Psychology of Violence

Cross-Informant Agreement on Mental Health Outcomes in Children With Maltreatment Histories: A Systematic Review

Elisa Romano, Kelly Weegar, Lyzon Babchishin, and Michael Saini

CITATION
http://dx.doi.org/10.1037/vio0000086
Cross-Informant Agreement on Mental Health Outcomes in Children With Maltreatment Histories: A Systematic Review

Elisa Romano, Kelly Weegar, and Lyzon Babchishin
University of Ottawa

Michael Saini
University of Toronto

Objective: This systematic review and meta-analysis examined cross-informant agreement concerning mental health outcomes for maltreated children. This study provides important information to individuals working with maltreated children (e.g., clinicians, child welfare practitioners) who must make critical decisions related to safety, mental health, and permanency. Method: We retrieved 29 studies that collected mental health data for maltreated children using standardized instruments. We were able to examine 4 informant pairs and 7 mental health outcomes. Results: Agreement across children’s mental health outcomes was highest for child–caregiver pairs (small to large, significant effect sizes), followed by child–teacher and caregiver–teacher (small to medium, significant effect sizes) and then caregiver–caregiver (small, nonsignificant effect sizes) pairs. Agreement generally was highest for overt behaviors. Conclusions: It is important to include children as informants, gather teacher and caregiver data when assessing child internalizing difficulties, and examine specific maltreatment outcomes (e.g., traumatic stress). Clinically, it is critical to understand why different informants perceive a child’s functioning in different ways for purposes of prognosis and treatment planning for the child and his or her family. It is also critical to triangulate data sources when assessing for maltreatment impact and to consider data points that both converge and diverge when making appropriate safety and treatment plans.

Keywords: meta-analysis, maltreatment, abuse, agreement, mental health

Supplemental materials: http://dx.doi.org/10.1037/vio0000086.supp

Childhood maltreatment is an unfortunate reality for many children (Finkelhor, Turner, Shattuck, & Hamby, 2015; Sedlak et al., 2010; Trocmé, Fallon, MacLaurin, Hélie, & Turcotte, 2010) and impacts an estimated 9.1 per 1,000 U.S. children per year (U.S. Department of Health and Human Services, 2015). Such experiences occur primarily within a child’s family and typically include sexual abuse, physical abuse, emotional abuse, exposure to intimate partner violence, and neglect (Trocmé et al., 2010). Negative maltreatment-related consequences can cut across a range of interrelated domains that are all important for children’s healthy development and functioning (Afifi et al., 2014; English et al., 2005; Hamby & Grych, 2013; Staudt, 2001; Trickett, Negriff, Ji, & Peckins, 2011; Veltman & Browne, 2001). These include, but are not limited to, brain development, language, cognitive development (e.g., concentration, memory), academic achievement, interpersonal relationships (e.g., child–caregiver attachment), and mental well-being (Cowell, Cicchetti, Rogoch, & Toth, 2015; Mills et al., 2013; Petrenko, Friend, Garrido, Taussig, & Culhane, 2012; Romano, Babchishin, Marquis, & Fréchette, 2015; Stone, 2007; Trout, Hagaman, Casey, Reid, & Epstein, 2008; Veltman & Browne, 2001). Research suggests that the frequency of clinically significant mental health difficulties among maltreated children ranges from 40 to 80% (Burge, 2007; Czincz & Romano, 2009), which contrasts sharply with rates of 14 to 18% reported for nonmaltreated children (Cowell, 2007; Czincz & Romano, 2009). Despite the negative impact associated with maltreatment as well as the importance of early detection and treatment of mental health consequences for maltreated children, no previous research has synthesized the degree of agreement across informants (e.g., caregivers, children, and teachers) when assessing for mental health problems and when developing treatment plans to best address maltreatment-related effects. To address this gap, the current systematic review examined cross-informant agreement on children’s mental health outcomes, which was defined through emotional (e.g., anxiety, mood) and behavioral (e.g., aggression) variables. Because of our focus on maltreated children, we also included trauma-related behaviors including dissociation, posttraumatic stress symptoms, and sexualized behavior.

Conceptual Considerations of Cross-Informant Agreement

Given the multitude of impairments that maltreated children often experience, the provision of clinical services is undeniable. A critical first step involves the completion of a clinical assessment
to evaluate the nature of a child’s presenting difficulties (e.g., onset, severity, associated level of distress and/or impairment) as well as developmental and contextual considerations (Manassis, 2014; Sattler & Hoge, 2006). To this extent, various methods (e.g., interviews, questionnaires) and informants (e.g., children, caregivers, teachers) are typically used and triangulated to obtain a comprehensive assessment and an appropriate treatment plan (McConaughy, 2013; Sattler & Hoge, 2006). The common use of a multimethod, multi-informant approach for the assessment of children’s mental health functioning raises issues around cross-informant correspondence, which is defined as the level of agreement in information provided by different individuals about a particular child’s functioning or well-being. In addition, it would seem important to understand the types of variables that might influence cross-informant agreement in either a positive or negative manner (i.e., moderating variables).

The operations triad model (OTM; De Los Reyes, Thomas, Goodman, & Kundey, 2013) provides a conceptual framework for interpreting the meaning of cross-informant data; it presents the three concepts of converging, diverging, and compensating operations that can be applied to the triangulation of multiple data sources. Converging operations posit that stronger assessment conclusions can be made from studies with converging data (e.g., high cross-informant agreement on a particular outcome). Diverging operations refers to situations where there is cross-informant disagreement that is based on meaningful reasons, rather than measurement error. For example, disagreement may be due to informants observing the child in different settings (e.g., home, school) and/or to difficulties occurring more frequently in one setting over another. Finally, compensating operations refers to situations where cross-informant disagreement is the result of methodological differences, including measurement error where informants might have completed different questionnaires that have varying psychometric properties.

