UiT The Arctic University of Norway

Faculty of Health Sciences

Department of Psychology

Providing a short and effective transdiagnostic treatment intervention and a valid outcome measure for adolescent with anxiety and depression. A randomized controlled trial of the SMART intervention and validation of the CORE-OM in adolescents aged 14 to 17

Veronica Lorentzen

A dissertation for the degree of Philosophiae Doctor, October 2022

Foto: Johannes Sirnes Lorentzen, 2021

Table of Contents

Summary 6
Abbreviations 7
List of papers 8
1 Introduction 9
1.1 Anxiety and depression in adolescents 9
1.1.1 Anxiety 11
1.1.2 Depression 11
1.1.3 Comorbidity in emotional disorders 12
1.2 Psychological treatment of emotional disorders in adolescents 13
1.3 Transdiagnostic treatment 16
1.3.1 Transdiagnostic interventions categories 17
1.3.2 Linking transdiagnostic interventions to transdiagnostic processes 17
1.4 Efficacy and effectiveness 18
1.5 Measurement-based care (MBC) 20
1.6 The Clinical Outcome in Routine Evaluation (CORE-OM) as a Routine outcome measure in adolescence 21
2 Thesis Aims 22
2.1 Paper 1 23
2.2 Paper 2 23
2.3 Paper 3 24
3 Methods 24
3.1 The SMART study 24
3.2 Treatment 24
3.3 Recruitment, sites and therapists 27
3.4 Training, supervision, treatment integrity, alliance and user satisfaction 27
3.5 Samples and data collection 29
3.6 Research ethics 30
3.7 Registration and financing of the study 31
3.8 Measures 32
3.8.1 Issues concerning choice of measurements for study entry and outcome measures 32
3.8.2 Measures used in the study 33
3.9 Analysis 37

4 Summary of the papers 40
4.1 Summary paper 1 40
4.2 Summary paper 2 41
4.3 Summary paper 3 43
5 General discussion 45
5.1 Methodological considerations: Strengths and limitations: 56
5.1.1 Strengths 56
5.1.2 Limitations 56
5.2 Implications 60
5.3 Future research 61
6 Concluding remarks 63
7 References 65

Papers 1-3

Acknowledgements

Thinking back on the years spent on this project, I am filled with feelings of sincere gratitude towards all the people who have contributed to bridge the gap between research and practice in this project. It really takes a village and a high degree of willingness; from patients to participate in research, from leaders to implement research, from therapists to be dedicated and open, from administrative staff to organize and collect data, from supervisors to oversee such a project.

The idea of doing research on SMART came to life through years of work as a clinical psychologist at CAMHS, especially working with children and adolescents struggling with anxiety, depression or both. Thanks to all the children and adolescents for sharing your struggles with sadness, loss, anxiety and worry, but also for sharing your courage, creativity and thoughtfulness. Being in therapy should also be about awareness of your own strengths and self-efficacy. Thank you for highlighting this to me when working with you.

First of all, this project could not have been realized without the participating adolescents and parents who said yes to share their treatment results to provide more knowledge to the field. I want you to know that not only did you give us the knowledge in this thesis, but the supervision and time spent to understand how we could be better performing CBT, has made us better at providing this. All of this made possible with the experiences you gave us, thus helping many other adolescents that came after you.

Thanks to my research group; it has been a bliss working together throughout the years both academically and personally. I cannot express my gratitude enough and I hope for continued opportunities for collaboration. More specifically, I would like to express my sincere gratitude to my main supervisor and project leader, Ingunn Skre, for your valuable, knowledgeable and wise guidance and support through every phase of the project. I value your professional experience and knowledge. Furthermore, you always do everything you can to enrich the life of others, and you rejoice on behalf of others, but you also set the standards and keep the facts straight. You have taught me so much in so many aspects of life and I will be forever grateful for your generosity. To my co-supervisor Simon-Peter Neumer, first of all thank you for SMART. When you lacked tools in CBT you made your own. You always take action with great professionality and generosity. SMART has been

a valuable tool in CBT, structured enough to hold on to the core ingredients of CBT, yet open enough to perform tailored treatment. And importantly it focuses on resources and goals in everyday life. Thank you for lending me SMART already in 2008. Since then, it has been widely used in CAMHS. In Tromsø there are regular SMART groups in CAMHS. Thank you so much for your generosity in this project within the field of research methods, emotional disorders, manualized treatment. I am forever grateful for your kindness, constructive and wise feedback, support, patience, creativity, and for all the laughs we had. To my co-supervisor Bjørn Helge Handegård I am amazed of your detailed knowledge in statistical methods. I am so grateful that you have been so generous, and patient in sharing your knowledge. You could translate whatever problem or challenge to statistics, and you are generous and kind. You are like an academic big brother who always have my back. Kenneth Fagermo my co PhD-student, you have done amazing, extensive work with punching and organizing all of the data, your sense of order and conscientiousness has been of outmost importance. I really appreciate your involvement in SMART. Thank you for all discussions, contributions and support. I am grateful for continuing our collaboration in supervision of SMART groups and the engagement you show in being a group leader in SMART.

Thanks, to all dedicated SMART therapists. Your dedication to your patients, to be really SMART therapists, to adhere to protocol while also being creative and playful in tailoring SMART to the needs of every adolescent. All this in the frame of routine clinical care. You are true SMART tailoress. Thanks to each and every one of you. Because you are a part of the data material I cannot mention specific names, but you know who you are and how I feel. Especially thanks to those of you who stayed over years in the project. A special thanks to the two most experienced SMART therapists who also were supervisors in Bodø and Stokmarknes. It's been an honor and so much fun that you were dedicated with the project all these years. You all really gave your heart to being SMART therapists and helping each and every adolescent, but also developing the method. I have learned so much from you and are forever grateful.

Thanks to all the leaders and administrative staff of CAMHS. With your willingness to take on such a project in ordinary care, you have believed in the aims of providing better services and that knowledge through performing research in ordinary clinical practice gives both new knowledge and the positive synergy effects of heightening the focus on treatment for emotional disorders. Through implementing evidence-based CBT you all

made a substantial effort to bridge the gap between research and clinical practice. First of all, thanks to Børge Mathiassen for your encouragement and belief that this was important to do from beginning to the end. Thanks for all the discussions and help to solutions throughout the project. To all the leaders in Child and Adolescent Mental Health Services (CAMHS), thank you for your enthusiasm, organizational skills and for being so cooperative and persevering in finding solutions and seeing this through. Lene Danielsen, Elsa Risjord, Hege Jacobsen, Jon Tomas Finnsson, Tarald Sæstad, Sari Venalainen, thank you! To all the administrative staff, you have been indispensable with your insistent energy and persistence. A special thanks to the key persons organizing the data collection. You rock!

A special expression of gratefulness to Northern Norway Regional Health Authority for financing the main study, and Regional Centre for Child and Youth Mental Health and Child Welfare North, for financing a pilot of the study. I would also like to thank my two workplaces Department of Psychology (IPS), UiT, The Arctic University of Norway and, University Hospital of North Norway (UNN), for continuing support for seeing this through.

I would like to thank all the employees and friends at CAMHS Tromsø who contributed one way or another to see this through. I also would like to thank my colleagues and friends at IPS and the clinical research group at IPS. I feel lucky to have such workplaces with colleagues like you. Thanks for all discussions, kindness, support and laughter.

Last but not least, a very special thanks to all my family and friends. Special thanks to my big family for being so empowered, strong, including, humorous, and filled with love and companionship. Oddmar, my love and life partner. I thank you the most for holding on to the importance of all other things than work in everyday life. Hanna and Johannes, you are the lights of my life, and I am so grateful and proud to be your mother. You teach me what life is all about every day. I`m deeply grateful. I love you very much.

Summary

Anxiety and depression are the most prevalent mental health problems among children and adolescents, and the consequences can be harmful if they are left untreated. Providing short and effective treatment for this group is paramount. When this effectiveness study of The Structured Material for Therapy (SMART) was conducted, it was the first Randomized Controlled Trial (RCT) of short-term transdiagnostic cognitive behavioral therapy (CBT) for adolescents with combined emotional problems, in regular clinical settings in Child and Adolescent Mental Health Services (CAMHS) in Norway. Measures for patients with emotional problems validated in the proper age group are important both for screening, tracking, providing feedback and documenting outcome for adolescent patients.

This thesis investigates the treatment effects in an RCT, of a six-session cognitive behavioral treatment of emotional disorders in adolescents (14-17 years) and investigates the factor structure and psychometric properties of the Clinical Outcomes in Routine Evaluations (CORE-OM) as a transdiagnostic outcome measure for adolescents with emotional disorders. The thesis also investigates the long-term effect and change trajectories of anxiety and depressive symptoms in adolescents, over the course of the SMART treatment and a 6 months follow-up period.

Results from this RCT are promising and indicate that a transdiagnostic CBT can be effective for youth with emotional problems treated in CAMHS, with significant decrease in overall emotional symptoms and significant increase in daily functioning both at post-treatment and at follow-up performed after 6 months. SMART can be considered as a first step in a stepped care model for treatment of anxiety and/or depression in CAMHS. Finally, the validation of CORE-OM revealed a somewhat differing factor solution and a higher cut-off score for adolescents than what is reported in adult samples, demonstrating the need for exploring psychometric properties of transdiagnostic instruments used to measure treatment effects in adolescents in CAMHS.

Abbreviations

BDI-II	Beck Depression Inventory, version II
CAMHS	Child and adolescent mental health services
CBT	Cognitive behavioral therapy
CFA	Confirmatory Factor Analysis
CGAS	Children's Global Assessment Scale
CORC	Child Outcomes Research Consortium
CORE-OM	Clinical Outcomes in Routine Evaluation – Outcome Measure
CTACS	Cognitive Therapy Adherence and Competence Scale
DAWBA	Development and Well-Being Assessment
EFA	Exploratory Factor Analysis
MASC	Multidimensional Anxiety scale for children
MBC	Measurement-Based Care
RCI	Reliable Change Index
RCT	Randomized Controlled Trial
ROM	Routine Outcome Monitoring
SDQ	Strengths and Difficulties Questionnaire
SMART	Structured Material for Therapy
WLC	Wait List Control
ES	Effect Size

List of papers

1. Lorentzen, Veronica; Fagermo, Kenneth; Handegård, Bjørn Helge; Skre, Ingunn; Neumer, Simon-Peter. A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14–17 years old in child and adolescent mental health services (CAMHS). BMC Psychology 2020; Volume 8 (1). ISSN 2050-7283.s 1 - 12.s doi: 10.1186/s40359-020-0393-x.

2. Lorentzen, Veronica; Handegård, Bjørn Helge; Lillevoll, Kjersti; Solem, Kenth; Moen, Connie Malén; Skre, Ingunn. CORE-OM as a routine outcome measure for adolescents with emotional disorders: factor structure and psychometric properties. BMC Psychology 2020; Volume 8 (86). ISSN 2050-7283.s 1 - 14.s doi: 10.1186/s40359-020-00459-5.

3. Lorentzen, Veronica, Fagermo, Kenneth, Handegård, Bjørn Helge; Neumer, Simon-Peter; Skre, Ingunn. Long-term effectiveness and trajectories of change after treatment with SMART, a transdiagnostic CBT for adolescents with emotional problems. BMC Psychology 2022; Volume 10 (167). ISSN 2050-7283.s 1 – 18.s doi: 10.1186/s40359-022-00872-y.

1 Introduction

By using a randomized controlled design in clinical practice, this thesis is a contribution to bridging the gap between research and practice, providing new knowledge regarding effectiveness, and long-term effects of the Structured Material for Therapy (SMART) program performed within the complex setting of ordinary care, with patients and employees in Child and Adolescent Mental Health Services (CAMHS). The outcome was evaluated with measures of symptoms and functioning, as well as measures of alliance, treatment integrity and user satisfaction, providing information about effectiveness and acceptability, as well as feasibility of this transdiagnostic intervention. The "transdiagnostic" approach, gives opportunities for novel insights into disorders within the emotional spectrum. Looking at trajectories for symptoms of anxiety and depression over the course of a transdiagnostic treatment and at 6-month follow-up, contributes to new knowledge about trajectories of change. The thesis also provides new knowledge about a transdiagnostic routine outcome measure for emotional symptoms and other problems for use in an adolescent population.

1.1 Anxiety and depression in adolescents

Feelings of anxiety and depression are a normal part of life. Anxiety is a necessary and important emotion. The reactions of "fight, flight and freeze" have kept us alive through history (Donahue, 2020). Worry and fearfulness signal to us that a danger or a sudden, threatening change is near so that we can make the appropriate response. Yet, anxiety can sometimes become an exaggerated and unhealthy response. Being a teenager with a life that is full of changes and uncertainties, anxiety may hum along like a background noise. For others, this time of life, anxiety can become a chronic, generalized, startling state interfering with daily functioning. School, leisure activities, making and keeping friends, and maintaining supportive relationships both within and outside the family can become a struggle. Sometimes anxiety can be a generalized, undefined, free-floating feeling of

uneasiness, as e.g., in generalized anxiety disorder (World Health Organization, 1992), at other times it develops into panic attacks and phobias. The object of anxiety varies. In phobia the anxiety provoking trigger is a specific object (spiders, waters, heights, balloons, etc.). In social phobia (World Health Organization, 1992) the trigger is being in social settings. In panic disorder the trigger is the body's own signals. The thoughts are often characterized by rumination and worries. The bodily reactions are fight, flight and freeze and the behavior is often dominated by avoidance.

Like anxiety, sadness is a part of normal life, but for some teens, the lows are more than just temporary feelings, they are symptoms of depression. Depression causes a persistent feeling of sadness and loss of interest (World Health Organization, 1992). It affects how the adolescent thinks (e.g.: negative feelings about self, others and the future). The emotional and cognitive changes involve feelings of sadness without any specific reason, frustration or anger, even small things matter, as well as hopelessness and feelings of emptiness. There is also a loss of interest or pleasure in usual activities or in being with family and friends. The self-esteem is low and the feelings of guilt and worthlessness dominate, leading to self-blame and self-criticism, sensitivity to rejection or failure, which again leads to drawback. The bodily symptoms are many: insomnia, fatigue, changes in appetite, use of alcohol or drugs, agitation and restlessness, slowed thinking, speaking or body movements and body aches. This often leads to social isolation, less attention to personal hygiene and the worst-case scenarios can be self-harm and suicide.

The most frequently diagnosed mental health disorders, both in the general population, and in child and adolescent mental health outpatient services (CAMHS) are anxiety and depression (Chavira et al., 2004; Costello et al., 2003; Gore et al., 2011). Symptoms of emotional problems negatively interfere with several aspects of functioning and quality of life (Copeland et al., 2014; Jaycox et al., 2009; Mendlowicz & Stein, 2000; Van Ameringen et al., 2003; Wittchen et al., 1998). There is an increasing prevalence of both anxiety and depressive disorders during adolescence (Merikangas & Knight, 2009; Silverman & Field, 2011). Anxiety and depression both have shared and separate etiology and features. The comorbidity and co-occurrence of anxiety and depression is high (Cummings et al., 2014). Hence, treatments that target both disorders could represent effective treatments for these complex disorders (Essau & Gabbidon, 2012).

1.1.1 Anxiety

Up to 10% of children and 20% of adolescents will meet the criteria of an anxiety disorder in the general population (Essau & Gabbidon, 2012). Adolescents with a primary anxiety diagnosis display more often than children co-occurring primary diagnosis of social anxiety disorder, diagnosis and symptoms of mood disorders, and uneven attendance to school (Waite & Creswell, 2014). As children proceed into adolescence, differences in the prevalence of particular anxiety disorders emerges, with decreased rates of separation anxiety disorder (SAD) (Cohen et al., 1993; Compton et al., 2000; Copeland et al., 2014; Costello et al., 2003) and increasing rates of agoraphobia, panic disorder, and obsessive compulsive disorder (OCD) among both boys and girls (Costello et al., 2003; Ford et al., 2003), with particularly generalized anxiety disorders (GAD) and social anxiety disorder in girls (Copeland et al., 2014; Costello et al., 2003). The same results are found in a recent Norwegian registry study, showing increasing incidence of diagnosed anxiety disorders over time, especially in girls (Ask et al., 2020). This gender difference was also present in a Danish cohort study showing that anxiety disorder was the most common diagnosis in girls (7.85%) with increasing incidence and age of onset (Dalsgaard et al., 2020). Overall, these studies showed increasing incidence of anxiety (Ask et al., 2020; Dalsgaard et al., 2020) with increases in overall rates across anxiety disorders from early adolescence to young adulthood (Copeland et al., 2014).

1.1.2 Depression

Depression is a leading cause of disability worldwide, and a major contribution to the global burden of disease (WHO, 2021). Adolescence puts you at high risk of developing depression. Before the age of 18, the percentage of adolescents with major depressive disorder range from 8 to 20% (Cheung & Dewa, 2006; Hankin et al., 1998; Kessler & Walters, 1998; Naicker et al., 2013). The Treatment for Adolescents with Depression Study (TADS) shows variations in the duration of the current major depressive episode with range from 3 to 572 weeks (0–11-year-olds), with a median duration of 40 weeks (Treatment for Adolescents with Depression Study, 2005). The mean duration for a depressive episode has been estimated to be between 4 and 9 months among clinically referred adolescents (Birmaher et al., 2002; Emslie et al., 2003). Around 2/3 of adolescents

diagnosed with depression at age 15 remain depressed at age 20. In this study the stability was apparent in both genders (Agerup et al., 2014).

The point prevalence of major depressive disorder (MDD) in adolescence is one in 20, and consequently accompanied by significant morbidity and risk of mortality (Essau & Dobson, 1999; Lewinsohn et al., 1993). In adolescence there is a dramatical increase in the rate of depression, with the rate in girls exceeding that in boys by about two to one at age 14 years (Dalsgaard et al., 2020). Relapse rates are relatively high (Lewinsohn et al., 1994). Suicide is the third leading cause of death in adolescents (Arias et al., 2003), and depression is an important contributor to suicidal behavior and death by suicide in adolescents (Brent et al., 1999; Brent et al., 1996). Hence, depressive symptoms in youths are risk factors for depressive disorder, suicidal risk, and long-term impairment into adulthood (Essau & Dobson, 1999; Weissman et al., 1999). Hence, both prevention and treatment of MDD in adolescents are of vital importance to public health. The field of research on depression is relatively new, from the 80`s and onward. Compared to other areas of research in the field of children and adolescents, depression itself, and treatment of depression are belated in advances compared to other fields of psychopathology research for this age group (Weisz, McCarty, et al., 2006).

1.1.3 Comorbidity in emotional disorders

Adolescents with comorbid anxiety and depression constitute a high-risk group, with complications affecting both functioning in life (Kendall et al., 2010), severity of symptoms (Cummings et al., 2014), quality of life, and long-term prognosis (Kendall et al., 2010). Young people that have a primary diagnosis of anxiety or depressive disorder are almost 30 times more likely than youngsters without this type of primary diagnosis, to suffer from the other disorder (Costello et al., 2003). Depression and anxiety are the most frequent co-occurring conditions in youth (Avenevoli et al., 2008). Although anxiety in early age can remit spontaneously, the majority of young people will relapse to the same or other mental disorders (other anxiety disorders, depression or substance abuse) over their lifetime. Across the range of anxiety disorders there is an increased risk of depression (Beesdo et al., 2009). This co-occurrence of anxiety and depression could imply common genetic vulnerability and underlying mechanisms that are shared between the disorders

(Garber & Weersing, 2010). As an example, a risk factor for developing both anxiety and depression is behavioral inhibition (Kagan et al., 1987). There are also cognitive commonalities such as dispositions towards negative beliefs of the world, emotional commonalities such as increased emotional reactivity, and behavioral commonalities such as avoidance as coping strategies for reducing negative emotions. The commonalities and the interrelations between anxiety and depression are so strong that some have suggested that psychiatric disorders could be an artifact (Maj, 2005), arising from the structure of the categorical classification system itself, rather than the co-occurrence of genuinely separable syndromes (van Loo & Romeijn, 2015).

