The Effect of Stop Now And Plan (SNAP™) Program on Children with Conduct Disorder in Community Settings: A Meta-Analysis

by

Fatma Osman

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Community Health Sciences

University of Manitoba

Winnipeg

Copyright © 2020 by Fatma Osman

ABSTRACT

The early identification and treatment of children with conduct disorder (CD) help in preventing future engagements in criminal activities. Literature shows that cognitive-behavioural treatments (CBT) may have desirable results in children under 12 years of age with CD. Stop Now, And Plan (SNAP^{TM}) as one of the CBT interventions is provided in community settings for children under 12 who act antisocially and aggressively. Some primary studies have been conducted to evaluate the SNAP^{TM} program's effectiveness. These evaluations have shown the program's effectiveness in reducing aggressive, conduct, rule-breaking, antisocial, and overall behavioural problems. The purpose of the study was to conduct a Meta-Analysis that combines data from the SNAP^{TM} primary evaluation studies to produce a precise total estimate of the program effect. Specifically, the study aimed to (1) determine the effect of SNAP^{TM} program on the children's CD in community settings, and (2) determine whether variations in treatment intensity and fidelity affect the program's outcomes.

The outcome summary data from each study that met the inclusion criteria were extracted by using a standardized data extraction form and were meta-analyzed by using the Review Manager. The Random-Effect Meta-Analysis methods were used to account for the heterogeneity in design, treatment intensity and fidelity among studies.

The study's results showed that the SNAP^{TM} program made a significant moderate reduction in the children's conduct problems (the summary effect size is 0.41 with 95% CI of 0.25 to 0.57, I^2 of 58%, and based on four studies with 231 participants) immediately after they participated in the program. This trend has continued to a significant large effect after six months of follow-up with a summary effect size of 0.61, 95% CI of 0.44 to 0.99, I^2 of 64%, and based on four studies with 318 participants. The results also showed that the program had a significant reduction in children's externalizing and internalizing behavioural problems as well as significant improvements in prosocial skills, immediately and at follow-up. Furthermore, the children who participated in the program were less likely involved in criminal convictions according to eleven years' follow-up observations.

Although the study did not systematically review the SNAP^{TM} program literature, it is the first Meta-Analysis study that examines the effectiveness of the SNAP^{TM} program. A need for well-conducted trials and more extended periods of follow-up are needed to further help in understanding the efficacy of the program.

ACKNOWLEDGEMENTS

I want to acknowledge my gratitude and thankfulness to my supervisor, Dr. Depeng Jiang, who guided and supported me during my MSc studies and thesis work. I would also like to express my appreciation for each member of my thesis committee, Dr. Rasheda Rabbani, Professor Sarah Heath and Dr. Javier Mignone, for their valuable comments and suggestions that contributed to the improvement of the thesis.

I would further like to thank my ultimate role model mother and the soul of my father, who was impatiently waiting for this moment. I would also like to thank my siblings and my children, Seham, Ali and Muhammad, for their ongoing support and patience. Thank you all so much! I could not have done it without you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Conduct disorder (CD) is one of the many terms used to describe children with persistent problems with aggressive and antisocial behaviours (Augimeri, 2005). CD interferes with children's healthy development and functioning in many contexts (e.g., home, school, and community), continuing into adulthood with considerable social, psychological, and fiscal costs (Institute of Medicine, 2004). In comparison with children that harbour different psychiatric disorders, children with early onset of CD have lower rates of education and employment, a higher likelihood to engage in criminal activities, as well as a greater chance of suffering from physical and mental disorders later in life (Koegl, Farrington, Augimeri, & Day, 2008). By averting the occurrence of a crime committed by an offender with antisocial behaviour history, crime prevention programs that target antisocial children have the potential to save society at least "1.4 million dollars per case" (Koegl, 2011). The early identification and treatment of children with early onset of CD help in preventing future engagements in criminal activities. Yet, CD is considered as one of the most difficult disorders to treat because its complexity requires a carefully designed treatment (Children's Mental Health Ontario, 2001). Findings from the literature show that a cognitive-behavioral, multifaceted treatment has desirable results in children under 12 years of age with conduct problems (Augimeri, Farrington, Koegl, & Day, 2007). Children with CD usually receive treatment in community settings where a considerable amount of resources is required to implement an effective treatment for them (Children's Mental Health Ontario, 2001). Stop Now And Plan (SNAP^{TM}) as one of the cognitive-behavioural interventions that are provided in community settings, "helps children and their parents regulate angry feelings by getting them to stop, think, and plan positive alternatives before they act impulsively" (Augimeri, Jiang, Koegl, & Carey, 2006, p. 6). Over the last three decades, the SNAP^{TM} program has been used to treat children with challenging behaviours or

conduct disorder. This intervention has gone through some internal evaluations conducted by the program developers, and external evaluations performed by third parties (Child Development Institute, 2019). These evaluation studies have demonstrated the program's effectiveness. Some studies have shown reductions in the children's externalizing behaviours (Augimeri et al., 2007; Augimeri et al., 2006; Lipman, Kenny, Brennan, O'Grady, & Augimeri, 2011; Lipman et al., 2008). Other evaluations have shown reductions in both externalizing and internalizing behaviours (Burke & Loeber, 2015, 2016; Hrynkiw-Augimeri, Pepler, & Goldberg, 1993; Pepler et al., 2010; Walsh, Pepler, & Levene, 2002). Some studies have also shown improvements in the children's prosocial skills (Burke & Loeber, 2015, 2016; Lipman et al., 2011; Lipman et al., 2008; Pepler et al., 2010; Walsh et al., 2002). The treatment gain was maintained up to 15-month of follow-up for children who received an intensive version of the SNAP^{TM} program (Augimeri et al., 2007; Augimeri et al., 2006). These evaluations, however, had some limitations such as; small sample sizes, low participation rates and high rates of attrition (Augimeri et al., 2007; Hrynkiw-Augimeri et al., 1993; Koegl et al., 2008; Lipman et al., 2011; Lipman et al., 2008; Pepler et al., 2010). Studies with small sample sizes need careful interpretation. The lack of statistically significant results in small studies does not mean the treatment is not effective (i.e., type II error). Small studies can also over-estimate the magnitude of the treatment effect or produce a false-positive effect (i.e., type I error) (Hackshaw, 2008). Furthermore, small studies limit the statistical analyses and power to detect post-treatment changes (Hackshaw, 2008; Pepler et al., 2010). To minimize or avoid the occurrence of type I and II errors, and also to provide more reliable evidence, larger confirmatory studies are needed (Hackshaw, 2008). Furthermore, there are some concerns about the high attrition rates especially among the control groups as significant, and marginally significant differences were detected in the baseline behaviour problems between the missing and

remaining children in the control groups (Augimeri et al., 2007; Koegl et al., 2008). The attrition of high-risk children in the control group "… may have had the effect of biasing the control group in favour of a more 'treatment amenable' group…", and again "If this is true, it would have resulted in a more conservative test of the overall program effect, in comparison with the other treatment groups." (Koegl et al., 2008, p. 431). Accordingly, the purpose of the study was to conduct a Meta-Analysis that combines data from the independent SNAP™ primary evaluation studies to increase the statistical power and produce a more precise overall estimate of the effect of SNAP™ program on children with conduct disorder in community settings. Specifically, the study sought to:

1. Determine the effect of SNAP™ program on the level and trend of children's conduct disorder in community settings
2. Determine whether variations in treatment intensity and fidelity affect the program's outcomes.

DEFINITION OF TERMS

Conduct disorder: according to the DSM-5 diagnostic criteria, the CD is "A repetitive and persistent pattern of behaviour in which the basic rights of others or major age-appropriate societal norms or rules are violated, …" (American Psychiatric Association, 2013, p. 146). Examples of CD behaviours include "frequent fighting, lying, stealing, fire-setting, cruelty to others (or to animals), and destroying property." (Alan E Kazdin, 2005, p. 6)

Externalizing Behavior Problems: refer to behaviour problems that "… reflect conflict with others and violation of social norms." (Lande et al., 2009, p. 3). Externalizing behaviours include rule-breaking behaviour, aggressive behaviour and delinquent behaviour. CD, oppositional defiant disorder (ODD), attention-deficit/hyperactivity disorder (ADHD) and antisocial behaviours are other examples of externalizing behaviours (Alan E Kazdin, 2005).

Internalizing Behavior Problems: represent behaviours that "… reflect mood disturbance, including anxiety, depression, and social withdrawal." (Lande et al., 2009, p. 3).

Juvenile Delinquency: the expression of behavioural acts that conflict with the law, like robbery, underage drinking, and running away from home (A. E. Kazdin, 2005).

Treatment Fidelity: the program fidelity represents the adherence level of the program activities to the treatment manual (Augimeri et al., 2007). Integrity and fidelity checklists are used to ensure reaching high levels of adherence.

Treatment intensity: SNAP™ program intensity refers to the number of sessions received from the program's core components (i.e., SNAP™ children and parent groups), besides some additional optional components (e.g., individual befriending, individual family counselling, child academic tutoring) (Koegl et al., 2008).

Youth criminal convictions: refer to "all offences committed between each child's 12th and 18th birthday" (Augimeri et al., 2007, p. 802). Some of these convictions are "contempt, failure to pay fines, terroristic threats, robbery, and aggravated assault with injury" (Burke & Loeber, 2015, p. 250). Breaking and entering, use of a weapon, fraud, mischief, drug and theft are other examples of reported youth criminal convictions (Augimeri, Pepler, Walsh, Jiang, & Dassinger, 2010).

LITERATURE REVIEW

SNAP™ Program Overview

A team of scientists and practitioners in the Child Development Institute, CDI, in Toronto, Canada (Formerly known as the Earlscourt Child and Family Centre) have developed the SNAP™ program. Precisely the team consisted of "Kenneth Goldberg, Kathryn Levene, Leena Augimeri, Elizabeth Leggett, Camille Hannays and Dr. Debra Pepler" (Child Development Institute, 2016a). Further members jointed the development team subsequently including "Kathy Williams, Nicola

Slater, Carl Riley, Christopher Koegl, Karen Sewell, Dr. David Day and Dr. Christopher Webster" (Child Development Institute, 2016a). The CDI originally developed SNAP™ as "a day treatment program" for children with CD, four decades ago (Child Development Institute, 2016a). In 1985, the program was further developed to become a comprehensive cognitive-behavioural intervention for children under 12 who act antisocially and aggressively. SNAP™ program, which is now known as the SNAP Under 12 Outreach Project (SNAP™ ORP) (Webster, Augimeri, & Koegl, 2002), is based on the cognitive-behavioural theory (CBT) that focuses on behavioural and cognitive alterations. Cognitive-Behavioural therapies are globally used to treat psychiatric disorders in patients of different ages (Dobson & Dozois, 2019). The program participants learn how to be self-aware when they are triggered and angry. They learn how to regulate and stop their angry feelings by using calming down strategies like taking deep breaths and counting to 10. Meanwhile, they are taught to think of making effective plans and solutions to make their problems smaller (Augimeri, Walsh, & Slater, 2011). In addition to CBT, the SNAP™ ORP program incorporates other theoretically sound models, namely: parent management training, social skills training, problem-solving, self-control and anger management approaches, cognitive self-instruction, and family management skills training (Burke & Loeber, 2015).

Referral to the SNAP™ ORP program is initially done by "teachers, social workers, police and other service providers" (Child Development Institute, 2016a). Referred children usually show some behavioural disorders such as disobedience, lying, verbal aggression, depressive symptoms, temper tantrums, fire setting, running away, cruelty to animals, stealing at home and outside, vandalism, resentfulness, and using weapons when fighting (Augimeri et al., 2006; Walsh et al., 2002). Unhealthy parent-child interactions, anxiety, delinquency, aggression and violence, antisocial values and conduct, poor self-control and problem-solving, cognitive distortions,

bullying, authority contact, school failure and isolation, are additional commonly exhibited behavioural disorders (Child Development Institute, 2016a). Only children whose ages are between 6 and 11 with a T-score of 70 or higher on the delinquency subscale of the Child Behavior Checklist (CBCL), or have had recent police contact as a result of CD, are admitted to the program (Farrington & Koegl, 2015).

Components of SNAP™ Program

SNAP™ ORP program consists of two core components offered to all involved children and their parents. These two components are represented in the SNAP™ children group, formally known as Transformer Club and SNAP™ parent group. These two groups receive manualized 90-minute sessions that occur concurrently after school in a community-based outpatient setting for 12 or 13 consecutive weeks. In these sessions, children learn cognitive-behavioural self-control and problem-solving techniques that enable them to stop and think about the negative consequences of their behaviour before acting impulsively. In the parents' sessions, parents learn effective child management techniques (As cited in Webster et al., 2002). In addition to the two core components, there are other optional components such as family counselling based on 'Stop Now And Plan Parenting' (SNAPP), in-home academic tutoring, school advocacy and teacher consultation, victim restitution, and individual befriending (Webster et al., 2002). Access to these additional components is based on the child's and family's needs and preferences (Koegl et al., 2008). The number of sessions received from the program's core components (i.e., SNAP™ children and parent groups), besides some additional optional components (e.g., individual befriending, individual family counselling, child academic tutoring) determine the treatment intensity (Koegl et al., 2008). Moreover, attendance of a minimum of 8 out of the 12 sessions of the SNAP™ core components is required for participants to graduate (Lipman et al., 2008).

The SNAP™ ORP program was initially designed as a "gender-neutral" intervention. However, due to the negative consequences observed in girls that participated in mixed-gender treatment groups, the SNAP™ Girls Connection program was developed in 1996 with three main structured components; SNAP™ Girls, SNAP™ Parenting and Girls Growing Up Healthy (Pepler et al., 2010). In addition to the theoretical frameworks mentioned earlier, the SNAP™ Girls Connection incorporates other theoretical approaches such as development theory, relationship theory, social learning theory and multisystemic approach to helping families (Walsh et al., 2002). The program aims to achieve long-term goals such as: keeping girls in school, reducing aggressive and antisocial behaviours and preventing possible negative trajectories such as teen pregnancy (Lipman, Kenny, & Wymouth, 2007; Walsh et al., 2002).

Moreover, the CDI developed another program as an extension to the SNAP™ ORP program, called SNAP™ youth programming. This program was first launched in 1994 to meet the needs of children that graduated from SNAP™. Participating adolescents can access different treatment components provided through "SNAP™ Youth Leadership Services" such as leadership clubs, employment counselling, summer leaders in training programs, school advocacy and tutoring, individual and family counselling, parent workshops, and victim restitution (Child Development Institute, 2016a). Table 1 outlines the various SNAP™ components, participants, type and number of required sessions.

SNAP™ Assessment and Evaluation Tools

The referred children and their families undergo a thorough screening process to determine their eligibility for admission to the program (Augimeri, 2005). To implement the SNAP™ program effectively, the treatment team needs to conduct a comprehensive, detailed assessment to identify the nature of the behavioural problems and assist in tailoring an adequate subsequent treatment

Table 1 SNAP™ components, participants, type and number of required sessions

SNAP™ component	**Participants**	**Type**	**Number of required sessions**
SNAP™ Boys group/Club	6-11 years old boys	Mandatory	12-13 weeks
SNAP™ Girls group/Club	6-11 years old girls	Mandatory	12-13 weeks
SNAP™ Parenting group	Children's parents/caregivers	Mandatory	12-13 weeks
Family Counselling	Family	Optional	Depends on preference and need
Individual Counselling/Mentoring	6-11 Children and Youth	Optional	Depends on preference and need
School Advocacy and Teacher Support	6-11 Children and youth	Optional	Depends on preference and need
Girls Growing up Healthy	6-11 Girls	Optional	Depends on preference and need
Academic Tutoring	6-11 Children and Youth	Optional	Depends on preference and need
Victim Restitution	6-11 Children and Youth	Optional	Depends on preference and need
Community Connections	6-11 Children and Youth	Optional	Depends on preference and need
Job Readiness	6-11 Children and Youth	Optional	Depends on preference and need
Long-Term Continued Care Services	6-11 Children and Youth	Optional	Depends on preference and need
SNAP™ Boys Youth Leadership Services	Boys youth	Optional	Depends on preference and need
SNAP™ Girls Youth Leadership Services	Girls youth	Optional	Depends on preference and need
Parent Problem-Solving Group	Children's parents/caregivers	Optional	Depends on preference and need

intervention that meets the clinical needs of the admitted children and their families (Hupp, Reitman, & Jewell, 2008).

Early Assessment Risk Lists (EARLs)

EARL is a comprehensive, evidence-based risk assessment tool structured for use by clinicians and other specialists to assess children under the age of 12 with behaviour problems signifying continued severe issues in the future (Augimeri, 2005; Augimeri et al., 2006; Augimeri et al., 2010; Child Development Institute, 2016a). The EARL has "… a threefold purpose: (1) to provide a platform for increasing clinicians' and researchers' general understanding of early childhood risk factors; (2) to offer structure that helps clinicians systematically identify and manage risks in order to plan appropriate treatment to improve clinical outcomes; and (3) as a result, to improve the reliability and validity in predicting the likelihood of antisocial children engaging in future delinquent behaviour …" (As cited in Augimeri et al., 2010, p. 14). Furthermore, there are separate risk assessment tools for young boys and girls, EARL-20B and EARL-21G, respectively. The EARL tool is gender-sensitive because the underlying risk factors and the kind of behaviour problems exhibited as well as the future consequences vary between boys and girls (Augimeri, 2005; Augimeri et al., 2006; Augimeri et al., 2010; Child Development Institute, 2016a; Moffitt, Caspi, Rutter, & Silva, 2001). In general, EARL tools have three main categories related to the child, family, and the child's responsiveness to the planned treatment. There are 20 items in the EARL tool for boys (EARL-20B), and 21 items for EARL for girls (EARL-21G). The rating of these items is as follows: (0) for not present, (1) for somewhat present, and (2) for present. The total score of EARLs ranges from (0-40) for boys, and from (0-42) for girls. The high scores are indicative of children's high-risk patterns (Augimeri, 2005; Augimeri et al., 2006; Augimeri et al., 2010). Moreover, EARL tools have been translated into many languages and used for over

eighteen years (Koegl, 2011). Drawing upon the work of other scholars, Farrington and Koegl (2015) stated that the EARL tool had established reliability and clinical, behavioural and criminal validity.

Child Behavior Checklist (CBCL)

The CBCL is a standardized measure used to assess children's behavioural and emotional problems, and social competence with high reliability and validity (Achenbach & Ruffle, 2000; Granic, O'Hara, Pepler, & Lewis, 2007). Furthermore, there are three parallel forms of the CBCL; the child's parents or caregiver completes one form (CBCL); the child's teacher completes a second form (Teacher Report Form: TRF); the youth themselves completes a third one (Youth Self-Report: YSR). Moreover, the tool has two versions; one for children from 2-3 years; and another for children from 4-18 years. It takes the informants from 10-20 minutes to complete the behavioural and emotional sections, and the optional competencies sections of the CBCL tool (Achenbach & Ruffle, 2000).

The release of the first version of the CBCL tool was in 1983. Two revised versions followed it in 1991 and 2001 (Bordin et al., 2013). Changes were made regarding the content and number of items used to construct the tool's scales (Bordin et al., 2013; Koegl, 2011; Koegl et al., 2008). As for the 2001 version of the tool for children from 4-18, the behaviour problem section of the three forms (CBCL, TRF, and YSR) consists of 118 items scored on a three-point Likert scale (0 = not true, 1= somewhat or sometimes true, 2 = very true or often true). This section provides scores for "eight narrow-band" syndrome scales namely:

1.Withdrawn/Depressed

2. Somatic Complaints

3. Anxious/Depressed

4. Social Problems

5. Thought Problems

6. Attention Problems

7. Rule-Breaking Behavior

8. Aggressive Behavior.

These syndrome scales are grouped into “three broad-band” scales, namely; internalizing behaviour problems, externalizing behaviour problems, and total behaviour problems. The internalizing behaviour problem scale represents the sum of three subscales being: withdrawn/depressed, somatic complaints, and anxious/depressed. Whereas, the externalizing behaviour problems scale represents the sum of two subscales: rule-breaking behaviour and aggressive behaviour. The total behaviour problem scale represents the sum of externalizing and internalizing behaviour problems, social problems, thought problems and attention problems. The attention problems syndrome scale comprises two subscales: inattention and hyperactivity-impulsivity.

Furthermore, the social competence section of the CBCL and YSR consists of 20 items. It compromises “three narrow-band” scales, namely: child’s activities, social relations, and school functioning, and “one broad-band” scale called total social competence. The child's activities scale measures how much time the child spends on sports, hobbies, games, or performance compared to their peers; how active the child is in the organizations, clubs, teams or groups to which the child belongs; how well the child carries out jobs or chores. The social relations scale measures the number of close friends the child has, the frequency of the child’s meetings with their friends, how well the child gets along with family members and other children, and how independent the child is when playing or working alone. The school functioning scale concerns the child’s problems—

both academic and non-academic—and their performance in academic subjects (As cited in Bordin et al., 2013). The items in the adaptive functioning section of the TRF provide scores for the child's academic performance in different subjects on a scale of one to five in which one indicates that the child's educational performance is far below grade, and five suggests that the child's performance is far above grade. This section also provides scores for four adaptive characteristics on a scale of one to seven. The adaptive characteristics scores determine the child's dedication to school work, appropriateness of behaviour in school, ability to learn, and the child's mood state "how happy compared to other students of the same age" (As cited in Bordin et al., 2013; Lipman et al., 2008).

Furthermore, the 2001 version of the tool for children from 4-18 added six DSM-oriented scales consistent with the DSM diagnostic categories from the Diagnostic and Statistical Manual of Mental Disorders (DSM). These DSM-oriented scales are as follows:

1. Affective Problems
2. Anxiety Problems
3. Somatic Problems
4. Attention-Deficit/Hyperactivity Problems, ADHD
5. Oppositional Defiant Problems
6. Conduct Problems (As cited in Bordin et al., 2013).

The informants of the CBCL tools report on the child's behaviours for the preceding six months. The obtained raw scores from the CBCL tools are then transformed into T-scores to compare them with children from the same gender and age. The standardized T-scores determine whether the child is scoring within the non-clinical, borderline, or clinical ranges. The children who score within the clinical range have severe emotional/behavioural problems and low social competence

(As cited in Bordin et al., 2013; Koegl, 2011). The CBCL tools have been used widely and translated into multiple languages. Furthermore, the CBCL tools have been commonly used to measure changes in children's disruptive behaviours, more specifically, changes in the externalizing behaviour and social relation problems after their participation in the SNAP™ program (Koegl et al., 2008; Walsh et al., 2002).

SNAP™ Research and Program Evaluation Studies

According to the Child Development Institute (2019), the SNAP™ boys and SNAP™ girls intervention programs have gone through some internal evaluations conducted by the program developers, and external evaluations performed by third parties. Some of these studies have been conducted with a primary objective of evaluating the program's effectiveness. Other studies have been conducted to analyze and assess the reliability and validity of the Early Assessment Risk List for Boys (EARL-20B) and Girls (EARL-21G) (Augimeri, 2005; Augimeri et al., 2010). Others have been conducted with a general purpose of measuring neural changes as well as individual differences between improving and non-improving participants of the SNAP™ intervention program (Byrd, Hawes, Burke, Loeber, & Pardini, 2018b; Granic et al., 2007; Levene, Walsh, Augimeri, & Pepler, 2004; Lewis et al., 2008; Woltering, Granic, Lamm, & Lewis, 2011; Woltering, Liao, Liu, & Granic, 2015; Woltering, Lishak, Hodgson, Granic, & Zelazo, 2016). Finally, one study was conducted to measure the SNAP™ program's monetary benefits and cost (Farrington & Koegl, 2015).

A number of the primary studies that were conducted to evaluate SNAP™ program were randomized controlled trials (RCTs) (Augimeri et al., 2007; Burke & Loeber, 2015, 2016; Day & Hrynkiw-Augimeri, 1996), and others used quasi-experimental designs (Koegl et al., 2008; Pepler et al., 2010). The program was also evaluated by using controlled before and after study designs

(Lipman et al., 2008), and uncontrolled pre-post observational study designs (Augimeri, 2005; Augimeri et al., 2006; Augimeri et al., 2010; Augimeri, Walsh, Donato, Blackman, & Piquero, 2018; Day, 1998, 2003; Day & Hunt, 1996; Hrynkiw-Augimeri et al., 1993; Koegl, 2011; Lipman et al., 2007; Walsh et al., 2002). Furthermore, there are other evaluation studies on the SNAP™ program that are conducted by using a mixed research design (Lipman et al., 2011).

Overall, SNAP™ evaluation studies have shown the programs' effectiveness in reducing aggressive, conduct, rule-breaking, antisocial, and total behavioural problems for the intervention groups. There were also significant improvements in parent-child relationships and parenting management skills. Some studies, moreover, claimed that children who received an intensive version of the SNAP™ program had maintained treatment gain up to 15-month of follow-up (Augimeri et al., 2007; Augimeri et al., 2006). There was also a significant association between the number of SNAP™ children group and family counselling sessions and reductions in the children's delinquency and minor regression (Koegl et al., 2008). Furthermore, the program has a significant positive effect on preventing targeted children from committing crimes (Koegl et al., 2008). However, there are some differences and limitations in these studies, as mentioned in the introduction of the study. In summary, the study contributed to our existing body of the SNAP™ program knowledge as it represented the first Meta-Analysis conducted to evaluate the effectiveness of the SNAP™ program.

RESEARCH QUESTIONS

Q1. What is the effect of the SNAP™ program on the level and trend of children's conduct disorder in community settings?

H1. Children participating in the SNAP™ program show a significant decrease in externalizing and internalizing behavioural problems as well as a significant improvement in prosocial skills.

Q2. Do variations in treatment intensity and fidelity affect the program's outcomes?

H2. Children who received intensive SNAP™ services would have fewer youth criminal convictions and more treatment gain at follow-up compared with those who received minimal core treatment components.

METHODS

Research Design and Procedure

The study aimed to determine the effect of the SNAP™ intervention program for children under 12 years of age with conduct disorder in community settings by combining the primary results of individual SNAP™ evaluation studies and performing a Meta-Analysis.

Criteria for considering studies for the Meta-Analysis and search methods for identification of studies:

The Child Development Institute (CDI), as a developer of the SNAP™ Program, requires that all organizations and professionals interested in implementing the program sign a SNAP licensing agreement with the institute. The SNAP™ Affiliate organizations undergo "… a multi-year process that includes assessing site readiness, ongoing training and consultation and a regulated quality assurance process." (Child Development Institute, 2016b) to ensure the treatment fidelity and integrity. Furthermore, the CDI keeps up on all SNAP™ research and program evaluation studies since its inception in 1985.

Accordingly, for the Meta-Analysis, the researcher determined the independent primary SNAP™ evaluation studies from a SNAP™ Research and Program Evaluation Studies Summary Chart prepared by the CDI (2019). The chart is continuously updated and has SNAP™ research and program evaluation studies (published and unpublished) up to April 2019. According to this chart, there were 28 research and evaluation studies on the SNAP™. However, only studies that met the

following inclusion criteria were selected and included in the Meta-Analysis:

1. Original studies that had directly evaluated the effectiveness of SNAP™ intervention program
2. The study's design was either RCT, quasi-experimental, controlled before and after or uncontrolled, pre-post designs.
3. Studies that reported summary statistics from either of the following measures:
    a. Child Behavior Checklist (CBCL)
    b. Teacher's Report Form (TRF)
    c. Criminal record search.
4. Primary data were available for extraction.

The researcher read the full text of all SNAP™ research and evaluation studies included in the CDI summary chart. At this stage, the researcher excluded SNAP™ research studies that did not intend to evaluate the program's effectiveness. Only quantitative studies were included and analyzed. Furthermore, the researcher excluded studies that are based on other SNAP™ evaluation studies. Any duplicate publication was identified and excluded at this stage to avoid overestimating the program effect. The study further used informal channels of communication to find other published/unpublished internal evaluation reports on the SNAP™ program. To be included in the Meta-Analysis, these internal reports had to be final reports and submitted by the same date at the last version of the CDI Summary Chart that was last updated in April 2019.

Types of participants:

The participants of SNAP™ intervention programs are children aged 6 to 12 years, from both genders and their parents. The participating children had been referred to the program because they had a T-score of 70 or higher on the externalizing scale of the CBCL or had police contact within

six months of their referral to the program. Children with a borderline T-score on the externalizing scale of the CBCL might still get access to the program depending on sites' flexibility and vacancies.

Types of interventions:

The intervention groups in SNAP™ program evaluation studies had received either the original version of SNAP™ ORP for boys and girls, SNAP™ program for boys, or SNAP™ Girls Connection program, or intensive/enhanced version of SNAP™ ORP, as described earlier. The control group, on the other hand, were children in the waiting-list or who had received a non-clinical recreation program before receiving the targeted SNAP™ program or received less intensive version of the program. The current study, however, has only included information about the control groups in the randomized controlled trials. Pre-, post- data were collected for only the intervention groups in the observational studies.

Types of outcome measures:

Most of the outcome measures used in SNAP™ program evaluation studies were standardized tools administered to parents, teachers, and youth. The following list represents the order from the most to the least commonly used measures in the potential SNAP™ evaluation research studies:

1. Child Behavior Checklist (CBCL)
2. Teacher's Report Form (TRF)
3. Early Assessment Risk Lists (EARLs); EARL-20B and EARL-21G
4. National Longitudinal Survey of Children and Youth (NLSCY)
5. Self-report Antisocial Behavior Scale (SRA-C)
6. Parent Report of Child's Antisocial Activity (SRA-P).
7. Parenting Dimensions Inventory (PDI)

8. Perceived Ineffectiveness Index (PII)
9. Parenting Stress Index (PSI)
10. Parental Depression (BDI)
11. Family Functioning (FACES-II)
12. Self-Control Scale (A self-report measure)
13. SSIS Self-Control (SSIS: Social Skills Improvement System), completed by parents.

Furthermore, the standardized measures used for SNAP^{TM} program evaluation studies have gone through several revisions and modifications, as described earlier. As a result, SNAP^{TM} program evaluation studies have used different versions of these measures across the studies' periods. Moreover, some studies have used modified versions of the tools' standardized subscales (Koegl et al., 2008; Walsh et al., 2002).

Primary outcomes:

SNAP^{TM} program primarily targets reductions in disruptive behaviours represented in externalizing behaviour and social relation problems and improvements in the social competences. To measure the effectiveness of the SNAP^{TM} program, the primary outcomes of the Meta-Analysis included all ***externalizing behaviours, delinquency, conduct problems, attention deficit hyperactivity disorder (ADHD), Oppositional Defiant Disorder (ODD), attention problems, social problems, social competence and justice system involvement variables.*** *The externalizing behaviour problem*s, as measured by the CBCL tools (i.e., CBCL, TRF), represent the sum of two subscales: *rule-breaking behaviour* and *aggressive behaviour*. *The adaptive functioning scale* on the TRF and the *total competence scales* on the CBCL measure children's *social competence* (Lipman et al., 2007). The delinquency subscale on the CBCL tool measures the delinquency variable. Participating children also report on the delinquent behaviour and attitudes subscales on

the SRA-C (Day, 2003; Day & Hrynkiw-Augimeri, 1996). The attention problems, ODD and ADHD, are measured by the CBCL tools (i.e., CBCL, TRF). The justice system involvements or criminal convictions are identified by the official criminal records by the end of the study or reported by the children's parents/guardians.

Secondary outcomes:

The secondary outcomes of the Meta-Analysis included *internalizing behaviour problems, total behaviour problems* and *parenting skills.* CBCL tools measure internalizing behaviour problems. The internalizing behaviour scale of the CBCL and TRF tools consists of three subscales; *withdrawn/depressed, somatic complaints and anxious/depressed*. The CBCL and TRF tools measure the total behaviour problem that represents the sum of externalizing and internalizing behaviour problems, social problems, thought problems and attention problems. NLSCY and PDI measure parenting skills.

Furthermore, all primary and secondary outcome variables (dependent variables) of the primary studies were measured and reported as continuous summary measurements except for the justice system involvement/ criminal conviction that was reported as a dichotomous variable.

Data collection and analysis

The study-level summary measures data, e.g., pre/post-intervention means, standard deviations and sample size from each study that met the inclusion criteria were extracted and analyzed by using the Review Manager software (RevMan v5.3.5, The Nordic Cochrane Center, The Cochrane Collaboration, Copenhagen, Denmark) to answer the study's research questions. Furthermore, the researcher used Microsoft Excel Worksheet to provide descriptive statistics about the included SNAP^{TM} evaluation studies.

Assessment of risk of bias in included studies

RevMan 5.3 program facilitated the construction of standard tables of characteristics and risk of bias of the included studies. The characteristics of the included studies table provided information about the methods, participants, interventions, outcomes, and other notes. Whereas the risk of bias table for RCTs consisted of six entries regarding selection, performance, detection, attrition and reporting biases. The researcher made decisions for each included RCT regarding random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective reporting. The table also has the seventh entry for any other noticed source of bias (Cochrane Community, 2019). Each bias entry or domain was provided with three response options; low risk, unclear risk or high risk. The overall risk of bias judgment is represented in the worst judgement across all the seventh entries. (Cochrane Community, 2019).

The RevMan 5.3 risk of bias assessment table primarily assesses the risk of bias for the RCTs; therefore, the researcher used another tool to evaluate and determine the risk of bias for the other types of study designs (Stroup et al., 2000). Specifically, the researcher used the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool (Sterne, 2016). ROBINS-I tool has the following seven bias domains:

1. Bias due to confounding
2. Bias in the selection of participants into the study
3. Bias in the classification of interventions
4. Bias due to departures from intended interventions
5. Bias due to missing data
6. Bias in measurement of outcomes

7. Bias in the selection of the reported result

Same as with the RCTs, the assessment process includes providing a judgment for the risk of bias in each of the seventh bias domains. The overall risk of bias judgment is represented in the worst judgement across all the seventh domains. However, ROBINS-I provides five response options to give a judgment about each bias domain. The response of "low risk of bias" option is given to studies that were comparable to well-performed RCTs; "moderate risk of bias" response option was granted to studies that had thorough evidence for non-randomized studies, but were not equivalent to well-performed RCTs; "serious risk of bias" option was given to studies that had some serious problems in the specified domain; "critical risk of bias" choice was for studies with no "useful evidence" and "too problematic" to add them in the analysis; finally, the "no information" option is given when there is no information to make a judgment about the specific bias domain (Sterne, 2016, p. 18).

Moreover, the risk of bias assessment for the current study was done at the study level, i.e., making a single judgment for the risk of bias for all the study's outcomes.

Measures of treatment effect:

The Meta-Analysis was conducted to determine the effect of the SNAP™ intervention program on the level and trend of children's conduct disorder in community settings (Question 1). The researcher calculated the effect size by using the standardized mean difference (SMD or Cohen's d) and its 95% confidence intervals (CI) for two reasons. First, except for the justice system involvement/ criminal conviction variable, all the primary and secondary outcome variables are measured on a continuous scale. Second, the number and content of the constructed scales and subscales of the outcome measures used in the potential studies have undergone some modifications during the studies' periods. Furthermore, Cohen's d statistic was converted to

Hedges' g statistic to avoid overestimating the effect sizes for studies with small sample sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009). For the justice system involvement/ criminal conviction variable, which was measured as a dichotomous variable, the odds ratio was calculated to measure the treatment effect.

According to Borenstein et al. (2009), the following formulas were used to calculate the SMD or Cohen's d and Hedges' g statistics:

1. For studies with independent groups

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_{within}}$$

where d is the Cohen's d; $\bar{X}_1$ and $\bar{X}_2$ are the two group' sample means; S_{within} is the pooled standard deviation.

$$s_{within} = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Where n_1 and n_2 are the two groups' sample sizes and S_2 are the two groups' standard deviations. Also, the formulas used to calculate the variance and standard error of d are as follows:

$$V_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

and

$$SE_d = \sqrt{V_d}$$

where V_d and SE_d are the variance and standard error of d respectively.

Furthermore, the formulas used to calculate Hedges' g statistics and its variance and standard error are as follows:

$$g = J \times d$$

where g is the Hedges' g; J is a correction factor calculated as follows:

$$J = 1 - \frac{3}{4df - 1}$$

where df is the degree of freedom calculated as $n_1 + n_2 - 2$ and Hedges' g statistics' variance and standard error are calculated as follows:

$$V_d = J^2 \times V_d$$

and

$$SE_g = \sqrt{V_g}$$

For uncontrolled studies (studies that used a pre-test, post-test scores):

The same formulas were used except for the pooled standard deviation and its variance, which are calculated as follows:

$$S_{within} = \frac{S_{diff}}{\sqrt{2(1-r)}}$$

where S_{diff} is the standard deviation of the difference, which was obtained from the authors or calculated using the following formula:

$$S_{diff} = \sqrt{S_1^2 + S_2^2 - 2 \times r \times S_1 \times S_2}$$

where r is the correlation between the pre-test and post-test scores, which was obtained from the authors or estimates from other related SNAP™ studies.

Data synthesis:

The researcher performed a random-effect Meta-Analysis because SNAP™ evaluation studies had different methodological designs, i.e., RCT, quasi-experimental, controlled before and after and uncontrolled, pre-post observational studies, besides; these kinds of studies met the inclusion

criteria of the research. According to the random-effect model, the real effects in SNAP^{TM} evaluation studies were assumed to be drawn from a distribution of true effects. The estimated SNAP^{TM} effect size represented the mean of that distribution. In addition to the within-study sampling error, the random-effect model took account of the between-studies variance (Borenstein et al., 2009).

