Interpreting Reasonable Suspicion: A National Sample of Early Childhood Professionals

Benjamin H. Levi MD PhD;[1] Nicole Verdiglione, MA;[2] Alysse Loomis, PhD;[3] Erik Lehman, MSc;[4] Whitney Barnett, PhD;[5] Roberta Andrade;[6] Richard Fiene, PhD.[7]

1. BH Levi, MD PhD, Professor, Departments of Humanities & Pediatrics, Penn State College of Medicine, 500 University Drive, Hershey, Pennsylvania; ORCID #: 0003-2380-7109. bhlevi@psu.edu
2. N Verdiglione, MA, Director of Research, iLookOut Project, Penn State College of Medicine, Hershey, Pennsylvania, nverdiglione@pennstatehealth.psu.edu
3. A Loomis, PhD, Associate Professor, Department of Social Work, University of Utah, alysse.loomis@utah.edu.
4. E Lehman, MSc, Research Statistician, Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania
5. W Barnett, Assistant Professor, Meharry Medical College, Department of Public Health, Nashville, Tennessee, whitney.barnett@mmc.edu.
6. R Andrade, PhD, Postdoctoral Fellow, iLookOut Project, Penn State College of Medicine, Hershey, Pennsylvania, rchardulorein a@pennstatehealth.psu.edu.
7. R Fiene, Professor Emeritus, Penn State University, rfiene@rikinstitute.com.

Short Title: Interpreting Reasonable Suspicion: A National Sample

Corresponding Author: Benjamin H. Levi

Abbreviations: CM (Child Maltreatment)

Key Words: Child maltreatment; Child abuse, Reasonable suspicion; Mandated reporting

Word Count: 2,981 (including headers)

Key Points

Question: How do staff at Head Start programs nationwide interpret the threshold for reporting possible child maltreatment (a.k.a., "reasonable suspicion")?

Findings: There was wide heterogeneity in how "reasonable suspicion" was interpreted. Among the 9,703 participants, the mean threshold was 54.5% likelihood, which suggests that child maltreatment must be more likely than not for there to be reasonable suspicion.

Meaning: This national study demonstrates significant variability in how early childhood professionals interpret "reasonable suspicion" –which raises concerns about the reliability and fairness of mandated reporting policies that rely on poorly defined thresholds.

ABSTRACT

Importance

Amid growing concern about the efficacy and fairness of child welfare systems, there is increased scrutiny of when and why people report "possible" child maltreatment (CM). Because unwarranted reports waste limited resources and only a fraction of children/families benefit from being reported, a poorly defined threshold for reporting negatively impacts child safety, family well-being, and resource allocation.

Objective

To examine how staff at Head Start programs nationwide interpret the threshold for reporting possible CM (a.k.a., "reasonable suspicion") and examine possible associations with knowledge and attitudes about CM and its reporting.

Design, Setting, and Participants

This national observational study of Head Start staff enrolled 10,670 English-speaking adults (≥18 years). Of these, 9,703 (90.9%) from 49 states and 6 US territories completed the iLookOut training and all pre/post training measures.

Exposure

Completion of the evidence-based, online, iLookOut mandated reporter training.

Main Outcomes and Measures

The primary outcome was likelihood of CM (0%–100%) that would warrant a report to child protective services. Findings were also examined for associations with demographic characteristics, prior training/reporting experience, and knowledge and attitudes about CM and its reporting.

Results

Before training, the mean likelihood constituting reasonable suspicion was 54.5% (range: 1%–100%) but with considerable variability (more than one-quarter of participants posited thresholds below 30%, another quarter 40-60%, and the remainder between 70-100%). Lower thresholds were associated with higher education and prior reporting experience (albeit with small effect sizes), as was greater knowledge about CM and its reporting ($p<.001$). Higher thresholds were associated ($p<.001$) with concerns about legal liability, parental backlash, child removal, the challenge of assessing risk, and distrust of child protective services. After training, the mean threshold increased slightly to 56.6% ($p<.001$), but variability persisted. Training attenuated all correlations between attitudes and thresholds. Both pre- and post-training, significant associations ($p<.001$) were found with participants' self-identified race.

Conclusions and Relevance

This national study demonstrates significant variability in how early childhood professionals interpret "reasonable suspicion" –though the mean response supported a "more likely than not" threshold. These findings raise concerns about the reliability and fairness of mandated reporting policies that rely on an undefined threshold and underscore the need for clearer, evidence-informed standards.

349/350 total

Introduction

Amid growing concern about the efficacy and fairness of child welfare systems in the United States,[1-4] there has been increased scrutiny of 'mandated reporting' of suspected child maltreatment (**CM** –i.e., physical, sexual, and emotional abuse, or neglect).[5-8] Reports of suspected CM serve as the entry point for most interactions with child welfare, affecting one-third of all children and over half of Black children in the U.S.[9] Most reports come from mandated reporters (clinicians, educators, law enforcement, early childhood providers, and others who work with children) who are legally required to report suspected CM once a statutory threshold –generally referred to as "reasonable suspicion"– has been met.[10] However, there is little interpretive guidance for what counts as reasonable suspicion and initial evidence that mandated reporters disagree about how likely CM must be to warrant a report.[11-15] This paper reports the first national examination of how 'reasonable suspicion' is interpreted by a workforce that takes care of children age 5 and younger, who are the most vulnerable to CM.[16]

Context for understanding the concern:

