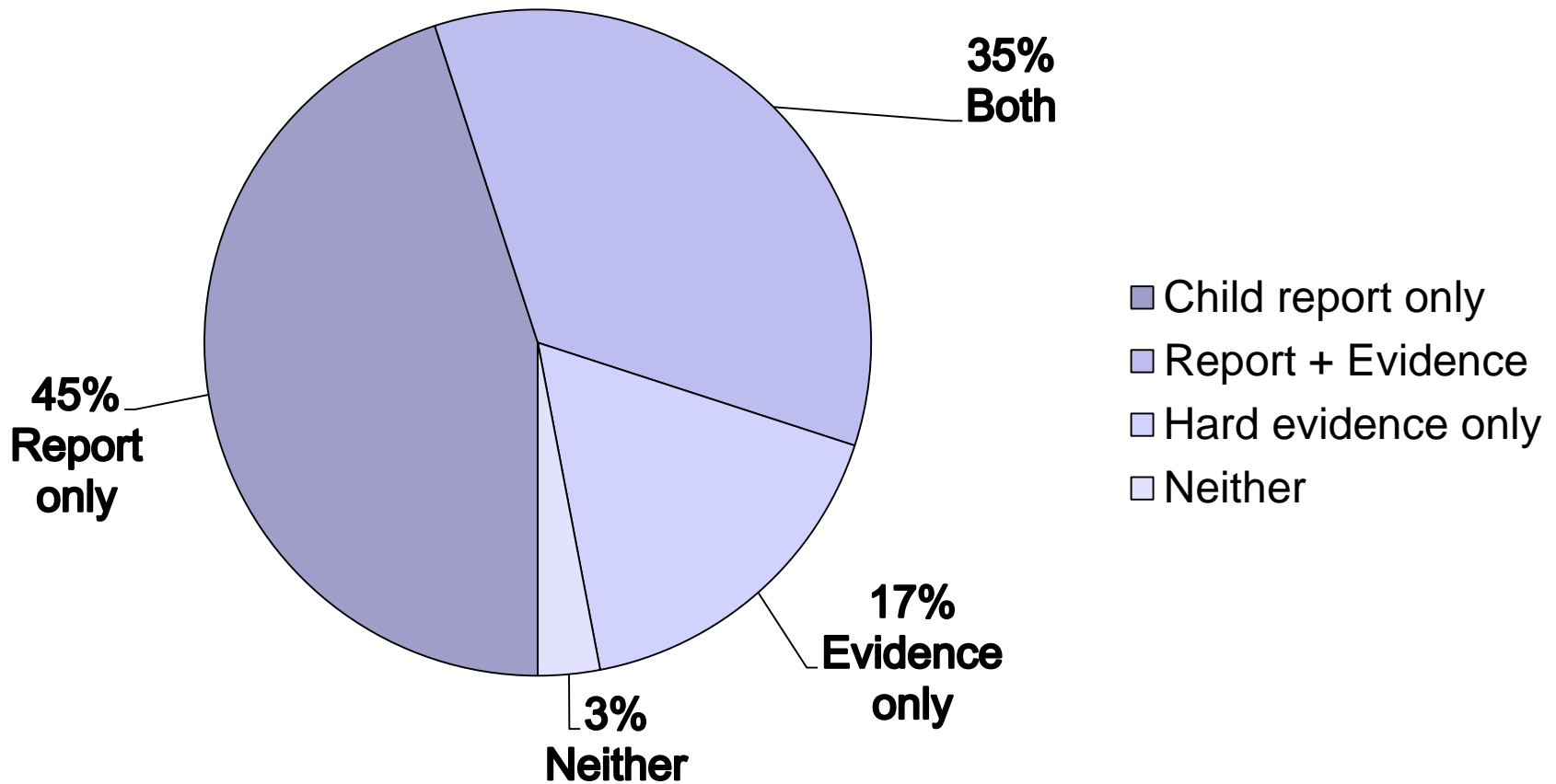
Three relevant studies:

- Elliott & Briere (1994)
- DiPietro, Runyan & Fredrickson (1997)
- Dubowitz, Black & Harrington (1992)

% of Combined Sample (N=677)



% of Substantiations (N=419)





Herman's Conclusions

Common beliefs are therefore wrong.