Dealing with missing data:

The information needed to estimate the magnitude and trend of the effect size of the SNAP^{TM} program depended on the study design. For example, the mean, standard deviation (SD), and sample size information were required to calculate the effect size for studies with independent groups, like in the RCT. Furthermore, information regarding mean, standard deviation of the difference (S_{diff}) and pretest-posttest correlation (r), or standard deviation within pretest-posttest groups are needed to calculate the effect size for studies that used pre-post scores, like in the observational studies. One of the challenges faced in performing meta-analyses is missing data. The following represents the recommended procedures that the researcher followed in dealing with the missing data:

If a study did not provide these kinds of information, and it was not possible to derive them from other presented information such as; t statistic, F statistic, or chi-square, the researcher reached out to the corresponding author in this regard. In case there was no response received, the researcher estimated these missing data from related studies (Borenstein et al., 2009). Lastly, the researcher excluded the study or the specific outcome with the missing information if none of the mentioned strategies worked.

The information that was missing in the included studies was some of the standard deviations of mean scores and standard deviations of the difference (S_{diff}) and pretest-posttest correlations (r).

The researcher was able to obtain all the missing standard deviations of mean scores and some of the standard deviations of the difference (S_{diff}) and pretest-posttest correlations (r) from the corresponding authors. Furthermore, for studies that did not report the pretest-posttest correlations, the researcher used the pretest-posttest correlations obtained from the authors of the other included SNAP^{TM} evaluation studies.

Assessment of heterogeneity:

The researcher performed the Q test to determine whether all studies shared a common effect size and that the heterogeneity in effect sizes was zero. The Q statistic is a standardized measure representing the total observed variance in the effect size from study to study, i.e., the observed weighted sum of squares (WSS), and it was calculated according to Borenstein et al. (2009) as follows:

$$Q=\sum_{i=1}^{k}\left(\frac{Y_i-M}{S_i}\right)^2$$

where Y_i is the effect size for study i; M is the pooled/summary effect; S_i is the variance of study i; and k is the number of studies.

Furthermore, the degree of freedom, df, of the Q statistic represents "… the expected WSS (under the assumption that all studies share a common effect),…" (Borenstein et al., 2009, p. 110). The difference between these two quantities yields the excess variation attributed to the real difference in the true effect size, i.e. the real heterogeneity. The distribution of the Q statistic is the same as the chi-square statistic with df equals to the number of the analyzed studies minus one (k-1). A significant Q test leads to rejecting the null hypothesis and concludes that the studies do not share a common effect size (Borenstein et al., 2009). The Q statistic and its p-value, however, serve only as a test of significance and not as an estimate of the amount of the true dispersion, especially with

small studies or a small number of studies. Accordingly, the I^2 statistic was conducted to determine what proportion of the observed variance is real, i.e., the true heterogeneity magnitude. The I^2 statistic is an absolute measure representing the ratio of the excess dispersion to the total dispersion and is calculated (Borenstein et al., 2009) as follows:

$$I^2 = \left(\frac{Q - df}{Q} \right) \times 100\%$$

This measure reflects the inconsistency between studies' results as it represents "the extent of overlap of the confidence intervals" of the studies' effect sizes (Borenstein et al., 2009, p. 118). The magnitude of the I^2 scale ranges from 0-100%. When I^2 moves toward 0, that means the observed variance is due to sampling error within studies; however, when I^2 moves away from 0, that means there is a real variance between studies. Higgins and colleagues (2003) projected tentative thresholds for the interpretation of I^2 values like 25%, 50% and 75% representing low, moderate and high heterogeneity, respectively (As cited in Borenstein et al., 2009).

Moreover, the researcher planned to perform a meta-regression to explain further the variations in the included studies with study-level covariates. However, the regression was not performed as there was not enough power, i.e., at least ten studies for each covariate (Borenstein et al., 2009; Deeks, Higgins, & Altman, 2011).

Subgroup analysis:

To explain the real variance in the effect size, the researcher planned to conduct subgroup analyses on pre-defined, study-level characteristics of the included studies. These characteristics represent the study-level confounding variables (SLCVs) needed to assess the differential effects of the SNAP^{TM} program. Based on SNAP^{TM} literature, the chosen confounding variables included the year of publication and enrollment, publication status and type, study location, study design, participant gender, comparisons, outcome measures and informants, evaluation agent/type, risk of

bias, and treatment intensity and fidelity variables. The subgroup analyses were also needed to determine whether variations in treatment intensity and fidelity affected the program's outcomes (Question 2).

Sensitivity analysis:

Sensitivity analysis is a technique used to assess the robustness of the study's results. Sensitivity analyses performed for Meta-Analysis studies are the same as those conducted for the primary studies (Borenstein et al., 2009). For the current study, the sensitivity analysis was conducted to determine the impact of removing one study at a time from the analysis on the overall effect size.

Assessment of publication bias:

The impact of publication bias on the magnitude and trend of the effect size is addressed through two methods: a funnel plot and a Trim and Fill procedure. The funnel plot is a graph of the effect size against the sample size or variance. It is a subjective method for detecting publication bias as the decision whether there is evidence of bias depends on the visual judgment on the symmetric distribution of the studies around the mean effect size in the shape of the funnel. The top of the funnel is where the large studies locate and cluster around the mean effect size. The small studies, on the other hand, spread widely at the bottom of the funnel. The symmetric shape of the funnel plot means there is no evidence of publication bias. Asymmetry in the shape of the funnel means that there is evidence of publication bias, and hence, the researcher would perform the Trim and Fill procedure. The Trim and Fill procedure aims to produce an unbiased effect size. During the Trim and Fill procedure, small studies with extreme effects would be trimmed; missing studies would be imputed; the effect size would be re-calculated, and the funnel plot would be examined for asymmetry. This process would continue until the symmetry funnel shape forms around the adjusted new effect size (Borenstein et al., 2009). The researcher planned to use the

Comprehensive Meta-Analysis (CMA) software, version 3 (2019), to perform the Trim and Fill procedure.

RESULTS

The purpose of the current study was to conduct a Meta-Analysis that combines data from the independent SNAP™ primary evaluation studies to answer the following questions and support or reject the accompanying hypotheses:

Q1. What is the effect of the SNAP™ program on the level and trend of children's conduct disorder in community settings?

H1. Children participating in the SNAP™ program show a significant decrease in externalizing and internalizing behavioural problems as well as a significant improvement in prosocial skills.

Q2. Do variations in treatment intensity and fidelity affect the program's outcomes?

H2. Children who received intensive SNAP™ services would have fewer youth criminal convictions and more treatment gain at follow-up compared with those who received minimal core treatment components.

The following sections present the study's results, starting with the study selection process, then providing descriptive statistics about the included studies, and finally showing the results of the Meta-Analysis that answer the study's research questions, and support/reject their hypotheses.

Study Selection Process

The study's search process, as mentioned earlier, was limited to the independent primary SNAP™ evaluation studies from a SNAP™ Research and Program Evaluation Studies Summary Chart prepared by the CDI (2019). The researcher further used informal channels of communication to find other published/unpublished final, internal evaluation reports on SNAP™ Program, which were submitted by April 2019. The CDI Summary Chart had 28 research and evaluation studies on the SNAP™ program as of April 2019. The researcher further was able to identify three more

final internal reports on SNAP™ through the other source (Figure 1). Therefore, the total number of records identified was 31, and after removing duplicates (n= 5), there were 26 records left for screening and checking for the study's eligibility criteria. All the duplicates were found in the CDI Summary Chart. These duplicates used the same participants' data that were used in other SNAP™ evaluation studies. In such cases where different studies used the same sample, the researcher chose the original study or the study that had reported the required information for the Meta-Analysis (Table 2).

Table 2 The included studies and their removed duplicates.

Chosen study	**Duplicate**
Augimeri et al. (2006)	Augimeri (2005)
Augimeri et al. (2007)	Day & Hrynkiw-Augimeri (1996); Koegl et al. (2008)
Burke & Loeber (2015)	Burke & Loeber (2016); Byrd, Hawes, Burke, Loeber, & Pardini (2018a)

At the screening stage, the number of studies that were removed was 10 (Byrd et al., 2018b; Farrington & Koegl, 2015; Granic et al., 2007; Koegl, 2011; Levene et al., 2004; Lewis et al., 2008; Lipman et al., 2011; Woltering et al., 2011; Woltering et al., 2015; Woltering et al., 2016). These studies did not meet the Meta-Analysis inclusion criteria. They had neither evaluated the SNAP™ effectiveness nor reported any of the Meta-Analysis primary or secondary outcomes. The total number of SNAP™ studies that met the Meta-Analysis inclusion criteria was 16. However, one study, Day (2003), was removed because it had critical missing data needed to perform the Meta-Analysis. Specifically, that study was missing the pretest-posttest correlation of parenting skill outcomes that no one of the included studies have provided. Therefore, the final total number of SNAP™ evaluation studies that were included in the Meta-Analysis was 15 (Figure 1).

- PRISMA flow diagram of study selection (Moher, 2009).

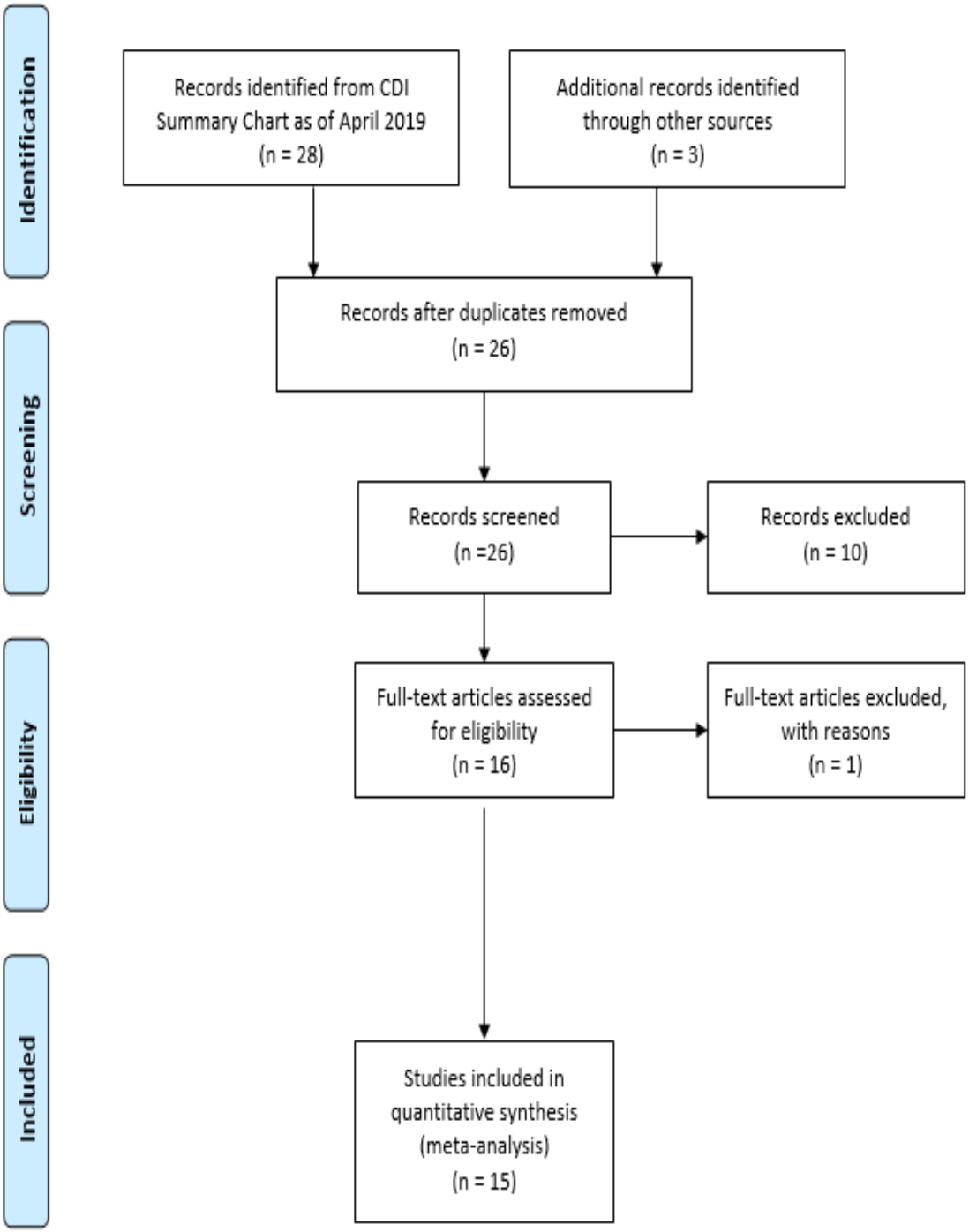


Figure 1 The selection process for including SNAP™ evaluation studies in the Meta-Analysis.

Descriptive Statistics About the Included Studies

1. The included studies' sample sizes, participants' gender and parental Marital Status

As displayed in Table 3, 2204 children had participated in the included SNAP™ studies. Most of them were males (N= 1527, 69%), whereas the females represented only 31% (N= 677) of the total sample. The participants' mean age was 9.2 (SD= 1.4). The study with the largest sample size (N= 343 children) was Augimeri et al. (2010); on the other hand, Augimeri et al. (2007) had the smallest sample size (N= 16 children). Nine studies reported summary data for both genders; three studies reported summary data for males only and three for females only.

Table 3 The included studies' sample sizes and participants' gender.

Study	Sample size	Mean age (SD)	Gender	
			Male	Female
Augimeri et al. (2018), C10	318	-	169	149
Burke & Loeber (2015), B4	130	8.9 (1.9)	130	NA
Smith-Moncrieffe (2015a), QR	89	8.8 (2.1)	56	33
Smith-Moncrieffe (2015b), TR	90	8.4 (1.5)	67	23
Smith-Moncrieffe (2015c), ER	85	9.5 (1.5)	60	25
Augimeri et al. (2010), C8	343	8.8 (1.6)	195	148
Pepler et al. (2010), A4	45	-	NA	45
Lipman et al. (2008), B1b	223	9.8 (1.7)	223	NA
Augimeri et al. (2007), A2	16	8.7 (1.4)	12	4
Lipman et al. (2007), B2	96	9.0 (1.7)	NA	96
Augimeri et al. (2006), C3	319	9.6 (1.4)	319	NA
Walsh et al. (2002), C9	98	8.9	NA	98
Day (1998), C5	203	9.9 (1.3)	173	30
Day & Hunt (1996), C4	85	9.1 (1.4)	69	16
Hrynkiw-Augimeri et al. (1993), C1	64	9.6	54	10
Total	**2204**	**9.2 (1.4)**	**1527**	**677**

SD: Standard Deviation; NA: Not Applicable

Furthermore, 39% of the participants' parents were single, and 38% were married or living with common-law partners. The percentage of parents that were separated, divorced or widowed was 18%; besides, 14% of the parents' marital status was reported as other (Table 4).

Table 4 Parental marital status.

Study	Parental Marital Status			
	Married/Common-Law % (n/N)	**Separated/Divorced/Widowed % (n/N)**	**Single % (n/N)**	**Other % (n/N)**
Augimeri et al. (2018), C10	-	-	-	-
Burke & Loeber (2015), B4	-	-	-	-
Smith-Moncrieffe (2015a), QR	82% (69/84)	11% (9/84)	7% (6/84)	0% (0/84)
Smith-Moncrieffe (2015b), TR	72% (51/71)	8% (6/71)	20% (14/71)	0% (0/71)
Smith-Moncrieffe (2015c), ER	26% (21/82)	26% (21/82)	48% (39/82)	1% (1/82)
Augimeri et al. (2010), C8	24% (82/343)	1% (4/343)	48% (164/343)	27% (93/343)
Pepler et al. (2010), A4	-	-	69% (55/80)	-
Lipman et al. (2008), B1b	28% (62/218)	33% (72/218)	24% (53/218)	14% (31/218)
Augimeri et al. (2007), A2	56% (9/16)	25% (4/16)	19% (3/16)	-
Lipman et al. (2007), B2	39% (37/96)	32% (31/96)	19% (18/96)	10% (10/96)
Augimeri et al. (2006), C3	-	-	48% (153/319)	-
Walsh et al. (2002), C9	35% (34/98)	-	54% (53/98)	11% (11/98)
Day (1998), C5	54% (91/168)	32% (54/168)	14% (23/168)	-
Day & Hunt (1996), C4	-	-	60% (51/85)	-
Hrynkiw-Augimeri et al. (1993), C1	25% (16/64)	16% (10/64)	55% (35/64)	5% (3/64)
Total	**38% (472/1240)**	**18% (211/1142)**	**39% (667/1724)**	**14% (149/1056)**

(%) Percentage; n: number of participants with an event; N: Study's Sample size.

2. Participants' baseline CBCL T-scores

As shown in Table 5, nine studies had reported some of the children's baseline CBCL T-scores. The baseline CBCL T-scores for the total sample of the included studies were in the clinical range in the aggression, conduct problems, externalizing problems and rule-breaking (73.3, 72.8, 70.7, and 69.0) respectively. The total sample had a borderline T-score (63.1) in the CBCL internalizing problems.

The highest T-score for externalizing problems was (75.3), reported in Burke & Loeber (2015), whereas the lowest (65.9) was reported in Smith-Moncrieffe (2015c). For internalizing problems, the highest T-score (66.66) was reported in Hrynkiw-Augimeri et al. (1993), and the lowest (59.6) was reported in Smith-Moncrieffe (2015b). Furthermore, the highest aggression T-score (79.4) was reported in Lipman et al. (2008), and the lowest (66.6) was reported in Smith-Moncrieffe (2015c). For rule-breaking, the highest T-score (72.9) was reported in Lipman et al. (2008), whereas the lowest T-score (64.9) was reported in Smith-Moncrieffe (2015c). Last, for the CBCL conduct problems, the highest T-score (77.2) was reported in Lipman et al. (2008), and the lowest T-score (67.2) was reported in Smith-Moncrieffe (2015c).

Accordingly, participants in the Lipman et al. (2008) study had the highest baseline T-score in the aggression, rule-breaking and conduct problems; on the other hand, participants in the Smith-Moncrieffe (2015c) study had the lowest baseline T-score in the externalizing problems, aggression, rule-breaking and conduct problems.

3. Study-level confounding variables (SLCVs)

The preplanned study-level confounding variables of the Meta-Analysis, i.e., Publication Year, Type, and Status, Evaluation Agent/Type, Study Design and Location, Participant Gender, Comparison Groups, Year of Enrollment, Outcome Measures and Informants; are displayed in

Tables 6, 7 and 8. All the included SNAP™ evaluation studies were submitted/published throughout the last three decades, specifically between 1993-2018. Almost half of them, 47%, were submitted/published between 2001-2010, about 33% were between 2011-2020, and 20% were between 1991-2000. About 60% of the included studies were published in journals; namely, Journal of Criminal Justice, Prevention Science, Journal of the Canadian Academy of Child and Adolescent Psychiatry, Journal of Child and Family Studies, Canadian Journal of Counselling, Canadian Journal of Criminology, Journal of Emotional and Behavioral Disorders and Canada's Mental Health. The rest of the included studies, 40%, were final evaluation reports or program evaluation research reports conducted by or submitted to different authorities; such as Public Safety Canada, The Provincial Centre of Excellence for Child and Youth Mental Health at Children's Hospital of Eastern Ontario (CHEO) and Offord Centre for Child Studies. Almost all the included studies, 93%, were available online, i.e., like journal articles or identified in the grey literature, except for one study, Lipman et al. (2007), was obtained from the CDI.

External/third parties conducted about 40% of the included SNAP™ evaluation studies, and 20% were performed by an internal evaluation team at the CDI. The remaining 40% were shared evaluation studies conducted by internal and external evaluation research teams. Regarding the study design, only two of the included studies, 13%, were RCTs, one, 7%, was quasi-experimental, 3, 20%, were controlled before and after, and nine studies, 60%, were uncontrolled pre-post evaluation studies. Furthermore, all of the included studies where conducted in Canada, except for one research, Burke and Loeber (2015) was conducted in the USA. More than half of the Canadian SNAP™ evaluation studies, 67%, were held in Toronto, Ontario, and 20% were conducted in Hamilton, Ontario. There was only one study conducted in Cree Nation Communities (Mistissini and Waswanipi), Quebec, and another in Edmonton, Alberta, Canada.

The participating children were enrolled in the program from its inception in 1985 until 2017. About 53% of the children were enrolled between 2001-2017, 40% were enrolled between 1985-2000, and one study, Burke and Loeber (2015), did not report the children's enrollment date. The comparison groups in the RCTs, quasi-experimental and the controlled before and after studies were children on the waitlists (27%) or children who received standard services, "non-clinical recreation" program (13%).

For the outcome measures used in the included studies, there were 13 studies, 87%, that had used the CBCL tool; these studies, however, had used different versions of the CBCL; specifically, the original CBCL version in 1983 and the updated versions in 1991 and 2001. Furthermore, a modified measure tool of the CBCL called Standardized Client Information System (SCIS) was used in Walsh et al. (2002) study. The outcome informants of both CBCL and SCIS were only parents. The TRF tool was used in seven of the included studies, 47%. Moreover, the Official Criminal Records (OCRs) were used in 4 included studies, 27%, and one study, Hrynkiw-Augimeri et al. (1993), used parents' reports on the children police contact after one year of follow-up.

The National Longitudinal Survey of Children and Youth (NLSCY) tool, which was used to measure the treatment effectiveness in regarding parenting skills (consistency, rational parenting, ineffective parenting, positive interaction), was only used in one of the included studies; specifically, Pepler et al. (2010).

For the outcome informants, 4 of the included studies reported results from only the children's parents/caregivers, 27%. Moreover, seven studies, 47%, were based on two informants; specifically, parents/caregivers and teachers, and 20%, three studies had results based on parents'/caregivers' reports and OCRs. Only one study, Day (1998), reported results found solely

on the OCRs.

4. SNAP™ Sessions

The two required components of the SNAP™ program are the Child SNAP™ Group Session and the Parent SNAP™ Group Session. Seven studies out of 15 have reported the average attendance to these two core components. As shown in Table 9, the average attendance to the child group session in these included studies was 8.8 (SD= 2.9). For the parents' sessions, the average attendance was 6.4 (SD=3.3). The highest attendance rates were in Smith-Moncrieffe (2015c) and (2015b) studies, followed by Augimeri et al. (2006), (2007) and Lipman et al. (2008). On average, the Children in Burke and Loeber (2015), and Smith-Moncrieffe (2015a) studies did not graduate from the program as their average mean attendance to child SNAP™ group session was less than eight sessions, i. e., the number of required sessions for graduation.

For the other additional/optional SNAP™ components, only two studies reported the average number of sessions/services received for the individual child befriending and family counselling components (Table 10). As shown in Table 10, on average, the participating children in Burke and Loeber (2015) and Augimeri et al. (2007) attended 4 (SD= 4.1) and 3.3 (SD= 6.3) of the individual child befriending and family counselling, respectively.

5. Treatment Fidelity

About half of the included SNAP™ evaluation studies reported the treatment fidelity (Table 11). The highest fidelity, 96.5%, was reached in Lipman et al. (2008) study, followed by Burke and Loeber's (2015) study that achieved 92% of the treatment fidelity. The treatment fidelity for Smith-Moncrieffe (2015c) and (2015b), Pepler et al. (2010) and Augimeri et al. (2007) studies ranged from 88.2% to 81.6%. The lowest fidelity, 53.4%, was reported in Smith-Moncrieffe (2015a).

Table 5 Participants' Baseline CBCL T-Scores (Externalizing, Internalizing, Aggression, Rule-Breaking and Conduct Problems).

Scale	CBCL-Externalizing Problems	CBCL-Internalizing Problems	CBCL-Aggression	CBCL-Rule-Breaking	CBCL-Conduct Problems
	Mean (SD)	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**
Burke & Loeber (2015), B4	75.3 (4.9)	63.4 (9.2)	79.1 (9.6)	72.4 (6.1)	76.0 (6.8)
Smith-Moncrieffe (2015b), TR	69.9 (7.0)	59.6 (9.5)	70.9 (8.9)	66.9 (7.6)	70.7 (7.3)
Smith-Moncrieffe (2015c), ER	65.9 (9.2)	60.8 (10.5)	66.6 (11.1)	64.9 (8.5)	67.2 (9.5)
Pepler et al. (2010), A4	71.4 (6.7)	65.4 (10.2)	72.9 (10.1)	67.8 (6.6)	72.4 (6.6)
Lipman et al. (2008), B1b	-	-	79.4 (10.3)	72.9 (6.9)	77.2 (8.0)
Smith-Moncrieffe (2015a), QR	67.1 (10.8)	60.9 (10.8)	67.6 (10.3)	66.8 (9.1)	70.0 (9.0)
Lipman et al. (2007), B2	75.1 (5.9)	64.6 (9.8)	78.7 (9.6)	71.0 (6.0)	76.0 (6.8)
Day & Hunt (1996), C4	69.0 (10.0)	63.5 (9.2)	71.5 (11.4)	-	-
Hrynkiw-Augimeri et al. (1993), C1	71.95	66.66	-	-	-
Total	**70.7 (7.5)**	**63.1 (9.3)**	**73.3 (10.1)**	**69.0 (7.1)**	**72.8 (7.7)**

SD: Standard Deviation; (-) Not reported; NA: Not Applicable.

Table 6 Study-level confounding variables: (Publication Year, Type, Status, Evaluation Agent/Type, Study Design and Location).

Study	Publication Year	Publication Type	Publication Status	Evaluation Agent/Type	Study Design	Study Location
Augimeri et al. (2018), C10	2018	Journal	Published	Shared	Retrospective uncontrolled pre-post	Toronto, ON, CA
Burke & Loeber (2015), B4	2014	Journal	Published	External	RCT	Pittsburgh, PA, USA
Smith-Moncrieffe (2015b), TR	2015	Report	Published	External	Prospective Controlled before and after	St. Leonard's Society of Toronto, Ontario, Canada
Smith-Moncrieffe (2015c), ER	2015	Report	Published	External	Prospective Controlled before and after	Edmonton, Alberta, Canada
Smith-Moncrieffe (2015a), QR	2015	Report	Published	External	Prospective Uncontrolled pre-post	Cree Nation Communities (Mistissini and Waswanipi) Quebec, Canada
Pepler et al. (2010), A4	2010	Journal	Published	Internal	Prospective Quasi-Experimental	Toronto, ON, CA
Lipman et al. (2008), B1b	2008	Journal	Published	External	Controlled before and after	Hamilton, ON, CA
Augimeri et al. (2010), C8	2009	Report	Published	Shared	Uncontrolled pre-post	Toronto, ON, CA and Hamilton, ON, CA Only for Criminal Outcome Data
Augimeri et al. (2007), A2	2007	Journal	Published	Internal	RCT	Toronto, ON, CA
Lipman et al. (2007), B2	2007	Report	Unpublished	External	Prospective uncontrolled pre-post	Hamilton, ON, CA
Augimeri et al. (2006), C3	2006	Report	Published	Shared	Retrospective uncontrolled pre-post	Toronto, ON, CA
Walsh et al. (2002), C9	2002	Journal	Published	Shared	Retrospective uncontrolled pre-post	Toronto, ON, CA
Day (1998), C5	1998	Journal	Published	Internal	Uncontrolled pre-post, long-term follow-up	Toronto, ON, CA
Day & Hunt (1996), C4	1996	Journal	Published	Shared	Prospective uncontrolled pre-post	Toronto, ON, CA
Hrynkiw-Augimeri et al. (1993), C1	1993	Journal	Published	Shared	Retrospective uncontrolled pre-post	Toronto, ON, CA

Table 7 Study-level confounding variables: (Participant Gender, Comparison Groups, Year of Enrollment, Outcome Measures and Informants).

Study	Participant Gender	Comparison Groups	Year of Enrollment	Outcome Measures	Outcome Informants
Augimeri et al. (2018), C10	Mixed-gender	SNAP program	2013-2017	CBCL (2001)	Parents
Burke & Loeber (2015), B4	Male Only	SNAP Group	Not Reported	CBCL (2001)	Parents
		SSG		OCRs	OCRs
Smith-Moncrieffe (2015b), TR	Mixed-gender	SNAP program	2010-2014	CBCL, TRF	Parents, Teachers
		Delayed Treatment Group (DTG)			
Smith-Moncrieffe (2015c), ER	Mixed-gender	SNAP program	2010-2014	CBCL, TRF	Parents, Teachers
		Delayed Treatment Group (DTG)			
Smith-Moncrieffe (2015a), QR	Mixed-gender	SNAP program	2010-2014	CBCL, TRF	Parents, Teachers
Pepler et al. (2010), A4	Female Only	SNAP® GC	2002-2004	CBCL (2001), TRF (1991), NLSCY (1999)	Parents, Teachers, Parents
		Waitlist Control Group			
Lipman et al. (2008), B1b	Male Only	ORP Group	2002–2005	CBCL, TRF (2001)	Parents, Teachers
		Waitlist Control Group	2005		
Augimeri et al. (2010), C8	Mixed-gender	SNAP® ORP & GC programs	2001-2009	CBCL (2001)	Parents
				OCRs	OCRs
Augimeri et al. (2007), A2	Mixed-gender	ITG-ORP	1994	CBCL (1991)	Parents
		DTG-CRC	1995	OCRs	OCRs
Lipman et al. (2007), B2	Female Only	SNAP® GC	2004	CBCL, TRF (2001)	Parents, Teachers
Augimeri et al. (2006), C3	Male Only	ORP program	1985-1999	CBCL (Different Versions)	Parents
Walsh et al. (2002), C9	Female Only	SNAP® GC	1996-2000	SCIS (1996)	Parents
Day (1998), C5	Mixed-gender	ORP program	1985-1992	OCRs (1993-1996)	OCRs
Day & Hunt (1996), C4	Mixed-gender	ORP program	1990-1991	CBCL (1983)	Parents
Hrynkiw-Augimeri et al. (1993), C1	Mixed-gender	ORP program	1985-1988	CBCL (1983), TRF (1986)	Parents, Teachers

ITG-ORP: Immediate Treatment Group-Under 12 Outreach Project; **DTG-CRC:** Delayed Treatment Group-Cool Runners Club; **CBCL:** Child Behavior Checklist; **TRF:** Teacher Report Form; **OCRs:** Official Criminal Records; **SNAP GC:** SNAP® Girls Connection; **SSG:** Standard Services Group; **SCIS**: Standardized Client Information System; **NLSCY**: National Longitudinal Survey of Children and Youth.

Table 8 General characteristics of included studies

Publication Year	**n**	**%**	**Participant Gender**	**n**	**%**
2011-2020	5	33%	Male Only	3	20%
2001-2010	7	47%	Female Only	3	20%
1991-2000	3	20%	Mixed gender	9	60%
Publication Type			**Comparison Group**		
Journal	9	60%	Standard Services Group (SSG)	2	13%
Report	6	40%	Waitlist Control Group	4	27%
Publication Status			**Year of Enrollment**		
Published	14	93%	1985-2000	6	40%
Unpublished	1	7%	2001-2017	8	53%
Evaluation Agent/Type			Not Reported	1	7%
External	6	40%	**Outcome Measures**		
Internal	3	20%	CBCL Only	3	20%
Shared	6	40%	CBCL and TRF	7	47%
Study Design			CBCL and OCRs	3	20%
Randomized Controlled Trial	2	13%	OCRs Only	1	7%
Quasi-Experimental	1	7%	SCIS	1	7%
Controlled Before and After	3	20%	NLSCY	1	7%
Uncontrolled Pre-Post	9	60%	**Outcome Informants**		
Study Location			One Informant-Parents	4	27%
Toronto, Ontario, Canada	10	67%	Two Informants-Parents and Teachers	7	47%
Hamilton, Ontario, Canada	3	20%	Two Informants-Parents and OCRs	3	20%
Cree Nation, Quebec, Canada	1	7%	One Informant-OCRs	1	7%
Edmonton, Alberta, Canada	1	7%			
Pittsburgh, Pennsylvania, USA	1	7%			

(%) Percentage; n: number of studies with an event; N: total number of the included studies (15); **CBCL:** Child Behavior Checklist; **TRF:** Teacher Report Form; **OCRs:** Official Criminal Records; **SCIS:** Standardized Client Information System; **NLSCY**: National Longitudinal Survey of Children and Youth.

Table 9 The Average attendance of the child and parents SNAP™ Group Session.

Study	**Child SNAP™ Group Session**	**Parent SNAP™ Group Session**
	Mean (SD)	
Burke & Loeber (2015), B4	6.25 (4.3)	5.02 (4.2)
Smith-Moncrieffe (2015a), QR	6.8 (3.2)	4.6 (3.6)
Smith-Moncrieffe (2015b), TR	10.1 (1.7)	9.2 (3.2)
Smith-Moncrieffe (2015c), ER	11.6 (1.9)	11.2 (2.8)
Lipman et al. (2008), B1b	8.8	7.6
Augimeri et al. (2007), A2	8.95 (2.8)	4.1 (3.8)
Hrynkiw-Augimeri et al. (2006), C3	9.4 (4.0)	3.2 (4.3)
Total	**8.8 (2.9)**	**6.4 (3.3)**

Table 10 The Average attendance of the individual child befriending and family counselling services.

Study	Individual Child Befriending	Family counselling
Burke & Loeber (2015), B4	2.5 (4.2)	4.12 (6.9)
Augimeri et al. (2007), A2	5.5 (4.4)	2.5 (3.6)
Total	**4 (4.1)**	**3.3 (6.3)**

Table 11 SNAP™ treatment fidelity.

Study	% Treatment Fidelity
Burke & Loeber (2015), B4	92%
Smith-Moncrieffe (2015a), QR	53.4%
Smith-Moncrieffe (2015b), TR	81.9%
Smith-Moncrieffe (2015c), ER	88.2%
Pepler et al. (2010), A4	83%
Lipman et al. (2008), B1b	96.5%
Augimeri et al. (2007), A2	81.6%

6. Funding sources

The funding sources of the included SNAP™ evaluation studies are displayed in Table 12. Public Safety Canada had provided funds to three SNAP™ evaluation studies, namely, Smith-Moncrieffe (2015a), (2015b) and (2015c), which represented 20% of the included studies. The Ontario Centre of Excellence for Child and Youth Mental Health at CHEO had provided grants to Augimeri et al. (2010) and (2006) that represented 13% of the included studies. The Trillium Foundation had also offered grants to 13% of the included SNAP™ evaluation studies, specifically to Pepler et al. (2010) and Day (1998). The Department of Health of the Commonwealth of Pennsylvania and the National Institute of Mental Health had provided two grants to Burke & Loeber (2015). The Hospital for Sick Children Foundation had also provided a fund to Pepler et al. (2010). The National Crime Prevention Strategy, Justice Canada and Hamilton Community Foundation had also provided funds to Lipman et al. (2008), Augimeri et al. (2007) and Lipman et al. (2007), respectively. However, there were three included studies, Walsh et al. (2002), Day & Hunt (1996) and Hrynkiw-Augimeri et al. (1993), that did not report their funding sources, which represented

in total 20% of the included studies. Moreover, only one included study, Augimeri et al. (2018), reported that they did not receive any specific funding, which represented 7% of the included studies.

Table 12 The included studies' funding sources

Study	Funding Source
Augimeri et al. (2018), C10	No specific funding received.
Burke & Loeber (2015), B4	The Department of Health of the Commonwealth of Pennsylvania (Grant#: 07-365-01), and the National Institute of Mental Health (Grant#: MH074148).
Smith-Moncrieffe (2015a), QR	Public Safety Canada (Research Report: 2015-R017)
Smith-Moncrieffe (2015b), TR	Public Safety Canada (Research Report: 2015-R017)
Smith-Moncrieffe (2015c), ER	Public Safety Canada (Research Report: 2015-R017)
Augimeri et al. (2010), C8	The Ontario Centre of Excellence for Child and Youth Mental Health at CHEO (Program Evaluation Grant#: RG-976)
Pepler et al. (2010), A4	The Hospital for Sick Children Foundation and the Trillium Foundation.
Lipman et al. (2008), B1b	The National Crime Prevention Strategy.
Augimeri et al. (2007), A2	Justice Canada (file#: 6114-20).
Lipman et al. (2007), B2	Hamilton Community Foundation
Augimeri et al. (2006), C3	The Ontario Centre of Excellence for Child and Youth Mental Health at CHEO (Program Evaluation Grant#: PEG162606-101)
Walsh et al. (2002), C9	No specific funding reported.
Day (1998), C5	The Trillium Foundation (Grant#: Not reported).
Day & Hunt (1996), C4	No specific funding reported.
Hrynkiw-Augimeri et al. (1993), C1	No specific funding reported.

Risk of bias within the included studies

A. Risk of bias in randomized controlled trials

The risk of bias assessment (RoB) for the two included RCTs; Burke and Loeber (2015) and Augimeri et al. (2007), is shown in Tables 13 and 14.

Six domains of the RoB assessment for **Burke and Loeber (2015)** study have been given a low risk of bias judgement for the reasons mentioned in Table 13. However, the study was rated as having a high risk of bias on one domain, specifically the "Blinding of participants and personnel

(performance bias)" domain as the study participants, i.e., children and parents, were aware of the allocated interventions, which might have introduced performance bias.

Table 13 Risk of bias assessment for Burke and Loeber (2015) study.

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"...participants were randomly assigned to study condition." P4.
Allocation concealment (selection bias)	Low risk	"Randomization was performed by the study investigators independently of the treatment providers using a random number generating computer program." P4.
Blinding of participants and personnel (performance bias)	High risk	"...data collection activities for this study was conducted by research project staff independent of SNAP service providers". P5. However, performance bias might have happened due to knowledge of the allocated interventions by participants.
Blinding of outcome assessment (detection bias)	Low risk	"All training and oversight of the interview and data collection activities for this study was conducted by research project staff independent of SNAP service providers." P5.
Incomplete outcome data (attrition bias)	Low risk	"As this evaluation was initiated as an intent-to-treat study, all participants were retained in the analyses after randomization, regardless of their level of service use." P6. Further, the authors stated that they "...used maximum likelihood estimators, which provide advantages in handling missing data..." P6.
Selective reporting (reporting bias)	Low risk	"Fixed effects of SNAP treatment in separate models predicting outcomes at waves 2 through 4" P7, were reported. Post-treatment means and standard deviations are not reported for all the study's outcomes.
Other bias	Low risk	The study appears to be free of other sources of bias.