CM occurs in every community and socio-economic class. There are >1,700 CM-related deaths and >500,000 confirmed victims of CM each year in the U.S.,[16] with the true incidence estimated to be 2-3 times that number.[17-20] CM constitutes a major public health problem whose physical, developmental, and mental health consequences extend into adulthood,[21-27] costing the US economy an estimated $450 billion per year.[28]

That said, reporting suspected CM is itself associated with a range of negative consequences, from family stress and disruption,[4,29-31] mental health problems,[32,33] risk-taking behaviors,[34] and

decreased well-being[35] to disproportionate enmeshment of marginalized communities for whom systemic racism and economic inequalities are already endemic.[3,4,36,37] Combined with substantiation rates of <15%[16] such harms raise questions as to whether current reporting practices yield a net positive outcome.[38,39]

Whatever one's beliefs about the child welfare system, two things are clear. First, many children and their families receive little to no benefit from being reported.[8,40] Second, the limited resources available for improving the lives of children experiencing maltreatment could be put to better use if we could decrease the volume of unnecessary reports that clog the system, diverting time and attention from those truly in need.[4,41,42] Increasing the signal-to-noise ratio for these reports would not only help concentrate resources on those children experiencing harm but also promote justice by minimizing the collateral damage to families where no maltreatment is occurring –potentially decreasing both false positives and false negatives.[43]

Clarifying When to Report

State-level variations in mandated reporter policies have been linked to differences in rates of substantiation (a proxy for accurate detection of CM) for children from marginalized backgrounds across maltreatment types;[44] and variations in statutory language have been associated with both how people interpret reporting thresholds[45,46] and substantiation rates for reports made to child protective services.[47] Regardless of state-specific wording, however, the threshold for reporting is rarely explained and at best normatively framed (e.g., when a reasonable person with the same information and vantage point would reach the conclusion that reasonable suspicion exists).[48]

But if in fact there is no normative standard, no agreed-upon general consensus about what reasonable suspicion means, inconsistencies will abound not only for when a child or family should (or shouldn't) be reported but also for determining when mandated reporters have met their professional and legal responsibilities for protecting children. Moreover, without a clear understanding of reasonable suspicion, it is challenging to reduce noise (i.e., inappropriate reporting) that both obscures the signal and can increase false positives. Compounding this problem is the commonly shared directive: 'when in doubt, report'[49] –not least because the more a person appreciates the myriad ways CM can present, the more often they will (quite legitimately) have doubt. The suddenly withdrawn child could be experiencing CM but just as easily could be expressing anxiety, reacting to the death of a pet or grandparent, or simply processing some non-CM-related issue. And though increased knowledge about signs and symptoms of CM can help a person more accurately identify CM, such knowledge is just as likely to fuel doubt as to dispel it.

A reporting threshold that is undefined and haphazardly interpreted is problematic because it fuels both under-reporting and over-reporting. Unreported CM can have profound ramifications, particularly for young children with neither the language nor physical ability to protect themselves.[50,51] Over-reporting not only dilutes scarce critical resources that should be directed towards children truly being harmed[8,40] but also leads to unnecessary interactions with child welfare, which can have negative and long-lasting consequences.[4,41,42]

Prior research with discrete groups of mandated reporters (including child abuse pediatricians) has shown inconsistent and widely varying interpretations of what it means to have reasonable suspicion of CM.[11,13,14,52,53] But to date there has been no examination of this threshold with a national sample of mandated reporters. The current study examines this issue with Head Start staff across the country who provide early care and education to children through age 5[54-56] and are known for their high professional standards and robust professional development.[57-60]

Methods

Participants and Procedures

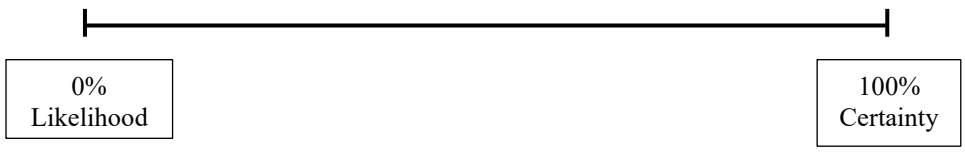
The data reported here were collected (11/1/20–6/30/25) as part of a study examining the impact of iLookOut for Child Abuse,[61] the online mandated reporter training recommended for Head Start staff. Participants (≥18 years of age, English-speaking, and Head Start staff) provided informed consent, completed demographic and study measures prior to the ~3-hour training, and repeated study measures post-training –all online in one sitting or across multiple sessions.

The study was approved by Penn State's Institutional Review Board, registered at the National Institutes of Health #NCT 02225301, and received a waiver of the need to collect a written signature (in accord with federal regulation 45 CFR part 46.116[d]). No protected health information was accessed for this study and the only conflict of interest involved Penn State policy regarding licensed products (see COI statement). The lead author (BHL) had responsibility for data integrity and accuracy of the analysis.

The primary study measure (see below) was adapted from previous studies, pilot tested for face validity, and tested via a cognitive interview process previously described.[62] The term 'child abuse' was defined as physical abuse, sexual abuse, psychological abuse, as well as neglect and was chosen over 'child maltreatment' because it is more commonly used by the lay public. To focus on the concept of reasonable suspicion, the question-item scenario did not mention the type of abuse, gender, age, or other demographic features –because such features are known to significantly influence assessments of CM probability.[63]

The 'estimated numerical probability' scale was chosen based on both its prior use[11,52,53] and the serious drawbacks of other scales.[64-68] In particular, common language terms denoting probability (possible, likely, very likely, etc.) are known to have widely variable interpretations[12,69,70] and rank ordering for probability has been shown to have low inter-rater as well as intra-rater reliability.[14,52,66,67] Accordingly, we used the following as our primary study measure:

Imagine that you encountered a child who has an injury or condition that might have been caused by abuse. You have gathered as much information as possible from the sources readily available. In order for you to feel you had reasonable suspicion that the child's injury or condition was caused by abuse, how great would that likelihood of abuse have to be? (That is, would you need to feel there was a 99% likelihood that abuse occurred before you felt you had reasonable suspicion? A 1% likelihood? Or something in between?)



