EVALUATING SCREENING GUIDELINES FOR DISRUPTIVE BEHAVIOUR PROBLEMS IN CHILDREN: A SYSTEMATIC REVIEW OF THE ACCURACY OF PARENTS' CONCERNS

Sarah Wells

Thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Interdisciplinary Health Sciences

School of Interdisciplinary Health Science
Faculty of Health Science
University of Ottawa

Abstract

The burden of suffering associated with disruptive behavior disorders (DBDs) is considerable and may influence the health outcomes of affected individuals across the life course. There is consensus that secondary prevention strategies which target early manifestations of DBDs in young children, known as disruptive behaviour problems (DBPs), are effective in facilitating early intervention and minimizing this burden of suffering. One example of such a strategy can be found within a current Canadian Paediatric Society practice guideline, which recommends that parents' concerns be used to inform decisions to screen for DBPs in children following the 24-month well-child visit until the age of 5 years (Charach, Bélanger, McLennan & Nixon, 2018). It is prudent to ensure accurate screening decisions within these visits, in effort to maximize the efficacy of screening for DBPs within such a program.

A systematic review was conducted to determine whether parents' concerns can provide primary care practitioners (PCPs) with enough information to justify a decision in favour of, or against, screening for DBPs, as is currently recommended. 6 electronic databases (Medline (Ovid), Embase (Ovid), Central (Ovid), CINAHL (EBSCOHost), PsycINFO (Ovid), and Eric (Ovid) and Scopus) were searched for diagnostic test accuracy studies that elicited parents' concerns about their child's behavioral, social and/or emotional development, and that measured the presence of DBPs in children aged 0-5 years. Studies that reported proportions of true positive, false positive, false negative, and true negative outcomes (or outcomes from which these measures could be generated) were included. Risk of Bias in eligible studies was assessed using the QUADAS-2. Results from included studies were synthesized to produce calibrated estimates of the accuracy of parents' concerns in the form of weighted kappa coefficients (Kraemer, 1992).

Just one study (Glascoe, MacLean & Stone, 1991) met all eligibility and exclusion criteria. Results from this study indicated that only moderate agreement (k(1,0) = 0.533, [0.501; 0.564]) could be found between the absence of DBPs and the absence of parents' concerns. Similarly, only fair agreement (k(0,0) = 0.255, [0.238;0.272]) could be found between the presence of DBPs and the presence of parents' concerns. These results provide some limited evidence that neither the presence, nor absence, of parents' concerns can provide enough accurate information to PCPs to justify decisions in favour of/against screening. However, the ability of systematic search methods to retrieve a sufficiently large body of homogeneous data relevant to this review's research objectives was limited and revealed a gap in the literature. Future research that seeks to replicate the methods of Glascoe et al. (1991) may facilitate more robust analyses of the accuracy of parents' concerns in this context.

Prospero Registration ID: CRD42021157492

Funding Statement

The authors received no financial support for the research or authorship of this review.

Keywords: Systematic Review; Disruptive Behaviours; Conduct Disorder; Oppositional Defiant Disorder; Diagnostic Screening; Diagnostic Test Accuracy

Acknowledgements

I would first like to express my gratitude to my supervisor, Raymond Baillargeon, for his guidance throughout my time as his student. I am deeply appreciative of his encouragement and support through the completion of this project amidst some very uncertain and difficult times. I would also like to thank my thesis advisory committee members, Matthew McInnes and Philippe Robaey, for their expertise and advice throughout the duration of this project.

I want to express my deepest gratitude to my friends and my family for their unwavering support, humour, and faith in me. Not many folks would agree to a wedding, a puppy, and a master's degree during a pandemic, and so to my husband, Alexander, I am also eternally grateful.

Lastly, I dedicate the work in this thesis to my late grandfather, Don – a man of few words and deep thoughts. He inspired constant curiosity in the mind of a budding young researcher, embodied perseverance in the face of hardship, and believed that the greatness of a man can be measured by the number of people he helps. May I be so lucky that his ethos weaves its way through the rest of my work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 – REVIEW OF CURRENT LITERATURE AND INTRODUCTION

1.1 An Overview of Disruptive Behavior Problems and Disorders

Disruptive behaviour problems (DBPs) may be conceptualized in a multidimensional framework of behaviours that shift accordingly with psychological development across the life course (Carter, Gray, Baillargeon & Wakschlag, 2013). Carter and colleagues present non-compliance, temper-loss, low concern for others, and aggression as four dimensions of DBPs that are present throughout the life course (Carter et al., 2013). The exact manifestations of these DBPs may shift in accordance with the developmental stage of the child in question (Carter et al., 2013). For example, both a reflexive "no" in response to caregiver requests, and an inability to receive and follow direction from work supervisors may be classified as non-compliant behaviour; however, these behaviours will likely occur at different stages in an individual's development (Carter et al., 2013).

The presence of DBPs within this framework are associated with symptoms of disruptive behavior disorders (DBDs) (Carter et al., 2013). DBDs are defined as "conditions involving problems in the self-control of emotions and behaviors" which "[manifest] in behaviors that violate the rights of others and/or that bring the individual into significant conflict with societal norms or authority figures" (APA, 2013, p. 461). For example, the DSM-5 includes such disorders as Oppositional Defiant Disorder (ODD) and conduct disorder (CD) among those categorized as DBDs (APA, 2013). Notably, early symptoms of several DBDs (ODD and CD included) tend to appear during young childhood, often as early as the preschool years (APA, 2013). As such, we may look to the early presentation of DBPs in infants and young children as both a reliable indicator of elevated risk (Copeland, Shanahan, Costello et al., 2009; Caye, Spadini, Karam et al., 2007), and as an opportunity to facilitate early prevention.

1.2 The Burden of Suffering

There is consensus in the literature that healthcare strategies aimed at the prevention of childhood mental illness may greatly reduce the burden of suffering across the life course (Charach, Mohammazadeh, Bélanger, et al., 2020; Waddell, McEwan, Shepherd, et al., 2005; Waddell, McEwan, Peters, et al. 2007; Offord et al., 1999). Offord et al. define burden of suffering as the cumulative impact of the morbidity, the continuity, the prevalence, and the financial and social costs associated with a disease (Offord, Kraemer, Kazdin et al., 1999). Early prevention strategies that target DBPs and DBDs may be especially important, given the considerable burden of suffering related to these disorders.

1.2.1 The Morbidity and Mortality of Disruptive Behavior Disorders

In the context of DBDs, morbidity related to the presence of these disorders in childhood is significant (Offord et al., 1999). Delinquency, academic underachievement, poor physical health, and comorbid mental health illnesses are common among children with DBDs (Offord & Bennett, 1994) - the impacts of which may persist across the life course. One systematic review and meta-analysis found that children with CD are at an elevated risk of experiencing a variety of negative, long-term psychological and physical health outcomes (Erskine, Norman, Ferrari et al., 2016). From these findings, children with CD had significantly higher odds of developing comorbid depression or anxiety ($OR = 2.10$, $p < 0.01$), and substance abuse related disorders ($OR=2.26$, $p<0.01$) in adulthood (Erskine et al., 2016). Similar outcomes have been reported for children with ODD. One longitudinal study reported that the odds of developing depression in adulthood (i.e., 18-21 years of age) were 2.4 times higher for individuals that had received a prior diagnosis of ODD in childhood (e.g., 9-10 years of age) (Copeland et al., 2009). This effect grew stronger as the age at which children were first diagnosed increased, where the odds of developing generalized anxiety disorder (GAD) or depression in adulthood were 6.3 and 2.8 times higher (respectively) for those individuals who

had received a prior diagnosis ODD in adolescence (i.e., 11-13 years of age) (Copeland et al., 2009). For individuals with DBDs, an elevated risk of comorbid mental health illness may contribute to a greater burden of suffering, particularly as the individual ages.

Additionally, individuals with DBDs tend to be more likely to engage in impulsive, risky, or delinquent behaviour. In the same systematic review as above, children with CD were found to have higher odds of smoking both cannabis and tobacco in adulthood (OR=1.83, OR=1.86 respectively), higher odds of pregnancy before 23 years of age (OR=2.69), higher odds of violent behavior and violent-crime related arrests (OR=3.52), and higher odds of failure to complete high school (OR=2.69) (Erskine et al., 2016). Higher rates of engagement in these behaviors may contribute to adverse physical and psychological health outcomes, and socioeconomic disadvantage for these individuals (Erskine et al., 2016). Further, there is an increased risk of early mortality among individuals with DBDs (Scott, Pedersen, Erskine et al., 2017). For example, one national cohort study found that individuals diagnosed with DBDs were "twice as likely to die [before the age of 32] compared to individuals without these disorders" (Scott et al., 2017, p. 258). This increase in risk was significantly higher in the presence of comorbid substance use disorders (mortality rate ratio (MRR)=7.88), and when the age at which a person was diagnosed with a DBD increased (MRR=7.37 for individuals older than 18 years) (Scott et al., 2017). Together, these findings emphasize the considerable burden of suffering experienced by individuals with DBDs, particularly for those individuals for whom diagnosis (and presumed intervention) does not occur until later in life.

1.2.2 The Prevalence of DBDs

Another factor that contributes to the burden of suffering associated with DBDs is the prevalence of these disorders within a population. One meta-analysis has estimated the global prevalence of mental health disorders (in general) among children and adolescents (i.e., 6-18 years of age) to be approximately 13.4% (95% CI: 11.3-15.9%) (Polanczyk, Salum, Sugaya et

al., 2015) . Within this age range, Polanczyk et al. (2015) estimated that children with a DBD accounted for 5.7% (95% CI: 4.0-8.1%) of the total population of children with mental health disorders (i.e., 5.7% of ~13% = 0.7%). A more recent meta-analysis reported the results of standardized diagnostic assessments (via DSM-IIIR or DSM-IV criteria) of younger children (aged 24-72 months) (Charach, Mohammadzadeh, Belanger et al. 2020). They estimated that the prevalence of mental health disorders (generally) in this population is 18.4% (95% CI: 12.3-24.4%) (Charach et al., 2020). This estimate is similar, albeit slightly higher than that which was reported by Polanczyk et al. (2015); however, no prevalence estimates specific to DBDs were reported for this younger cohort.

Also, it is essential to note that there is considerable between-studies heterogeneity reported by the reviews cited above (i.e., differences in sample demographics, severity and/or level of impairment among participants, clinical assessment methods, healthcare contexts etc.). This heterogeneity was identified by the authors as a limitation to the validity of the pooled prevalence estimates that have been published (Charach et al., 2020; Polanczyk et al., 2015). Further, it is difficult to estimate the true global prevalence of DBDs among younger age ranges in the absence of concrete data. As such, the presentation of these estimates here serves to highlight that, while it is known that a small proportion of older children with mental health disorders have one or more DBDs, our understanding of how the prevalence of DBDs among younger children impacts the burden of suffering requires more research.

1.2.3 The Financial and Social Costs of DBDs

Lastly, the financial and social costs associated with DBDs contribute to the overall burden of suffering related to these disorders. In general, the healthcare costs associated with childhood mental health disorders can be substantial and may pose a barrier to the accessibility of timely treatment (Reardon, Harvey, Baranowska et al., 2016). The financial costs associated with DBDs are no exception to this. For example, higher levels of personal financial costs (e.g.,

individual and family behavioural therapy) have been associated with ODD independent of other comorbid DBDs and in the presence of comorbid DBDs (Christenson, Crane, Malloy & Parker, 2016). Further, these personal healthcare costs tend to increase as children approach adulthood in many jurisdictions, where public funding becomes less readily available for adult psychiatric treatment (Christenson et al., 2016).

In addition to the personal healthcare costs associated with DBDs, a high level of public healthcare spending can be attributed to the long-term psychological healthcare needed to support individuals with DBDs. For example, children with CD tend to engage in high levels of public health service use (i.e., prescription medication, publicly-funded counselling, diagnostic testing etc.), resulting in greater public health expenditures across the life course relative to their peers (Foster & Jones, 2005; Matza, Paramore & Prasad, 2005; Leibson & Long, 2003). Given this, the financial strain associated with DBDs further impacts the burden of suffering associated with these disorders, not just for the affected individual, but for the larger community as well.

1.3 Minimizing the Burden of Suffering: Levels of Prevention and Prevention Strategies

1.3.1 Levels of Prevention

Overall, the burden of suffering related to DBDs is considerable. Preventive strategies that target DBDs and related DBPs may offer an opportunity to minimize this burden (Chacko, Granski, Horn et al., 2018; Bayer, Hiscock, Scalzo et al., 2009). Preventive strategies can be categorized into three main approaches: primary prevention, secondary prevention, and tertiary prevention (Jekel, Katz, Elmore, et al., 2007). Primary prevention strategies are those that target all individuals in a population who are susceptible to a disease, and which aim to prevent the onset of a disease. For example, routine childhood vaccination programs target all children within a specified age range and aim to prevent the onset of communicable diseases that are common within this population (i.e., chicken pox, measles etc.).

Secondary prevention strategies are those that seek to detect individuals with early (i.e., subclinical) forms of a disease, and which aim to prevent the onset of overt, clinical symptoms of disease. For example, breast cancer screening via mammography aims to identify small areas of dense, potentially cancerous tissue among individuals without overt, clinical symptoms of breast cancer (e.g., palpable tumours etc.). This is done with the intent of initiating early treatment, preventing the onset of overt, more severe disease symptoms for the individual, and improving health outcomes.

Lastly, tertiary prevention measures target individuals with clinical symptoms of a disease and seek to either reduce the severity of disease, slow or stop the progression of disease, and/or prevent negative health outcomes associated with disease progression. An example of a tertiary prevention strategy would be a program in which hospital in-patients who have had an ischemic stroke and show signs of motor function impairment receive intensive physical rehabilitation therapy very soon after admission to hospital. A program like this would seek to reduce or limit the severity of impairments to motor functioning that occurred following the stroke and prevent long-term negative health outcomes associated with these symptoms (e.g., reduced mobility negatively impacting quality of life).

1.3.2 Universal and. Targeted Screening

Screening programs that seek to identify subclinical or clinical disease symptoms can overlap somewhat with the first two levels of prevention strategies. Most secondary prevention strategies involve some form of disease screening. One example of such a program that fits well within this framework would be targeted screening, where only those individuals at elevated risk of disease are selected to be screened for subclinical disease. One example of such a program would be one which recommends that only women who are at an elevated risk for breast cancer (e.g., those with dense breast tissue, those with family history etc.) within a

certain target age range (e.g., 30-50 years of age) should receive annual breast cancer screening via mammography.

However, secondary prevention strategies could also incorporate universal screening, where all individuals in a target population (e.g., a specific age range) are screened for subclinical disease. An example of a universal screening program would be one which recommends that all women between the ages of 50-74 years of age receive breast cancer screening via mammography every 2 years, regardless of their individual level of risk. This concept overlaps somewhat with primary prevention strategies, which seek to expose all individuals in a population (regardless of individual-level risk) to an intervention in effort to prevent to prevent disease onset. In this case, however, the goal of such intervention may be the early identification of subclinical disease and prevention of severe disease onset. Universal screening programs, therefore, overlap somewhat between primary and secondary level prevention strategies.

1.4 Prevention of DBDs and DBPs: Current Recommendations

One systematic review of randomized clinical trials of disruptive behavior interventions identified secondary prevention strategies as the most effective at preventing DBPs in preschool-aged children relative to other levels of prevention (Bayer, Hiscock, Scalzo et al., 2009). Three separate secondary prevention strategies involving targeted intervention were identified by Bayer et al. (2009) as interventions that most effectively reduced negative outcomes among preschool children at risk for DBPs (e.g., the US Family Checkup programme, the Triple P programme, and the Incredible Years Group Parenting programme). However, among the interventions reviewed by Bayer et al. (2009), secondary prevention strategies that promoted early, routine monitoring of children's psychological development were conspicuously absent. For example, secondary prevention strategies that incorporate early developmental surveillance during Well-Baby visits have been proposed elsewhere as one

strategy by which DBPs and DBDs can be detected (Lipkin, Macias & AAP Council on Children with Disabilities, Section on Developmental and Behavioral Pediatrics, 2020; Committee on Practice and Ambulatory Medicine & Bright Futures Periodicity Schedule Workgroup, 2020). Given that DBPs tend to onset within the first two years of life and have been shown to persist as age increases (Baillargeon, Keenan & Cao, 2012; Petitclerc & Tremblay, 2009), such early monitoring programs may represent an excellent opportunity to prevent the onset of severe disease and encourage a reduction in suffering through early intervention.