Since the overall risk of bias is represented in the worst judgement across all the seventh entries, the overall RoB judgment for Burke and Loeber (2015) was that the study had a high risk of bias. The results of the RoB assessment for **Augimeri et al. (2007)** are the same as those for Burke and Loeber (2015) RoB assessment, except that Augimeri et al. (2007) had two domains that were rated as having a high risk of bias (Table 14). One is the "Blinding of participants and personnel (performance bias)" domain, for the same reason as for Burke and Loeber (2015). The other domain was "Incomplete outcome data (attrition bias)" because the study had a higher attrition

rate among “the more serious cases” only in the control group. Therefore, the overall RoB judgment for Augimeri et al. (2007) was that the study had a high risk of bias.

Table 14 Risk of bias assessment for Augimeri et al. (2007) study.

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The original design was to match 32 children, case-by-case in 16 pairs, on age, sex, and severity of delinquency (T-score) on the CBCL. One member of each pair was randomly assigned to the Experimental Group (EG) and the other to the Control Group (CG)." P4.
Allocation concealment (selection bias)	Low risk	"One member of each pair was randomly assigned to the Experimental Group (EG) and the other to the Control Group (CG)." P4.
Blinding of participants and personnel (performance bias)	High risk	"The research assistant, who collected the outcome data, was unaware of the assigned condition of the participants." P4. But performance bias might have happened due to knowledge of the allocated interventions by participants.
Blinding of outcome assessment (detection bias)	Low risk	"The research assistant, who collected the outcome data, was unaware of the assigned condition of the participants." P4.
Incomplete outcome data (attrition bias)	High risk	“a slightly higher rate of attrition in the CG compared to the EG. As mentioned earlier, this was likely due to the frustration experienced by CG families who wanted and expected immediate clinical services (but instead received the Cool Runners Club) from Time 1 to Time 2.” P7. “… those who were lost would have been among the more serious cases in the group, which would lead toward a more “treatment amenable” CG at Time 5.” P7.
Selective reporting (reporting bias)	Low risk	Means and t-statistic at four time periods were reported. No standard deviation nor sample size at the different periods were reported.
Other bias	Low risk	The study appears to be free of other sources of bias.

B. Risk of bias in non-randomized studies

Table 15 displays study-level RoB judgments in the remaining 13 studies, which were included in the Meta-Analysis. The bias domains were judged as low for studies that were comparable to well-performed RT regarding these specific domains. The judgment of Moderate risk of bias was made to studies that were sound as non-randomized studies, but they were not comparable to well-performed RT regarding these specific domains. Furthermore, the bias domains were judged as

serious when the studies had some important problems in these specific domains. The "No Information" (NI) response was chosen when the researcher was unable to make a judgment due to the lack of information for that specific bias domain. The overall study-level RoB judgment was made based on the worst judgment across the seven bias domains.

Augimeri et al. (2018) study was overall judged as having a moderate risk because the worst judgment received was moderate on three bias domains. The first domain was the "bias in selection of participants into the study"; it was rated moderate because the start of follow-up and start of intervention did not coincide for most participants. However, adjustment techniques were likely used to correct for this bias. The second domain that was rated moderate was the "bias due to missing data" as the children were excluded due to missing pre-test data on either the CBCL or the Social Skills Improvement System (SSIS). The third domain was the "Bias in measurement of outcomes" because the outcome measure might have been influenced by knowledge of the intervention received, and the outcome assessors were probably aware of the intervention that was received by study participants. All the included studies, therefore, were rated moderate on this domain due to the same reason.

The three studies, **Smith-Moncrieffe (2015a), (2015b) and (2015c),** were overall judged as having a serious risk of bias because the worst judgment made across their domains was serious on the "Bias due to missing data" domain. The judgment was made as the outcome data were not available for all, or nearly all participants, and participants were probably excluded due to missing data on variables needed for the analysis. Besides, there was no sufficient data to make a judgment on the "Bias due to confounding" domain only for Smith-Moncrieffe (2015a) study. The three studies were finally rated as having a moderate risk on the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Augimeri et al. (2010) study was overall judged as having a moderate risk of bias because the worst judgment received was moderate on four bias domains. The first domain was the "bias in selection of participants into the study" because the start of follow-up and start of intervention did not coincide with most participants. However, adjustment techniques were likely used to correct for this bias. The second domain that was rated moderate was the "bias due to missing data" because some participants were excluded due to missing data on the EARL items. The study was also judged to be at a moderate risk on the "Bias in selection of the reported result" as the reported effect estimates were likely to be selected based on the results from multiple analyses of the intervention-outcome relationship and different subgroup analyses. The fourth domain rated as moderate was the "Bias in measurement of outcomes," and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Pepler et al. (2010) study was overall judged to be at moderate risk because the worst judgment received was moderate on three bias domains. The first domain was the "bias due to missing data" because the outcome data were not available for all, or nearly all, participants. However, there was evidence that results were robust to the presence of missing data because there were no differences between girls who had complete data and those who had missing follow-up data on any pre-test scores. The second domain was the "Bias in measurement of outcomes," and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Lipman et al. (2008) and (2007) studies were overall judged as having a serious risk of bias because the worst judgment made across their domains was serious on the "Bias due to missing data" domain. The judgment was made as the outcome data were not available for all, or nearly all, participants, and participants were probably excluded due to missing data on variables needed for the analysis. Specifically, boys with pre-test police contact were more likely not be included in

the pre-post group analysis in Lipman et al. (2008) study. Also, children with pre-post data were worse on externalizing behaviour than those with only pre-test data. So, the results might not be robust to the presence of missing data. Lipman et al. (2008) study was judged to be at serious risk on the "bias in selection of participants into the study" domain. The start of follow-up and start of intervention did not coincide for most participants, and adjustment techniques were probably not used to correct for this bias. Lipman et al. (2007), on the other hand, was rated as having serious risk of bias on "Bias due to confounding" domain, because the authors did not use an appropriate analysis method that controlled for all the important confounding domains and for time-varying confounding. Moreover, the two studies were judged to be at a moderate risk on the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Augimeri et al. (2006) study was overall judged to be at moderate risk because the worst judgment received was moderate on two bias domains. The first domain was on the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study. The second domain was the "Bias in selection of the reported result," as the reported effect estimates were likely to be selected based on the results from multiple analyses of the intervention-outcome relationship and different subgroup analyses. Specifically, the treatment effects were reported according to high, moderate and low delinquency classes and the difference between enhanced and standard ORP treatment groups.

Walsh et al. (2002) study was overall judged to be at moderate risk because the worst judgment received was moderate on two bias domains. The first domain was the "Bias due to missing data" because the outcome data were not available for all. Some participants had only six months of follow-up scores, and others had only 12 months of follow-up scores. The second domain was

Table 15 Risk of bias assessment for non-randomized studies.

Bias	Bias due to confounding	Bias in selection of participants into the study	Bias in classification of interventions	Bias due to deviations from intended interventions	Bias due to missing data	Bias in measurement of outcomes	Bias in selection of the reported result	Overall study-level RoB judgment
Study	**Risk of bias judgement**							
Augimeri et al. (2018), C10	Low	Moderate	Low	Low	Moderate	Moderate	Low	Moderate
Smith-Moncrieffe (2015a), QR	NI	Low	Low	Low	Serious	Moderate	Low	Serious
Smith-Moncrieffe (2015b), TR	Low	Low	Low	Low	Serious	Moderate	Low	Serious
Smith-Moncrieffe (2015c), ER	Low	Low	Low	Low	Serious	Moderate	Low	Serious
Augimeri et al. (2010), C8	Low	Moderate	Low	Low	Moderate	Moderate	Moderate	Moderate
Pepler et al. (2010), A4	Low	Low	Low	Low	Moderate	Moderate	Moderate	Moderate
Lipman et al. (2008), B1b	Low	Serious	Low	Low	Serious	Moderate	Low	Serious
Lipman et al. (2007), B2	Serious	Low	Low	Low	Serious	Moderate	Low	Serious
Augimeri et al. (2006), C3	Low	Low	Low	Low	Low	Moderate	Moderate	Moderate
Walsh et al. (2002), C9	Low	Low	Low	Low	Moderate	Moderate	Low	Moderate
Day (1998), C5	Low	Low	Low	NI	Low	Moderate	Low	Moderate
Day & Hunt (1996), C4	Low	Low	Low	Low	Moderate	Moderate	Low	Moderate
Hrynkiw-Augimeri et al. (1993), C1	Low	Low	Low	Low	Moderate	Moderate	Low	Moderate

NI (No Information): insufficient information provided to determine the risk of bias.

the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Day (1998) study also was overall judged to be at moderate risk because the worst judgment received was moderate on the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Day and Hunt (1996) study was overall judged to be at moderate risk because the worst judgment received was moderate on two bias domains. The first domain was the "Bias due to missing data" because the outcome data were not available for all participants. Participants with follow-up data had less antisocial behaviour at baseline than those with no follow-up data. The other domain was the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Hrynkiw-Augimeri et al. (1993) study was overall judged to be at moderate risk because the worst judgment received was moderate on two bias domains. The first domain was "Bias due to missing data." However, even though the attrition rate was high, and the outcome data were not available for all, there were no significant differences between those who stayed and those who dropped out. The second domain was the "Bias in measurement of outcomes" domain, and the reason was the same as mentioned earlier with the Augimeri et al. (2018) study.

Meta-Analysis results

This section presents the synthesis results in order to answer the research questions and test their accompanying hypotheses. The results for the study's primary outcomes are presented first, followed by the study's secondary outcomes. Sensitivity analysis results are presented last. Furthermore, the RCTs synthesis results are presented first, and then the synthesis results from the observational studies are presented second.

1. Results from the RCTs

As mentioned above, only two out of the fifteen included SNAP™ evaluation studies are RCTs. Furthermore, there was only one outcome that was common and available for extracting its data from these two RCTs. Specifically, both studies reported the official criminal records for children who reached the age of criminal responsibility by the end of the study. As shown in Figure 2, children who participated in the SNAP™ program had a significant 56% less chance of getting involved in criminal activity after receiving SNAP™ treatment in comparison to those who had not (Odds ratio= 0.44; p= 0.04; 95% CI: 0.21-0.95). Furthermore, the heterogeneity in effect sizes was not significant (p= 0.69) and (I^2= 0%).

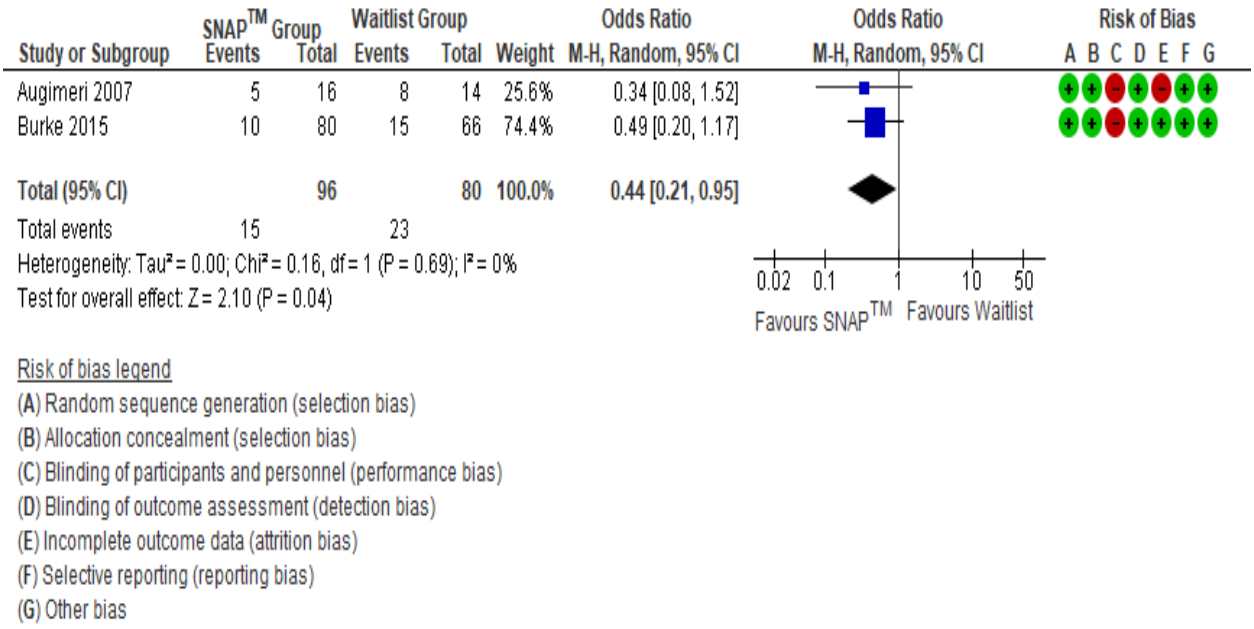


Figure 2 Forest plot of Official Criminal Records by the End of the Study.

2. Results from the non-randomized studies (observational studies)

A. Primary outcomes

1. Externalizing behaviours (CBCL, Parent's report and TRF, Teacher's report)

The forest plots in Figures 3 and 4 show the synthesis results of the CBCL-externalizing problem at post-treatment (about three months after admission), and six months of follow-up. As shown in Figure 3, only four included studies reported the results of externalizing behaviour problems outcome at post-treatment. There was a decrease in the children's' externalizing behaviour problems after three months of the SNAP^{TM} sessions (SMD= - 0.41), and this improvement was statistically significant (p< 0.0001; 95% CI: - 0.59, - 0.22). The heterogeneity in effect sizes was significant (p= 0.03) and moderate (I^2= 68%). After performing the possible subgroup analyses specifically, publication type; publication year; evaluation agent/type; participant gender; and risk of bias, three studies, Smith-Moncrieffe (2015a, 2015b, 2015c), were sub-grouped together versus one study, Pepler et al. (2010) (See Table 1 in appendix A). The subgroup differences were significant and high (Chi^2 = 7.42, df = 1 (p= 0.006), I^2 = 86.5%). Both subgroups had significant improvements; however, the treatment effect was larger for the subgroup that included journal articles published between 2001 and 2010, evaluated internally, had female-only and with a moderate risk of bias (SMD= - 0.67). Other subgroup analyses were performed specifically by the study location, treatment intensity and fidelity; however, their subgroup differences were not significant (See Table 1, appendix A).

There was a further significant decrease in the children's' externalizing behaviour problems from baseline to six months of follow-up (SMD= - 0.55; 95% CI: - 0.73, - 0.36; p< 0.00001); the heterogeneity in effect sizes was significant (p= 0.01) and moderate (I^2= 73%) (Figure 4). Therefore, subgroup analyses were performed (Table 2 in appendix A). High and significant

heterogeneity was observed in the subgroups by study design, evaluation agent/type and publication year. Prospective studies that were evaluated internally and published between 2001 and 2010 had a larger treatment effect, SMD= -0.75, -0.77 and -0.75, respectively, than the other subgroups of studies.

According to the teachers' reports, the children had also some non-significant improvements in their externalizing behaviour immediately after the 3-month SNAP™ sessions and 6-month of follow-up, (SMD= -0.04; p= 0.71; 95% CI: -0.22, 0.15), and (SMD= - 0.14; p= 0.11; 95% CI: -0.31, 0.03), respectively. Only the heterogeneity in the immediate effect sizes was significant (p= 0.004) and high (I^2= 77%) (Figures 5 & 6). Therefore, subgroup analyses were performed for TFR-externalizing behaviours at post-treatment (see Table 1, Appendix B). The subgroup differences were significant and high for subgroups by treatment fidelity, treatment intensity and study location. Studies with treatment fidelity more than 80%, where participant children attended more than 8-session of the Child SNAP™ Group and were performed in Toronto, Ontario, Canada had better, larger treatment effects than the other subgroups of studies.

1.1. Rule-breaking (CBCL, Parent's report and TRF, Teacher's report)

Four included studies reported the CBCL rule-breaking after the completion of the SNAP™ core sessions (Figure 7). According to the parents' reports the children had a significant decrease in the rule-breaking problems (SMD= - 0.36; p= 0.0002; 95% CI: - 0.55, - 0.17). Further significant reduction, (Figure 8), was achieved also after 6-month of follow-up (SMD= - 0.66; p< 0.00001; 95% CI: - 0.79, - 0.54). Only the heterogeneity in immediate effect sizes was significant (p= 0.02) and moderate (I^2= 71%). Therefore, subgroup analyses were performed for the CBCL rule-breaking at post-treatment (see Table 3, Appendix A). The subgroup differences were significant and high for subgroups by publication type, publication year, evaluation agent/type, participant

gender, and risk of bias. Again, the subgroup that included a journal article published between 2001 and 2010, evaluated internally, had female-only and with a moderate risk of bias had a better outcome (SMD= - 0.59; 95% CI: - 0.80, - 0.38), than the other subgroup of studies (SMD= - 0.28; 95% CI: - 0.43, - 0.13).

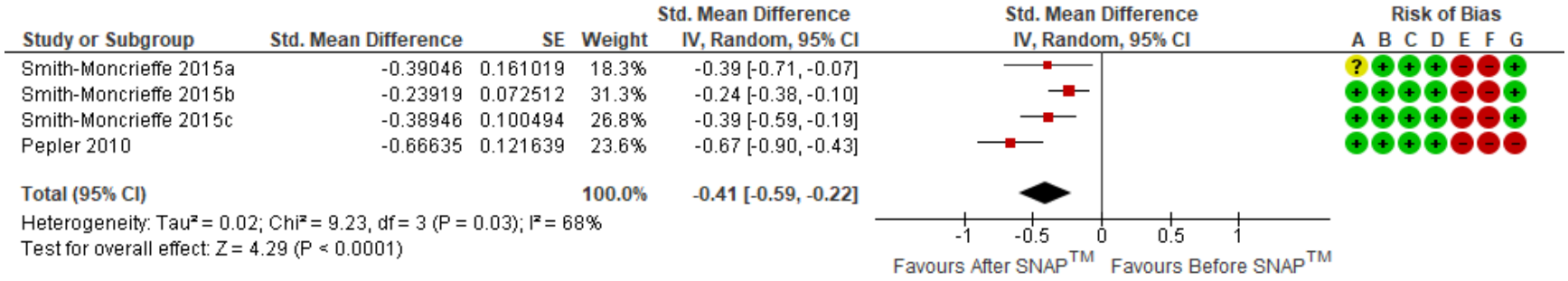
On the other hand, according to teachers' reports, the children had some non-significant increase, Figure 9, in the rule-breaking problems after 3-month of their participation in the SNAP^{TM} sessions (SMD= 0.04; p= 0.71; 95% CI: - 0.15, 0.22). But this increase did not last long, because the children had significant improvements in the rule-breaking problems after a period of 6-month of follow-up (SMD= - 0.12; p= 0.04; 95% CI: - 0.24, - 0.01), according to their teachers' reports (Figure 10). Only the heterogeneity in immediate effect sizes was significant (p= 0.001) and high (I^2= 81%). Therefore, subgroup analyses were performed for the TRF-rule-breaking at post-treatment (see Table 2, Appendix B). The heterogeneity in the effect sizes was significant and high for subgroups by treatment fidelity, treatment intensity and study location. Even though the treatment effect was non-significant, studies with treatment fidelity more than 80%, where participant children attended more than 8-session of the Child SNAP™ Group and were performed in Toronto, Ontario, Canada had a better outcome than the other subgroups of studies. Conversely, the subgroup that included one study with treatment fidelity less than 80%, where participant children attended less than 8-session of the Child SNAP™ Group and were performed in the Cree Nation, Quebec, Canada had a significant increase in the TRF-rule-breaking problem (SMD= 0.27, p< 0.0001; 95% CI: 0.14, 0.41).

1.2. Aggression (CBCL, Parent's report and TRF, Teacher's report)

The forest plots in Figures 11 and 12 show the synthesis results of the CBCL-aggression problem at post-treatment or three months after admission and six months of follow-up. As shown in Figure

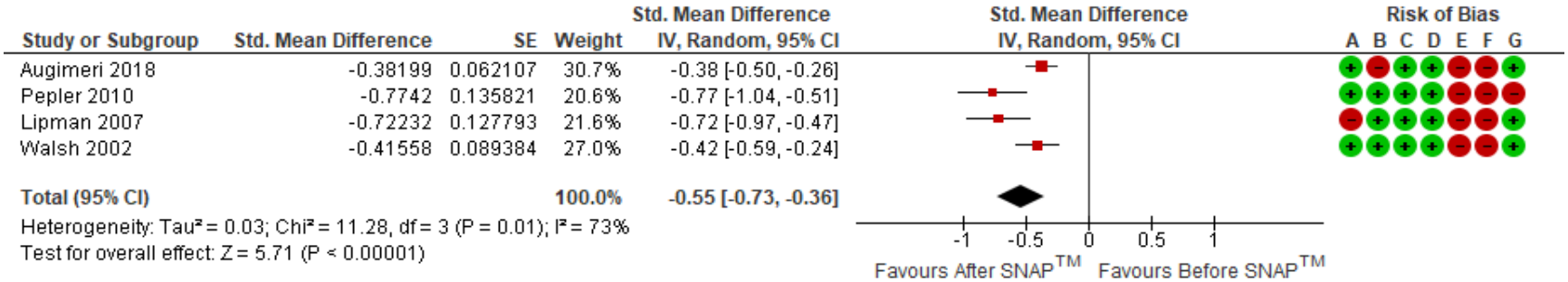
11, only four included studies reported the results of aggression outcome at post-treatment. There was a significant decrease in the children's' aggression after three months of the SNAP™ sessions (SMD= - 0.36; p< 0.00001; 95% CI: - 0.51, - 0.21). Furthermore, the heterogeneity in effect sizes was low (I^2= 45%) non-significant (p= 0.14). There was a further significant reduction in the children's aggression problem after six months of follow-up (SMD= - 0.73; p< 0.00001; 95% CI: - 0.86, - 0.60). Furthermore, the heterogeneity in effect sizes was zero (I^2= 0%) and non-significant (p= 0.86) (Figure 12).

According to teachers' reports, the children had also some non-significant improvements in their aggression immediately after the 3-month SNAP™ sessions and marginal improvement after 6-month of follow-up (SMD= - 0.05; p= 0.65; 95% CI: - 0.27, 0.17) and (SMD= - 0.14; p= 0.07; 95% CI: - 0.28, 0.01), respectively. Furthermore, only the heterogeneity in immediate effect sizes was significant (p= 0.0005) and high (I^2= 83%) (Figures 13 & 14). Therefore, subgroup analyses were performed for the TRF-aggression problem at post-treatment (see Table 3, Appendix B). The heterogeneity in the effect sizes was significant and high for subgroups by treatment fidelity, treatment intensity and study location. Studies with treatment fidelity more than 80%, where participant children attended more than 8-session of the Child SNAP™ Group and were performed in Toronto, Ontario, Canada had significant immediate reductions in the TRF-aggression problems (SMD= - 0.14; 95% CI: - 0.25, - 0.04), (SMD= - 0.15; 95% CI: - 0.30, 0.00) and (SMD= - 0.20; 95% CI: - 0.35, - 0.04), respectively. On the other hand, the subgroup that included one study with treatment fidelity less than 80%, where participant children attended less than 8-session of the Child SNAP™ Group and were performed in the Cree Nation, Quebec, Canada had a significant immediate increase in the TRF-aggression problem (SMD= 0.25, p= 0.002; 95% CI: 0.09, 0.41) according to the children's teachers.



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 3 Forest plot of CBCL-Externalizing Problem (Post-Treatment).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 4 Forest plot of CBCL-Externalizing Problem (6-Month Follow-up).

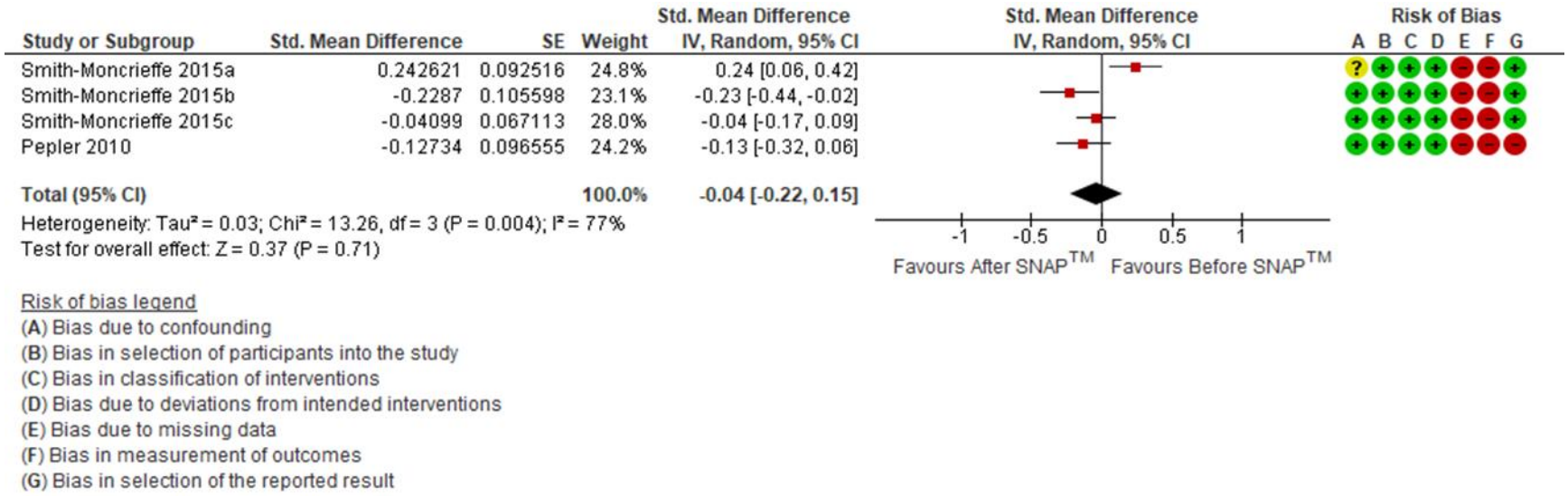


Figure 5 Forest plot of TRF-Externalizing Problem (Post-Treatment).

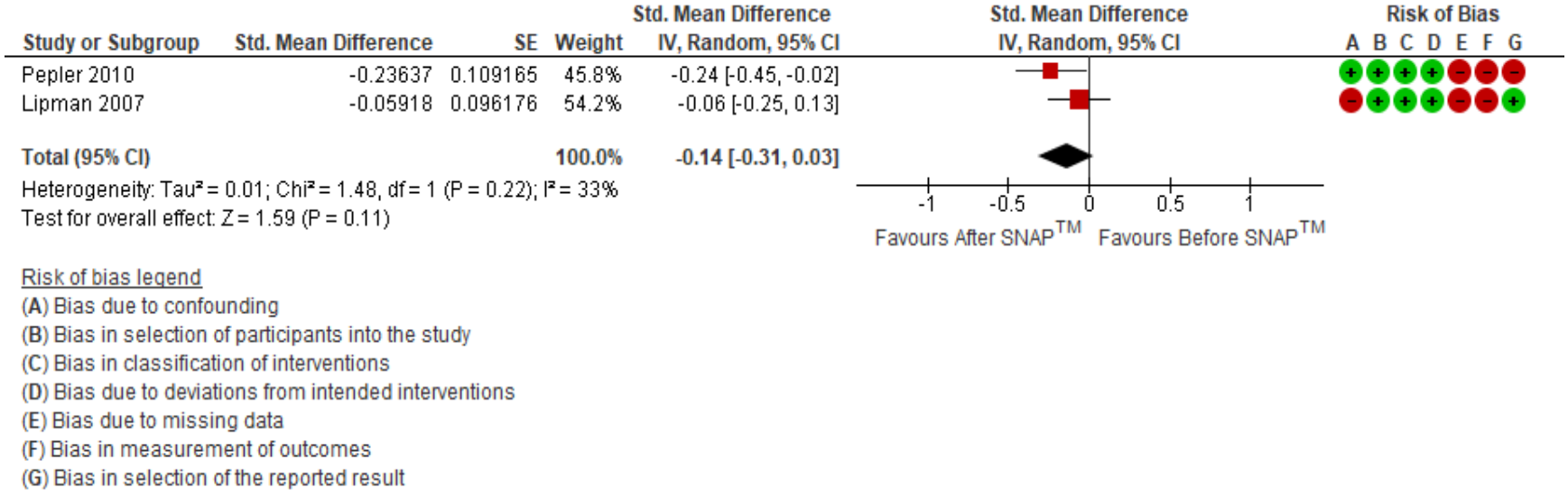
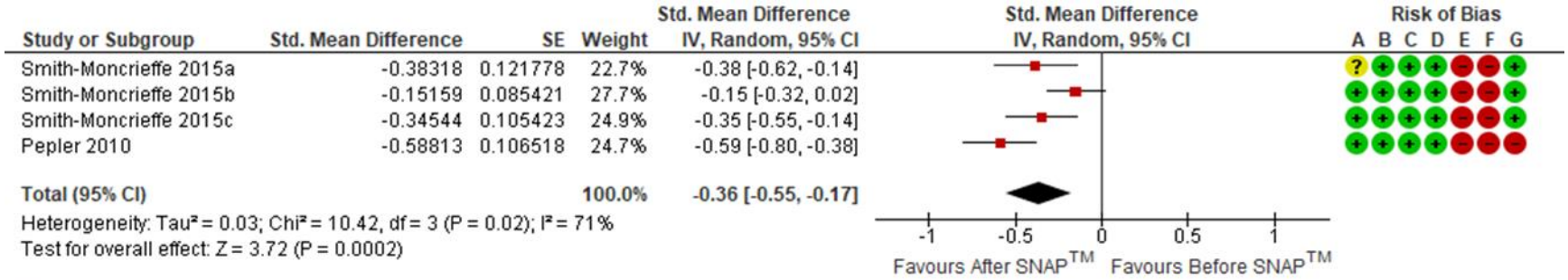
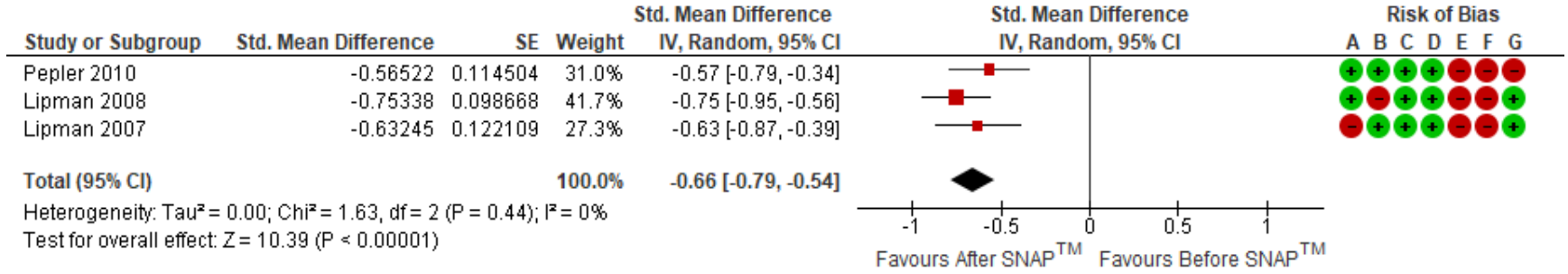


Figure 6 Forest plot of TRF-Externalizing Problem (6 Months of Follow-up).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 7 Forest plot of CBCL-Rule-Breaking (Post-Treatment).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 8 Forest plot of CBCL-Rule-Breaking (6-Month Follow-up).

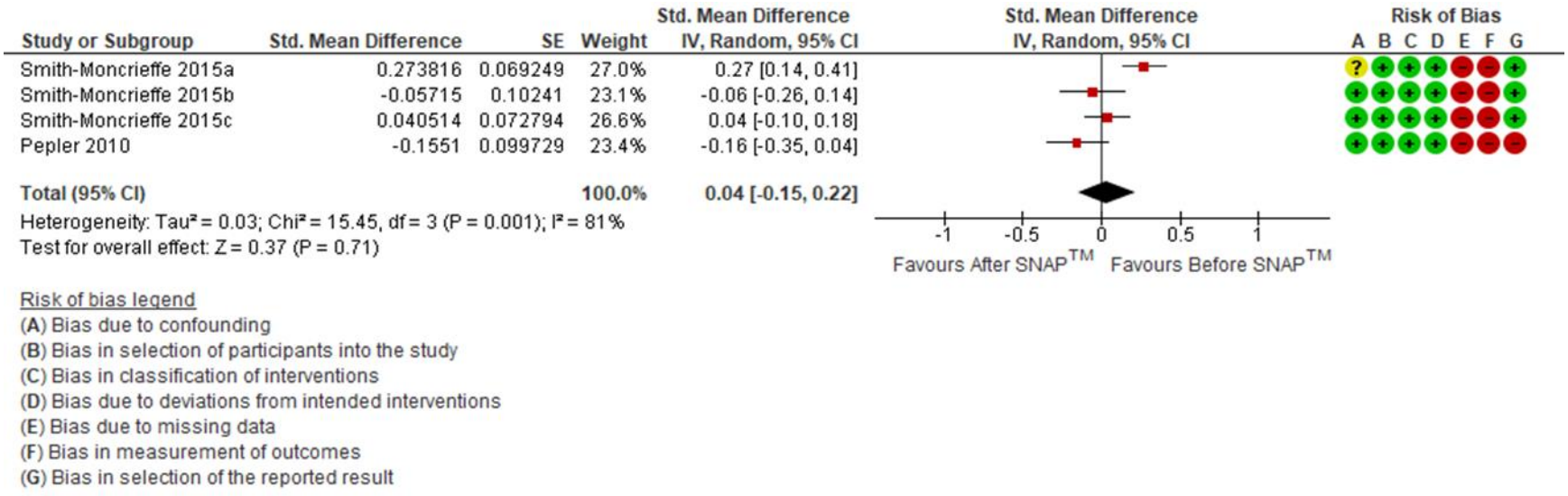


Figure 9 Forest plot of TRF-Rule-Breaking (Post-Treatment).

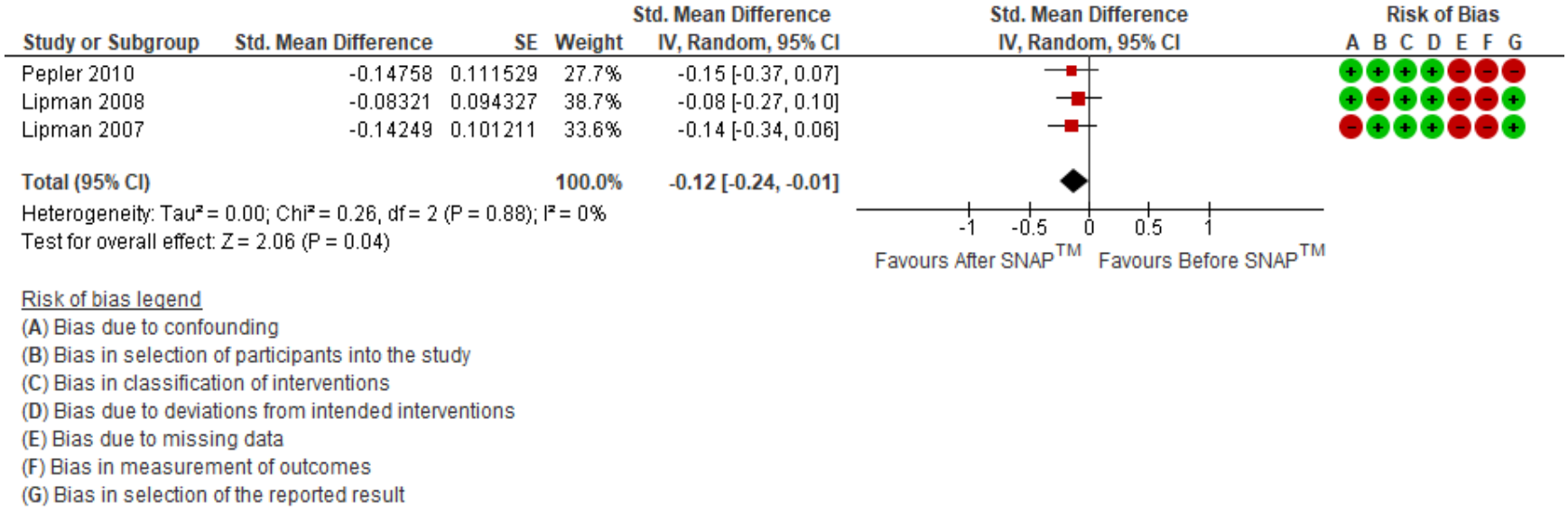


Figure 10 Forest plot of TRF-Rule-Breaking (6 Months Follow-up).