Descriptive statistics were generated, including means, medians, and standard deviations for continuous variables and frequency tables for discrete variables. Associations between (self-reported) demographic factors and reasonable suspicion were characterized by using contingency table analysis; significance levels were determined by Pearson's chi-squared statistics and two-sample t tests. Comparisons between estimated probability and both total knowledge score and attitudinal items used quantile regression analysis of the median (due to non-normally distributed responses).

Results

Demographics

Of 10,670 Head Start staff who registered, 90.9% (9,703) completed the training and pre- and post-tests. Respondents predominantly identified as female (95.7%), full-time employees (91.1%), and teachers or teachers' aides (65.2%) but were diverse in terms of race/ethnicity (58.3% White, 25.1% Black, and 1.8% Asian; with 28.0% identifying as Hispanic), age (26.7% age 18-29; 37.1% age 30-44; 36.3% age >44), parental status, highest education, years working, and location (spanning 49 states and 6 U.S. Territories). (See Table 1 for demographics) Most respondents had never themselves reported suspected CM (68%) or been involved in a team decision to report (62%) and few had themselves reported (8.9%) or been part of a team decision to report (13.1%) more than twice. (See Table 2)

Pre-intervention Likelihood Constituting Reasonable Suspicion

Prior to completing iLookOut, the 'mean likelihood of CM' (aka, **threshold**) Head Start staff identified as constituting *reasonable suspicion* was 54.5%. Over a quarter of respondents

posited a threshold of ≤30% likelihood, another quarter posited 40-60%, and four out of ten set the threshold between 70-100% likelihood. (See **Fig. 1**)

Statistically significant differences ($p<.001$) were found between the threshold set by Head Start staff who identified as White (51.0%) versus Hispanic (57.3%), Black (58.3%), and Asian (61.2%), however, the effect sizes for those differences (Cohen's d 0.21, 0.25, and 0.28 respectively) demonstrated significant heterogeneity within each of these groups. Thresholds were also lower ($p<.001$) among Head Start staff with prior training about CM (52.8%) versus no training (58.1%) and those who had previously reported suspected CM (Yes-50.6%, No-57.2%).

Significant differences also were found with respect to age (59.1% for 18-29 year-olds, 56.2% for 30-44 year-olds, 51.2% for >44 year-olds); marital status (59% for singles, 51.1% for widowed, 52.5% for married or cohabitating, 50.6% for separated or divorced); age of oldest child (ranging from 51.4% to 57.8%); type of Head Start program (ranging from 50.7% to 58.7%); and size of Head Start program (ranging from 49.1% to 57.5%). However, all the effect sizes were small (Cohen's d <.28) with most being very small (Cohen's d <.10). No statistically significant differences were found for gender, parental status, or employment status.

The one area where pre-intervention differences were not negligible was 'highest level of education' (see **Table 3**). Here, thresholds ranged from 47.7% to 58.8%, *more* education directly correlated with *lower* thresholds, and the largest effect sizes were seen comparing staff with graduate degrees to those with Associate's degrees or with 'high school or less' (Cohen's d 0.3 and 0.37, respectively).

Post-intervention Likelihood Constituting Reasonable Suspicion

After completing iLookOut, the threshold Head Start staff identified as constituting *reasonable suspicion* was 56.6%, which was significantly ($p<.001$) but only slightly higher than the pre-intervention threshold of 54.5%. Similar to pre-intervention, responses were wide-ranging, though fewer post-intervention responses posited thresholds $\leq$30% likelihood. (See **Fig. 1**) Statistically significant differences ($p<.001$) were found between thresholds set by those identifying as White (54.1%) versus Hispanic (59.8%), Black (58.5%), or Asian (58.9%), however, effect sizes for those differences (Cohen's d of 0.21, 0.09, and 0.11, respectively) again demonstrated significant heterogeneity within each of these groups.

Statistically significant differences ($p<.001$) were also identified between staff with prior training about CM (55.4%) versus no training (59.4%); across different ages (60.4% for 18-29 year-olds, 58.2% for 30-44 year-olds, 53.8% for >44 year-olds); marital status (60% for singles, 53.4% for widowed, 55.3% for married or cohabitating, 54.2% for separated or divorced); age of oldest child (ranging from 53.6% to 59.5%); highest level of education (ranging from 52.8% to 60%); having previously ever reported suspected CM (Yes-53.2%, No-59.3%); type of Head Start program (ranging from 55.6% to 59.0%); and size of Head Start program (ranging from 53.4% to 58.9%). In all cases, however, pre-to-post changes in likelihood threshold had effect sizes <.25, with most being <.10. As with pre-intervention responses, no statistically significant differences were found for gender, parental status, or employment status.

Correlations with Knowledge and Attitudes

Notably, Head Start staff with higher knowledge scores set significantly ($p<.001$) lower thresholds for what counts as reasonable suspicion both pre- and post-intervention, though the differences were smaller post-intervention. By contrast, higher thresholds were significantly ($p<.001$) correlated with concern about being held legally liable for mistaken reports, fear of backlash from parents, doubt that child protective services will respond effectively, concern that a child might be removed, and belief that it's too hard to be sure CM occurred (see **Table 4** for full pre/post correlations).