A current Canadian Paediatric Society practice guideline represents another example of early monitoring as a secondary prevention strategy in a Canadian health context (Charach, Bélanger, McLennan et al., 2017). It recommends that parents' concerns be used to inform the screening decisions for DBPs in children beginning at the 24-month well-child visit and continuing until the child reaches 5 years of age. This guideline specifically advocates for primary care providers (PCPs) to routinely elicit parents' concerns to assist in deciding in favour of/against screening for DBPs, and that *if* parents' concerns are present, then PCPs should always decide in favour of screening. Otherwise, when parents' concerns are not present, PCPs should decide against screening. In this systematic review of the literature, we sought to determine what, if any, evidence exists in the literature that would support the use of parents' concerns in the context of clinical screening decisions for DBPs.

1.5 Research Objective & Questions

1.5.1 Research Objective

The purpose of this research was to determine, via a systematic review of the literature, whether parents' concerns can provide PCPs with enough information to decide in favour of/against screening for DBPs.

1.5.2 Research Questions

1. To determine if the mere presence of parents' concerns justifies deciding in favour of screening for DBPs in children?
2. To determine if the mere absence of parents' concerns justifies deciding against screening for DBPs in children?

CHAPTER 2 – RESEARCH DESIGN AND METHODS

2.1 Protocol & Registration

A protocol for this review was created and registered on Prospero. It can be accessed electronically (registration number: CRD42021157492).

2.2 PITO Elements & Eligibility Criteria

The population of interest in this review were children between the ages of 0-5 years, and their parent/caregiver. The index test was the presence or absence of parents' psychological concerns about their child's behaviour. Herein, psychological concerns are defined as any concern about the social, emotional, and/or behavioral development of the child. The reference standard was the presence of absence of DBPs in our population of interest.

The outcomes of interest that were assessed by this study are as follows:

1. The proportion of children with a DBP and whose parents have behavioural concerns (true positives, or TP)
2. The proportion of children with a DBP whose parents do not have behavioural concerns (false negatives, or FN)
3. The proportion of children without a DBP and whose parents do not have behavioural concerns (true negatives, or TN)
4. The proportion of children without a DBP and whose parents have behavioural concerns (false positives, or FP)

Studies that met the following eligibility criteria were included for review:

1. had investigated DBPs in children between the ages of 0-5 years
2. had systematically elicited parents' psychological concerns about their child via a validated index test (e.g., the Parental Evaluation of Developmental Status (PEDS) (Glascoe, 1997), or equivalent), or via a custom index test designed to elicit parents' concerns for study use only
3. the study confirmed the presence/absence of a DBP in children via a reference standard that had been validated for use within the population of interest (e.g., Eyberg Child Behavior Inventory (ECBI) (Eyberg & Pincus, 1999), or equivalent)
4. the study reported the 4 outcomes of interest identified by this review or reported outcome measures from which the 4 outcomes of interest could be generated.

We excluded studies that met one or more of the following exclusion criteria:

1. the study population of interest did not match the review's population of interest (i.e., sampled only children 5 years and older, or did not report data for the population of interest that was separable from older populations)
2. the study reported the wrong outcomes (i.e., reported the wrong outcomes of interest, or did not report outcomes from which the 4 outcomes of this review could be generated)
3. the study measured heterogeneous mental health problems (i.e., measured a combination of problems from different developmental domains (e.g., DBPs and physical health problems)), or did not measure DBPs at all
4. the study elicited heterogeneous concerns from parents (e.g., concerns about psychological and physical development in their child), or elicited concerns from parents that were not psychological in nature (i.e., not about social, emotional, or behavioral development)

5. the study used a pseudo-retrospective or pseudo-prospective sampling method (Kraemer, 1992) (see below)
6. the study was not a primary analysis of data (e.g., re-analysed data from a previous study)

There are two primary arguments justifying these eligibility/exclusion criteria. These can be categorized as 1) Sampling Methods, and 2) Defining Parents' Concerns.

2.2.1 Sampling Methods

Among diagnostic test accuracy (DTA) studies that have assessed the accuracy of parents' concerns with respect to DBPs, there are three sampling methods by which participant data can be generated: naturalistic, prospective, and retrospective sampling. A description of each of these three sampling methods, according to the methods of Kraemer (1992), are provided below.

Naturalistic sampling simply involves drawing a representative sample from the population of interest (N0) and testing every participant with both the reference standard and the index test irrespective of the participant's results on either test. From the results of these tests, a 2x2 contingency table can be populated and the 4 outcomes of interest (i.e., TP, FN, FP,TN) can be estimated. These outcomes are then used to estimate the values of P, Q, sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) (see *Data Synthesis and Analysis*). Briefly, P is the proportion of children who have a DBP within the population of interest, and Q is the proportion of parents' who have concerns within the population of interest. SE is the conditional probability of a parent having concerns if their child has a DBP; SP is the conditional probability of a parent *not* having concerns if their child does *not* have a DBP; PPV is the conditional probability of a child having a DBP if

their parent has concerns; NPV is the conditional probability of a child *not* having a DBP if their parent does *not* have concerns.

Prospective DTA studies first draw a representative sample of parents from the population of interest and test all individuals in the sample with the index test (N0). From these index test results, we can estimate the proportion of parents in the population of interest with a positive index test result – that is, those who have concerns about their child's behaviour (Q). A second, random sample of children whose parents gave a positive response on the index test (N1) is then drawn from N0. A third, random sample of children whose parents gave a negative response on the index test (N2) is also drawn from N0. All children from the N1 and N2 samples are then administered the reference standard. From these reference standard results, we can estimate the conditional probability of a child having a positive reference standard result (i.e., DBP) if their parent had a positive index test result (i.e., concerns) – that is, we can estimate PPV. Similarly, we can estimate the conditional probability of a child having a negative reference standard result (i.e., no DBP) if their parent had a negative index test result (i.e., no concerns) – that is, we can estimate NPV. Having estimates of Q, PPV, and NPV, we can then proceed to estimate the 4 outcomes of interest, SE, SP, and P (Kraemer, 1992). However, in the absence of the first sampling stage (N0) (i.e., "pseudo-prospective sampling" (Kraemer, 1992)), we cannot generate a prior estimate of Q. Without Q, we cannot calculate later estimates of the 4 outcomes of interest, SE, SP, and P (Kraemer, 1992), nor can we calibrate these measures (see *Data Analysis and Synthesis*) (Kraemer & Gibbons, 2009). Therefore, pseudo-prospective DTA studies were unsuitable for this study's analysis protocol and were excluded from this review.

Inversely, retrospective DTA studies first draw a representative sample of children from the population of interest and test all individuals in the sample with the reference standard (N0). From these reference standard results, we can estimate the proportion of children in the

population of interest with a positive reference standard result – that is, those who have a DBP (P). A second, random sample of children who received a positive result on the reference standard (N1) is then drawn from N0. A third, random sample of children who received a negative result on the reference standard (N2) is also drawn from N0. All parents of the children from the N1 and N2 samples are then administered the index test. From these index test results, we can then estimate the probability of a parent having a positive index test result (i.e., concerns) if their child has a positive reference standard result (i.e., a DBP) – that is, we can estimate SE. Similarly, we can estimate the conditional probability of a parent having a negative index test result (i.e., no concerns) if their child has a negative reference standard result (i.e., no DBP) – that is, we can estimate SP. Having estimates of P, SE, and SP, we can then proceed to estimate the 4 outcomes of interest, PPV, NPV, and Q. However, in the absence of the first sampling stage (N0) (i.e., "pseudo-retrospective sampling" (Kraemer, 1992)), we cannot generate a prior estimate of P. Without P, we cannot calculate later estimates of the 4 outcomes of interest, NPV, PPV, and Q (Kraemer, 1992), nor can we calibrate these measures (*see Data Analysis and Synthesis*) (Kraemer & Gibbons, 2009). (Kraemer & Gibbons, 2009). Therefore, pseudo-retrospective DTA studies were unsuitable for this study's analysis protocol and were excluded from this review.

2.2.2 Defining Parents' Concerns

Determining which types of parents' concerns to consider can miss the target condition (i.e., DBPs), either by mapping too broadly (i.e., including too many types of concerns) or by mapping too narrowly (i.e., excluding too many types of concerns). There are limitations to either approach. For example, broadening the definition of behavioural concerns too much, such that many concerns that are included *do not* map onto DBPs (but, rather, on to other developmental problems), may artificially reduce our SP estimate. Alternatively, narrowing the definition of behavioural concerns too much, such that many concerns are excluded that *do* map

onto DBPs may artificially reduce our SE estimate. Either situation will introduce bias. In effort to minimize the number of behavioral concerns that may be missed by defining this construct too narrowly, this review defined parents' concerns as concerns about the psychological development of their child. As mentioned, psychological development has been defined as the social, emotional, and/or behavioral development of the child. Therefore, studies that elicited concerns from parents that extended beyond the scope of this construct were excluded from this review.

2.2.3 Deviations from PITO Elements in Review Protocol

There are some key areas in which the final study design deviated from the planned methods identified within the protocol. First, as mentioned above, the population of interest in this review were children 0-5 years of age and their parent/caregiver. This age range is narrower than what had been published as part of our protocol – originally the population of interest included children aged 0-12 years. In pilot testing of our search strategy, the broader age range yielded close to 30000 records in total, which reviewers were concerned would be unmanageable given our available resources. Therefore, upon further refinement of our search strategy, we narrowed the age range limitation to better mirror the clinical context addressed by the Canadian Pediatric Society (which specifically addressed screening for children under 5 years).

Second, the dimensions of DBPs defined by this review's target condition also differed slightly from what was first published in our protocol. As mentioned, the current target condition was defined as the presence or absence of a DBP in a young child. Here, DBPs were operationalized to incorporate 4 dimensions of disruptive behavior (see pg. 1). Key words that reflected these dimensions of behavior were thus used in our search strategy (see Appendix A). By comparison, the search strategy published alongside the protocol included dimensions of attention deficit/hyperactivity disorder (ADHD) among these key words. This was done,

originally, to reflect the inclusion of dimensions of ADHD alongside the behavioral dimensions of DBPs. Following pilot testing of the search strategy, reviewers again were concerned that the volume of initial records would be unmanageable given our available resources. In addition, reviewers felt, given that ADHD is not classified as a DBD in the DSM-5, that the exclusion of dimensions of ADHD from both the operationalization of DBPs and from the search strategy would again better reflect the CPS guidelines. Therefore, the current operationalization of the target condition, and the ways in which the target condition is reflected in the current search strategy, are narrower than those published alongside the protocol in that they exclude dimensions of ADHD.

2.3 Systematic Search Strategy

Eligible studies were identified using a systematic search strategy that was developed with the assistance of a resource librarian (NL) (see Appendix A). The search strategy was adapted for use in Medline(Ovid), Embase(Ovid), Central(Ovid), CINAHL(EBSCOHost), PsycINFO(Ovid), and Eric(Ovid) and Scopus. Search results were first uploaded to Covidence Systematic Review Software (Covidence, Melbourne, VIC, Australia), in which all steps of the eligibility review process were conducted. Following the import of search results, duplicates were removed. Title-and-Abstract Screening was then completed, where potentially relevant studies were selected for full-text review. Titles and abstracts were each screened independently by 2 reviewers, with 6 reviewers in total contributing to title and abstract screening (SW, HP, TC, IM, NY & RB). Consensus on decisions to include or exclude was required between the 2 reviewers assessing each record, and conflicts were resolved via discussion.

Following Title-and-Abstract Screening, full-text articles were independently screened for eligibility by 2 reviewers, with three reviewers in total contributing to this stage of review (SW, RB & HP). Studies were excluded from our analysis if they did not meet our inclusion

criteria (see below). Like the Title and Abstract screening phase, consensus was required between the 2 reviewers on decisions to include or exclude a study. In situations where a reviewer wished to vote to exclude a record, the reviewer had to select one primary reason for ineligibility in Covidence. Any disagreement between reviewers was resolved via discussion, both for disagreements about whether to include or exclude a study, and for disagreements about the primary reason for exclusion.

Following Full-Text Review, backward citation searching was completed by manually checking the reference lists of the included studies (see Briscoe, Bethel & Rogers, 2018). One reviewer (SW) imported all reference material from included studies into Covidence for completion of screening. Title and Abstract screening, and full-text screening during this phase were completed independently by two reviewers (SW & RB). Conflicts were resolved via discussion. If a study(s) was missing data required for analysis, the review team contacted the study author(s) via email to request the missing data during this review phase.

2.4 Data Extraction

It is important to note here that the specific data extraction and analysis methods used by review depended on the sampling methods used by each included study. That is, for example, the extraction and analysis of data would have been slightly different for a study with a prospective sampling design than for one with a retrospective sampling design. Similarly, the exact methods required to analyze data from a study with a naturalistic sampling design are also unique. Given the results of this review, all extraction and analysis methods described hereon in are specific to data from studies that used a naturalistic sampling design.

A data extraction form specific to naturalistic sampling methods was adapted from the "Data Collection form for intervention review – RCTs and non-RCTs" data collection template produced by the Cochrane Collaboration (2020). An example of the adapted data extraction form for a naturalistic sampling design can be found in Appendix B. Data pertaining to the

outcomes of interest (i.e., TP, TN, FP, and FN), participant demographics (i.e., age, sex, gender, … etc.), publication details (i.e., authors, date of publication, … etc.), and study characteristics (i.e., number of participants, reference standard used, index test used etc.) were independently extracted from included studies and entered into the data extraction form.

2.5 Assessment of Risk of Bias

As per the PRISMA-DTA guidelines, risk of bias was assessed in included studies based on their applicability to the current problem of interest (McInnes et al., 2018). Applicability of studies to the current review and other sources of methodological bias were assessed using the QUADAS-2 tool (Whiting, Rutjes, Westwood, et al., 2011[a]). According to the QUADAS-2, the applicability of a DTA study is defined as the "the extent to which primary studies are applicable to the review's research questions" and bias is defined as "systematic flaws or limitations in the design or conduct of a study [which] distort the results" (Whiting et al., 2011[a], p.529).

For each included study, two reviewers (SW & RB) independently completed the QUADAS-2 assessment to determine the risk of bias within each of the following 4 domains: patient selection methods, the index test, the reference standard, and the Flow and Timing of test administration. Reviewers also used the QUADAS-2 tool to determine whether each included study was applicable to our review questions. This was assessed across each of the following 3 domains: the patient selection methods, the index test, and the reference standard. Following independent assessment, any discrepancies in reviewers' judgments for each of the domains were resolved via discussion. The QUADAS-2 poses signalling questions for each risk of bias domain, and each applicability domain (Whiting et al., 2011[b]) to assist reviewers in their assessment. Reviewers judged the risk of bias as "low", "unclear", or "high" (Whiting et al., 2011[b]) based on their responses to each signalling question. As per the QUADAS-2 guidance document, a response of "no" to any signalling question indicated the potential for

bias; a response of "unclear" was appropriate when the included study reported insufficient data to permit a judgment from reviewers; a response of "yes" to any signalling question indicated a low risk of bias (Whiting et al., 2011[b]).

2.6 Data Analysis & Synthesis

2.6.1 Naturalistic Sampling Designs

To determine whether the mere presence of parents' concerns justifies deciding in favour of screening for DBPs in children, reviewers needed to determine how accurately parents' concerns could rule in the presence of a DBP in children. To do this, reviewers calculated the concerns' positive predictive value (PPV), specificity (SP), and jackknife estimates of RR1 based on the four outcomes of interest extracted from each included study. RR1 was defined as the ratio of the odds of having a DBP among those children who have parents with concerns, over the odds of having a DBP among all children (Kraemer, 1992). Standard errors, and 95% confidence intervals were obtained for all measures listed above. The formulae used to calculate SP, PPV and RR1 for a naturalistic sampling design can be found in Figure 1.