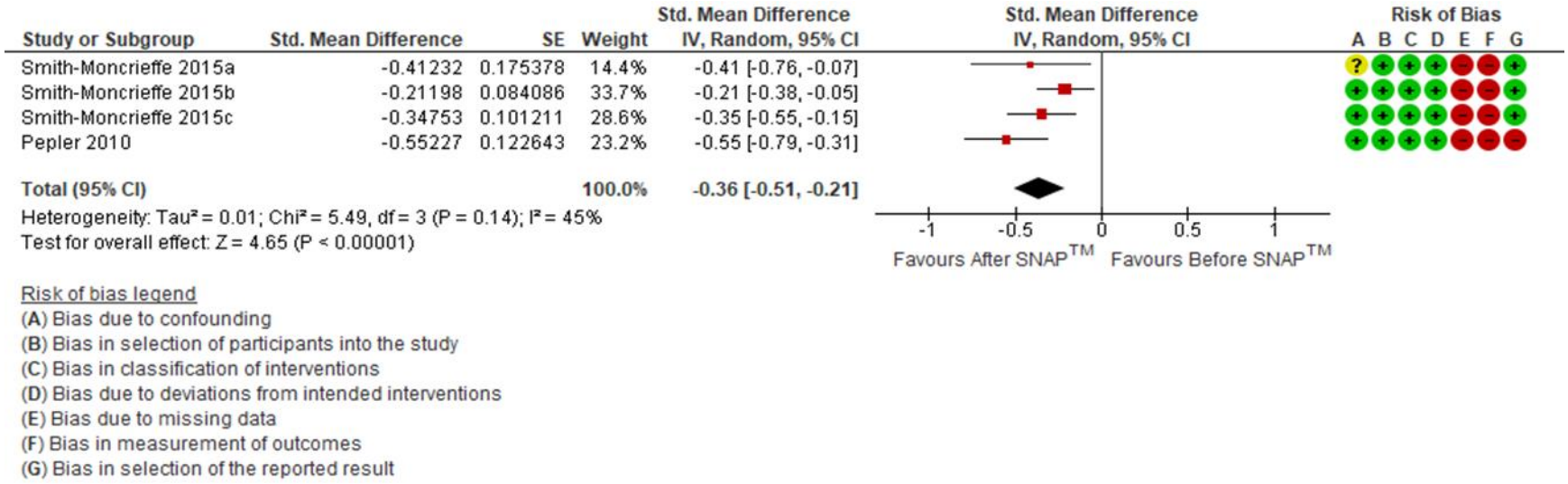


Figure 11 Forest plot of CBCL-Aggression (Post-Treatment).

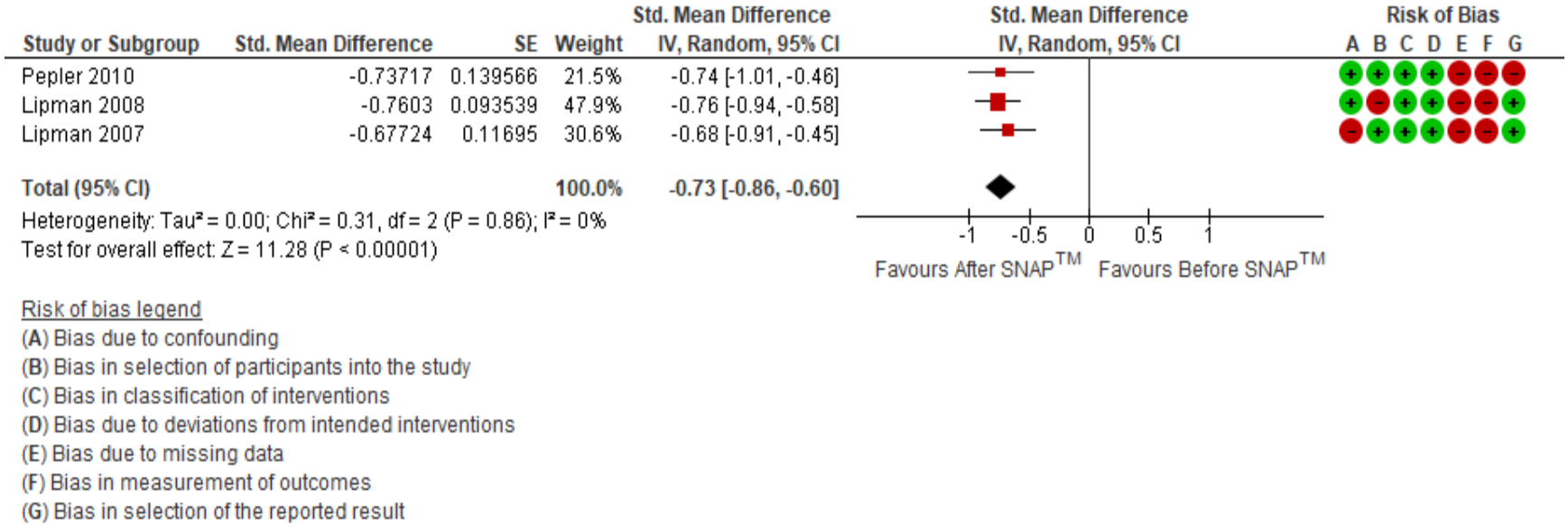


Figure 12 Forest plot of CBCL-Aggression (6-Month Follow-up).

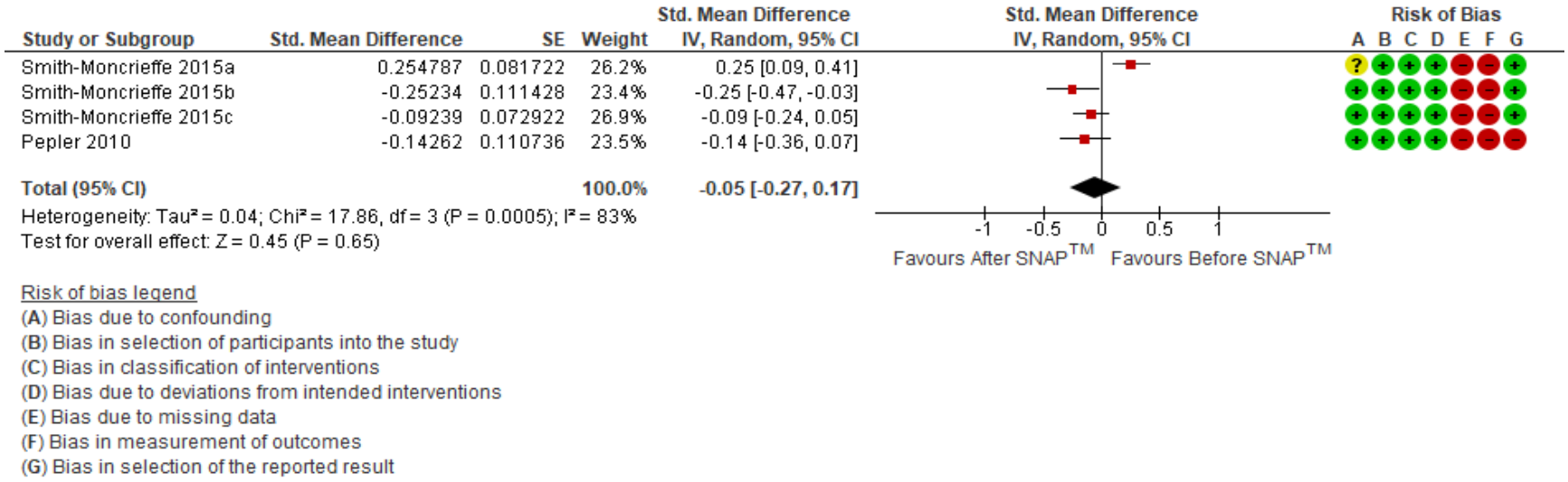


Figure 13 Forest plot of TRF-Aggression (Post-Treatment).

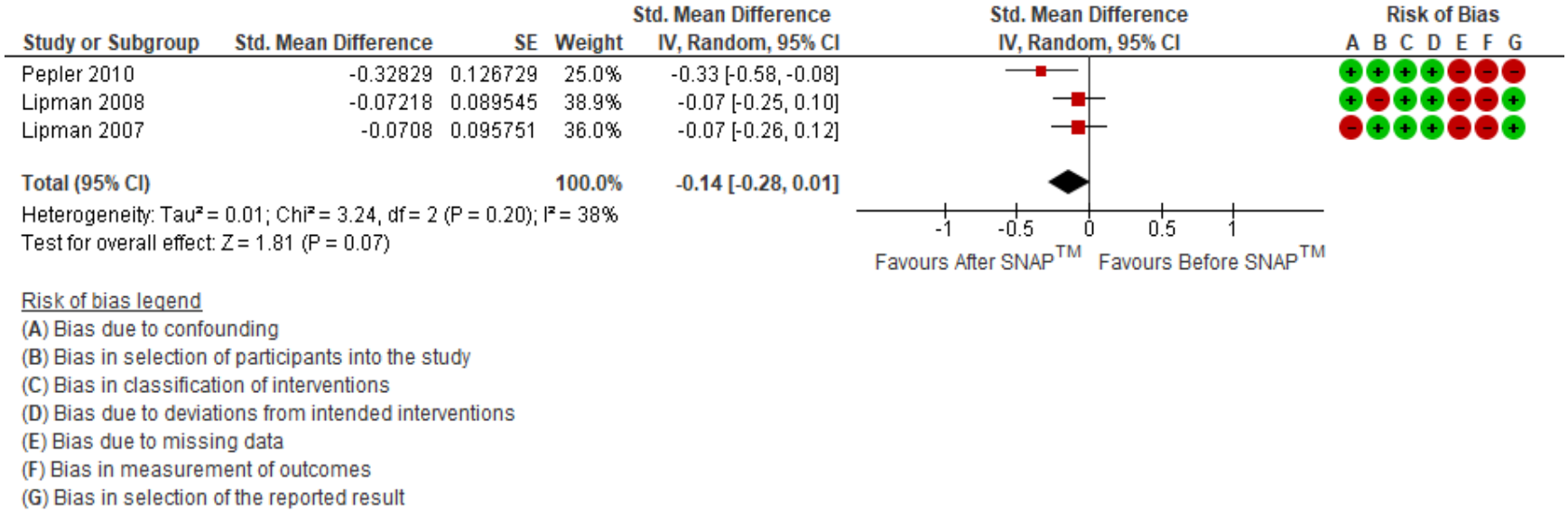


Figure 14 Forest plot of TRF-Aggression (6 Months Follow-up).

2. Delinquency (CBCL, Parent's report)

Only two included studies reported the delinquency outcome (Figure 15). There was a significant decrease in the children's delinquent behaviour after 9-month of their participation in the program (SMD= - 0.86; p< 0.00001; 95% CI: - 1.17, - 0.55). Furthermore, the heterogeneity in effect sizes was significant (p= 0.004) and high (I^2= 88%). Therefore, subgroup analyses were performed (see Table 4, Appendix A). The possible subgroup analyses were performed by publication type, publication year, study design and participant gender. Both subgroups had significant improvements in the children's CBCL-delinquency; however, the treatment effect was larger for the subgroup that included a journal article published between 1991 and 2000, performed prospectively and had mixed-gender (SMD= - 1.03).

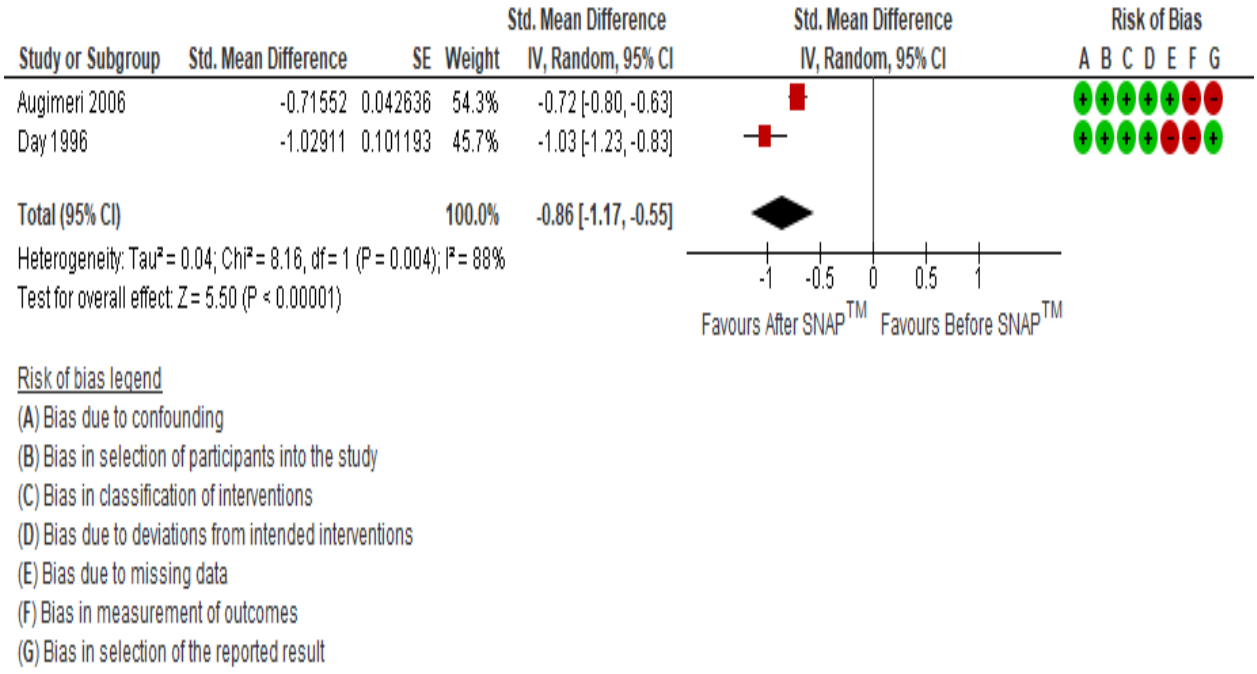


Figure 15 Forest plot of CBCL-Delinquency (9-Month Follow-up).

3. DSM Conduct Disorder (CBCL, Parent's report and TRF, Teacher's report)

Four included studies reported the CBCL-DSM conduct problems after the completion of the SNAPTM core sessions (Figure 16). According to the parents' reports the children had a significant decrease in the conduct problems (SMD= - 0.41; p< 0.00001; 95% CI: - 0.57, - 0.25). A further significant reduction, (Figure 17), was achieved also after 6-month of follow-up (SMD= - 0.61; p< 0.00001; 95% CI: - 0.79, - 0.44). Furthermore, only the heterogeneity in the 6-month follow-up effect sizes was significant (p= 0.04) and moderate (I^2= 64%). Therefore, subgroup analyses were performed for the CBCL-DSM Conduct Disorder at 6-month of follow-up (see Table 5, Appendix A). The heterogeneity in the effect sizes was significant and high for subgroups by study design, year of enrollment, outcome measures and evaluation agent/type. Even though the treatment effect was significant in all these subgroups, studies with a prospective design, enrollment year between 2001 and 2017, used the outcome measure CBCL (2001) and were evaluated externally had a larger treatment effect than the other subgroup of studies.

According to teachers' reports (Figure 18), the children had also some non-significant improvement in their DSM conduct disorder immediately after the 3-month SNAP^{TM} sessions (SMD= - 0.04; p= 0.68; 95% CI: - 0.20, 0.13). After 6-month of follow-up, a significant decrease had been observed in the children's DSM conduct disorder (SMD= - 0.19; p= 0.003; 95% CI: - 0.32, - 0.07), according to their teachers (Figure 19). Furthermore, only the heterogeneity in immediate effect sizes was significant (p= 0.02) and moderate (I^2= 68%), and accordingly, the subgroup analyses were performed (see Table 4, Appendix B). The heterogeneity in the effect sizes was significant and high for subgroups by treatment fidelity, treatment intensity and study location. Studies with treatment fidelity more than 80% and were performed in Toronto, Ontario, Canada had significant immediate reductions in the children's TRF- DSM Conduct Disorder

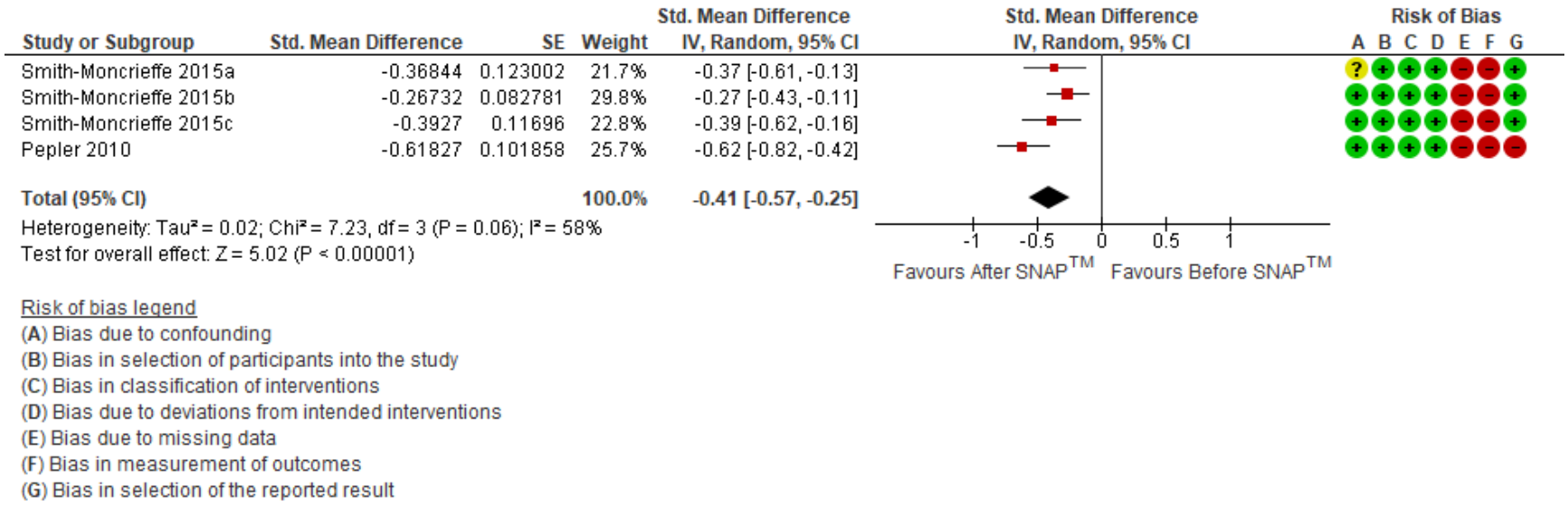


Figure 16 Forest plot of CBCL-DSM Conduct Disorder (Post-Treatment).

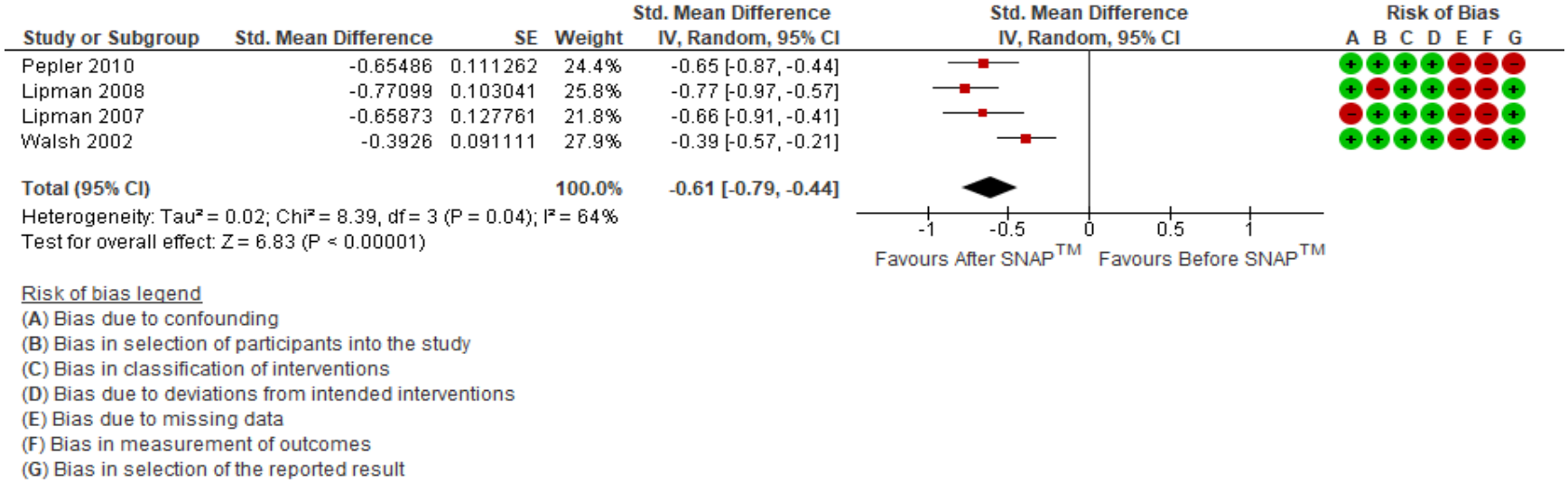


Figure 17 Forest plot of CBCL-DSM Conduct Disorder (6 Months Follow-up).

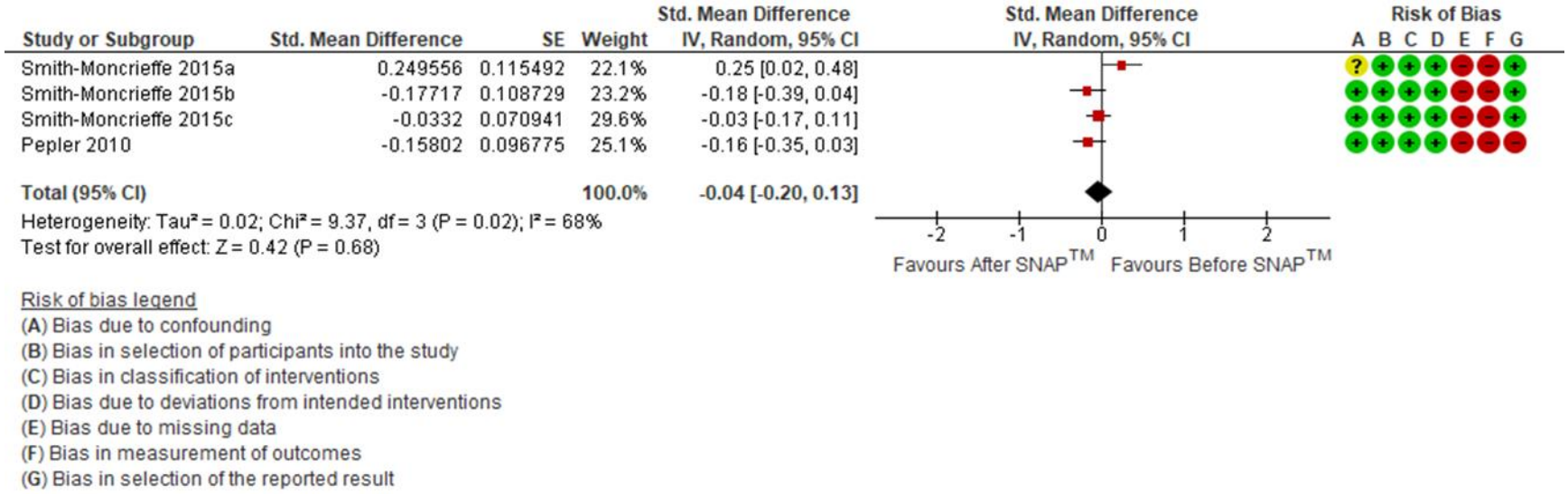


Figure 18 Forest plot of TRF-DSM Conduct Disorder (Post-Treatment).

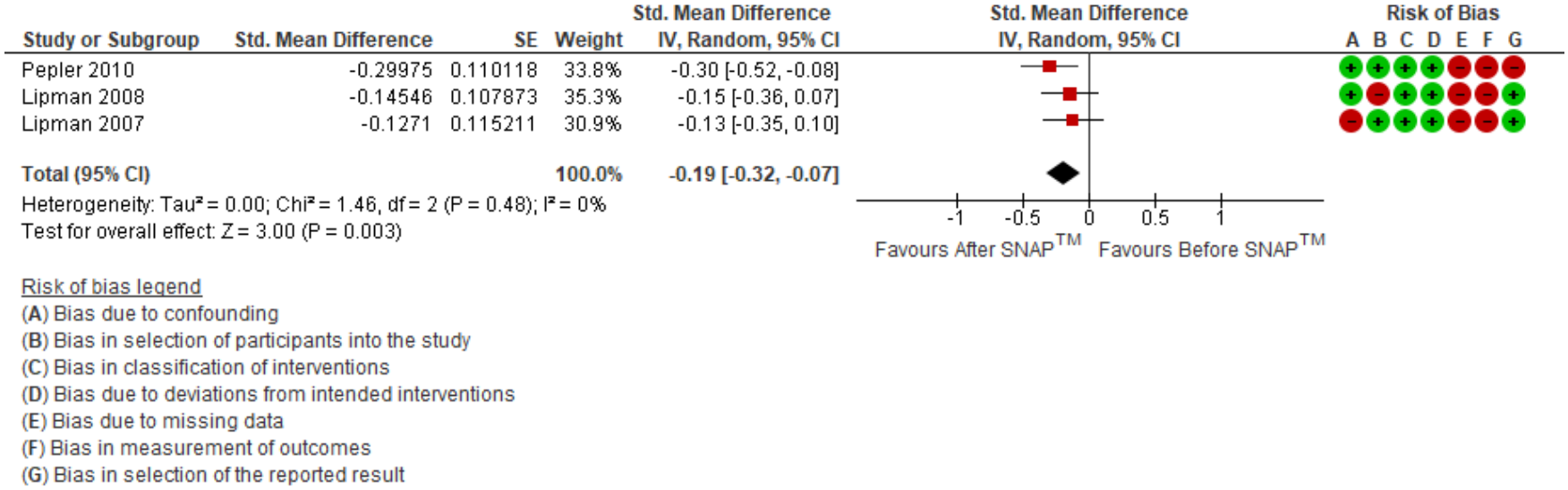


Figure 19 Forest plot of TRF-DSM Conduct Disorder (6 Months Follow-up).

(SMD= - 0.10; 95% CI: - 0.20, 0.00) and (SMD= - 0.17; 95% CI: - 0.31, - 0.02), respectively. On the other hand, the subgroup that included one study with treatment fidelity less than 80%, where participant children attended less than 8-session of the Child SNAP™ Group and were performed in the Cree Nation, Quebec, Canada had a significant immediate increase in the TRF- DSM Conduct Disorder (SMD= 0.25, p= 0.03; 95% CI: 0.02, 0.48) according to the children's teachers.

4. **Attention Deficit Hyperactivity Disorder (ADHD) (CBCL, Parent's report and TRF, Teacher's report)**

Four included studies reported the CBCL-ADHD after the completion of the SNAP™ core sessions (Figure 20). According to the parents' reports the children had a significant decrease in the ADHD (SMD= - 0.30; p< 0.00001; 95% CI: - 0.42, - 0.18). Further significant reduction, (Figure 21), was achieved also after 6-month of follow-up (SMD= - 0.44; p< 0.00001; 95% CI: - 0.59, - 0.29). According to teachers' reports (Figure 22), the children had also some significant improvements in their ADHD immediately after the 3-month of their participation in SNAP™ sessions (SMD= -0.11; p= 0.02; 95% CI: - 0.20, - 0.02). After 6-month of follow-up (Figure 23), there was a further reduction in the children's TRF-ADHD; however, this improvement was not statistically significant (SMD= - 0.08; p= 0.31; 95% CI: - 0.24, 0.07). Furthermore, the heterogeneity in effect sizes from both reports (CBCL and TRF) was non-significant (Figures 20-23).

5. **Oppositional Defiant Disorder (ODD) (CBCL, Parent's report)**

For the ODD, only results from the parents' reports (CBCL) were available. After the completion of the SNAP™ core sessions (Figure 24), the children had a significant reduction in the ODD (SMD= - 0.23; p= 0.0005; 95% CI: - 0.36, - 0.10), based on three studies. A further significant decrease in the ODD, Figure 25, was also reached after 6 months of follow-up (SMD= - 0.53; p<

0.00001; 95% CI: - 0.71, - 0.34). Moreover, the heterogeneity in the immediate and 6-month follow-up effect sizes was non-significant.

6. Attention problems (CBCL, Parent's report and TRF, Teacher's report)

Four included studies reported the CBCL-attention problems after the completion of the SNAP™ core sessions (Figure 26). According to the parents' reports the children had a significant decrease in the attention problems (SMD= - 0.28; p= 0.0004; 95% CI: - 0.43, - 0.12). Besides, the heterogeneity in effect sizes was significant (p= 0.05) and moderate (I^2= 61%). After performing the possible subgroup analyses, the heterogeneity in the effect sizes for all subgroups was non-significant, and its magnitude was either low or 0% (see Table 6, Appendix A).

According to the parents' reports, a further significant reduction was achieved after 6-month of follow-up (SMD= - 0.31; p< 0.00001; 95% CI: - 0.42, - 0.20), and the heterogeneity in effect sizes was non-significant and I^2 was 0% (Figure 27).

Furthermore, according to teachers' reports, the children had some non-significant reduction, Figure 28, in their attention problems immediately after their 3-month of their participation in the SNAP™ sessions (SMD= - 0.05; p= 0.56; 95% CI: - 0.24, 0.13). Subgroup analyses were performed for the TRF-attention problems at post-treatment as the heterogeneity in effect sizes was significant (p= 0.005), and high (I^2= 76%). The heterogeneity in the effect sizes was significant and high for subgroups by treatment fidelity, treatment intensity and study location (see Table 5, Appendix B). Studies with treatment fidelity more than 80%, and participant children attended more than 8-session of the Child SNAP™ had significant immediate reductions in the TRF-attention problems (SMD= - 0.14; 95% CI: - 0.26, - 0.03) and (SMD= - 0.17; 95% CI: - 0.30, - 0.03), respectively. On the other hand, a significant immediate increase in the TRF-attention problems (SMD= 0.15; 95% CI: 0.03, 0.27) was observed in the subgroup that included a study,

which was conducted in the Cree Nation, Quebec, Canada, where participant children attended less than 8-session of the Child SNAP™ Group. The study's treatment fidelity was also less than 80%.

A further non-significant decrease in the children's attention problems was reported after a period of 6-month of follow-up (SMD= - 0.12; p= 0.18; 95% CI: - 0.29, 0.05), according to their teachers' reports. Furthermore, the heterogeneity in effect sizes of TRF-aggression problems at a 6-month follow-up was non-significant (p= 0.59), and the I^2 was 0% (Figure 29).

7. Social problems (CBCL, Parent's report and TRF, Teacher's report)

Only three included SNAP™ evaluation studies had reported the participating children's CBCL-social problems. According to the parents' reports there was a significant decrease in the children's social problems (SMD= - 0.50; p< 0.00001; 95% CI: - 0.68, - 0.32) after a period of 6-month of their participation in the program (Figure 30). The teachers' reports showed some non-significant increase in the children's social problems (SMD= 0.02; p= 0.83; 95% CI: - 0.14, 0.18) immediately following the 12-week core SNAP™ Child Group sessions (Figures 31). Nonetheless, a non-significant reduction was later observed in the children's social problems after 6 months of their participation in the SNAP™ (SMD= - 0.15; p= 0.25; 95% CI: - 0.41, 0.11), according to their teachers (Figures 32). Furthermore, the heterogeneity in effect sizes according to both reports (CBCL and TRF) was non-significant and low (Figures 30-32).

8. CBCL-Total social competence

The CBCL-total competence outcome at post-treatment was reported in three included studies (Figure 33). A marginal increase or improvement (SMD= 0.10; p= 0.09; 95% CI: - 0.02, 0.23) in the children total social competence as reported by the parents/guardians. Furthermore, a significant improvement (SMD= 0.28; p< 0.0001; 95% CI: 0.15, 0.40) was reported after a 6-month period of follow-up (Figure 34). The heterogeneity in effect sizes for both periods (post-

and 6-month follow-up) was zero (I^2= 0%), and non-significant.

9. TRF-Adaptive functioning

The TRF-adaptive functioning outcome at post-treatment was reported in three included studies (Figure 35). A non-significant decrease (SMD= - 0.03; p= 0.76; 95% CI: - 0.19, 0.14) in the children's adaptive functioning as reported by their teachers. A non-significant increase or improvement (SMD= 0.08; p= 0.60; 95% CI: - 0.23, 0.40) was reported after a 6-month period of follow-up (Figure 36). Furthermore, the heterogeneity in effect sizes was significant (p= 0.04) and high (I^2= 77%) only at the 6-month follow-up analysis. After performing the possible subgroup analyses (see Table 6, Appendix B), the subgroup that included a study, which was a published journal article and involved males only, had a better and significant outcome where the children had improvement in their TRF-adaptive functioning (SMD= 0.24; 95% CI: 0.04, 0.45).

10. Justice system involvements/criminal convictions

Two of the included observational studies reported data about the children's official criminal records by the end of the study for those who were between ages 12 and 18, and also pre-admission police contact data. As shown in Figure 37, the SNAP^{TM} program had made a significant 85% decrease in the chance of getting the participating youth, who reached the age of criminal responsibility, involved in criminal activities based on a follow-up period ranged from 1 to 11 years from their admission to the program (Odds ratio= 0.15; p= 0.002; 95% CI: 0.04-0.51). However, there was a significant (p= 0.02), and high heterogeneity (I^2= 81%) in the studies' effect sizes. Therefore, subgroup analyses were performed (see Table 7, Appendix A). The subgroup that included a study, which was evaluated by a team composed of internal and external researchers (shared) and based on the parents/guardians' reports as a mean of the outcome measure, had a better outcome where the children had improvement more than the children in the other subgroup (Odds ratio= 0.07; 95% CI: 0.03, 0.19). Nonetheless, both subgroups had significant

improvements.

B. Secondary outcomes

1. Internalizing behaviour problems (CBCL, Parent's report and TRF, Teacher's report)

Four included studies reported the CBCL-internalizing behaviour problems after the completion of the SNAP™ core sessions (Figures 38-39). According to the parents' reports the children had a significant decrease in their internalizing behaviour problems (SMD= - 0.37; p< 0.0001; 95% CI: - 0.55, - 0.19) immediately and after a 6-month of follow-up (SMD= - 0.48; p< 0.00001; 95% CI: - 0.68, - 0.29). The heterogeneity in effect sizes was significant (p= 0.03) and moderate (I^2= 67%) for the post-treatment analysis. After performing the possible subgroup analyses, the heterogeneity in the effect sizes for all subgroups was non-significant, and its magnitude was either moderate or 0% (see Table 8, Appendix A).

According to the teachers' reports (Figures 40-41), the children had also some improvements, but non-significant in their internalizing behaviour problems immediately after the 3-month of their participation in SNAP™ sessions and after 6-month of follow-up, (SMD= - 0.00; p= 0.98; 95% CI: - 0.26, 0.26) and (SMD= - 0.16; p= 0.13; 95% CI: - 0.37, 0.05), respectively. Furthermore, the heterogeneity in effect sizes was significant (p< 0.0001) and high (I^2= 88%) for the post-treatment analysis; accordingly, the subgroup analyses were performed. The heterogeneity in the effect sizes was significant and high for the subgroups by treatment fidelity, treatment intensity and study location (see Table 5, Appendix B). Studies that were conducted in Toronto, Ontario, Canada had significant immediate reductions in the TRF-internalizing behaviour problems (SMD= - 0.21; 95% CI: - 0.35, - 0.08). On the contrary, a significant immediate increase (SMD= 0.45; 95% CI: 0.22, 0.67) was observed in the subgroup that included a study conducted in the Cree Nation, Quebec, Canada, where children attended less than 8-session of the required Child SNAP™ Group sessions,

and the treatment fidelity was less than 80%.

1.1. CBCL-Withdrawn/depressed

Only three of the included studies reported the CBCL-withdrawn/depressed behaviour problems after the completion of the SNAP™ core sessions (Figure 42). According to the parents' reports the children had some non-significant decrease in the withdrawn/depressed behaviour problems (SMD= - 0.17; p= 0.24; 95% CI: - 0.46, 0.12). Furthermore, the heterogeneity in effect sizes was significant (p= 002), and high (I^2= 85%). After performing the possible subgroup analyses, only one subgroup, which included a study conducted in Edmonton, Alberta, Canada, had a significant and better outcome (SMD= - 0.39; 95% CI: - 0.54, - 0.24), according to their parents/guardians (see Table 9, Appendix A).

1.2. CBCL-Somatic complaints

Three of the included studies reported the CBCL-somatic complaint behaviour problems after the completion of the SNAP™ core sessions (Figure 43). According to the parents' reports the children had some non-significant decrease in the somatic complaint behaviour problems (SMD= - 0.20; p= 0.29; 95% CI: - 0.56, 0.17). Furthermore, the heterogeneity in effect sizes was significant and high (I^2= 90%); accordingly, the subgroup analysis was performed. The results (see Table 10, Appendix A), showed that one subgroup had a significant immediate decrease in the CBCL-somatic complaint problems (SMD= - 0.60; 95% CI: - 0.84, - 0.36). Again, this subgroup comprised of one study with treatment fidelity less than 80%, were conducted in the Cree Nation, Quebec, Canada, and the participating children had attended less than 8-session of the Child SNAP™ Group.

1.3. CBCL-Anxious/depressed

Three of the included studies reported the CBCL-anxious/depressed problems after the completion of the SNAP™ core sessions (Figure 44). According to the parents' reports the children had a

significant immediate decrease in the anxious/depressed behaviour problems (SMD= - 0.40; p= 0.002; 95% CI: - 0.66, - 0.14). Furthermore, the heterogeneity in effect sizes was significant (p= 0.02), and high (I^2= 74%). After performing the possible subgroup analyses (see Table 11, Appendix A), significant improvements in the Children's CBCL-anxious/depressed problems were reported in two studies, one was conducted in the Cree Nation, Quebec, Canada (SMD= -0.65; 95% CI: -0.95, -0.35), and the other was conducted in Edmonton, Ontario, Canada (SMD= -0.44; 95% CI: - 0.64, - 0.23).