Discussion

This national study of nearly 10,000 early childhood professionals demonstrates remarkably wide variability in how people interpret the threshold that is supposed to trigger reports of suspected child maltreatment (CM). These findings raise questions about not only the efficacy of using reasonable suspicion as the "standard" for mandated reporting but also the justifiability of policies that rely on it –whether to protect at-risk children or hold adults accountable for failing to report suspected CM. The implications of these findings are amplified by the fact that Head Start staff are considerably more prepared than most mandated reporters to fulfill their responsibility to protect children[71-73] and that this variability persists even after completing a rigorously developed, evidence-based, online mandated reporter training.[62,74,75] Similar results have been found with other mandated reporters,[14] including child abuse pediatricians.[52]

That being said, the correlations found in the current study may help us better understand what drives some of this variability. Notably, individuals with greater knowledge about CM and its reporting set lower thresholds for when to report at-risk children to child protection agencies.

This correlation may reflect a clearer understanding that reporting does not require proof; or that reports are screened by trained professionals prior to any action being taken; or that with a better understanding of its prevalence and risk factors, weaker signals for CM seem worth reporting. The somewhat lower post-intervention correlation with knowledge may be related to iLookOut's emphasis on critical thinking, which prompts learners to consider a range of other factors when deciding when to report. What remains unclear is how these variable thresholds correspond with subsequent identification of CM (i.e., true positives).

Not surprisingly, individuals who feared retaliation from families, those who expressed lower confidence in the effectiveness of child protection agencies, and those who worried that reporting could negatively impact a child's wellbeing (due to either their parents' or agency intervention) set higher thresholds for the likelihood of CM that should warrant a report. But again, completing iLookOut mitigated the strength of those correlations (see **Table 3**), suggesting that a well-designed training may be able to (i) counter-balance negative perceptions of the child welfare system by highlighting the risks of not reporting suspected CM and thereby (ii) disentangle how people gauge reasonable suspicion. If so, this is encouraging because despite its well-documented limitations,[2,4,76,77] the current child welfare system remains the primary mechanism for identifying and responding to CM.

Interestingly, the threshold for what counts as reasonable suspicion was not correlated (either pre- or post-intervention) with a person's confidence to identify signs and symptoms of CM, sense of preparedness to report suspected CM, desire to fulfill professional responsibility, or belief as to whether reporting improves a child's safety or leads to helpful services. This may

suggest that the threshold for reporting is not correlated with people's doubt about either their abilities or the potential benefit of reporting –though it may also reflect the large heterogeneity in people's views. Future research could examine whether likelihood thresholds or people's attitudes regarding their capacity and preparedness to report are differentially related to accuracy of reporting (e.g., substantiated reports) –as this could help identify useful targets for efforts aiming to minimize overreporting.

It is notable that completing iLookOut resulted in statistically significant increases in post-intervention thresholds set by Head Start staff who identified as White (3.1% increase) or Hispanic (2.5% increase) but not Black (0.1%) or Asian (-2.3%) –though within-group variability far exceeded between-group differences. Completing iLookOut also attenuated the association between higher thresholds for reasonable suspicion and concerns about reporting suspected CM (backlash from parents, harm to or removal of the child, or being held legally liable for mistaken reports) as well as beliefs that reporting is a waste of time or that it is too hard to be sure that CM occurred.

Taken together, these findings point to the need for a standard that can be accurately defined, effectively taught and implemented, and relied upon as the measure for when the possibility of CM must be reported. The present national dataset also confirms prior findings that most people set the threshold for reasonable suspicion above 50% –which implies that CM must be more likely than not before a report is warranted. Arguably, the "appropriate" threshold is the one that best 1) identifies CM (i.e., true positives), 2) minimizes unwarranted reports (with their

accompanying harms and costs), and 3) utilizes the limited resources available for protecting children and supporting their well-being.

Admittedly, the matter of reasonable suspicion is but one of many problems with the child welfare system, including the broader relationship between structural inequities and conditions/behavior that threaten children's well-being. But as the trigger and entry point to the child welfare system, a better standard (read: consistent, just, effective, evidence-driven, and implementable) for reasonable suspicion is needed for all mandated reporters –along with evidence-based training that can be broadly and effectively implemented.

Limitations

This study measured self-reported conceptual thresholds rather than observed reporting behavior. The hypothetical scenario may not fully capture real-world decision-making. Numeric probability may not perfectly operationalize suspicion. The sample was limited to Head Start staff and may not generalize to other mandated reporters.

Conclusion

In this national study, early childhood professionals demonstrated marked variability in how they interpreted "reasonable suspicion." iLookOut, an evidence-based training resulted in small changes but did not eliminate the heterogeneity. Both pre- and post-training, respondents on average indicated that CM would have to be more likely than not (>50%) before they felt they had reasonable suspicion to report. These findings suggest that statutory standards for reporting

suspected CM lack a shared operational meaning, which may have implications for the consistency, equity, and fairness of mandated reporting practices.

ACKNOWLEDGEMENTS

Contributor Statements
Dr. Benjamin Levi conceptualized and designed the study, supervised and coordinated data collection and statistical analysis, drafted the initial manuscript, critically reviewed and revised the manuscript.
Nicole Verdiglione contributed to the study design, supervised and coordinated data collection and statistical analysis, and critically reviewed and revised the manuscript.
Dr. Elysse Loomis contributed to data interpretation and critically reviewed and revised the manuscript.
Erik Lehman oversaw statistical analysis of the study data and reviewed and revised the manuscript.
Whitney Barnett contributed to data interpretation and critically reviewed and revised the manuscript.
Dr. Roberta Andrade contributed to data interpretation and critically reviewed and revised the manuscript.
Dr. Richard Fiene contributed to data interpretation and critically reviewed and revised the manuscript.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Data availability: Aggregate, deidentified will be available upon reasonable request.