The OTM model suggests that examining cross-informant agreement can have several important implications. First, differences in the reporting of children’s mental health difficulties can help to identify the degree of problems, as experienced by various informants (Achenbach, Krukowsk, Dumenci, & Ivanova, 2005). Similarly, cross-informant agreement can help determine the degree to which particular problems are reported by all informants versus only one (or a few) informant, thereby indicating the extent to which problems are circumscribed to specific situations or interaction partners. Second, cross-informant agreement has implications for incremental validity, defined as improvements derived from obtaining additional information. For example, if a child’s difficulties only occur in one setting (e.g., home but not school), then obtaining a teacher report may not provide incrementally valid information. In contrast, if a child experiences difficulties across settings, then data from multiple informants would add incremental validity because one could then determine the extent to which multiple informants are consistent or inconsistent in their reports on the nature, frequency, and severity of the child’s difficulties (De Los Reyes et al., 2015). Third, cross-informant agreement can have treatment implications. Dirks, De Los Reyes, Briggs-Gowan, Cella, and Wakschlag (2012) suggested “that symptoms occurring in different situations or settings, or as perceived by different informants, may constitute distinct phenotypes, and that our ability to understand these phenotypes holds the promise of advancing the diagnosis and treatment of psychopathology” (p. 559). For example, past research has found that cross-informant disagreement on the presence of anxiety was associated with slower treatment improvement in 8- to 14-year-olds with an anxiety disorder (Kendall, Panichelli-Mindel, Sugarman, & Callahan, 2003), and low adolescent-parent agreement on adolescent mental health impairment has been found to be a strong predictor of continued impairment (Ferdinand, van der Ende, & Verhulst, 2004, 2006). No previous research review has considered the degree of cross-informant agreement for assessing mental health problems among maltreated children. Therefore, there remains a gap in our understanding about how mental health problems are viewed by different informants when a child has experienced abuse and/or neglect.

Research Syntheses on Cross-Informant Agreement

Issues of cross-informant agreement were first addressed over 25 years ago in a seminal meta-analysis of 119 studies published between 1967 and 1985 on correspondence for mental health problems among 1.5- to 19-year-olds (Achenbach, McConaughy, & Howell, 1987). Informants included parents, teachers, mental health workers, observers, peers, and youth themselves, all of whom were recruited from a variety of settings including the general population as well as inpatient and outpatient clinics. Studies were excluded if the youth were known to have major biological abnormalities (e.g., autism or a severe intellectual impairment) likely to impact behavior. Pearson correlations between various informant pairs, including parent–parent (22 studies, $df^p = 3,448$), child–parent (11 studies, $df = 1,339$), parent–teacher (26 studies, $df = 12,730$), and child–teacher (17 studies, $df = 4,578$), were all found to be statistically significant; however, they were modest with a mean correlation of $r = .28$. The mean correlation was higher ($r = .60$) between informants who had a similar relationship to the child (e.g., pairs of parents) or who interacted with a child in a similar setting (e.g., pairs of teachers). These findings suggest that ratings by different informants may not be interchangeable and can vary considerably, which may not be altogether surprising given that individuals have different perspectives based on the context and the types of interactions they have with children. Achenbach et al. (1987) also examined the degree to which cross-informant agreement might vary as a function of the child’s age (6–11 vs. 12–19 years), sex, clinical status (deviant vs. nonreferred/general population), parent (agreement between mothers and other informants vs. fathers and other informants), and types of problems (overcontrolled vs. undercontrolled). There were two significant findings: Cross-informant agreement on mental health outcomes was higher for 6–11 compared to 12–19 year olds (31 and 26 samples, respectively), and for undercontrolled (i.e., externalizing) compared to overcontrolled (i.e., internalizing) problems (159 and 121 samples, respectively).

Similarly, Duhig, Renk, Epstein, and Phares (2000) examined 60 published studies between 1990 and 1997 on interparental agreement (mother-father) for internalizing, externalizing, and total behavior problems among 3- to 19-year-olds. Study samples included children/adolescents from the general population as well as “clinically and medically challenged” populations as long as

\[ df = \sum(n - 3) \text{ across all samples.} \]
information on their emotional and behavioral functioning was provided. Results indicated moderate interparental agreement for children’s internalizing problems \((r = .45)\) and high interparental agreement for externalizing problems \((r = .63)\). Significant moderators of interparental agreement included child age (greater agreement with increasing age) and socioeconomic status (greater agreement for middle or mixed vs. lower status). More recently, Renk and Phares (2004) conducted a meta-analysis on cross-informant correspondence that included 74 studies. However, the focus was not on emotional and behavioral outcomes but rather, on social competence ratings of children/adolescents from at least two informants. Widely used and psychometrically sound measures of social competence were used (e.g., sociometric measures, social problem-solving measures, rating scales). Results indicated small mean weighted effect sizes between the self-reports of children/adolescents and the reports of other informants, with a range from \(r = .21\) (child/adolescent–parent) to \(r = .30\) (child/adolescent–peer). For other informant pairs, the mean weighted effect sizes were moderate and ranged from \(r = .38\) (teacher–parent) to \(r = .48\) (teacher–peer). Several potential moderating variables were examined, namely the type of measure completed by the informants, the age and sex of the child/adolescent, and the location of the assessment. Note that analyses were descriptive in nature because of the limited data across included studies on moderating variables. Findings suggested that cross-informant correspondence on social competence was higher with the use of sociometric measures rather than rating scales and with the completion of measures at school rather than in the laboratory (for teacher–peer pairings only). For age, children/adolescent-parent agreement was higher during middle childhood (compared to adolescence), whereas the agreement between children/adolescents and peers as well as between teachers and peers was higher during adolescence (compared to childhood).

