Discussion

DISCUSSION

In this thesis we examined multiple informant ratings of children's emotional and behavioural problems. In chapter 2 and chapter 3 we showed how ratings of multiple informants changed across age and across time. The sections **Normative development** and **Prevalence** of internalizing and externalizing problems will discuss implications of multiple informants' ratings based on the findings in these chapters. Chapters 4, 5, 6, and 7 examined how discrepancies between informants were associated with adverse outcomes. We used different methods to investigate these associations. In the section **Methods for analysing ratings from multiple informants** we will review these and alternative methods. The next section **The accelerated longitudinal design** describes the strengths and weaknesses of the design of the Zuid-Holland study, which we used in chapters 2, 3, 6, and 7. The section **Clinical implications** focuses on selecting the optimal informant and integrating ratings of multiple informants. The discussion ends with recommendations for **Future research**.

Normative development

In Chapter 2 we demonstrated that parent-reported internalizing problems increased across age and that parent-reported externalizing problems decreased across age. We showed that, similar to parent reports, self-reported internalizing problems increased across age, but that self-reported externalizing problems also increased across age in contrast with parent reports. Thus, changes in internalizing and externalizing problems across age depend on the informant who reported the problems. These findings are in line with the findings from our longitudinal study in chapter 3, in which we observed increasing differences across age between parents and youths reporting internalizing and externalizing problems. Our studies underline the importance of taking the informant into account when evaluating normative development of internalizing and externalizing problems. For example, in the study by Bongers, Koot, van der Ende, and Verhulst (2003), based on the same data as we used in chapter 3, parent reports of internalizing problems, not surprisingly, showed an increase across age, and externalizing problems decreased across age. The title of this paper that we chose was "The normative development of child and adolescent problem behavior", omitting the informant, in this case the parent, in the title. As we have shown in chapters 2 and 3, there is no single normative development of child problem behaviours, but multiple ways of regarding normative development depending on the source of information.

Prevalence

The differences between informants we observed in chapters 2 and 3 are also important in evaluating prevalence studies (Bronsard et al., 2016; Polanczyk, Salum, Sugaya, Caye, &

Rohde, 2015; Roberts, Attkisson, & Rosenblatt, 1998). These studies often employ a two-stage design (Pickles, Dunn, & Vazquez-Barquero, 1995). In the first stage, the screening stage, a sample is approached with a questionnaire to determine the level of problem behaviour of children. Next, two groups are selected, namely one consisting of, often all, children with high scores and one of a random sample of children with scores in the normal range. The children in the two groups will then be interviewed to obtain psychiatric diagnoses. Based on a combination of the two stages, prevalence rates of psychiatric disorders can be estimated. In both stages it is possible to obtain ratings from multiple informants. For example, in our prevalence study among 13-18-years-olds (Verhulst, Van der Ende, Ferdinand, & Kasius, 1997), we obtained parent, teacher, and self-reports of problem behaviour in the screening stage. To arrive at a single score indicating problem behaviour, we transformed the raw score of each informant into z-scores and averaged these z scores. In this way a score was obtained even if ratings of one or two informants were missing. The computed average, however, hampered examining the influence of the different informants in the final prevalence estimates. A cut-off based on the averaged z scores was used to select all children with high scores and a random sample of children with low scores for the interviewing stage. We conducted interviews with the selected children and their parents separately. To arrive at a single prevalence rate, we used two often used approaches for combining the results of the interviews. One approach is to use an "or" rule: a child receives a diagnosis if according to one or both informants of a pair of informants the criteria for a diagnosis are fulfilled. Alternatively, the "and" rule states that a child only receives a diagnosis if two informants agree on the criteria for a diagnosis. The former approach will obviously result in higher prevalence estimates than the latter approach. In our study, the prevalence of any disorder was 35.5% with the "or" rule, whereas with the "and" rule the prevalence was 4.0%. A disadvantage is that the "or" and "and" approaches will quickly become more complex when there are more than two informants (Martel, Markon, & Smith, 2017). Reviews of prevalence studies on psychiatric disorders among children also suggest that prevalence rates depend (i.e., unequal prevalence estimates based on ratings from different informants) on the informant of the problems (Bronsard et al., 2016; Polanczyk et al., 2015; Roberts et al., 1998). For example, the prevalence rate of any disorder was 15% when the child was the only informant, 14% when the parent was the only informant, and 9% when the teacher was the only informant (Roberts et al., 1998). Thus, for general groupings of disorders and typical raters, child reports lead to the highest prevalence rates, whereas teacher reports tend to lead to the lowest prevalence rates. However, for specific disorders, different rankings of prevalence rates across informants may appear. For example, Thomas, Sanders, Doust, Beller, and Glasziou (2015) presented a different ranking of prevalence estimates of ADHD than for general groupings of disorders. Now teacher reports yield the highest prevalence rates, then parent reports, and the lowest are from child reports.