Conflict of Interest Statement: In accordance with Penn State University's intellectual property policy, Dr. Levi receives royalties related to iLookOut-related products that have been licensed for commercial use. The licensing and royalty arrangements are managed by Penn State and all iLookOut-related research and publications are in compliance with institutional policies. Dr. Levi had no access to the study's raw data and was not directly involved in the actual analysis of study data.

AI Contributions: ChatGPT was used to review the completed manuscript to check for adherence to journal submission requirements.

Funding: This parent study for this research was funded by the National Institute for Child Health and Human Development (NICHD), grant number 5R01HD088448-10A1.

Compensation:
Dr. Levi received salary support from NICHD as the Principal Investigator for the parent grant.
Ms. Verdiglione received salary support from the NICHD parent grant.
Mr. Lehman received salary support from the NICHD parent grant.
Dr. Barnett received salary support from the NICHD parent grant.

References

1. Raz M. Abusive policies: how the American child welfare system lost its way. UNC Press Books; 2020.
2. Dettlaff AJ, Weber K, Pendleton M, Boyd R, Bettencourt B, Burton L. It is not a broken system, it is a system that needs to be broken: the upEND movement to abolish the child welfare system. J Public Child Welfare. 2020;14(5):500-517.
3. Baughman C, Coles T, Feinberg J, Newton H. The surveillance tentacles of the child welfare system. Columbia J Race Law. 2021;11(3):501-532.
4. Roberts D. Torn apart: how the child welfare system destroys Black families and how abolition can build a safer world. Basic Books; 2022.
5. McTavish JR, Kimber M, Devries K, Colombini M, MacGregor JC, Wathen N, et al. Children's and caregivers' perspectives about mandatory reporting of child maltreatment: a meta-synthesis of qualitative studies. BMJ Open. 2019;9(4):e025741.
6. Edwards F. Family surveillance: police and the reporting of child abuse and neglect. RSF. 2019;5(1):50-70.
7. Fong K. Neighborhood inequality in the prevalence of reported and substantiated child maltreatment. Child Abuse Negl. 2019;90:13-21.
8. Rosenberg R, Williams SC, Martinez V, Ball JC. Mandated reporting policies and the detection of child abuse and neglect. Child Youth Serv Rev. 2024;159:107499.
9. Kim H, Wildeman C, Jonson-Reid M, Drake B. Lifetime prevalence of investigating child maltreatment among US children. Am J Public Health. 2017;107(2):274-280. doi:10.2105/ajph.2016.303545
10. Sedlak AJ, Heaton L, Evans M. Trends in child abuse reporting. In: Krugman RD, Korbin JE, eds. Handbook of child maltreatment. Springer; 2022:3-34.
11. Levi BH, Brown G. Reasonable suspicion: a study of Pennsylvania pediatricians regarding child abuse. Pediatrics. 2005;116(1):e5-e12.
12. Levi BH, Brown G, Erb C. Reasonable suspicion: a pilot study of pediatric residents. Child Abuse Negl. 2006;30(4):345-356.
13. Levi BH, Dellasega C, Boehmer S. What is reasonable suspicion of child abuse? J Public Child Welfare. 2012;6(5):569-589.
14. Levi BH, Crowell K, Walsh K, Dellasega C. How childcare providers interpret "reasonable suspicion" of child abuse. Child Youth Care Forum. 2015;44(6):875-891.
15. Levi BH, Portwood SG. Reasonable suspicion of child abuse: finding a common language. J Law Med Ethics. 2011;39(1):62-69.
16. US Department of Health and Human Services, Administration for Children and Families, Children's Bureau. Child maltreatment 2024. Published 2026. Accessed April 9, 2025. https://acf.gov/cb/data-research/child-maltreatment
17. Pérez-Fuentes G, Olfson M, Villegas L, Morcillo C, Wang S, Blanco C. Prevalence and correlates of child sexual abuse: a national study. Compr Psychiatry. 2013;54(1):16-27.
18. Finkelhor D, Turner HA, Shattuck A, Hamby SL. Prevalence of childhood exposure to violence, crime, and abuse: results from the National Survey of Children's Exposure to Violence. JAMA Pediatr. 2015;169(8):746-754.
19. Mills R, Kisely S, Alati R, Strathearn L, Najman J. Self-reported and agency-notified child sexual abuse in a population-based birth cohort. J Psychiatr Res. 2016;74:87-93.