1. Hard, definitive evidence is common (33%).
2. Child report is not central (only 45% of subs).
3. Hard, definitive evidence is sufficient for decision making (52% of subs).



How are decisions based on psychosocial evidence made?

- Informal clinical method
- Holistic judgment of the data
- Implicit rather than explicit decision rules

This is a highly subjective process.



How valid or accurate are evaluation decisions based on psychosocial evidence?

- Difficult question to address
 - Requires knowledge of case validity
- Direct assessment vs. indirect assessment



Indirect Assessment of Validity

- Use of reliability to estimate validity

Reliability = measure of agreement
in decision making

Validity = measure of accuracy in
decision making

- Reliability is easier to assess

Possible to observe that two
evaluators disagree about a case
without knowing which one is correct.



Classical Test Theory

Reliability is a necessary, but insufficient condition for validity

Poor reliability = poor validity

Good reliability = ? validity



Published Studies on Reliability of Professional Judgments

- Finlayson & Koocher, 1991
- Horner, Guyer & Kalter, 1992, 1993a, 1993b
- Jackson & Nuttall, 1993
- McGraw & Smith, 1992
- Realmuto, Jensen & Wescoe, 1990
- Realmuto & Wescoe, 1992




Common Research Design

- Review of same set of case materials
- Rating of credibility of case or decision whether or not to substantiate
- Reliability is high to degree participants agree on ratings

Reliability Across Studies

| | <u>Reliability</u> | <u>Kappa Interpretation*</u> |
|------------------------------------|--------------------|------------------------------|
| Realmuto et al. (1990) | -- | -- |
| Finlayson & Koocher (1991) | .42 | Moderate |
| Horner et al. (1992, 1993a, 1993b) | .08 | Poor |
| Realmuto & Wescoe (1992) | .36 | Fair |
| McGraw & Smith (1992) | .14 | Poor |
| Jackson & Nuttall (1993) | -- | -- |
| * Landis & Koch (1977) | | |



Therefore, on average, forensic evaluators make the wrong judgment in about 1 of every 4 cases.

Unacceptable high error rate

Direct Assessment of Accuracy

Hershkowitz, Fisher, Lamb & Horowitz
(2007)

Participants: All Israeli youth investigators
serving in 2003 ($N = 42$)

Target cases:

24 cases in which children alleged CSA

12 “plausible”

12 “implausible”



Case Material: Written transcript of one child forensic interview for each case

All other information and evidence withheld



Herman's Re-Analysis: Accuracy of Professional Judgments

Overall

61%

Hit rate based on child's disclosure statement

39%

Error rate based on child's disclosure statement

Most errors were false positives



Herman's Conclusion

Consistency across studies:

24% minimum error rate in prior studies

39% error rate in Hershkowitz, et al.

Unacceptably high

Invalid methodology

Unethical practice



Our Rebuttal




Rebuttal Argument

1. Over-estimation of rate of hard evidence (32%)
2. Reliability studies unrepresentative of forensic practice and lack ecological validity
3. Unintended consequences for children

Rates of Diagnostic Medical Findings

| | <u>Data Collection Year</u> | <u>Sample Size</u> | <u>% Diagnostic Medical Findings</u> |
|-----------------------------|-------------------------------------|------------------------|----------------------------------------------|
| Elliott & Briere (1994) | 1992-1993 | 399 | 16% |
| | | | |
| Di Pietro, et al. (1992) | 1991-1992 | 179 | 35% |
| | | | |
| Dubowitz et al. (1992) | 1989-1990 | 99 | 28% |



“ . . . many of the anatomical findings that were reported as abnormal in the earlier studies are now considered by clinicians and researchers to be nonspecific genital variations.” p. 648

Hager, et al., 2002



Current Accepted Rate of Diagnostic Medical Findings

In non-acute physical examinations (> 72 hours from abuse), diagnostic medical findings are seen in only 4-5% of children reporting CSA.