Prognosis for comorbid anxiety and depression is worse than for either disorder alone, with a higher risk of recurrence, longer duration, greater impairment and less favorable response to treatments (Birmaher et al., 1996; Ezpeleta et al., 2006; Liber et al., 2010; O'Neil & Kendall, 2012; Rapee et al., 2013).

1.2 Psychological treatment of emotional disorders in adolescents

This study is to our knowledge the first to study the effectiveness of transdiagnostic treatment in adolescents within a context of routine clinical care. To draw relevant comparisons to other studies to illustrate the field is challenging, because the studies performed vary in methodological factors, treatment method, treatment length, inclusion criteria, object of research, and measuring efficacy and/or effectiveness. However, some guidelines and meta-analyses stands out as being representative for the tendencies and paradoxes shown in the field. Firstly, clinical practice guidelines recommend psychotherapy, pharmacotherapy, or a combination of both, in treating youth with depression (Cheung et al., 2018; Dolle & Schulte-Körne, 2013; McDermott et al., 2010). Although cognitive-behavioral therapy (CBT) is the most frequently evaluated and effective treatment for children and adolescents with anxiety symptoms, meta-analyses evaluating the effectiveness of psychological interventions for depression in adolescents found no evidence for the superiority of a single approach (In-Albon & Schneider, 2007; McKinnon et al., 2018). CBT is the only recommended first-line treatment by the Royal

Australian and New Zealand College of Psychiatrists treatment guidelines for anxiety disorders (Andrews et al., 2018). In the United Kingdom, the National Institute for Health and Care Excellence guidelines for social anxiety recommends CBT as first-line treatment for young people (NHS, 2013).

A very central meta-analysis by Weisz et al. (2017), sums up the field spanning over five decades concerning overall effects of youth psychological therapy. The study showed beneficial overall effects of youth psychological treatment that are moderate in magnitude and relatively durable with a post-treatment mean effect size (ES) of 0.46 (Hedges' g corrected for small samples) and mean follow-up ES at 0.36. (Weisz et al., 2017). The meta-analysis offered mixed evidence, with no overall treatment type moderator effect, but showed an interesting interaction between treatment type and informant, showing that youth-focused behavioral treatments (including CBT) produced the most robust cross-informant evidence of beneficial effects (Weisz et al., 2017) with significant effects across youth, parent, and teacher informants (Weiss & Weisz, 1995; Weisz et al., 1995). Effects also differed markedly by target problem. Target problem was the most potent moderator of treatment effects for youth psychological therapy, showing largest effects for anxiety with post-treatment ES at 0.61 and ES at 0.55 at follow-up. On the other side of the continuum of ES was depression with post-treatment effect sizes at 0.29 and follow-up at 0.22. The weakest ES was effects of overall youth psychological therapy on youth with multiple problems, showing ES post-treatment as low as 0.15 and 0.02 at follow-up (Weisz et al., 2017). However, inclusion of the studies of multiple problems was not exclusively based on anxiety and/or depression. The definition of multiple problems rested on studies that targeted multiple problems in concurrent treatment within one treatment period. The inclusion criteria were wide concerning type of problems, research setting, age, treatment and so on. The findings suggested that concurrent treatment of multiple problems showed noteworthy smaller mean ES than treatment of any of the treatments targeted at single problems, not significantly different from zero at post-treatment or follow-up. Further, the results can be caused by other methodological challenges which could have led to lower the ES for this approach (Weisz et al., 2017). Anyhow, the findings of these studies combined could show that efforts to treat multiple problems simultaneously has proven less successful than to treat problems in a more narrow way (Craske et al., 2007). Given that comorbidity, especially in anxiety and depression, is high (Cummings et al., 2014), there is a need to both treat and investigate complex

psychological problems in a different way. On the other hand, it is also a problem that studies that target single problems do not focus on common comorbidities, this could represent a problem because of the comorbid and pervasive nature of problems in clinically referred youths (Angold et al., 1999; Costello et al., 2003). The lacking body of research concerning treatment on multiple, concurrent problems, combined with the weak effects could imply a need for new ways to address youth comorbidity (Barlow et al., 2010; Bearman & Weisz, 2015; Ehrenreich-May & Chu, 2013; Weisz et al., 2015).

CBT is shown to be both an effective and empirically supported treatment for children and adolescents with anxiety, depression, OCD and PTSD (Post Traumatic Stress Disorder) (Dorsey et al., 2017; Freeman et al., 2018; Higa-McMillan et al., 2016; Silverman et al., 2008; Weersing et al., 2017; Wergeland et al., 2021). However, the evidence mainly consists of efficacy studies comprised of of RCTs performed in research settings (Lee et al., 2013; Weisz, Kuppens, et al., 2013).

In her thesis, Sørheim Nilsen (2017) sums up the studies of effectiveness of treatments performed in CAMHS routine care. They show positive results on a group level, although to a moderate extent (Sørheim Nilsen et al., 2015). In most studies, surprisingly few individuals obtain clinically significant improvements (Sørheim Nilsen, 2017). When it comes to effectiveness studies in children and adolescents with internalizing disorders, few reviews and meta-analyses have been conducted (Wergeland et al., 2021). In a recent meta-analysis and review of effectiveness studies performed in CAMHS with children and adolescents, the conclusion was that the quality of the studies was fair, but the heterogeneity high, and the analysis concludes that CBT delivered in routine care is efficacious in reducing internalized disorders and that the outcomes are comparable with results obtained in efficacy studies (Wergeland et al., 2021).

When it comes to treatment length it is difficult to find representative data on average contact time and treatment response. However, naturalistic outcome data from 1641 adolescents in 60 mental health services in England reported an average contact length of 31 weeks (Krause et al., 2022). Some studies indicate that briefer treatments may potentially be as effective as lengthier ones, and that treatment duration was not significantly correlated to outcome (Weisz, McCarty, et al., 2006). In a recent meta-analysis on CBT for internalizing disorders in routine care, number of sessions or weeks of treatment did not moderate the effect size and remission rates (Wergeland et al., 2021).

This could give some indications that shorter treatments could be as effective as lengthier treatments.

1.3 Transdiagnostic treatment

As a natural answer to the high comorbidity of problems, and according to the multidimensional model for mental disorders, integrative intervention approaches with a focus on shared transdiagnostic processes has grown forward (Cook et al., 2017; Emmelkamp et al., 2014; McHugh & Barlow, 2010). According to Mansell et al. (2009) one can define a psychological process as "transdiagnostic" if: 1) you can measure it in both clinical and non-clinical samples, 2) it emerges across diagnostic boundaries, and 3) it illuminates the development and maintenance of several disorders simultaneously. Transdiagnostic processes have been targeted in the development of a large array of treatment protocols categorized as transdiagnostic (Gros, 2014; Harvey et al., 2011; Norton, 2012; Sauer-Zavala et al., 2020). According to Carlucci et al. (2021) these treatment protocols, compared to protocols aimed for treating single-disorders, offer several advantages both in a clinical and practical manner. These include: 1) treating comorbid symptoms and sub-threshold symptoms that do not meet diagnostic criteria for specific diagnostic categories (Craske et al., 2007; Gibbons & DeRubeis, 2008), 2) potentially increase therapeutic effects by targeting the core mechanisms of several disorders (Brown et al., 2001; McManus et al., 2010), 3) cost benefit, in terms of scalability and cost-effects when delivered in different formats (i.e.: web-based or group-based format), and 4) to save time and resources when training clinicians (Steele et al., 2018). Transdiagnostic treatment for emotional disorders focuses on the cognitive, behavioral, and physiological processes that are shared or common across diverse disorders. The process of careful differential diagnosing is not necessary, and represents an integrative approach (Mansell et al., 2008).

1.3.1 Transdiagnostic interventions categories

As stated in Dalgleish et al. (2020), transdiagnostic psychological interventions fall into two broad categories: universal and modular approaches (Meidlinger & Hope, 2017; Sauer-Zavala et al., 2017). Universal interventions such as the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders (Barlow et al., 2010), provide the same set of therapeutic elements to all clients. The elements are selected to have broad applicability across diagnoses (Dalgleish et al., 2020). This gives several advantages when it comes to practical issues such as time spent on diagnostic processes or resources spent on selecting the right intervention elements beforehand. Clinical training also becomes more efficient (Steele et al., 2018). However, there is a risk that the universal approach do not fit all, and that they give too much attention to irrelevant topics, or too little attention to topics relevant for the patient. The other approach to transdiagnostic intervention is the modular approach, such as "The Modular Approach to Therapy for Anxiety, Depression", Trauma, or Conduct Problems (MATCH-ADTC) for children (Chorpita et al., 2005). According to Dalgleish and colleagues (2020), modular approaches are comprised of sets of evidence-based, self-contained functional units (therapy modules), that can be delivered flexibly and operate independently. This has the advantage that module selection and order are tailored to the needs of each client so that it potentially involves a better goodness-of-fit between the therapy and how the individual presents clinically. The downside to this is that delivering modular interventions are more demanding than universal protocols, because the therapists must match the needs and requirements of each client when choosing and delivering the modules (Dalgleish et al., 2020). The SMART program is both a universal and modular approach to transdiagnostic treatment. It is developed in a universal manner, but can also be applied as a modular approach.

1.3.2 Linking transdiagnostic interventions to transdiagnostic processes

As stated by Queen et al. (2014), there are several shared risk factors in anxiety and depression regarding both psychological, biological and environmental factors. For instance common neuroendocrine (Weems et al., 2005) and neurotransmitter dysregulation (Flores et al., 2004; Fox et al., 2005), cognitive risk factors as rumination and worry (Olatunji et al., 2013), and behavioral avoidance (Aldao et al., 2010). In accordance with these findings, negative affect has shown to be a latent factor underlying both depressive

and anxiety disorders (Brown et al., 1998; Trosper et al., 2012). Pharmacological and psychosocial interventions have shown similar responses to e.g., selective serotonin reuptake inhibitor (SSRI) (March et al., 2004; Walkup et al., 2008). These common elements in emotional disorders are targeted in evidence-based youth treatments with the purpose of developing cognitive behavioral skills (Chorpita & Daleiden, 2009). In accordance with these findings, CBT treatment trials targeted to treat single disorders have shown so called "spill-over effects" onto comorbid anxiety and depressive disorders (Kendall et al., 2004). One example of this is a meta-analysis of CBT designed for treating primary depression, which demonstrates effects for the targeted disorder, depression, but also display "spillover effects" on anxiety symptom reduction (Weisz, McCarty, et al., 2006). Prior research efforts on evidence-based treatments for youths have been unable to examine the concurrent trajectories of primary anxiety and depressive symptoms across the course of treatment (Queen et al., 2014). Transdiagnostic treatment creates an opportunity to examine whether there are separate trajectories of change in this group when it comes to depression, anxiety, and comorbid anxiety and depression following a shared intervention.

1.4 Efficacy and effectiveness

To determine whether interventions demonstrate good results in regular clinical settings, effectiveness studies are important (Glasgow et al., 2003; Marchand et al., 2011). In efficacy trials, the interventions can be examined under strict, highly controlled and optimal conditions, while performing the same interventions in regular clinical settings, researchers naturally have reduced control over the research setting (Marchand et al., 2011). In regular clinical settings, the service providers have less time and resources to perform optimal training and supervision. The clinical problems of the patients are also more complex. From a research perspective this represents non-optimal conditions (Weisz, McCarty, et al., 2006). The randomized controlled study design (RCT) is considered the "gold standard" for evaluating health care interventions (Schulz et al., 2010), giving the best opportunity to control for diversity in the intervention groups. There is an ongoing debate concerning what kind of research design should constitute evidence-based practice

in psychology. A central issue in this debate is to what degree the body of evidence attained by RCT`s can be generalized to routine clinical practice (Hunsley & Lee, 2007; Norcross et al., 2006), further whether implementation of methods attained by RCT`s can achieve better clinical improvement, or if the experimental control of efficacy trials leads to a decrease in the external validity and limit the generalizability of the results (La Greca et al., 2009). The question as to whether therapists, clients or training are equal or differ in research clinics and community clinics in significant ways, rises many underlying issues (Hunsley, 2007; Lee et al., 2013; Weisz, Ng, et al., 2013; Weisz, Ugueto, et al., 2013). One might question whether manualized treatment are feasible in regular clinical practice, and to what degree can it be delivered under the constraints in regular clinical settings (Hunsley & Lee, 2007). All these concerns stem from scientific standards required in an RCT, including more homogenous samples of patients due to inclusion criteria where the targeted disorder(s) of the treatment is carefully diagnosed before treatment begins. In RCTs inclusion is often dichotomously defined, by patients having the diagnosis or not, rather than looking at symptoms in a continuous way. Also, the patients are randomly assigned, and the therapists are trained for the targeted intervention (Hunsley, 2007; Nathan & Gorman, 2015; Weisz, Doss, et al., 2005; Weisz, Ng, et al., 2013). In regular clinical practice it is less common that the diagnosis alone indicates the treatment. The connection between diagnostic assessment and treatment is less stringent, due to variation of available interventions for a diagnosis and individual competencies and preferences by therapists. There is also currently a demand for user involvement, where patients can choose treatment, and even choose therapist, based on information and perceived problems and goals. Therapists in research clinics have more focus on treatment integrity through supervision, more access to extensive training and treatment monitoring than their colleagues in regular clinical settings (Smith et al., 2017). In the design of effectiveness studies, the therapists delivering the interventions are already working in diverse clinical settings where they are expected to apply many differential methods to meet a broad array of disorders and referral problems of diverse ordinarily referred patients. The therapists are often generalists with extensive caseloads (Hunsley, 2007; Weisz, Ng, et al., 2013; Weisz, Ugueto, et al., 2013), not only concerning clinical methods and supervision, but also concerning tasks involving assessment, treatment monitoring, and rescheduling of missed appointments or follow-ups (Wergeland et al., 2021). In sum, it is more demanding to perform the methodological rigor demanded in a RCT in community clinics than in research clinics. As such, when implementing treatment programs that are developed and

evaluated under settings consisting of highly controlled conditions there is a risk that they may give poorer results when delivered in routine clinical care (Baker-Ericzén et al., 2010; Weisz, Jensen-Doss, et al., 2006; Weisz, Ng, et al., 2013; Wergeland et al., 2021).

Efficacy and effectiveness studies are often discussed as if they were polar opposites. A more useful approach is viewing them as research designs with different foci. Overall efforts in understanding and evaluating the potential effect and impact of a treatment should include both efficacy and effectiveness studies (Hunsley et al., 2014), where efficacy can test the initial effect of an intervention under optimal conditions and an effectiveness study can investigate to what degree the intervention keeps its impact when applied in daily routine in community clinics. However, there are different opinions concerning the sequence of the two designs, where some question the confidence in results obtained in an uncontrolled effectiveness study without established efficacy evidence (Hunsley & Lee, 2007), while others state that the demand of several efficacy studies preceding effectiveness studies is more appropriate in evaluating pharmaceuticals than it is for psychological interventions (Weisz, Jensen, et al., 2005).

1.5 Measurement-based care (MBC)

Although it is established that psychotherapy works, it is also known that it does not work for everyone. Some patients even deteriorate in treatment (Lambert, 2013). One of the most important tasks for psychotherapy is to gain more knowledge of what works for whom and how to treat those individuals who deteriorate or have no benefit from therapy (Lambert, 2010; Prescott et al., 2017). In this task, using a tool to bridge the gap between research and clinical practice could involve the use of Routine Outcome Monitoring (ROM) (Lambert, 2010; Prescott et al., 2017). In ROM, instruments are usually applied in a measurement feedback system, where the aim is to increase treatment personalization and clinical responsiveness. Such systems are the core tools in Measurement-Based Care (MBC). ROM implies administrating outcome assessments periodically to track client treatment progress. The running aim is to interpret ROM data in order to predict trajectories of change, flagging clients who may be at risk for negative outcomes and

suggesting appropriate clinical interventions. These tools are supposed to lift clinical decision-making by giving systematic feedback on clients' current functioning and predicted trajectory (Bugatti & Boswell, 2022). The need for such data-based feedback is supported by existing findings that clinical judgment alone has limited predictive accuracy (Hannan et al., 2005; Walfish et al., 2012). The effectiveness of MBC is supported by a significant body of research (Bugatti & Boswell, 2022). MBC facilitates the achievement of improved treatment outcomes (Lambert et al., 2018), while also improving clinicians' ability to identify and respond to deviations from trajectories of change that otherwise could lead to negative treatment outcomes (Hannan et al., 2005). In this sense, ROM has important qualities for future psychotherapy research; ROM can be conducted in routine clinical practice and based on standardized measurement systems, they allow clinicians to take an active part in research by using data from their own patients in clinical practice (Castonguay, 2013). The use of ROM has been shown to improve outcome in therapy in numerous of studies (Amble et al., 2015; Bickman et al., 2016; Bickman et al., 2011; Shimokawa et al., 2010; Wampold, 2015). As the patients themselves are the one providing feedback on their own progress, ROM has been attributed to increased user involvement in mental health services (Ulvestad et al., 2007). In addition to giving insights into process and patterns of individual change, for both the patient and the therapist, systems for collecting routine practice data can give insights in organizational development (Barkham et al., 2010). There exist several systems for collecting, using and analyzing outcome data, for instance Outcome Questionnaire system (OQ-system) (Lambert, 2015), the Partners for Change Outcome Measurement System (PCOMS) (Duncan & Reese, 2015), the Clinical Outcome in Routine Evaluation (CORE) system (Barkham et al., 2001) and several others (Kraus et al., 2005; Pinsof & Chambers, 2009). As noted by Wampold (2015), the main aim is not to test what system works best but to search for the efficacious components in ROM.

1.6 The Clinical Outcome in Routine Evaluation (CORE-OM) as a Routine outcome measure in adolescence

There is a call for targeted ROM for adolescents in CAMHS. Valid and reliable routine outcome measures are key tools in detecting and preventing treatment failure (Lambert et

al., 2003). In addition, there is a call for transdiagnostic measures that address comorbidity and are sensitive to change. The self-report Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) is a well-established measure used in adult psychotherapy research and different mental health facilities and counselling services (Barkham et al., 1998; Barkham et al., 2001; Evans et al., 2002). The CORE-OM items were selected based on their clinical significance and sensitivity to detect change in psychological status (Barkham et al., 1998; Barkham et al., 2001). The measure is a questionnaire consisting of 34-items with a 5-point Likert scale from 0 (not at all) to 4 (most of the time). Theoretically the CORE-OM consists of four dimensions: Well-being (4 items); Functioning (12 items); Problems/symptoms (12 items); and Risk (6 items) (Evans et al., 2000; Skre et al., 2013). 'Well-being' refers to a patient's sense of life quality and emotional health. The items in 'Problems/symptoms' concern issues such as anxiety and depression symptoms, reactions to trauma, and physical complaints. 'Functioning' relates to interpersonal, social and general functioning in daily life. There have been found high correlations among these three domains (Elfström et al., 2013; Evans et al., 2002; Lyne et al., 2006), and it has been recommended that these three dimensions should be combined into a general psychological distress scale called All-items-minus Risk (Bedford et al., 2010). The 'Risk/harm' domain concerns items of harm to self and suicidal ideation (risk-to-self items) and violent behavior and threats towards other people (risk-to-others items) (Barkham et al., 2006; Evans et al., 2000). The aim is to help the clinician detect and monitor a patient's intentions of self-harm, suicide and violence. It has been recommended that Risk should be monitored separately (Bedford et al., 2010). The second study in this thesis aimed at examining the psychometric properties and factor structure of the CORE-OM in adolescent samples.

2 Thesis Aims

This thesis aims to provide new knowledge that can contribute to the evidence base of mental health services offered to youths with emotional problems. To reach this aim we performed a randomized controlled trial of the transdiagnostic treatment SMART, a six-

session cognitive behavioral treatment for emotional disorders in adolescents 14-17 years, within the context of regular child and adolescent mental health services (CAMHS). The thesis also aimed to validate a routine outcome measure by investigating the factor structure and psychometric properties of the transdiagnostic measure CORE-OM, as a tool in measurement-based care for adolescent populations. Regarding the randomized controlled trial, the first aim of this thesis was to investigate the short-term effectiveness of the transdiagnostic SMART treatment on emotional symptoms and functioning. Secondly, the aim was to investigate the long-term effectiveness and trajectories of change in emotional symptoms and functioning over the course of the intervention and a 6 months follow-up period.