For each included study, these estimates were then calibrated on a common scale (ranging from 0 to 1) using a weighted kappa coefficient, k(0,0) (Kraemer, 1992). The formulae used to calibrate SP, PPV and RR1 can also be found in Figure 1. Note that, once calibrated, all three estimates yielded the same jackknife estimate of k(0,0) (Kraemer, 1992). One may interpret a weighted k(0,0) value as a quality index of the reproducibility of a positive diagnosis, where k(0,0) equal to 0.0 indicates that any agreement between a positive result on the reference standard and a positive result on the index test is due to chance alone (Kraemer, 1992, p. 17). Conversely, k(0,0) equal to 1.0 indicates 100% corrected-for-chance agreement between a positive result on the reference standard and on the index test. As defined by Landis & Koch

Figure 1.

Necessary Formulae for Data Analysis for a Naturalistic Sampling Design

$$SP = \frac{TN}{P\prime} \qquad SE = \frac{TP}{P}$$

$$PPV = \frac{TP}{Q} \qquad NPV = \frac{TN}{Q\prime}$$

$$RR1 = \frac{SE}{(1-SP)} \qquad NPV = \frac{SP}{(1-SE)}$$

$$k(0,0) = \frac{(SP-Q')}{Q} \qquad k(0,0) = \frac{(PPV-P)}{P\prime} \qquad k(0,0) = \frac{(RR1-1)}{(RR1+({}^{P\prime}/_{P}))}$$

$$k(1,0) = \frac{(SE-Q)}{Q\prime} \qquad k(1,0) = \frac{(NPV-P\prime)}{P} \qquad k(1,0) = \frac{(RR3-1)}{(RR3+({}^{P}/_{P\prime}))}$$

Note. P refers to the proportion of children with a DBP (*P=TP+FN*, and *P'* = (1 - P)). *Q* refers to the proportion of children of concerned parents (*Q=TP+FP*, and *Q'* = (1 - Q)). *TP* refers to the proportion of children who have a DBP and have parents with concerns; *FP* refers to the proportion of children without a DBP and have parents with concerns; *FN* refers to the proportion of children with a DBP and have parents without concerns; *TN* refers to the proportion of children without a DBP and have parents without concerns.

(1977), the relative strength of agreement represented by any weighted kappa statistic ranges from poor ($k \leq 0.00$) to almost perfect ($0.8 < k \leq 1.00$) (see Table 1). Values of k(0,0) must be at least substantial (i.e., $k > 0.6$) to demonstrate that parents' concerns can provide a reliable means to rule *in* the presence of a DBP in the absence of additional information (Landis & Koch, 1977).

To determine if the mere absence of parents' concerns justifies deciding against screening for DBPs in children, reviewers needed to determine how accurately parents' concerns could rule out the presence of a DBP in children. This was done by calculating the concerns' negative predictive value (NPV), sensitivity (SE), and jackknife estimates of RR3 based on the four outcomes of interest extracted from each included study. RR3 was defined as the ratio of the odds of not having a DBP among children who have parents with no concerns,

Table 1

Strength of Agreement Represented by Weighted Kappa Statistics

Kappa statistic	Strength of Agreement
$k \leq 0.00$	Poor
$0.00 < k \leq 0.20$	Slight
$0.20 < k \leq 0.40$	Fair
$0.40 < k \leq 0.60$	Moderate
$0.60 < k \leq 0.80$	Substantial
$0.80 < k \leq 1.00$	Almost Perfect

Note. This table is adapted from the thresholds defined and published by Landis & Koch (1977, p.165)

over the odds of not having a DBP among all children (Kramer, 1992). Standard errors and 95% confidence intervals were also obtained for all measures listed above. The formulae used to calculate SE, NPV and RR3 for a naturalistic sampling design can also be found in Figure 1.

These estimates were also calibrated on a common scale (ranging from 0 to 1) using a weighted kappa coefficient, k(1,0) (Kraemer, 1992). Formulae for calibrating SE, NPV, and RR3 can be found in Figure 1. Note that, once calibrated, all three estimates yielded the same jackknife estimate of k(1,0) (Kramer, 1992). One may interpret a weighted k(1,0) value as a quality index of the reproducibility of a negative diagnosis, where values of k(1,0) equal to 0.0 indicate that any agreement between a negative result on the reference standard and a negative result on the index test is due to chance alone (Kraemer, 1992, p. 17). Conversely, values of k(1,0) equal to 1.0 indicate 100% corrected-for-chance agreement between a negative result on the reference standard and on the index test. Values of k(1,0) exceeding 0.6 indicate that parents' concerns can provide a reliable means to rule *out* the presence of a DBP (Landis & Koch, 1977).

2.7 Meta- and Subgroup Analyses

In the review protocol, a meta-analysis plan was proposed which used a bivariate, multilevel model of k(0,0) and k(1,0) estimates to construct a summary Receiver Operating Characteristic (sROC) curve (see Reitsma, Glas & Rutjes et al., 2005). The area under the sROC curve would have then been calculated to determine the overall accuracy of parents' concerns. However, one scenario in which the proposed bivariate model may fail to converge arises when too little data is available for meta-analysis (i.e., an insufficient number of studies is returned from our search that meet our eligibility criteria). There is little consensus on what constitutes "too little" data; however, by convention, attempting to use data from fewer than 4 studies will likely render the model unstable. Therefore, reviewers defined, *a priori*, that a minimum threshold of 4 included studies was required to complete a meta-analysis. As is discussed in Chapter 3, the final systematic search did not yield a sufficiently large dataset from which summary estimates of k(1,0) and k(0,0) could be modeled. Therefore, the meta-analysis method published in the protocol was not used to meta-analyze data from the included study.

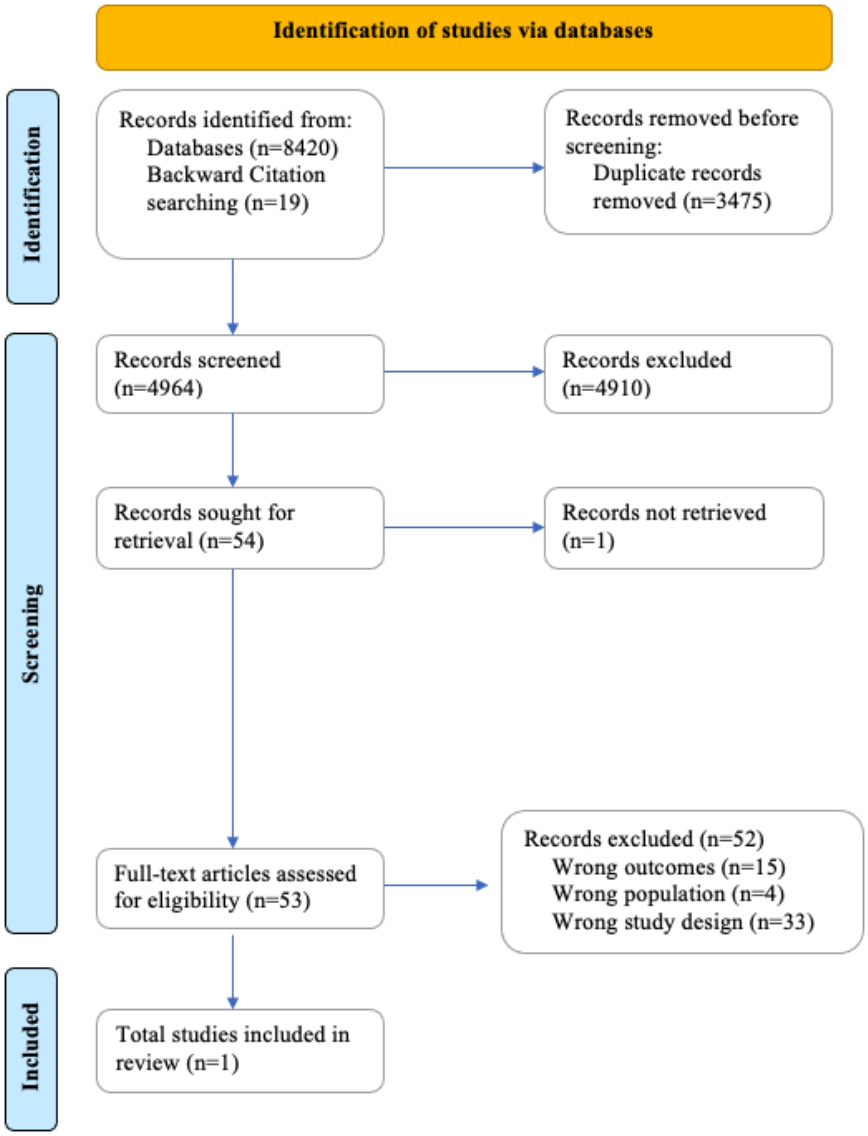
CHAPTER 3 – DATA ANALYSIS AND RESULTS

3.1 Study Selection

Figure 2 was adapted from the PRISMA flow diagram published by Page et al. (2021) and summarizes the results of the screening and review process. Systematic searches returned 8420 articles, and backward citation searching returned an additional 19 articles. 3475 articles were removed as duplicates. Titles and abstracts from the remaining 4964 records were each screened independently by 2 reviewers, with 6 reviewers in total contributing to this screening phase (SW, HP, TC, IM, NY & RB). 4910 records were considered irrelevant, and thus excluded at this stage. Of the remaining 54 articles, 1 full-text version (Melo & Rodriguez, 1980) was unable to be retrieved. 53 full-text articles were independently screened for eligibility by 2 reviewers, with 3 reviewers in total contributing to this screening phase (SW,

Figure 2.

Prisma Flow Diagram



Note. This diagram was adapted from the *Prisma Flow Diagram* published by Page et al., 2021.

HP & RB). 52 out of 53 full-text records were ineligible for inclusion for a variety of reasons. Just one article (Glascoe et al., 1991) met all inclusion criteria. Very few studies failed to meet inclusion criteria for one reason alone. However, constraints within the Covidence software required reviewers to select only one, primary reason for exclusion on which reviewers could reach a consensus (see Figure 2).

Table 2 summarizes the decisions made via this procedure and gives, for each of the 52 excluded records, a description of how the selected reason to exclude a study given in

Covidence aligned with our exclusion criteria. If a study was missing data that were required for analysis, SW contacted the study author(s) to request the missing data. One such study was retrieved via our search (Reijneveld et al., 2008). In this study, sample data published for participants under the age of 5 could not be analyzed separately from sample data corresponding to participants over the age of 5. As such, we could not generate necessary outcome measurements (i.e., TPs, FNs, TNs, and FPs) for our population of interest using only the data published by Reijneveld et al. (2008). A request for the disaggregated data was sent to the study's authors. Unfortunately, no response was received by the time of submission. Therefore, the paper by Reijneveld et al. (2008) was excluded from the analysis.

Three other studies were identified in Covidence as having samples did not match our population of interest and were therefore excluded from this review. First, the study by August et al. included "children [that] ranged in age from 7-12 years" (1998, p. 346), and thus contained a sample that was older than our target population. Similarly, the study by Poduska sampled from "all first-graders in 9 public elementary schools … [that were] participants in 2-school based interventions targeting early learning and aggressive behavior" (2000, p. 585). However, the study did not report the mean age or age range of children included in their sample. In North America, most children in first-grade are between 6 and 7 years of age. Hence, the sample studied by Poduska (2000) appeared to be older than our population of interest and the study was therefore excluded from this review. Lastly, the study by Feldman et al. (2000) was erroneously categorized in Covidence as containing the wrong population of interest. However, the children included in the study by Feldman et al. (2000) were all 2 years of age, and thus were within this review's population of interest. Nonetheless, all children included in this sample had been diagnosed with, or were at risk for, developmental delays (Feldman et al., 2000). Therefore, this study did not match this review's target condition (DBPs), and therefore

Table 2

Description of Primary Reasons for Exclusion in Covidence

Exclusion criterion met based on reason-for-exclusion given in Covidence	Description of primary reason-for-exclusion in Covidence	Excluded studies
Wrong Outcomes (*n*=15)	Study did not report proportions of TPs, TNs, FPs, or FNs as outcome, or did not report similar outcomes that would allow us to generate estimates of proportions of TPs, TNs, FPs, and FNs (*n*=15)	August et al., 1995; Berger-Jenkins et al., 2019; Bengi-Arslan et al., 1997; Bergold et al., 2019; Briggs-Gowan, 1996; Briggs-Gowan et al., 2013; Campbell et al., 1984; Holtz, 2010; Lavigne et al., 2019; Nærde et al., 2014; Plamondon et al., 2018; Sabol et al., 2022; Wakschlag et al., 2012; Wakschlag et al., 2014; Winsper & Wolke, 2014
Wrong Population of Interest (*n*=4)	Results from children in the target population (i.e., under the age of 5 years) were not available or could not be analyzed (*n*=4)	August et al., 1998; Feldman et al., 2000; Poduska, 2000; Reijneveld et al., 2008
Wrong Study Design (*n*=33)	Incomplete sampling methods (*n*=7)	Fanton et al., 2008; Glascoe, 1998; Gray et al., 2012; Herman-Staab, 1994; Ireton, 1996; Ogden & Hagen, 2008; Szczepaniak et al., 2013
	Re-analysis of existing data (*n*=2)	Glascoe, 1999[a]; Glascoe, 2003
	Elicited heterogeneous concerns from parents (*n*=4)	Barkley et al., 2002; Karabekiroglu et al., 2013; Kruizinga et al., 2012; Wendland et al., 2014
	Measured heterogeneous developmental problems (*n*=6)	Alakortes et al., 2017; Briggs-Gowan & Carter, 2008; Chen et al., 2004; Ellingson et al., 2004; Godoy et al., 2014; Sheldrick et al., 2012
	Absent or inappropriate index test (*n*=7)	Barkauskiene et al., 2009; Feeney-Kettler, 2009; Lorber et al., 2014; Sand, 1972; Sawyer et al., 1996; Thomas et al., 1991; Ware et al., 2001
	Absent or inappropriate reference standard (*n*=7)	Alink et al., 2006; Edelstein et al., 2022; Glascoe et al., 1989; Glascoe, 1999[b]; Ilić et al., 2020; Pavaluri et al., 1995; Studts et al., 2017

the study was excluded on the grounds that the developmental problems measured were too heterogeneous.

3.2 Heterogeneity in Parents' Concerns

As previously mentioned, a relatively broad definition of parents' concerns was employed by this review, where studies that elicited parents' psychological concerns (i.e., social, emotional, or behavioral concerns) about their young children were eligible. Parents'

concerns were considered *too* heterogeneous if they addressed psychological development and other domains of development in their child simultaneously. For example, a study that elicited concerns from parents that addressed physical or cognitive development were not eligible for inclusion. Four studies were identified during full text-screening that elicited parent concerns which were considered too heterogeneous for the scope of this review. Summaries of how parents' concerns were defined and measured by each of these 4 studies, and why each study was ultimately excluded from this review are presented below.

3.2.1 Barkley, Shelton, Crosswait, et al. (2002)

In the study by Barkley et al. (2002), there were no measurement tools used that could be considered index tests that elicited parents' concerns. The only measure that came close to an index test (as defined by this review) was the Normative Adaptive Behavior Checklist (NABC). However, this tool did not appear to elicit specific concerns from parents about their child's psychological development. Further, the NABC assesses global developmental functioning across six very different domains: fine and gross-motor skill development, sensory-motor skills, language, self-help skills, independent living skills, and social skills (Barkley et al., 2002, p. 51). Therefore, reviewers considered the data gathered by Barkley et al. (2002) via the NABC to be too heterogeneous for this review.

3.2.2 Karabekiroglu, Uslu, Kapci-Seyitoglu, et al. (2013)

In the study by Karabekiroglu et al. (2013), there were again no measurement tools used that could be considered index tests that elicited parents' concerns. The only measure that appeared close to an index test was a series of questionnaire items that asked parents to "state the mental and/or developmental problems that [parents] thought to exist in [their] child" (2013, Table 2, p. 165). Importantly, the questionnaire items did not appear to elicit specific psychological concerns or worries from parents. Further, items on this questionnaire appeared to address areas of global developmental functioning (i.e., speech and language,

learning/comprehension, sleeping, eating/feeding behavior) and areas of social-emotional development (i.e., irritability, phobias/fearfulness, aggression, hyperactivity, attention problems, excessive crying, and social withdrawal) (Karabekiroglu et al., 2013, Table 2). Therefore, the "index test" used by Karabekiroglu et al. (2013) was considered too heterogeneous for the scope of this review.

