Measuring Informant Discrepancies in Clinical Child Research

Andres De Los Reyes
Yale University

Alan E. Kazdin
Yale University School of Medicine

Discrepancies among informants' ratings of child psychopathology have important implications for diagnosis, assessment, and treatment. Typically, parents and children complete measures (e.g., self-report checklists, diagnostic instruments) to assess child dysfunction. Ratings gathered from these sources reveal relatively little agreement on the nature and extent of the child's social, emotional, and behavioral problems. This article reviews and illustrates the most frequently used methods of measuring informant discrepancies in the clinical child literature (i.e., raw difference, standardized difference, and residual difference scores) and outlines key considerations to influence their selection. The authors conclude that frequently used methods of measuring informant discrepancies are not interchangeable and recommend that future investigations examining informant discrepancies in clinical child research use the standardized difference score as their measure of informant discrepancies.

Different informants' (i.e., mothers, children, and teachers) ratings of child psychopathology are often discrepant from one another (e.g., Achenbach, McConaughy, & Howell, 1987; Youngstrom, Loeber, & Stouthamer-Loeber, 2000). Reviews consistently reveal little agreement between the ratings provided by parents and children (e.g., *r*s often in .20s). Discrepancies among informants can make it difficult to integrate information from multiple informants in the clinical assessment of children, can raise fundamental questions about the prevalence of psychopathology in youth, and can lead to differences in who is identified as meeting criteria for a given disorder (Kazdin, 1989; Offord et al., 1996; Yeh & Weisz, 2001). Thus, the consequences of discrepancies among informants' endorsements of diagnoses in children can be enormous, given their relation to the prevalence and identification of childhood disorder.

Investigators have examined the discrepancies between informants' ratings by comparing how two (or more) informants rate the same child (e.g., Chi & Hinshaw, 2002; Youngstrom et al., 2000). The distinct mathematical properties exhibited by the different measures can lead to different conclusions, as a function of the differences between the variances of informants' ratings or the correlation between informants' ratings. This article elaborates several options related to the evaluation of discrepancies among informants' ratings. We review the most frequently used methods

of measuring informant discrepancies in the clinical child literature, illustrate these methods using a clinic-referred sample of children with aggression and their mothers, and provide recommendations for measuring informant discrepancies in clinical child research.

Current Ways of Measuring Informant Discrepancies

The three most commonly used methods to measure informant discrepancies are the difference between two informants' raw or unstandardized ratings, the difference between two informants' standardized ratings, and the residual difference between two informants' ratings. First, the raw difference score is simply the subtraction of one informant's rating from another informant's rating. Second, the standardized difference score is the subtraction of one informant's standardized rating from another informant's standardized rating. Each informant's rating is converted into a z score by placing each informant's rating on a distribution, relative to the rest of the same informants' ratings in the sample. The informant's converted rating is then subtracted from the other informant's converted rating (e.g., z score of parent's rating minus z score of child's rating). The standardized difference score is most often used to place both informants' ratings on the same metric or scale of variability (e.g., the z distribution), because doing so enhances the interpretability of the score (i.e., the standardized difference score has a mean of 0 and a standard deviation of 1).[1] Finally, the residual difference score is computed by using one informant's rating to predict another informant's rating and measuring the difference between the rating predicted by the predictor informant's rating and the predicted informant's actual rating. One informant's rating is used as the dependent variable and the other informant's rating is used as the independent variable in a regres-

[1] Of course, the standardized difference score may not have a mean of 0 in instances in which missing data exist for one of the informants in a pair (e.g., mother and child are informants of interest, and the child for one of the parent–child pairs in the dataset did not provide ratings).

sion model, and the independent variable rating is used to predict the dependent variable rating. The difference between the rating predicted by the independent variable informant's rating and the dependent variable informant's actual rating becomes the residual difference score. Typically, the score is then standardized (e.g., converted into a z score) by dividing it by the standard deviation of the residual difference scores in their unstandardized form (e.g., Chi & Hinshaw, 2002).