2. Total behaviour problems (CBCL, Parent's report and TRF, Teacher's report)

Three included studies reported the CBCL-total behaviour problems after the completion of the SNAP™ core sessions (Figure 45). According to the parents' reports the children had a significant immediate decrease in the total behaviour problems (SMD= - 0.40; p< 0.0001; 95% CI: - 0.59, -0.21). A further significant reduction, (Figure 46), was also achieved after 6-month of follow-up (SMD= - 0.74; p< 0.00001; 95% CI: - 0.92, - 0.55). According to teachers' reports, the children had some non-significant decrease in their total behaviour problems (SMD= - 0.04; p= 0.75; 95% CI: - 0.28, 0.20) immediately after the 3-month of their participation in SNAP™ sessions (Figure 47). Teachers had also reported a further non-significant decrease (SMD= - 0.09; p= 0.16; 95% CI: - 0.23, 0.04) in the same outcome after 6-month of follow-up (Figure 48). Furthermore, the heterogeneity in effect sizes for both reports (CBCL and TRF) was significant (p= 0.009), and high (I^2= 79%) only for immediate effects of TRF- total behaviour problems (Figures 45-48). The subgroup analyses were performed accordingly. The heterogeneity in the effect sizes was significant and high for subgroups by treatment fidelity, treatment intensity and study location (see Table 8, Appendix B). Studies that were conducted in Toronto and Edmonton, Ontario, Canada had marginal immediate reductions in the TRF-total behaviour problems (SMD= - 0.24; 95% CI:

- 0.48, 0.01) and (SMD= - 0.11; 95% CI: - 0.23, 0.00), respectively. Furthermore, Studies with treatment fidelity more than 80% had a significant immediate decrease in the children's total behaviour problems (SMD= - 0.13; 95% CI: - 0.24, - 0.03), according to the teachers. However, the subgroup that composed of only the Cree Nation, Quebec's study had a significant immediate increase in the total behaviour problems (SMD= 0.24; 95% CI: 0.01, 0.46), according to the children's teachers.

3. Parenting skills

Only one study, Pepler et al. (2010), of the included studies has reported changes in parenting skills after they participated in SNAP™ parents' sessions. Therefore, this outcome could not be analyzed any further in the current Meta-Analysis.

Sensitivity analysis

The sensitivity analysis was performed for each outcome, with at least three included studies.

- *The findings that were based on the children's parents/guardians' reports (CBCL):*

By removing one review at a time from the analysis, there was no effect on the overall CBCL-effect sizes' direction or statistical significance except for three CBCL-outcomes. The first and the second outcomes were the total competence and withdrawn-depressed at post-treatment (Figures 49 & 50). By removing one study specifically Smith-Moncrieffe (2015b), the overall effect size changed from non-significant to significant (SMD= 0.17; p= 0.04; 95% CI: 0.01, 0.34) and (SMD= - 0.30; p= 0.01; 95% CI: - 0.54, - 0.06), respectively. The only difference between this study and the others is that it was conducted in Toronto, Ontario. The third outcome was the anxious/depressed at post-treatment (Figure 51). By removing one study specifically Smith-Moncrieffe (2015c), the overall effect size changed from significant to non-significant (SMD= - 0.40; p= 0.10; 95% CI: - 0.87, 0.08). The only difference between this study and the others is that

it was conducted in Edmonton, Alberta.

- *The findings that were based on the teachers' reports (TRF):*

The sensitivity analyses showed no effect on the overall TRF-effect sizes' direction or significance, except for the following TRF-outcomes:

The first outcome was the internalizing problems at post-treatment (Figure 52). By removing one study specifically Smith-Moncrieffe (2015a), the overall effect size changed from non-significant to marginally significant (SMD= - 0.13; p= 0.06; 95% CI: - 0.27, 0.01). The difference between this study and the others is that it was conducted in the Cree Nation, Quebec, its treatment fidelity was less than 80%, and its participating children attended less than 8-session of the Child SNAP™ Group.

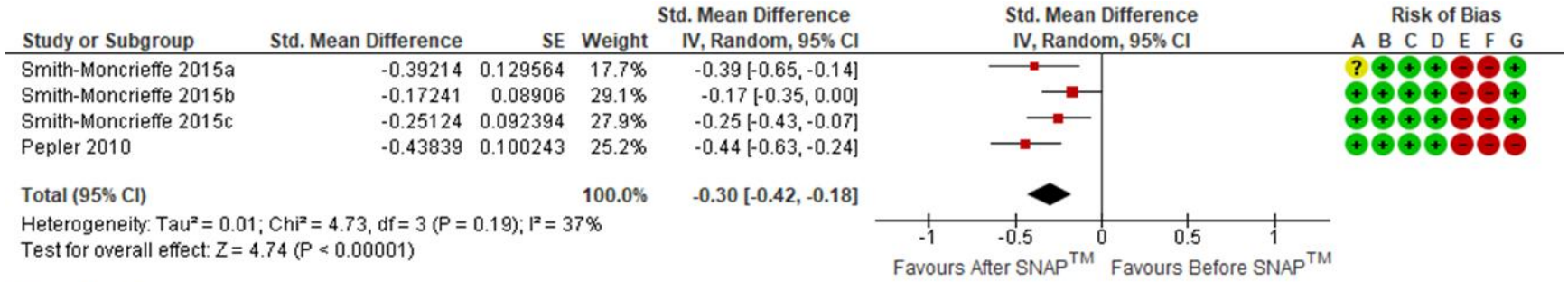
Furthermore, the sensitivity analyses for the TRF-externalizing problem, total problem, aggression, attention, and DSM conduct disorder at post-treatment, showed that by removing one study specifically Smith-Moncrieffe (2015a), the overall immediate effect sizes changed from non-significant to significant (SMD= - 0.11; p= 0.04; 95% CI: - 0.21, - 0.00), (SMD= - 0.13; p= 0.01; 95% CI: - 0.24, - 0.03), (SMD= - 0.14; p= 0.008; 95% CI: - 0.25, - 0.04), (SMD= - 0.14; p= 0.02; 95% CI: - 0.26, - 0.03), (SMD= - 0.10; p= 0.05; 95% CI: - 0.20, 0.00), respectively (Figures 53-57). Moreover, the heterogeneity in the effect sizes of these outcomes changed from significant to non-significant, and its magnitude changed from moderate and high to low or zero.

Last, the sensitivity analysis for the TRF- DSM ADHD at post-treatment (Figure 58) showed that by removing one study specifically Smith-Moncrieffe (2015c), the overall effect size changed from significant to non-significant (SMD= - 0.07; p= 0.32; 95% CI: - 0.22, 0.07).

Assessment of publication bias

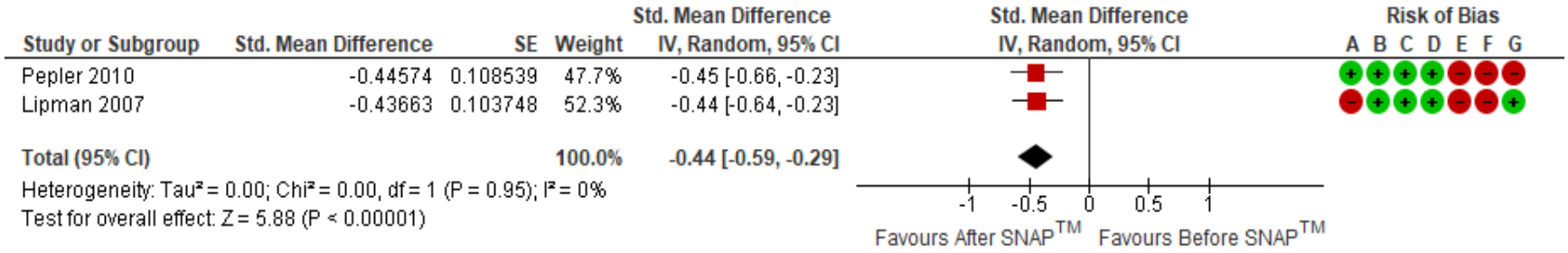
Higgins and Deeks (2011), suggested that the "tests for funnel plot asymmetry should be used only

when there are at least 10 studies included in the meta-analysis, because when there are fewer studies the power of the tests is too low to distinguish chance from real asymmetry" (Chapter 10; Section 10.4.3.1). Since the current Meta-Analysis had a maximum of four studies per outcome, it was not possible to test for the funnel plots asymmetry to determine the impact of publication bias on the overall effect sizes for all the study's outcomes. Therefore, the researcher refrained from performing any other tests to identify the publication bias as it was not possible or recommended.



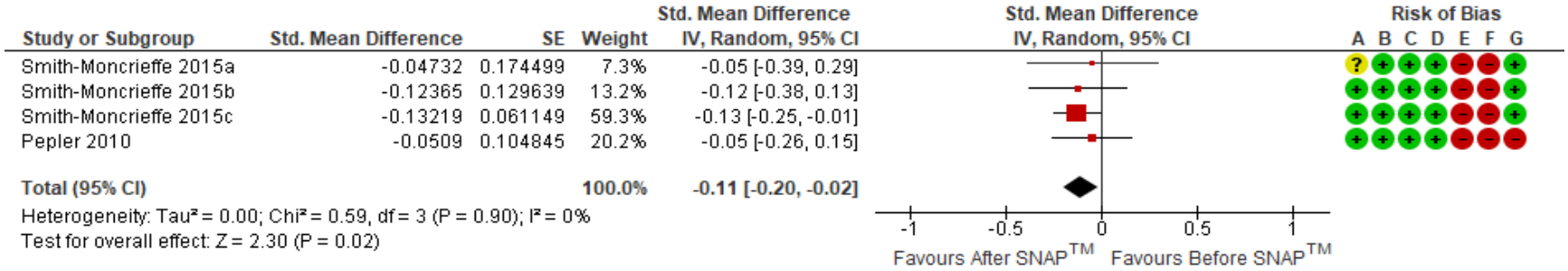
Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 20 Forest plot of CBCL-DSM ADHD (Post-Treatment).



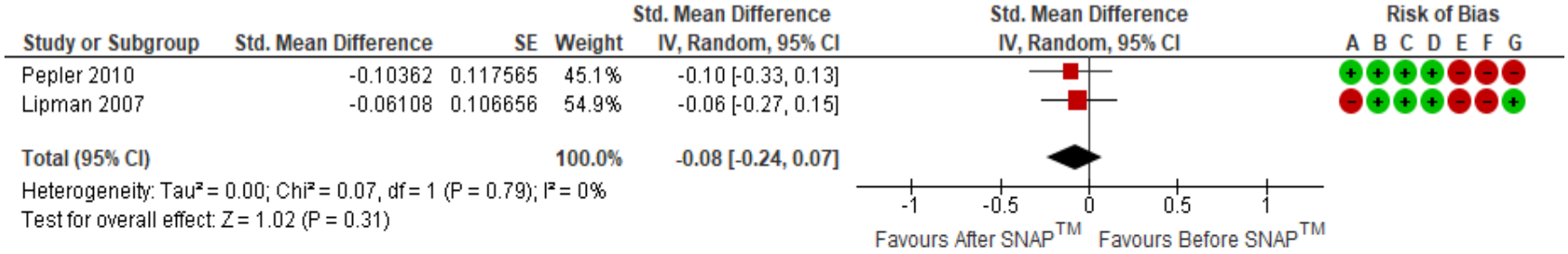
Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 21 Forest plot of CBCL-DSM ADHD (6-Month Follow-up).



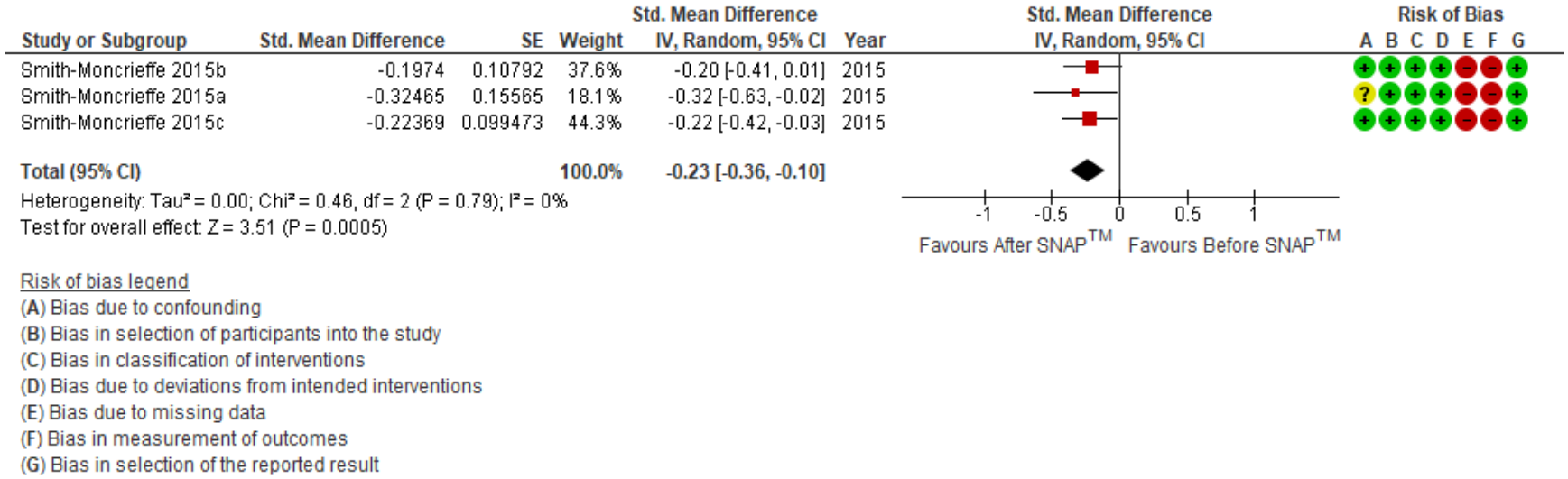
Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 22 Forest plot of TRF- DSM ADHD (Post-Treatment).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

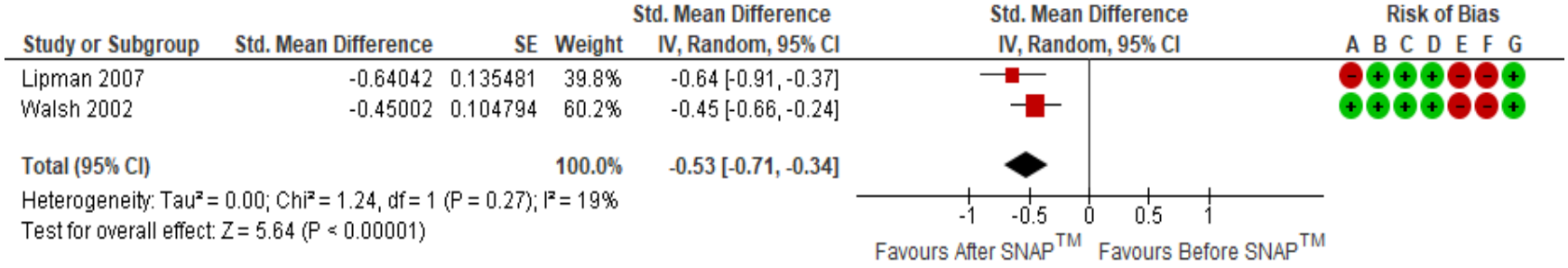
Figure 23 Forest plot of TRF- DSM ADHD (6 Months Follow-up).



Risk of bias legend

(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 24 plot of CBCL-Oppositional Subscale (Post-Treatment).



Risk of bias legend

(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 25 Forest plot of CBCL-Oppositional Subscale (6-Month Follow-up).

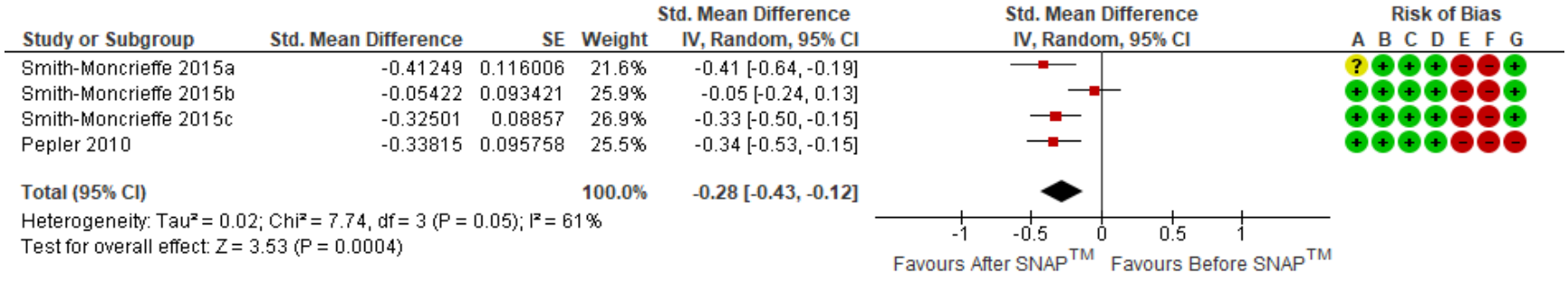


Figure 26 Forest plot of CBCL-Attention (Post-Treatment).

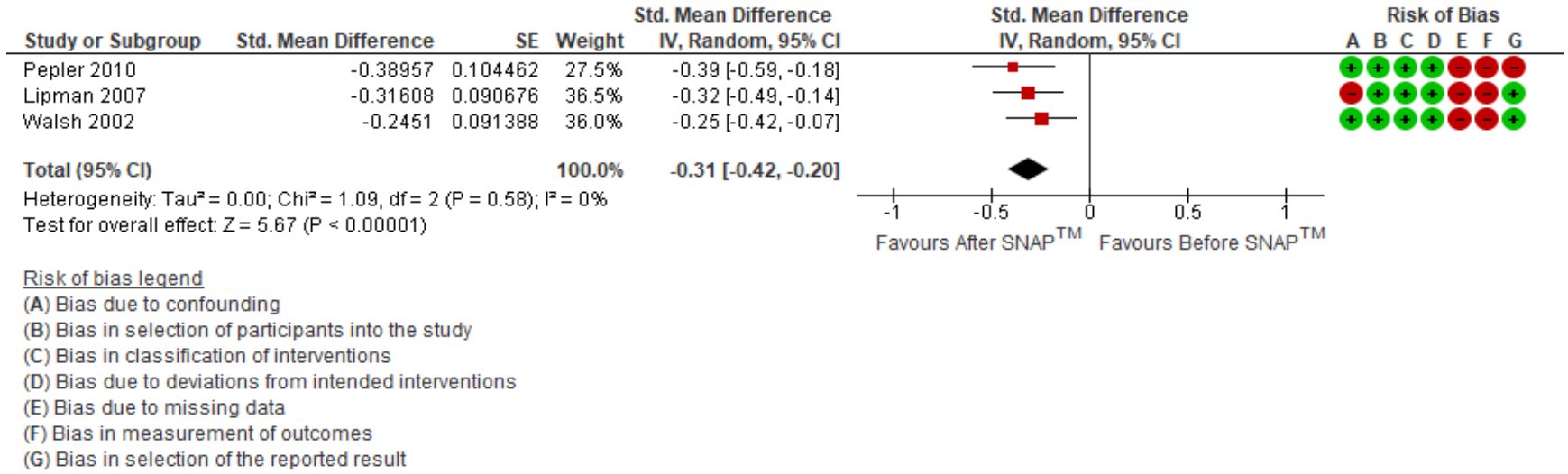
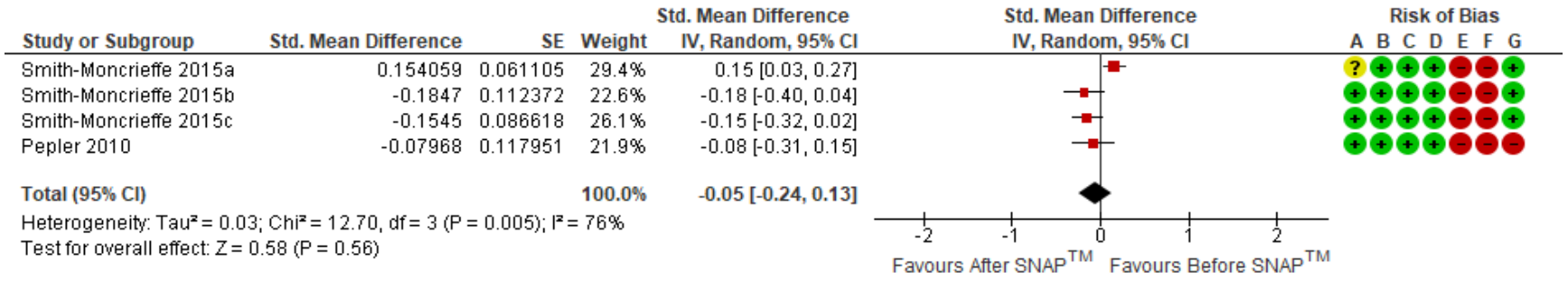


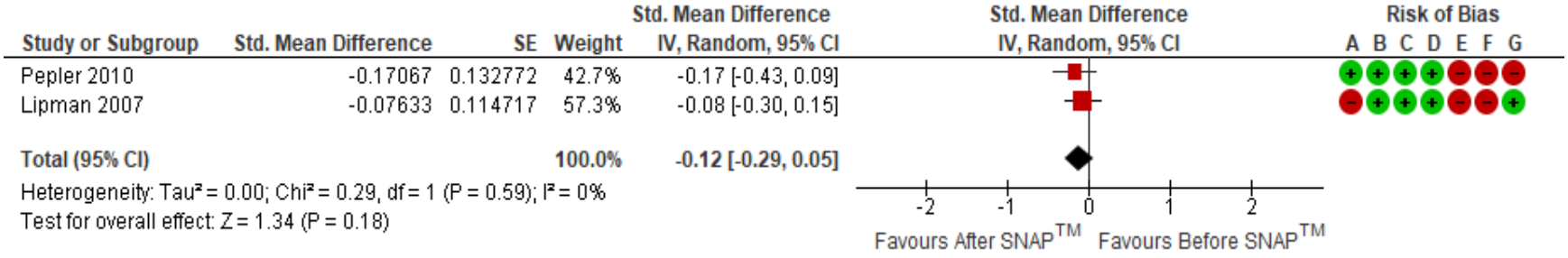
Figure 27 Forest plot of CBCL-Attention (6-Month Follow-up).



Risk of bias legend

(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 28 Forest plot of TRF-Attention (Post-Treatment).



Risk of bias legend

(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 29 Forest plot of TRF-Attention (6 Months Follow-up).

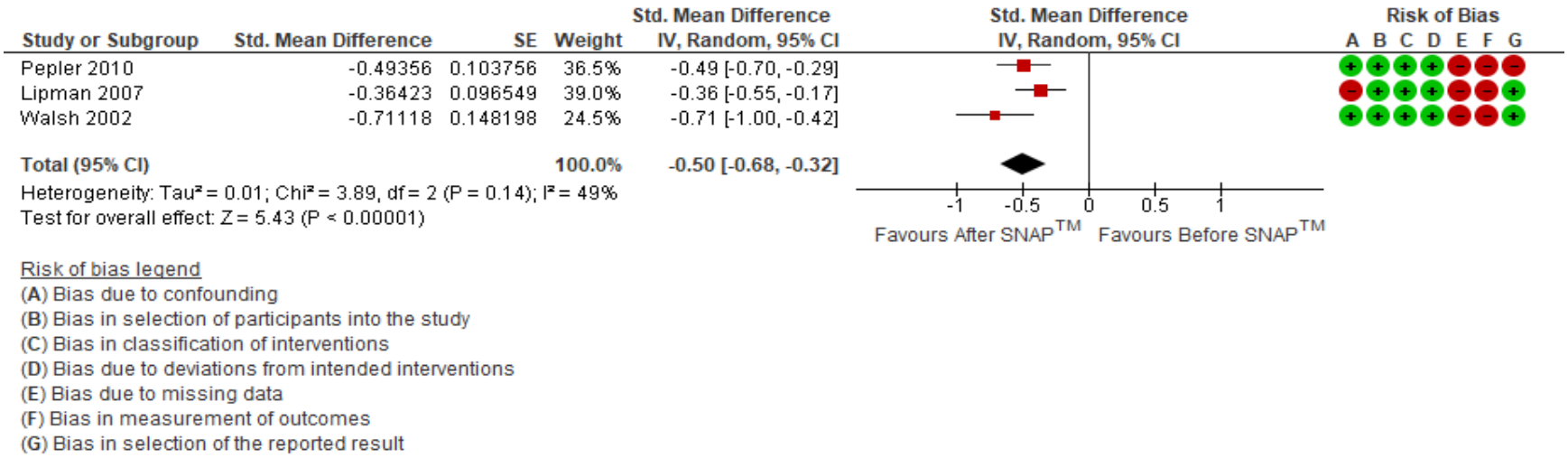


Figure 30 Forest plot of CBCL-Social Problems (6-Month Follow-up).

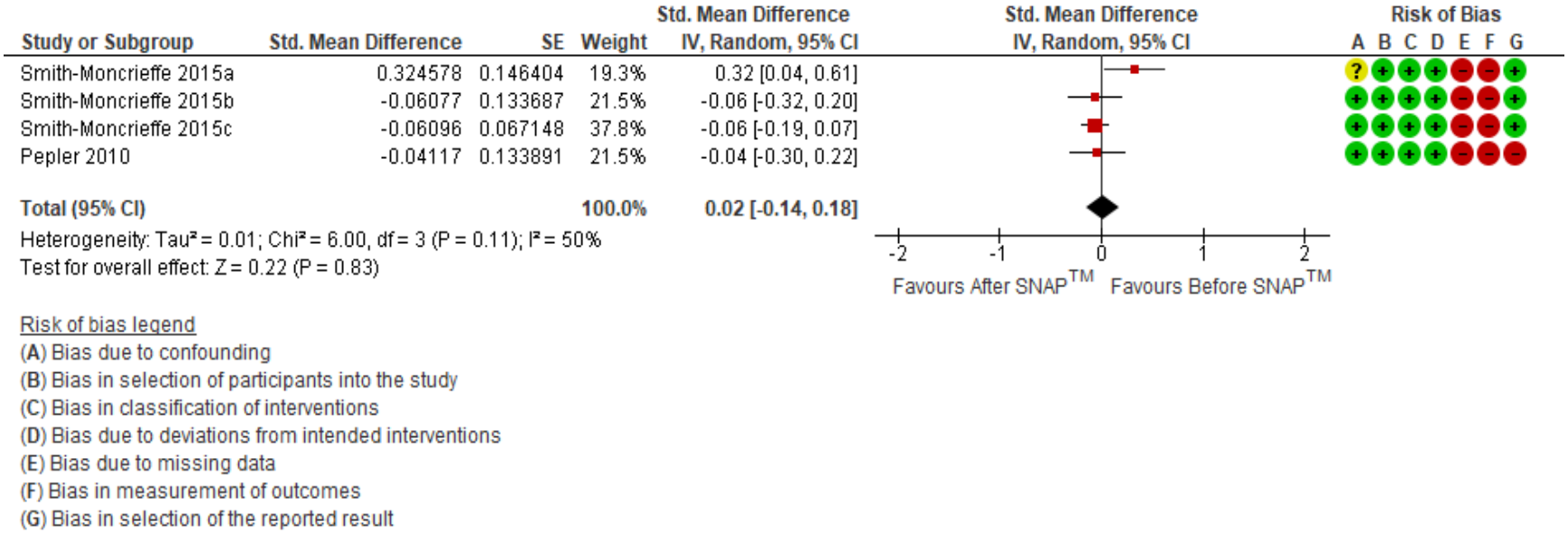


Figure 31 Forest plot of TRF-Social Problems (Post-Treatment).

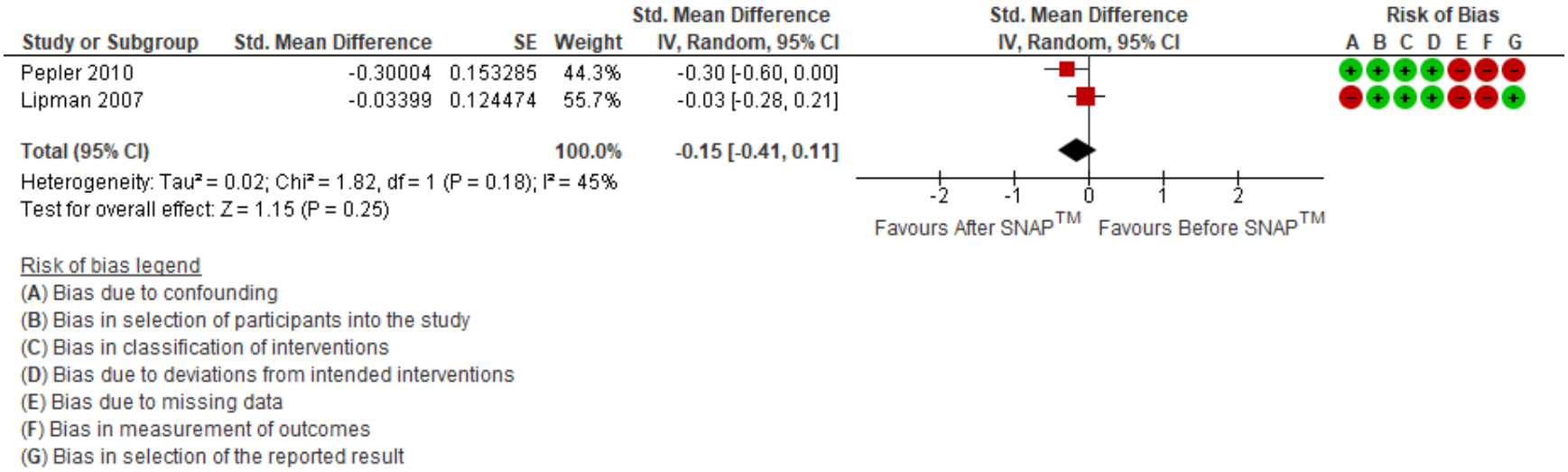


Figure 32 Forest plot of TRF-Social Problems (6 Months Follow-up).

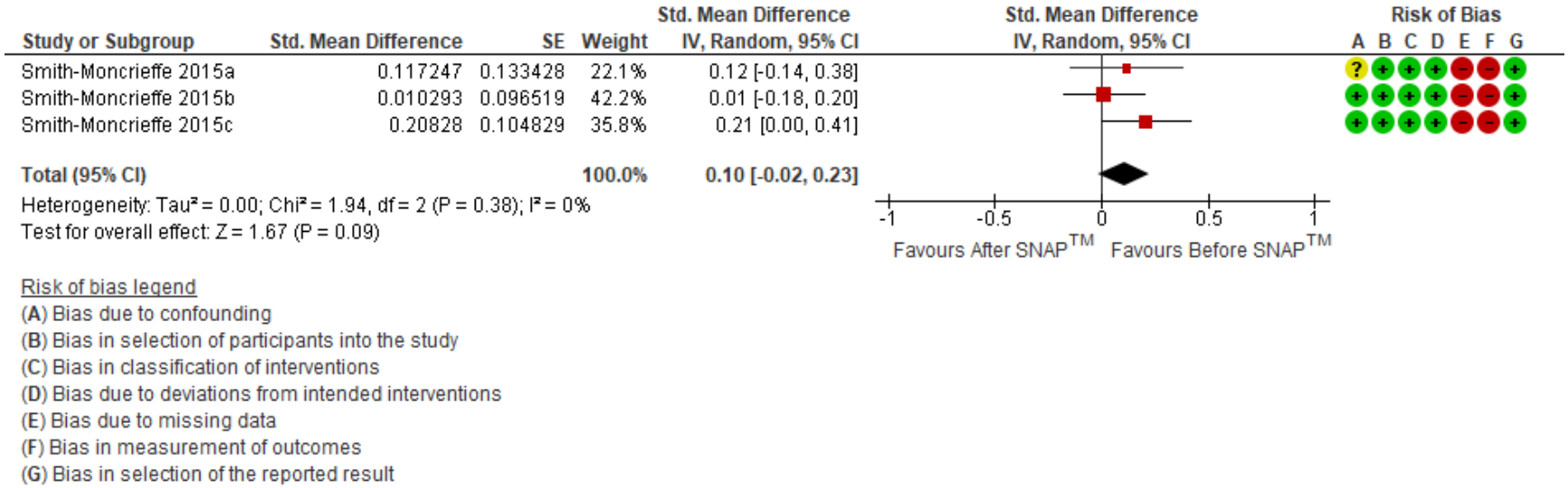


Figure 33 Forest plot of CBCL-Total Competence (Post-Treatment).

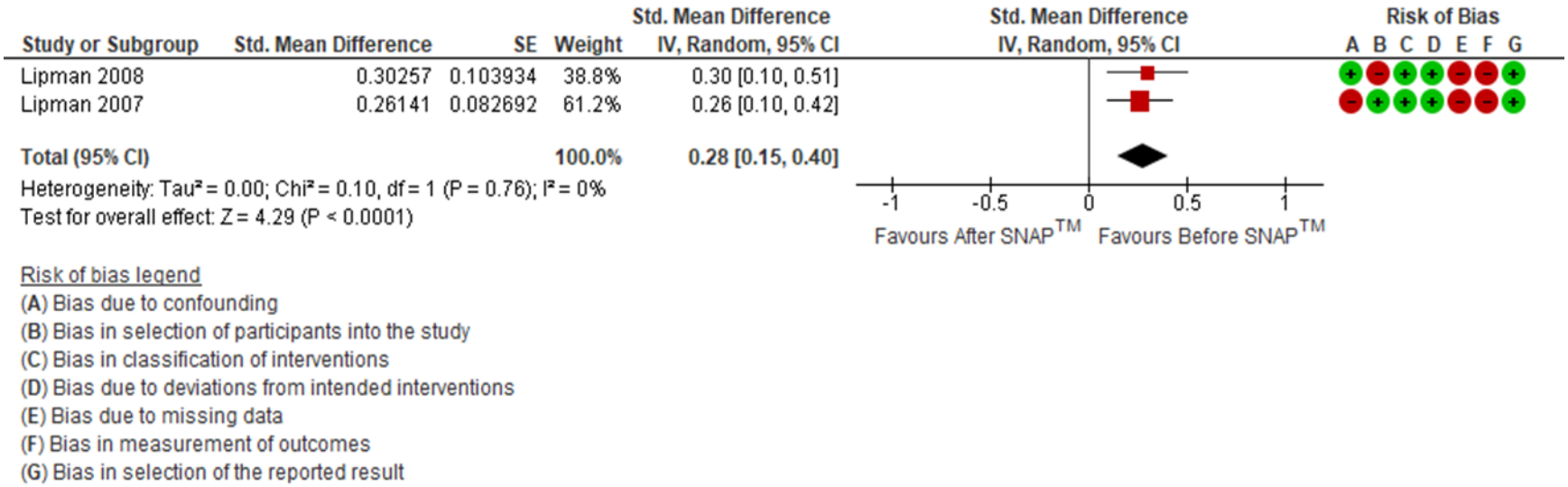


Figure 34 Forest plot of CBCL-Total Competence (6-Month Follow-up).

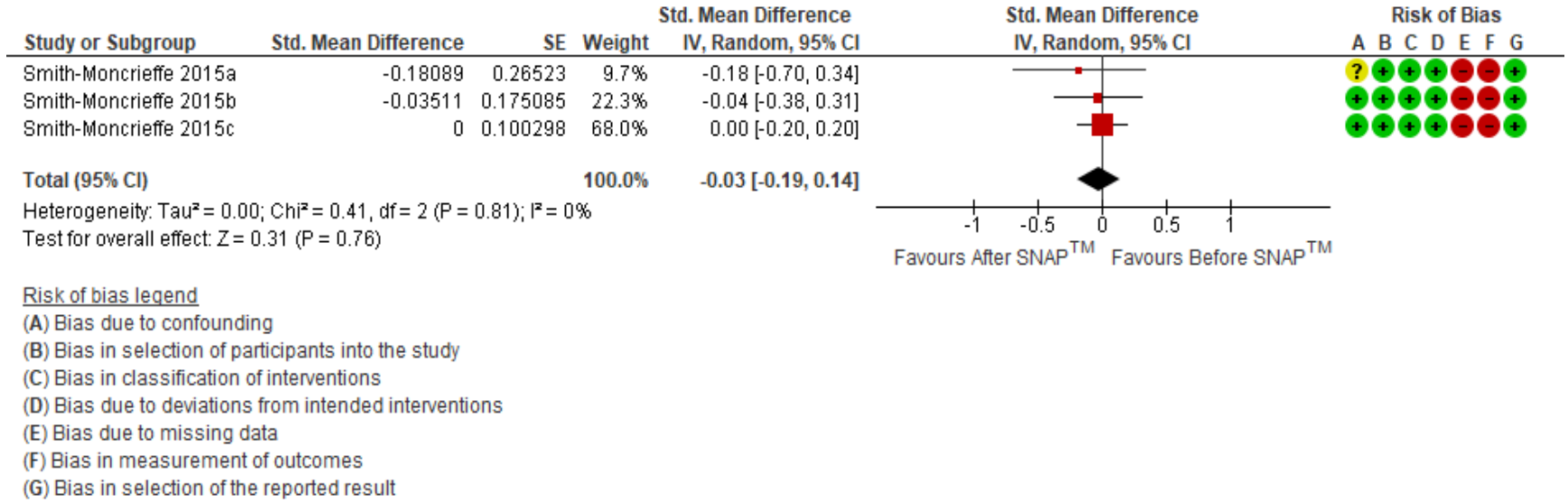


Figure 35 Forest plot of TRF-Adaptive Functioning (Post-Treatment).