3.2.3 Kruizinga, Jansen, de Haan, et al. (2012)

In the study by Kruizinga et al. (2012), there were again no measurement tools used that could be considered index tests that elicited parents' concerns. The only measure that came close to an index test (as defined by this review) was an item which assessed whether parents "worried about their child's upbringing" (Kruizinga et al., 2012, p. 5). It was unclear to reviewers whether this measure was contained within a validated screening tool, or if it was a standalone item. Nonetheless, reviewers felt that "upbringing" as a construct was too broad and did not fit this review's definition of parents' concerns.

3.2.4 Wendland, Danet, Gacoin, et al. (2014)

The study by Wendland et al. sought to examine the psychometric properties of a French translation of the BITSEA, which is a screening tool for the identification of elevated levels of social-emotional behavior problems or delayed levels of competence in children aged 1-3 years (2014). The BITSEA contains 2 single-item questions that assess parents' worries, in addition to subscales that measure behavioral problems. Pertinent to the scope of this review, the BITSEA-A item asks parents to report on the level of worry they have about their young child's behavior, emotions, or relationships (Wendland et al., 2014). In their analysis, Wendland et al. calculated correlations between BITSEA subscale scores and the level of parental worry as reported by parent responses to the BITSEA-A item (2014). Unfortunately, this study was erroneously categorized during the full-text screening phase as having reported heterogeneous concerns. Given that the BITSEA-A screening item does elicit specific concerns from parents,

the study did in fact meet this inclusion criterion. Nonetheless, the fact that only correlational data was reported by Wendland et al. (2014) (from which we could not generate estimates of proportions of TPs, TNs, FPs, and FNs) rendered this study ineligible for inclusion.

3.3 Included Study Characteristics: Glascoe, MacLean & Stone, 1991

3.3.1 Demographic and Study Characteristics

Demographic and study characteristic data for the included study by Glascoe et al. (1991) was extracted by reviewers (SW & RB). This cross-sectional study enrolled 99 parent-child dyads from five sampling sites, where patients who had visited one of the five sampling sites to seek non-acute medical care were eligible to participate in the study (Glascoe et al., 1991). Sampling sites consisted of "two urban teaching hospitals" and "three private pediatric practices" (Glascoe et al., 1991, p. 8). Four eligible parent/child dyads declined participation, leading to a sample size of n=95. Children enrolled were between 24-78 months of age, with a mean age of 48 months (Glascoe et al., 1991). Statements that disclosed the study's funding were not included in the publication by Glascoe et al. (1991).

3.3.2 Eliciting Parents' Concerns: The Index Test

First, an index test was administered verbally to parents (Glascoe et al., 1991). The first question on the index test "asked each parent 'Please tell me any concerns about your child's learning and development?'" (1991, p. 9). This generated positive responses specific to behavioral concerns from 8 out of 95 parents. All parents were then asked, "Do you have any concerns about the way [your child] behaves" (Glascoe et al., 1991, p.9). This second question generated 26 positive responses (Glascoe et al., 1991). Results from these 2 index test questions were pooled to yield a total of 34 parents that had indicated that they had concerns about their child's behavior. Note that these and other questions were later published in a formal, validated tool intended to elicit information about parents' concerns – the PEDS (Glascoe, 1997).

The PEDS itself has demonstrated moderate internal consistency (Cronbach's alpha *a* = 0.692), with a tendency toward higher internal consistency as the age of participants increases (Glascoe, 2013[a], Table 3-6, p. 111). The concurrent validity of the PEDS has been well studied, and associations between parent concerns on the PEDS and diagnostic measures of development are significant (Glascoe, 2013[b]). However, recent studies of the predictive validity of the tool are limited. Only reports of strong associations between parents' concerns and later observations of academic failure and diagnoses of autism spectrum disorder in children are reported by the tool's author (Glascoe, 2013[b]).

3.3.3 Measuring Disruptive Behavior Problems: The Reference Standard

Following administration of the index test questions, the presence/absence of DBPs were confirmed using the Eyberg Child Behavior Inventory (ECBI) as a reference standard. The ECBI is a parent-report questionnaire comprised of 2 subscales: the problem subscale, and the intensity subscale. Relevant to this review, the problem scale asks parents to identify which behavior problems have been observed (if any) in their child, out of a total of 36 behaviors problems common among children with conduct problems. The ECBI was standardized on children between 2-12 years of age (Robinson, Eyberg & Ross, 1980), which is consistent with this review's population of interest. The instrument has demonstrated good reliability, with an internal consistency coefficient (Cronbach's alpha) of r=0.98 for the problem subscale (Robinson et al., 1980). Also, the ECBI has demonstrated good predictive validity, where children with conduct problems tend to have problem subscale scores (mean=15.0) that are significantly higher than children without conduct problems (mean=5.6, $p < 0.01$) (Robinson et al., 1980). Glascoe et al. (1991) defined as the cut-off for a failed test (i.e., indicating the presence of DBPs) as the presence of 16 or more behavior problems. A total of 20 children received a score of 16 or higher (Glascoe et al., 1991). Unfortunately, due to the methods by which the results of the index test questions were pooled together, it is unknown which of the

children with positive ECBI scores received a positive score on the first index question, and which of these children received a positive score for the second index test question.

3.4 Risk of Bias in Included Studies: Glascoe, MacLean & Stone, 1991

Table 3 summarizes the results of the risk of bias assessment completed for the study by Glascoe et al. (1991) via the QUADAS-2 assessment tool. Two reviewers (SW & RB) independently completed the QUADAS-2 assessment for the included study. Appendix C contains a copy of the QUADAS-2 assessment form, including all signalling questions posed to reviewers for both risk of bias concerns and applicability concerns across all domains.

3.4.1 Domain 1: Patient Selection

Risk of Bias in Patient Recruitment and Enrolment

The QUADAS-2 tool assessed whether the selection of patients by Glascoe et al. (1991) could have introduced bias. Three signalling questions identified key methodological areas where risk of bias could have been introduced.

The first signalling question asked whether the study used "consecutive" or random sampling as a strategy to minimize the risk of sampling bias (Whiting, Rutjes, Westwood, et al., 2011[b], p. 4). The study by Glascoe et al. used convenience sampling to enroll participants, where "parents of children seeking health services who were between 24 and 78 months of age and who were not acutely ill were asked to participate" (1991, p. 8). As mentioned, all

Table 3

Revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2): Results for the study by Glascoe, MacLean & Stone, 1991

Risk of bias				Applicability concerns		
Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
☺	?	☹	☺	☺	☺	☺

☺ Low risk ☹ High Risk ? Unclear Risk

Note. Table adapted from the QUADAS-2 results template published by Whiting et al. (2011[a])

participants were parent-child dyads who had been recruited while seeking pediatric care from one of the five recruiting sites (Glascoe et al., 1991). While convenience sampling carries some risk of sampling bias, reviewers did not find evidence in the text to suggest that participants were enrolled during the recruiting window in a manner that would impart additional risk of bias (e.g., cherry picking). Therefore, reviewers considered the risk of bias in this area to be low.

Second, the QUADAS-2 asked reviewers to determine whether a case-controlled design had been avoided by the study in question. The study by Glascoe et al. (1991) used a naturalistic sampling design, where parents' concerns and children's behavior problems were assessed in tandem. The risk of bias pertaining to this question was therefore judged to be low.

Lastly, reviewers were asked to determine whether any unjustifiable exclusion criteria were implemented that could have introduced sampling bias. The study by Glascoe et al. (1991) had two explicit eligibility criteria. First, as mentioned, children were not eligible for inclusion if they were seeking acute medical care, and second, children were not eligible if they were younger than 24 months or older than 78 months of age (Glascoe et al., 1991). These exclusion criteria were considered justifiable in the eyes of reviewers, given that the context of this review and of the practice guidelines in question (Charach et al., 2017) are also specific to well-child primary care visits; that is, routine, non-acute healthcare visits for young children (i.e., between 2-5 years of age).

One additional area not addressed by the QUADAS-2 tool may also contribute to the risk of bias in this study. 4 out of 99 recruited parent-child -dyads declined to participate in the study by Glascoe et al. (1991). Further, no data was reported by Glascoe et al. (1991) about the demographic or clinical features of these parent-child dyads to indicate that they differed significantly (or not) from those patients included in the study. Given the lack of transparency about potential differences between those individuals who participated and those who did not,

reviewers determined that a risk of nonresponse bias (Goodwin, 2008) could not be ruled out. However, given that the return rate for this study was relatively high (i.e., responses gathered from 95/99 from eligible participants), reviewers considered this risk to be low. Overall, risk of bias within the patient selection domain was judged to be low (see Table 3).

Applicability of the Patient Sample

The QUADAS-2 included a signalling question about whether the study included patients that did not match our review question. As mentioned, patients who participated in the study by Glascoe et al. (1991) were sampled from pediatric care clinics and consisted of only those patients who were seeking non-acute care. Reviewers found that this clinical context was sufficiently similar to the clinical context detailed in the practice guidelines by Charach et al. (2017), which informed our review questions. Therefore, reviewers had low concern about the overall applicability of the participants sampled by Glascoe et al. (1991) (see Table 3).

3.4.2 Domain 2: Index Test

Risk of Bias Within the Index Test

The QUADAS-2 tool also assessed whether the index test, how it was conducted, and how it was interpreted by Glascoe et al. (1991) could have introduced bias. Two signalling questions assessed this. Reviewers were first asked to evaluate if the results of the index test used by Glascoe et al. (1991) had been interpreted without knowledge of the results of the reference standard. The index test questions were administered before the reference standard (the ECBI) (Glascoe et al., 1991). Also, even if the results from the index test questions had been scored after the ECBI was administered, reviewers believe that the format of the index questions were consistent with those contained in the PEDS – that is, responses from parents are given by circling "yes" or "no", or "a little". In this format, parents' concerns can be assessed with relative objectivity, requiring little interpretation on behalf of researchers scoring

the results. Therefore, reviewers considered the risk of bias pertaining to this question to be low.

Second, reviewers were asked to determine if a test threshold was used for the index test, and (if so) if the test's threshold had been pre-specified. As mentioned above, the index test used by Glascoe et al. (1991) did not calculate or sum a total score for which a test threshold was necessary. Rather, as mentioned above, the total number of parents who reported concerns about their child's behavior was calculated by summing the number of participants who reported concerns about behavior in response to either the first (n=8) or second (n=26) index test question (Glascoe et al., 1991). Therefore, the risk of bias pertaining to this signalling question was found to be low.

Outside of the signalling questions asked by the QUADAS-2, however, reviewers identified additional areas of concern where the conduct of the index test may have introduced bias. If we assume that Glascoe et al. (1991) used a version of the PEDS that is consistent with the published version of the tool as their index test there is some ambiguity in how Glascoe et al. (1991) dichotomized parents' responses to the index test questions to arrive at that number of concerns (i.e., 34). The published version of the PEDS explicitly codes parents' responses to these same index test questions in one of three categories: "no", the parent is not concerned; the parent is "a little" concerned; or "yes" the parent is concerned (Glascoe, 1997). If Glascoe et al. (1991) used a similar scale to record index test responses, they may have combined responses of "yes" or "a little" to either index test question into one "yes" category for a dichotomous response variable. Any such strategy to determine what constituted a "yes" response on these index test questions may have influenced the total number of concerns measured, and thus poses a risk of bias. Therefore, the overall risk of bias in the index test domain was found to be unclear (see Table 3).

Applicability of the Index Test

The QUADAS-2 tool also included a signalling question that assessed whether the index test, its conduct, or its interpretation differed from our review questions. As mentioned previously, the second index test question posed by Glascoe et al. (1991) specifically elicited parents' concerns about their child's behavior. As for the first index test question, only parents' responses that specifically referred to behavioral concerns were included in the analysis by Glascoe et al. (1991). Reviewers therefore had low concern about the applicability of the index test used by Glascoe et al. (1991).

3.4.3 Domain 3: Reference Standard

Risk of Bias Within the Reference Standard

The QUADAS-2 tool also asked reviewers to determine if the reference standard used by Glascoe et al. (1991), its conduct, or its interpretation could have introduced bias. This was assessed via 2 signalling questions.

The first signalling question asked whether the reference standard was likely to classify the target condition correctly. According to the scoring guide that accompanies the QUADAS-2, "estimates of test accuracy are based on the assumption that the reference standard is 100% sensitive" (Whiting, Rutjes, Westwood, et al., 2011[b], p. 6). To the knowledge of reviewers, there is no instrument that is 100% accurate in screening for behavior problems in young children. As such, any bias that results from the reference standard chosen by Glascoe et al. (1991) should be assessed based on the reported validity and reliability of the tool itself.

Glascoe et al. (1991) used the Eyberg Child Behavior Inventory (ECBI) as a reference standard. Thus, reviewers felt that the reference standard was likely to classify the target condition (behavior problems) correctly. Therefore, the risk of bias pertaining to the first signalling question was judged to be low.

The second signaling question asked reviewers to determine whether the reference standard results were interpreted without knowledge of the results of the index test. According to Glascoe et al., "after parents stated their concerns, they were asked to complete the Eyberg Child Behavior Inventory (ECBI)" (1991, p. 9). The ECBI is designed such that the scoring of parent responses is intended to be objective, where scoring consists of summing the number of "yes" responses given by parents (for the problem scale) (Eyberg et al., 1978). Therefore, there is little interpretation to be done on behalf of a researcher who scores the reference standard. Therefore, reviewers considered there to be a low risk of bias attributable to the interpretation of the reference standard results.

However, many additional aspects of the methods by which the ECBI was administered and scored remain unclear, and therefore represent potential sources of bias. First, it is unknown whether the same researcher administered both the index test and the ECBI to parents. Reviewers considered this to be the probable scenario, given that the time interval between tests was implied to be short. Also, it is unclear whether researchers assisted parents when they completed the ECBI. Hence, the effects of observer-expectancy bias on parents' responses to the reference standard cannot be ruled out.

Second, Glascoe et al. (1991) defined a cutoff threshold for the ECBI of 16, where 16 or more behavior problems was considered a positive reference standard result. This cutoff, however, is inconsistent with the cutoff threshold defined by the authors of the ECBI, who have identified a "tentative cutoff point of 11" for the behavior problem scale (Eyberg & Ross, 1978, p. 115). Therefore, reviewers had concerns that the cutoff threshold for behavior problems that was used by Glascoe et al. (1991) may have biased (downward) the number of children identified as having behavior problems via the reference standard.

Lastly, the order of test administration could have introduced a response bias; that is, parents' previous responses to the index test may influence their later responses to the ECBI

via a variety of mechanisms. The anchoring effect describes the tendency of an established reference point to influence an individual's decisions. For example, if a parent previously expressed that they did not have concerns about their child's behavior (i.e., via the index test), he or she may have been less likely to report the presence of specific behavior problems. Inversely, had the parents completed the ECBI first, a different type of response bias might have been introduced. Overall, reviewers considered there to be a high risk of bias within the reference standard domain (see Table 3).

Applicability of the Reference Standard

The QUADAS-2 tool asked reviewers to assess whether the target condition (i.e., DBPs), as defined by the reference standard used by Glascoe et al. (1991) (i.e., the ECBI), matched the target condition defined in our review question. While the problems assessed by the ECBI may refer to behaviors common to many DBDs (i.e., ODD, Conduct Disorder, ADHD etc.), the authors of the ECBI only state that "the ECBI is an internally valid scale which measures a unitary construct, 'conduct problem'" (Robinson et al., 1980, p. 26). Whether this tool possesses similar construct validity for our target condition (DBPs) is unknown. Reviewers considered the similarity between these two constructs to be sufficient such that there were low concerns about its applicability to the review questions.

3.4.4 Domain 4: Risk of Bias within Study Flow and Timing

Lastly, the QUADAS-2 asked reviewers to assess whether "patient flow could have introduced bias" (Whiting et al., 2011[b], p.6). The first signalling question asked reviewers if there was an appropriate time interval between the administration of the index test and of the reference standard. Reviewers were also asked to determine if any interventions were delivered between the administration of the index test and the reference standard. Glascoe et al. simply state that "after parents stated their concerns [via the index test], they completed the ECBI" (1991, p. 9), implying that both tests were administered within a short time interval and that no

intervention occurred in between the two tests. Assuming this to be true, reviewers considered there to be a low risk of bias pertaining to this question.