Thus, prevalence rates depend largely on the informant of a child's problems and also on how ratings of multiple informants are combined. These differences between informants make it hard to compare prevalence rates across studies. It is therefore recommended to report separate estimates for different informants in addition to estimates based on combining informants (Roberts et al., 1998).

Methods for analysing ratings from multiple informants

In our studies, we used various methods to incorporate ratings from multiple informants in the analysis. To investigate how informant discrepancies predicted adverse outcomes, we used in chapter 4 regression analyses, which included difference scores as predictors in addition to the scores of single raters. In chapter 2 and 3 we employed multilevel models to examine informant differences across age. In chapter 5, we investigated determinants of differences between mother- and self-reported problems of young children with latent profile analysis. In chapter 6 we used cross-lagged models to illustrate differences in the associations between parent- and teacher-reported problems and academic functioning. We used confirmatory factor analysis - multi-trait multi-method models (CFA-MTMM) to examine how discrepancies between self-, parent-, and teacher-reported problems predicted disorders in adulthood in chapter 7. This array of available methods to investigate informant discrepancies raises the question which method is the best for which question. To answer this question, Martel et al. (2017) reviewed a vast number of methods to manage ratings from multiple informants and concluded that most methods have limitations. For example, manifest difference scores as we used in chapter 4 are just reformulations of the separate informant scores (Laird & De Los Reyes, 2013; Laird & Weems, 2011) and have no additional information. An alternative approach to difference scores is polynomial regression (Laird & De Los Reyes, 2013), which provides more direct tests of informant discrepancies predicting adverse outcomes. In chapter 3 we examined absolute differences between informant ratings across age in a longitudinal study. With absolute differences it is difficult to determine which informant has higher scores than another informant. Multivariate multilevel regression analysis (Baldwin, Imel, Braithwaite, & Atkins, 2014) allows modelling changes of separate informant ratings across age or time and differences between informants at the same time.

A benefit of combining ratings from multiple informants is the increase in precision of the estimate of the score (Goldwasser & Fitzmaurice, 2001). More precision can be achieved, for example, by simple averaging the scores of multiple informants (Javaras, Goldsmith, & Laird, 2011) or by constructing latent variables (Barbot et al., 2016). The underlying assumption, however, is that the ratings are exchangeable. This assumption may not hold as previous research has shown that there is only low to moderate agreement between ratings of multiple informants (Achenbach, McConaughy, & Howell, 1987; De Los Reyes et al., 2015). In this case, examining informants' ratings separately would be

valuable as well as using methods that can show the contribution of each informant and discrepancies between the informants at the same time such as multivariate regression analysis (Goldwasser & Fitzmaurice, 2001) and CFA-MTMM analysis (Geiser, Hintz, Leonard Burns, & Servera, 2019). The latter approach comprises methods specifically suited for structurally different ratings (Eid et al., 2008), which cannot easily be replaced by another (informants providing the ratings are specifically selected) as opposed to interchangeable ratings, which are sampled from similar ratings (e.g., students who rate the teaching quality of a teacher). Ratings of, for example, parents and teachers regarding children's emotional and behavioural problems are considered structurally different, because they observe children in different contexts (e.g., at home or school) and have different insights (e.g., professional versus not professional) (Alexander, McKnight, Disabato, & Kashdan, 2017). These methods based on MTMM analysis enable separating trait from informant effects and can be valuable in evaluating determinants or outcomes of informant discrepancies.