To provide a better overview of the different studies, a summary of the aims of each paper is presented here:

2.1 Paper 1

In the first paper, we investigated the short-term effectiveness of the 6-weeks SMART program in CAMHS, by performing a RCT in three community CAMHS in Norway. We investigated effects in internalized symptoms, symptoms of depression, anxiety and general functioning. User satisfaction and alliance were also investigated.

2.2 Paper 2

In the second paper we aimed to validate the CORE-OM in adolescent populations, by examining the factor structure and psychometric properties of the CORE-OM in data obtained from the CAMHS sample in paper 1, and a sample of high school students. Factor structure, reliability and cut-off point were investigated.

2.3 Paper 3

In the third paper the long-term effectiveness of the SMART program, and the separate trajectories of anxiety, depression and combined anxiety and depression symptoms were investigated over the course of the transdiagnostic intervention described in Paper 1, and through 6-month follow-up.

3 Methods

3.1 The SMART study

The data presented and discussed in this thesis is a product of a multi-center study with a randomized controlled trial, following CONSORT guidelines (Schulz et al., 2010). In this study there were three participating sites; CAMHS Tromsø, CAMHS Stokmarknes and CAMHS Bodø. Assessments were completed by youths at pre-waiting list, pre-treatment, post-treatment and at 6-month follow-up. A block randomization was used in which groups of 5 youths meeting the inclusion criteria were randomized with a 1:1 allocation to either wait-list control (WLC) or direct treatment. To ensure blinding of the therapist, researchers and the participants to the allocation process, the participants were enrolled and assigned to treatment or WLC by administrative staff. The mean duration of WLC was 6.8 weeks, and the mean duration of the treatment condition was 10.3 weeks (Lorentzen et al., 2020). The overall explanations for prolonged treatment time were patients not showing up for treatment due to various reasons, vacations, and also therapist sick leave.

3.2 Treatment

The treatment received by the adolescents was the SMART program (Neumer & Junge-Hoffmeister, 2010a, 2010b). This is a Norwegian version of the GO! program, originally

developed and evaluated in Germany (Junge, 2003). The treatment was selected because the PhD candidate had used the program over several years in CAMHS as a clinical psychologist with good results and positive feedback from the patients, and aimed to do research on the program (Lorentzen, 2008). The SMART manual was at the start of the study the only freely available transdiagnostic, modularized CBT manual available in Norwegian. It is developed for 14–25 years old adolescents and younger adults. SMART is based on well supported methods for treating anxiety and depression, with a strong emphasis on cognitive restructuring, exposure and activation. Originally, SMART is an 8-week manual-based modularized CBT program for groups. In this study the program was delivered as individual therapy over 6 sessions. The special features of the program are as follows: definition of individual treatment goals, activation of personal resources, behavioral experiments, information about emotional problems and related coping strategies. The materials in the complete version are organized in five modules (introduction, depression, anxiety, assertiveness training, and summary, in a total of eight sessions). The modularized and flexible organization of the materials allow for the program to be shortened to four or six sessions by selecting modules, and can be used both in a group or an individual therapy format. These options have been applied in our study. All modules, except the assertiveness module (2 sessions) at the end of the original program, were employed as an individual standard brief therapy in the outpatient clinics. The reason for not employing the assertiveness module was that the WLC and the treatment group should initially be of the same duration, and due to health authorities' regulations, waiting time could not exceed 6 weeks for this patient group. We chose the modules that were targeting specific depression and anxiety symptoms. Figure 1 shows the four modules delivered over 6 sessions, each with a duration of 1.5 hours. The reason for delivering SMART in the current project as individual therapy was that the small treatment units in Northern Norway make it difficult to collect a sufficient number of adolescents for group treatment within a limited timeframe.

Figure 1

SMART as used in the study comprises of four of six sessions consisting of 1.5 hour

Topic	**Introduction**	**Depression**	**Anxiety**	**Summary**
Session nr.	1	2+3	4+5	6
Topics in session	Four-component of Thoughts, bodily reactions, feelings, behavior model is introduced Personal strengths Schedule of the week with monitoring behavior and feelings used for planning positive activities, and recognizing patterns. All topics in session 1 is used every session during the program	Homework from introduction Behavioral experiment Automatic thoughts and classical thought traps Cognitive restructuring of automatic thoughts Information about depression (causes, cycles, depressive triade, attributional errors in depression)	Information about anxiety. Vicious circle of anxiety, thoughts, feelings, bodily reactions, fight, flight and freeze in anxiety. Sustaining factors in anxiety Avoidance and exposure Anxiety diagnosis Progressive muscle relaxion (Bodyscan) Tools to manage anxiety	Summary, feedback
Homework	Resources/skills Personal goals (Social, education/work, personally) Schedule and planning of positive activity	Schedule and planning of positive activity Tools to identify Automatic thoughts and classical thought traps and applying Cognitive restructuring of automatic thoughts List of positive activities	Schedule and planning of positive activity Exposure Progressive muscle relaxion (Bodyscan) Tools to manage anxiety	

*Module 4 The assertiveness training is not included in this table, as it is not used in the study.

3.3 Recruitment, sites and therapists

The recruitment of the participating sites was done by the author, travelling to a number of different CAMHS presenting the project. The final participating clinics were three CAMHS in the north of Norway, Bodø, Tromsø and Stokmarknes, representative for public child and adolescent mental health outpatient clinics, covering both urban and rural parts of Northern Norway. Adolescents are usually referred to the CAMHS by general practitioners. Teams in CAMHS are multidisciplinary and work with diverse ages from 0 until 18 years old, and have different methods for treating a range of diagnoses in their caseload. Overall, twenty therapists participated in delivering the SMART treatment (*M* age = 39.18 years, *SD* = 10.93, range 24–57, 100% females).
The therapists had on average 6.8 years of clinical experience (*SD* = 8.23, range 0–32 years). Of the 20 therapists, 11 were psychologists, 2 psychology students, 4 pedagogues, 2 social educators and 1 was a public health nurse. Two of the therapists had a two-year specific education and training in CBT.

3.4 Training, supervision, treatment integrity, alliance and user satisfaction

The therapists had a 2-day training course in the use of the SMART manual, led by the PhD candidate. The training consisted of lectures, hands-on training, and role play. When therapists started using the SMART manual, they had bi-weekly supervision based on the Cognitive Therapy Adherence and Competence Scale (CTACS) (Barber et al., 2003). The supervision was mainly performed online by Skype by the PhD candidate and the copyright holder, who both are trained and experienced CBT therapists and CBT supervisors through the Norwegian Association for Cognitive and Behavioral Therapy. In the last part of the project, supervision in the Stokmarknes and Bodø sites was performed by two highly experienced SMART therapists, one in Stokmarknes, and one in Bodø.

Before supervision sessions, supervisor and therapists usually scored a video session with the CTACS and proceeded to discuss the scores and how to improve competence and

adherence. CTACS is a widely employed 21-item scale to measure therapist competence and adherence, and was used in the present studies as a part of the training and supervision of treatment integrity. The supervision also consisted of scoring the assessments, adherence to protocol and tailoring the CBT to each patient by using cognitive case conceptualization and feedback from sessions with patients (Neumer & Junge-Hoffmeister, 2010a). To continuously strengthening treatment integrity, periodic full-day booster sessions gathering all therapists were arranged. The administrative staff also participated in booster sessions where they shared challenges and solutions to adhere to the study protocol. The PhD candidate and the second author on paper 1 and 3, weekly monitored the adherence to protocol and had regular meetings to ensure protocol adherence. Routines to adhere to protocol was developed for each clinic and continuously evaluated and adjusted by the PhD student and the second author on paper 1 and 3 together with leaders, administrative staff, and therapists.

Evaluation of sessions (Neumer & Junge-Hoffmeister, 2010a) was used for measuring treatment integrity, alliance and user satisfaction. Following the SMART manual, this measure was completed after each session and at the end of therapy. Patients rated aspects of each session on topics concerning specific content and satisfaction with the session on a Likert scale from 1 "very unsatisfied" to 5 "excellent". The aim is to give the therapist opportunity to adjust therapy, alliance, content and relevance in close collaboration with the patient. In order to secure that treatment integrity was measured as near as possible to adherence of the elements in the treatment manual, the patient rated the *Evaluation of sessions* questionnaire addressing specific elements of each session (Likert scale from 1 "very unsatisfied" to 5 "excellent"). An adherence score was calculated, reflecting the degree (in percentage) to which each of the elements in the SMART manual had been included in the sessions (see Table 1, page 5 in paper 1). Three items from the *Evaluation of sessions* questionnaire were used for measuring *therapeutic alliance*: "I liked today's session" and "I felt understood by the therapist" (Likert scale from 1 to 5, ranging from negative to positive). Furthermore, items from the End of the therapy questionnaire were used as measures of *therapeutic alliance*: "Therapist's competence and presentation were", "Therapist's understanding was" and "Therapist's openness was" (Likert scale from 1 to 5, ranging from negative to positive). Finally, as a measurement of *user satisfaction*, two questions from the *Evaluation of therapy* questionnaire were employed:

"Overall the course was" (Likert scale from 1 to 6, ranging from negative to positive) and "I would recommend this course to others" (Yes/No).

3.5 Samples and data collection

The initial sample is the same in paper 1 and 3, and constitutes the clinical sample of paper 2. The overall analyzed sample comprised of 145 adolescents 14–17 years old ($M = 15.72$, $SD = 1.14$, 90.3% females), recruited from referrals to three Norwegian public child and adolescent mental health outpatient clinics (CAMHS) between January 2012 and November 2016. Information about the study was presented to the participants during the routine intake procedure of the clinic. All adolescents and parents of children under 16 years, and adolescents over 16 years signed informed consent and received the Strengths and Difficulties Questionnaire (SDQ).

Being the same sample, the inclusion criteria in the clinical sample were the same in paper 1 and 3 and for the clinical sample in study 2. Inclusion criteria were 1) age between 14 and 17 years; 2) a probable diagnosis of emotional disorder as indicated by a score of at least 6 on SDQ emotional problems subscale; and 3) maintenance of a maximum waiting time for necessary medical care of 6 weeks given by Norwegian health authorities. Exclusion criteria were 1) a diagnosis of pervasive developmental disorder (PDD); 2) psychotic symptoms; 3) anxiolytic or anti-depressant medication effects during the treatment period; and 4) patients who did not speak the Norwegian language. In paper 2, the inclusion criterion for the non-clinical sample was being a student at a junior or senior high school, while the exclusion criterion was that the adolescent could not read or write Norwegian fluently. The inclusion criteria in the clinical sample were the same as in paper 1 and 3.

A total of 199 adolescents were found eligible for the study and were asked for informed consent. Of these, 36 did not consent, 7 were excluded due to exclusion criteria (1,2,3,4), 11 withdrew from the study. The remaining sample of $N = 163$ were block randomized into direct treatment, or six-week wait-list. The participants who did not receive the intervention are accounted for in paper 1 in the consort diagram. Of the persons that did not receive the intervention, 11 adolescents withdrew their consent. When withdrawing, in

line with the consent (approved by the regional ethics committee), the patient`s data was deleted and were not be used as data in the analysis when withdrawing. Four of the patients did not meet the inclusion criteria for participating in the study (Suicidality, psychosis and PDD). Three of the patients chose not to receive the intervention, since they were referred to other interventions in CAMHS as they were unable to wait for six weeks. These three patients were originally assigned to the wait-list but it turned out that they could not wait six weeks before receiving necessary medical help. This violated the inclusion criteria of being able to wait for six weeks, hence they did not meet the inclusion criteria for participating in the study anymore. This will be further elaborated in the discussion section.

In paper 2, the clinical sample was the same as in paper 1 and 3, with a small difference in the sample size (n=140, (age 14–17, M = 15.72, SD = 1.15), out of which there were 13 boys (9.3%) and 127 girls (90.7%)). The non-clinical sample (n = 531) was recruited from four junior high schools and four senior high schools in both urban and rural areas in North Norway. The schools were randomized and drawn, and data were collected until at least 65 participants from each class grade were included. In the non-clinical sample (age 14–18, M = 15.91, SD = 1.45), there were 273 (51.4%) boys and 258 (48.6%) girls. The response rate on school level was 83.3% (5 junior high schools of a total of 6 consented) and 71.4% (5 out of 7 senior high schools consented). This information was not included in paper 2 due to limited space. Although the sampling procedure was systematic, the non-clinical sample should be viewed as a convenience sample. This sample was recruited as a part of a master thesis supervised by the PhD candidate, exploring psychometric properties of CORE-10. For more detailed information see Solem & Moen (2015).

3.6 Research ethics

The study was performed in compliance with the Helsinki Declaration for research on humans (World Medical Association, 2013) and its later amendments or comparable ethical standards. Permission to perform the study was applied for and was approved by the Regional Ethics Committee (REC North) in 2011 (Reference number 2011/1937). The study of CORE-OM in paper 2 was also applied for in REC north, and the response was

that we did not need additional approval from REC North because it was a psychometric investigation of instruments/anonymous routine data.

According to the Health Research Act (Helseforskningsloven, 2008) and national ethical guidelines (NEM; National Committee for Medical and Health Research Ethics; NESH; The National Committee for Research Ethics in the Social Sciences and Humanities) (NESH-Guidelines, 2016), all participants need to understand all aspects of a research project, including the purpose and consequences of participation. In this project, we included adolescents 14-17 years old. The adolescents under 16 years do not have the legal competence to give informed consent alone and are thus, by definition, labelled a "vulnerable group" (NESH-Guidelines, 2016). The clinical sample from CAMHS in Studies 1-3 participated and consented according to the regulations governing the research project, with written parental consent provided for those under age 16 in addition to the participants (REC North, Reference number 2011/1937). In the events of high scores on symptom and risk items, these were addressed by the responsible therapist or counsellor. We obtained informed consent from all individual participants included in the study, and informed consent from both parents and adolescents under 16 years of age.

In the school sample the school was told to inform the students where they could get help if they needed someone to talk to, that it was voluntary to participate, and that their response would be kept anonymous (see Solem & Moen (2015) for further information).

3.7 Registration and financing of the study

The study was planned in 2011 and, and received preliminary funding in 2012 from the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU North), and from January 2013 and onwards the study was fully funded by the Northern Norway Regional Health Authority (Helse Nord) health trust research fund.

After receiving permission for the study from the regional committee, the project was registered in EUTRO, the database for internal control and registering of ongoing research projects at the faculty of health sciences at UiT- The Arctic University of Norway. Registration was done in 2012, and the ID of the current research project in EUTRO is

2620.00012. Going back in the correspondence concerning the registration of the project, it is obvious that registration of research projects for internal control in 2012 was new both to the Faculty of Health Sciences and at The Department of Psychology, since all registration of ongoing projects at the department was performed simultaneously. The project was registered in Clinical Trials.gov (Identifier: NCT02150265) on May 29, 2014, two years after the project preliminarily commenced, and one year after the project was fully financed and fully launched, when the project leader became aware of the necessity to register the project in this database.

According to the registration the primary outcome measure is: "Status of primary emotional disorder based on the Development and Well-Being Assessment (DAWBA) interview schedule with parents and youths [Time Frame: At recruitment, before treatment start and at 6 months follow-up]. The dissonance between the trial registration and the actual study could solely be explained by the lack of knowledge of the importance of trial registration, as well as the sensitivity of time in registration of clinical trials. The research group specified early in the planning process how the analysis with the SDQ actually should be performed. Important decisions on details that have been made during the study should have been registered in time, and updated consequently.

3.8 Measures

3.8.1 Issues concerning choice of measurements for study entry and outcome measures

The change in primary measure was decided before the analysis of the data. Additionally, the power analysis in the initial study protocol was performed with continuous variables. The inclusion criteria, SDQ, is a part of the DAWBA package. The clinician-generated diagnoses in this study and the results from the DAWBA bands differed, and it was decided that the most reliable and resource manageable solution was to use DAWBA bands. In the studies constituting this thesis the DAWBA diagnosis is solely used for description of prevalence in the sample, and not as an outcome variable. The candidate has parallel to the studies presented in this thesis, supervised two master theses looking at the prediction of the instruments on diagnosis used in this study (Nakken, 2021; Skogly, 2015). These data will be further analyzed and described in a future paper along with clinician generated diagnosis and

DAWBA. When it comes to the multi-informant possibilities in SDQ and DAWBA, patients' ratings only, were used due to data quality from the informants. Collaboration with clinicians and management was central in the planning and execution of the study all the way to the choice of instruments. It was important that the study and instruments were feasible for the clinics. As a measure of depression, the BDI II (Beck et al., 1996) was chosen since this was already in frequent use and acceptable for the clinics. The other measure considered to measure depression was the Norwegian translation of the Mood and Feelings Questionnaire (MFQ) (Richter & Sund, 2013) that also has documented good psychometric properties, however only for adolescents 13-15 years of age. A Norwegian validation for adolescents of BDI II, based on the present sample and a school sample is planned and approved through REC north (reference number 2011/1937). As a measure of functioning, CGAS was already routinely in use in one of the clinics. MASC was chosen as an anxiety measure, and CORE-OM as a transdiagnostic measure for emotional problems. CORE-OM was already in use in the university clinic at UiT for adolescents (Solem & Moen, 2015), and there was a need to validate the instrument accordingly. Further descriptions of the instruments follow.

The same measures were employed in study 1 and 3, and were administered at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up. Study 2 was a validation of CORE-OM only.

3.8.2 Measures used in the study

Development and Well-Being Assessment (DAWBA) (Goodman, Ford, Richards, et al., 2000), translated to Norwegian language by Heiervang and colleagues (Brøndbo et al., 2013). DAWBA was a routine instrument in the intake procedures at the participating clinics during the recruitment period of the study. It is a diagnostic interview performed by web in this study. The interview can be performed both multi-informant (parents, school), or only by the patient. It contains both open- and closed-ended questions. In this study, only closed-ended information from the patients was used with the purpose of assessing diagnosis. It was administered at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up, as part of the DAWBA package. When completed online, DAWBA uses computer algorithms to suggest the likelihood of diagnoses resulting in band levels corresponding to the prevalence of the disorder (Goodman et al., 2011). The bands range

from levels 0–5 and are dichotomously combined to either 'absent' (levels 0–3; < 0.1 to 15% probability of disorder) or 'present' (Levels 4–5; ~ 50 to > 70% probability of disorder). As mentioned in paper 1 (page 3), Goodman and colleagues (2000) reported that DAWBA could discriminate between community and clinical samples of youth. Goodman et al. (2011) found that the DAWBA bands were well suited to find an approximate prevalence of disorders. Comparing the computer-generated DAWBA bands to clinician-rated diagnoses, Goodman et al. (2011) found that DAWBA underestimates the actual prevalence on a group level. Agreement on an individual level showed kappa values that were usually between 0.4–0.7, sensitivity 0.4–0.8, specificity 0.98–0.99, positive predictive values 0.5–0.8 and negative predictive values 0.96–0.99.

The Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997), Norwegian language versions published in 1999, based on a translation and back-translation by Heiervang and colleagues (Eidstuen & Kornør, 2017). The main inclusion criteria and the primary outcome measure for emotional symptoms was the SDQ emotional symptoms subscale. The SDQ was administered at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up, as part of the DAWBA package. The version used in Study 1 and 3 was the self-rated SDQ for 11 to 17-year-olds with five subscales. SDQ is a questionnaire, using a 3-point Likert scale, from 0 (not true) to 1 (somewhat true) to 2 (certainly true), giving a maximum score of 10 covering emotional and behavioral symptoms as well as computing an impact score. In this study we used the emotional symptoms subscale as inclusion criteria and a primary outcome measure. Goodman (2001) suggested a cut-off on the emotional problem subscale of 6/7. The scale has shown acceptable reliability and adequate internal consistency (Brøndbo et al., 2011; Bøe et al., 2016; Goodman, 2001). We used the norms from a Norwegian study (Rønning et al., 2004) to select 6 as the cut-off for study-entry. A cut-off of 6 has been shown to be appropriate to identify adolescents with emotional problems (Van Roy et al., 2006). In Rønning (2004) there were different prevalence on emotional disorders for boys and girls. Scores above cut-off on the SDQ can be used as a "warning signal" for initiating treatment (Goodman, 1997, 2001). SDQ is an overly used screening instrument and has satisfactory psychometric properties (Bøe et al., 2016; Goodman, 2001; Muris et al., 2003). The SDQ includes both multi-informant possibilities, and measures of impact (Goodman, 1999). However, adults and teachers are found to be insensitive to youths' emotional symptoms, and might have limited insight concerning important arenas in adolescent´s daily life (e.g., the parent's impression of

their child's school situation). Hence, youth self-report of emotional symptoms is considered to be the best source for identifying emotional problems (Aebi et al., 2017; Angold, 1989; Berg-Nielsen et al., 2003; Goodman, Ford, Richards, et al., 2000).