The mathematical properties of the different measures of informant discrepancies dictate that the relations among informant discrepancies and variables of interest (e.g., informant characteristics) will be distinct, depending on differences between the variances of informants' ratings used to create the measures or the correlation between informants' ratings. For instance, when the variances of groups of informants' ratings are equal, raw difference scores and standardized difference scores will yield the same estimates of the relation between informant discrepancies and variables of interest. However, raw difference scores and standardized difference scores will yield maximally different estimates of the relation between informant discrepancies and variables of interest when the variance of one group of informants' ratings is maximally larger than the other group of informants' ratings. Moreover, the correlation between informants' ratings will determine the estimates of the relation between informant discrepancies and variables of interest produced by residual difference scores. For example, when the correlation between informants' ratings is low (e.g., variance predicted in the dependent variable informant's rating by the independent variable informant's rating is less than 5%), the residual difference score will be almost completely predicted by the dependent variable informant's rating, and the relation between the independent variable informant's rating and the residual difference score will be zero. As the correlation between informants' ratings increases, the relation between the residual difference score and the dependent variable informant's rating will decrease, and the relation between the independent variable informant's rating and the residual difference score will remain zero. The relation between the independent variable informant's rating and the residual difference score is always zero, because by definition, the residual difference score is composed of variance in the dependent variable informant's rating that the independent variable informant's rating is not related to.

Illustration of the Reviewed Measures of Informant Discrepancies

To illustrate key differences among measures of informant discrepancies, we drew parents' and children's ratings from a sample of clinic-referred children with aggression and their mothers. This illustration conveys how the different measures vary in how they relate to both the informants' ratings that are used to create them and to demographic and family characteristics.

Method

Participants

Participation was initiated by families who contacted a triage center at a child psychiatry service that serves a large catchment area or by direct contact with the clinic. Children referred for oppositional, aggressive, or antisocial behavior were seen at the Yale Child Conduct Clinic, an outpa-

tient treatment service for children and families. After referral, children and families completed an initial evaluation to assess child, parent, and family functioning, and then began treatment. The study included 389 children (96 girls and 293 boys) and families. Children ranged in age from 6 to 16 years ($M = 10.1$, $SD = 1.9$); 236 (60.7%) of the children were European American, 116 (29.8%) were African American, 16 (4.1%) were Hispanic American, and 21 (5.4%) were of other groups or mixed background. Diagnoses of the children were obtained from the parent version of the Research Diagnostic Interview (RDI; Kazdin, Siegel, & Bass, 1992). The interview assessed the presence, absence, and duration of child symptoms on the basis of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 1987). The child version of the RDI was also administered but was not used for the purposes of delineating the diagnoses below. The reliability of Axis-I diagnoses was routinely assessed by independent observers for randomly selected cases ($n = 30$) and yielded high agreement ($\kappa = .90$ across all diagnoses). Principal Axis-I diagnoses included conduct disorder (CD; 42.4%), oppositional defiant disorder (ODD; 30.3%), attention-deficit/hyperactivity disorder (ADHD; 4.9%), major depressive disorder (15.9%), other disorders (2.3%), or no diagnosable Axis-I disorder (4.1%).

The primary caretaker of the child ranged in age from 22 to 62 years ($M = 36.5$, $SD = 6.7$) and included biological (90.9%), adoptive (2.6%), stepmother (1.8%), foster mother (1%), or other relatives (3.7%); 49.9% of the children were from single-parent families. The head of the household's occupational and educational level (Hollingshead, 1975) was distributed from lower to higher sociooccupational classes: Class I = 12.3%, Class II = 21.7%, Class III = 27.7%, Class IV = 24.1%, and Class V = 14.1%. Median monthly family income was from $1,501 to $2,000 (range from $0–$500 to > $5,000).

Assessment

Measures were completed before treatment and drew on multiple assessment formats (interviews, questionnaires) and informants (parents and children). Four domains of functioning were assessed.