Finally, De Los Reyes et al. (2015) conducted a meta-analysis on 344 cross-informant correspondence studies on children’s mental health published between 1989 and 2014. The study focused on internalizing and externalizing behaviors, as assessed through quantitative measures such as behavior checklists and self-report measures. Results indicated an overall mean correlation of \(r = .28\), with slightly higher agreement for externalizing \((r = .30)\) compared to internalizing \((r = .25)\) problems. There was greater agreement across internalizing and externalizing problems between informants who observe the child in the same setting (i.e., parents), compared to the parent–child, parent–teacher, and teacher–child pairings. Agreement was also higher when informants made use of dimensional, rather than categorical, measures. Child age did not emerge as a statistically significant moderator.

In sum, past research syntheses have suggested only modest agreement between informants on children’s emotional and behavioral outcomes. This is not altogether surprising given that each informant has a unique relationship with a child and interacts with him/her in different settings, which contributes to situational variability (Dirks et al., 2012; McConaughy, 2013). Past syntheses have also suggested that cross-informant agreement may be influenced by sociodemographic, child, parent, and family characteristics as well as by the outcome that is being examined (with higher agreement for observable behaviors). Several findings are equivocal (e.g., child age) and are likely due to methodological differences, such as the particular outcomes assessed, whether age was examined as a continuous or dichotomous variable, and the types of informant pairs that were included. For example, Duhig et al. (2000) examined interparental agreement on children’s behavioral outcomes while De Los Reyes et al. (2015) examined agreement on these outcomes after collapsing various informant pairs. Undoubtedly, the extent to which an informant judges a child’s emotional or behavioral functioning to be problematic is also influenced by such factors as attributions (whether the cause of the behavior is perceived to be dispositional vs. environmental), biases (threshold for deciding whether the behavior warrants treatment intervention), and contexts (situations in which informants observe the behavior and/or in which observations are made; De Los Reyes, 2011; Dirks et al., 2012). As such, it is generally accepted that each informant has the potential to offer reliable and valid information, despite low cross-informant agreement.

**Study Objective and Hypotheses**

While cross-informant agreement research on children’s mental health outcomes exists, we did not come across any syntheses that focused specifically on children with maltreatment histories. As such, our objective was to examine cross-informant agreement for mental health outcomes among maltreated children. We focused on mental health outcomes that are most frequently assessed among maltreated children, but we also examined more unique outcomes, such as trauma-related symptoms and sexual behaviors, which have not been explored in past syntheses. Based on past research, we had the following hypotheses:

**Hypothesis 1:** Cross-informant agreement generally would be low across all informant pairs, although agreement among individuals who interacted with a child in a similar context/relationship was expected to be higher (e.g., parent–parent).

**Hypothesis 2:** Agreement among informant pairs would be higher for externalizing, compared to internalizing, behaviors.

**Method**

**Search Strategy**

The information retrieval process included a search of eight electronic databases: MEDLINE, PsycINFO, EMBASE, Applied Social Sciences Index and Abstracts, Education Resources Information Center, Social Work Abstracts, Social Sciences Abstracts, and Dissertation Abstracts International. Across these databases, we developed a systematic search strategy which included child or adolescent or infant or boy or girl or toddler or baby or preschool or preschool or teen and abuse or maltreatment or neglect or physical abuse or sex abuse or witness or child protect or child welfare or threat or danger and psychometrics or questionnaires or cross informant or agreement or correlation or assessment or reliability or rating or construct or variance or model or comparison. Search term combinations were modified to match the controlled vocabulary specific to each database. We also reviewed the reference lists of all relevant articles that were obtained, including those from previously published reviews.

We considered published and unpublished English studies (e.g., dissertations) ranging from January 1980 to June 2014. Studies needed to include at least two informants who provided data on the mental health functioning of maltreated children. Informants could...
span such individuals as children (0–18 years), caregivers, teachers, and clinicians. Mental health functioning included emotional and behavioral variables. Emotional variables could encompass anxiety/worries, phobias, depression, mood, withdrawal, crying, clinginess, separation anxiety, negative affect, and posttraumatic stress. Behavioral variables could include hyperactivity, inattention, attention-deficit/hyperactivity disorder, conduct problems, anger, aggression, oppositional behavior, defiant behavior, sexualized behavior, self-injury, alcohol/drug use, gang involvement, and school truancy.

We used a standard definition of child maltreatment, specifically acts of commission and/or omission toward a child in the context of a relationship of responsibility, trust, or power that result in actual or potential harm to the child’s physical or psychological well-being (World Health Organization, 2002). As such, child maltreatment could include specific acts of physical abuse, sexual abuse, neglect, and emotional abuse, including exposure to intimate partner violence. We focused on adult perpetrators who could be either family or nonfamily members; peer victimization studies were not included. Acceptable measures of determining child maltreatment included official reports or administrative data on the substantiation (or confirmation or indication) of an allegation of child maltreatment. The allegation could come from single or multiple informants (e.g., teacher, physician) or could be made based on self-report by a child or caregiver. In this review, substantiated and nonsubstantiated reports of child maltreatment were included to compare agreement levels across substantiation determinations if possible.

Studies were included if they made use of standardized measures to assess any of the mental health outcomes described above. These measures included standardized questionnaires as well as observational tools and interviews, with the collected data for the two latter methods being scored according to a specific set of procedures. Methodologies included experimental, quasiexperimental, cross-sectional, and longitudinal cohort studies. Qualitative studies were excluded given our focus on standardized measures.