The sensitivity and specificity differ between studies with different samples (Brøndbo et al., 2011; Goodman, Ford, Simmons, et al., 2000; Hysing et al., 2007; Mathai et al., 2004). In the CAMHS North Study in Norway (Brøndbo et al., 2011), the emotional problem scale showed the lowest sensitivity. In this study the participants already had status as patients in CAMHS and SDQ emotional symptoms score was not used for diagnostic purposes, but to assess who would want to receive SMART treatment for emotional problems.

The Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983), Norwegian language version (Hanssen-Bauer et al., 2007), was used as a secondary outcome measure for general level of function and was scored pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up. The CGAS is a therapist-scored rating scale evaluation global functioning on a continuum ranging from 0 to 100, where higher scores indicate a higher level of function. CGAS has shown good psychometric properties (Schorre & Vandvik, 2004). In the present study each child's clinical profile was scored blindly by a group of at least 3 trained clinicians. The average of the clinician scores was employed, to ensure the stability of the CGAS scores. The clinicians were experienced CGAS raters, through routine clinical practice. In the present sample this was reflected through a high degree of reliability across CGAS raters (Intraclass Correlation= .97).

Clinical Outcome in Routine Evaluation-Outcome Measure (CORE-OM) (Evans et al., 2000), Norwegian language version (Skre et al., 2013), was used as a secondary outcome measure for general symptom pressure and risk of suicide and self-harm and was distributed at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up. CORE-OM is a questionnaire with 34 items employing a 5-point Likert scale from 0 to 4, with higher scores indicating an increased symptom pressure. Skre et al. (2013) suggested a cut-off point of 1 for discriminating between clinical and non-clinical populations in adult populations. It has been demonstrated that CORE can be read and understood by Norwegian users from 14 years old and upwards (Skre et al., 2013), and can thus be applied in this sample. CORE has shown good psychometric properties (Evans et al., 2000; Skre et al., 2013). A validation study on an adult Norwegian sample concluded that

CORE-OM has the same psychometric properties as the English version (Evans et al., 2000; Skre et al., 2013). However, confirmatory factor analyses in both adult Norwegian and British samples indicated an unintended "method factor" formed by positively formulated items across theoretically derived subdomains in the CORE-OM (Lyne et al., 2006; Skre et al., 2013).

Beck Depression Inventory, second edition (BDI-II) (Beck et al., 1996), Norwegian language version, was distributed at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up. In this study it was used as a secondary outcome measure for depression. BDI-II is a 21-item self-report questionnaire that can be used from ages 13 to 80 years. Items are rated on a 4-point Likert scale from 0 to 3, giving a maximum score of 63 (Beck et al., 1996). Suggested cut-off ranges are between 14 and 19 for mild depression, 20–28 for moderate depression and 29–63 for severe depression. BDI-II has shown good psychometric properties (Osman et al., 1997; Steer et al., 1998; Aasen, 2001). However, it has not to date been validated in a Norwegian adolescent sample.

Multidimensional Anxiety Scale for Children (MASC) (March et al., 1997), Norwegian language version. It was translated to Norwegian by Arnulf Kolstad at the NTNU: Norwegian University of Science and Technology in cooperation with the owner of the test Multi-Health Systems (Villabø & Neumer, 2017). There is no documentation of year or procedures for the translation. The candidate was licensed by Multi-Health Systems to use the Norwegian translation. MASC was used as a secondary outcome measure for the degree of anxiety, and was distributed at pre-wait-list, pre-treatment, post-treatment, and at six-month follow-up. The MASC is a 39 item self-report questionnaire measuring symptoms of anxiety, and was developed for children and adolescents between ages 8 and 17 years. Items form 6 subscales and are rated on a 4-point Likert scale from 0 to 3, with higher scores indicating a higher degree of anxiety. In the present study, we used the total score, converted to a t-distribution centered at approximately 50. MASC has shown good psychometric properties (March et al., 1997; Rynn et al., 2006).

Evaluation of sessions (Neumer & Junge-Hoffmeister, 2010a) was used for measuring treatment integrity, alliance and user satisfaction, and was distributed after each session and at the end of therapy. The SMART manual suggests an evaluation after each module, where the patient rates aspects of specific content and satisfaction with the session on a Likert scale from 1 "very unsatisfied" to 5 "excellent". However, in this study we

performed Evaluation of sessions after every session, as a feedback tool with the intention that the therapist adjusts therapy, alliance, content and relevance in collaboration with the patient.

3.9 Analysis

Statistical power and expected attendance rate were underlying the considerations for calculating the initial sample size of 160. As mentioned in paper 1, the minimum required sample size for the comparison of group averages in two groups (two-tailed t-test with a 0.01 significance level, statistical power 0.80, and an expected effect size on difference in mean scores between groups $d = 0.60$) was calculated to be 58–67 patients in each group (Machin et al., 2011). The estimated effect size was set as an average effect size of 0.67, estimated by the study summarizing existing research on CBT with children and adolescents by Weisz and colleagues (2006). The attendance rate was estimated to be 80%. As mentioned in paper 1 this was based on data from a counselling service for young people in the same geographic area that had a no-show rate of 12–17% (Wang et al., 2007). Consequently, we needed to recruit 160, to treat the necessary 126 patients.

In paper 1 missing data was handled by reporting the results as an intention-to-treat analysis (Hollis & Campbell, 1999) as suggested in the CONSORT 2010 statement (Schulz et al., 2010). In paper 1, when data was missing, outcome variables were imputed using multiple imputation (MI). When it is reasonable to assume that data are missing at random, MI has been suggested to be the recommended imputation technique (Del Re et al., 2013). Imputations of missing data were based on predictive mean matching using the MICE package for R (Van Buuren & Groothuis-Oudshoorn, 2010). Each imputation was chosen from a random draw among the 5 observations that were closest to the value predicted by the imputation model. Both outcome and demographic variables were used to predict (other) outcome variables. The linear mixed model's procedure analyzed each of the 50 imputed datasets separately, and the results were pooled using standard procedures.

In paper 1, linear mixed model analysis (Singer & Willett, 2003) was used to test for the effects of the SMART treatment in paper 1 and 2 in somewhat different ways. In paper 1,

the data were analyzed as hierarchical with measurement occasions (level 1; pre and post) nested within individuals (level 2). A random intercept was included in the model, but no random slope was included because measurement was only performed twice. A test of the significance of the time by group interaction is used as a test of whether the SMART group and the control group change in different ways from pre- to post-treatment. To adjust for the individual probability of being randomized directly into treatment (the SMART group), a separate analysis was done. This analysis however, showed no difference from the main analysis.

By using the pooled standard deviation of the pre-measurement for the standardization, effect sizes with 95% confidence intervals were computed as a standardized difference between the group's gain scores (Hedges' g) (Hunter & Schmidt, 2004). To test pre-treatment differences between the groups linear mixed models, one-way ANOVA, or chi-square tests was used, depending on the situation. The same methods were used to compare dropouts after pre-treatment with nondropouts on demographic variables and pre-treatment outcome variables. To asses clinical and significant change on the SDQ in both paper 1 and 3 the Reliable Change Index (RCI) was used (Jacobson & Truax, 1991) together with the clinical status on the SDQ emotional problems variable at post-treatment or follow-up (5 or less vs. 6 or more). The cut-off was based on a study by Rønning et al. (2004) with adolescents from the same region as the present study. In paper 1, IBM SPSS v24 were used for all analyses, and .05 was generally set as the significance level.

In paper 2 the analysis was essentially divided into five parts: 1. Factor analysis: Here we split the sample into two random halves. The first half was used to do an exploratory factor analysis (EFA), and the second half was used to test the model selected in the EFA in a confirmatory factor analysis. The model selection process was based on standard model fit criteria. 2. Reliability analysis: Internal consistency for the scales in the selected model was evaluated by computing McDonald's omega for ordinal items. We also investigated potential "weak" items within the scales by studying the omega if item deleted and the item to total correlations. 3. Measurement invariance: We tested measurement invariance for the gender variable using the non-clinical sample. Here a configural model and a scalar model was compared, and since full scalar invariance was not achieved, partial scalar invariance was also evaluated by highlighting a few items that might have caused the scalar non-invariance. 4. Gender differences: The difference between boys and girls was compared using a Z-test based on the latent means from the

final partial scalar model. 5. A clinical cut-off score was computed based on a formula from Jacobson and Truax (1991):

$$cut - off = \frac{M_c \cdot S_n + M_n \cdot S_c}{S_c \cdot S_n}$$

where M_c and M_n are the descriptive means for the clinical and non-clinical groups respectively, and S_c and S_n are the standard deviations in these two groups.

In paper 3 we also used linear mixed model analysis (Singer & Willett, 2003). This time to test the group by time and time effects; measurement occasion (level 1; pre, post and follow-up) is nested within individuals (level 2). In the analysis of paper 3, time was treated as a continuous variable, represented as the number of days that had passed since the baseline for the different measurement occasions. A random slope model was used. In order to get model-based predictions for each treatment condition on each measurement occasion, we computed estimated marginal means and standard errors in a linear mixed model analysis where time was treated as categorical. To compute the effect sizes at follow-up in paper 3, effect sizes for the time by group effect and the time effect were computed as the unstandardized coefficient (computed by the LMM analysis) divided by the pooled within-group standard deviation at baseline. Since the unstandardized coefficient was given in change difference per day or change per day, this ratio was multiplied by the average number of days from baseline to the follow-up measurement in the total sample, see Feingold (2013). Accordingly, follow-up effect sizes were calculated after 327 days had elapsed since baseline (minimum = 151, maximum = 632) (Lorentzen et al., 2022).

4 Summary of the papers

4.1 Summary paper 1

Lorentzen, Veronica; Fagermo, Kenneth; Handegård, Bjørn Helge; Skre, Ingunn; Neumer, Simon-Peter. A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14–17 years old in child and adolescent mental health services (CAMHS). BMC Psychology 2020; Volum 8 (1). ISSN 2050-7283.s 1 - 12.s doi: 10.1186/s40359-020-0393-x

Objectives

This study aims to investigate effectiveness of a 6-week, transdiagnostic cognitive behavioral therapy (CBT) for anxiety and depression in adolescents, the Structured Material for Therapy (SMART), in naturalistic settings of child and adolescent mental health outpatient services (CAMHS). The objective of Study 1 was to examine the effectiveness of the SMART in adolescents aged 14 to 17 who had been referred to CAMHS, and who reported clinically significant emotional symptoms. Adherence was measured to ensure that the therapy was delivered according to the SMART protocol.

The effectiveness was investigated by comparing pre to post changes in the treatment and the wait-list group on several domains: general clinical status measured with the CORE-OM, psychological functioning rated on the CGAS, and emotional symptoms defined by the SDQ, depression as measured by the BDI and anxiety as measured by the MASC

Results

The Adherence score indicated high levels of adherence to the manual with a high completion of the elements of the SMART model. Those elements receiving lower scores were group exercises in the original program. These were adjusted in different ways in the individual treatment provided in this study. The vast majority of patients were female, hence comparison between genders could not be performed. A third of the patients in the treatment group (CBT) improved their score on the main outcome measure (SDQ), compared to slightly more than a tenth in the WLC. According to criteria for clinical improvement or recovery, clinically significant and reliable change was experienced by

17.7% in the CBT condition, compared to 5.8% of patients in the WLC condition. No patients experienced deterioration. For patients in the treatment group, in addition to the main outcome measure of internalization symptoms (SDQ), statistically significant treatment effects were demonstrated on the measures of general functioning (CGAS) and anxiety symptoms (MASC). On measures of depressive symptoms, (BDI-II) and general clinical status (CORE-OM), no significant change was observed.

Conclusion

Results from this RCT of a manualized 6 session transdiagnostic treatment delivered in CAMHS are promising. They indicate that short-term CBT for youths with emotional problems may be more effective than no treatment. Further investigations may be needed, but SMART can be considered as an option of a first step in a stepped care model of treatment of anxiety and/or depression in CAMHS. The rates of recovery highlight that two-thirds of the patients may need additional treatment to support further improvement. Furthermore, the effectiveness of 6 session SMART to improve depressive symptoms was not supported, although there were significantly positive effects for anxiety, functioning and emotional problems. Further research is needed employing the full-scale SMART program to tailor the optimal additional interventions for these patients.

4.2 Summary paper 2

Lorentzen, Veronica; Handegård, Bjørn Helge; Lillevoll, Kjersti; Solem, Kenth; Moen, Connie Malén; Skre, Ingunn. CORE-OM as a routine outcome measure for adolescents with emotional disorders: factor structure and psychometric properties. BMC Psychology 2020;
Volume 8 (86). ISSN 2050-7283.s 1 - 14.s doi: 10.1186/s40359-020-00459-5

Objectives

There is a need for valid routine outcome measures that can be used in measurement-based care for adolescents.

To bridge this gap, this study examined the factor structure and psychometric properties of the CORE-OM in two samples of Norwegian adolescents (age 14–18): youths referred for treatment of emotional problems in CAMHS ($N = 140$) and high school students ($N = 531$). No previous study has evaluated the factor structure and psychometric properties of CORE-OM in an adolescent population. Since CORE-OM and the abridged version, CORE-10, are widely used as instruments for measurement-based care and session-to-session outcome tracking, also in clinics and counselling services receiving youths, an age-specific validation is needed.

The aims of this paper are to study the CORE-OM factor structure, reliability, gender invariance, comparing factor means for boys and girls, and presenting adolescent clinical cut-off scores.

Results

The best fitting model from the CFA could only partially confirm the theoretical model the CORE-OM was developed upon. The model included five factors, as opposed to four factors in the theoretical model, and the content of the factors differed to some extent from the original: 1) General problems that included most of the items from the Symptoms/Problems scale and some of the items from the original Functioning scale. The original Risk-scale was split in two: 2) Risk to self, and 3) Risk to others. A new factor emerged containing all items that were positively framed: 4) Positive resources, and finally items from the original Function scale relating to interpersonal problems loaded on: 5) Problems with others. The items originally belonging to the theoretically constructed Well-being scale were distributed on the scale's General problems, Positive resources and Functioning in the confirmatory factor analysis. The clinical cut-off score based on the calculated total score of all items was higher than cut-off scores reported from adult samples. Gender differences were found, both for the all-item total and for the general problems cut-off score. The measurement invariance analysis for gender showed that gender comparisons should not be performed without modification of the scale.

Conclusion

These factor analyses of the CORE-OM in adolescent samples yielded a five-factor solution, proposing new subscales concerning positive resources and problems with others. A new 17 item general problems/symptoms scale, where positively framed items were removed, is believed to be a more valid measure of psychological distress than the original scale that included positively framed items. Developers of self-report instruments should be reluctant to reverse items, if the intention is not to study reversed items separately. The general emotional problem and the positive resources scale should be modified if means for gender should be compared. However, the low number of boys in the clinical sample may constrain the generalizability of gender specific results from this study. The higher cut-off scores found in these adolescent samples, as compared to adult samples, could reflect that adolescents generally have higher levels of distress or lower threshold to report problems.

4.3 Summary paper 3

Lorentzen, Veronica; Fagermo, Kenneth; Handegård, Bjørn Helge; Neumer, Simon-Peter; Skre, Ingunn. Long-term effectiveness and trajectories of change after treatment with SMART, a transdiagnostic CBT for adolescents with emotional problems. BMC Psychology 2022; Volume 10 (167). ISSN 2050-7283.s 1 - 18.s doi: 10.1186/s40359-022-00872

Objectives

This study followed the separate symptom trajectories for youths with diagnoses of anxiety, depression, and comorbid anxiety and depression through the course of a transdiagnostic CBT intervention, and at 6-month follow-up.

Trajectories of symptom (SDQ emotional, BDI-II, MASC and CORE 17 items) and functioning (CGAS) change from baseline to follow-up were examined. Furthermore, treatment by time (pre-therapy, post-therapy and 6-month follow-up) interactions, effects of time (WLC vs CBT), and diagnostic group (anxiety vs depression vs combined anxiety and depression vs no diagnosis) by time interactions were evaluated.

Results

The results showed a highly significant change in the sample for all outcome variables. Effect sizes on the time variable were largest for general functioning (CGAS), with a predicted change after 327 days corresponding to 2.19 standard deviations (SD) increase on the CGAS. Furthermore, results showed a decrease equal to 2.10 SD change in emotional problems measured on the SDQ. Both depressive symptoms (BDI-II), anxiety symptoms (MASC), and general symptom load (CORE-17 items) decreased corresponding to a change of around 1 SD. The results showed no significant time by diagnostic group interactions for any of the outcomes, indicating similar patterns of change in the separate diagnostic groups. The intervention group by time interactions from baseline to follow-up indicated that waiting 6 weeks for treatment had no impact on the effect of the treatment. However, symptom score before treatment predicted symptom score at follow-up.

Conclusion

The main finding from this study following adolescent patients with emotional problems from pre-treatment, through 6 weeks transdiagnostic CBT, to 6-month follow-up was that six weeks of transdiagnostic treatment with the SMART program for emotional problems showed promising effects at 6-month follow-up for some of the patients. The patients showed a large reduction in general emotional symptoms as measured by the SDQ emotional scale, and growth in daily functioning as measured by the CGAS. Furthermore, this transdiagnostic treatment for anxiety and depression had equal effects on depressive and anxiety symptoms at follow-up. The fact that there were no significant differences in change trajectories between groups may indicate that the treatment seems to target all groups well. However, since symptom score before treatment predicted symptom score at follow-up, we advise that the treatment of the comorbid group with both anxiety and depression, which also had the highest symptom score before treatment, should be given special consideration. Furthermore, the effect sizes in a within-subjects design should be interpreted with caution. Further investigations should also be done to address the needs of the patients who did not show improvements. The high attrition rate from post to follow-up must be taken into account when interpreting the results.

5 General discussion

The main aim of this thesis was to provide new knowledge that can contribute to the evidence base of mental health services offered to adolescents. In the three studies included in this thesis, the short- and long-term effectiveness and trajectories of change of a brief CBT for adolescents with emotional problems have been investigated. Furthermore, the validity of an outcome measure for use in youth populations was investigated.

In paper 1 the post-effects of a 6-week SMART treatment showed statistically significant treatment effects for anxiety symptoms and general functioning compared to wait-list controls at post-treatment. Clinically significant change in emotional problems (SDQ) was observed significantly more frequently in the treatment condition at post-treatment. In paper 3 the main findings at 6-month follow-up were, firstly, that six weeks of transdiagnostic treatment for emotional problems yielded recovery in nearly half the patients; and clinically significant or reliable change in nearly one third, according to the main inclusion criterium SDQ emotional symptoms scale. Secondly, almost half the patients had normal functioning according to the outcome measure CGAS at follow-up. More specifically, the within effect sizes of change from baseline to follow-up in emotional symptoms, functioning, general psychological distress, depressive and anxiety symptoms were all relatively large. According to the results in paper 1 the patients reported satisfaction with the overall treatment and with the alliance to the therapist. However, the attrition was high at follow-up and further treatment for the patients who didn't have effect should be given attention both clinically and in further research.

Magnitude of effects: After six weeks of SMART treatment, the magnitude of the post-treatment effect sizes was moderate to small on the measures of emotional problems, general functioning and anxiety symptoms. The depressive symptoms and general clinical outcome improved, but there was no statistically significant difference between the treatment and wait-list conditions on these measures. The rates of recovery were somewhat lower than what is expected for CBT efficacy trials for single-disorder treatments (Weisz et al., 2017), where the majority of studies show moderate to large effects targeting anxiety and depression (James et al., 2020; Kendall & Peterman, 2015).