20. Kisely S, Najman JM, Strathearn L. Childhood maltreatment and intimate partner violence: 30-year follow-up of a birth cohort study. J Aggress Maltreat Trauma. 2024;33(7):844-863.
21. Gilbert R, Widom CS, Browne K, Fergusson D, Webb E, Janson S. Burden and consequences of child maltreatment in high-income countries. Lancet. 2009;373(9657). doi:10.1016/S0140-6736(08)61706-7
22. Spratt EG, Friedenberg S, LaRosa A, et al. The effects of early neglect on cognitive, language, and behavioral functioning in childhood. Psychology. 2012;3(2). doi:10.4236/psych.2012.32026
23. Cowell RA, Cicchetti D, Rogosch FA, Toth SL. Childhood maltreatment and its effect on neurocognitive functioning: timing and chronicity matter. Dev Psychopathol. 2015;27(2). doi:10.1017/S0954579415000139
24. Shields ME, Hovdestad WE, Gilbert CP, Tonmyr LE. Childhood maltreatment as a risk factor for COPD: findings from a population-based survey. Int J Chron Obstruct Pulmon Dis. 2016;11. doi:10.2147/COPD.S107549
25. Muniz CN, Fox B, Miley LN, DeLisi M, Cigarran GP, Birnbaum A. The effects of adverse childhood experiences on internalizing versus externalizing outcomes. Crim Justice Behav. 2019;46(4). doi:10.1177/0093854819826213
26. Lopes S, Hallak JEC, Sousa JPM, Osório FL. Adverse childhood experiences and chronic lung diseases in adulthood: a systematic review and meta-analysis. Eur J Psychotraumatol. 2020;11(1). doi:10.1080/20008198.2020.1720336
27. Soares AG, Howe LD, Heron J, et al. How does childhood maltreatment influence cardiovascular disease? Int J Epidemiol. 2021. doi:10.1093/ije/dyab085
28. Peterson C, Florence C, Klevens J. The economic burden of child maltreatment in the United States, 2015. Child Abuse Negl. 2018;86. doi:10.1016/j.chiabu.2018.09.018
29. Besharov DJ. Responding to child sexual abuse: the need for a balanced approach. Future Child. 1994;4(2):135-155.
30. Unrau YA, Seita JR, Putney KS. Former foster youth remember multiple placement moves: a journey of loss and hope. Child Youth Serv Rev. 2008;30(11):1256-1266.
31. Sankaran V, Church C, Mitchell M. A cure worse than the disease: the impact of removal on children and their families. Marq Law Rev. 2018;102:1161.
32. Bederian-Gardner D, Hobbs SD, Ogle CM, et al. Instability in the lives of foster and nonfoster youth: mental health impediments and attachment insecurities. Child Youth Serv Rev. 2018;84:159-167.
33. McGuire A, Cho B, Huffhines L, et al. The relation between dimensions of maltreatment, placement instability, and mental health among youth in foster care. Child Abuse Negl. 2018;86:10-21.
34. Stott T. Placement instability and risky behaviors of youth aging out of foster care. Child Adolesc Soc Work J. 2011;29(1):61-83.
35. Rubin DM, O'Reilly ALR, Luan X, Localio AR. The impact of placement stability on behavioral well-being for children in foster care. Pediatrics. 2007;119(2):336-344.
36. Dettlaff AJ, Boyd R. Racial disproportionality and disparities in the child welfare system. Ann Am Acad Pol Soc Sci. 2020;692(1):253-274.
37. Barth RP, Berrick JD, Garcia AR, Drake B, Jonson-Reid M, Gyourko JR, et al. Research to consider while effectively re-designing child welfare services. Res Soc Work Pract. 2022;32(5):483-498.

38. Wulczyn F, Barth RP, Yuan YYT, Harden BJ, Landsverk J. Beyond common sense: child welfare, child well-being, and the evidence for policy reform. Routledge; 2017.
39. Melton GB. Mandated reporting: a policy without reason. Child Abuse Negl. 2005;29(1):9-18.
40. Raz M, Gupta-Kagan J, Asnes AG. Using child abuse specialists to reduce unnecessary child protective services reports and investigations. JAMA Pediatr. 2023;177(12):1249-1250.
41. Merritt DH. How do families experience and interact with CPS? Ann Am Acad Pol Soc Sci. 2020;692(1):203-226.
42. Evangelist M, Thomas MM, Waldfogel J. Child protective services contact and youth outcomes. Child Abuse Negl. 2023;136:105994.
43. Lyon TD. Child maltreatment, the law, and two types of error. Child Maltreat. 2023;28(3):403-406. doi:10.1177/10775595231176454
44. Williams SC, Rosenberg R, Martinez V, Ball JC. Are mandated reporting policies contributing to disparities in child welfare? Fam Soc. 2025;106(2):447-462.
45. Blacker DM. Reporting of child sexual abuse: the effects of varying definitions of reasonable suspicion on psychologists' reporting behavior [dissertation]. California School of Professional Psychology; 1998.
46. Flieger CL. Reporting child physical abuse: the effects of varying legal definitions of reasonable suspicion on psychologists' reporting behavior [dissertation]. California School of Professional Psychology; 1999.
47. Piersiak HA, Levi BH, Humphreys KL. Statutory threshold wording is associated with child maltreatment reporting. Child Maltreat. 2023;28(3):517-526.
48. Mandated Reporter. What qualifies as reasonable suspicion of child abuse? Accessed April 9, 2025. https://mandatedreporter.com/blog/what-qualifies-as-reasonable-suspicion-of-child-abuse/
49. State and institutional guidance on mandated reporting (multiple sources). Accessed April 9, 2025. Includes: Texas Elementary Principals and Supervisors Association; Illinois Department of Children and Family Services; Wisconsin Child Welfare Professional Development System; Arizona Department of Child Safety; Kentucky Cabinet for Health and Family Services; Princeton University; Pennsylvania State University; Wyoming Department of Education; Yale University.
50. Sylvestre A, Bussières ÈL, Bouchard C. Language problems among abused and neglected children: a meta-analytic review. Child Maltreat. 2016;21(1):47-58.
51. Sylvestre A, Di Sante M, Julien C, Bouchard C, Mérette C. Developmental trajectories of speech and language in neglected children aged 3 to 5 years. Child Abuse Negl. 2023;146:106448.
52. Levi BH, Crowell K. Child abuse experts disagree about the threshold for mandated reporting. Clin Pediatr (Phila). 2011;50(4):321-329.
53. Crowell K, Levi BH. Mandated reporting thresholds for community professionals. Child Welfare. 2012;91(1):35-54.
54. Office of Head Start. Head Start programs overview. Accessed April 9, 2026. https://headstart.gov/programs/article/head-start-programs
55. Hines JM. An overview of Head Start program studies. J Instr Pedagog. 2017;18.
56. Bailey MJ, Sun S, Timpe B. Prep school for poor kids: the long-run impacts of Head Start on human capital and economic self-sufficiency. Am Econ Rev. 2021;111(12):3963-4001.