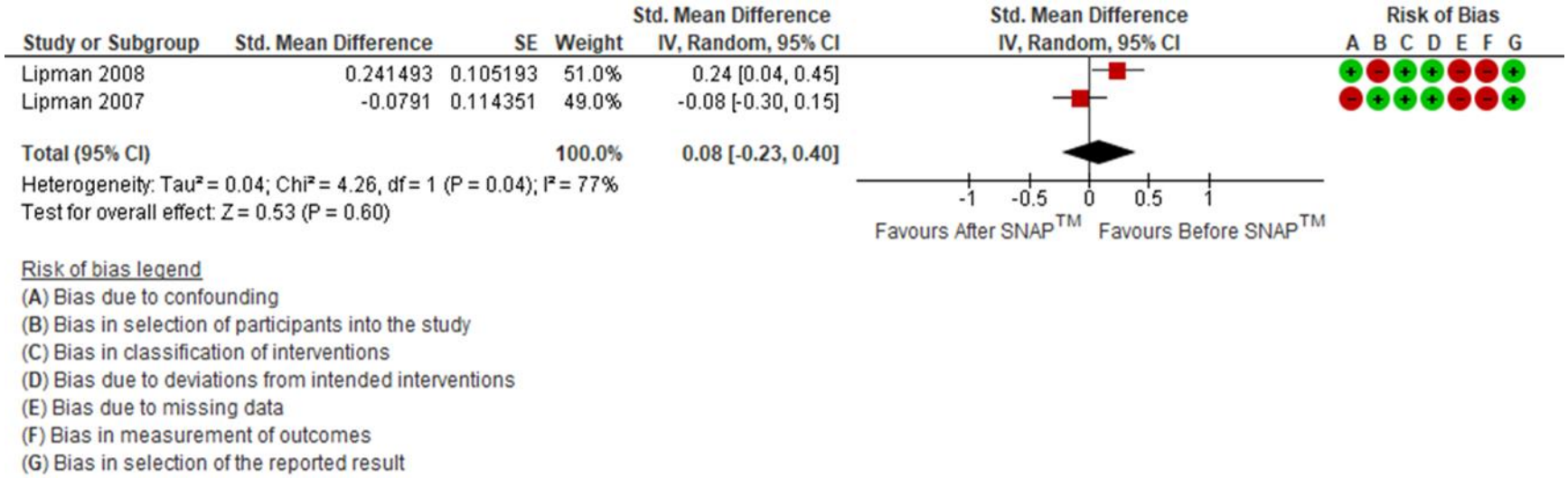


Figure 36 Forest plot of TRF-Adaptive Functioning (6-Month Follow-up).

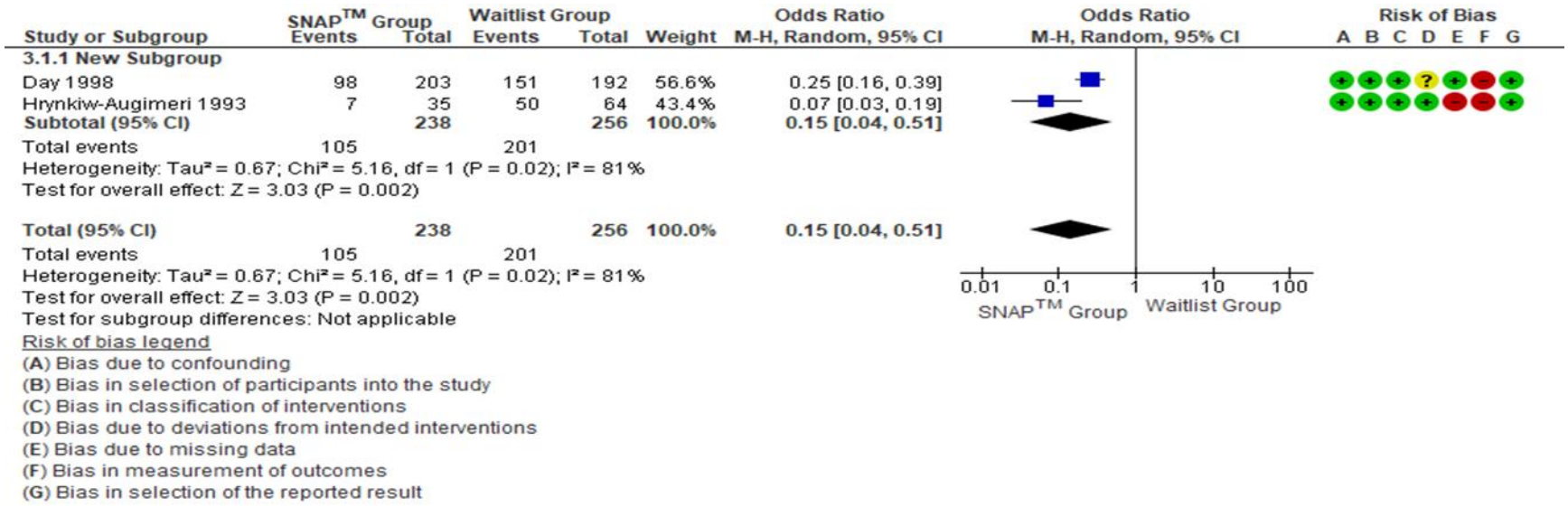


Figure 37 Forest plot of Official Criminal Records by the End of the Study.

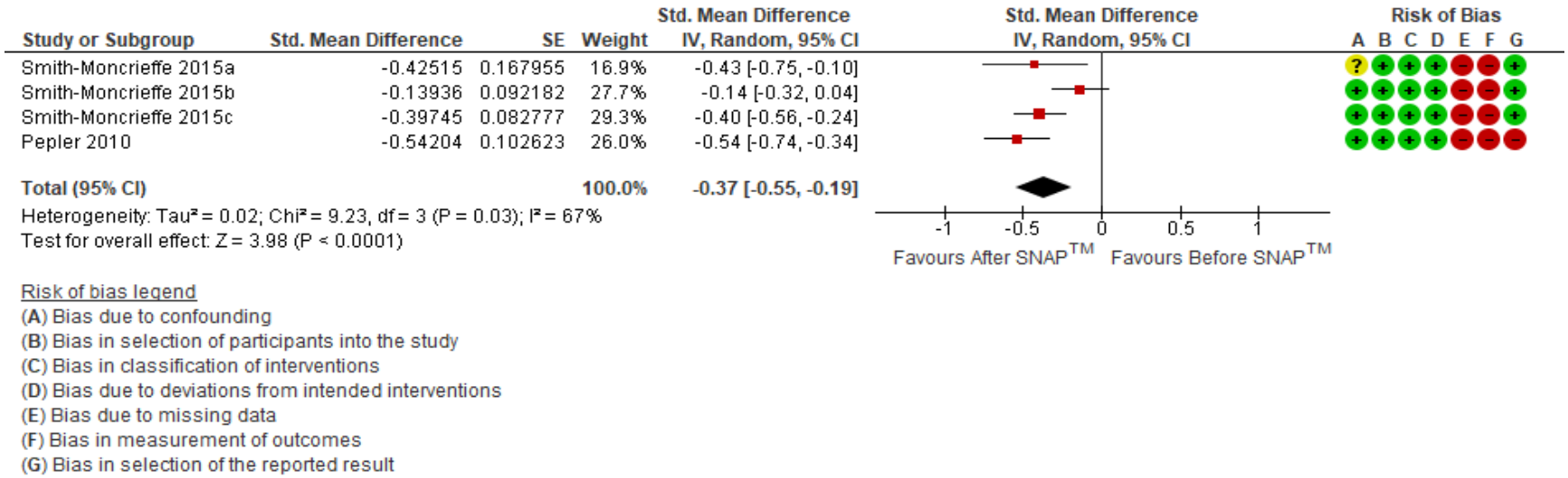


Figure 38 Forest plot of CBCL-Internalizing Problems (Post-Treatment).

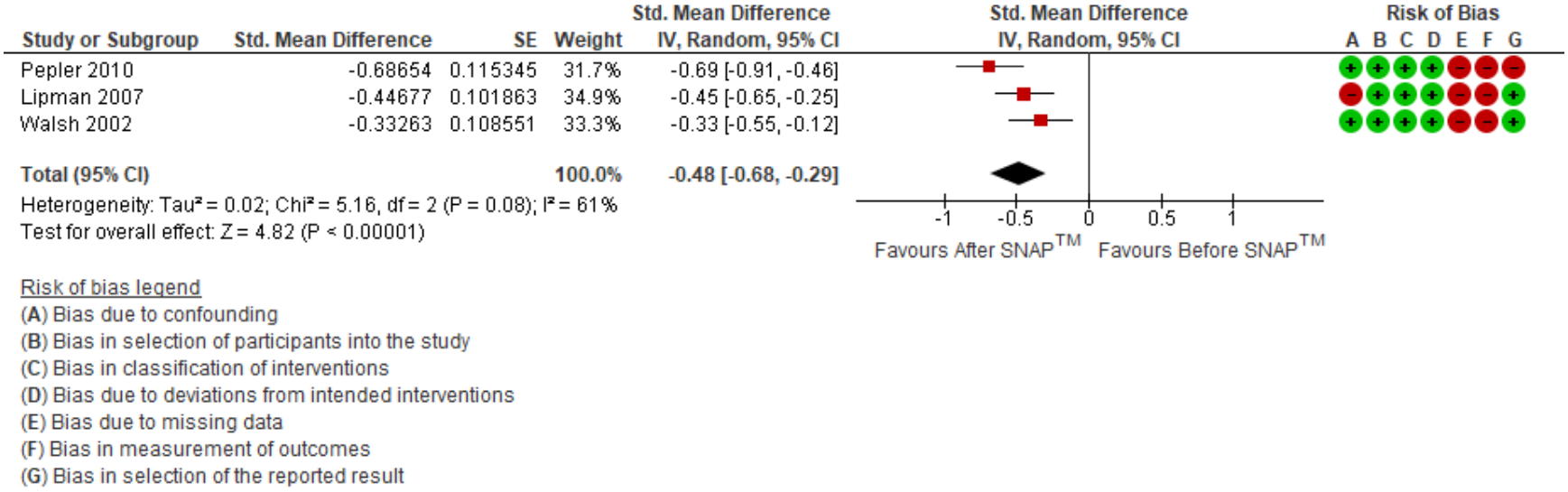


Figure 39 Forest plot of CBCL-Internalizing Problems (6-Month Follow-up).

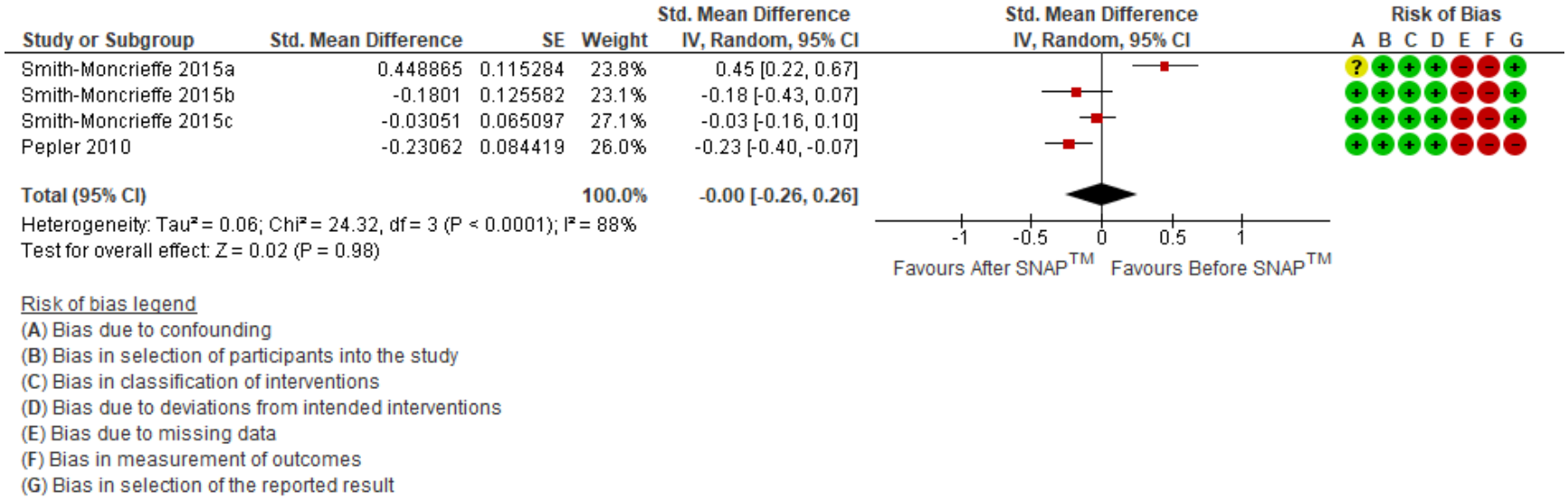


Figure 40 Forest plot of TRF-Internalizing Problems (Post-Treatment).

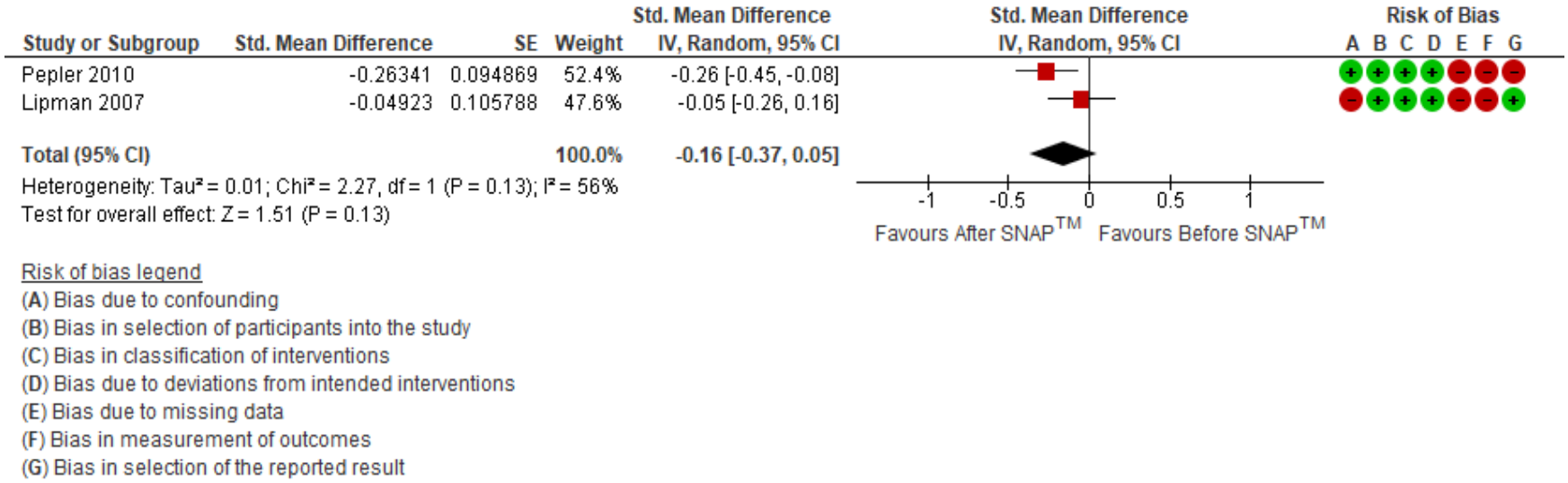


Figure 41 Forest plot of TRF-Internalizing Problems (6 Months Follow-up).

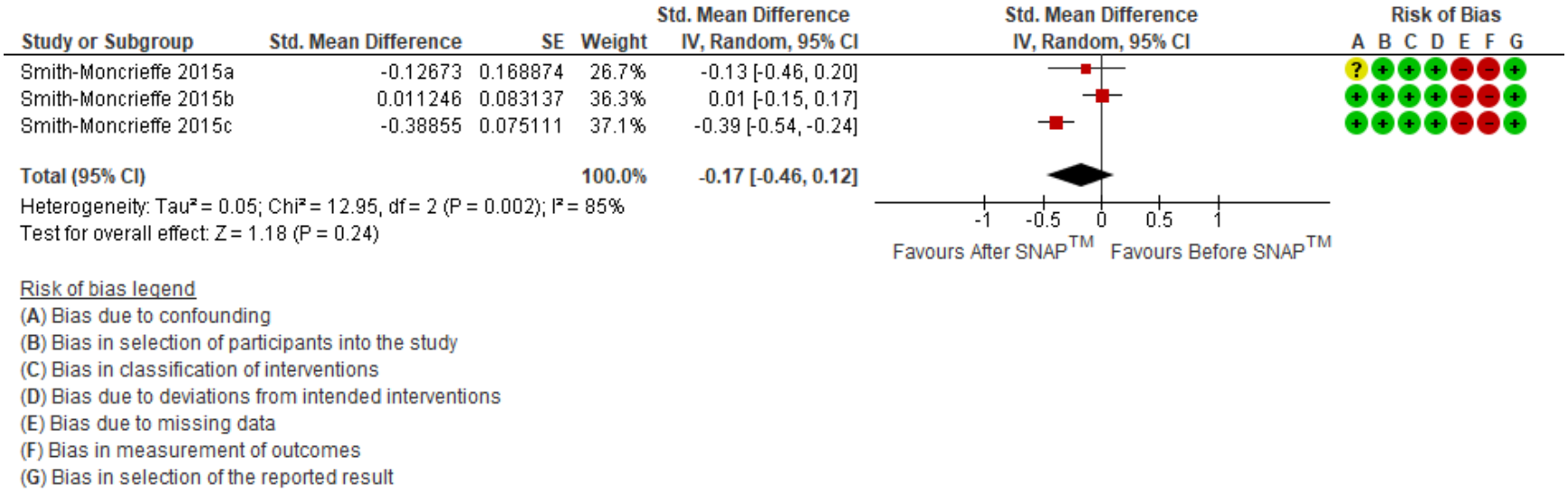


Figure 42 Forest plot of CBCL-Withdrawn-Depressed (Post-Treatment).

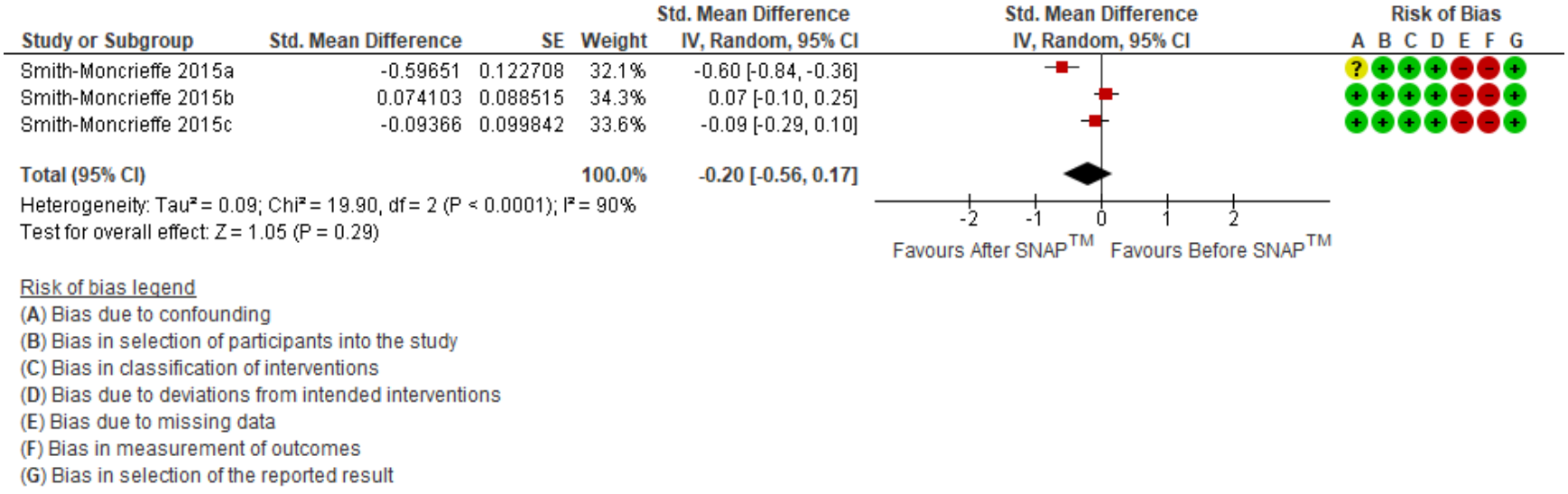
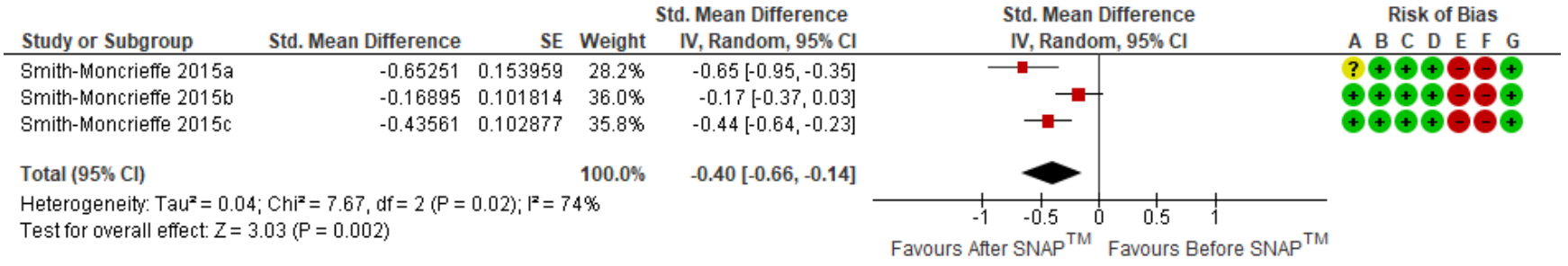
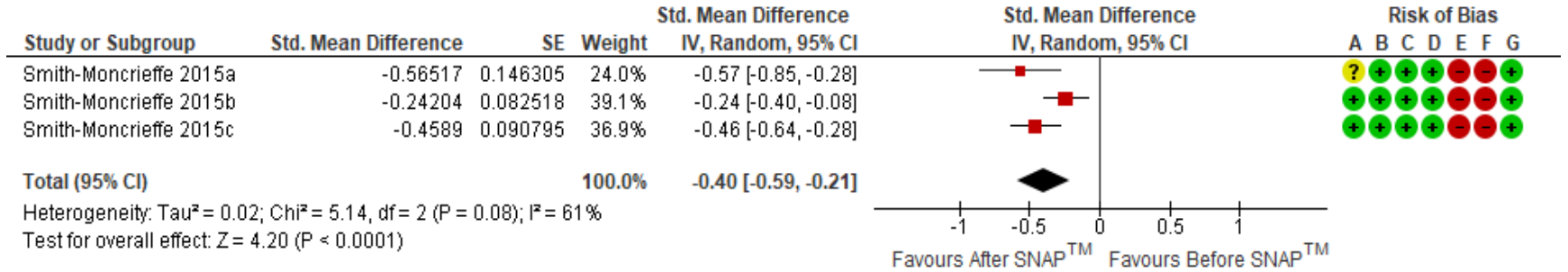


Figure 43 Forest plot of CBCL-Somatic Complaints (Post-Treatment).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 44 Forest plot of CBCL-Anxious-Depressed (Post-Treatment).



Risk of bias legend
(A) Bias due to confounding
(B) Bias in selection of participants into the study
(C) Bias in classification of interventions
(D) Bias due to deviations from intended interventions
(E) Bias due to missing data
(F) Bias in measurement of outcomes
(G) Bias in selection of the reported result

Figure 45 Forest plot of CBCL-Total Problems (Post-Treatment).

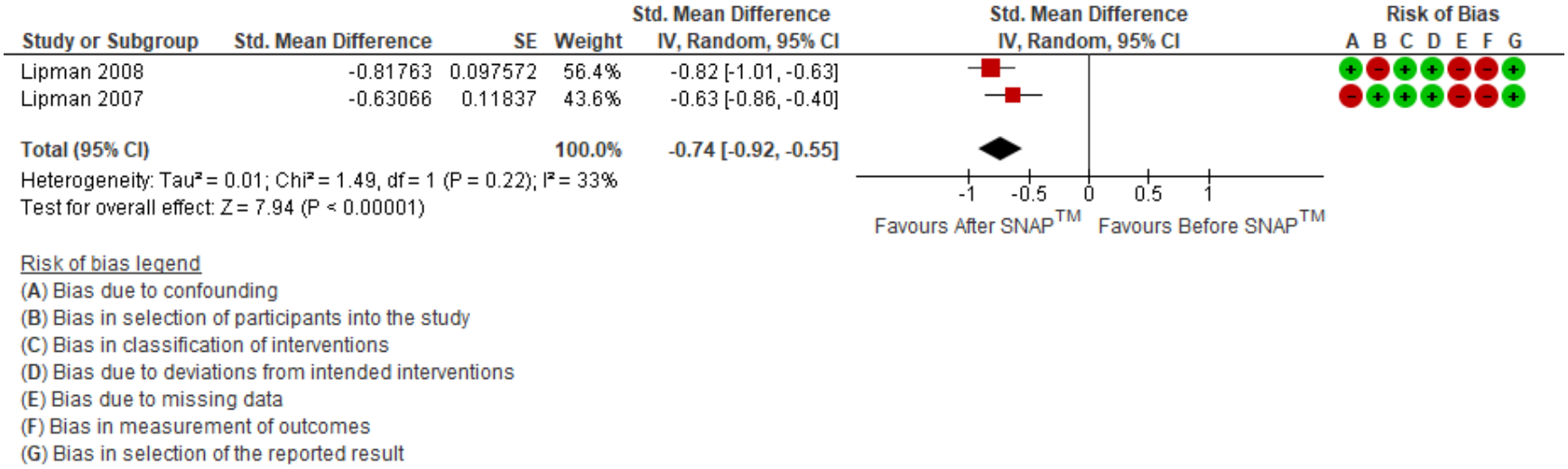


Figure 46 Forest plot of CBCL-Total Problems (6-Month Follow-up).

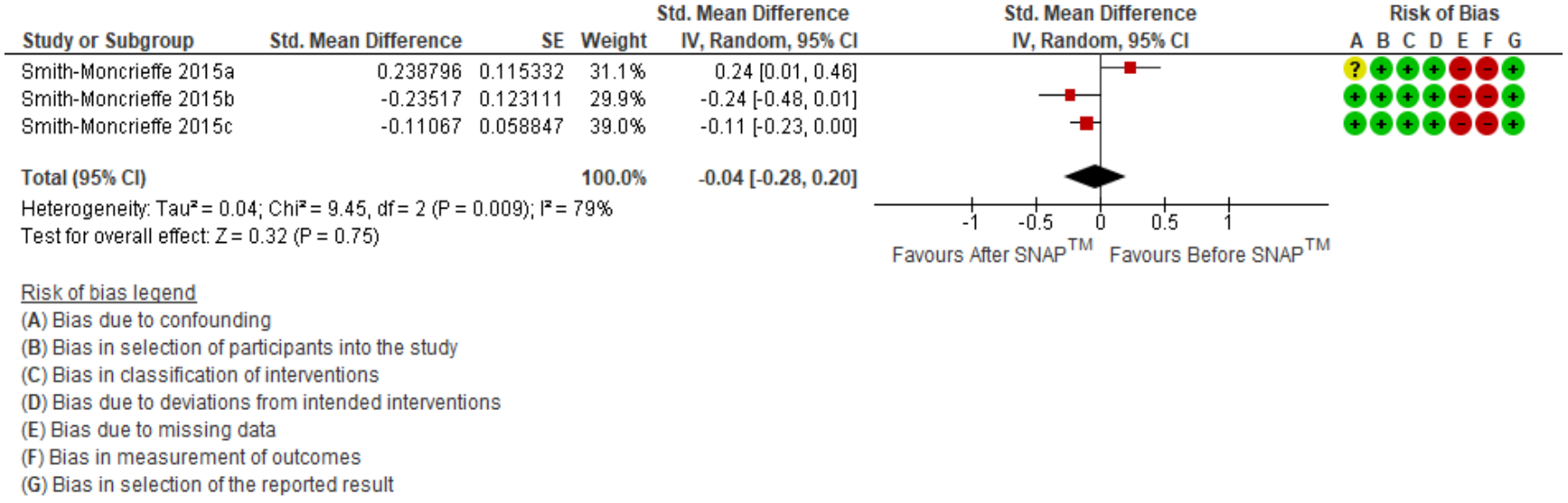


Figure 47 Forest plot of TRF-Total Problems (Post-Treatment).

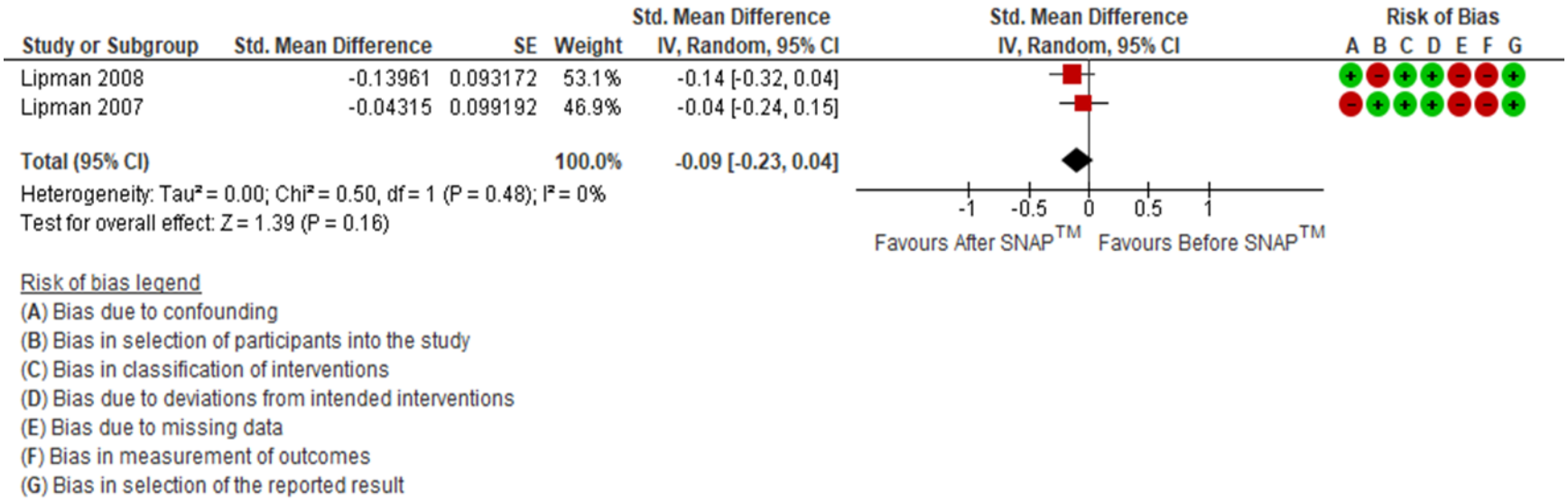


Figure 48 Forest plot of TRF-Total Problems (6 Months Follow-up).

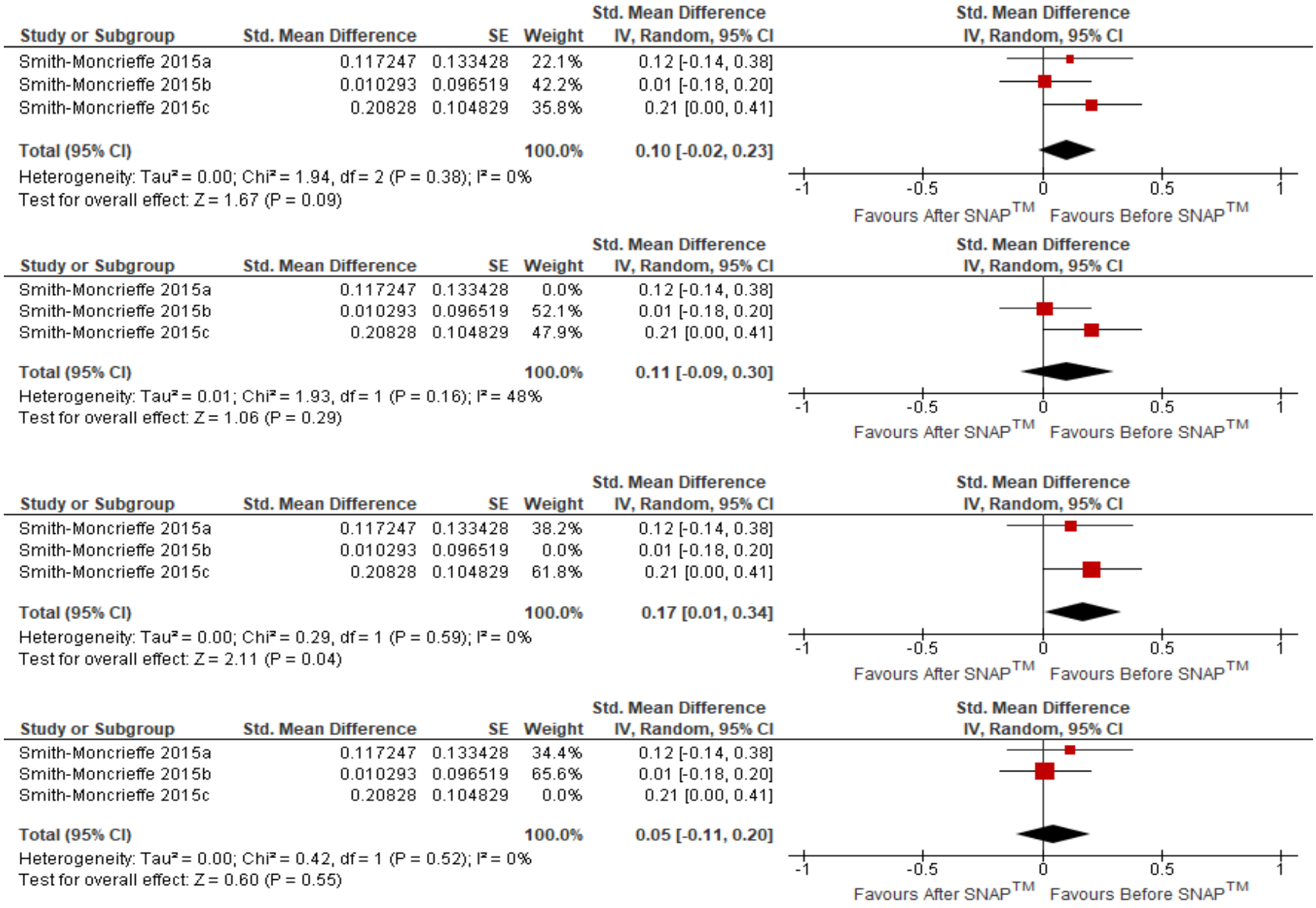


Figure 49 CBCL-Total Competence (Post-Treatment): Sensitivity Analysis.

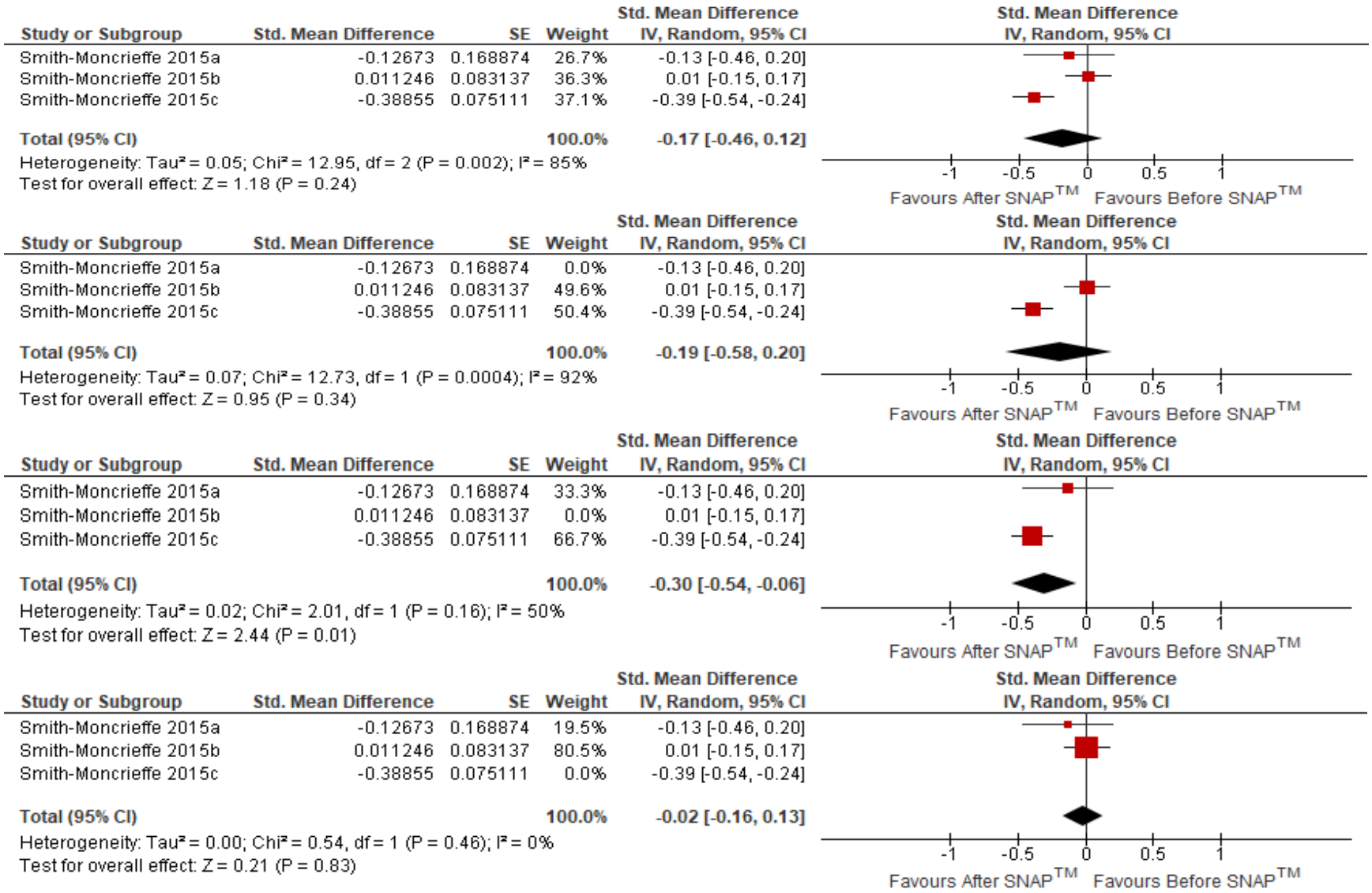


Figure 50 CBCL-Withdrawn-Depressed (Post-Treatment): Sensitivity Analysis.