However, it must be taken into consideration that these studies were efficacy studies, and in our effectiveness study, we targeted both anxiety and depression with a short treatment. Further, the SDQ emotional scale and MASC showed comparable effects to those found by Weisz and colleagues (2017), and lay slightly above the effects shown by Weisz and colleagues (2013) of therapies delivered in the setting of ordinary clinical care. The present results are promising compared to effect sizes in effectiveness studies performed in ordinary clinical practice (Weisz, Kuppens, et al., 2013). Compared to a report showing treatment as usual after six months of treatment in one of the participating CAMHS (CORC, 2018), our study showed comparable effects after only six sessions of treatment according to the rate of change on the SDQ and CGAS.

In the 6-month follow-up, the improvement seems to have continued from post-treatment to follow-up. The magnitude of the observed changes indicates a significant drop of emotional symptoms, both anxiety and depression, and psychological distress, on all measures, and a heightening of psychological functioning. The analysis from baseline to follow-up showed a highly significant change in the overall sample for all outcome variables (emotional symptoms, functioning, general psychological distress, depressive and anxiety symptoms), and the within effect sizes were all well above what Cohen characterized as large effects. There are few comparable effectiveness studies with similar participants, clinical settings, treatment duration, and transdiagnostic treatment, but the changes shown in our study are in line with those found in a meta-analysis of durability of effects of treatments for emotional disorders after 1-year follow-up (Rith-Najarian et al., 2019). These findings of so- called “sleeper effects” with growing ES from post to follow-up are also consistent with the findings of long-term outcomes for youth CBTs targeting anxiety on the CGAS and MASC in the Coping Cat study conducted in CAMHS in Norway consisting of a 12-session treatment solely focusing on anxiety treatment (Villabø et al., 2018). It is important to note that the ES from study 1 was calculated between the treatment and the wait-list group, while the ES in paper 3 was calculated as a within group ES, and thus comparing these ES should be done with caution. The results for the present study are also comparable to those found at 6-month follow-up after CBT for depression, as measured with the BDI presented by (Stice et al., 2010). In addition, the effects are comparable to similar studies showing moderate to large effects on CBT targeting anxiety, depression and trauma (Gutermann et al., 2017; James et al., 2020; Kendall & Peterman, 2015).

Treatment length and additional treatment: Given that the relapse rate could be as high as one third in youths (Ginsburg et al., 2014; Kennard et al., 2009), SMART shows promising results also at 6-month follow-up despite the compressed and time-limited character of the SMART intervention. When it comes to treatment length it is difficult to find comparable data on average contact time and treatment response. However, naturalistic outcome data from 1641 adolescents in 60 mental health services in England reported an average contact length of 31 weeks (Krause et al., 2022). When it comes to the duration of a depressive episode, the numbers vary. The TADS study shows variations in the duration of the current major depressive episode with range from 3 to 572 weeks (0–11-year-olds), with a median duration of 40 weeks (Treatment for Adolescents with Depression Study, 2005). The mean duration for a depressive episode has been estimated to be between 4-9 months among clinically referred adolescents (Birmaher et al., 2002; Emslie et al., 2003). In our study it makes clinical sense that the patients who had higher scores also had more treatment. In SMART the Pearson correlations between the symptom scores and number of sessions between post and follow-up were moderate to high (and significant). This indicates that participants with more serious emotional problems were given more treatment and that the routine care was able to tailor the intervention to the patients' needs despite the fact that treatment provided through the study had finished.

Clinical and reliable change: In paper 1 the post-treatment effect, the RCI (Jacobson & Truax, 1991), was used as a measure of clinically and reliable change. The RCI was used together with status over and below cut-off, as a parameter to indicate whether the change could be characterized as reliable. On the SDQ emotional scale "reliable change" corresponds to 4 points improvement, "clinical change" corresponds to a change from above to below cut-off and "clinical and reliable change" includes both the categories mentioned above," detoriation" corresponds to four-point increase on the SDQ emotional scale. In our sample at post-treatment, nearly one-sixth of the participants showed clinical and reliable change while nearly two-thirds showed neither clinical nor reliable change. None of the patients deteriorated. Studies from ordinary care report poorer outcome following treatment (Manteuffel et al., 2008; Warren et al., 2009). Poorer outcome following treatment was also shown in the CORC (2018) report mentioned earlier, where 27% of the CAMHS patients showed deterioration. Further, in our sample, general functioning measured by the CGAS, showed that the adolescents in the treatment condition improved nearly one category, indicating clinically meaningful change after a

brief intervention measured post-treatment. The results at 6-month follow-up showed recovery in nearly half the patients measured at follow-up, and clinically significant or reliable change in one third, according to the main inclusion criterium SDQ emotional symptoms. It is discussed that the RCI represents a very conservative measure of change, with indications that the majority of the children and adolescents in CAMHS do not experience clinical improvement when this criterion is being applied (Garland et al., 2013; Warren et al., 2010). In the meta-analysis mentioned earlier (Bear et al., 2020) with 11739 young people in CAMHS receiving treatment as usual, 38% showed reliable improvement, 44% no reliable change and 6% showed reliable deterioration. In accordance with the findings in our study and the overall results of studies on clinical improvements in CAMHS (Bear et al., 2020; CORC, 2018; Sørheim Nilsen et al., 2015; Sørheim Nilsen et al., 2016), further attention from the research field to patients not improving should be encouraged.

Paper 3 reports that 18 of 58 individuals showed clinical and reliable change with SDQ data at this measurement occasion. However, this is not based on an ITT analysis. A more conservative estimate of the percentage of cases with a clinical and reliable change would result from imputing a "no" status on both the clinical change and the reliable change variable for the 87 individuals without follow-up SDQ data. Assuming that an ITT-analysis can be carried out on the 145 that either got the SMART treatment or the wait-list condition, an ITT analysis table will look like this:

		Reliable change	
		No	Yes
No clinically significant change	n	113	4
	% total	77.9%	2.8%
Clinically significant change	n	10	18
	% total	6.9%	12.4%

12.4% constitutes a minimal value of the percentage that shows both clinical and reliable change.

The analysis performed on every measure and point of measurement is based on available data. Although 18 of the 58 patients we had data for on the SDQ showed clinical and reliable change, the attrition was high. A minimal value of percentage of both clinical and reliable change would then be 12.4% assuming no change in the participants with missing data.

Trajectories of change: The change trajectories for the young patients receiving this transdiagnostic treatment, were similar for pure anxiety, pure depression, anxiety and depression combined, and patients with emotional problems without a specific diagnosis.

Concerning change from baseline to follow-up for various diagnostic groups (no anxiety or depression diagnosis; anxiety; depression; anxiety and depression), the results showed that the groups shared similar patterns of change with large, significant effects in reductions of emotional symptoms and a significant increase in daily functioning in all outcome variables indicating similar patterns of change in the four diagnostic groups, with similar reductions in every group. The main finding is that the groups show decreasing symptoms over time, and increasing functioning from baseline to follow-up. Given these results it seems that transdiagnostic treatment delivered and targeted to anxiety and depression gives similar rates of change across disorders. However, symptom level at pre-treatment in the groups predicted symptom level in the groups at follow-up, with the result that the comorbid group with highest level initially, also had the highest symptom level after treatment. Consequently, the further treatment needs for patients with comorbid anxiety and depression should be explored more extensively.

Waiting time: There were no significant differences for any of the outcome variables in longitudinal trajectories from baseline to follow-up for the two treatment conditions. Large efforts have been made in reducing waiting time because generally waiting lists have been regarded as negative (Cayirli & Veral, 2003). Research on health services for physical health reports that waiting time is connected to poorer functioning both socially and physically, as well as lower quality of life and poorer health status (Oudhoff et al., 2004; Oudhoff et al., 2007; Sampalis et al., 2001). In our study the patients had read about the treatment. They knew that they were going to receive treatment and also what kind of treatment. They also knew when they were going to receive the treatment. The waiting time only lasted for 6 weeks. The patients with more serious problems were excluded due to governmental restrictions on waiting. It is not possible to generalize these results, but it

would be interesting to investigate further what happened in the WLC while they waited for treatment. We have no data to indicate whether a longer waiting time would have a negative effect on long term outcome. It must be taken into account that the attrition at follow- up in our study was relatively high, and as discussed in paper 3, the attrition can be caused by reasons we have no information of, so the analysis should be interpreted with caution. The analysis performed is based on the assumption that data is missing at random (MAR), but if data is not missing at random (NMAR) there is a possibility that the mechanisms that led to attrition could have been different in the SMART- group compared to the wait-list group. E.g., that the patients in the wait-list group potentially could have had another symptom development than the patients in the SMART- group.

Providing a valid measure: To measure, screen, monitor and as feedback tools for adolescents and therapists, we need instruments that are valid and have good psychometric properties. In the validation of CORE-OM in paper 2 we found several results that could make this instrument valid when used with adolescent patients. In sum we found a new factor solution and a higher cut-off score than reported in adult samples. The EFA resulted in a five-factor solution, and the factor contents were interpreted as general problems, positive resources, risk to self, risk to others, and problems with others. The CFA model fit for this model was good. The measurement invariance analysis for gender showed that gender comparisons should not be performed without modification of the scale. The clinical cut-off score based on the all-item total was higher than in an adult sample.

The reason for including negative items in scales is to reduce the acquiescence bias (Baumgartner & Steenkamp, 2001). This happens when people are not attentive to the actual content of statements, but just agrees as an automatic response pattern, without regarding the content. Other reasons could be laziness or indifference (Podsakoff et al., 2003). One way to prevent this from happening is to reduce response speed, to try to promote cognitive reasoning by reversing items to minimize this bias, hence mixing positive and negative items (Podsakoff et al., 2003). In this way the intention is to further increase the validity of measurements by promoting cognitive efforts to understand the construct in the instrument. (Weijters & Baumgartner, 2012). However, some studies have shown that this combining of positive and negative items does not reduce acquiescence bias (Sauro & Lewis, 2011).

One of the main findings in paper 2 was that a different model than the one described by the developers of CORE-OM emerged. The initial developer describes the instrument as a

four-dimensional measure with dimensions of Subjective well-being, Problems, Functioning and Risk (Evans et al., 2000). We found improved model fit for a five-factor model over factor solutions with less factors; General problems, Positive resources, Risk to self, Risk to others and Problems with others. So, what could explain why data from the present youth sample yielded a different factor structure? In a study on an adult sample Lyne et al. (2006) showed a method factor related to positive and negative wording of the questionnaire items. This is consistent with the findings in our sample where all the eight positively keyed items loaded on the same factor in the five-factor EFA. Adolescents that have both positive and negative resources, will relate positively keyed questions to their positive resources, likewise they will relate negatively keyed questions to their negative resources. Because of this, positively keyed items will have higher correlations to other positively keyed items. This will probably happen even though the intention was to reverse the items on a problem scale. According to the tripartite theory of anxiety and depression (Clark & Watson, 1991), negative affect and lack of positive affect may represent separate dimensions of internalizing problems. The current factor solution supports that negative affect and lack of positive affect are different phenomena. These findings highlight the problematic status of assuming that a low score on a positively keyed item reflects the same as a high score on a negatively keyed item, and challenges the validity of the Symptoms/Problems scale in the original CORE-OM measure. Generally, there was high reliability for all scales except Risk to others. Since it contains only two items, it should be expanded if it is to be included in the scale, or else this scale may be subject to excessive measurement errors.

As discussed in paper 2, the main difference between the 17-item general problems scale from the present study and the 28-item non-risk scale is the exclusion of the positively keyed items from the 17-item version. As discussed earlier, one of the reasons for including both positively and negatively keyed items in a questionnaire is to reduce acquiescence bias (response style bias, respondents tending to agree with statements) (Suárez-Álvarez et al., 2018). However, positively and negatively keyed items may involve different cognitive processes (Marsh, 1986; Marsh, 1996), and this is one of the reasons that a positive item latent variable showed up in the EFA. It is a paradox that including some positively keyed items in a questionnaire consisting mostly of negatively keyed items, in order to mitigate acquiescence bias, seems to confuse the responders and therefore makes the instrument less valid and scales less reliable. This finding is in line with other studies showing problems with the mixing

positive and negative items resulting in inconsistently responses (Colosi, 2005; Garg, 1996). It is hypothesized that this happens because it is difficult to cognitively establish the difference between positive and negative items (Sauro & Lewis, 2011; Van Sonderen et al., 2013). Potential method effects associated with negatively worded items has been found in several scales (Ebesutani et al., 2012; Tomas et al., 2013; Van Dam et al., 2012; Ye & Wallace, 2014) with negative items that tend to be intercorrelated. As a consequence it has been recommended to use positive items only (Schriesheim & Eisenbach, 1995). However, this problem with reversed items does not apply to all cultures (Wong et al., 2003).

Gender differences: From puberty onwards, there is a higher and increasing frequency of depression and anxiety disorders among girls compared to boys (FHI, 2016). Results from recent studies in gender differences in adolescent depression displaying increasing differences from the age of 13 and upwards are similar to results from studies from 1980 and 1990 (Nolen-Hoeksema & Girgus, 1994; Salk et al., 2016). Trajectories show that girls accelerate early, while boys accelerate late in adolescent. Overall, in adolescence there is a higher frequency of girls with depression (Salk et al., 2016). The period of adolescence contains multiple changes on many levels. As young people enter puberty, they experience new stressors concerning education, sexuality and peer conflict (Berg et al., 2017; Schaffhuser et al., 2017). Although the prevalence varies with population, measures etc., there is a consistent finding that girls are more prone to emotional disorders, whereas boys are more prone to experience externalizing symptoms (Campbell et al., 2021; Green et al., 2005; Van Droogenbroeck et al., 2018). A promising theory to account for these differences is the "Gender intensification Hypothesis" which poses that gender differences in depression is formed with the onset of puberty as a "Kick off" for normative changes between gender roles. (Conley & Rudolph, 2009) Early onset of puberty and early maturation in girls is a risk factor for depression (Hamlat et al., 2019; Yoon et al., 2022). In paper 2 the measurements invariance for gender indicate that one cannot compare means for boys and girls without modifying the scales. Changes in gender roles in adolescence can result in a better fit for girls than boys in questions reflecting emotional problems. One example of an item showing that it is problematic to compare means for gender is item 14 in CORE-OM ("I have felt like crying"). Our analysis showed that boys and girls report very differently on such issues. There is a genuine risk that you do not measure symptoms of emotional disorders, but differences in gender expression of

problems e.g., crying. It is therefore problematic to include such an item in a scale where similar cut off for boys and girls is applied.

The results from paper 2 concerning gender differences in the non-clinical sample did not only show gender differences on item level and cut off, but also different reporting on the latent variables between the genders. Boys and girls differed on four of the five latent variables. The exception is the risk to self-variable where there was a non-significant gender difference. In the non-clinical group few adolescents had thoughts of self-harm. On the general factor the girls scored higher than boys, which is consistent with other findings (Evans et al., 2002; Palmieri et al., 2009). On the risk to others factor boys scored significantly higher. This is also consistent with other validations (Skre et al., 2013; Uji et al., 2012). Girls showed significantly lower scores than boys on the positive resources' latent variable, while the girls showed higher scores on the problems with others factor. The items in this scale involve feelings of being criticized, humiliated, made shameful or having no friends, as such it relates to emotional relations with others. In line with the gender intensification hypothesis, transition from childhood to adolescence, with girls reaching puberty earlier, peer relations become more significant in search for aspects concerning identity, such as attitudes, activities and well-being (Brown & Bakken, 2011). In these aspects girls often display higher levels of distress in peers, more fear of rejection and higher empathy than boys (Rose & Rudolph, 2006) causing significant emotional symptoms (Schaffhuser et al., 2017). Girls often use emotional coping skills, while boys tend to devaluate such emotional expressions (Copeland & Hess, 1995) or seek distraction in physical activities (Plenty et al., 2014). This way of reporting could also have cultural explanations. In the Japanese version of CORE-OM the female participants showed lower scores on "close relationships" subscales (Uji et al., 2012).

However, the low number of boys in the clinical sample challenges the generalizability of the gender specific findings in the present study.

Clinical cut-off score: Several studies show that when confronted with the same stressors, girls perceive more difficulties and report symptoms more frequently than boys (Frydenberg & Lewis, 1997; Jose & Ratcliffe, 2004). In paper 2 we concluded with recommending higher cut-off scores in adolescents than adults. This finding needs to be replicated, but it corresponds well with the finding that youths also score higher than adults on the BDI (Albert & Beck, 1975; Teri, 1982). Based on our study in paper 2 of

CORE-OM, the 17-item factor is a recommended as a measure of general problems and psychological distress, as it is a more reliable measure when the positively keyed items do not interfere with this factor. The reason for this is that the reversed items included in All items minus Risk 28-item score makes up a method factor that underestimate the level of emotional distress experienced by the users. Gender differences is seen both in All-items in the original version and the 17-items general distress scale, with girls scoring higher than boys, and higher than in adult samples (Skre et al., 2013). It must be further investigated how to take these differences into account as different solutions will give different consequences. The prevalence for emotional disorders in adolescent girls is higher than in boys (Salk et al., 2016). If the solution is to set a higher cut-off score for girls for example at the 90-percentile, we could lose girls who are in need for treatment, further if the prevalence of anxiety and depression in boys is close to 5%, we could overestimate the number of boys in need of psychological treatment. As mentioned above, there is a good chance that boys underreport or report different than girls on different items measuring emotional problems. But setting a lower cut-off to accommodate the boys lower scoring or use gender specific scores as suggested by (Connell et al., 2007) is not obviously the best solution .

The gender differences discussed in the above section can also be relevant for the inclusion with the SDQ emotional symptom scale. Concerning representativity in our study, there were too few boys ($n = 14$) in the clinical samples to perform analyses of gender differences. Maybe the inclusion criteria in our study could have altered how many boys who participated. Studies show evidence of partial gender non-equivalence, with a tendency for girls to more often endorse items measuring symptoms of emotional problems and prosocial behavior on the SDQ (Bøe et al., 2016). This could have implications for the SDQ as a screening instrument and could have affected the recruitment of boys in the study, indicating that perhaps a lower cut-off score could have been employed for inclusion of boys. The accuracy measures of a screening test may vary due to the prevalence of a disorder, which informant who gives information and the sample studied. Ideally, and maybe in the future, data could be merged to consist of large enough samples to divide them into subgroups making it possible to compare different clinical samples according to problems, age gender and so on. Related to this study, it would be interesting to investigate the psychometric properties and gender differences in self-reported SDQ scores on the emotional problem scale with adolescent 14-18 in a clinical population referred with emotional problems. The emotional scale also

consists of few items so if using the SDQ. An alternative could be to use a combination of the emotional problem scale and the impact scale. In a study by Goodman (1999) Impact scores were better than symptom scores at discriminating between the community and clinic samples; discrimination based on the single “Is there a problem?” item was almost as good. The SDQ burden rating correlated well with a standardized interview rating of burden. The impact supplement of the extended SDQ appears to provide useful additional information without taking up much more of respondents' time. Emotional problems had the strongest association with perceived difficulties and impact caseness. (Goodman, 1999). The sensitivity and specificity differ between studies with different samples (Goodman, Ford, Simmons, et al., 2000; Hysing et al., 2007; Mathai et al., 2004). In the CAMHS North Study in Norway the emotional problem scale showed the lowest sensitivity (Brøndbo et al., 2011). According to this, maybe a combination of emotional problem scale and impact scale as primary outcome and study entry would have yielded different results in who participated and give another picture of the results obtained. Concerning the aspect of false positive with having 6 as a cut off, without impact scale. It is of relevance that this was a clinical sample where the participants already had status as patients with already established treatment needs, and that they were asked to participate in the study. They could freely choose another treatment alternative in CAMHS. Hence, the consequence of false positive inclusion would be rather different than in a community sample screening for people who have treatment needs. SDQ emotional problem scale was not used for diagnostic purposes or accuracy, but more as a warning signal for treatment of emotional problems (Goodman, 1997, 2001).