The accelerated longitudinal design

We used data from two longitudinal studies for investigating informant discrepancies. The Zuid-Holland study, which provided the data for our analyses in chapters 3, 4, 6, and 7, is a multi-cohort study. In this study age cohorts from 4 to 16 years are followed across time. The number of assessment waves was seven, which resulted in a total follow-up time of 24 years. During the first eight years five assessments took place, each two years apart. The sixth assessment was conducted six years after the fifth, and the last assessment ten years after the sixth. This type of design is a cohort-sequential or accelerated longitudinal design, in which multiple cohorts are overlapping time periods. For example, a participant of four years at baseline is assessed from four years until 28 years, whereas a participant of 16 years is assessed from 16 until 40 years. Thus only the age range between 16 and 28 years can comprise the same assessments. An advantage of this design is that with a short longitudinal period a long developmental period can be investigated. In our study, the total follow-up time of 24 years covers the developmental period from four to forty years. With data from the first five waves of the Zuid-Holland study Stanger, Achenbach, and Verhulst (1997) showed different growth patterns of aggressive and delinquent behaviour across ages four to 18 years while the follow-up time was only 8 years. We used the accelerated design of the Zuid-Holland study in chapter 3 on informant differences across age and in chapter 7 on the bidirectional pathways between academic functioning and problem behaviours. There are, however, some weaknesses pertaining to the accelerated design. Although a common trend across cohorts over time can be estimated, the assumption that the trends are equal among the cohorts has to be tested. We tested this assumption in chapter 7 where we did not find differences between the cohorts in the effects of academic functioning on problem behaviour or vice versa. Another weakness

of the accelerated design is that it is difficult to examine determinants of the commend trend, because baseline assessments of determinants are connected with the cohort ages at baseline. In chapter 4 on informant discrepancies predicting adverse outcomes and in chapter 6 on MTMM models we used a single cohort design. We used narrow age ranges to avoid overlapping ages and cohorts.

In chapter 5 on latent profiles of informant differences between young children and parents, we used data from the Generation R study (Kooijman et al., 2016). This is a single birth cohort that is followed from foetal life onwards. Currently participants are about thirteen years old, but we analysed data from assessments until the participants were 6 years old. All assessments were conducted when the participants were at the same age, although due to logistical reasons of collecting data from more than 6,000 participants there is sometimes variation in the assessment times. Statistical techniques, however, can adjust for this variation.

Clinical implications

An important issue in the area of multi-informant studies is still open for debate. Who is the best informant? More specifically, are there different conditions, such as purpose of the assessment, type of problem, and gender and age of the child in which a certain informant is preferred above another?

One way to determine optimal informants for assessing children's problems is to ask the opinion from mental health professionals or informants themselves. For example, Loeber, Green, and Lahey (1990) asked clinicians and researchers in the field of child and adolescent psychiatry who according to their judgment is the best informant of children's emotional and behavioural problems. Phares (1997) used similar questions to obtain the opinion of mothers and fathers. There was much agreement among the participants of both surveys on the ranking of informants. They suggested that children were the least useful reporters of their externalizing problems. Teachers were better reporters than mothers of attention problems and hyperactivity, whereas mothers were better reporters than teachers of oppositional behaviour. Regarding internalizing problems, mothers were seen as the best informants, better than children, whereas teachers were seen as the least useful. Generally mothers and teachers were most often considered as the best informants, yet Loeber et al. (1990) and Phares (1997) suggested that ratings from multiple informants are needed evaluate children's problems. A limitation of both studies is that the respondents of the survey did not answer questions about the reasons of preferring one informant above another. Further, the findings of the two studies, however, are based on personal judgments and do therefore not necessarily reflect evidence from empirical studies.