The first domain, *mothers' and children's ratings of child externalizing symptoms*, was measured using the RDI. The RDI has both a parent and child version (RDI-P and RDI-C, respectively), administered by separate trained interviewers who were each unaware of the results of the other administration. Items on the RDI were rated 0 (*definitely no*), 1 (*perhaps, sometimes, or maybe*), or 2 (*definitely yes*). The RDI measured the total number of child externalizing symptoms as rated by both the mother and child. A total externalizing symptom rating was computed for both mother about child and child about self by counting the total number of CD, ODD, and ADHD symptoms endorsed by the mother and child. All symptoms rated a 1 or 2 were counted as endorsed symptoms.[2]

The second domain, *maternal psychopathology*, was measured with the Hopkins Symptom Checklist (SCL-90; Derogatis & Cleary, 1977), a 90-item scale reflecting the degree of discomfort across several symptom areas. The total score was used as an overall index of psychiatric dysfunction of the parent. The psychometric properties of the SCL-90 have been studied extensively. The third domain, *maternal stress,* was measured using the Parenting Stress Index (PSI; Abidin, 1990), a 120-item scale reflecting areas of perceived stress related to the child (e.g., demandingness, mood) and to the mother's views of her own functioning (e.g., restrictions of role, social isolation). The items yield a total perceived stress

---

[2] The conduct disorder section of both the RDI child and parent versions contains a question on sexual activity. Mothers were always asked this question, regardless of the age of the child. However, only children 12 years or older were asked this question. For those children under 12 years of age, this question was prorated on the basis of the number and severity of symptoms endorsed in the child self-rated conduct disorder section.

score, which was used for this study. Reliability and validity data for the PSI have been reported elsewhere (Abidin, 1990). The final domain, *socioeconomic status and child characteristics*, included two measures: an index of the head of the household's socioeconomic status (SES; Hollingshead's, 1975, Head of the Household 4-Factor Index [$M = 36.1$, $SD = 14.4$]) and child age.

Mother–Child Symptom Ratings and Their Discrepancies

Three measures were computed to assess mother–child discrepancies: the raw difference score, standardized difference score, and residual difference score. The raw difference score was created by subtracting the RDI-C total externalizing symptoms from the RDI-P total externalizing symptoms. The standardized difference score was created by first converting the RDI-C total externalizing symptoms and the RDI-P total externalizing symptoms into z scores and then subtracting the z score of the RDI-C total externalizing symptoms from the z score of the RDI-P total externalizing symptoms. Lastly, the residual difference score was created from a regression model, in which the RDI-P total externalizing symptoms was the dependent variable and the RDI-C total externalizing symptoms was the independent variable. The RDI-C total externalizing symptoms was used to predict the RDI-P total externalizing symptoms, and the difference between the rating predicted by the RDI-C total externalizing symptoms and the actual RDI-P total externalizing symptoms rating became the residual difference score. The residual difference score was then standardized by dividing it by the standard deviation of the residual difference scores.

Results

Correlations Between Measures of Discrepancies, Mother-, and Child-Rated Symptoms

Means and standard deviations for the child-completed and mother-completed measures are presented in Table 1. Correlations between mothers' ratings of children, children's self-ratings, and measures of informant discrepancies are presented in Table 2. The Pearson product–moment correlation between the RDI-P total ex-

Table 1

Means and Standard Deviations of Measures, and t Tests to Compare Parent and Child Scores

	Parent		Child		RDI comparisons
Measure	M	SD	M	SD	t(776)
Research Diagnostic Interview					
Total CD symptoms	3.33	2.28	3.34	2.88	< 1
Total ODD symptoms	7.22	1.83	4.90	2.48	14.83**
Total ADHD symptoms	10.01	3.34	6.27	4.04	14.06**
Total externalizing symptoms	20.55	5.48	14.51	8.20[a]	12.09**
SCL-90 total score	55.41	46.19			
PSI total score	261.51	45.58			

Note. $N = 389$. RDI = Research Diagnostic Interview; CD = conduct disorder; ODD = oppositional defiant disorder; ADHD = attention-deficit/hyperactivity disorder; SCL-90 = Hopkins Symptom Checklist; PSI = Parenting Stress Index.
[a] Variance of RDI-C (child version) total externalizing symptoms > variance of RDI-P (parent version) total externalizing symptoms, Levene's $F(1, 776) = 70.98$, $p < .001$.
** $p < .001$.