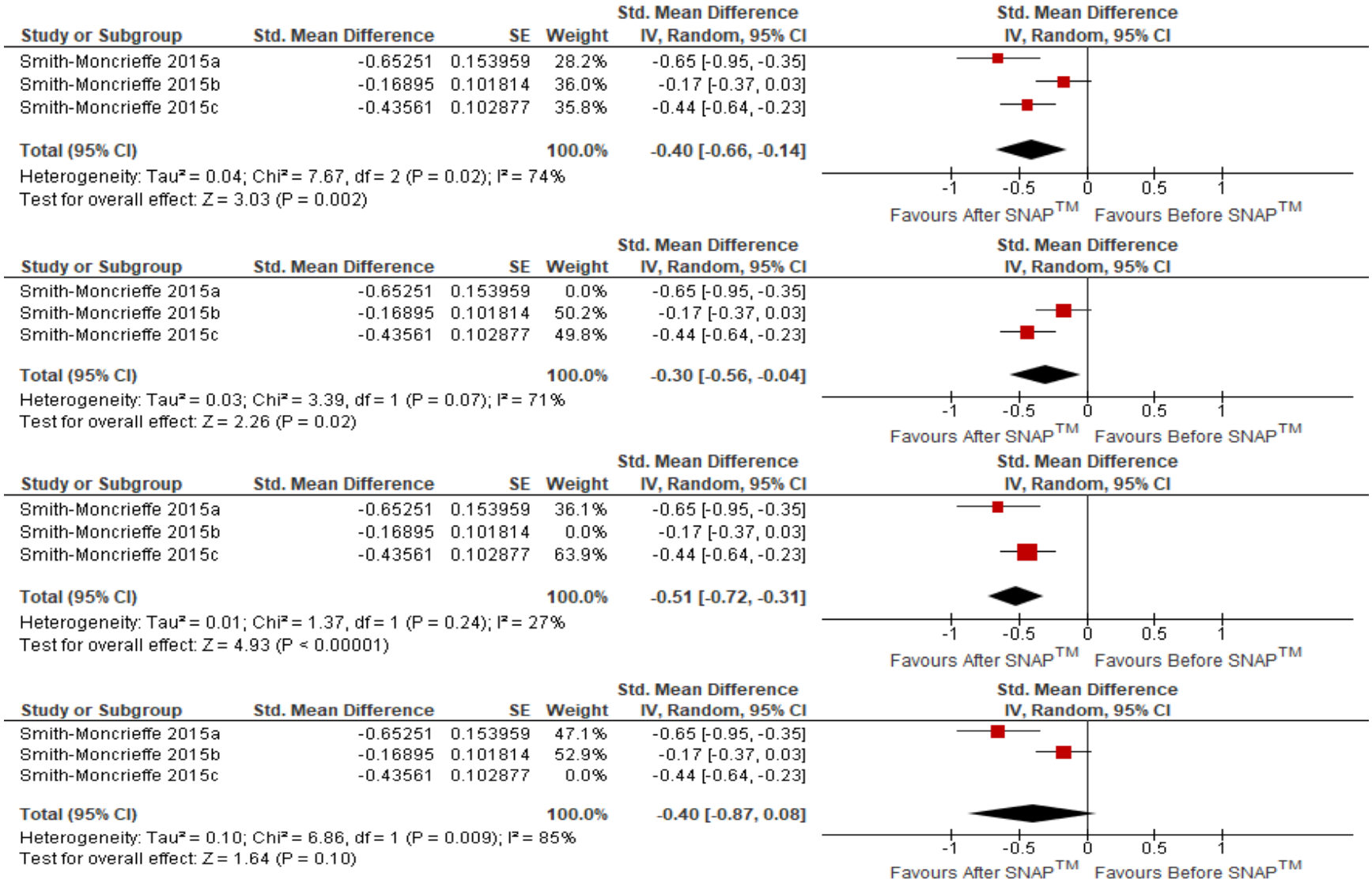


Figure 51 CBCL-Anxious-Depressed (Post-Treatment): Sensitivity Analysis.

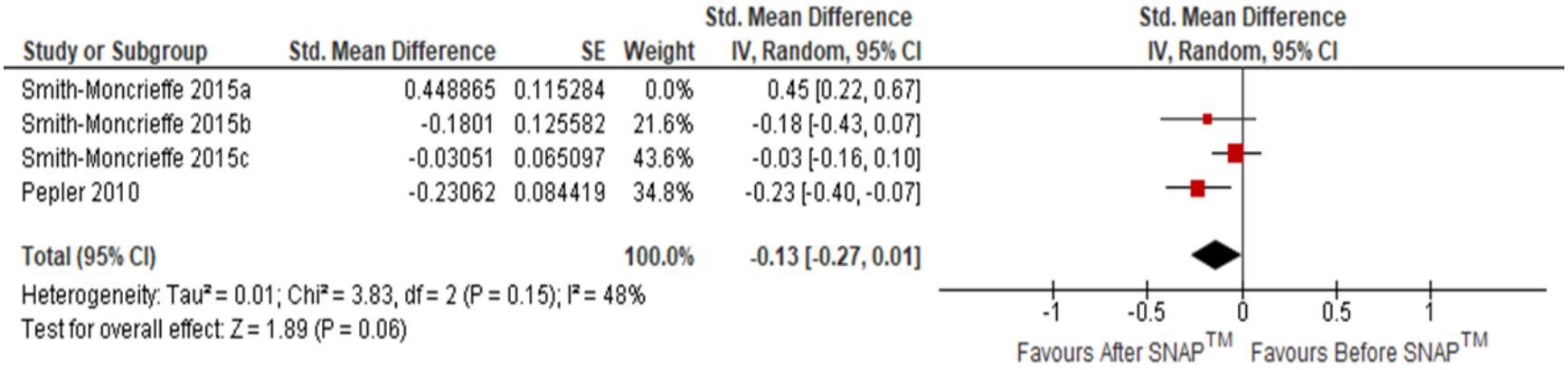


Figure 52 TRF-Internalizing Problems (Post-Treatment): Sensitivity Analysis.

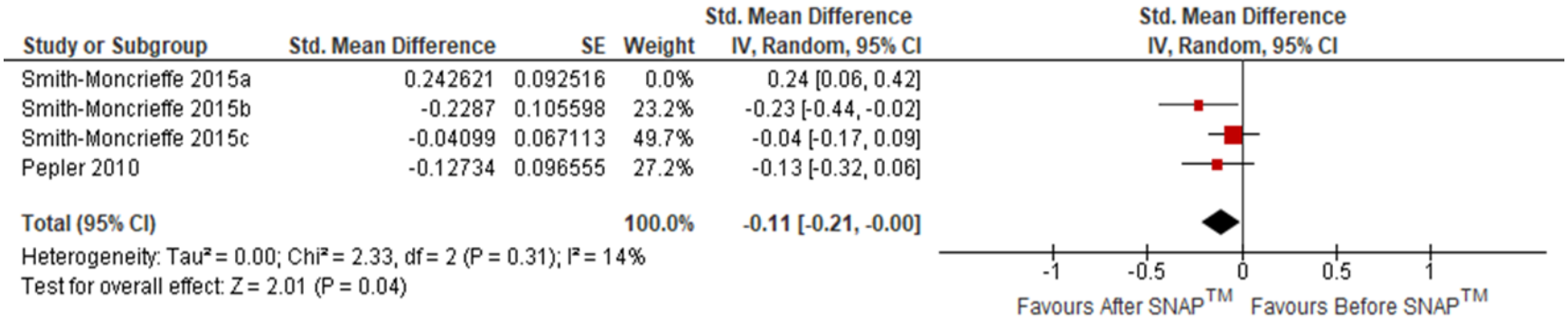


Figure 53 TRF-Externalizing Problem (Post-Treatment): Sensitivity Analysis.

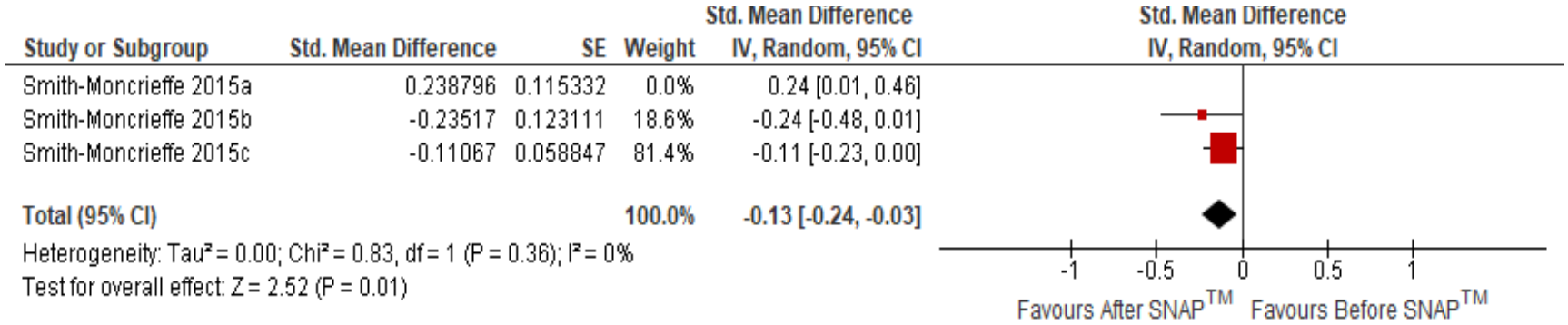


Figure 54 TRF-Total Behaviour Problem (Post-Treatment): Sensitivity Analysis.

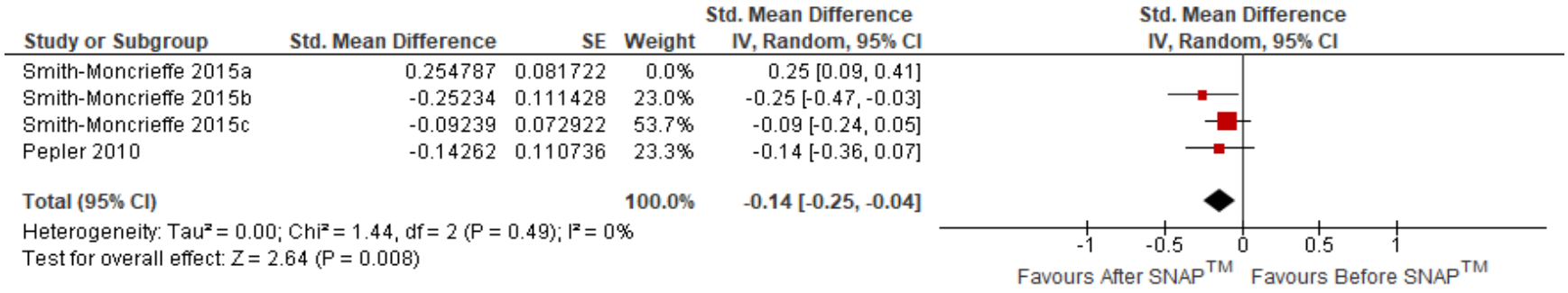


Figure 55 TRF-Aggression (Post-Treatment): Sensitivity Analysis.

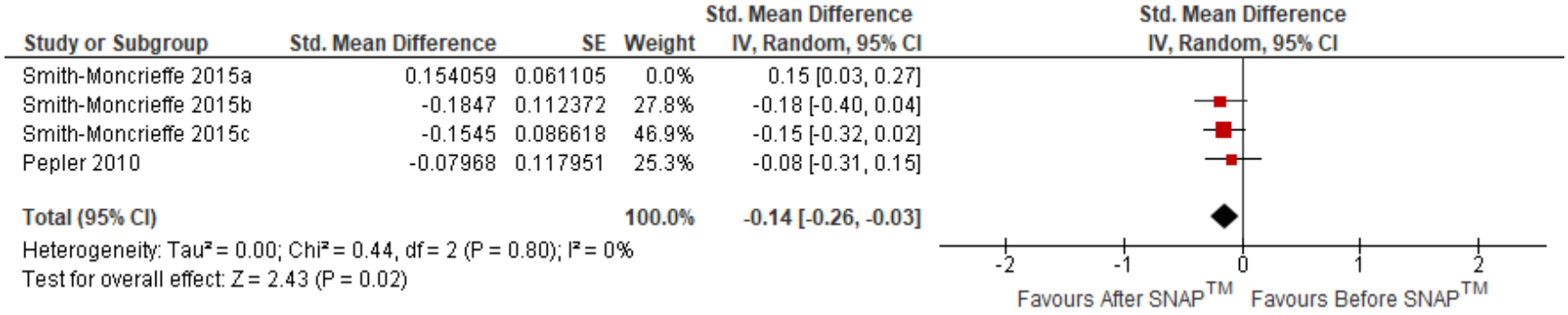


Figure 56 TFR-Attention (Post-Treatment): Sensitivity Analysis.

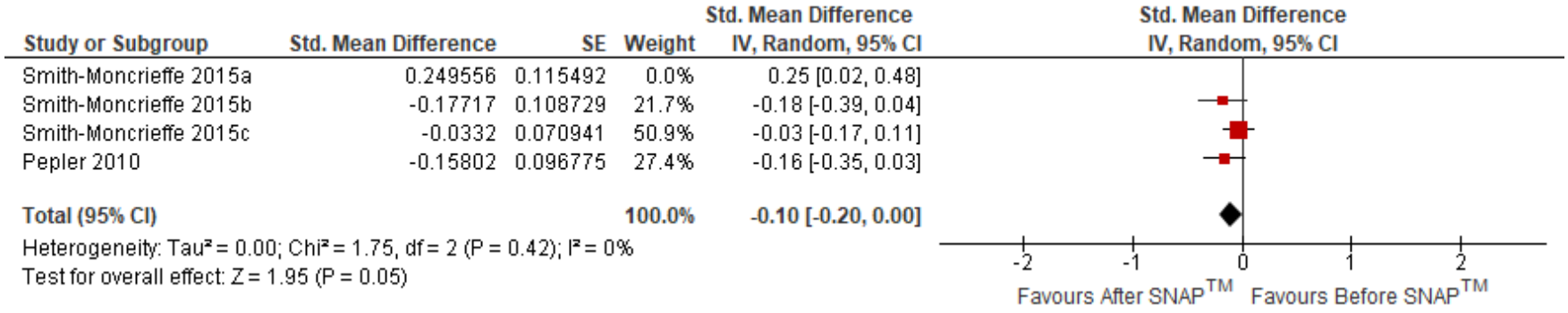


Figure 57 TRF-DSM Conduct Disorder (Post-Treatment): Sensitivity Analysis.

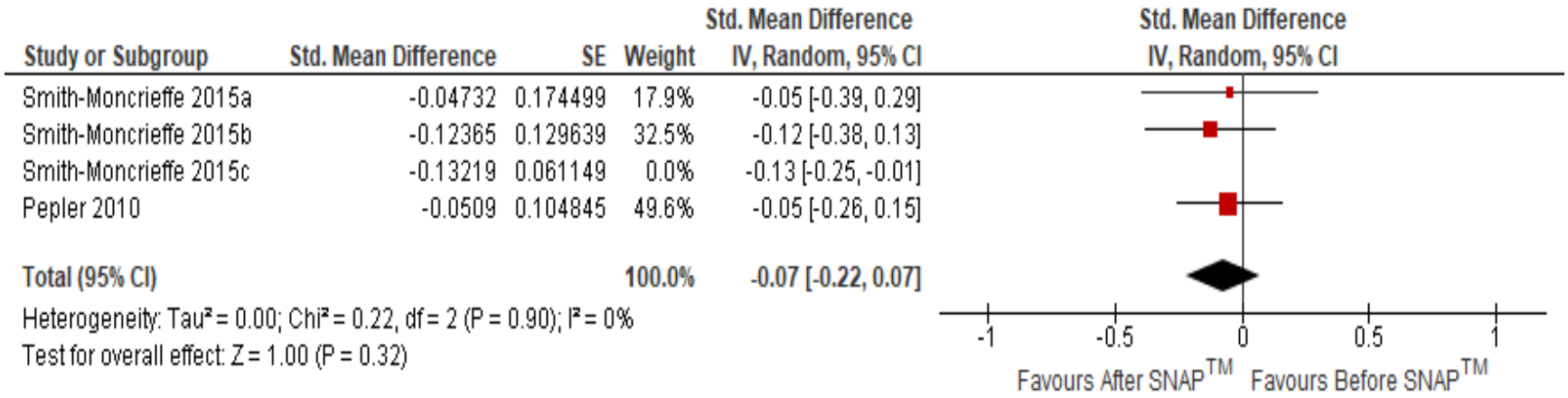


Figure 58 TRF- DSM ADHD (Post-Treatment): Sensitivity Analysis.

DISCUSSION

The current study aimed to determine the effect of the SNAP™ program on the level and trend of children's conduct disorder in community settings and to determine whether variations in treatment intensity and fidelity affect the program's outcomes.

The current study's results showed that SNAP™ was significantly effective in reducing criminal convictions for youth, who reached the age of criminal responsibility, after 15-18 months of their participation in the program (odds ratio= 0.44; p= 0.04; 95% CI: 0.21-0.95), according to the RCTs. This trend had continued up to eleven years of follow-up (Odds ratio= 0.15; p= 0.002; 95% CI: 0.04-0.51), according to the observational studies. Furthermore, there was a significant reduction in the children's behaviour problems and an increase in their social competence levels immediately after the three-month treatment, and this trend continued up to six-month of follow-up.

Only two randomized controlled trials were identified in the included studies, and these studies have only one common outcome, justice system involvements/criminal conviction by the end of the study, which was synthesized. There was no significant heterogeneity between these two RCTs. The first RCT, Augimeri et al. (2007), was conducted in Toronto, ON, Canada. On average, children who participated in this study were considered graduates where they had attended 8.95 (SD= 2.8) of the 12-SNAP™ child session. They had also joined another optional session, specifically the individual child befriending (M= 5.5; SD= 4.4). The parents/guardians of these children had attended, on average only 4.1 (SD= 3.8) of the 12-SNAP™ parent session and 2.5 (SD= 3.6) of the optional family counselling sessions. The treatment fidelity rate of this study was 81.6%. The second trial, Burke and Loeber (2015), was conducted in Pittsburgh, PA, USA. Children who participated in this study had attended, on average, 6.25 (SD= 4.3) of the 12-

SNAP^{TM} child session. Their average attendance rate was lower than 8-session, so on average, those children had not graduated from the program. However, they had attended, on average, 2.5 (SD= 4.2) of the optional individual child sessions, which is considered as the most clinically intensive supplemental child component of the SNAP^{TM} program (Augimeri et al., 2006). Their parents/guardians had attended an average of 5.02 (SD= 4.2) of the 12-SNAP^{TM} parent session, and 4.12 (SD= 6.9) of the optional family counselling sessions. Therefore, the parents/guardians in the second trial had attended, on average more sessions than the parents/guardians in the first trial. Furthermore, the treatment fidelity rate of Burke and Loeber (2015) study was 92%.

Koegl et al. (2008) found that attending nine or more sessions of SNAP^{TM} child sessions is significantly associated with having fewer criminal convictions later when reaching the age of criminal responsibility. Attending more sessions of the most clinically intensive child components and their parents/guardians participating in more parents' sessions, might have helped the children in the second trial have fewer criminal convictions by the time of follow-up.

Furthermore, two included observational studies had also reported on the children's justice system involvement/criminal conviction by the end of the study. The synthesized results from the observational studies had supported the results obtained from the RCTs. The heterogeneity between these two studies, however, was high, 81%, and the test was statistically significant (p= 0.02). The performed subgroup analysis showed that both studies had overall significant effects. However, the study/subgroup that was evaluated by a shared team of internal and external researchers, and findings based on the parents/guardians' reports had a larger effect size (Odds ratio= 0.07; 95% CI: 0.03, 0.19) than the other study/subgroup. Both of these studies, Hrynkiw-Augimeri et al. (1993) and Day (1998), were conducted in Toronto, ON, Canada. The participating children in Hrynkiw-Augimeri et al. (1993) study had attended an average of 9.4 (SD= 4.0) of

the12-SNAP^{TM} child session, which was higher than the attendance rate reported in the included RCTs. Their parents/guardians, on the other hand, had attended an average of 3.2 (SD= 4.3) of the 12-SNAP^{TM} parent session. Day (1998) did not report the children's and their parents'/guardians' attendance rates. Also, both studies did not report the fidelity of the treatment in their reports.

The rest of the current study's primary and secondary outcomes showed from small to large improvements in the children's anti-social behaviours immediately after 3-month of their participation in the SNAP^{TM} and after 6-month of follow-up.

For the primary outcomes, there were significant, small to moderate immediate effects in children's externalizing behaviour (SMD= - 0.41), rule-breaking (SMD= - 0.36), aggression (SMD= - 0.36), DSM-conduct disorder (SMD= - 0.41), DSM-ADDHD (SMD= - 0.30), ODD (SMD= - 0.23) and attention problems (SMD= - 0.28) based on the parents'/guardians' reports. Moreover, there was a significant, small immediate effect on children's DSM-ADDHD (SMD= - 0.11), based on the teachers' reports.

Furthermore, there were further significant, medium to large effects after a 6-month of follow-up in the children's externalizing behaviour (SMD= - 0.55), rule-breaking (SMD= - 0.66), aggression (SMD= - 0.73), DSM-conduct disorder (SMD= - 0.61), ODD (SMD= - 0.53), social problems (SMD= - 0.50), based on the parents/guardians' reports. Moreover, there were significant, small to moderate effects after a 6-month of follow-up in the children's DSM-ADDHD (SMD= - 0.44), attention problems (SMD= - 0.31) and total competence (SMD= 0.28), based on the parents/guardians' reports. Significant and small effects after a 6-month of follow-up were also observed by the teachers on the children's rule-breaking (SMD= - 0.12), DSM-conduct disorder (SMD= - 0.19) and DSM-ADDHD (SMD= - 0.08).

Furthermore, there was a significant, large improvement (SMD= - 0.86) in children's delinquent

behaviours after a 9-month follow-up period, according to their parents/guardians.

The primary outcomes that were not significant immediately after completing the program treatment were total competence (SMD= 0.10; p= 0.09), according to the parents/guardians, and externalizing behaviour (SMD= - 0.04; P= 0.71), rule-breaking (SMD= 0.04; p= 0.71), aggression (SMD= - 0.05; p= 0.65), DSM-conduct disorder (SMD= - 0.04; p= 0.68), attention problems (SMD= - 0.05; p= 0.56), social problems (SMD= 0.02; p= 0.83) and adaptive functioning (SMD= - 0.03; p= 0.76), according to the teachers. Furthermore, the primary outcomes that were not significant after a 6-month of follow-up were externalizing problems (SMD= - 0.14; p= 0.11), aggression (SMD= - 0.14; p= 0.07), attention problems (SMD= - 0.12; p= 0.18), social problems (SMD= - 0.15; p= 0.25) and adaptive functioning (SMD= - 0.08; p= 0.60), according to the teachers.

For the secondary outcomes, there were significant, moderate immediate effects in children's internalizing behaviour (SMD= - 0.37), anxious/depressed (SMD= - 0.40) and total behaviour problems (SMD= - 0.40), according to the parents/guardians. There was a further significant, small effect after a 6-month follow-up in the children's internalizing behaviour (SMD= - 0.48), and a significant, large effect (SMD= - 0.74) in the children's total behaviour problems, according to their parents/guardians.

The secondary outcomes that were not significant immediately after completing the program treatment were withdrawn/depressed (SMD= - 0.17; p= 0.24) and somatic complaints (SMD= - 0.20; p= 0.29), according to the parents/guardians. The treatment effect was not significant immediately or at the follow-up for two secondary outcomes as reported by the teachers specifically, internalizing behaviours ($SMD_{(Post)}$= - 0.00; p= 0.48, and $SMD_{(Follow\text{-}up)}$= - 0.16; p= 0.13), and total behaviour problems ($SMD_{(Post)}$= - 0.04; p= 0.75, and $SMD_{(Follow\text{-}up)}$= - 0.09; p=

0.16).

These results were somehow close to results obtained from similar studies that evaluated different treatments for children with challenging behaviours. For example, in their study of the effect of Good Behaviour Game on anti-social behaviour in school settings, Flower, McKenna, Bunuan, Muething and Vega (2014) found moderate to high significant effects immediately after the introduction of the treatment. A small reduction was further seen at follow-up periods, but that was not statistically significant.

There was heterogeneity in the effect sizes of some of the primary and secondary outcomes. Although the subgroup analyses that were performed had a low power as some subgroups had only one study, they might have explained some of the heterogeneity observed. Studies that were published, journal articles, conducted prospectively, located in Toronto, Ontario, Canada and used the CBCL-outcome measure had better and significant effect sizes. Besides, reviews with treatment fidelity more than 80% and their participants attended ≥ 8-session of the 12-Child SNAP^{TM} session, their participants had better outcomes, according to their teachers only. Moreover, studies that were obtained from the grey literature had small and non-significant effect sizes.

These findings are similar to the results of a review conducted by Hopewell, McDonald, Clarke, and Egger (2007) to determine the impact of grey literature in meta-analyses of randomized trials of health care interventions. The review found that published trials tend to have a 9% larger treatment effect than grey literature trials. Furthermore, results that are positive and significant are more likely to be published than non-significant ones (McAuley, Pham, Tugwell, & Moher, 2000). Moreover, prospective study designs usually have larger treatment effects than the retrospective ones because the prospectively collected data are generally complete, accurate and consistently

measured (Hulley, 2007).

Furthermore, studies that had treatment fidelity more than 80% had high adherence to the treatment manual. Besides, children who attended more than eight sessions of the 12-SNAP™ Child Group session had better outcomes because they are believed to have received the needed skills as planned and recommended by the program's inventor. All these factors helped in achieving significant improvements in the children's behaviour problems.

The sensitivity analyses showed that by removing the study that was conducted in the Cree Nation, Quebec, Canada, Smith-Moncrieffe (2015a), some effect sizes changed from non-significant to significant such as the overall immediate effect size of the TRF-externalizing behaviour (SMD= - 0.10; p= 0.04), TRF-aggression (SMD= - 0.14, p= 0.008), TRF-DSM-conduct disorder (SMD= - 0.10; p= 0.05), TRF-attention problems (SMD= - 0.14; p= 0.02) and TRF-total behaviour problems (SMD= - 0.13; p= 0.01). For the TFR-internalizing problem, by removing the Cree Nation study, the overall immediate effect size turned from non-significant to marginally significant (SMD= - 0.13; p= 0.06). Furthermore, by removing the same study, the overall immediate effect size of the TRF-DSM-ADDHD outcome changed from significant to non-significant (SMD= - 0.07; p= 0.32). Results obtained from Smith-Moncrieffe (2015a) should be interpreted with caution as the treatment fidelity of this study was low (53.4%), and its participants had attended less than 8-session of the 12 Child SNAP™ Group sessions. That was further explained by the study's author, where she acknowledged that "the Cree Nation site experienced a number of fidelity challenges related to quality of implementation, child/parent participation and matching children's risk with the appropriate treatment dosage.". The study also called for a standardized practice for reporting and collecting TRF data as the pre-post evaluation data were reported in different academic years by different teachers. That might have contributed to receiving imprecise data from two different

raters (Lipman et al., 2008; Smith-Moncrieffe, 2015a). The study was also judged as having a serious risk of bias as the outcome data were not available for all or nearly all participants, and participants were probably excluded due to missing data on variables needed for the analysis. Besides, there was not enough data to make a judgment on the "Bias due to confounding" domain. The same thing happened with the other two studies that were conducted by the same research team, specifically, Smith-Moncrieffe (2015b) and (2015c). By removing the study of Smith-Moncrieffe (2015b), the overall immediate effect size of the CBCL-total competence and CBCL-withdrawn/depressed turned from non-significant to significant (SMD= 0.17; p= 0.04) and (SMD= - 0.30; p= 0.01), respectively. Furthermore, by removing the study of Smith-Moncrieffe (2015c), the overall immediate effect size of the CBCL-anxious/depressed changed from significant to non-significant (SMD= - 0.40; p= 0.10). The first study, Smith-Moncrieffe (2015b), was conducted in Toronto, Ontario, Canada, and the second, Smith-Moncrieffe (2015c), was conducted in Edmonton, Alberta, Canada. Interestingly, both studies had treatment fidelity more than 80%, and their participants had attended ≥ 8-session of the 12 Child SNAP™ sessions. These studies were reports published in the grey literature. They were rated to have a serious risk of bias. The judgment was made as the outcome data were not available for all or nearly all participants, and participants were probably excluded due to missing data on variables needed for the analysis.

Missing critical data from the two included RCTs and receiving no answers from the authors in this regard made it difficult for the researcher to synthesize their results. The same problem happened with the observational studies, the selective reporting and missing critical data, in addition to the small number of the found studies could have affected the current study's results.

Furthermore, even though SNAP™ parents' sessions are one of the core components of the SNAP™ treatment, only one study had reported changes in parents' skills after they participate in

the program. Changes in “consistency in parenting” and “perceived ineffective parenting” were not statistically significant (Pepler et al., 2010). Further studies are needed to synthesize their results and identify the real change in their effect sizes.

Only one reviewer has evaluated the quality of the included studies. About 61.5% of the include observational studies were rated as having a moderate risk of bias. These studies represented 53.3% of the total included studies (i.e., RCTs + observational studies). The remaining four of the observational studies were rated as having a serious risk of bias. One of these studies was a journal article, and the remaining three were performed by the same research group and were found in the grey literature. Moreover, the researcher has rated the two included RCTs as having a high risk of bias because the RoB assessment tool for the RCTs has only three options for judging the risk of bias: low, high or unclear risk of bias.

CONCLUSION

The following represents the study's answers to its research questions based on the parents/guardians' reports and the official criminal records:

Q1. What is the effect of the SNAP™ program on the level and trend of children's conduct disorder in community settings?

The SNAP™ program made a significant moderate reduction in the children conduct problems (SMD= - 0.41; p< 0.00001; 95% CI: - 0.57, - 0.25; I^2= 58%; 4 studies; 231 participants) immediately after their participation in the program's three-month sessions. This trend has continued to a large, significant effect after the 6-month period of follow-up (SMD= - 0.61; p< 0.00001; 95% CI: - 0.79, - 0.44; I^2= 64%; 4 studies; 318 participants).

H1. Children participating in the SNAP™ program show a significant decrease in externalizing and internalizing behavioural problems as well as significant improvements in prosocial skills.

The SNAP™ program made significant, small to moderate immediate reductions in children's externalizing behaviours (SMD= - 0.41; p< 0.0001; 95% CI: - 0.59, - 0.22; I^2= 68%; 4 studies; 231 participants), rule-breaking (SMD= - 0.36; p= 0.0002; 95% CI: - 0.55, - 0.17; I^2= 71%; 4 studies; 231 participants), aggression (SMD= - 0.36; p< 0.00001; 95% CI: - 0.51, - 0.21; I^2= 45%; 4 studies; 231 participants), internalizing behaviours (SMD= - 0.37; p< 0.0001; 95% CI: - 0.55, - 0.19; I^2= 68%; 4 studies; 231 participants), anxious/depressed (SMD= - 0.40; p= 0.002; 95% CI: - 0.66, - 0.14; I^2= 74%; 3 studies; 191 participants). This trend has continued to significant, medium to large effects after 6-month follow-up in children's externalizing behaviours (SMD= - 0.55; p< 0.00001; 95% CI: - 0.73, - 0.36; I^2= 73%; 4 studies; 401 participants), rule-breaking (SMD= - 0.66; p< 0.00001; 95% CI: - 0.79, - 0.54; I^2= 0%; 3 studies; 246 participants), aggression (SMD= - 0.73;

p< 0.00001; 95% CI: - 0.86, - 0.60; I^2= 0%; 3 studies; 246 participants), internalizing behaviours (SMD= - 0.48; p< 0.00001; 95% CI: - 0.68, - 0.29; I^2= 61%; 3 studies; 184 participants) and social problems (SMD= - 0.50; p< 0.00001; 95% CI: - 0.68, - 0.32; I^2= 49%; 3 studies; 173 participants). For the total social competence, the program made a significant, small increase (SMD= 0.28; p< 0.0001; 95% CI: 0.15, 0.40; I^2= 0%; 2 studies; 193 participants), after a 6-month follow-up. Moreover, the program made a significant, large decrease in the children's delinquent behaviours (SMD= - 0.86; p< 0.00001; 95% CI: - 1.17, - 0.55; I^2= 88%; 2 studies; 386 participants), after a 9-month follow-up.

Therefore, the current study supports the hypothesis that children participating in the SNAP™ program showed a significant decrease in externalizing and internalizing behavioural problems as well as significant improvements in prosocial skills immediately after receiving the treatment or after 6-month of follow-up.

Q2. Do variations in treatment intensity and fidelity affect the program's outcomes?

According to the results of the subgroup analyses, the variations in treatment intensity and fidelity did not affect the program's outcomes, especially for the outcomes that were measured by the CBCL tool. However, they had affected the outcomes that were measured by the TRF tool. The studies with treatment fidelity more than 80% and their participants attended ≥ 8-session of the 12 Child SNAP™ Group sessions had better outcomes than the studies that had treatment fidelity less than 80%, and their participants attended less than 8-session of the 12 Child SNAP™ Group sessions. This conclusion should be interpreted with caution as the subgroup analyses had low power. Some subgroup analyses gave misleading results as some subgroups had only one study that had a serious risk of bias, and low treatment fidelity and intensity.

H2. Children who received intensive SNAP™ services would have fewer youth criminal convictions and more treatment gain at follow-up compared with those who received minimal core treatment components.

Based on the RCTs findings, the current study found that the participants who reached the age of criminal responsibility had a significant 56% less chance of getting involved in illegal activities compared to those who had not (Odds ratio= 0.44; p= 0.04; 95% CI: 0.21-0.95; I^2= 0%; 2 studies; 176 participants), based on a follow-up period ranged from 15- to 18-month. This significant trend continued to up to 85% reduction in the chance of getting involved in criminal activities based on a follow-up period ranged from 1 to 11 years from their admission to the program, according to the observational studies (Odds ratio= 0.15; p= 0.002; 95% CI: 0.04-0.51; I^2= 81%; 2 studies; 494 participants). However, the researcher was not able to determine the effect of the program's intensity as there were no variations in the effect sizes of the two included RCTs. For the two observational studies that reported this outcome, there was missing information about the treatment intensity that prevented the researcher from determining its effect on the children's criminal activities.

In summary, the current study differs from other studies that evaluated the SNAP™ program as it identified, assessed and summarized the findings of all relevant individual studies over the SNAP™ program. Therefore, it made the available evidence more accessible to interested professionals and decision-makers. The results of the current study help to plan research and frame guidelines for using the SNAP™ program to treat children with conduct disorder.

RECOMMENDATIONS

The following represents the main recommendations the researcher made from the current study:

1. The researcher recommends the adoption and use of the SNAP™ program to treat children with challenging behaviours as it has the potential of preventing future engagements in criminal activities.
2. The researcher recommends organizations interested in implementing the SNAP™ program to ensure high levels of adherence to the treatment manual to achieve high fidelity rates that, according to the study's findings, have a significant effect on the program's outcome.
3. The researcher also recommends the SNAP™ users, i.e., children and their parents/guardians, maintain a high attendance rate to the SNAP™ core components. As well as attend more sessions of the most clinically intensive optional SNAP™ components, i.e., individual child befriending and family counselling, as they have significant effects on altering the child's antisocial behaviours.
4. As the current study is a way of objectively summarizing the SNAP™ research evidence, the researcher recommends the CDI, clinicians and researchers consider the findings of the present study when planning for research and framing guidelines for the SNAP™ program adoption and implementation.
5. Although the study did not systematically review the SNAP™ program literature, there is a need for well-conducted trials as there were only two RCTs identified with a high risk of bias in the SNAP™ Research and Program Evaluation Studies Summary Chart (2019).
6. Prospective evaluation studies of the SNAP™ program should follow-up participating children for a more extended period as the current study found very few studies reporting

about the treatment effect after a 6-month follow-up.

7. Information about the parenting skills was not reported in the included SNAP™ evaluation studies, even though the SNAP™ parent group session is one of the two main components of the treatment. The researcher recommends reporting these results to identify the effect of the SNAP™ parent group sessions on the children's parents/guardians.
8. The researcher recommends performing a Systematic Review and Meta-Analysis study to determine whether the current research has missed any other SNAP™ evaluation studies.
9. Performing a Meta-Analysis study is overwhelming and time-consuming, especially for a novice scholar. Therefore, the researcher recommends the involvement of a research team that includes a content expert, a statistician and a couple of reviewers in performing such kind of studies.

ETHICAL CONSIDERATIONS

The study is "a type of research synthesis" (Cooper & Dent, 2011, p. 417), where the summary statistics of the primary individual studies were collected and synthesized to determine the overall SNAP^{TM} effect size. Studies that met the inclusion criteria were published and unpublished SNAP^{TM} evaluation studies. Furthermore, studies based on data collected from previous studies were excluded to avoid overestimating the program's effect size. Moreover, the study was submitted to and approved by the University of Manitoba Health Research Ethics Board (HREB) before its implementation.