57. Institute of Medicine; National Research Council. The early childhood care and education workforce: challenges and opportunities. National Academies Press; 2012.
58. Doromal JB, Bassok D, Bellows L, Markowitz AJ. Hard-to-staff centers: exploring center-level variation in child care teacher turnover. Early Child Res Q. 2022;61:170-178.
59. Amadon S, Lin YC, Padilla C. Turnover in the center-based child care workforce. Office of Planning, Research, and Evaluation; 2023.
60. Carr RC, Vernon-Feagans L, Burchinal MR. Head Start in low-wealth rural communities: evidence from the Family Life Project. Early Educ Dev. 2023;34(7):1590-1611.
61. Office of Head Start. Mandated reporting resources. Accessed April 9, 2026. https://headstart.gov/
62. Panlilio CC, Famularo L, Masters J, Dore S, Verdiglione N, Yang C, et al. Integrating validity evidence to revise a child abuse knowledge test. Am J Eval. 2022;43(4):559-583.
63. Kahneman D. Thinking, fast and slow. Macmillan; 2011.
64. Hakel MD. How often is often? Am Psychol. 1968;23(7):533.
65. Rapoport A, Wallsten TS, Erev I, Cohen BL. Revision of opinion with verbally and numerically expressed uncertainties. Acta Psychol (Amst). 1990;74(1):61-79.
66. Teigen KH. Studies in subjective probability III: the unimportance of alternatives. Scand J Psychol. 1983;24:97-105.
67. Teigen KH. When are low-probability events judged to be probable? Acta Psychol (Amst). 1988;68:157-174.
68. Wallsten TS, Budescu DV, Zwick R. Comparing calibration and coherence of probability judgments. Manage Sci. 1993;39(2):176-190.
69. Brun W, Teigen KH. Verbal probabilities: ambiguous or context dependent? Organ Behav Decis Process. 1988;41:390-404.
70. Lichtenstein S, Newman JR. Empirical scaling of verbal probability phrases. Psychon Sci. 1967;9:563-564.
71. Green BL, Ayoub C, Bartlett JD, Von Ende A, Furrer C, Chazan-Cohen R, et al. The effect of Early Head Start on child welfare involvement. Child Youth Serv Rev. 2014;42:127-135.
72. Zhai F, Waldfogel J, Brooks-Gunn J. Estimating the effects of Head Start on parenting and maltreatment. Child Youth Serv Rev. 2013;35(7):1119-1129.
73. Wallace LE, Manz PH. Addressing the needs of infants exposed to maltreatment. Child Youth Serv Rev. 2023;146:106808.
74. Levi BH, Mundy M, Palm C, Verdiglione N, Fiene R, Mincemoyer C. An interactive online learning program on child abuse reporting. J Educ Online. 2021;18(2).
75. Humphreys KL, Piersiak HA, Panlilio CC, Lehman EB, Verdiglione N, Dore S, et al. A randomized controlled trial of mandated reporter training. Child Abuse Negl. 2021;117:105033.
76. Berrick JD, Barth RP, Drake B, Jonson-Reid M, Garcia A, Greeson JK, Gyourko J. Is child welfare oppressive? J Teach Soc Work. 2025;45(2):362-383.
77. Raz M, Gupta-Kagan J, Asnes AG. Should I call child protection?—guidelines for clinicians. JAMA Pediatr. 2024;178(11):1095-1096.

Table 1. Demographics

Characteristic	No. (%) (n=9703)
Age, y	
18-29	2588 (26.7)
30-44	3597 (37.1)
>44	3518 (36.3)
Gender	
Male	353 (3.6)
Female	9288 (95.7)
Gender nonconforming	33 (0.3)
Other	29 (0.3)
Race	
Black/African American	2432 (25.1)
White	5658 (58.3)
Asian	174 (1.8)
Other	1439 (14.8)
Ethnicity	
Hispanic or Latino	2714 (28.0)
Not Hispanic or Latino	6989 (72.0)
Parental status	
Yes	5964 (61.5)
No	3474 (35.8)
Missing	265 (2.7)
Age of youngest child	
No child	3474 (35.8)
5 years or younger	1916 (19.8)
6-17 years old	2349 (24.2)
Older than 17 years	1699 (17.5)
Missing	265 (2.7)
Marital status	
Married or cohabitating	4294 (44.3)
Separated or divorced	908 (9.4)
Widowed	247 (2.6)
Single	4084 (42.1)
Missing	170 (1.8)
Highest level of education	
High school diploma/GED	2792 (28.8)
Associate's degree or CDA	3334 (34.4)
Bachelor's degree	2696 (27.8)
Graduate degree	881 (9.1)