**Study Selection**

Figure 1 shows that the comprehensive electronic database search yielded 6,841 records. The removal of duplicate and non-English records resulted in 4,334 records, which underwent an initial screening for inclusion. Three dyads reviewed all abstracts and retained records for studies (a) in which the abstract refer to an empirical study, (b) that included 0- to 18-year-olds who had experienced maltreatment, and (c) which included information on mental health outcomes (i.e., emotional and/or behavioral prob-
lems). After the first 100 records were screened, the percent agreement for each dyad was calculated (i.e., total number of concordant observations divided by the total number of paired observations). The first author (E. R.) then met with each dyad to discuss discrepant ratings, following which two dyads proceeded to complete the initial screening for the remaining records because their percent agreements were acceptable (>80%). The other dyad screened an additional 100 abstracts, after which point percent agreement was calculated and found to be >80%. At this point, this last dyad completed the initial screening for the remaining records. Once dyads completed the initial screening for all abstracts, any discrepancies were reviewed by study authors (E. R., M. S.) who then made a decision about whether the record would be included. This initial screening process resulted in the retention of 815 records.

In the second screening, three dyads were again responsible for one third of the records. Records were retained if the study (a) included a standardized tool to gather data and (b) included more than one informant. Any discrepant ratings were reviewed by study authors (E. R., M. S.) who then made a decision about whether the record would be included. In the end, 173 titles and abstracts were retained. In attempting to retrieve these studies, 24 were not reviewed because the full-text article could not be located or access was not permitted.

The remaining 149 studies were divided across three dyads and underwent a third screening. Records were retained if (a) the study sample included at least 80% maltreated children and/or adolescents or if the study included a separate subsample of maltreated youth and (b) the study provided data from two informants on the same emotional or behavioral outcome. Any discrepant ratings were reviewed by study authors (E. R., M. S.) who then made a decision about whether the record would be included. This final screening process resulted in 91 studies being retained for purposes of data extraction. The remaining information for Figure 1 can be found below.

Data Extraction

Using a predesigned data extraction form, the 91 included studies were divided among six coders who extracted study identification information as well as data on emotional/behavioral outcomes. Note that coders received prior training on the use of the data extraction form; study authors resolved any queries from the coders. In conducting the data extraction, additional exclusions were necessary. Fifty-eight studies could not be included in any data analyses because of insufficient data. These studies included data for both maltreated and nonmaltreated children but did not differentiate results between the two groups (e.g., Mahoney, Donnelly, Boxer, & Lewis, 2003). There also were studies which did not have multiple informant data for the same behavioral or emotional outcome (e.g., Sedlar, 2001). Removal of these studies resulted in a sample of 33 studies.

There were several different informant pairs across the 33 studies, but it was necessary to ensure a minimum of three data points for each informant pair to conduct meta-analyses. As such, we regrouped the various combinations into the four following informant groups: (a) child–caregiver, which included child–biological parent, child–foster parent, and child–caregiver informant pairs; (b) child–teacher; (c) caregiver–teacher, which included caregiver–teacher and biological parent–teacher informant pairs; and (d) caregiver–caregiver, which included caregiver mother–caregiver father and biological mother–biological father informant pairs. Within these groups, we used the term caregiver when the exact nature of the parent’s relationship to the child was unspecified in the research study. There were insufficient numbers of effect sizes for other types of cross-informant pairs so four studies were removed prior to meta-analyses, namely child–peers (Salzinger, Feldman, Hammer, & Rosario, 1993), child–mental health professional (Fricker & Smith, 2001), teacher–mental health professional (Simmer-Dvorch, 1998), and teacher–child care provider (Hukkanen, Sourander, Bergroth, & Piha, 1999). Removal of these studies resulted in a final sample of 29 studies (see Figure 1).

Several studies presented data on teacher–mental health professional (Scarpa Scerbo & Kolko, 1995) and caregiver–mental health professional (Porter, 2003; Scarpa Scerbo & Kolko, 1995) pairings; however, these pairs were not included in the meta-analysis due to insufficient data. Nonetheless, these two studies were still included because they had data for other informant pairs included in the current review.

For mental health outcomes, it was necessary for purposes of analysis to create categories to regroup the multitude of outcomes across the 29 included studies. We created the seven following categories: (a) Total behaviors included outcomes labeled as such and typically consisted of total scores on behavioral checklists (e.g., Conners’ Rating Scales, Child Behavior Checklist); (b) externalizing behaviors related to children’s negative outward behavior and included aggression, anger, delinquent behavior, hyperactivity, impulsivity, inattention, and oppositional behavior (e.g., scores on the Child Behavior Checklist Externalizing scale); (c) internalizing behaviors related to children’s internal psychological states and included anxiety, anxious/depressed, avoidance, depression, sleep problems, somatic complaints, and somatization (e.g., scores on the Child Behavior Checklist Internalizing scale); (d) trauma-related behaviors included outcomes often associated with childhood trauma, such as dissociation, posttraumatic stress, and suicidality (e.g., scores on the Trauma Symptom Checklist for Children dissociation and posttraumatic stress subscales); (e) social behaviors included outcomes related to children’s relational behaviors, such as popularity, social withdrawal, and sharing (e.g., scores on the Child Behavior Checklist Social Problems subscale); (f) sexual behaviors included outcomes related to atypical sexual thoughts, feelings, or behaviors (e.g., scores on the Trauma Symptom Checklist for Children Sexual Concerns subscale); and (g) emotionality included behaviors related to children’s emotional reactivity and/or regulation (e.g., scores on the emotional liability subscale of the Conners’ Rating Scales).