Although several studies included ratings from multiple informants, only a few suggested a ranking of informants. For example, Loeber, Green, Lahey, and Stouthamer-Loe-

ber (1989) found in a sample of 177 boys of 7-12 years that regarding disruptive problems parents and teachers both contributed unique information and thus were both important in assessing disruptive problems. However, children themselves were less useful as informants. In chapter 6 we suggested that in the prediction of internalizing disorders, teachers provide additional information to self-reports of internalizing problems, whereas parents contribute to self-reports of externalizing problems in the prediction of externalizing disorders. In order to give a more systematically derived advice on selecting informants, Smith (2007) reviewed numerous studies, including clinical and general population samples addressing informant ratings. Based on findings from these studies rankings of parents, teachers, and children as informants were given for different conditions: younger versus older children, internalizing versus externalizing problems, and inpatients versus outpatients despite including general population samples. A rater was considered a best informant if previous studies revealed that the scores of the rater showed incremental validity or clinical utility (Smith, 2007). For younger children it was recommended that parents were the best informants for both internalizing and externalizing problems. Regarding internalizing problems, children were the second-best informants, whereas teachers were the second-best informants for externalizing problems. For older children, however, children were the best informants, followed by parents regarding internalizing problems. When rating externalizing problems, teachers were the best informants. In addition to these rankings based on empirical findings, Smith (2007) suggested that the ratings of preferred informants are meant as reference points to compare with the ratings of other informants.

If we need multiple informants, it brings us to another issue: How should we integrate ratings from multiple informants to be useful in clinical practice? To guide users of the Achenbach System of Empirically Based Assessment (ASEBA), Achenbach and Rescorla (2001) provide Q correlations that indicate the level of agreement between forms from multiple informants. Q correlations are correlations between item scores of scales from two different forms of one individual. For example, the item scores of parent-reported depressive problems are correlated with the same item scores of self-reported problems. If a form comprises eight scales and responses were obtained from a child, a mother, and a teacher there will be 24 Q correlations for this child. These Q correlations can be compared with mean Q correlations from reference groups (Achenbach & Rescorla, 2001). In this way a clinician can see whether the agreement between two informants is on the same level, is below average, or is above average compared to the reference group. These comparisons can support the clinician in examining the sources of disagreement between informants. A disadvantage of Q correlations is that a correlation between two informants can be perfect while summed scores can be very different. Moreover, a correlation can be near zero, while the summed scores can be similar. The side by side display of item scores of multiple informants and the Q correlations are helpful in guiding a clini-

cian in evaluating informant differences, but additional information may be beneficial. For example, like we have suggested in chapter 2, it would be helpful to have norms of difference scores. Thus, not only scores of single informants are compared with scores of a reference group, also differences between scores of two informants, e.g., of a parent and a teacher, are compared with differences in a reference group. Moreover, reports about specific items or scales where informants agree or disagree would support clinicians in obtaining follow-up information from informants.

Future research

Previous research has shown that there is only low to moderate agreement between informants of children's emotional and behavioural problems, despite using parallel instruments (Achenbach et al., 1987; De Los Reyes et al., 2015; Kaurin, Egloff, Stringaris, & Wessa, 2016). Instead of seeking instruments that are parallel, thus incorporating similar items that must be endorsed by multiple, but different informants, it might be better to have instruments that are different between informants (Kaurin et al., 2016). Current instruments already included items specifically for different informants. Compared with the CBCL, for example, the TRF (Achenbach & Rescorla, 2001) comprise additional items on ADHD. "Fidgets" and "Disturbs other children" are examples of items appearing on the TRF, but not on the CBCL. In this way, instruments can be optimized to the different situations in which informants observe children and to the particular views informants may have of children's problems. However, the CBCL and TRF are similar methods, they are both paper-and-pencil forms. It would also be beneficial to add other types of methods, such as interviews, observations, tests, wearables, and smartphones. To continue with the example of ADHD, additional methods could be observations in the classroom or a Continuous Performance Test (Epstein et al., 2003), which is a neuropsychological task to measure impulsivity and inattention. These different methods with specific contents for informants tap views of a person's behaviour from various angles. Different methods that provide disagreeing ratings of one's behaviour are favourable (Kraemer et al., 2003). Highly correlated methods only point reliably to the same behaviour, but methods with low correlations validly point to behaviour from different viewpoints. Kraemer et al. (2003) compared this to the Global Positioning System (GPS). When satellites are in one line they all provide the same position, which is a reliable position, but which may the wrong position. The right position, thus the valid position, can only be found when the satellites have different positions in the orbit. However, different instruments imply greater challenges in aggregating them. Kraemer et al. (2003) suggested to use Principal Component Analysis (PCA), which leads to three dimensions: Trait describes the characteristic of the person under study (e.g., anxiety or depression), context gives the conditions influencing the trait (e.g., home versus school), and perspective denotes the views of the informant that influences the assessment of the trait (e.g., parents versus