Years worked as an early childhood professional	
<1	1443 (14.9)
1-2	1194 (12.3)
3-5	1668 (17.2)
6-10	1672 (17.2)
11-15	1110 (11.4)
>15	2616 (27.0)
Employment basis	
Volunteer	52 (0.5)
Paid full-time	8840 (91.1)
Paid part-time	612 (6.3)
Paid seasonal	105 (1.1)
Paid substitute	94 (1.0)
Primary job responsibilities	
Teacher (hands-on)	6,326 (65.2)
Visiting Support	1099 (11.3)
Administrator	769 (7.9)
Support Staff	617 (6.4)
Classroom Support	330 (3.4)
Other	563 (5.8)
Formal training in early care and education	
Yes	7544 (77.8)
No	2159 (22.2)
Prior training on child maltreatment	
Yes	5447 (56.1)
No	4256 (43.9)
Program capacity	
<100 children	4512 (46.5)
100-199 children	2267 (23.4)
200-499 children	1827 (18.8)
≥500 children	1097 (11.3)
Kind of program	
Head Start (HS)	3864 (39.8)
Early HS	2810 (29.0)
Both HS/Early HS	2215 (22.8)
Child care program	593 (6.1)
Other	221 (2.3)
Region of US	
Northeast	1524 (15.7)
Southeast	2328 (24.0)
Midwest	2408 (24.8)

Southwest	1919 (19.8)
West	1207 (12.4)
Other	295 (3.0)
Missing	22 (0.2)

Percentages may not total 100 because of rounding.

Table 2. Prior Experience Reporting Suspected CM

	No. (%) (total=9703)
Times you were part of a team decision to report	
0	6,023 (62.1)
1-2	2,022 (20.8)
3-4	586 (6.0)
5-10	547 (5.6)
>10	149 (1.5)
Missing	376 (3.8)
Times you (personally) made a report	
0	6,621 (68.2)
1-2	1,972 (20.3)
3-4	433 (4.5)
5-10	362 (3.7)
>10	72 (0.7)
Missing	243 (2.5)

Percentages may not total 100 because of rounding.

Table 3. Educational Level and Threshold for Reasonable Suspicion

Highest Educational Level	No.	Pretraining Mean	Posttraining Mean	P value
High school or less	2,792	58.8	60.0	.03
Associate's degree or CDA	3,334	56.9	57.9	.05
Bachelor's degree	2,696	51.6	54.8	<.001
Graduate degree	881	47.7	52.8	<.001

P values were calculated using paired t tests comparing pretraining and posttraining mean thresholds for reasonable suspicion within each educational level.

Table 4. Predictors of "Reasonable Suspicion" Threshold

Predictor Variable	Effect on Threshold		Interpretation
	Pre	Post	
Knowledge score (0–25)	–3.33%$^{\alpha}$	–1.67%$^{\alpha}$	Higher knowledge lowers threshold; reduced posttraining
Early childhood professionals should be required to report suspected abuse	–1.43%$^{\alpha}$	0.00^{ns}	Stronger agreement lowers threshold; effect disappears posttraining
Reporting promotes children's long-term interests	0.00^{ns}	0.00^{ns}	No effect pre or post
Concern about being held legally liable for mistaken reports	+3.33%$^{\alpha}$	+2.00%$^{\alpha}$	Legal concerns raise threshold; reduced posttraining
Confidence identifying signs of abuse	0.00%^{ns}	0.00%^{ns}	No effect pre or post
Preparedness to report abuse	0.00%^{ns}	0.00%^{ns}	No effect pre or post
Low confidence that child protective services will respond effectively	+2.86%$^{\alpha}$	+1.25%$^{\alpha}$	Less trust raises threshold; reduced posttraining
Fear parents may harm child if reported	+3.33%$^{\alpha}$	+1.25%^{ns}	Fear raises threshold; reduced posttraining
Fear of backlash from family/others	+4.29%$^{\alpha}$	+1.43%$^{\alpha}$	Fear raises threshold; reduced posttraining
Report even if supervisor disagrees	–2.00%$^{\alpha}$	0.00^{ns}	Stronger conviction lowers threshold pre, null post
Desire to fulfill professional duty	–1.25%^{ns}	0.00^{ns}	No significant effect
Would not report if child might be removed	+4.29%$^{\alpha}$	+1.25%$^{\alpha}$	Concern raises threshold; reduced posttraining
Too hard to be sure abuse occurred	+4.29%$^{\alpha}$	+1.67%$^{\alpha}$	Concern raises threshold; reduced posttraining
Reporting is a waste of time	+3.33%$^{\alpha}$	0.00^{ns}	Cynicism raises threshold;effect disappears posttraining
Reporting improves children's safety	0.00%^{ns}	0.00%^{ns}	No effect pre or post
Reporting leads to helpful services	–1.43%^{ns}	0.00%$^{\alpha}$	Belief in helpfulness lowers threshold; disappears posttraining

$\alpha = p < 0.001$; ns = not statistically significant
Negative = lower threshold (more likely to report); Positive = higher threshold (less likely to report).

Quantile regression analysis estimates the magnitude and direction of change each 'predictor variable' exerts on the threshold for reasonable suspicion (with each predictor being modeled separately). E.g., for "Knowledge score Pre," the estimate of -3.33 means that for every 1-point increase in the pretraining knowledge score, there is a 3.33% decrease in the median response for the percentage of likelihood that constitutes reasonable suspicion. For the "Knowledge score Post," every 1-point increase in knowledge is c

Figure 1. Percent Likelihood Needed for "Reasonable Suspicion"

Figure 1. Percent Likelihood Needed for "Reasonable Suspicion"