Data Analytic Plan

Meta-analyses were used to examine the degree of variation occurring around the mean effect size using Comprehensive Meta-Analysis (Version 3.0). Studies with effect size statistics for at least one informant pair were transferred to a predesigned dataset for analysis. Studies which did not include effect size statistics but nonetheless presented data directly comparing two informants (e.g., mean, standard deviation, and sample size for each informant pair; correlation value; t value; chi-square value; or F value) were converted to effect sizes and included in the dataset. Note that it
was possible for a study to contain data on more than one behavioral or emotional outcome. It was also possible for a study to include data from more than one pair of informants. In such cases, effect sizes were calculated for each relevant outcome and each informant pair but separated in the synthesis to ensure independence of effect sizes across studies. In studies with repeat assessments, only data from the initial assessment were included. To combine the findings of multiple independent studies, a weighted mean effect size \((r)\) was calculated to adjust for differences in sample sizes across studies, as suggested by Achenbach et al. (1987). Statistical heterogeneity in the outcome measures was assessed using the \(Q\) statistic and the associated \(p\) value for each analysis (Higgins, 2008).

As part of this study, our intention was to examine potential moderator variables of cross-informant agreement at the level of child/family and maltreatment characteristics. For maltreatment characteristics, data were extracted on the following: maltreatment type; whether the child experienced multiple types of maltreatment; perpetrator sex and the relationship to the child; whether the child was in a child welfare, clinical, or community setting; whether the maltreatment was substantiated; and whether a caregiver informant was also the perpetrator. For child/family characteristics, data were extracted on child sex, age, and ethnicity as well as on family socioeconomic status. However, the large amount of missing data made it impossible to examine moderators of cross-informant agreement for each mental health outcome and to conduct statistical analyses. Even after collapsing all mental health outcomes together to compare potential moderators for each of the four informant pairs, a number of the cells had only a few effect sizes (or none at all) that could be extracted from the included studies. For example, there were no studies that examined the potential role of child sex on child–teacher agreement across mental health outcomes. For other informant pairs, data were only available for either boys (in the case of caregiver–teacher agreement) or girls (in the case of caregiver–caregiver agreement).

**Results**

**Description of Included Studies**

The supplemental materials indicate that the majority of studies were conducted in the United States (84%), followed by Canada (10%), Australia (3%), and the Netherlands (3%). For sample characteristics, most studies were conducted with clinic-referred samples (55%), followed by child welfare (31%), community (8%), and mixed (6%; child welfare/clinical or child welfare/community) samples. Overall sample sizes ranged from 15 to 1,046, with the mean ages of children ranging from 4.4 to 15.9 years. The majority of studies (82%) included samples composed of both boys and girls, while 11% were composed primarily of boys and 7% primarily of girls.

For maltreatment characteristics, 52% of studies included children whose maltreatment had been substantiated, while 45% had children with suspected but unsubstantiated maltreatment. The remaining study (3%) had a mixed sample with equal numbers of children having experienced either substantiated or suspected maltreatment. Most studies (48%) focused on children who had primarily experienced sexual abuse, followed by studies which included various maltreatment experiences (21%), primarily physical abuse (17%), primarily neglect (10%), and primarily exposure to domestic violence (4%). Even though a study’s sample might have been composed of children who primarily experienced a specific type of maltreatment (i.e., \(\approx 80\%\) of the sample), we were also interested in the number of studies that included children with multiple victimization experiences. Only 14% of the studies included that the children in their samples had experienced more than one maltreatment type (Kugler et al., 2013; Lamers-Winkelman, Willemse, & Visser, 2012; Porter, 2003; Sines, 2002). The remaining studies were more or less evenly divided either in not providing any information on whether children had experienced multiple types of victimization (41%) or in stating that the sample of children had primarily experienced only one maltreatment type (45%).

With regard to perpetrator characteristics, most of the studies (45%) did not provide any information on perpetrator sex. In contrast, 31% of studies noted that perpetrators were primarily male, while 3% included samples where the perpetrators were primarily female. Finally, 21% of studies sampled children where there was a mix in terms of perpetrator sex. For the child-perpetrator relationship, most studies (45%) did not provide this information, while 34% had samples of children where there was a mix in terms of the relationship with the perpetrator (family and nonfamily members). Finally, 14% of studies indicated that the perpetrators were primarily family members, while 7% included children who experienced maltreatment primarily by nonfamily members. Out of the 28 studies that included caregivers as informants, 14% indicated that one of the caregivers providing information about the child’s mental health was in fact the perpetrator of the maltreatment, and the maltreatment in these studies was mainly of the type where children tend to remain in the home (i.e., neglect, exposure to domestic violence). In 50% of the studies, it was clearly indicated that the caregiver informant was not the perpetrator, and the maltreatment in these studies was primarily the type where children would be removed from the home and away from the perpetrator (i.e., sexual abuse). For the remaining studies (36%), there was either no clear information about the status of the caregiver as a perpetrator or the information that was provided would suggest that some of the caregiver informants were the perpetrators (e.g., children were still living in the biological home).

The emotional and behavioral outcomes assessed across the studies varied; most (66%) collected data on more than one outcome, with the more frequent categories being broad internalizing and externalizing behaviors. Data collection consisted almost entirely of standardized self-report measures, with the most frequent being the various versions of Achenbach’s behavior checklists (Youth Self-Report, parent-completed Child Behavior Checklist, and Teacher Report Form), the parent- and teacher-completed versions of Conners’ Rating Scales–Revised, and Briere’s trauma checklists (Trauma Symptom Checklist for Children and Trauma Symptom Checklist for Young Children).