In the second and third signaling questions, the QUADAS-2 asked reviewers to determine if all participants were assessed with a reference standard, and if all participants were assessed using the same reference standard. Only one reference standard was used (i.e., the ECBI) by Glascoe et al. (1991). Furthermore, data from all 95 individuals to whom the ECBI and the index test were administered were reported by Glascoe et al. (1991, Table 2, p. 10). Therefore, there were no missing data for either the reference standard or the index test. Reviewers therefore considered there to be a low risk of bias pertaining to these signaling questions.

The last signaling question asked reviewers to determine whether all data from patients enrolled in the study by Glascoe et al. (1991) were included in their analysis. As mentioned, there was no evidence of missing data in the study by Glascoe et al. (1991). Therefore, reviewers considered there to be a low risk of bias pertaining to this question. Overall, reviewers found there to be a low risk of bias in the study flow and timing domain (see Table 3).

3.5 Results of Individual Studies: Glascoe, MacLean & Stone, 1991

Table 4 presents the frequencies of true positive, false negative, true negative, and false negative outcomes extracted from the study by Glascoe et al. (1991, Table 2, p.10). Table 5 presents the proportions of true positive, false negative, true negative, and false positive outcomes, their standard errors, and 95% confidence intervals. There is an asymmetry in the number of parents who have concerns and the number of children who have disruptive behavior problems, with more of the former than the latter. A notably larger proportion of parents expressed concerns about their child's behavior (0.358) than there were children who had behavioral problems (0.211). Indeed, the odds that a parent had concerns were 2.091

Table 4

Frequencies of true positive, false positive, false negative, and true negative outcomes for the study by Glascoe, MacLean & Stone (1991)

		Behavioral Concerns		Marginal Frequency
		Yes	No	
Disruptive Behavior Problems	Yes	14	6	20
	No	20	55	75
Marginal Frequency		34	61	95

[i.e., (.358 / .642) / (.211 / .789)] times higher than the odds that a child had a DBP (Jackknife se=0.058, Jackknife 95% CI: 1.977-2.205). Given this asymmetry, it is not wholly surprising that there is a relatively large proportion of false positive outcomes present in the sample (0.211). Also, there is a statistically significant proportion of false negative outcomes (0.063). Overall, parents made errors distinguishing between children with a DBP and children without in approximately 27.4% of cases.

3.6 Results of Data Synthesis & Analysis

Table 6 presents measures that indicate the accuracy with which parents' concerns can rule in the presence of a DBP in young children. An estimated PPV of 0.412 indicates that, among all parents who expressed concerns, 41.2% had a child with a DBP. An estimated SP of 0.733 indicates that, among all children without a DBP, 73.3% of their parents did not express behavioral concerns. RR1 was estimated to be 2.626, indicating that the odds that a child has a DBP are 2.6 times higher among parents who report concerns (i.e., 14/20 = 0.700), compared to the same odds irrespective of parents expressing concerns or not (i.e., 20/75 = 0.267). Following Kraemer (1992), these measures were calibrated using a k(0,0) weighted kappa coefficient. An estimated k(0,0) value of 0.255 indicates only a fair agreement (i.e., $0.2 < k \leq 0.4$) between a positive reference standard result (i.e., presence of a DBP, as measured via the

Table 5

Proportions of true positive, false positive, false negative, and true negative outcomes for the study by Glascoe, MacLean & Stone (1991)

		Behavioral Concerns		Marginal
		Yes	No	Proportion
Disruptive Behavior Problems	Yes	0.147 (0.036) [0.076 - 0.219]	0.063 (0.025) [0.014 - 0.112]	0.211 (0.042) [0.129 – 0.293]
	No	0.211 (0.042) [0.129 - 0.293]	0.579 (0.051) [0.480 - 0.678]	0.789 (0.042) [0.708 – 0.872]
Marginal Proportion		0.358 (0.049) [0.262 - 0.454]	0.642 (0.049) [0.546 - 0.739]	1.00

Note. Standard errors are given in parentheses; 95% confidence intervals are given in square parentheses

ECBI) and a positive index test result (i.e., presence of parents' concerns, as measured via the index test).

Table 7 presents measures that indicate the accuracy with which parents' concerns can rule out the presence of a DBP in young children. An estimated NPV of 0.902 indicates that, among all parents who did not express concerns, 90.2% had a child without a DBP. An estimated SE of 0.7 indicates that, among all children with a DBP, 70% of their parents expressed behavioral concerns. RR3 was estimated to be 2.448, indicating that the odds that a child does *not* have a DBP are 2.5 times higher among parents that did *not* report concerns (i.e.,

Table 6

Measures of the Accuracy of Parents' Concerns at Ruling In the Presence of a Disruptive Behavior Problem in Young Children: Results from Glascoe, MacLean & Stone (1991)

PPV	**SP**	**RR1**	**k(0,0)**
0.412 (0.084)	0.733 (0.051)	2.626 (0.068)	0.255 (0.009)
[0.246 – 0.577]	[0.633 – 0.833]	[2.493 – 2.760]	[0.238 – 0.272]

Note. PPV is the positive predictive value of parents' concerns; SP is the specificity of parents' concerns; Jackknife estimates of RR1 and k(0,0) are given, where RR1 is a relative risk ratio, and k(0,0) is a weighted kappa coefficient. Standard errors are given in parentheses; 95% confidence intervals are given in square parentheses.

Table 7

Measures of the Accuracy of Parents' Concerns at Ruling Out the Presence of a Disruptive Behavior Problem in Young Children: Results from Glascoe, MacLean & Stone (1991)

NPV	SE	RR3	k(1,0)
0.902 (0.038)	0.700 (0.102)	2.448 (0.100)	0.533 (0.01)
[0.827 – 0.976]	[0.499 – 0.901]	[2.252 – 2.644]	[0.501 – 0.564]

Note. NPV is the negative predictive value of parents' concerns; SE is the sensitivity of parents' concerns; Jackknife estimated of RR3 and k(1,0) are given, where RR3 is a relative risk ratio; k(1,0) is a weighted kappa coefficient. Standard errors are given in parentheses, and 95% confidence intervals are given in square parentheses.

55/6 = 9.17), compared to the same odds irrespective of the presence or absence of parents' concerns (i.e., 75/20 = 3.75). Again, following Kraemer (1992), these measures were calibrated using a k(1,0) weighted kappa coefficient. An estimated k(1,0) value of 0.533 indicates only a moderate agreement (i.e., $0.4 < k \leq 0.6$) between a negative reference standard result (i.e., absence of a DBP, as measured via the ECBI) and a negative index test result (i.e., absence of parents' concerns, as measured via the index test).

CHAPTER 4 – DISCUSSION

4.1 Overview

The importance of early screening programs for DBPs is well-established. Selective screening strategies, like those proposed by the latest CPS guidelines (Charach et al., 2017), may offer an opportunity for early detection and prevention of DBPs in young children. Within a Canadian health context, however, it is prudent to ensure that accurate screening occurs. This review sought to determine whether parents' concerns can provide enough accurate information to PCPs, such that they can efficiently decide in favour of or against screening for DBPs in young children. A limited number of results were retrieved by this review. This lack of data may indicate that high-quality evidence to demonstrate the accuracy of parents' concerns in distinguishing between children with DBPs and those without is difficult to retrieve

using systematic review methods. These results also leave room for questions about the existence of equal or more accurate methods that can be used to determine when screening for DBPs in young children is warranted.

4.2 Efficiency of Systematic Review Methodology: Scope and Applications

Our systematic search strategy produced 8420 records and resulted in 4964 abstracts screened for eligibility by reviewers. Just 53 of these records were eligible to be reviewed in full. Out of the 53 full-text studies reviewed, only one study (Glascoe et al., 1991) met all eligibility criteria. Despite the large number of records generated by our systematic search strategy, this review yielded more limited data than was previously identified in a recent rapid review (Baillargeon et al., 2022). Interestingly, most records that were excluded during full-text screening were excluded due to problems within the studies' designs (n=33). Further, most of these records were excluded due to heterogeneity in how studies defined or measured parents' concerns and DBPs in children (*n*=24/33). This result points to a need for additional research that a) generates consensus on how behavior problems and parents' concerns, as constructs, are defined, and b) develops gold standard screening tools that reliably measure these constructs. Until such time that behavioral concerns and DBPs are defined and measured consistently within the literature, systematic review methods may continue to be inefficient in retrieving data that addresses these constructs.

Given the results of this review, it is possible that diagnostic test accuracy methods may not be the most important method by which we can assess the predictive validity of parents' concerns. Instead, we may benefit from more observational data that assesses the longitudinal trajectory of FP and FN cases. Such work could provide much needed, objective data about the harms associated with both outcomes, and may provide an empirical basis on which we can identify which outcomes should be addressed with greater political attention. Alternatively,

with unlimited resources, one could study the clinical outcomes associated with both FP and FN cases via RCT; however, such a method is likely not be feasible given ethical constraints.

4.3 The Accuracy of Parent's Concerns: Quality and Scope of Evidence

Ultimately, reviewers did not find good-quality evidence in the results of Glascoe et al. (1991) to demonstrate that parents' concerns can accurately distinguish between children with DBPs and those without. While the reported sensitivity of parents' concerns (SE = 0.700, [0.499; 0.901]) was relatively high, calibration of this result demonstrated only moderate agreement (k(1,0) = 0.533, [0.501; 0.564]) between a negative reference standard result (i.e., absence of DBPs) and a negative index test result (i.e., absence of parents' concerns). Given that the estimated k(1,0) value is less than 0.6, this result suggests that the absence of parents' concerns may not provide enough accurate information to PCPs to rule *out* the presence of a DBP and decide against screening (Landis & Koch, 1977). Similarly, the specificity of parents' concerns (SP = 0.733, [0.633; 0.833]) was also relatively high, but calibration of this result demonstrated only fair agreement (k(0,0) = 0.255, [0.238;0.272]) between a positive reference standard result (i.e., presence of DBPs) and a positive index test result (i.e., presence of parents' concerns). Given that the estimated value of k(0,0) is again less than 0.6, this result also suggests that the presence of parents' concerns may not provide enough accurate information to PCPs to rule *in* the presence of DBPs and decide in favour of screening.

Given the ratio of k(1,0) to k(0,0) (i.e., .533 / .255 = 2.091), parents' concerns appear to be relatively better at ruling *out* the presence of DBPs than they are at ruling *in*. In fact, any test that is better at ruling *out* illness will have a value of k(1,0) that is higher than k(0,0). When interpreting the relationship between these quality indices, it is important to assess the relative harms associated with a low value of either quality index. These harms are influenced by the relative prevalence of concerns (i.e., P and P'), and by the SE and SP of the test. As mentioned,

while the SE of parents' concerns (0.700) was only slightly smaller than the SP (0.733), the quality of the specificity (k(0,0)=0.255) was lower than that of the sensitivity (k(1,0)=0.533). Additionally, there are more children without DBPs (P'=0.789) in the sample from Glascoe et al. (1991) than there are children with DBPs (P=0.211), and more FP cases (i.e., .211 or 20 / 95) than FN cases (i.e., .063 or 6 / 95). Since the relative prevalence of P to P' are so disparate, and the SE and SP of parents' concerns are roughly equal, parents' concerns in this context are more likely to generate FP outcomes than FN outcomes.

There are significant harms associated with the use of a test that is relatively poor at ruling *in* illness (i.e., with a low k(0,0) value) and that which is likely to generate a high volume of FPs. This is due to the relative number of children affected by these kinds of errors, where more well children are likely to be misclassified as having a DBP. These harms can include unnecessary testing and increased emotional burden for a larger proportion of (well) children and their families. Indirectly, this scenario can also lead to harm for the smaller proportion of children with DBPs, where a greater demand for screening and diagnostic services can reduce the accessibility of diagnostic testing for children who really need it. Thus, the harms associated with using a test that has a relatively poor ability to rule *in* illness (i.e., low k(0,0)) are likely to impact a majority of children, particularly in a system where there are finite resources available at a diagnostic level. In this context of the data from Glascoe et al. (1991), it is therefore concerning that the quality of the specificity of parents' concerns (k(0,0)) is so low.

This is not to say that there are no harms associated with the use of a test that is poor at ruling *out* illness (i.e., where k(1,0) < k(0,0)). In the context of the data from Glascoe et al. (1991), parents' concerns were still unlikely to provide enough accurate information to rule *out* the presence of DBPs given that the value of k(1,0) was smaller than 0.6 (Landis & Koch, 1977). A poor ability to rule *out* illness in this context can still cause harm for the minority children who do have DBPs. Given the significant burden of suffering associated with the

presence of DBPs, failure to decide in favour of screening when necessary may cost children with DBPs access to timely interventions – interventions which, importantly, are effective in attenuating the risks of long-term harm. While these costs impact the minority of children in this sample (i.e., children with DBPs, P=0.211), the scale of these costs is substantial for those affected, and therefore should not be underestimated.

In addition to the results of our data analysis, reviewers found there to be a high risk of bias within the methods of Glascoe et al. (1991), which limits the strength of the evidence presented. There remains some ambiguity in how exactly parents' concerns were elicited in the study by Glascoe et al. (1991), and by what index test. This points, again, to ambiguity in how parents' behavioral concerns are defined and measured in this field. Further, the order of test administration, ambiguity in researcher blinding methods also contribute to a high risk of bias within the reported findings. Lastly, a discrepancy exists between the cut-off threshold defined by Glascoe et al. (1991) and the cut-off threshold defined by the authors of the reference standard (i.e., the ECBI). This may have biased downward the estimates of accuracy reported here. Overall, the quality of the evidence retrieved by this review further emphasizes the need for more high-quality research that address whether parents' concerns can accurately distinguish between children with DBPs and children without. Until such a time that good quality evidence on this subject is accessible via methods like systematic reviews, researchers may remain without the necessary tools to determine whether the presence or absence of parents' concerns justifies deciding in favour of or against screening for DBPs in young children.

4.4 Limitations

4.4.1 Limitations Due to Construct Heterogeneity

There is a known limitation introduced by the way in which the threshold for parents' concerns is defined. As mentioned previously, defining parents' concerns too narrowly, such

that many concerns are excluded that do map to behavioral problems, can impact the validity of marginal proportions (i.e., the relative prevalence of behavior problems, P and P'). This, in turn can artificially reduce an SE estimate, and may also lead to improper estimates of k(1,0) and of the marginal odds ratios. Any change to the marginal odds estimates will impact the ratio of kappa coefficients, which then can impact the interpretation of whether a test is better at ruling *in* or ruling *out* illness. While studies that elicited behavioral, social, or emotional concerns from parents were eligible for inclusion in this review, the working definition of parents' concerns that was implemented by Glascoe et al. (1991) remained quite narrow relative to the definition of concerns implemented by this review. In the study by Glascoe et al. (1991), parent concerns about domains of development outside of their child's behavior were excluded from their analysis. Therefore, it is possible that the estimates of SE and k(1,0) generated from Glascoe et al.'s data (1991) and reported here, may be skewed downward.

4.4.2 The Absence of a Gold-Standard

In addition, the absence of a gold standard from which we may verify the presence/absence of DBPs limits our ability to estimate the accuracy of any index test. Not only does the absence of a gold standard contribute to the variability in the methods by which any study assesses the presence of DBPs in young children, the use of a non-"gold standard" test in diagnostic test accuracy may yield biased estimates of the prevalence of cases and non-cases (Baillargeon et al., 2004). Further, over- or underestimation of the prevalence of the target condition will result in over- or underestimation of the reliability of an index test (Baillargeon et al., 2004). The absence of a gold-standard tool therefore limits the reliability of the accuracy estimates presented by this review (as generated from data by Glascoe et al. (1991)). Reviewers caution that the results presented here should not be interpreted as

absolute indicators of the accuracy of parents' concerns, but rather as *estimates* that can be biased by classification errors in both the reference standard and the index test.