Table 2

Correlations Between Informants' Ratings, Measures of Informant Discrepancies, and Informant Characteristics for the Total Sample

Variable	Raw difference score	Standardized difference score	Residual difference score
RDI-P total externalizing symptoms	.47**	.66**	.99**
RDI-C total externalizing symptoms	−.81**	−.66**	.00
SCL-90 total score	.06	.11*	.21**
PSI total score	.19**	.26**	.39**
SES	−.01	−.04	−.12*
Child age	−.29**	−.23**	.01

Note. $N = 389$. RDI = Research Diagnostic Interview; P = parent version; C = child version; SCL-90 = Hopkins Symptom Checklist; PSI = Parenting Stress Index; SES = Hollingshead (1975) Head of Household 4-Factor Index of socioeconomic status.
* $p < .05$.   ** $p < .001$.

ternalizing symptoms and RDI-C total externalizing symptoms was significant, but low, $r(387) = .14$, $p < .01$. To illustrate the differences in the relationships among measures of informant discrepancies and mother-rated and child self-rated externalizing symptoms, we computed correlations between the raw difference score, standardized difference score, and residual difference score and the RDI-P total externalizing symptoms and RDI-C total externalizing symptoms.

The residual difference score correlated almost perfectly with the RDI-P total externalizing symptoms but did not correlate at all with the RDI-C total externalizing symptoms (see Table 2).[3] Thus, any correlation between the residual difference score and another variable will be essentially identical to the correlation between that same variable and the mothers' ratings. In addition, the variance accounted for in the residual difference score by the children's self-ratings is partialed out, reflected by the children's self-ratings correlating with the residual difference score at $r = .00$. Therefore, any variable or characteristic that may influence the children's ratings and make them discrepant from the mothers' ratings will not be detected by correlations with the residual difference score.

Different relationships from those illustrated with the residual difference score and mother-rated and child self-rated externalizing symptoms were evident with the raw difference score (see Table 2). The raw difference score correlated with the RDI-P total

---

[3] The high correlation between the residual difference score and the RDI-P total externalizing symptoms is partially a product of the low correlation between the RDI-P total externalizing symptoms and the RDI-C total externalizing symptoms. Indeed, the correlation between these two ratings was at $r = .14$, making the variance accounted for in the mothers' ratings by the children's self-ratings less than 2%. However, if one were to create the same residual difference score using mother and child ratings that correlated far higher than what has typically been found in previous research (see Achenbach et al., 1987), for instance, $r = .40$, the variance accounted for in the mothers' ratings by the children's self-ratings would still be quite low (16%). Thus, the resulting residual difference score in this instance would still suffer from a high correlation with the mothers' ratings.

externalizing symptoms and RDI-C total externalizing symptoms in the expected directions (i.e., positive for mothers' ratings, negative for children's self-ratings) because the raw difference scores were created by subtracting the children's self-ratings from the mothers' ratings of children. Thus, as expected, greater positive raw difference scores were associated with both greater mothers' ratings and lower children's self-ratings, whereas greater negative raw difference scores were associated with both greater children's self-ratings and lower mothers' ratings. There were substantial differences between the magnitudes of the correlations between the mothers' ratings and children's self-ratings and the raw difference score; this is consistent with a concern of raw difference scores that when the variances between the ratings from which they were derived differ, the ratings with the greater variances will always contribute more to the resulting difference score (Edwards, 1994). Indeed, there was significantly greater variability in the children's self-ratings than in the mothers' ratings (see Table 1), resulting in the children's self-ratings having a substantially high correlation with the raw difference scores, relative to the mothers' ratings.