**Description of Outcomes and Informant Pairs**

The supplemental materials indicate that the most frequent data were collected for internalizing behaviors (72%; 21 studies), followed by externalizing behaviors (55%; 16 studies), total behavior problems (28%; eight studies), social problems (21%; six studies), trauma-related symptoms (17%; five studies), emotionality (10%;
three studies), and sexual concerns (7%; two studies). For informant pairs, most of the data were gathered from children and their caregivers (55%; 16 studies), followed by caregiver–teacher (48%; 14 studies), child–teacher (14%; four studies), and caregiver–caregiver (7%; two studies).

**Cross-Informant Agreement on Mental Health Outcomes for Maltreated Children**

Across all informant pairs and mental health outcomes (29 studies, 110 samples), the mean weighted effect size ($r$) was in the medium range ($r = .31$). This suggests moderate cross-informant agreement on mental health outcomes for maltreated children. To better understand the heterogeneity across informant pairings, Table 1 presents the mean weighted effect size ($r$) for the seven mental health outcomes collapsed across the informant pairs, as well as for each of the four informant pairs. It should be noted that several effect sizes were based on data from only one informant pairing (e.g., trauma-related behavior for child–teacher pairs). Collapsed across all pairs, results indicated statistically significant cross-informant agreement for all mental health outcomes with the exception of emotionality. Mean effect sizes were either small (total behavior, trauma-related symptoms, social behaviors, and sexual behaviors) or medium (externalizing, internalizing).

When mental health outcomes were examined individually for the four informant pairings, mean effect size ranges showed greater variability. For child–caregiver pairs, the statistically significant mean effect sizes were either small (total behavior, trauma-related symptoms, sexual behaviors) or large (externalizing, internalizing, and social behaviors). This informant pair had the highest mean effect sizes (and thus the greatest agreement on mental health outcomes for maltreated children) in comparison with the other three informant pairs. Note that we did not test whether the differences between pairs were statistically significant; however, we calculated standardized mean differences so the comparisons are based on a common metric using Cohen’s notion of small, medium, and large effects. For child–teacher pairs, the statistically significant mean effect sizes ranged from small (total behavior, externalizing, internalizing) to medium (trauma-related internalizing). Statistically significant mean effect sizes for caregiver–teacher pairs were in the small (total behavior, social) and small to medium (externalizing, internalizing) range. Finally, the mean effect sizes for caregiver–caregiver agreement on maltreated children’s mental health outcomes were very small and not statistically significant.

For mental health outcomes, externalizing, internalizing, and social behaviors had the largest, statistically significant mean effect sizes across informant pairs. The mean effect sizes for these

### Table 1

**Cross-Informant Effect Sizes for Each Informant Pair by Each Outcome**

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Informant pair</th>
<th>Across all informant pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total behavior</td>
<td>Child–caregiver</td>
<td>.17***</td>
</tr>
<tr>
<td></td>
<td>Child–teacher</td>
<td>.20***</td>
</tr>
<tr>
<td></td>
<td>Caregiver–teacher</td>
<td>.16***</td>
</tr>
<tr>
<td></td>
<td>Caregiver–caregiver</td>
<td>.02*</td>
</tr>
<tr>
<td>Number of studies (samples)</td>
<td>2 (2)</td>
<td>2 (2)</td>
</tr>
<tr>
<td></td>
<td>5 (5)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td>8 (10)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>3.22</td>
<td>5.48**</td>
</tr>
<tr>
<td></td>
<td>23.14***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35.05***</td>
<td></td>
</tr>
<tr>
<td>Externalizing behavior</td>
<td>Mean weighted $r$</td>
<td>.72***</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>6 (6)</td>
</tr>
<tr>
<td></td>
<td>10 (18)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16 (32)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>400.60***</td>
<td>20.15***</td>
</tr>
<tr>
<td></td>
<td>845.04***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1,872.05***</td>
<td></td>
</tr>
<tr>
<td>Internalizing behavior</td>
<td>Mean weighted $r$</td>
<td>.61***</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>12 (19)</td>
</tr>
<tr>
<td></td>
<td>9 (13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>21 (40)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>1,519.59***</td>
<td>32.50***</td>
</tr>
<tr>
<td></td>
<td>278.61***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2,345.77***</td>
<td></td>
</tr>
<tr>
<td>Trauma-related behavior</td>
<td>Mean weighted $r$</td>
<td>.24***</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>4 (10)</td>
</tr>
<tr>
<td></td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5 (13)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>327.30***</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>362.42***</td>
<td></td>
</tr>
<tr>
<td>Social behavior</td>
<td>Mean weighted $r$</td>
<td>.71***</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td>3 (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6 (9)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>—</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>125.79***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>174.95***</td>
<td></td>
</tr>
<tr>
<td>Sexual behavior</td>
<td>Mean weighted $r$</td>
<td>.25***</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>2 (2)</td>
</tr>
<tr>
<td></td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.40</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>.25**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.2.40</td>
<td></td>
</tr>
<tr>
<td>Emotionality</td>
<td>Mean weighted $r$</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Number of studies (samples)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>3 (4)</td>
<td></td>
</tr>
<tr>
<td>$Q$ statistic</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>.05</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Conventional values of effect size (Cohen, 1988) are small ($r = .10$), medium ($r = .30$), and large ($r = .50$).

* This value represents an individual effect size (and not a mean) because it is based on data from only one informant pairing. As such, a $Q$ statistic could not be calculated.

$p < .05$. ** $p < .01$. *** $p < .001$. 
outcomes generally fell in the small to medium or large range, except for caregiver–caregiver pairings. The mean effect sizes were small or small to medium across all informant pairs for total behavior problems \((r = .02\) to \(0.20\)) and trauma-related symptoms \((r = .07\) to \(0.24\)). There was a small to medium mean effect size for sexual behavior for the child–caregiver informant pair \((r = .25)\).