4.4.3 Limitations Imposed on Search Strategy

In attempt to capture as much published data as possible that addressed the accuracy of parents' concerns, the review authors implemented an intentionally broad search strategy. However, limitations were imposed within narrow eligibility criteria in attempt to generate relatively homogenous data from which we could generate summary estimates of accuracy. For example, age limits on the population of interest (i.e., children under 5 years of age) were imposed in effort to mirror the context in which the clinical practice guidelines by Charach et al. (2017) were recommended. However, this search limitation likely reduced the number of records that were ultimately eligible for this review, and thus compromised our ability to generate summary estimates.

Given the limited number of studies ultimately included in this review, in hindsight it may have been prudent to preserve the broader, 0–12-year age range that was identified in the review protocol. In doing so, a larger volume of studies may have ultimately been eligible for inclusion in this review. This may also have facilitated collection of enough data to successfully generate summary measures of accuracy, in accordance with the planned meta-analysis. However, a recent rapid review by Baillargeon et al. (2022) did assess the accuracy of parents' concerns for a broader target population (0-12-years) and it did not yield many more included studies (n=2). Given these previously published results, and concerns about the lack of available resources to manage a potential increase in the volume of results, reviewers ultimately decided to maintain the narrowed the age range criteria.

An additional limitation was imposed on search results such that dimensions of disruptive behavior which closely align with Attention Deficit/Hyperactivity Disorder (ADHD) (i.e., impulsivity, hyperactivity) were not included among key words. This was done to

accommodate time constraints imposed by the volume of initial results produced by our search strategy. This search limitation was also intended to better align the target condition of this review (i.e., DBPs) with the four dimensions of disruptive behavior that were proposed by Carter et al. (2013) (i.e., non-compliance, temper-loss, low concern for others, and aggression). However, while ADHD is not traditionally categorized as a DBD on its own, there is a high degree of comorbidity between DBDs and ADHD (APA, 2013; Loeber, Burke & Pardini, 2009). Therefore, it's possible that DTA studies of disruptive behavior related to ADHD alone could have contained data that were relevant to the accuracy of parents' concerns about disruptive behavior in their children. Hence, it is possible that these search limitations further reduced the number of eligible records retrieved by our search strategy. As a result, this may have limited our ability to generate summary estimates.

Lastly, in the interests of conserving time and resources, reviewers decided not to perform a comprehensive search of grey literature. This may have further limited the number of eligible studies retrieved by this review.

4.5 Interdisciplinary Contributions to Research and Future Directions

The current CPS screening guidelines for DBPs promote the psychological health and development of young children through routine and early screening (Charach et al., 2017). This prevention strategy bridges the fields of public health, psychology, and medicine. In the context of these guidelines and this review, the assessment of the accuracy of parents' concerns further employed research and analysis methods from each of these three fields. This interdisciplinary approach attempted to bridge some of the research gaps that exist between the assessment of diagnostic test accuracy in medicine, and the assessment of screening test accuracy in psychology and public health.

While the results of this review did not provide robust answers to our research questions, they uncovered a gap in the current literature that may be addressed by future, interdisciplinary

research. As the prevention of psychological disorders becomes more central to public health policy, the development of accurate screening measures that can detect early symptoms may be in the interests of stakeholders. This may be particularly important for disorders that have an early age of onset and are associated with a considerable burden of suffering, like DBPs. With respect to the findings of this review, future research may benefit from addressing the accuracy of parents' concerns in a broader context. A focus on broader age ranges, and other developmental domains may facilitate the construction of a better-defined framework from which parents' concerns about their children's development can be categorized. Further, additional research that builds on the work of Glascoe et al. (1991) is needed to develop a summary estimate of the accuracy of parents' concerns about their child's behavior.

As it stands, while parents' concerns seem to be quite poor at ruling *in* the presence of DBPs, eliciting parents' concerns to gather information about the developmental status of children may still be useful in the context of selective developmental screening strategies. In fact, despite limitations to the reliability of parents' concerns that are demonstrated by these findings, previous research has demonstrated that the expected benefit conferred by selective screening strategies that use parents' concerns is still greater than the expected benefit conferred by an alternative, universal screening strategy (Baillargeon, 2021).

However, there are a wide variety of methods by which parents' concerns could be elicited that may improve their reliability within such selective screening strategies. In the context of the current Canadian practice guidelines, policy makers could consider implementing more explicit recommendations that better define *how* parents' concerns should be elicited by PCPs. For example, parents' concerns have been shown to be more reliable when elicited systematically by PCPs (i.e., via a questionnaire like the PEDS) (Glascoe, 2013a). The inclusion of specific methods or tools to the current recommendations that assist clinicians in

eliciting parents' concerns may help minimize screening errors and minimize variability from clinician to clinician.

Considering the evidence presented in this review, current guidelines could also better identify which children are most in need of screening by incorporating an if/and algorithm into their practice recommendations. The current guidelines first instruct PCPs to elicit parents' concerns about their child's behavior. Given that parents' concerns are relatively poor at ruling *in* the presence of DBPs, policy makers could (for example) recommend that PCPs follow up with a second question that has a relatively better ability to rule *in* DBPs (I.e., has a higher k(0,0) value) than do parents' concerns. Following this, PCPs could be instructed to proceed with screening only if they receive positive responses from parents on both questions. Future studies would benefit from further examining whether there are other metrics that are more accurate in their ability to rule *in* the presence of DBPs than parents' concerns, and whether the implementation of such a decision-making procedure into practice guidelines would enhance the reliability with which we can identify children in need of early screening and intervention.

4.6 Conclusions

Ultimately, any conclusive evidence regarding the accuracy of parents' concerns remains ambiguous following the results of this review. However, the findings presented here are not intended to imply that parents' concerns are wholly unreliable as a means by which we can assess the need to screen for DBPs in young children; rather, these results simply demonstrate that there is minimal evidence upon which we may verify that parents' concerns are reliable. What evidence has been collected from the results of Glascoe et al. (1991) offers some, low-quality evidence that parents' concerns may not provide PCPs with enough accurate information to rule *out* DBPs in young children, and thus decide against screening. Further, these results also indicate that parents' concerns are quite poor at ruling *in* the presence of DBPs. This is concerning from a population health perspective, in that these findings cannot

with great confidence confirm that the methods recommended in current practice guidelines are likely to optimize the early identification of DBPs in young children.

Additionally, the results of this review suggest that there is a paucity of good-quality evidence that replicates the work by Glascoe et al. (1991), and that which is retrievable using systematic review methods. Therefore, significant questions remain about the extent to which the presence of parents' concerns justifies screening for DBPs in young children, and vice versa. Considering these results, future research that seeks to refine the methods by which we define and elicit parents' concerns about child behavior, and that which seeks to replicate the findings of Glascoe et al. (1991) is needed. Without such research, our ability to assess the efficacy of selective screening guidelines, like those published by the Canadian Pediatric Society (2017), remains limited.

Funding Statement

The authors received no financial support for the research or authorship of this review.

CHAPTER 5 – REFERENCES

Alakortes, J., Kovaniemi, S., Carter, A.S., Bloigu, R., Moilanen, I.K., & Ebeling, H.E. (2017). Do child healthcare professionals and parents recognize social-emotional and behavioral problems in 1-year-old infants? *European Child & Adolescent Psychiatry, 26,* 481-495. doi: 10.1007/s00787-016-0909-3

Alink, L.R.A., Mesman, J., Van Zeijl, J., Stolk, M., Juffer, F., Koot, H.M., Bakermans-Kranenburg, M.J., & Van Ijzendoorn, M.H. (2006). The Early Childhood Aggression Curve: Development of Physical Aggression in 10- to 50-Month-Old Children. *Child Development, 77*(4), 954-966. https://doi.org/10.1111/j.1467-8624.2006.00912.x

American Academy of Pediatrics Committee on Practice and Ambulatory Medicine & Bright Futures Periodicity Schedule Workgroup. (2020) Recommendations for Preventive Pediatric Health Care: Policy Statement. *Pediatrics*, *145*(3), e20170254. https://doi.org/10.1542/peds.2017-0254

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.* Arlington, VA: American Psychiatric Association.

August, G.J., Realmuto, G.M., Crosby, R.D., & MacDonald, A.W. (1995). Community-Based Multiple-Gate Screening of Children at Risk for Conduct Disorder. *Journal of Abnormal Child Psychology, 23*(4). doi: 10.1007/BF01447212

August, G.J., Braswell, L., & Thuras, P. (1998). Diagnostic Stability of ADHD in a Community Sample of School-Aged Children Screened for Disruptive Behavior. *Journal of Abnormal Child Psychiatry, 26*(5), 345-356. doi: 10.1023/A:1021999722211

Baillargeon, R.H. (2021). Clinical practice guidelines for monitoring children's behavioral development at the 18-month well-baby visit: A decision analysis comparing the expected benefit of two alternative strategies. *Journal of Evaluation in Clinical Practice*, *27*(1) 1-7. doi: 10.1111/jep.13385

Baillargeon, R.H., Charette, M., Tessier, F., & Brand, K.P. (2022). Clinical Practice Guidelines About Screening for Disruptive Behavior Problems at Well-Child Visits: A Rapid Review of the Literature on the Accuracy of Parents' Behavioral Concerns. *OBM Integrative and Complementary Medicine, 7*(3), 1-18. doi: 10.21926/obm.icm.2203031

Baillargeon, R.H., Keenan, K., & Cao, G. (2012). The Development of Opposition-Defiance in Toddlerhood: A Population-Based Cohort Study. *Journal of Developmental and Behavioural Pediatrics, 33,* 608-617. doi: 10.1097/DBP.0b013e3182653c6c

Baillargeon, R.H., Tremblay, R.E., Willms, D., Romano, E., Lee, K., & Wu, H. (2004). Modeling intraindividual change over time in the absence of a "Gold Standard". *Psychology Science, 1,* 7-34.

Barkley, R., Shelton, T., Crosswait, C., Moorehouse, M., Fletcher, K., Barrett, S., Jenkins, L., & Metevia, L. (2002). Preschool children with disruptive behavior: Three-year outcome as a function of adaptive disability. *Development and Psychopathology, 14.* 45-67.

Barkauskienė, R., Bongarzoni, A.D., Bieliauskaitė, R., Jusienė, R., & Raižienė, S. (2009). Attention deficit/hyperactivity disorder: Possibilities of early diagnostics. *Medicina, 45*(10), 764-771. MDPI AG. Retrieved from http://dx.doi.org/10.3390/medicina45100099

Bayer, J., Hiscock, H., Scalzo, K., Mathers, M., McDonald, M., Morris, A., Birdseye, J., & Wake, M. (2009). Systematic review of preventive interventions for children's mental health: what would work in Australian contexts? *Australian & New Zealand Journal of Psychiatry, 43*(8), 695-710. https://doi.org/10.1080/00048670903001893

Bengi-Arslan, L., Verhulst, F.C., van der Ende, J., & Erol, N. (1997). Understanding childhood (problem) behaviors from a cultural perspective: comparison of problem behaviors and competencies in Turkish immigrant, Turkish, and Dutch children. *Social Psychiatry and Psychiatric Epidemiology, 32*(8)*,* 477-484. doi: 10.1007/BF00789143

Berger-Jenkins, E., Monk, C., D'Onfro, K., Sultana, M., Brandt, L., Ankam, J.,

Vazquez, N., Lane, M., & Meyer, D. (2019). Screening for Both Child Behavior and Social Determinants of Health in Pediatric Primary Care. *Journal of Developmental and Behavioral Pediatrics, 40*(6), 415-425. doi: 10.1097/DBP.0000000000000676

Bergold, S., Christiansen, H., & Steinmayr, R. (2019). Interrater agreement and discrepancy when assessing problem behaviors, social-emotional skills, and developmental status of kindergarten children. *Journal of Clinical Psychology, 75,* 2210-2232. DOI: 10.1002/jclp.22840

Briggs-Gowan, M. J. (1996). *A parent assessment of social-emotional and behavior problems and competence for infants and toddlers: Reliability, validity, and associations with maternal symptoms and parenting stress* (Order No. 9635397) [Doctoral Dissertation, Yale University]. Available from ProQuest Dissertations & Theses Global.

Briggs-Gowan, M.J. & Carter, A.S. (2008). Social-Emotional Screening Status in Early Childhood Predicts Elementary School Outcomes. *Pediatrics, 121*(5), 957-962, doi:10.1542/peds.2007-1948.

Briggs-Gowan, M.J., Carter, A.S., McCarthy, K., Augustyn, M., Caronna, E., & Clark, R. (2012). Clinical Validity of a Brief Measure of Early Childhood Social-Emotional/Behavioral Problems. *Journal of Pediatric Psychology, 38*(5), 577-587. doi:10.1093/jpepsy/jst014

Briscoe, S., Bethel, A., & Rogers, M. (2018). Conduct and reporting of citation searching in Cochran systematic reviews: A cross-sectional study. *Research Synthesis Methods, 11*, 169-180. doi: 10.1002/jrsm.1355

Campbell, S.B., Breaux, A.M., Ewing, L.J., & Szumowski, E.K. (1984). A One-Year Follow-up Study of Parent-referred Hyperactive Preschool Children. *Journal of the American Academy of Child Psychiatry, 23*(3), 243-249. doi: 10.1016/S0002-7138(09)60499-4

Carter, A.S., Gray, S.A.O., Baillargeon, R.H., & Wakschlag, L.S. (2013). A

multidimensional Approach to Disruptive Behaviors: Informing Life Span Research from an Early Childhood Perspective. In PH Tolan & BI Leventhal (Eds.)*, Disruptive Behavior Disorders, Advances in Development and Psychopathology: Brain Research Symposium Series* (pp. 103-135). New York: Springer Science+Business Media.

Caye, A., Spadini, A.V., Karam, R.G., Grevet, E.H., Rovaris, D.L., Bau, C.H.D., Rohde, L.A., & Kieling, C. (2016). Predictors of persistence of ADHD into adulthood: a systematic review of the literature. *European Child and Adolescent Psychiatry, 25*, 1151-1159. doi: 10.1007/s00787-016-0831-8

Chacko, A., Granski, M., Horn, E. P., Levy, M. D., Dahl, V., Lacks, R. S., & Ramakrishnan, A. (2018). Prevention of disruptive behavior problems in children. In M. M. Martel (Ed.), *Developmental Pathways to Disruptive, Impulse-Control and Conduct Disorders* (pp. 347–380). New York, NY: Academic Press. https://doi.org/10.1016/B978-0-12-811323-3.00014-6

Charach, A., Bélanger, S. A., McLennan, J. D. & Nixon, M. K. (2017). Le dépistage des comportements perturbateurs en première ligne chez les enfants d'âge préscolaire. *Paediatrics and Child Health (Canada), 22*(8), 478-493. https://doi.org/10.1093/pch/pxx128.

Charach, A., Mohammadzadeh, F., Bélanger, S. A., Easson, A., Lipman, E. L., McLennan, J. D., Parkin, P., & Szatmari, P. (2020). Identification of Preschool Children with Mental Health Problems in Primary Care: Systematic Review and Meta-analysis. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Académie canadienne de psychiatrie de l'enfant et de l'adolescent*, *29*(2), 76–105. ISSN : 1719-8429.

Chen, I., Lee, H., Yeh, G., Lai, C., & Chen, S. (2004). The Relationship Between Parental Concerns and Professional Assessment in Development Delay in Infants and Children – A Hospital-based Study. *Journal of the Chinese Medical Association, 67*(5)*,* 239-244. PMID: 15357111

Christenson, J.D., Crane, D.R., Malloy, J. & Parker, S. (2016). The Cost of Oppositional Defiant Disorder and Disruptive Behavior: A Review of the Literature. *Journal of Child and Family Studies, 25,* 2649-2658. doi: 10.1007/s10826-016-0430-9.