A dilemma in deciding a clinical diagnosis is what should constitute the gold standard? According to Costello, Egger, and Angold (2005), clinical structured interviews are the closest we come to deciding a clinical diagnosis. On the other hand, there is poor agreement between structured interviews and diagnosis assigned by clinicians. Knowledge concerning the most valid methods is lacking (Lewczyk et al., 2003). It is suggested that the assignment of clinical experts aided by a structured interview such as the DAWBA may be considered the best available reference for comparison (Brøndbo et al., 2011). In our data material there was a difference between the assigned diagnosis by the clinician and the DAWBA bands. In addition, DAWBA clinical expert diagnosis was not available at that time. This was, together with the reasoning mentioned in the introduction, reasons to use the SDQ emotional problem scale as a primary measure and using DAWBA bands in the manner it is used in the study. This being said, this is a transdiagnostic treatment for emotional problems, not a diagnosis-

specific treatment, resting on carefully diagnosed patients. Although some of our participants did not receive a probable diagnose, this study was performed in CAMHS, where there is no prerequisite of diagnosis before receiving treatment. Assessing treatment needs was performed according to the service needs and regulations in CAMHS, which are not predicted solely by diagnosed disorders, but also include people who exhibit difficulties and need help for psychosocial impairment without having an assigned diagnoses (Hesedirektoratet, 2015).

5.1 Methodological considerations: Strengths and limitations:

5.1.1 Strengths

When this study was performed it was, to the best of our knowledge, the first RCT testing the effect of a 6-session transdiagnostic CBT treatment in an adolescent sample suffering mainly from combined emotional disorders, receiving the intervention in a CAMHS naturalistic setting. This thesis gives new and relevant knowledge concerning treatment of emotional problems in CAMHS in several ways: Firstly, this is one of few studies focusing on adolescents. Secondly, the study was performed in routine clinical settings with routinely referred patients. Thirdly, therapists represent the clinical personnel providing therapeutic interventions in Norwegian CAMHS. Fourthly, the studies provide information about clinical status and functioning at 6-month follow-up, and trajectories of change for anxiety, depression and comorbid anxiety and depression. And finally, to our knowledge, the validation of the CORE-OM was the first validation performed in adolescent samples. In this way the thesis contributes to insights in both short-term CBT for adolescents, to more lasting effects of transdiagnostic treatment, and to the importance of validating outcome tools in age specific samples, and in representative samples of high-school students. The study results are therefore expected to have external validity. Furthermore, the sample size is fairly large for a RCT in CAMHS. Effects are tested with a wide selection of outcome measures.

5.1.2 Limitations

The present study was performed with regularly referred patients in CAMHS who were entitled to receive treatment under time constraints. The reason for not including the 18

persons that did not receive the intervention was that they either withdrew their consent to use their data, or that they did not belong to the inclusion criteria of the study. This is described in detail earlier in the thesis. These exclusions should ideally have been performed prior to randomization. Maybe clinical interviews in person would have yielded more precise assessments prior to inclusion than computerized interviews of SDQ and DAWBA. With these 18 individuals included in the sample we could have investigated whether the more conservative calculations from a sample ITT would give other results and other conclusions than we have reached in paper 1 and 3. However, for the reasons mentioned above this data is not included.

We made considerable efforts to ensure equal treatment duration in the intervention group and the wait-list, by eliminating the assertiveness module. It is easier to control waiting time than it is to make sure that patients attend to weekly sessions. Patients get sick, have school tests, vacations and other reasons for absence. Hence, a possible limitation of the study is that the mean duration for the treatment condition was 3.5 weeks longer than for the WLC, which could influence the results in the intervention condition either through maturation, that the patients have longer time to get better, or through history, that external events could affect the treatment. In planning of a new study, we have learned that 6 weeks will probably in practice require 9 weeks to administer. However, in CAMHS the patients usually are not able to wait so long according to the governmental restrictions on waiting time. Other possible limitation regarding representativeness was that the governmental restrictions on waiting time prevented us from enrolling part of the referred patients in the study. Furthermore, the wait-list control group was after enrollment ensured that they would commence treatment in 6 weeks. The expectation of a nearby intervention might have led to relief and symptom reduction in the wait-list group, making it more challenging for the clinical effect of the SMART intervention to outweigh the expectancy effect in the wait-list group. As mentioned earlier and in paper 1, a report from the Child Outcomes Research Consortium (CORC, 2018) reported outcomes from patients receiving treatment in one of our participating CAMHS from 2013 to 2016. These scores indicate that the present sample in paper 1 and 3 also included acute and serious disorders, and thus could be considered representative of moderate to severe referred patients. However, the clinical sample may not be representative of the entire CAMHS population of patients with emotional disorders, since patients evaluated as acutely suicidal were excluded from the sample because they could not wait 6-weeks before commencing treatment. Patients

who are acutely suicidal need more extensive care and monitoring in ordinary CAMHS in this phase than what is offered in the SMART program. Thus, excluding these patients from the study was both a correct clinical and ethical decision, not only in this study, but also in regular care. However, the same argument is not completely valid considering the clinical sample in paper 2. The CORE-OM was mainly developed to monitor problems and risk in patients receiving outpatient treatment and counselling. Although the present clinical sample probably has a high density of the phenomena that the CORE-OM was designed to monitor, perhaps problems concerning risk to self and others was underrepresented, since the high-risk cases were excluded.

In the validation of CORE-OM, the mean and standard deviation in the male clinical sample used in the Jacobson and Truax formula have large standard errors due to the low rate of males in the clinical sample, hence, the clinical cut-off scores for boys may be uncertain. Hence, generalization of our results to both genders should be done with caution.

Adolescents that were 18-year-old were included in the non-clinical sample, but not in the clinical sample. The reason for this was that the age span in the clinical sample was 14–17, since in Norway patients from age 18 are referred to mental health services for the adult population. In the non-clinical sample, the age span was 14–18, since in Norwegian high school, enrolment in different grades is based on the year of birth, hence the inclusion of 18 years old in the non-clinical sample. However, the mean age in the two samples was similar, and therefore the inclusion of 18-year-olds in one of the validation samples and not in the other may have had only marginal influence on the results.

Another limitation is the use of a wait-list as a control condition, as opposed to an active control condition. According to the protocol the two experimental conditions should have equivalent length. The whole length of the original SMART program was 8 weeks, but to accommodate the ethical and legal issues concerning the maximum waiting time of 6 weeks, we had to shorten the treatment. The assertiveness module in SMART was removed to create equal length in both conditions. This was warranted, but does not reflect the comparative effect of full-scale SMART treatment. On the other hand, the deletion of one module of the SMART program is an intended modification that this intervention gives to the practitioner. Furthermore, SMART is both transdiagnostic and modularized, and in this study, SMART was delivered in a linear order to all patients, despite some

research showing that delivering in a modular, flexible fashion gives better results (Weisz, Kuppens, et al., 2013). Although SMART shows promising results, the effectiveness documented in the present studies should be qualified as preliminary, pending future trials of the full-scale program delivered in a modularized fashion in order to assess "active ingredients". The treatment should also in future research be compared to an active control group.

Another limitation of the study is the high attrition rate at follow-up – only 83 out of 145 patients completed the assessment at follow-up, corresponding to a response rate of 57%, hence the attrition rate being 43%. The analysis applied information from all 145 subjects when estimating effects. The linear mixed model analysis require that missing data are missing at random (MAR). However, since 42 of the 62 missing cases at follow-up were listed as missing due to a variety of administrative challenges, we can state that at least the majority of these cases were not missing because of any systematic features of the missing patients. We show that 83 patients contributed to data at follow-up, but for some variables the number of participants contributing with data were lower depending on what measure were analyzed (for example we had only data for 58 patients for the SDQ emotional scale and 64 on CGAS at follow-up). Consequences of attrition could lead to biases in the estimates, resulting in wrong conclusions. If for example patients do not answer information on certain measures because they have too high scores on anxiety and depression measures, this could contribute to wrong estimates and conclusions. In paper 3 we have concluded that waiting six weeks for treatment did not give different effects than going straight to treatment, and that different diagnostic groups show about the same rate of change from baseline to follow-up. With less attrition we could be more confident of these conclusions. For SDQ emotional we had less data at follow-up and at the same time we saw that CGAS had a higher effect. We cannot rule out that the results could have been different with lower attrition, but there is a high possibility that most attrition was of administrative sort and not linked systematically to the adolescents, hence MAR and not NMAR.

In paper 3 the primary outcome measure was the SDQ emotional symptoms scale. Since the patients were screened and included into the study based on their above cut-off score on the SDQ emotional scale, all had scores in the clinical range at inclusion. Consequently, one would expect that change at follow-up partly could be explained by regression to the mean. Similar arguments could be used regarding the change in daily

functioning score as measured with the CGAS. However, the magnitude of the effect sizes for both SDQ emotional symptoms and CGAS were too large to solely be explained by regression to the mean. On the other hand, the distributions at inclusion on the secondary measures general psychological distress (CORE-17), depression (BDI-II) and anxiety (MASC) were more heterogeneous, and with possibilities for so-called flooring effects for those scoring in the lower range at intake. Nevertheless, effect sizes at follow-up were fairly large, also for all the secondary outcome measures. Since the study was not designed for comparing the development in different diagnostic groups, the sample has low power for analyzing time by diagnostic group interactions, and these results should be interpreted with caution.

5.2 Implications

If a six session transdiagnostic treatment can be acceptable, and have lasting impact, it is a scalable and likely cost-effective treatment to be considered as the first step in a stepped care model in CAMHS for youths with emotional disorders. The high attrition must be taken into account when interpreting the results. The results also illuminate the need for further treatment for some of the patients. These patients should have both research and clinical attention.

In Norway there are no national treatment guidelines for child and adolescent anxiety and/or depression. The patients' pathways give no recommendations for treatment of child and adolescent anxiety and/or depression (Helsedirektoratet, 2018). Even though several methods have shown effects both in CAMHS and in primary care, there is no clear guideline for where and how to treat anxiety and/or depression. For example, The National Institute for Health and Care Excellence (NICE) published guidelines on depression in children and young people in June 2019 (NICE, 2019). The guideline covers the identification and management of depression in young people aged 5–18 years. Based on the stepped-care model, it aims to improve recognition and assessment and promote effective treatments for mild and moderate-to-severe depression. Clearer guidelines and guidelines for treatment of anxiety and/or depression should also be considered in

Norway. The stepped care model should be considered used more systematic than it is today.

Routine measurement along with frequent feedback from the users of the service provided has shown to improve outcomes in CAMHS (Bickman et al., 2011; Lambert & Shimokawa, 2011). Integrating outcome data in decision making and everyday practice can improve quality of treatment (Edbrooke-Childs et al., 2015). Improving the quality in mental health services require measurements to be administered frequently and that it is made relevant in feedback to the users (Wolpert, 2014). The feedback in this study has been made useful to the therapists and patients through active use both in supervision and treatment. SMART has many assumed effective components based on CBT, but some of the components in CBT can also be defined within the framework of MBC. Not only are symptom and function scores used pre-treatment, but there are evaluations session by session where patients rate several aspects of each session. The evaluation consists of a broad base of questions ranging from semantic understanding of the elements from the CBT model, how they feel about the format, alliance, to evaluation of the therapist and so on. This is used as feedback to the therapist to improve the tailoring of the model and one can assume that this leads both to better tailored treatment for the patient and also makes the therapist a better provider of CBT. When combined with symptom measures as the CORE-OM, this study also is an example of deliberate practice within CBT. Within this framework an instrument with good psychometric properties is warranted. CORE-17 with the suggested modifications includes factors which are clinically meaningful to monitor during a treatment and can be used in a framework of measurement-based care and deliberate practice.

5.3 Future research

Further evaluation is needed of the full-scale program including the assertiveness module to find the optimal combination with additional interventions. The study should be replicated with an active control group also at follow-up to rule out possible placebo effects and to evaluate what incremental effects the SMART treatment can contribute with

compared to treatment already established in the CAMHS. Due to the short duration of the intervention, the results of an additional cost-utility analysis would be interesting. This study was not designed to test moderators. It would be interesting to design a study to look in more detail at the effects of the different elements in the treatment, characteristics of the patient, and contextual factors. There are several research questions that can be answered in the existing data set. Overall, there are several overarching issues that need to be discussed when it comes to further research.

Presently, SMART is a part of the treatment offered in the participating CAMHS. This could indicate that the SMART program implemented in connection with the present research project has been considered clinically useful for the participating clinics and for the patients in CAMHS. The treatment outcome was evaluated not only with effectiveness measures, but also measures of alliance, treatment integrity and user satisfaction, providing information about effectiveness and acceptability, as well as feasibility of this transdiagnostic intervention. The completion ratio of the treatment was also high. This raises important questions for further research, firstly, what are the effects of the methodologic rigor in itself? Secondly, what were the ingredients responsible for successful implementation. As mentioned in the introduction, there are distinctive differences between research clinics and ordinarily CAMHS settings when it comes to the preconditions necessary to adhere to the strict methodological rigor demanded in performing an RCT. A remaining question is to what degree we can separate the effects, supervision, adherence, format, training and so on. For example, it could be interesting to perform research on different aspects of supervision. A big part of the supervision was devoted to tailor the treatment to the characteristics of the patient based on cognitive case conceptualizations and feedback, but also tailoring characteristics of the therapists to strengthen their identity as performers of manualized CBT. This could be investigated in a naturalistic study or a dismantling study where supervision, adherence and training could be differentiated.

Studies show that by improving the transformation from research to practice, there will not only be better quality of the services, but the effect of the intervention will also improve, hence better results will be produced (Durlak & DuPre, 2008). Generally, findings suggest that studies performed in less controlled research settings gain poorer results than studies performed in highly controlled research settings, resulting in lower treatment quality

(Dusenbury et al., 2005; Gottfredson & Gottfredson, 2002). As mentioned earlier this raises the question whether SMART would be as effective without the strict framework of feedback, adherence and supervision provided in the study. This could be investigated by performing a naturalistic observational study with SMART as it is delivered today as part of routine outcome measurement and quality assurance in the outpatient clinics.

6 Concluding remarks

This study supports the growing evidence of transdiagnostic treatments as an option in treating depression and anxiety in CAMHS. However, the attrition at follow-up was high so the results should be interpreted with caution. There also is a need for further research on the participants not showing effects, and reasons and psychological status of the people not participating in follow-up.

Further, the SMART program has built in the core components of CBT, but also components of measurement-based care, with feedback tools such as goal setting, alliance, adherence, user satisfaction in sessions and semantic understanding of the sessions. The use of these components as it is intended with the patient, with high adherence and good tailoring, is a prerequisite for generalization of the results from this study.

According to the factors found in this study in the CORE-OM, we recommend the 17-item factor as a more reliable measure of general problems. Comparing means for gender in non-clinical samples should not be done without modification of the general emotional problem and the positive resources scales. Mixing of positively and negatively keyed items should be avoided, if the intention is not to measure separate dimensions. This should be objectives for future revisions of the scale.

The effects of treatment will vary with the measures used, population studied, clinical setting, inclusion, length and so on. Ideally and maybe in the future, there could be agreement on how to report such studies, data could be merged to consist of large enough samples to divide them into subgroups to compare different clinical samples, aspects of

treatment, patients and service providers. As suggested in an international consensus statement concerning recommendations for reporting on treatment trials for child and adolescent anxiety disorders, it is offered recommendations for selection and reporting of outcome measures in clinical trials to guide further research, improve communication and to take full advantage of data sharing possibilities (Creswell et al., 2021). With increasing number of studies on anxiety and depression in different clinical settings, formats, and service provider levels maybe in the future we can have enough evidence to guide what the steps in stepped care should consist of, resulting in clinical guidelines for depression and anxiety treatment for children and adolescence in CAMHS settings.

7 References

Aebi, M., Kuhn, C., Banaschewski, T., Grimmer, Y., Poustka, L., Steinhausen, H.-C., & Goodman, R. (2017). The contribution of parent and youth information to identify mental health disorders or problems in adolescents. *Child and adolescent psychiatry and mental health*, *11*(1), 1-12.

Agerup, T., Lydersen, S., Wallander, J., & Sund, A. (2014). Longitudinal course of diagnosed depression from ages 15 to 20 in a community sample: Patterns and parental risk factors. *Child Psychiatry & Human Development*, *45*(6), 753-764. https://doi.org/10.1007/s10578-014-0444-8

Albert, N., & Beck, A. T. (1975). Incidence of depression in early adolescence: A preliminary study. *Journal of Youth and Adolescence*, *4*(4), 301-307. https://doi.org/10.1007/BF01536895

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review*, *30*(2), 217-237. https://doi.org/10.1016/j.cpr.2009.11.004

Amble, I., Gude, T., Stubdal, S., Andersen, B. J., & Wampold, B. E. (2015). The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: A multisite randomized clinical trial in a naturalistic setting. *Psychotherapy Research*, *25*(6), 669-677. https://doi.org/10.1080/10503307.2014.928756

Andrews, G., Bell, C., Boyce, P., Gale, C., Lampe, L., Marwat, O., Rapee, R., & Wilkins, G. (2018). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of panic disorder, social anxiety disorder and generalised anxiety disorder. *Australian & New Zealand Journal of Psychiatry*, *52*(12), 1109-1172.

Angold, A. (1989). Structured assessments of psychopathology in children and adolescents. *The instruments of psychiatric research*.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(1), 57-87.

Arias, E., MacDorman, M. F., Strobino, D. M., & Guyer, B. (2003). Annual summary of vital statistics--2002. *Pediatrics*, *112*(6 Pt 1), 1215-1230. https://doi.org/10.1542/peds.112.6.1215

Ask, H., Handal, M., Hauge, L. J., Reichborn-Kjennerud, T., & Skurtveit, S. (2020). Incidence of diagnosed pediatric anxiety disorders and use of prescription drugs: a nation-wide registry study. *European child & adolescent psychiatry*, *29*(8), 1063-1073.

Avenevoli, S., Knight, E., Kessler, R. C., & Merikangas, K. R. (2008). Epidemiology of depression in children and adolescents.

Baker-Ericzén, M. J., Hurlburt, M. S., Brookman-Frazee, L., Jenkins, M. M., & Hough, R. L. (2010). Comparing child, parent, and family characteristics in usual care and empirically supported treatment research samples for children with disruptive behavior disorders. *Journal of Emotional and Behavioral Disorders*, *18*(2), 82-99.

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, *13*(2), 205-221.

Barkham, M., Evans, C., Margison, F., & McGrath, G. (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health*, *7*(1), 35.

Barkham, M., Hardy, G. E., & Mellor-Clark, J. (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. John Wiley & Sons.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of consulting and clinical psychology*, *69*(2), 184.

Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE System. *Counselling and Psychotherapy Research*, *6*(1), 3-15.

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & May, J. T. E. (2010). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. Oxford University Press.

Baumgartner, H., & Steenkamp, J.-B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of marketing research*, *38*(2), 143-156.

Bear, H. A., Edbrooke-Childs, J., Norton, S., Krause, K. R., & Wolpert, M. (2020). Systematic Review and Meta-analysis: Outcomes of Routine Specialist Mental Health Care for Young People With Depression and/or Anxiety. *J Am Acad Child Adolesc Psychiatry*, *59*(7), 810-841. https://doi.org/10.1016/j.jaac.2019.12.002

Bearman, S. K., & Weisz, J. R. (2015). Comprehensive treatments for youth comorbidity–evidence-guided approaches to a complicated problem. *Child and Adolescent Mental Health*, *20*(3), 131-141.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory (BDI-II)* (Vol. 10). Pearson London, UK.

Bedford, A., Watson, R., Lyne, J., Tibbles, J., Davies, F., & Deary, I. J. (2010). Mokken scaling and principal components analyses of the CORE-OM in a large clinical sample. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *17*(1), 51-62.

Beesdo, K., Knappe, S., & Pine, D. S. (2009). Anxiety and anxiety disorders in children and adolescents: developmental issues and implications for DSM-V. *The Psychiatric clinics of North America*, *32 3*, 483-524.