If we collapse the mental health outcomes across each informant pairing, a classic fail-safe \(N\) analysis suggests that 21,839 studies with a null effect would be needed to change the child–caregiver results from statistically significant \((p < .001)\) to nonsignificant. Orwin's fail-safe \(N\) analysis indicates that 379 studies would be needed to bring the correlation for this informant pairing under 0.05 across the mental health outcomes that were examined (Borenstein, Hedges, Higgins, & Rothstein, 2009). For child–teacher pairings, a classic fail-safe \(N\) analysis suggests that 618 studies with a null effect would be needed to change results from statistically significant \((p < .001)\) to nonsignificant. Orwin’s fail-safe \(N\) analysis indicates that 42 studies would be needed to bring the correlation for this informant pairing under 0.05. For caregiver–teacher pairings, a classic fail-safe \(N\) analysis suggests that 7,200 studies with a null effect would be needed to change results from statistically significant \((p < .001)\) to nonsignificant. Orwin’s fail-safe \(N\) analysis indicates that 172 studies would be needed to bring the correlation for this informant pairing under 0.05. Finally, for caregiver–caregiver pairings, a classic fail-safe \(N\) analysis suggests that no additional studies with a null effect would be needed to change results from statistically significant \((p < .001)\) to nonsignificant. Orwin’s fail-safe \(N\) analysis was not applicable as the effect size \((r)\) results for the caregiver–caregiver pairing were between 0.04 and 0.06.

Discussion

This systematic review included 29 studies conducted from 1980 to 2014 that fit criteria for examining cross-informant agreement on mental health outcomes for maltreated children (0–18 years). Based on available data, we were able to examine agreement across four informant pairs (child–caregiver, child–teacher, caregiver–teacher, caregiver–caregiver). We were also able to examine seven mental health outcomes (total behaviors, externalizing behaviors, internalizing behaviors, trauma-related behaviors, social behaviors, sexual behaviors, and emotionality).

Cross-Informant Agreement

Findings on agreement collapsed across all informant pairings and mental health outcomes was moderate \((r = .31)\), which is in line with previous syntheses that obtained overall effect sizes (Achenbach et al., 1987; De Los Reyes et al., 2015). Compared with the current study, however, previous syntheses have not focused exclusively on children with maltreatment histories, and they have not been as comprehensive in terms of the informant pairings and mental health outcomes examined. As such, it is difficult to make direct comparisons across study findings.

Turning to specific informant pairings, our findings indicated that agreement on maltreated children’s mental health outcomes was highest among child–caregiver pairs (small to large, significant effect sizes), followed by child–teacher and caregiver–teacher pairs (small to medium, significant effect sizes). For caregiver–caregiver pairs, the effect sizes were small and not statistically significant. Comparisons with past cross-informant agreement syntheses are limited because of the different child populations that were examined. Moreover, past cross-informant syntheses have also differed on such methodological variables as the types of informant pairings and outcomes examined. For example, Duhig et al. (2000) did not examine pairings that included children as informants. Also, past syntheses, including the recent meta-analysis by De Los Reyes et al. (2015), have focused on broad behavioral categories (i.e., internalizing, externalizing). In contrast, we included more maltreatment-related outcomes in addition to these general behaviors, such as trauma symptoms and sexualized behaviors. Within this context, however, our findings on the relative lack of agreement between caregivers, compared to other informant pairings, is contrary to our first hypothesis and with past research (Achenbach et al., 1987; De Los Reyes et al., 2015; Duhig et al., 2000). We would have expected higher agreement given that these informants have a similar caregiving relationship with the child and presumably, are exposed to more similar behaviors than other pairings (e.g., caregiver–teacher). There may be several possible explanations for this counterintuitive finding. The most likely explanation is that methodologically, only two studies (Lipovsky, Saunders, & Murphy, 1989; Manion et al., 1998) collected data from both caregivers so there were instances when we could not calculate mean effect sizes but instead had to rely on individual effect sizes. As such, our statistical power for this informant pairing was highly limited, and the findings could change with additional data on caregiver–caregiver agreement.

Examining the results from the point of view of mental health outcomes, agreement across informant pairing (with the exception of caregiver–caregiver pairs) was highest for more easily observable behaviors, such as externalizing and social behaviors (generally large effect sizes). These findings were in line with our second hypothesis and generally are consistent with past research (Achenbach et al., 1987; De Los Reyes et al., 2015; Duhig et al., 2000). Interestingly, our findings for internalizing behavior also indicated high cross-informant agreement among child–caregiver pairings as well as moderate agreement among child–teacher pairings. It is not surprising that children would be aware of their internal states, but this finding also suggests that teachers and caregivers may be particularly astute at identifying internalizing behaviors (e.g., mood, anxiety) among children. For teachers, this may be a function of the considerable exposure they have to a wide variety of children, which may make them quite adept at identifying problematic versus normative behavior.

Limitations and Research Implications

The limitations of the present study were primarily the result of what we were unable to do analytically because of missing data across the included studies. For cross-informant agreement, we could not examine all potentially informative pairings for this population of maltreated children (e.g., child–child welfare practitioner, caregiver–child welfare practitioner), and our conclusions for caregiver–caregiver pairings must be regarded as preliminary given the limited data on this pairing. This is concerning given the critical role that caregivers play in the identification of children’s
mental health difficulties and in the management of these difficulties through intervention efforts.