The Cochrane Collaboration. (2020). Data Collection form for intervention review – RCTs and non-RCTs. Retrieved December 4, 2020 from https://dplp.cochrane.org/data-extraction-forms

Copeland, W.E., Shanahan, L., Costello, J. & Angold, A. (2009). Childhood and Adolescent Psychiatric Disorders as Predictors of Young Adult Disorders. *Archives of General Psychiatry, 66*(7), 764-772. doi: 10.1001/archgenpsychiatry.2009.85

Edelstein, M.L., Moen, A., Benson, J.L., Smucker, R., & Perkins-Parks, S. (2022). Development and Implementation of a Function-Based Clinical Interview to Evaluate Childhood Behavior Problems. *Cognitive and Behavioral Practice*. In press, online publication. https://doi.org/10.1016/j.cbpra.2022.01.003

Ellingson, K.D., Briggs-Gowan, M.J., Carter, A.S., & Horwitz, S.M. (2004). Parent Identification of Early Emerging Child Behavior Problems. *Archives of Pediatric & Adolescent* Medicine, 158, 766-772. doi: 10.1001/archpedi.158.8.766

Erskine, H.E., Norman, R.E., Ferrari, A.J., Chan, G.C.K., Copeland, W.E., Whiteford, H.A., & Scott, J.G. (2016.). Long-Term Outcomes of Attention-Deficit/Hyperactivity Disorder and Conduct Disorder: A Systematic Review and Meta-Analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(10), 841-850. doi: 10.1016/j.jaac.2016.06.016

Eyberg, S. & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Behavior Inventory-Revised: Professional Manual*. Psychological Assessment Resources.

Fanton, J.H., MacDonald, B., & Harvey, E.A. (2008). Preschool Parent-Pediatrician Consultation and Predictive Referral Patterns for Problematic Behaviors. *Journal of*

Developmental and Behavioral Pediatrics, 29(6), 475-482. DOI: 10.1097/DBP.0b013e31818d4345

Feeney-Kettler, K.A. (2008). *Early identification of preschool students at risk for emotional and behavioral disorders: Development and validation of a parent-teacher screener* (Order No. 3327961) [Doctoral Dissertation, University of Wisconsin – Madison]. Available from ProQuest Dissertations & Theses Global.

Feldman, M.A., Hancock, C.L., Rielly, N., Minnes, P., & Cairns, C. (2000). Behavior Problems in Children with or At Risk for Developmental Delay. *Journal of Child and Family Studies, 9*(2). 247-261. https://doi.org/10.1023/A:1009427306953

Foster, E.M., Jones, D.E. & The Conduct Problems Prevention Research Group. (2005). The High Costs of Aggression: Public Expenditures Resulting from Conduct Disorder. *American Journal of Public Health, 95*(10), 1767-1772. doi: 10.2105/AJPH.2004.061424.

Glascoe, F. (1997). *Parents' Evaluations of Developmental Status: A method for detecting and addressing developmental and behavioral problems in children*. Ellsworth & Vandermeer Press.

Glascoe, F. P. (1998). Behavioral and developmental problems in children. *Patient Care*, *32*(8), 63+. https://link-gale-com.proxy.bib.uottawa.ca/apps/doc/A20620524/AONE?u=otta77973&sid=bookmark-AONE&xid=937a869e

Glascoe, F.P. (1999[a]). A method for deciding how to respond to parents' concerns about developmental status. *Ambulatory Child Health, 5*(3), 197-208. https://search-ebscohost-com.proxy.bib.uottawa.ca/login.aspx?direct=true&db=a9h&AN=6313231&site=ehost-live

Glascoe, F.P. (1999[b]). Using Parents' Concerns to Detect and Address Developmental and Behavioral Problems. *Journal of the Society of Pediatric Nurses, 4*(1), 24-35. doi: 10.1111/j.1744-6155.1999.tb00077.x

Glascoe, F.P. (2003). Parents' Evaluation of Developmental Status: How Well Do Parents' Concerns Identify Children with Behavioral and Emotional Problems? *Clinical Pediatrics, 42*(2). 133-138. doi: 10.1177/000992280304200206

Glascoe, F. P. (2013a). Ch. 3: The Reliability of PEDS. In *Collaborating with Parents: Using Parents' Evaluation of Developmental Status to Detect and Address Developmental and Behavioral Problems* (2nd ed., pp. 103–114). chapter, PEDStest.com, LLC.

Glascoe, F. P. (2013b). Ch. 4: The Validity of PEDS. In *Collaborating with Parents: Using Parents' Evaluation of Developmental Status to Detect and Address Developmental and Behavioral Problems* (2nd ed., pp. 115–135). chapter, PEDStest.com, LLC.

Glascoe, F.P., Altemeier, W.A. & MacLean., W.E. (1989). The Importance of Parents' Concerns About Their Child's Development. *American Journal of Diseases in Children, 143*(8), 955-958.

Glascoe, F.P., MacLean, W.E., & Stone, W.L. (1991). The Importance of Parents' Concerns About Their Child's Behavior. *Clinical Pediatrics, 30*(1), 8-11. doi: 10.1177/000992289103000101

Godoy, L., Carter, A.S., Silver, R.B., Dickstein, S., & Seifer, R. (2014). Infants and Toddlers Left Behind: Mental Health Screening and Consultation in Primary Care. *Journal of Developmental & Behavioral Pediatrics, 35*(5)*,* 334-353. doi: 10.1097/DBP.0000000000000060

Goodwin, C. J. (2008). Observational and Survey Research Methods. In *Research in psychology: Methods and design* (5th ed., pp. 423–460). chapter, Wiley.

Gray, S.A.O., Carter, A.S., Briggs-Gowan, M.J., Hill, C., Danis, B., Keenan, K., & Wakschlag, L.S. (2012). Preschool Children's Observed Disruptive Behavior: Variations Across Sex, Interactional Context, and Disruptive Psychopathology. *Journal of Clinical Child & Adolescent Psychology, 41*(4), 499-507. doi: 10.1080/15374416.2012.675570

Herman-Staab, B. (1994). Screening, Management, and Appropriate Referral for Pediatric Behavior Problems. *The Nurse Practitioner, 19*(7), 40-49. doi: 10.1097/00006205-199407000-00009

Holtz, C. A. (2010). *Screening of Behavior Problems in Young Children from Low-Income Families: The Development of a New Assessment Tool* (Order No. 3375129) [Doctoral Dissertation, Marquette University]. Available from ProQuest Dissertations & Theses Global.

Ilić, S.B., Nikolić, S.J., Ilić-Stošović, D.D., & Golubović, Š.S. (2020). Early identification of children with developmental delay and behavioral problems according to parents concerns in the Republic of Serbia. *Early Childhood Development and Care, 190*(16), 2605-2611. https://doi.org/10.1080/03004430.2019.1595610

Ireton, H. (1996). The Child Development Review: Monitoring Children's Development Using Parents' and Pediatrician's Observations. *Infants & Young Children, 9*(1), 42-52. DOI: 10.1097/00001163-199607000-00006

Jekel, J. F., Katz, D. L., Elmore, J. G., & Wild, D. M. (2007). Understanding the Quality of Data in Clinical Medicine. In *Epidemiology, Biostatistics, and Preventive Medicine* (3rd ed., p. 108-114). Philadelphia, PA: Saunders Elsevier.

Karabekiroglu, K., Uslu, R., Kapci-Seyitoglu, E.G., Özbaran, B., Öztop, D.B., Özel-Özcan, Ö., Dogangün, B., Gülen-Sismanlar, S., Görker, I., Fidan, T., Bahali, K.M., Barut, Y., Gürkan, K., Kilic, B.G., Ay, P., Taskin, B., Bilgin, N., & Cengel-Kültür, E.S. (2013). A

nationwide study of social-emotional problems in young children in Turkey. *Infant Behavior and Development, 36.* 162-170. http://dx.doi.org/10.1016/j.infbeh.2012.11.007

Kraemer, H.C. & Gibbons, R.D. (2009). Where Do We Go Wrong in Assessing Risk Factors, Diagnostic and Prognostic Tests? The Problems of Two-by-two Association. *Psychiatric Annals, 39*(7), 711-718. https://doi.org/10.3928/00485713-20090625-05

Kraemer, H.C. (1992). *Evaluating Medical Tests : Objective and Quantitative Guidelines*. SAGE Publications.

Kruizinga, I., Jansen, W., de Haan, C.L., & Raat, H. (2012). Reliability and Validity of the KIPPPI: An Early Detection Tool for Psychosocial Problems in Toddlers. *PLoS ONE, 7*(11), e49633. DOI: 10.1371/journal.pone.0049633

Landis, J.R. & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

Lavigne, J.V., Bryant, F.B., Hopkins, J., & Gouze, K.R. (2019). Age 4 Predictors of Oppositional Defiant Disorder in Early Grammar School. *Journal of Clinical Child & Adolescent Psychology, 48*(1), 93-107. https://doi.org/10.1080/15374416.2017.1280806

Leibson, C.L. & Long, K.H. (2003). Economic Implications of Attention-Deficit Hyperactivity Disorder for Healthcare Systems. *Pharmacoeconomics, 21*(17), 1239-1262. https://doi.org/10.2165/00019053-200321170-00002.

Lipkin, P.H., Macias, M.M. & AAP Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics. (2020). Promoting Optimal Development: Identifying Infants and Young Children With Developmental Disorders Through Developmental Surveillance and Screening, *Pediatrics*, *145*(1), 1-19. https://doi.org/10.1542/peds.2019-3449

Loeber, R., Burke, J.D., & Pardini, D.A. (2009). Development and Etiology of Disruptive and Delinquent Behavior. *Annual Review of Clinical Psychology, 5*, 291-310. doi: 10.1146/annurev.clinpsy.032408.153631

Lorber, M.F., Del Vecchio, T., & Smith Slep, A.M. (2014). Infant Externalizing Behavior as a Self-Organizing Construct. *Developmental Psychology, 50*(7), 1854-1861. http://dx.doi.org/10.1037/a0036985

Matza, L.S., Paramore, C., & Prasad, M. (2005). A review of the economic burden of ADHD. *Cost Effectiveness and Resource Allocation, 3*(5), 1-9. doi: 10.1186/1478-7547-3-5.

McInnes, M.D.F., Moher, D., Thombs, B.D., McGrath, T.A., Bossuyt, P.M., & the PRISMA-DTA Group (2018). Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA, 319*(4), 388-396. doi:10.1001/jama.2017.19163

Nærde, A., Ogden, T., Janson, H., & Zachrisson, H.D. (2014). Normative Development of Physical Aggression From 8 to 26 Months. *Developmental Psychology, 50*(6), 1710-1720. Doi: 10.1037/a0036324

Ogden, T. & Hagen, K.A. (2008). Treatment Effectiveness of Parent Management Training in Norway: A Randomized Controlled Trial of Children with Conduct Problems. *Journal of Consulting and Clinical Psychology, 76*(4), 607-621. doi: 10.1037/0022-006X.76.4.607

Offord, D.R. & Bennett, K.J. (1994). Conduct Disorder: Long-Term Outcomes and Intervention Effectiveness. *Journal of the American Academy of Child & Adolescent Psychiatry, 33*(8), 1069-1078. https://doi.org/10.1097/00004583-199410000-00001

Offord, D.R., Kraemer, H.C., Kazdin, A.E., Jensen, P.S., Harrington, R. & Gardner, J.S. (1999). Lowering the Burden of Suffering: Monitoring the Benefits of Clinical, Targeted, and Universal Approaches. In D. Keating & C. Hertzman (Eds.), *Developmental Health and*

the Wealth of Nations (pp. 293-310). New York, NY: The Guilford Press.

Pavuluri, M.N., Luk, S., Clarkson, J., & McGee, R. (1995). A community study of preschool behavior disorder in New Zealand. *Australian and New Zealand Journal of Psychiatry, 29,* 454-462. doi: 10.3109/00048679509064954

Petitclerc, A. & Tremblay, R.E. (2009). Childhood Disruptive Behaviour Disorders: Review of Their Origin, Development, and Prevention. *Canadian Journal of Psychiatry, 54*(4), 222-231. https://doi.org/10.1177/070674370905400403

Plamondon, A., Browne, D.T., Madigan, S., & Jenkins, J.M. (2018). Disentangling Child-Specific and Family-Wide Processes Underlying Negative Mother-Child Transactions. *Journal of Abnormal Child Psychiatry, 46,* 437-447. Doi: 10.1007/s10802-017-0309-2

Poduska, J.M. (2000). Parents' Perceptions of Their First-Graders' Need for Mental Health and Educational Services. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(5), 584-591. doi: 10.1097/00004583-200005000-00012

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. doi: 10.1111/jcpp.1238.

Reardon, T., Harvey, K., Baranowska, M., O'Brien, D., Smith, L., & Creswell, C. (2017). What do parents perceive are the barriers and facilitators to accessing psychological treatment for mental health problems in children and adolescents? A systematic review of qualitative and quantitative studies. *European Child & Adolescent Psychiatry, 26,* 623-647. doi: 10.1007/s00787-016-0930-6

Reitsma, J.B., Glas, A.S., Rutjes, A.W.S., Scholten, R.J.P.M., Bossuyt, P.M., & Zwinderman, A.H. (2005). Bivariate analysis of sensitivity and specificity produces

informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology, 58,* 982-990. doi: 10.1016/j.jclinepi.2005.02.022

Robinson, E.A., Eyberg, S.M., & Ross, A.W. (1980). The Standardization of an Inventory of Child Conduct Problem Behaviors. *Journal of Clinical Child Psychology, 9*(1), 22-28. https://doi.org/10.1080/15374418009532938

Sabol, T.J., Kessler, C.L., Rogers, L.O., Petitclerc, A., Silver, J., Briggs-Gowan-M., & Wakschlag, L.S. (2022). A window into racial and socioeconomic status disparities in preschool disciplinary action using developmental methodology. *Annals of the New York Academy of Sciences, 1508*, 123-136. Doi: 10.1111/nyas.14687

Sand, E.A. (1972). Symptômes de comportement d'enfants belges observés par leur mère II : Associations entre symptômes (3 à 9 ans). *Revue de neuropsychiatrie infantile et d'hygiène mentale de l'enfance, 20*(3), 253-265.

Sawyer, M.G., Mudge, J., Carty, V., Baghurst, P., & McMichael, A. (1996). A prospective study of childhood emotional and behavioral problems in Port Pirie, South Australia. *Australian and New Zealand Journal of Psychiatry, 30*(6), 781-787. doi: 10.3109/00048679609065045

Scott, J.G., Pederson M.G., Erskine, H.E., Bikic, A., Demontis, D., McGrath, J.J., & Dalsgaard, S. (2017). Mortality in individuals with disruptive behavior disorders diagnosed by specialist services – A nationwide cohort study. *Psychiatry Research, 251,* 255-260. https://doi.org/10.1016/j.psychres.2017.02.029

Sheldrick, B.C., Neger, E.N., & Perrin, E.C. (2012). Concerns About Development, Behavior, and Learning Among Parents Seeking Pediatric Care. *Journal of Developmental & Behavioral Pediatrics, 33*(2), 156-160. doi: 10.1097/DBP.0b013e3182420f4a

Szczepaniak, D., McHenry, M.S., Nutakki, K., Bauer, N.S., & Downs, S.M. (2013). The Prevalence of At-Risk Development in Children 30 to 60 Months Old Presenting With Disruptive Behaviors. *Clinical Pediatrics, 52*(10), 942-949. doi: 10.1177/0009922813493832

Thomas, B.H., Byrne, C., Offord, D.R., & Boyle, M.H. (1991). Prevalence of Behavioral Symptoms and the Relationship of Child, Parent, and Family Variables in 4- and 5-Year-Olds: Results from the Ontario Child Health Study. *Developmental and Behavioral Pediatrics, 12*(3), 177-184. doi: 10.1097/00004703-199106000-00006

Waddell, C., McEwan, K., Shepherd, C., Offord, D. & Hua, J. (2005). A Public Health Strategy to Improve the Mental Health of Canadian Children. *Canadian Journal of Psychiatry, 50*(4), 226-233. https://doi.org/10.1177/070674370505000406

Waddell, C., McEwan, K., Peters, R.D., Hua, J.M., & Garland, O. (2007). Preventing Mental Disorders in Children: A Public Health Priority. *Canadian Journal of Public Health, 98*(3), 174-178. doi: 10.1007/BF03403707

Wakschlag, L.S., Briggs-Gowan, M.J., Choi, S.W., Nichols, S.R., Kestler, J., Burns, J.L., Carter, A.S., & Henry, D. (2014). Advancing a Multidimensional, Developmental Spectrum Approach to Preschool Disruptive Behavior. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(1), 82-96. doi: 10.1016/j.jaac.2013.10.011

Wakschlag, L.S., Choi, S.W., Carter, A.S., Hullsiek, H., Burns, J., McCarthy, K., Leibenluft, E., & Briggs-Gowan, M.J. (2012). Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *The Journal of Child Psychology and Psychiatry, 53*(11), 1099-1108. Doi: 10.1111/j.1469-7610.2012.02595.x

Ware, H.S., Jouriles, E.N., Spiller, L.C., McDonald, R., Swank, P.R., & Norwood, W.D. (2001). Conduct Problems Among Children at Battered Women's Shelters: Prevalence and Stability of Maternal Reports. *Journal of Family Violence, 16*(3), 291-307. doi:

10.1023/A:1011190316783

Wendland, J., Danet, M., Gacoin, E., Didane, N., Bodreau, N., Saïas, T., Le Bail, M., Cazenave, M., Molina, T., Puccinelli, O., Chirac, O., Medeiros, M., Gérardin, P., Cohen, D., & Guédeney, A. (2014). French Version of the Brief Infant-Toddler Social and Emotional Assessment Questionnaire – BITSEA. *Journal of Pediatric Psychology, 39*(5). 562-575. doi: doi:10.1093/jpepsy/jsu016

Whiting, P.F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., & Bossuyt, P.M.M. (2011[a]). QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Annals of Internal Medicine*, *155*(8), 529-536. https://doi.org/10.7326/0003-4819-155-8-201110180-00009

Whiting, P. F., Rutjes, A.W.S., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M.G., Sterne, J.A.C., & Bossuyt, P.M.M. (2011[b]). QUADAS-2 background document. Bristol; University of Bristol.