professionals). Martel et al. (2017) suggested Multitrait-Multimethod (MTMM) analysis based on structural equation models and machine learning as methods to overcome weaknesses of earlier strategies to integrate ratings form multiple informants. We used MTMM analysis in chapter 6 with ratings from children, parents, and teachers. Some variations of MTMM analysis also enable to incorporate dissimilar instruments (Geiser et al., 2019), although empirical studies on this topic are scarce. Machine learning methods are data-driven techniques, which are successfully used in prediction and classification of psychiatric disorders and other areas within child and adolescent psychiatry (Monuteaux & Stamoulis, 2016). For example, (Bone et al., 2016) examined whether machine learning methods could improve screening algorithms in classifying individuals with autism. They found that the new algorithms were more efficient than the conventional ones. To date, however, studies focusing on multiple informants are lacking. Much more research is needed to investigate when and when not these methods work in integrating data from multiple informants.

Creating visualizations of informant reports is another area in which advancements are needed. Although major tests (e.g., Achenbach & Rescorla, 2001; Conners, 2008; Kamphaus & Reynolds, 2015) provide computer scoring, offering additional information in addition to scoring forms on paper, which are still provided and used as well. These paper forms, however, seem to hamper developments in computer scoring, because the output of computer scoring must resemble the scoring forms on paper. Modern web-based visualizations supported by advanced computer algorithms could offer insights in informant discrepancies, which cannot be revealed on paper forms. An example could be a multidimensional plot of types of problems (e.g., internalizing and externalizing) in which informants are plotted as well. In this way the plot shows which informant differs (or is similar to) from another informant for certain types of problems. Additional information could also be shown in the plot, for example mean discrepancies between informants based on norm data or previous findings. Likewise, information on how predictive discrepancies are of adverse outcomes. This information may also be based on findings from previous studies. With only two types of problems and three informants this is still a simple plot, but it will soon be more complex with additional types of problems, informants, and details. However, computer-based plots can be made very user friendly. With only a few clicks, details can be removed or added from a plot enabling views of the data a user is interested in.

Conclusions

We showed that differences between informant ratings of problem behaviour became larger across age. Therefore, it is important that studies on normative development and prevalence studies take the informant into account. Further, we demonstrated that discrepancies between informant ratings contributed to the prediction of later disorders

and adverse outcomes over and above ratings provided by individual informants. Future research using new statistical methods could provide more insight how to best combine ratings from multiple informants.

REFERENCES

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*(2), 213-232.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Alexander, L. A., McKnight, P. E., Disabato, D. J., & Kashdan, T. B. (2017). When and How to use Multiple Informants to Improve Clinical Assessments. *Journal of Psychopathology and Behavioral Assessment, 39*(4), 669-679.

Baldwin, S. A., Imel, Z. E., Braithwaite, S. R., & Atkins, D. C. (2014). Analyzing multiple outcomes in clinical research using multivariate multilevel models. *Journal of consulting and clinical psychology, 82*(5), 920-930.

Barbot, B., Bick, J., Bentley, M. J., Balestracci, K. M. B., Woolston, J. L., Adnopoz, J. A., & Grigorenko, E. L. (2016). Changes in mental health outcomes with the intensive in-home child and adolescent psychiatric service: a multi-informant, latent consensus approach. *International Journal of Methods in Psychiatric Research, 25*(1), 33-43.

Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *Journal of child psychology and psychiatry, 57*(8), 927-937.

Bongers, I. L., Koot, H. M., van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology, 112*(2), 179-192.

Bronsard, G., Alessandrini, M., Fond, G., Loundou, A., Auquier, P., Tordjman, S., & Boyer, L. (2016). The Prevalence of Mental Disorders Among Children and Adolescents in the Child Welfare System: A Systematic Review and Meta-Analysis. *Medicine, 95*(7), e2622-e2622.

Conners, C. K. (2008). *Conners 3rd Edition*. Toronto, Canada: Multi-Health Systems.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin, 141*(4), 858-900.