Citations: Hager, et al, 2002
 Adams, 2003
 Berkoff, et al., 2008

Summary of Reliability Studies

| <u>Study</u> | <u>Sample</u> | <u>Case Material Presented</u> |
|---------------------------------------------------|--------------------------------------------------------------|--------------------------------------------------------------------------------------|
| Finlayson & Koocher (1991) | 269 pediatric psychologists “who worked with children” | 2 page vignettes |
| Hershkowitz, Fisher, Lamb & Horowitz (2007) | 42 Israeli child forensic evaluators | Written transcripts of single interview |
| Horner, Guyer & Kalter (1992, 1993a, 1993b) | 129 “mental health professionals” | Comprehensive oral summary of 1 case involving a 3-year-old in custody case |

Summary of Reliability Studies, continued

| | | |
|-----------------------------|------------------------------------------------------------------------------------------------------|------------------|
| Jackson & Nuttall (1993) | 656 professionals “with varying degrees of experience in child sexual abuse” | 1 page vignettes |
| McGraw & Smith (1992) | Undisclosed number of CPS workers who investigated 18 cases of alleged sexual abuse in 1986 and 1987 | Case files |

Summary of Reliability Studies, continued

| | | |
|--------------------------------------|----------------------------------------------------------------------|-------------------------------------|
| Realmuto, Jensen, & Wescoe (1990) | 1 “senior female child psychiatric resident” | 29-minute anatomical doll interview |
| Realmuto & Wescoe (1992) | 14 professionals with at least 10 years of experience with CSA cases | 20-minute anatomical doll interview |



Critique of Seven Reliability Studies

Samples with limited or no forensic experience:

- Finlayson & Koocher (1991)
- Horner, et al. (1992, 1993)
- Jackson & Nuttall (1993)
- Realmuto, et al. (1990)



Critique of Seven Reliability Studies

Lacking ecological validity:

- Finlayson & Koocher (1991)
- Hershkowitz, et al. (2007)
- Jackson & Nuttall (1993)
- Realmuto, et al. (1990)
- Realmuto & Wescoe (1992)

- McGraw & Smith (1992)



The **CHIC** Forensic Evaluation Model

C Comprehensive

H Hypothesis testing

I Idiographic (vs. Nomothetic)

C Corroborative

Everson & Faller (in press)
Base rates, multiple indicators
and comprehensive forensic
evaluations, *Journal of Child
Sexual Abuse*, 21 (1).



Unintended Consequence

- With minimal planning, perpetrators could sexually exploit children and teenagers with impunity.



The New Worry



Study of Sensitivity and Specificity Bias

Research questions:

- Can sensitivity/specificity biases be reliably assessed?
- Are such biases predictive of professional judgments?

Everson, Sandoval,
Berson & Crowson
(in preparation)



Research Sample

| | |
|-----------------------------------|-----|
| Child Protective Services (CPS) | 492 |
| Law Enforcement (LE) | 281 |
| Child Forensic Evaluators (CFE) | 40 |
| Child Forensic Interviewers (CFI) | 212 |
| Medical Personnel (MED) | 44 |



Research Sample, continued

| | |
|--------------------|-----|
| Mental Health (MH) | 167 |
|--------------------|-----|

| | |
|-----------------------|----|
| Victim Advocates (VA) | 65 |
|-----------------------|----|

| | |
|------------------------------------|----|
| Assistant District Attorneys (ADA) | 80 |
|------------------------------------|----|

| | |
|--------------------------------|----|
| District Court Attorneys (ATT) | 49 |
|--------------------------------|----|

| | |
|-----------------------------|----|
| District Court Judges (JUD) | 35 |
|-----------------------------|----|



Child Forensic Attitude Survey (CFAS)

Brief attitude survey:

- Sensitivity Emphasis
- Specificity Emphasis
- Level of Skepticism

Decision-Making Exercises

| Case Type | Description | n |
|-------------------|--------------------|----------|
| Highly Probable | 2 ½ page vignette | 213 |
| Probable | 3 hour exercise | 128 |
| Ambiguous #1 | 8 page case review | 325 |
| Ambiguous #2 | 1 page vignette | 291 |
| Improbable | 1 page vignette | 295 |
| Highly Improbable | 2 page vignette | 213 |

% “Likely True” Ratings by Attitude Group

| Case Type | Sensitivity | Balanced | Specificity |
|-------------------|-------------|----------|-------------|
| Highly probable | 95 | 95 | 80 |
| Probable | 91 | 82 | 64 |
| Ambiguous #1 | 67 | 48 | 33 |
| Ambiguous #2 | 60 | 41 | 31 |
| Improbable | 23 | 14 | 7 |
| Highly improbable | 2 | 0 | 1 |



So where are we?