Substantially different relationships from those illustrated with either the residual difference or raw difference scores and mother-rated and child self-rated externalizing symptoms were evident with the standardized difference score (see Table 2). Similar to the raw difference score, the standardized difference score correlated with the RDI-P total externalizing symptoms and RDI-C total externalizing symptoms in the expected directions (i.e., positive for mothers' ratings, negative for children's self-ratings) because the standardized difference scores were created by subtracting the children's self-ratings from the mothers' ratings of children. As expected, greater positive standardized difference scores were associated with both greater mothers' ratings and lower children's self-ratings, whereas greater negative standardized difference scores were associated with both greater children's self-ratings and lower mothers' ratings. However, the standardized difference score was the only measure reviewed for which the mothers' ratings and children's self-ratings correlated equally with the measure; this was expected, given that the mothers' and children's ratings were placed on the same metric (z distribution) and thus had the same variance.

Correlations Between Measures of Discrepancies and Child and Family Characteristics

The measures vary in their relations to child and family characteristics (see Table 2) and can be distinguished by their relations to the RDI-P total externalizing symptoms. For instance, on three of the informant characteristics examined, the correlations between the residual difference score and these characteristics were not significantly different from these same characteristics' correlations with the RDI-P total externalizing symptoms, SCL-90 total score, $r(387) = .22$, $p < .001$; PSI total score, $r(387) = .39$, $p < .001$; SES, $r(387) = -.13$, $p < .001$. Moreover, the residual difference score did not correlate with child age. Because child age was a characteristic distributed across children in the sample (i.e., not all mothers in the sample had children of the same age), our illustration suggests that child-specific characteristics related to informant discrepancies were lost when using the residual difference score.[4]

The raw difference score was inconsistent in its correlations between child and family characteristics. However, unlike the residual difference score, the significant correlations that were found could be distinguished from the RDI-P total externalizing symptoms. The raw difference score correlated with child age and the PSI total score, and the correlations between the raw difference score and both child age and the PSI total score were significantly different from the correlations between the RDI-P total externalizing symptoms and child age, $r(387) = .07$, ns, Williams (1959) $t(386) = -7.27$, $p < .001$; and the PSI total score, $r(387) = .39$, $p < .001$, Williams $t(386) = -4.01$, $p < .001$.[5]

Lastly, the standardized difference score produced the most consistent relationships with child and family characteristics, and unlike the residual difference score, each relation could be distinguished from the RDI-P total externalizing symptoms. With the exception of SES, the standardized difference score correlated with each of the informant characteristics examined (child age, SCL-90 total score, and PSI total score). Moreover, the correlations between the standardized difference score and child age, the SCL-90 total score, and the PSI total score were significantly different from the correlations between the RDI-P total externalizing symptoms and child age, $r(387) = .07$, ns, Williams $t(386) = -7.61$, $p < .001$; the SCL-90 total score, $r(387) = .22$, $p < .001$, Williams $t(386) = -2.69$, $p < .01$; and the PSI total score, $r(387) = .39$, $p < .001$, Williams $t(386) = -3.35$, $p < .001$.

Discussion and Recommendations

The mathematical properties of the different measures of informant discrepancies reviewed can lead to different conclusions, as a function of the differences between the variances of informants' ratings or the correlation between informants' ratings. This was illustrated with an empirical demonstration that showed that the different measures of discrepancies produce distinct estimates of the relations between informant discrepancies and informant characteristics, such as maternal psychopathology and stress, SES, and child age. The three measures reviewed are definitely not interchangeable measures of informant discrepancies, and future work

---

[4] As an additional illustration of the residual difference score, we computed a second score, entering the RDI-P total externalizing symptoms as the independent variable and entering the RDI-C total externalizing symptoms as the dependent variable. Unlike the illustration of the first residual difference score, this score did not correlate with any of the maternal characteristics (psychopathology, stress, SES) because the variability of the mothers' ratings was partialed out to create this score. Indeed, the only relation found between any of the informant characteristics and this second residual difference score was for child age, and this relation was nearly identical to, and not significantly different from, the relation between the RDI-C total externalizing symptoms and child age, $rs(387) = .36$ vs. 37, ps $< .001$. This illustration is consistent with our illustration of the first residual difference score and provides additional support for the differences between the measures of informant discrepancies presented in the above illustration.