It is difficult to explain the lack of information from two caregivers. Given that 48% of the included studies focused on family perpetrated maltreatment, it may be that one of the caregivers was the perpetrator and therefore, not available to participate in the study (e.g., in the case of sexual abuse where the perpetrator may no longer be living in the family home). Moving forward, it would be important for additional studies on cross-informant agreement for maltreated children to include caregiver pairings. In so doing, information needs to be gathered on maltreatment characteristics as a way to guide the interpretation of agreement levels for this informant pair. For example, one might expect that the amount of maltreatment-related mental health impairment reported for a child would vary if the caregiver informant were also the perpetrator of that maltreatment. One might also speculate that caregiver–caregiver agreement on maltreated children’s mental health outcomes would be higher in substantiated cases where there is clear evidence that the child had experienced maltreatment. In addition to prioritizing the collection of data from both caregivers (whenever possible), the inclusion of other useful pairings needs to be considered, specifically that of child welfare practitioners and mental health clinicians given their frequent involvement in the lives of children with maltreatment histories.

Turning to mental health outcomes, the majority of included studies focused on broad behavioral categories, such as internalizing and externalizing difficulties. While important, it is imperative that future studies also examine outcomes that are directly related to having experienced maltreatment. Examples include posttraumatic stress symptoms, sexualized behaviors (in the case of sexual abuse), and emotion regulation. Finally, we could not examine in any meaningful way potential moderators of cross-informant agreement as data across the included studies were limited and unsystematic. Moving forward, it is imperative that research on maltreated children (whether specifically designed as a cross-informant study or not) collect pertinent data in a comprehensive and systematic manner. At a minimum, such information needs to include sociodemographic variables and a detailed description of the maltreatment, including whether the maltreatment was substantiated, whether the child experienced multiple types of victimization (and what those specific types were), and perpetrator characteristics (number, sex, relationship to the child). By doing so, additional efforts at synthesizing existing knowledge can draw firm conclusions and, in this way, advance our understanding of informant agreement on maltreated children’s mental health outcomes.

**Clinical and Policy Implications**

The first general implication is that the assessment of mental health outcomes for maltreated children needs to gather data from multiple individuals who play a role in the child’s life. This is based on our findings that mental health difficulties were identified and agreed upon (to varying degrees) by individuals who interact with maltreated children in various settings (e.g., home, school). In such circumstances, the OTM model suggests that collecting data from multiple informants adds to our understanding of the child and is incrementally valid because the child is exhibiting difficulties across different settings with which he or she interacts. A more specific implication relates to the importance of including the maltreated child as an informant to the greatest extent possible (taking into account age and developmental considerations). The two informant pairings with the greatest agreement were those that included the child, suggesting that the child is an important informant (De Los Reyes et al., 2013).

From a policy perspective, these results support Article 12 of the United Nations Convention on the Rights of the Child, which states that children have the right to express their opinions in matters that affect them and to have their opinions considered by decision makers (www.unicef.org/crc/index_30160.html). As such, rather than depending solely on adults to provide details about children’s emotional and behavioral needs, procedures need to be established and/or strengthened to ensure that children are provided with opportunities to be part of the decision—making process with regard to such matters as their well-being and receipt of clinical services.

The second general implication is that one can expect more or less agreement depending on the informant pair. As such, clinicians could be better informed about potential overlap (or lack thereof) in data gathered from various informants and about the significance of such findings. The OTM model states that one can make stronger conclusions when there is converging information from two informants (De Los Reyes et al., 2013). For our specific results, this suggests that the greatest amount of confidence can be placed on data derived from child–caregiver pairs, followed by child–teacher and caregiver–teacher data, at least as these data pertain to mental health outcomes for maltreated children. It is important to emphasize that stronger agreement between informants does not indicate that the mental health outcomes of a maltreated child are any more real than outcomes for children where cross-informant agreement is weaker. It simply means that the presentation of the mental health difficulty may be more obvious and/or more widespread (which would make it easier for more than one individual to observe). Agreement does not speak to the validity of a difficulty’s presence but really about the importance of bringing multiple perspectives to obtain a more complete picture of a child’s well-being and functioning.

From a clinical perspective, it is critical to understand why different informants perceive a child’s functioning in different ways; such an understanding is informative for purposes of prognosis and treatment planning via a vis the child and his or her family. From a child welfare perspective, it is critical to triangulate data sources when assessing for impacts of maltreatment and to consider data points that both converge and diverge when making appropriate safety and treatment plans. In our study, agreement between caregivers was low, although it is important to again emphasize the preliminary nature of these findings given the limited available data. Nonetheless, this finding may be an indication of the complex dynamics which can exist between caregivers whose child has experienced maltreatment. This dynamic is undoubtedly influenced by characteristics of the child’s maltreatment experience. It may also be influenced by the fact that families in which maltreatment has occurred tend to also struggle with greater dysfunction and parental mental health impairment (Hamby & Grych, 2013; Leinonen, Solantaus, & Punamaki, 2003; Turner, Finkelhor, & Ormrod, 2012), which may compromise caregivers’ ability to accurately report on their child’s mental well-being.
In sum, this synthesis is relevant across the different sectors in which maltreated children are involved. For clinicians who evaluate these children’s mental health functioning, understanding the level of cross-informant agreement as a function of the informant pair and the particular outcome has implications for the way in which data are interpreted and reported in an assessment. This is important because these assessments hold significant weight not only for treatment planning but often for the child welfare and judicial systems, which are often required to make decisions related to custody and permanency planning in the case of a child welfare-involved child.

References


Czincz, I., & Romano, E. (2009). Examining how the mental health needs of children who have experienced maltreatment are addressed within Ontario Children’s Aid Societies. Canadian Journal of Family and Youth, 2, 25–51.


Kugler, B., Burke, N., Bloom, M., Truax, T., Kaercher, L., & Storch, E. (2013). Caregiver-teacher agreement on emotional and behavioral problems in traumatized youth in residential treatment. Residential Treat-


Received May 15, 2016
Revision received October 14, 2016
Accepted October 25, 2016

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.