Winsper, C. & Wolke, D. (2014). Infant and Toddler Crying, Sleeping and Feeding Problems, and Trajectories of Dysregulated Behavior Across Childhood. *Journal of Abnormal Child Psychiatry, 42,* 831-843. Doi: 10.1007/s10802-013-9813-1

CHAPTER 6 – APPENDICES

6.1 Appendix A: Search Strategy Adaptations by Database and Date Searched

6.1.1 Medline(Ovid) – Mar. 22, 2022

1	Child, Preschool/	972275
2	(child* or kid? or toddler? or girl? or boy? or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*).ti,ab.	1847261
3	1 or 2	2237264
4	Child Behavior Disorders/ or "attention deficit and disruptive behavior disorders"/ or conduct disorder/	26367
5	((dysfunc* or problem* or disrupt* or defian* or dysregulate*) adj behav*).ti,ab.	10927
6	((disrupt* or defian*) adj2 problem*).ti,ab.	573
7	exp Aggression/	41898
8	(aggression or aggressive*).ti,ab.	230200
9	"low concern for others".ti,ab.	9
10	("callous?unemotional traits" or "callous, unemotional traits").ti,ab.	681
11	(temper adj tantrum*).ti,ab.	250
12	(disobedien* or non-complian* or "non compliant" or "non compliance").ti,ab.	7772
13	problem behavior/	3290
14	4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13	288961
15	exp Diagnosis/ or mass screening/	9051820
16	diagnosis.fs.	2796540
17	(screen* or test or tests or testing).ti,ab.	3247904
18	((measur* or screen*) adj2 tool*).ti,ab.	45726
19	15 or 16 or 17 or 18	11855852
20	((parent* or mother* or father* or guardian* or caregiver*) adj2 (concern* or worr* or identifi* or complain* or opinion* or observ* or screen* or report* or insight* or attunement)).ti,ab.	47763
21	3 and 14 and 19 and 20	**1782**

**A total of 1782 records were retrieved from Medline.*

6.1.2 PsycInfo(Ovid) – Mar. 23, 2022

1	behavior disorders/ or conduct disorder/ or disruptive behavior disorders/ or oppositional defiant disorder/	15119
2	((dysfunc* or problem* or disrupt* or defian* or dysregulate*) adj behav*).ti,ab.	20854
3	((disrupt* or defian*) adj2 problem*).ti,ab.	841
4	Behavior Problems/ or Externalizing Symptoms/	31433
5	Aggressive Behavior/ or Relational Aggression/	27392
6	(aggression or aggressive*).ti,ab.	82240
7	exp callous-unemotional traits/	585
8	"low concern for others".ti,ab.	9
9	tantrums/ or anger/	10670
10	(tantrum* or (temper adj tantrum*)).ti,ab.	983
11	(disobedien* or non-complian* or "non compliant" or "non compliance").ti,ab.	2263
12	or/1-11	145189
13	(child* or kid? or toddler? or girl? or boy? or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*).ti,ab.	813452
14	exp Diagnosis/ or Screening/ or Screening Tests/	246417
15	(screen* or test or tests or testing).ti,ab.	749173
16	((measur* or screen*) adj2 tool*).ti,ab.	13491
17	or/14-16	938556
18	((parent* or mother* or father* or guardian*) adj2 (concern* or worr* or identifi* or complain* or opinion* or observ* or screen* or report* or insight* or attunement)).ti,ab.	36862
19	12 and 13 and 17 and 18	**833**

**A total of 833 records were retrieved from PsycInfo.*

6.1.3 Embase(Ovid) – Mar. 23, 2022

1	behavior disorder/ or externalizing disorder/	61421
2	((dysfunc* or problem* or disrupt* or defian* or dysregulate*) adj behav*).ti,ab.	13323
3	((disrupt* or defian*) adj2 problem*).ti,ab.	701
4	problem behavior/ or disruptive behavior/	10461
5	exp aggression/	125010
6	(aggression or aggressive*).ti,ab.	342284
7	"low concern for others".ti,ab.	9
8	("callous?unemotional traits" or "callous, unemotional traits").ti,ab.	807
9	anger/	20282
10	(tantrum* or (temper adj tantrum*)).ti,ab.	823
11	(disobedien* or non-complian* or "non compliant" or "non compliance").ti,ab.	15388
12	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11	495411
13	exp preschool child/ or exp toddler/	670947
14	(child* or kid? or toddler? or girl? or boy? or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*).ti,ab.	2581857
15	13 or 14	2829967
16	((parent* or mother* or father* or guardian*) adj2 (concern* or worr* or identifi* or complain* or opinion* or observ* or screen* or report* or insight* or attunement)).ti,ab.	54540
17	screening/ or developmental screening/	193051
18	diagnosis/	1540669
19	diagnosis.fs.	3537986
20	(screen* or test or tests or testing).ti,ab.	4716785
21	((measur* or screen*) adj2 tool*).ti,ab.	67427
22	17 or 18 or 19 or 20 or 21	8894168
23	12 and 15 and 16 and 22	**1658**

**A total of 1658 records were retrieved from Embase.*

6.1.4 CINAHL – Mar. 23, 2022

S1	(MH "Child Behavior Disorders") OR (MH "Social Behavior Disorders")	12,130
S2	(TI (((dysfunc* or problem* or disrupt* or defian* or dysregulate*) N1 behav*))) OR (AB (((dysfunc* or problem* or disrupt* or defian* or dysregulate*) N1 behav*)))	15,914
S3	TI ((disrupt* N2 problem*) or (defian* N2 problem*)) OR AB ((disrupt* N2 problem*) or (defian* N2 problem*))	382
S4	(MH "Disruptive Behavior")	3,881
S5	MH Aggression/	11,827
S6	TI (aggression or (aggressive N1 behav*) or aggressive*) OR AB (aggression or (aggressive N1 behav*) or aggressive*)	47,141
S7	TI "low concern for others" OR AB "low concern for others"	3
S8	TI "callous-unemotional traits" OR AB "callous-unemotional traits"	394
S9	TI temper N1 tantrum* OR AB temper N1 tantrum*	99
S10	TI ((disobedien*) or (non-complian*) or (« non compliant ») or (« non compliance »)) AND AB ((disobedien*) or (non-complian*) or (« non compliant ») or (« non compliance »))	158
S11	S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10	76,074
S12	(MH "Child, Preschool")	222,802
S13	TI (child* or kid? Or toddler? Or girl? Or boy? Or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*) OR AB (child* or kid? Or toddler? Or girl? Or boy? Or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*)	671,978
S14	S12 OR S13	733,928
S15	(MH "Diagnosis, Developmental")	434
S16	MH "Health Screening"	52,101
S17	TI (screen* or test or tests or testing) OR AB (screen* or test or tests or testing)	847,719
S18	TI ((measur* or screen*) N2 tool*) OR AB ((measur* or screen*) N2 tool*)	22,229
S19	S15 OR S16 OR S17 OR S18	867,487
S20	TI ((parent* or mother* or father* or guardian* or caregiver*) N2 (concern* or worr* or 67dentify* or complain* or opinion* or observ* or screen* or report* or insight* or attunement)) OR AB ((parent* or mother* or father* or guardian* or caregiver*) N2 (concern* or worr* or 67dentify* or complain* or opinion* or observ* or screen* or report* or insight* or attunement))	32,806
S21	S11 AND S14 AND S19 AND S20	**618**

**A total of 618 records were retrieved from CINAHL*

6.1.5 Cochrane Central Register of Controlled Trials – Mar. 26, 2022

1	"attention deficit and disruptive behavior disorders"/ or conduct disorder/ or child behavior disorders/	1363
2	((dysfunc* or problem* or disrupt* or defian* or dysregulate* or problem*) adj behav*).ti,ab.	1629
3	((disrupt* or defian* or behav*) adj2 problem*).ti,ab.	3928
4	exp aggression/	1389
5	(aggression or aggressive*).ti,ab.	12053
6	"low concern for others".ti,ab.	0
7	("callous?unemotional traits" or "callous, unemotional traits").ti,ab.	54
8	(temper adj tantrum*).ti,ab.	33
9	(disobedien* or non-complian* or "non compliant" or "non compliance").ti,ab.	1692
10	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	18760
11	child, preschool/	31125
12	(child* or kid? or toddler? or girl? or boy? or preschool* or nurser* or pre-school* or prekindergarten* or kindergarten* or pediat* or paediat*).ti,ab.	165430
13	11 or 12	170829
14	exp diagnosis/	356347
15	mass screening/	3407
16	diagnosis.fs.	0
17	(screen* or test or tests or testing).ti,ab.	390002
18	((measur* or screen*) adj2 tool*).ti,ab.	4075
19	14 or 15 or 16 or 17 or 18	667435
20	((parent* or mother* or father* or guardian*) adj2 (concern* or worr* or identifi* or complain* or opinion* or observ* or screen* or report* or insight* or attunement)).ti,ab.	6575
21	10 and 13 and 19 and 20	**392**

** A total of 392 records were retrieved from Cochrane Central*

6.1.6 Scopus – Mar 23, 2022

```
[title-abs-key ( ( behav W/0 disorder* ) OR ( "disruptive behav* problem*" ) OR ( "child behav*
disorder*" ) OR ( problem W/0 behavior* ) ) OR TITLE-ABS-
KEY ( ( dysfunc* OR problem* OR disrupt* OR defian* OR dysregulate* ) W/1 behav* ) OR TITLE-ABS-
KEY ( ( disrupt* OR defian* OR conduct ) W/2 problem* ) OR TITLE-ABS-
KEY ( aggression OR aggressive* ) OR TITLE-ABS-KEY ( "low concern for others" OR "callous-unemotional
traits" ) OR TITLE-ABS-KEY ( ( temper W/1 tantrum* ) OR tantrum* OR anger ) OR TITLE-ABS-
KEY ( disobedien* OR "non?complian*" OR "non w/0 complian*" ) ] AND [title-abs-
key ( child* OR kid? OR toddler? OR girl? OR boy? OR preschool* OR nurser* OR pre-
school* OR prekindergarten* OR kindergarten* OR pediat* OR paediat* ) ] AND [title-abs-
key ( diagnos* OR screen* OR test OR tests OR testing ) OR TITLE-ABS-
KEY ( ( measur* OR screen* ) W/2 tool* ) ] AND [title-abs-
key ((parent* OR mother* OR father* OR guardian* OR caregiver*) W/2 (concern* OR worr* OR identifi
* OR complain* OR opinion* OR observ* OR screen* OR report* OR insight* OR attunement))]
```

**A total of 3134 records were retrieved from Scopus*

6.*2 Appendix B: Data extraction form for Studies with a Naturalistic Sampling Design*

ORGANIZATIONAL DATA					
REF ID		REVIEWER DATE		CHECKED BY	
Author/year					
Journal/source					
Country of origin					
Publication type					
STUDY CHARACTERISTICS					
Sample size					
Number of participants excluded					
Recruitment method					
Setting					
Location					
Funding					
Attrition					
PARTICIPANT CHARACTERISTICS (if reported)					

Age (mean, sd)	
Ethnicity (%)	
Sex (%)	
Annual household income	
Diagnostic status of participants at rdx (pre-study)	
Comorbid conditions?	
METHODOLOGY	
Behavioural problem(s) studied	
Index test used	
Reference standard used	
Timeline/flow of assessments	
Outcomes measures reported/assessed	

OUTCOME CHARACTERISTICS (if reported)	
Outcome 1: **true positives**	Primary Secondary Not Defined
Definition of outcome	
Timing/flow of assessment	

# Participants evaluated for outcome	
Reasons for drop out/participant exclusion	
Source of information	
Outcome 2: **false positives**	Primary secondary not defined
Definition of outcome	
Timing/flow of assessment	
# Participants evaluated for outcome	
Reasons for drop out/participant exclusion	
Source of information	
Outcome 3: **true negatives**	Primary secondary not defined
Definition of outcome	
Timing/flow of assessment	
# Participants evaluated for outcome	
Reasons for drop out/participant exclusion	
Source of information	
Outcome 4: **false negatives**	Primary secondary not defined
Definition of outcome	
Timing/flow of assessment	
# Participants evaluated for outcome	
Reasons for drop out/participant exclusion	

Source of information	

6.3 Appendix C: Quality Assessment of Diagnostic Accuracy Studies – 2 (QUADAS-2)

Phase 1: State the review question:

Patients (setting, intended use of index test, presentation, prior testing):
Index test(s):
Reference standard and target condition:

Phase 2: Draw a flow diagram for the primary study

Phase 3: Risk of bias and applicability judgments

QUADAS-2 is structured so that 4 key domains are each rated in terms of the risk of bias and the concern regarding applicability to the research question (as defined above). Each key domain has a set of signalling questions to help reach the judgments regarding bias and applicability.

DOMAIN 1: PATIENT SELECTION

A. Risk of Bias

Describe methods of patient selection:

- ❖ Was a consecutive or random sample of patients enrolled? Yes/No/Unclear
- ❖ Was a case-control design avoided? Yes/No/Unclear
- ❖ Did the study avoid inappropriate exclusions? Yes/No/Unclear

Could the selection of patients have introduced bias? **RISK: LOW/HIGH/UNCLEAR**

B. Concerns regarding applicability

Describe included patients (prior testing, presentation, intended use of index test and setting):

Is there concern that the included patients do not match the review question? **CONCERN: LOW/HIGH/UNCLEAR**

DOMAIN 2: INDEX TEST(S)

If more than one index test was used, please complete for each test.

A. Risk of Bias

Describe the index test and how it was conducted and interpreted:

- ❖ Were the index test results interpreted without knowledge of the results of the reference standard? Yes/No/Unclear
- ❖ If a threshold was used, was it pre-specified? Yes/No/Unclear

Could the conduct or interpretation of the index test have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question? **CONCERN: LOW /HIGH/UNCLEAR**

DOMAIN 3: REFERENCE STANDARD

A. Risk of Bias

Describe the reference standard and how it was conducted and interpreted:

- ❖ Is the reference standard likely to correctly classify the target condition? Yes/No/Unclear
- ❖ Were the reference standard results interpreted without knowledge of the results of the index test? Yes/No/Unclear

Could the reference standard, its conduct, or its interpretation have introduced bias? **RISK: LOW /HIGH/UNCLEAR**

B. Concerns regarding applicability

Is there concern that the target condition as defined by the reference standard does not match the review question? **CONCERN: LOW /HIGH/UNCLEAR**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram):

Describe the time interval and any interventions between index test(s) and reference standard:

- ❖ Was there an appropriate interval between index test(s) and reference standard? Yes/No/Unclear
- ❖ Did all patients receive a reference standard? Yes/No/Unclear
- ❖ Did patients receive the same reference standard? Yes/No/Unclear
- ❖ Were all patients included in the analysis? Yes/No/Unclear

Could the patient flow have introduced bias? **RISK: LOW /HIGH/UNCLEAR**