LIMITATIONS OF THE STUDY

The study did not execute a comprehensive systematic review of the literature due to resource limitations. The studies' search method was restricted only on a SNAP^{TM} Research and Program Evaluation Studies Summary Chart, prepared by the CDI (2019). However, it assumed that the CDI as a developer, licensor and quality assurance approver is aware of and knowledgeable of all SNAP^{TM} research and evaluation studies since its inception in 1985. The CDI continuously updates this summary chart with published and unpublished SNAP^{TM} studies. The study further used informal channels of communication to find other published/unpublished internal SNAP^{TM} evaluation reports.

Furthermore, the study was limited to studies that met the inclusion criteria. Only quantitative studies and studies that directly evaluated the program effectiveness were included. Finally, even though the extracted data for the study was done by using a standardized data extraction form, the whole process was carried out by one person (i.e., the researcher). The researcher followed specific criteria and guidelines prepared for extracting primary summary data by the Cochrane Handbook of Systematic Reviews of Interventions (2011). Higgins and Deeks (2011), however,

recommended that at least two reviewers perform this process to minimize potential errors and biases resulted from a single reviewer.

POTENTIAL IMPLICATIONS OF THE STUDY

The study was the first Meta-Analysis study that measured the effectiveness of the SNAP™ program. It tried to overcome the limitations of the independent studies by increasing the number of participants, and accordingly, the statistical power. Furthermore, it attempted to reduce the quantity of primary data by combining the results of the SNAP™ evaluation studies and producing an overall effect size of the intervention. It aimed to increase the efficiency of the existing data and explain the inconsistency in the included studies (Gopalakrishnan & Ganeshkumar, 2013). Knowledge of this thesis work will lay the groundwork for following children exposed initially to the SNAP™ into the future until they reach mid- to late adolescence when the risk of criminal offending is highest. Finally, the results of the study will benefit others interested in implementing or supporting a program of this nature in the future. The thesis work will inform other researchers about the suggested Meta-Analysis approach to use to review evaluations of crime prevention programs with the presence of similar limitations and challenges. Furthermore, reaching an in-depth understanding of the implementation of the intervention will help decision-makers in Canada and other countries to scale up sustainable, evidence-based mental health interventions and thereby reduce the mental health treatment burden.

REFERENCE LIST

Achenbach, T. M., & Ruffle, T. M. (2000). The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review, 21*(8), 265-271.

American Psychiatric Association. (2013). Disruptive, Impulse-Control, and Conduct Disorders. In *Diagnostic and Statistical Manual of Mental Disorders*. (Reprinted from: Fifth Edition).

Augimeri, L. K. (2005). *Aggressive and antisocial Young Children: Risk Assessment and Management Utilizing the Early Assessment Risk List for Boys (EARL-20B).* (Ph.D. Dissertation), Ontario Institute for Studies in Eduction, University of Toronto.

Augimeri, L. K., Farrington, D. P., Koegl, C. J., & Day, D. M. (2007). The SNAP™ Under 12 Outreach Project: Effects of a Community Based Program for Children with Conduct Problems. *Journal of Child and Family Studies, 16*(6), 799-807. doi:10.1007/s10826-006-9126-x

Augimeri, L. K., Jiang, D., Koegl, C. J., & Carey, J. (2006). *Differential effects of the SNAP™ under 12 outreach project (SNAP™ ORP) associated with client risk and treatment intensity*. Program Evaluation Report submitted to The Provincial Centre of Excellence for Child and Youth Mental Health at CHEO (Grant: # PEG162606-101). Retrieved from http://www.excellenceforchildandyouth.ca/sites/default/files/gai_attach/peg-631_final_outcomes_report.pdf

Augimeri, L. K., Pepler, D., Walsh, M. M., Jiang, D., & Dassinger, C. (2010). *Aggressive and antisocial young children: Risk prediction, assessment and clinical risk management*. Program Evaluation Report submitted to The Provincial Centre of Excellence for Child

and Youth Mental Health at CHEO (Grant: # RG-976). Retrieved from http://www.excellenceforchildandyouth.ca/sites/default/files/gai_attach/RG-976_Final_Outcomes_Report.pdf

Augimeri, L. K., Walsh, M., Donato, A., Blackman, A., & Piquero, A. R. (2018). SNAP (Stop Now And Plan): Helping children improve their self-control and externalizing behavior problems. *Journal of Criminal Justice, 56*, 43-49.

Augimeri, L. K., Walsh, M., & Slater, N. (2011). Rolling out SNAP® an evidence-based intervention: A summary of implementation, evaluation and research. *International Journal of Child, Youth and Family Studies, 2*(2.1), 330-352. doi:10.18357/ijcyfs22.120117710

Bordin, I. A., Rocha, M. M., Paula, C. S., Teixeira, M. C. T., Achenbach, T. M., Rescorla, L. A., & Silvares, E. F. (2013). Child Behavior Checklist (CBCL), Youth Self-Report (YSR) and Teacher's Report Form (TRF): an overview of the development of the original and Brazilian versions. *Cadernos de Saúde Pública, 29*, 13-28.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*: John Wiley & Sons.

Burke, J., & Loeber, R. (2015). The Effectiveness of the Stop Now and Plan (SNAP) Program for Boys at Risk for Violence and Delinquency. *Prevention Science, 16*(2), 242-253. doi:10.1007/s11121-014-0490-2

Burke, J., & Loeber, R. (2016). Mechanisms of Behavioral and Affective Treatment Outcomes in a Cognitive Behavioral Intervention for Boys. *Journal of Abnormal Child Psychology, 44*(1), 179-189. doi:10.1007/s10802-015-9975-0

Byrd, A. L., Hawes, S. W., Burke, J. D., Loeber, R., & Pardini, D. A. (2018a). Boys with conduct problems and callous-unemotional traits: Neural response to reward and punishment and associations with treatment response. *Developmental Cognitive Neuroscience, 30,* 51-59.

Byrd, A. L., Hawes, S. W., Burke, J. D., Loeber, R., & Pardini, D. A. (2018b). Boys with conduct problems and callous-unemotional traits: neural response to reward and punishment and associations with treatment response. *Developmental Cognitive Neuroscience, 30,* 51-59.

Child Development Institute. (2016a). A brief history of SNAP®. Retrieved from https://childdevelop.ca/snap/about-snap/brief-history-snap

Child Development Institute. (2016b). About SNAP® Affiliates. Retrieved from https://childdevelop.ca/snap/snap-affiliates?gclid=CjwKCAjw-ITqBRB7EiwAZ1c5U9s6A2gjnQ_iDf65CQ4URs__TBX6nQzCP1iImvS3C34HHJlr_ZqxLBoCQ6gQAvD_BwE

Child Development Institute. (2019). *SNAP® Research Outcomes.*

Children's Mental Health Ontario. (2001). Children and Adolescents with Conduct Disorder: Findings from the Literature and Clinical Consultation in Ontario. In. Retrieved from http://www.cmho.org/documents/EBP_conduct_disorder_findings.pdf

Cochrane Community. (2019). RevMan 5. Retrieved from https://community.cochrane.org/help/tools-and-software/revman-5

Cooper, H., & Dent, A. (2011). Ethical Issues in the Conduct and Reporting of Meta-Analysis. In *Handbook of Ethics in Quantitative Methodology*: Routledge.

Day, D. M. (1998). Risk for court contact and predictors of an early age for a first court contact among a sample of high risk youths: A survival analysis approach. *Canadian Journal of Criminology, 40*(4), 421-446.

Day, D. M. (2003). *Addressing Antisocial Behaviour in Children: An Evaluation of the "Durham Under Twelve" Pilot Project*. Kinark child and Family Services, Durham Office, Oshawa, Ontario.

Day, D. M., & Hrynkiw-Augimeri, L. (1996). *Serving children at risk for juvenile delinquency: An evaluation of the Earlscourt Under 12 Outreach Project (SNAP® ORP)*. Final Report. Submitted to the Department of Justice. Earlscourt Child and Family Centre.

Day, D. M., & Hunt, A. C. (1996). A Multivariate Assessment of a Risk Model for Juvenile Delinquency with an "Under 12 Offender" Sample. *Journal of Emotional and Behavioral Disorders, 4*(2), 66-72. doi:10.1177/106342669600400201

Deeks, J. J., Higgins, J. P., & Altman, D. G. (2011). In J. P. Higgins & S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration. Retrieved from http://handbook-5-1.cochrane.org/

Dobson, K. S., & Dozois, D. J. (2019). *Handbook of cognitive-behavioral therapies*: Guilford Publications.

Farrington, D. P., & Koegl, C. J. (2015). Monetary benefits and costs of the Stop Now and Plan Program for boys aged 6–11, based on the prevention of later offending. *Journal of Quantitative Criminology, 31*(2), 263-287.

Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S., & Vega, R. (2014). Effects of the Good Behavior Game on Challenging Behaviors in School Settings. *Review of Educational Research, 84*(4), 546-571. doi:10.3102/0034654314536781

Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *J Family Med Prim Care, 2*(1), 9-14. doi:10.4103/2249-4863.109934

Granic, I., O'Hara, A., Pepler, D., & Lewis, M. D. (2007). A dynamic systems analysis of parent–child changes associated with successful "real-world" interventions for aggressive children. *Journal of Abnormal Child Psychology, 35*(5), 845-857.

Hackshaw, A. (2008). Small studies: strengths and limitations. In A. Hackshaw (Ed.), (Vol. 32, pp. 1141-1143).

Higgins, J. P., & Deeks, J. J. (2011). In J. P. Higgins & S. Green (Eds.), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0: The Cochrane Collaboration. Retrieved from http://handbook-5-1.cochrane.org/.

Hopewell, S., McDonald, S., Clarke, M. J., & Egger, M. (2007). Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database of Systematic Reviews*(2). doi:10.1002/14651858.MR000010.pub3

Hrynkiw-Augimeri, L., Pepler, D., & Goldberg, K. (1993). An outreach program for children having police contact. *Canada's Mental Health, 41*(2), 7-12.

Hulley, S. B. (2007). *Designing clinical research*: Lippincott Williams & Wilkins.

Hupp, S. D., Reitman, D., & Jewell, J. D. (2008). Cognitive behavioral theory. *Handbook of clinical psychology, 2*, 263-287.

Kazdin, A. E. (2005). *Parent Management Training : Treatment for Oppositional, Aggressive, and Antisocial Behavior in Children and Adolescents*. Cary, UNITED STATES: Oxford University Press USA - OSO.

Kazdin, A. E. (2005). *Parent Management Training : Treatment for Oppositional, Aggressive, and Antisocial Behavior in Children and Adolescents.* In. Retrieved from http://search.ebscohost.com.uml.idm.oclc.org/login.aspx?direct=true&db=nlebk&AN=138170&site=ehost-live

Koegl, C. J. (2011). *High-risk antisocial children: Predicting future criminal and health outcomes.* (Ph.D. Dissertation), University of Cambridge.

Koegl, C. J., Farrington, D. P., Augimeri, L. K., & Day, D. M. (2008). Evaluation of a Targeted Cognitive-Behavioral Program for Children with Conduct Problems — The SNAP® Under 12 Outreach Project: Service Intensity, Age and Gender Effects on Short- and Long-Term Outcomes. *Clinical Child Psychology and Psychiatry, 13*(3), 419-434. doi:10.1177/1359104508090606

Lande, M. B., Adams, H., Falkner, B., Waldstein, S. R., Schwartz, G. J., Szilagyi, P. G., . . . Palumbo, D. (2009). Parental assessments of internalizing and externalizing behavior and executive function in children with primary hypertension. *The Journal of pediatrics, 154*(2), 207–212. doi:10.1016/j.jpeds.2008.08.017

Levene, K. S., Walsh, M. M., Augimeri, L. K., & Pepler, D. J. (2004). Linking identification and treatment of early risk factors for female delinquency. In *Girls and Aggression* (pp. 147-163): Springer.

Lewis, M. D., Granic, I., Lamm, C., Zelazo, P. D., Stieben, J., Todd, R. M., . . . Pepler, D. (2008). Changes in the neural bases of emotion regulation associated with clinical improvement in children with behavior problems. *Development and psychopathology, 20*(3), 913-939.

Lipman, E. L., Kenny, M., Brennan, E., O'Grady, S., & Augimeri, L. k. (2011). Helping boys at-risk of criminal activity: qualitative results of a multi-component intervention. *BioMed Central Public Health, 11,* 364-364. doi:10.1186/1471-2458-11-364

Lipman, E. L., Kenny, M., Sniderman, C., O'Grady, S., Augimeri, L. k., Khayutin, S., & Boyle, M. H. (2008). Evaluation of a community-based program for young boys at-risk of antisocial behaviour: results and issues. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 17*(1), 12-19.

Lipman, E. L., Kenny, M., & Wymouth, M. (2007). *Banyan Community Service Girls Connection Program: Final Evaluation Report.* Offord Centre for Child Studies.

McAuley, L., Pham, B., Tugwell, P., & Moher, D. (2000). Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet, 356*(9237), 1228-1231. doi:10.1016/s0140-6736(00)02786-0

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. New York, NY, US: Cambridge University Press.

Pepler, D., Walsh, M., Yuile, A., Levene, K., Jiang, D., Vaughan, A., & Webber, J. (2010). Bridging the Gender Gap: Interventions with Aggressive Girls and Their Parents. *Prevention Science, 11*(3), 229-238. doi:10.1007/s11121-009-0167-4

Smith-Moncrieffe, D. (2015a). *Third Annual Evaluation Report – SNAP Program Implementation at Cree Nation Communities* (2015-R017). Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2015-r017/2015-r017-en.pdf

Smith-Moncrieffe, D. (2015b). *Third Annual Evaluation Report – SNAP Program Implementation at St. Leonard's Society of Toronto* (2015-R017). Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2015-r017/2015-r017-en.pdf

Smith-Moncrieffe, D. (2015c). *Third Annual Evaluation Report – SNAP Program Implementation at Uncles and Aunts at Large (Edmonton)* (2015-R017). Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/2015-r017/2015-r017-en.pdf

Sterne, J. A., Higgins, J.P., Elbers, R.G., Reeves, B.C. and the development group for ROBINS-I. (2016). Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I): detailed guidance, updated 12 October 2016.

Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., . . . for the Meta-Analysis Of Observational Studies in Epidemiology, G. (2000). Meta-analysis of Observational Studies in Epidemiology: A Proposal for Reporting. *JAMA*, *283*(15), 2008-2012. doi:10.1001/jama.283.15.2008

Walsh, M. M., Pepler, D. J., & Levene, K. S. (2002). A Model Intervention for Girls with Disruptive Behaviour Disorders: The Earlscourt Girls Connection. *Canadian Journal of Counselling*, *36*(4), 297-311.

Webster, C. D., Augimeri, L. K., & Koegl, C. J. (2002). The Under 12 Outreach Project for antisocial boys: A research based clinical program. In *Multi-problem violent youth: A foundation for comparative research on needs, interventions and outcomes.* (pp. 207-218). Amsterdam, Netherlands: IOS Press.

Woltering, S., Granic, I., Lamm, C., & Lewis, M. D. (2011). Neural changes associated with treatment outcome in children with externalizing problems. *Biological psychiatry*, *70*(9), 873-879.

Woltering, S., Liao, V., Liu, Z.-X., & Granic, I. (2015). Neural Rhythms of Change: Long-Term Improvement after Successful Treatment in Children with Disruptive Behavior Problems. *Neural Plasticity, 2015*(2015). doi:10.1155/2015/873197

Woltering, S., Lishak, V., Hodgson, N., Granic, I., & Zelazo, P. D. (2016). Executive function in children with externalizing and comorbid internalizing behavior problems. *Journal of Child Psychology and Psychiatry, 57*(1), 30-38.

APPENDIX A

Table 16 CBCL-Externalizing Problem (Post-Treatment): Subgroup Analysis

Subgroup	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	191	-0.30 [-0.41, -0.19]	0%
Journal; 2001-2010; Internal; Female Only; Moderate	1	40	-0.67 [-0.90, -0.43]	N/A
Test for subgroup differences: $Chi^2 = 7.42$, df = 1 (P = 0.006), $I^2 = 86.5\%$				
By Study Location				
Toronto, Ontario, CA	2	105	-0.44 [-0.86, -0.02]	89%
Quebec, Ontario, CA	1	55	-0.39 [-0.71, -0.07]	N/A
Edmonton, Alberta, CA	1	71	-0.39 [-0.59, -0.19]	N/A
Test for subgroup differences: $Chi^2 = 0.05$, df = 2 (P = 0.97), $I^2 = 0\%$				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	136	-0.30 [-0.44, -0.15]	32%
Attended < 8 of Child SNAP™ Group Session	1	55	-0.39 [-0.71, -0.07]	N/A
Test for subgroup differences: $Chi^2 = 0.27$, df = 1 (P = 0.60), $I^2 = 0\%$				
By Treatment Fidelity				
≥ 80%	3	176	-0.42 [-0.65, -0.18]	78%
< 80%	1	55	-0.39 [-0.71, -0.07]	N/A
Test for subgroup differences: $Chi^2 = 0.02$, df = 1 (P = 0.90), $I^2 = 0\%$				

Table 17 CBCL-Externalizing Problem (6-Month Follow-up): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Year				
2011-2020	1	215	-0.38 [-0.50, -0.26]	N/A
2001-2010	2	114	-0.75 [-0.93, -0.56]	0%
1991-2000	1	72	-0.42 [-0.59, -0.24]	N/A
Test for subgroup differences: $Chi^2 = 11.21$, df = 2 (P = 0.004), $I^2 = 82.2\%$				
By Publication Type; Publication Status; Study Location; Risk of Bias				
Report; Unpublished; Hamilton, Ontario, CA; Serious	1	80	-0.72 [-0.97, -0.47]	N/A
Journal; Published; Toronto, Ontario, CA; Moderate	3	321	-0.49 [-0.69, -0.30]	71%
Test for subgroup differences: $Chi^2 = 2.02$, df = 1 (P = 0.16), $I^2 = 50.4\%$				
By Evaluation Agent/Type				
Internal	1	34	-0.77 [-1.04, -0.51]	N/A
External	1	80	-0.72 [-0.97, -0.47]	N/A
Shared	2	287	-0.39 [-0.49, -0.29]	0%

Test for subgroup differences: Chi² = 11.19, df = 2 (P = 0.004), I² = 82.1%				
By Study Design				
Retrospective	2	287	-0.39 [-0.49, -0.29]	0%
Prospective	2	114	-0.75 [-0.93, -0.56]	0%
Test for subgroup differences: Chi² = 11.11, df = 1 (P = 0.0009), I² = 91.0%				
By Participant Gender				
Female	3	186	-0.62 [-0.86, -0.38]	70%
Mixed-Gender	1	215	-0.38 [-0.50, -0.26]	N/A
Test for subgroup differences: Chi² = 3.01, df = 1 (P = 0.08), I² = 66.7%				
By Year of Enrollment; outcome Measures				
1985-2000; SCIS (1996)	1	72	-0.42 [-0.59, -0.24]	N/A
2001-2017; CBCL (2001)	3	329	-0.61 [-0.89, -0.33]	81%
Test for subgroup differences: Chi² = 1.29, df = 1 (P = 0.26), I² = 22.3%				

Table 18 CBCL-Rule-Breaking (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	191	-0.28 [-0.43, -0.13]	39%
Journal; 2001-2010; Internal; Female Only; Moderate	1	40	-0.59 [-0.80, -0.38]	N/A
Test for subgroup differences: Chi² = 5.68, df = 1 (P = 0.02), I² = 82.4%				
By Study Location				
Toronto, Ontario, CA	2	105	-0.37 [-0.79, 0.06]	90%
Quebec, Ontario, CA	1	55	-0.38 [-0.62, -0.14]	N/A
Edmonton, Alberta, CA	1	71	-0.35 [-0.55, -0.14]	N/A
Test for subgroup differences: Chi² = 0.06, df = 2 (P = 0.97), I² = 0%				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	136	-0.24 [-0.43, -0.05]	51%
Attended < 8 of Child SNAP™ Group Session	1	55	-0.38 [-0.62, -0.14]	N/A
Test for subgroup differences: Chi² = 0.87, df = 1 (P = 0.35), I² = 0%				
By Treatment Fidelity				
≥ 80%	3	176	-0.36 [-0.61, -0.10]	80%
< 80%	1	55	-0.38 [-0.62, -0.14]	N/A
Test for subgroup differences: Chi² = 0.02, df = 1 (P = 0.88), I² = 0%				

Table 19 CBCL-Delinquency (9-Month Follow-up): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Study Design; Participant Gender				
Report; 2001-2010; Retrospective; Male only	1	318	-0.72 [-0.80, -0.63]	N/A
Journal; 1991-2000; Prospective; Mixed-Gender	1	68	-1.03 [-1.23, -0.83]	N/A
Test for subgroup differences: $Chi^2 = 8.16$, df = 1 (P = 0.004), $I^2 = 87.7\%$				

Table 5 CBCL-DSM Conduct Disorder (6-Month Follow-up): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Status				
Report; Unpublished	1	80	-0.66 [-0.91, -0.41]	N/A
Journal; Published	3	238	-0.60 [-0.83, -0.37]	75%
Test for subgroup differences: $Chi^2 = 0.11$, df = 1 (P = 0.74), $I^2 = 0\%$				
By Study Design; Year of Enrollment; Outcome Measures				
Retrospective; 1985-2000; SCIS (1996)	1	72	-0.39 [-0.57, -0.21]	N/A
Prospective; 2001-2017; CBCL (2001)	3	246	-0.70 [-0.83, -0.57]	0%
Test for subgroup differences: $Chi^2 = 7.65$, df = 1 (P = 0.006), $I^2 = 86.9\%$				
By Evaluation Agent/Type				
Internal	1	34	-0.65 [-0.87, -0.44]	N/A
External	2	212	-0.73 [-0.88, -0.57]	0%
Shared	1	72	-0.39 [-0.57, -0.21]	N/A
Test for subgroup differences: $Chi^2 = 7.92$, df = 2 (P = 0.02), $I^2 = 74.7\%$				
By Participant Gender				
Female Only	3	186	-0.55 [-0.74, -0.37]	56%
Male Only	1	55	-0.77 [-0.97, -0.57]	N/A
Test for subgroup differences: $Chi^2 = 2.38$, df = 1 (P = 0.12), $I^2 = 57.9\%$				
By Study Location; Risk of Bias				
Toronto, Ontario, CA; Moderate	2		-0.52 [-0.77, -0.26]	70%
Hamilton, Ontario, CA; Serious	2		-0.73 [-0.88, -0.57]	0%
Test for subgroup differences: $Chi^2 = 1.89$, df = 1 (P = 0.17), $I^2 = 47.0\%$				

Table 6 CBCL-Attention (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	191	-0.26 [-0.47, -0.05]	72%
Journal; 2001-2010; Internal; Female Only; Moderate	1	40	-0.34 [-0.53, -0.15]	N/A

Test for subgroup differences: Chi² = 0.31, df = 1 (P = 0.58), I² = 0%				
By Study Location				
Toronto, Ontario, CA	2	105	-0.20 [-0.47, 0.08]	78%
Quebec, Ontario, CA	1	55	-0.41 [-0.64, -0.19]	N/A
Edmonton, Alberta, CA	1	71	-0.33 [-0.50, -0.15]	N/A
Test for subgroup differences: Chi² = 1.40, df = 2 (P = 0.50), I² = 0%				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	136	-0.19 [-0.46, 0.07]	77%
Attended < 8 of Child SNAP™ Group Session	1	55	-0.41 [-0.64, -0.19]	N/A
Test for subgroup differences: Chi² = 1.54, df = 1 (P = 0.21), I² = 35.1%				
By Treatment Fidelity				
≥ 80%	3	176	-0.24 [-0.42, -0.06]	66%
< 80%	1	55	-0.41 [-0.64, -0.19]	N/A
Test for subgroup differences: Chi² = 1.36, df = 1 (P = 0.24), I² = 26.7%				

Table 7 Official Criminal Records by the End of the Study: Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate OR [95% CI]	I^2
By Evaluation Agent/Type; Outcome Measures				
Internal; Official Criminal Records	1	395	0.25 [0.16, 0.39]	N/A
Shared; Parents' Reports	1	99	0.07 [0.03, 0.19]	N/A
Test for subgroup differences: Chi² = 5.16, df = 1 (P = 0.02), I² = 80.6%				

Table 8 CBCL-Internalizing Problems (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	191	-0.31 [-0.50, -0.11]	60%
Journal; 2001-2010; Internal; Female Only; Moderate	1	40	-0.54 [-0.74, -0.34]	N/A
Test for subgroup differences: Chi² = 2.78, df = 1 (P = 0.10), I² = 64.1%				
By Study Location				
Toronto, Ontario, CA	2	105	-0.34 [-0.73, 0.06]	88%
Quebec, Ontario, CA	1	55	-0.43 [-0.75, -0.10]	N/A
Edmonton, Alberta, CA	1	71	-0.40 [-0.56, -0.24]	N/A
Test for subgroup differences: Chi² = 0.11, df = 2 (P = 0.95), I² = 0%				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	136	-0.27 [-0.52, -0.02]	77%
Attended < 8 of Child SNAP™ Group Session	1	55	-0.43 [-0.75, -0.10]	N/A
Test for subgroup differences: Chi² = 0.53, df = 1 (P = 0.47), I² = 0%				

By Treatment Fidelity				
≥ 80%	3	176	-0.36 [-0.58, -0.14]	78%
< 80%	1	55	-0.43 [-0.75, -0.10]	N/A
Test for subgroup differences: $Chi^2 = 0.11$, df = 1 (P = 0.74), $I^2 = 0\%$				

Table 9 CBCL-Withdrawn-Depressed (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Study Location				
Toronto, Ontario, CA	1	65	0.01 [-0.15, 0.17]	N/A
Quebec, Ontario, CA	1	55	-0.13 [-0.46, 0.20]	N/A
Edmonton, Alberta, CA	1	71	-0.39 [-0.54, -0.24]	N/A
Test for subgroup differences: $Chi^2 = 12.95$, df = 2 (P = 0.002), $I^2 = 84.6\%$				
By Treatment Intensity and Fidelity				
Attended ≥ 8 of Child SNAP™ Group Session and Fidelity ≥ 80%	2	136	-0.19 [-0.58, 0.20]	92%
Attended < 8 of Child SNAP™ Group Session and Fidelity < 80%	1	55	-0.13 [-0.46, 0.20]	N/A
Test for subgroup differences: $Chi^2 = 0.06$, df = 1 (P = 0.81), $I^2 = 0\%$				

Table 10 CBCL-Somatic Complaints (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Study Location				
Toronto, Ontario, CA	1	65	0.07 [-0.10, 0.25]	N/A
Quebec, Ontario, CA	1	55	-0.60 [-0.84, -0.36]	N/A
Edmonton, Alberta, CA	1	71	-0.09 [-0.29, 0.10]	N/A
Test for subgroup differences: $Chi^2 = 19.90$, df = 2 (P < 0.0001), $I^2 = 89.9\%$				
By Treatment Intensity and Fidelity				
Attended ≥ 8 of Child SNAP™ Group Session and Fidelity ≥ 80%	2	136	-0.00 [-0.17, 0.16]	37%
Attended < 8 of Child SNAP™ Group Session and Fidelity < 80%	1	55	-0.60 [-0.84, -0.36]	N/A
Test for subgroup differences: $Chi^2 = 15.95$, df = 1 (P < 0.0001), $I^2 = 93.7\%$				

Table 11 CBCL-Anxious-Depressed (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Study Location				
Toronto, Ontario, CA	1	65	-0.17 [-0.37, 0.03]	N/A
Quebec, Ontario, CA	1	55	-0.65 [-0.95, -0.35]	N/A
Edmonton, Alberta, CA	1	71	-0.44 [-0.64, -0.23]	N/A

Test for subgroup differences: $Chi^2 = 7.67$, df = 2 (P = 0.02), $I^2 = 73.9\%$				
By Treatment Intensity and Fidelity				
Attended ≥ 8 of Child SNAP™ Group Session and Fidelity ≥ 80%	2	136	-0.30 [-0.56, -0.04]	71%
Attended < 8 of Child SNAP™ Group Session and Fidelity < 80%	1	55	-0.65 [-0.95, -0.35]	N/A
Test for subgroup differences: $Chi^2 = 2.96$, df = 1 (P = 0.09), $I^2 = 66.3\%$				

APPENDIX B

Table 1 TRF-Externalizing Problem (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	-0.01 [-0.25, 0.24]	83%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.13 [-0.32, 0.06]	N/A
Test for subgroup differences: $Chi^2 = 0.60$, df = 1 (P = 0.44), $I^2 = 0\%$				
By Study Location				
Toronto, Ontario, CA	2	83	-0.17 [-0.31, -0.03]	0%
Quebec, Ontario, CA	1	20	0.24 [0.06, 0.42]	N/A
Edmonton, Alberta, CA	1	74	-0.04 [-0.17, 0.09]	N/A
Test for subgroup differences: $Chi^2 = 12.76$, df = 2 (P = 0.002), $I^2 = 84.3\%$				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	124	-0.12 [-0.30, 0.06]	56%
Attended < 8 of Child SNAP™ Group Session	1	20	0.24 [0.06, 0.42]	N/A
Test for subgroup differences: $Chi^2 = 7.59$, df = 1 (P = 0.006), $I^2 = 86.8\%$				
By Treatment Fidelity				
≥ 80%	3	157	-0.11 [-0.21, -0.00]	80%
< 80%	1	20	0.24 [0.06, 0.42]	N/A
Test for subgroup differences: $Chi^2 = 10.74$, df = 1 (P = 0.001), $I^2 = 90.7\%$				

Table 2 TRF-Rule-Breaking (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	0.09 [-0.10, 0.29]	78%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.16 [-0.35, 0.04]	N/A
Test for subgroup differences: $Chi^2 = 3.16$, df = 1 (P = 0.08), $I^2 = 68.4\%$				
By Study Location				
Toronto, Ontario, CA	2	83	-0.11 [-0.25, 0.03]	0%
Quebec, Ontario, CA	1	20	0.27 [0.14, 0.41]	N/A
Edmonton, Alberta, CA	1	74	0.04 [-0.10, 0.18]	N/A
Test for subgroup differences: $Chi^2 = 14.99$, df = 2 (P = 0.0006), $I^2 = 86.7\%$				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	124	0.01 [-0.11, 0.12]	0%
Attended < 8 of Child SNAP™ Group Session	1	20	0.27 [0.14, 0.41]	N/A
Test for subgroup differences: $Chi^2 = 8.51$, df = 1 (P = 0.004), $I^2 = 88.3\%$				

By Treatment Fidelity				
≥ 80%	3	157	-0.04 [-0.16, 0.07]	22%
< 80%	1	20	0.27 [0.14, 0.41]	N/A
Test for subgroup differences: Chi^2 = 11.98, df = 1 (P = 0.0005), I^2 = 91.7%				

Table 3 TRF-Aggression (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	-0.02 [-0.31, 0.26]	88%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.14 [-0.36, 0.07]	N/A
Test for subgroup differences: Chi^2 = 0.42, df = 1 (P = 0.51), I^2 = 0%				
By Study Location				
Toronto, Ontario, CA	2	83	-0.20 [-0.35, -0.04]	0%
Quebec, Ontario, CA	1	20	0.25 [0.09, 0.41]	N/A
Edmonton, Alberta, CA	1	74	-0.09 [-0.24, 0.05]	N/A
Test for subgroup differences: Chi^2 = 17.37, df = 2 (P = 0.0002), I^2 = 88.5%				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	124	-0.15 [-0.30, 0.00]	31%
Attended < 8 of Child SNAP™ Group Session	1	20	0.25 [0.09, 0.41]	N/A
Test for subgroup differences: Chi^2 = 13.03, df = 1 (P = 0.0003), I^2 = 92.3%				
By Treatment Fidelity				
≥ 80%	3	157	-0.14 [-0.25, -0.04]	0%
< 80%	1	20	0.25 [0.09, 0.41]	N/A
Test for subgroup differences: Chi^2 = 16.42, df = 1 (P < 0.0001), I^2 = 93.9%				

Table 4 TRF-DSM Conduct Disorder (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	0.01 [-0.21, 0.22]	73%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.16 [-0.35, 0.03]	N/A
Test for subgroup differences: Chi^2 = 1.28, df = 1 (P = 0.26), I^2 = 22.0%				
By Study Location				
Toronto, Ontario, CA	2	83	-0.17 [-0.31, -0.02]	0%
Quebec, Ontario, CA	1	20	0.25 [0.02, 0.48]	N/A
Edmonton, Alberta, CA	1	74	-0.03 [-0.17, 0.11]	N/A
Test for subgroup differences: Chi^2 = 9.35, df = 2 (P = 0.009), I^2 = 78.6%				
By Treatment Intensity				

Attended ≥ 8 of Child SNAP™ Group Session	2	124	-0.08 [-0.21, 0.05]	19%
Attended < 8 of Child SNAP™ Group Session	1	20	0.25 [0.02, 0.48]	N/A
Test for subgroup differences: Chi^2 = 6.10, df = 1 (P = 0.01), I^2 = 83.6%				
By Treatment Fidelity				
≥ 80%	3	157	-0.10 [-0.20, 0.00]	0%
< 80%	1	20	0.25 [0.02, 0.48]	N/A
Test for subgroup differences: Chi^2 = 7.62, df = 1 (P = 0.006), I^2 = 86.9%				

Table 5 TRF-Attention (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	-0.05 [-0.29, 0.19]	84%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.08 [-0.31, 0.15]	N/A
Test for subgroup differences: Chi^2 = 0.03, df = 1 (P = 0.86), I^2 = 0%				
By Study Location				
Toronto, Ontario, CA	2	83	-0.13 [-0.29, 0.02]	0%
Quebec, Ontario, CA	1	20	0.15 [0.03, 0.27]	N/A
Edmonton, Alberta, CA	1	74	-0.15 [-0.32, 0.02]	N/A
Test for subgroup differences: Chi^2 = 12.28, df = 2 (P = 0.002), I^2 = 83.7%				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	124	-0.17 [-0.30, -0.03]	0%
Attended < 8 of Child SNAP™ Group Session	1	20	0.15 [0.03, 0.27]	N/A
Test for subgroup differences: Chi^2 = 12.12, df = 1 (P = 0.0005), I^2 = 91.7%				
By Treatment Fidelity				
≥ 80%	3	157	-0.14 [-0.26, -0.03]	0%
< 80%	1	20	0.15 [0.03, 0.27]	N/A
Test for subgroup differences: Chi^2 = 12.25, df = 1 (P = 0.0005), I^2 = 91.8%				

Table 6 TRF-Adaptive Functioning (6 Months Follow-up): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Status; Participant Gender				
Report; Unpublished; Female Only	1	83	-0.08 [-0.30, 0.15]	N/A
Journal; Published; Male Only	1	101	0.24 [0.04, 0.45]	N/A
Test for subgroup differences: Chi^2 = 4.26, df = 1 (P = 0.04), I^2 = 76.5%				

Table 7 TRF-Internalizing Problems (Post-Treatment): Subgroup Analysis

Subgroups	Studies	Participants	Effect Estimate SMD [95% CI]	I^2
By Publication Type; Publication Year; Evaluation Agent/Type; Participant Gender; Risk of Bias				
Report; 2011-2020; External; Mixed-Gender; Serious	3	144	0.08 [-0.25, 0.41]	88%
Journal; 2001-2010; Internal; Female Only; Moderate	1	33	-0.23 [-0.40, -0.07]	N/A
Test for subgroup differences: $Chi^2 = 2.68$, df = 1 (P = 0.10), $I^2 = 62.7\%$				
By Study Location				
Toronto, Ontario, CA	2	83	-0.21 [-0.35, -0.08]	0%
Quebec, Ontario, CA	1	20	0.45 [0.22, 0.67]	N/A
Edmonton, Alberta, CA	1	74	-0.03 [-0.16, 0.10]	N/A
Test for subgroup differences: $Chi^2 = 24.21$, df = 2 (P < 0.00001), $I^2 = 91.7\%$				
By Treatment Intensity				
Attended ≥ 8 of Child SNAP™ Group Session	2	124	-0.07 [-0.19, 0.06]	11%
Attended < 8 of Child SNAP™ Group Session	1	20	0.45 [0.22, 0.67]	N/A
Test for subgroup differences: $Chi^2 = 15.28$, df = 1 (P < 0.0001), $I^2 = 93.5\%$				
By Treatment Fidelity				
≥ 80%	3	157	-0.13 [-0.27, 0.01]	48%
< 80%	1	20	0.45 [0.22, 0.67]	N/A
Test for subgroup differences: $Chi^2 = 18.54$, df = 1 (P < 0.0001), $I^2 = 94.6\%$				

Table 8 TRF-Total Problems (Post-Treatment): Subgroup Analysis

By Study Location				
Toronto, Ontario, CA	1	50	-0.24 [-0.48, 0.01]	N/A
Quebec, Ontario, CA	1	20	0.24 [0.01, 0.46]	N/A
Edmonton, Alberta, CA	1	74	-0.11 [-0.23, 0.00]	N/A
Test for subgroup differences: $Chi^2 = 9.45$, df = 2 (P = 0.009), $I^2 = 78.8\%$				
By Treatment Intensity and Fidelity				
Attended ≥ 8 of Child SNAP™ Group Session and Fidelity ≥ 80%	2	136	-0.13 [-0.24, -0.03]	0%
Attended < 8 of Child SNAP™ Group Session and Fidelity < 80%	1	55	0.24 [0.01, 0.46]	N/A
Test for subgroup differences: $Chi^2 = 8.61$, df = 1 (P = 0.003), $I^2 = 88.4\%$				