[5] The Williams t test for comparing differences between dependent correlations was used to compare differences between the relations between measures of discrepancy and informant characteristics and the relations between the RDI-P total externalizing symptoms and informant characteristics, as recommended by previous investigators (see Steiger, 1980).

examining informant discrepancies in clinical child research must pay careful attention to the measure of informant discrepancies chosen.

Of the three measures reviewed, the standardized difference score was the only measure that correlated equally with each of the informants' ratings from which it was created. Moreover, the standardized difference score produced the most consistent estimates among informant discrepancies and informant characteristics and was statistically discernable from the informants' ratings from which it was created. Given its mathematical properties and relations to child and family characteristics, we recommend that future investigations use the standardized difference score as their measure of informant discrepancies. Indeed, that informants' ratings correlated equally with the standardized difference score is consistent with the notion that no one informant can be considered a "gold standard" by which to interpret another informant's ratings as representing an over- or underestimated rating of child psychopathology (see Richters, 1992). Thus, in investigations wishing to examine the relations between informant discrepancies and variables of interest (e.g., informant characteristics), it may be important that the measure of discrepancy chosen correlates equally with the informants' ratings used to create the measure. However, there may be other instances not illustrated in this article in which the standardized difference score may not be an optimal measure of informant discrepancies, and we encourage future investigations to identify instances in which other measures may be conceptually and methodologically optimal.

There are limitations to the present review and illustration. The focus on a clinic sample of children may yield results that would not apply to community samples. Differences among the various scores (measures of discrepancies) might be influenced by characteristics of the sample, because of the high rates of behavior problems. However, much of the literature on discrepancies among informants has used clinic samples, and that was the basis for selecting the present sample. Sampling and generality aside, the broader issue is the primary message of the article. Specifically, measures of informant discrepancies currently used in research are not interchangeable and yield different conclusions when correlated with various characteristics of the informants (e.g., child, parent). Further research on the factors that account for and contribute to informant discrepancies ought to consider what the appropriate measure of discrepancies is. One measure may be better suited for a particular purpose than another. Our goal was to convey that an accumulating body of research cannot be expected to produce consistent results because the measures used among the studies are not interchangeable.

References

Abidin, R. R. (1990). *Parenting Stress Index clinical manual*. Charlottesville, VA: Pediatric Psychology Press.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

Chi, T. C., & Hinshaw, S. P. (2002). Mother–child relationships of children with ADHD: The role of maternal depressive symptoms and depression-related distortions. *Journal of Abnormal Child Psychology, 30,* 387–400.

Derogatis, L. R., & Cleary, P. A. (1977). Confirmation of the dimensional structure of the SCL-90: A study in construct validation. *Journal of Clinical Psychology, 33,* 981–989.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58,* 51–100.

Hollingshead, A. B. (1975). *Four-factor index of social status.* New Haven, CT: Yale University.

Kazdin, A. E. (1989). Identifying depression in children: A comparison of alternative selection criteria. *Journal of Abnormal Child Psychology, 17,* 437–455.

Kazdin, A. E., Siegel, T., & Bass, D. (1992). Cognitive problem-solving skills training and parent management training in the treatment of antisocial behavior in children. *Journal of Consulting and Clinical Psychology, 60,* 733–747.

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 1078–1085.

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112,* 485–499.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245–251.

Williams, E. J. (1959). The comparison of regression variables. *Journal of the Royal Statistical Society, Series B, 21,* 396–399.

Yeh, M., & Weisz, J. R. (2001). Why are we here at the clinic? Parent–child (dis)agreement on referral problems at outpatient treatment entry. *Journal of Consulting and Clinical Psychology, 69,* 1018–1025.

Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology, 68,* 1038–1050.