Berg, M. T., Simons, R. L., Barr, A., Beach, S. R., & Philibert, R. A. (2017). Childhood/Adolescent stressors and allostatic load in adulthood: Support for a calibration model. *Social Science & Medicine*, *193*, 130-139.

Berg-Nielsen, T. S., Vika, A., & Dahl, A. A. (2003). When adolescents disagree with their mothers: CBCL-YSR discrepancies related to maternal depression and adolescent self-esteem. *Child: care, health and development*, *29*(3), 207-213.

Bickman, L., Douglas, S. R., De Andrade, A. R. V., Tomlinson, M., Gleacher, A., Olin, S., & Hoagwood, K. (2016). Implementing a measurement feedback system: A tale of two sites. *Administration and Policy in Mental Health and Mental Health Services Research*, *43*(3), 410-425.

Bickman, L., Kelley, S. D., Breda, C., de Andrade, A. R., & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: Results of a randomized trial. *Psychiatric Services*, *62*(12), 1423-1429.

Birmaher, B., Arbelaez, C., & Brent, D. (2002). Course and outcome of child and adolescent major depressive disorder. *Child and Adolescent Psychiatric Clinics*, *11*(3), 619-637.

Birmaher, B., Ryan, N. D., Williamson, D. E., Brent, D. A., Kaufman, J., Dahl, R. E., Perel, J., & Nelson, B. (1996). Childhood and adolescent depression: a review of the past 10 years. Part I. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(11), 1427-1439.

Brent, D. A., Baugher, M., Bridge, J., Chen, T., & Chiappetta, L. (1999). Age-and sex-related risk factors for adolescent suicide. *Journal of the American Academy of Child & Adolescent Psychiatry*, *38*(12), 1497-1505.

Brent, D. A., Moritz, G., Bridge, J., Perper, J., & Canobbio, R. (1996). The impact of adolescent suicide on siblings and parents: a longitudinal follow-up. *Suicide Life Threat Behav*, *26*(3), 253-259. https://www.ncbi.nlm.nih.gov/pubmed/8897664

Brown, B. B., & Bakken, J. P. (2011). Parenting and peer relationships: Reinvigorating research on family–peer linkages in adolescence. *Journal of research on adolescence*, *21*(1), 153-165.

Brown, T. A., Chorpita, B. F., & Barlow, D. H. (1998). Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. *Journal of abnormal psychology*, *107*(2), 179.

Brown, T. A., Di Nardo, P. A., Lehman, C. L., & Campbell, L. A. (2001). Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. *Journal of abnormal psychology*, *110*(1), 49.

Brøndbo, P. H., Mathiassen, B., Martinussen, M., Handegard, B. H., & Kvernmo, S. (2013). Agreement on diagnoses of mental health problems between an online clinical assignment and a routine clinical assignment. *Journal of Telemedicine and Telecare*, *19*(2), 113-119.

Brøndbo, P. H., Mathiassen, B., Martinussen, M., Heiervang, E., Eriksen, M., Moe, T. F., Sæther, G., & Kvernmo, S. (2011). The strengths and difficulties questionnaire as a screening instrument for norwegian child and adolescent mental health services, application of UK scoring algorithms. *Child Adolesc Psychiatry Ment Health*, *5*, 32. https://doi.org/10.1186/1753-2000-5-32

Bugatti, M., & Boswell, J. F. (2022). Clinician perceptions of nomothetic and individualized patient-reported outcome measures in measurement-based care. *Psychotherapy Research*, 1-12.

Bøe, T., Hysing, M., Skogen, J. C., & Breivik, K. (2016). The Strengths and Difficulties Questionnaire (SDQ): Factor structure and gender equivalence in Norwegian adolescents. *PloS one*, *11*(5), e0152202.

Campbell, O. L., Bann, D., & Patalay, P. (2021). The gender gap in adolescent mental health: a cross-national investigation of 566,829 adolescents across 73 countries. *SSM-population health*, *13*, 100742.

Carlucci, L., Saggino, A., & Balsamo, M. (2021). On the efficacy of the unified protocol for transdiagnostic treatment of emotional disorders: A systematic review and meta-analysis. *Clinical psychology review*, *87*, 101999.

Castonguay, L. G. (2013). Psychotherapy outcome: An issue worth re-revisiting 50 years later.

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: a review of literature. *Production and operations management*, *12*(4), 519-549.

Chavira, D. A., Stein, M. B., Bailey, K., & Stein, M. T. (2004). Child anxiety in primary care: Prevalent but untreated. *Depression and anxiety*, *20*(4), 155-164.

Cheung, A. H., & Dewa, C. S. (2006). Canadian community health survey: major depressive disorder and suicidality in adolescents. *Healthc Policy*, *2*(2), 76-89. https://www.ncbi.nlm.nih.gov/pubmed/19305706

Cheung, A. H., Zuckerbrot, R. A., Jensen, P. S., Laraque, D., Stein, R. E., Levitt, A., Birmaher, B., Campo, J., Clarke, G., & Emslie, G. (2018). Guidelines for adolescent depression in primary care (GLAD-PC): Part II. Treatment and ongoing management. *Pediatrics*, *141*(3).

Chorpita, B. F., & Daleiden, E. L. (2009). Mapping evidence-based treatments for children and adolescents: application of the distillation and matching model to 615 treatments from 322 randomized trials. *Journal of consulting and clinical psychology*, *77*(3), 566.

Chorpita, B. F., Daleiden, E. L., & Weisz, J. R. (2005). Modularity in the design and application of therapeutic interventions. *Applied and Preventive Psychology*, *11*(3), 141-156.
Clark, L., & Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of abnormal psychology*, *100*(3), 316-336.
Cohen, P., Cohen, J., Kasen, S., Velez, C. N., Hartmark, C., Johnson, J., Rojas, M., Brook, J., & Streuning, E. L. (1993). An epidemiological study of disorders in late childhood and adolescence--I. Age- and gender-specific prevalence. *J Child Psychol Psychiatry*, *34*(6), 851-867. https://doi.org/10.1111/j.1469-7610.1993.tb01094.x
Colosi, R. (2005). Negatively worded questions cause respondent confusion. *Proceedings of the Survey Research Methods Section, American Statistical Association (2005)*, 2896-2903.
Compton, S. N., Nelson, A. H., & March, J. S. (2000). Social phobia and separation anxiety symptoms in community and clinical samples of children and adolescents. *J Am Acad Child Adolesc Psychiatry*, *39*(8), 1040-1046. https://doi.org/10.1097/00004583-200008000-00020
Conley, C. S., & Rudolph, K. D. (2009). The emerging sex difference in adolescent depression: Interacting contributions of puberty and peer stress. *Development and psychopathology*, *21*(2), 593-620.
Connell, J., Barkham, M., Stiles, W. B., Twigg, E., Singleton, N., Evans, O., & Miles, J. N. (2007). Distribution of CORE–OM scores in a general population, clinical cut-off points and comparison with the CIS–R. *The British Journal of Psychiatry*, *190*(1), 69-74.
Cook, S. C., Schwartz, A. C., & Kaslow, N. J. (2017). Evidence-based psychotherapy: Advantages and challenges. *Neurotherapeutics*, *14*(3), 537-545.
Copeland, E. P., & Hess, R. S. (1995). Differences in young adolescents' coping strategies based on gender and ethnicity. *The Journal of Early Adolescence*, *15*(2), 203-219.
Copeland, W. E., Angold, A., Shanahan, L., & Costello, E. J. (2014). Longitudinal patterns of anxiety from childhood to adulthood: the Great Smoky Mountains Study. *J Am Acad Child Adolesc Psychiatry*, *53*(1), 21-33. https://doi.org/10.1016/j.jaac.2013.09.017
CORC. (2018). *University Hospital of Northern Norway CORC report.* [Unpublshed report]. London: CORC, Anna Freud National Centre for Children and Families.
Costello, E. J., Egger, H., & Angold, A. (2005). 10-Year Research Update Review: The Epidemiology of Child and Adolescent Psychiatric Disorders: I. Methods and Public Health Burden. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*(10), 972-986. https://doi.org/https://doi.org/10.1097/01.chi.0000172552.41596.6f
Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch Gen Psychiatry*, *60*(8), 837-844. https://doi.org/10.1001/archpsyc.60.8.837
Craske, M. G., Farchione, T. J., Allen, L. B., Barrios, V., Stoyanova, M., & Rose, R. (2007). Cognitive behavioral therapy for panic disorder and comorbidity: more of the same or less of more? *Behaviour Research and Therapy*, *45*(6), 1095-1109.
Creswell, C., Nauta, M. H., Hudson, J. L., March, S., Reardon, T., Arendt, K., Bodden, D., Cobham, V. E., Donovan, C., & Halldorsson, B. (2021). Research Review: Recommendations for reporting on treatment trials for child and adolescent anxiety disorders–an international consensus statement. *Journal of Child Psychology and Psychiatry*, *62*(3), 255-269.

Cummings, C. M., Caporino, N. E., & Kendall, P. C. (2014). Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychol Bull*, *140*(3), 816-845. https://doi.org/10.1037/a0034733
Dalgleish, T., Black, M., Johnston, D., & Bevan, A. (2020). Transdiagnostic approaches to mental health problems: Current status and future directions. *Journal of consulting and clinical psychology*, *88*(3), 179.
Dalsgaard, S., Thorsteinsson, E., Trabjerg, B. B., Schullehner, J., Plana-Ripoll, O., Brikell, I., Wimberley, T., Thygesen, M., Madsen, K. B., & Timmerman, A. (2020). Incidence rates and cumulative incidences of the full spectrum of diagnosed mental disorders in childhood and adolescence. *JAMA psychiatry*, *77*(2), 155-164.
Del Re, A., Maisel, N. C., Blodgett, J. C., & Finney, J. W. (2013). Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. *BMJ open*, *3*(11), e003464.
Dolle, K., & Schulte-Körne, G. (2013). The treatment of depressive disorders in children and adolescents. *Deutsches Ärzteblatt International*, *110*(50), 854.
Donahue, J. J. (2020). Fight-Flight-Freeze System. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1590-1595). Springer International Publishing. https://doi.org/10.1007/978-3-319-24612-3_751
Dorsey, S., McLaughlin, K. A., Kerns, S. E. U., Harrison, J. P., Lambert, H. K., Briggs, E. C., Revillion Cox, J., & Amaya-Jackson, L. (2017). Evidence Base Update for Psychosocial Treatments for Children and Adolescents Exposed to Traumatic Events. *J Clin Child Adolesc Psychol*, *46*(3), 303-330. https://doi.org/10.1080/15374416.2016.1220309
Duncan, B. L., & Reese, R. J. (2015). The Partners for Change Outcome Management System (PCOMS) revisiting the client's frame of reference. *Psychotherapy*, *52*(4), 391.
Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, *41*(3), 327-350.
Dusenbury, L., Brannigan, R., Hansen, W. B., Walsh, J., & Falco, M. (2005). Quality of implementation: developing measures crucial to understanding the diffusion of preventive interventions. *Health Education Research*, *20*(3), 308-313.
Ebesutani, C., Drescher, C. F., Reise, S. P., Heiden, L., Hight, T. L., Damon, J. D., & Young, J. (2012). The loneliness questionnaire–short version: An evaluation of reverse-worded and non-reverse-worded items via item response theory. *Journal of Personality Assessment*, *94*(4), 427-437.
Edbrooke-Childs, J., Jacob, J., Law, D., Deighton, J., & Wolpert, M. (2015). Interpreting standardized and idiographic outcome measures in CAMHS: what does change mean and how does it relate to functioning and experience? *Child and Adolescent Mental Health*, *20*(3), 142-148.
Ehrenreich-May, J., & Chu, B. C. (2013). Overview of transdiagnostic mechanisms and treatments for youth psychopathology. *Transdiagnostic treatments for children and adolescents: Principles and practices*, 3-14.
Eidstuen, J. S., & Kornør, H. (2017). Diagnostisk nøyaktighet ved algoritmen til den norske versjonen av Strengths and Difficulties Questionnaire (SDQ). *PsykTestBarn*, *1:2*.
Elfström, M., Evans, C., Lundgren, J., Johansson, B., Hakeberg, M., & Carlsson, S. (2013). Validation of the Swedish version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM). *Clinical Psychology & Psychotherapy*, *20*(5), 447-455.
Emmelkamp, P. M., David, D., Beckers, T., Muris, P., Cuijpers, P., Lutz, W., Andersson, G., Araya, R., Banos Rivera, R. M., & Barkham, M. (2014). Advancing psychotherapy

and evidence-based psychological interventions. *International journal of methods in psychiatric research*, *23*(S1), 58-91.
Emslie, G. J., Mayes, T. L., Laptook, R. S., & Batt, M. (2003). Predictors of response to treatment in children and adolescents with mood disorders. *Psychiatric Clinics*, *26*(2), 435-456.
Essau, C., & Dobson, K. (1999). Depressive disorders in children and adolescents: epidemiology, course and treatment. *Epidemiology of depressive disorders. Jason Aronson Inc., Northvale*, 69-103.
Essau, C. A., & Gabbidon, J. (2012). Epidemiology, Comorbidity and Mental Health Services Utilization. In C. A. Essau & T. H. Ollendick (Eds.), *The Wiley-Blackwell handbook of the treatment of childhood and adolescent anxiety* (pp. 23-42). John Wiley & Sons.
Evans, C., Connell, J., Barkham, M., Margison, F., McGRATH, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE–OM. *The British Journal of Psychiatry*, *180*(1), 51-60.
Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., Audin, K., Connell, J., & Graeme McGrath, C. (2000). CORE: clinical outcomes in routine evaluation. *Journal of Mental Health*, *9*(3), 247-255.
Ezpeleta, L., Domenech, J. M., & Angold, A. (2006). A comparison of pure and comorbid CD/ODD and depression. *Journal of Child Psychology and Psychiatry*, *47*(7), 704-712.
Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. *Review of General Psychology*, *17*(1), 111-121.
FHI. (2016). *Quality of life and mental health among children and adolescents in Norway.* Public Health Report - Health Status in Norway [online document]. https://www.fhi.no/en/op/hin/mental-health/mental-health-children-adolescents/: Oslo: Institute of Public Health Retrieved from https://www.fhi.no/en/op/hin/mental-health/mental-health-children-adolescents/
Flores, B., Musselman, D., DeBattista, C., Garlow, S., Schatzberg, A., & Nemeroff, C. (2004). Biology of mood disorders. *Textbook of psychopharmacology [3rd ed]. Washington DC: American Psychiatric Publishing*, *718*, 717-763.
Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: the prevalence of DSM-IV disorders. *J Am Acad Child Adolesc Psychiatry*, *42*(10), 1203-1211. https://doi.org/10.1097/00004583-200310000-00011
Fox, N. A., Nichols, K. E., Henderson, H. A., Rubin, K., Schmidt, L., Hamer, D., Ernst, M., & Pine, D. S. (2005). Evidence for a gene-environment interaction in predicting behavioral inhibition in middle childhood. *Psychological Science*, *16*(12), 921-926.
Freeman, J., Benito, K., Herren, J., Kemp, J., Sung, J., Georgiadis, C., Arora, A., Walther, M., & Garcia, A. (2018). Evidence Base Update of Psychosocial Treatments for Pediatric Obsessive-Compulsive Disorder: Evaluating, Improving, and Transporting What Works. *J Clin Child Adolesc Psychol*, *47*(5), 669-698. https://doi.org/10.1080/15374416.2018.1496443
Frydenberg, E., & Lewis, R. (1997). Coping with Stresses and Concerns during Adolescence: A Longitudinal Study.
Garber, J., & Weersing, V. R. (2010). Comorbidity of Anxiety and Depression in Youth: Implications for Treatment and Prevention. *Clin Psychol (New York)*, *17*(4), 293-306. https://doi.org/10.1111/j.1468-2850.2010.01221.x
Garg, R. K. (1996). The influence of positive and negative wording and issue involvement on responses to Likert scales in marketing research. *Market Research Society. Journal.*, *38*(3), 1-11.

Garland, A. F., Haine-Schlagel, R., Brookman-Frazee, L., Baker-Ericzen, M., Trask, E., & Fawley-King, K. (2013). Improving community-based mental health care for children: Translating knowledge into action. *Administration and Policy in Mental Health and Mental Health Services Research*, *40*(1), 6-22.

Gibbons, C. J., & DeRubeis, R. J. (2008). Anxiety symptom focus in sessions of cognitive therapy for depression. *Behavior Therapy*, *39*(2), 117-125.

Ginsburg, G. S., Becker, E. M., Keeton, C. P., Sakolsky, D., Piacentini, J., Albano, A. M., Compton, S. N., Iyengar, S., Sullivan, K., & Caporino, N. (2014). Naturalistic follow-up of youths treated for pediatric anxiety disorders. *JAMA psychiatry*, *71*(3), 310-318.

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American journal of public health*, *93*(8), 1261-1267.

Goodman, A., Heiervang, E., Collishaw, S., & Goodman, R. (2011). The 'DAWBA bands' as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples. *Social psychiatry and psychiatric epidemiology*, *46*(6), 521-532.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581-586.

Goodman, R. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *Journal of Child Psychology and Psychiatry*, *40*(5), 791-799. https://doi.org/https://doi.org/10.1111/1469-7610.00494

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345.

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 645-655.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*(6), 534-539.

Gore, F. M., Bloem, P. J., Patton, G. C., Ferguson, J., Joseph, V., Coffey, C., Sawyer, S. M., & Mathers, C. D. (2011). Global burden of disease in young people aged 10–24 years: a systematic analysis. *The Lancet*, *377*(9783), 2093-2102.

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of research in crime and delinquency*, *39*(1), 3-35.

Green, H., McGinnity, Á., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain, 2004*. Palgrave macmillan Basingstoke.

Gros, D. F. (2014). Development and initial evaluation of Transdiagnostic Behavior Therapy (TBT) for veterans with affective disorders. *Psychiatry research*, *220*(1-2), 275-282.

Gutermann, J., Schwartzkopff, L., & Steil, R. (2017). Meta-analysis of the long-term treatment effects of psychological interventions in youth with PTSD symptoms. *Clinical child and family psychology review*, *20*(4), 422-434.

Hamlat, E. J., Snyder, H. R., Young, J. F., & Hankin, B. L. (2019). Pubertal timing as a transdiagnostic risk for psychopathology in youth. *Clinical Psychological Science*, *7*(3), 411-429.