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural Equation Modeling of Multitrait-Multimethod Data: Different Models for Different Types of Methods. *Psychological Methods, 13*(3), 230-253.

Epstein, J. N., Erkanli, A., Conners, C. K., Klaric, J., Costello, J. E., & Angold, A. (2003). Relations Between Continuous Performance Test Performance Measures and ADHD Behaviors. *Journal of Abnormal Child Psychology, 31*(5), 543-554.

Geiser, C., Hintz, F. A., Leonard Burns, G., & Servera, M. (2019). Structural equation modeling of multiple-indicator multimethod-multioccasion data: A primer. *Personality and Individual Differences, 136*, 79-89.

Goldwasser, M. A., & Fitzmaurice, G. M. (2001). Multivariate linear regression analysis of childhood psychopathology using multiple informant data. *International Journal of Methods in Psychiatric Research, 10*, 1-10.

Javaras, K. N., Goldsmith, H. H., & Laird, N. M. (2011). Estimating the effect of a predictor measured by two informants on a continuous outcome: A comparison of methods. *Epidemiology, 22*(3), 390-399.

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior Assessment System for Children - Third Edition (BASC-3)*. Bloomington, MN: Pearson.

Kaurin, A., Egloff, B., Stringaris, A., & Wessa, M. (2016). Only complementary voices tell the truth: a reevaluation of validity in multi-informant approaches of child and adolescent clinical assessments. *Journal of Neural Transmission, 123*(8), 981-990.

Kooijman, M. N., Kruithof, C. J., van Duijn, C. M., Duijts, L., Franco, O. H., van IJzendoorn, M. H., . . . Jaddoe, V. W. V. (2016). The Generation R Study: design and cohort update 2017. *European Journal of Epidemiology, 31*(12), 1243-1264.

Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*(9), 1566-1577.

Laird, R. D., & De Los Reyes, A. (2013). Testing Informant Discrepancies as Predictors of Early Adolescent Psychopathology: Why Difference Scores Cannot Tell You What You Want to Know and How Polynomial Regression May. *Journal of Abnormal Child Psychology, 41*(1), 1-14.

Laird, R. D., & Weems, C. F. (2011). The Equivalence of Regression Models Using Difference Scores and Models Using Separate Scores for Each Informant: Implications for the Study of Informant Discrepancies. *Psychological Assessment, 23*(2), 388-397.

Loeber, R., Green, S. M., & Lahey, B. B. (1990). Mental health professionals' perception of the utility of children, mothers, and teachers as informants on childhood psychopathology. *Journal of Clinical Child Psychology, 19*(2), 136-143.

Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology, 1*(4), 317-337.

Martel, M. M., Markon, K., & Smith, G. T. (2017). Research Review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of child psychology and psychiatry, 58*(2), 116-128.

Monuteaux, M. C., & Stamoulis, C. (2016). Machine Learning: A Primer for Child Psychiatrists. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(10), 835-836.

Phares, V. (1997). Accuracy of informants: Do parents think that mother knows best? *Journal of Abnormal Child Psychology, 25*(2), 165-171.

Pickles, A., Dunn, G., & Vazquez-Barquero, J. L. (1995). Screening for stratification in two-phase ('two- stage') epidemiological surveys. *Statistical Methods in Medical Research, 4*(1), 73-89.

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry, 56*(3), 345-365.

Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry, 155*(6), 715-725.

Smith, S. R. (2007). Making Sense of Multiple Informants in Child and Adolescent Psychopathology: A Guide for Clinicians. *Journal of Psychoeducational Assessment, 25*(2), 139-149.

Stanger, C., Achenbach, T. M., & Verhulst, F. C. (1997). Accelerated longitudinal comparisons of aggressive versus delinquent syndromes. *Development and Psychopathology, 9*(1), 43-58.

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-analysis. *Pediatrics, 135*(4), e994.

Verhulst, F. C., Van der Ende, J., Ferdinand, R. F., & Kasius, M. C. (1997). The prevalence of DSM-III-R diagnoses in a national sample of Dutch adolescents. *Archives of General Psychiatry, 54*(4), 329-336.