Hankin, B. L., Abramson, L. Y., Moffitt, T. E., Silva, P. A., McGee, R., & Angell, K. E. (1998). Development of depression from preadolescence to young adulthood:

emerging gender differences in a 10-year longitudinal study. *J Abnorm Psychol*, *107*(1), 128-140. https://doi.org/10.1037//0021-843x.107.1.128
Hannan, C., Lambert, M. J., Harmon, C., Nielsen, S. L., Smart, D. W., Shimokawa, K., & Sutton, S. W. (2005). A lab test and algorithms for identifying clients at risk for treatment failure. *Journal of clinical psychology*, *61*(2), 155-163.
Hanssen-Bauer, K., Gowers, S., Aalen, O. O., Bilenberg, N., Brann, P., Garralda, E., Merry, S., & Heyerdahl, S. (2007). Cross-national reliability of clinician-rated outcome measures in child and adolescent mental health services. *Administration and Policy in Mental Health and Mental Health Services Research*, *34*(6), 513-518.
Harvey, A. G., Murray, G., Chandler, R. A., & Soehner, A. (2011). Sleep disturbance as transdiagnostic: consideration of neurobiological mechanisms. *Clinical psychology review*, *31*(2), 225-235.
Helsedirektoratet. (2018). *Pakkeforløp for utredning og behandling i psykisk helsevern, barn og unge [nettdokument]*. Oslo: Helsedirektoratet (sist faglig oppdatert 08. desember 2021, lest 04. april 2022). Retrieved from https://www.helsedirektoratet.no/pakkeforlop/psykiske-lidelser-barn-og-unge
Lov om medisinsk og helsefaglig forskning (helseforskningsloven), (2008). https://lovdata.no/dokument/NL/lov/2008-06-20-44
Hesedirektoratet. (2015). *Prioriteringsveileder – psykisk helsevern for barn og unge [nettdokument]*. Tilgjengelig fra https://www.helsedirektoratet.no/veiledere/prioriteringsveiledere/psykisk-helsevern-for-barn-og-unge: Oslo: Helsedirektoratet
Higa-McMillan, C. K., Francis, S. E., Rith-Najarian, L., & Chorpita, B. F. (2016). Evidence base update: 50 years of research on treatment for child and adolescent anxiety. *Journal of Clinical Child & Adolescent Psychology*, *45*(2), 91-113.
Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Bmj*, *319*(7211), 670-674.
Hunsley, J. (2007). Addressing key challenges in evidence-based practice in psychology. *Professional Psychology: Research and Practice*, *38*(2), 113.
Hunsley, J., Elliott, K., & Therrien, Z. (2014). The efficacy and effectiveness of psychological treatments for mood, anxiety, and related disorders. *Canadian Psychology/Psychologie canadienne*, *55*(3), 161.
Hunsley, J., & Lee, C. M. (2007). informed benchmarks for psychological treatments: Efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice*, *38*(1), 21.
Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
Hysing, M., Elgen, I., Gillberg, C., Lie, S. A., & Lundervold, A. J. (2007). Chronic physical illness and mental health in children. Results from a large-scale population study. *Journal of Child Psychology and Psychiatry*, *48*(8), 785-792.
In-Albon, T., & Schneider, S. (2007). Psychotherapy of childhood anxiety disorders: A meta-analysis. *Psychotherapy and psychosomatics*, *76*(1), 15-24.
Jacobson, N. S., & Truax, P. (1991). Clinical Significance: A Statistical Approach to Denning Meaningful Change in Psychotherapy Research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12-19.
James, A. C., Reardon, T., Soler, A., James, G., & Creswell, C. (2020). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane database of systematic reviews*(11).
Jaycox, L. H., Stein, B. D., Paddock, S., Miles, J. N., Chandra, A., Meredith, L. S., Tanielian, T., Hickey, S., & Burnam, M. A. (2009). Impact of teen depression on academic,

social, and physical functioning. *Pediatrics*, *124*(4), e596-605. https://doi.org/10.1542/peds.2008-3348
Jose, P. E., & Ratcliffe, V. (2004). Stressor frequency and perceived intensity as predictors of internalizing symptoms: Gender and age differences in adolescence. *New Zealand Journal of Psychology*, *33*(3), 145-154.
Junge, J. (2003). *Primäre Prävention von Angst und Depression bei Jugendlichen: Entwicklung und Evaluation des Programms GO!-" Gesundheit und Optimismus"*
Kagan, J., Reznick, J. S., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child development*, 1459-1473.
Kendall, P. C., Compton, S. N., Walkup, J. T., Birmaher, B., Albano, A. M., Sherrill, J., Ginsburg, G., Rynn, M., McCracken, J., & Gosch, E. (2010). Clinical characteristics of anxiety disordered youth. *Journal of anxiety disorders*, *24*(3), 360-365.
Kendall, P. C., & Peterman, J. S. (2015). CBT for adolescents with anxiety: Mature yet still developing. *American Journal of Psychiatry*, *172*(6), 519-530.
Kendall, P. C., Safford, S., Flannery-Schroeder, E., & Webb, A. (2004). Child anxiety treatment: outcomes in adolescence and impact on substance use and depression at 7.4-year follow-up. *Journal of consulting and clinical psychology*, *72*(2), 276.
Kennard, B. D., Silva, S. G., Tonev, S., Rohde, P., Hughes, J. L., Vitiello, B., Kratochvil, C. J., Curry, J. F., Emslie, G. J., & Reinecke, M. (2009). Remission and recovery in the Treatment for Adolescents with Depression Study (TADS): acute and long-term outcomes. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(2), 186-195.
Kessler, R. C., & Walters, E. E. (1998). Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the National Comorbidity Survey. *Depress Anxiety*, *7*(1), 3-14. https://doi.org/10.1002/(sici)1520-6394(1998)7:1<3::aid-da2>3.0.co;2-f
Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of clinical psychology*, *61*(3), 285-314.
Krause, K. R., Edbrooke-Childs, J., Singleton, R., & Wolpert, M. (2022). Are we comparing apples with oranges? Assessing improvement across symptoms, functioning, and goal progress for adolescent anxiety and depression. *Child Psychiatry & Human Development*, *53*(4), 737-753.
La Greca, A. M., Silverman, W. K., & Lochman, J. E. (2009). Moving beyond efficacy and effectiveness in child and adolescent intervention research. *Journal of consulting and clinical psychology*, *77*(3), 373.
Lambert, M. J. (2010). *Prevention of treatment failure: The use of measuring, monitoring, and feedback in clinical practice*. American Psychological Association.
Lambert, M. J. (2013). Outcome in psychotherapy: the past and important advances.
Lambert, M. J. (2015). Progress feedback and the OQ-system: The past and the future. *Psychotherapy*, *52*(4), 381.
Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy*, *48*(1), 72.
Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? a meta-analysis. *Clinical Psychology: Science and Practice*, *10*(3), 288.
Lambert, M. J., Whipple, J. L., & Kleinstäuber, M. (2018). Collecting and delivering progress feedback: A meta-analysis of routine outcome monitoring. *Psychotherapy*, *55*(4), 520.

Lee, C. M., Horvath, C., & Hunsley, J. (2013). Does it work in the real world? The effectiveness of treatments for psychological problems in children and adolescents. *Professional Psychology: Research and Practice*, *44*(2), 81.

Lewczyk, C. M., Garland, A. F., Hurlburt, M. S., Gearity, J., & Hough, R. L. (2003). Comparing DISC-IV and clinician diagnoses among youths receiving public mental health services. *J Am Acad Child Adolesc Psychiatry*, *42*(3), 349-356. https://doi.org/10.1097/00004583-200303000-00016

Lewinsohn, P. M., Clarke, G. N., Seeley, J. R., & Rohde, P. (1994). Major depression in community adolescents: age at onset, episode duration, and time to recurrence. *J Am Acad Child Adolesc Psychiatry*, *33*(6), 809-818. https://doi.org/10.1097/00004583-199407000-00006

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *J Abnorm Psychol*, *102*(1), 133-144. https://doi.org/10.1037//0021-843x.102.1.133

Liber, J. M., van Widenfelt, B. M., van der Leeden, A. J., Goedhart, A. W., Utens, E. M., & Treffers, P. D. (2010). The relation of severity and comorbidity to treatment outcome with Cognitive Behavioral Therapy for childhood anxiety disorders. *J Abnorm Child Psychol*, *38*(5), 683-694. https://doi.org/10.1007/s10802-010-9394-1

Lorentzen, V. (2008). *"Ja takk begge deler?" En utprøving av SMART, manualisert gruppebehandling for angst og depresjon kombinert med individuell oppfølging for ungdom 15–18 år.* Universitetet i Tromsø.

Lorentzen, V., Fagermo, K., Handegard, B. H., Skre, I., & Neumer, S. P. (2020). A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14-17 years old in child and adolescent mental health services (CAMHS). *BMC Psychol*, *8*(1), 25. https://doi.org/10.1186/s40359-020-0393-x

Lorentzen, V., Fagermo, K., Handegård, B. H., Neumer, S.-P., & Skre, I. (2022). Long-term effectiveness and trajectories of change after treatment with SMART, a transdiagnostic CBT for adolescents with emotional problems. *BMC Psychology*, *10*(1), 1-18.

Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology*, *45*(2), 185-203.

Machin, D., Campbell, M. J., Tan, S. B., & Tan, S. H. (2011). *Sample size tables for clinical studies*. John Wiley & Sons.

Maj, M. (2005). "Psychiatric comorbidity": an artefact of current diagnostic systems? *Br J Psychiatry*, *186*, 182-184. https://doi.org/10.1192/bjp.186.3.182

Mansell, W., Harvey, A., Watkins, E., & Shafran, R. (2009). Conceptual foundations of the transdiagnostic approach to CBT. *Journal of Cognitive Psychotherapy*, *23*(1), 6-19.

Mansell, W., Harvey, A., Watkins, E. R., & Shafran, R. (2008). Cognitive behavioral processes across psychological disorders: A review of the utility and validity of the transdiagnostic approach. *International Journal of Cognitive Therapy*, *1*(3), 181-191.

Manteuffel, B., Stephens, R. L., Sondheimer, D. L., & Fisher, S. K. (2008). Characteristics, service experiences, and outcomes of transition-aged youth in systems of care: Programmatic and policy implications. *The Journal of Behavioral Health Services & Research*, *35*(4), 469-487.

March, J., Silva, S., Petrycki, S., Curry, J., Wells, K., Fairbank, J., Burns, B., Domino, M., McNulty, S., & Vitiello, B. (2004). Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA*, *292*(7), 807-820.

March, J. S., Parker, J. D., Sullivan, K., Stallings, P., & Conners, C. K. (1997). The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(4), 554-565.

Marchand, E., Stice, E., Rohde, P., & Becker, C. B. (2011). Moving from efficacy to effectiveness trials in prevention research. *Behaviour Research and Therapy*, *49*(1), 32-41.

Marsh, H. (1986). Negative Item Bias in Ratings Scales for Preadolescent Children:. A Cognitive-Developmental Phenomenon. *Developmental Psychology*, 37-49.

Marsh, H. (1996). Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? *Journal of personality and social psychology*, *70*(4), 810-819.

Mathai, J., Anderson, P., & Bourne, A. (2004). Comparing psychiatric diagnoses generated by the Strengths and Difficulties Questionnaire with diagnoses made by clinicians. *Australian & New Zealand Journal of Psychiatry*, *38*(8), 639-643.

McDermott, B., Baigent, M., Chanen, A., Fraser, L., Graetz, B., Hayman, N., & Spence, S. (2010). Clinical practice guidelines: Depression in adolescents and young adults. Beyondblue: The National Depression Initiative. In: Melbourne.

McHugh, R. K., & Barlow, D. H. (2010). The dissemination and implementation of evidence-based psychological treatments: a review of current efforts. *American Psychologist*, *65*(2), 73.

McKinnon, A., Keers, R., Coleman, J. R., Lester, K. J., Roberts, S., Arendt, K., Bögels, S. M., Cooper, P., Creswell, C., & Hartman, C. A. (2018). The impact of treatment delivery format on response to cognitive behaviour therapy for preadolescent children with anxiety disorders. *Journal of Child Psychology and Psychiatry*, *59*(7), 763-772.

McManus, F., Shafran, R., & Cooper, Z. (2010). What does a transdiagnostic approach have to offer the treatment of anxiety disorders? *British Journal of Clinical Psychology*, *49*(4), 491-505.

Meidlinger, P. C., & Hope, D. A. (2017). The new transdiagnostic cognitive behavioral treatments: Commentary for clinicians and clinical researchers. *Journal of anxiety disorders*, *46*, 101-109.

Mendlowicz, M. V., & Stein, M. B. (2000). Quality of life in individuals with anxiety disorders. *American Journal of Psychiatry*, *157*(5), 669-682.

Merikangas, K. R., & Knight, E. (2009). The epidemiology of depression in adolescents. In S. Nolen-Hoeksema & L. M. Hilt (Eds.), *Handbook of depression in adolescents* (pp. 53-74). Routledge.

Muris, P., Meesters, C., & van den Berg, F. (2003). The strengths and difficulties questionnaire (SDQ). *European child & adolescent psychiatry*, *12*(1), 1-8.

Naicker, K., Galambos, N. L., Zeng, Y., Senthilselvan, A., & Colman, I. (2013). Social, demographic, and health outcomes in the 10 years following adolescent depression. *J Adolesc Health*, *52*(5), 533-538. https://doi.org/10.1016/j.jadohealth.2012.12.016

Nakken, M. A. (2021). Predictive Validity of the Norwegian Beck Depression Inventory-II (BDI-II), Multidimensional Anxiety Scale for Children (MASC) and the Strengths and Difficulties Questionnaire (SDQ) Impact Score on Diagnostic Membership of Depression, Anxiety and Comorbid Anxiety and Depression. *Thesis in Clinical Psychology, Arctic university of Norway*.

Nathan, P. E., & Gorman, J. M. (2015). *A guide to treatments that work*. Oxford University Press.

Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi, (2016). https://www.forskningsetikk.no/ressurser/publikasjoner/forskningsetiske-retningslinjer-for-samfunnsvitenskap-humaniora-juss-og-teologi/
Neumer, S.-P., & Junge-Hoffmeister, J. (2010a). SMART. *Forebygging av emosjonelle problemer hos ungdom. Veileder. Oslo: Kommuneforlaget*.
Neumer, S.-P., & Junge-Hoffmeister, J. (2010b). *SMART Arbeidsbok*. Kommuneforlaget AS.
NHS. (2013). Social anxiety disorder: recognition, assessment and treatment.
NICE, N. I. f. H. a. C. E. (2019). *Depression in children and young people: identification and management (NG134)*. National Institute for Health and Care Excellence Retrieved from www.nice.org.uk/guidance/ng134
Nolen-Hoeksema, S., & Girgus, J. S. (1994). The emergence of gender differences in depression during adolescence. *Psychol Bull*, *115*(3), 424-443. https://doi.org/10.1037/0033-2909.115.3.424
Norcross, J. C., Beutler, L. E., & Levant, R. F. (2006). *Evidence-based practices in mental health: Debate and dialogue on the fundamental questions*. American Psychological Association.
Norton, P. J. (2012). *Group cognitive-behavioral therapy of anxiety: A transdiagnostic treatment manual*. Guilford Press.
O'Neil, K. A., & Kendall, P. C. (2012). Role of Comorbid Depression and Co-occurring Depressive Symptoms in Outcomes for Anxiety-Disordered Youth Treated with Cognitive-Behavioral Therapy. *Child & Family Behavior Therapy*, *34*(3), 197-209. https://doi.org/10.1080/07317107.2012.707086
Olatunji, B. O., Naragon-Gainey, K., & Wolitzky-Taylor, K. B. (2013). Specificity of rumination in anxiety and depression: A multimodal meta-analysis. *Clinical Psychology: Science and Practice*, *20*(3), 225.
Osman, A., Downs, W. R., Barrios, F. X., Kopper, B. A., Gutierrez, P. M., & Chiros, C. E. (1997). Factor structure and psychometric characteristics of the Beck Depression Inventory-II. *Journal of psychopathology and behavioral assessment*, *19*(4), 359-376.
Oudhoff, J. P., Timmermans, D. R., Bijnen, A. B., & Van Der Wal, G. (2004). Waiting for elective general surgery: physical, psychological and social consequences. *ANZ journal of surgery*, *74*(5), 361-367.
Oudhoff, J. P., Timmermans, D. R., Knol, D. L., Bijnen, A. B., & Van der Wal, G. (2007). Waiting for elective surgery: effect on physical problems and postoperative recovery. *ANZ journal of surgery*, *77*(10), 892-898.
Palmieri, G., Evans, C., Hansen, V., Brancaleoni, G., Ferrari, S., Porcelli, P., Reitano, F., & Rigatelli, M. (2009). Validation of the Italian version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *16*(5), 444-449.
Pinsof, W. M., & Chambers, A. L. (2009). Empirically informed systemic psychotherapy: Tracking client change and therapist behavior during therapy.
Plenty, S., Östberg, V., Almquist, Y. B., Augustine, L., & Modin, B. (2014). Psychosocial working conditions: An analysis of emotional symptoms and conduct problems amongst adolescent students. *Journal of adolescence*, *37*(4), 407-417.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879.
Prescott, D. S., Maeschalck, C. L., & Miller, S. D. (2017). *Feedback-informed treatment in clinical practice: Reaching for excellence*. JSTOR.

Queen, A. H., Barlow, D. H., & Ehrenreich-May, J. (2014). The trajectories of adolescent anxiety and depressive symptoms over the course of a transdiagnostic treatment. *J Anxiety Disord*, *28*(6), 511-521. https://doi.org/10.1016/j.janxdis.2014.05.007

Rapee, R. M., Lyneham, H. J., Hudson, J. L., Kangas, M., Wuthrich, V. M., & Schniering, C. A. (2013). Effect of comorbidity on treatment of anxious children and adolescents: Results from a large, combined sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(1), 47-56.

Richter, J., & Sund, A. M. (2013). Måleegenskaper ved den norske versjonen av the Mood and Feelings Questionnaire-Humøret Ditt (MFQ).

Rith-Najarian, L. R., Mesri, B., Park, A. L., Sun, M., Chavira, D. A., & Chorpita, B. F. (2019). Durability of cognitive behavioral therapy effects for youth and adolescents with anxiety, depression, or traumatic stress: A meta-analysis on long-term follow-ups. *Behavior Therapy*, *50*(1), 225-240.

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: potential trade-offs for the emotional and behavioral development of girls and boys. *Psychological bulletin*, *132*(1), 98.

Rynn, M. A., Barber, J. P., Khalid-Khan, S., Siqueland, L., Dembiski, M., McCarthy, K. S., & Gallop, R. (2006). The psychometric properties of the MASC in a pediatric psychiatric sample. *Journal of anxiety disorders*, *20*(2), 139-157.

Rønning, J. A., Handegaard, B. H., Sourander, A., & Mørch, W.-T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European child & adolescent psychiatry*, *13*(2), 73-82.

Salk, R. H., Petersen, J. L., Abramson, L. Y., & Hyde, J. S. (2016). The contemporary face of gender differences and similarities in depression throughout adolescence: Development and chronicity. *Journal of affective disorders*, *205*, 28-35.

Sampalis, J., Boukas, S., Liberman, M., Reid, T., & Dupuis, G. (2001). Impact of waiting time on the quality of life of patients awaiting coronary artery bypass grafting. *Cmaj*, *165*(4), 429-433.

Sauer-Zavala, S., Bentley, K. H., Steele, S. J., Tirpak, J. W., Ametaj, A. A., Nauphal, M., Cardona, N., Wang, M., Farchione, T. J., & Barlow, D. H. (2020). Treating depressive disorders with the Unified Protocol: A preliminary randomized evaluation. *Journal of affective disorders*, *264*, 438-445.

Sauer-Zavala, S., Gutner, C. A., Farchione, T. J., Boettcher, H. T., Bullis, J. R., & Barlow, D. H. (2017). Current definitions of "transdiagnostic" in treatment development: A search for consensus. *Behavior Therapy*, *48*(1), 128-138.

Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? Proceedings of the SIGCHI conference on human factors in computing systems,

Schaffhuser, K., Allemand, M., & Schwarz, B. (2017). The development of self-representations during the transition to early adolescence: The role of gender, puberty, and school transition. *The Journal of Early Adolescence*, *37*(6), 774-804.

Schorre, B. E. H., & Vandvik, I. H. (2004). Global assessment of psychosocial functioning in child and adolescent psychiatry. *European child & adolescent psychiatry*, *13*(5), 273-286.

Schriesheim, C. A., & Eisenbach, R. J. (1995). An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures. *Journal of Management*, *21*(6), 1177-1193.

Schulz, K. F., Altman, D. G., Moher, D., & Group*, C. (2010). CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Annals of internal medicine*, *152*(11), 726-732.

Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General psychiatry*, *40*(11), 1228-1231.

Shimokawa, K., Lambert, M. J., & Smart, D. W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of consulting and clinical psychology*, *78*(3), 298.

Silverman, W. K., & Field, A. P. (2011). *Anxiety disorders in children and adolescents* (2nd ed.). Cambridge University Press.

Silverman, W. K., Pina, A. A., & Viswesvaran, C. (2008). Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *37*(1), 105-130.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press.

Skogly, S. M. (2015). Kan Multidimensional Anxiety Scale for Children (MASC) og Beck Depression Inventory-II (BDI-II) predikere emosjonelle lidelser hos norske ungdommer i poliklinisk behandling? *Thesis in Clinical Psychology, Arctic university of Norway*.

Skre, I., Friborg, O., Elgarøy, S., Evans, C., Myklebust, L. H., Lillevoll, K., Sørgaard, K., & Hansen, V. (2013). The factor structure and psychometric properties of the Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM) in Norwegian clinical and non-clinical samples. *BMC psychiatry*, *13*(1), 1-14.

Smith, M. M., McLeod, B. D., Southam-Gerow, M. A., Jensen-Doss, A., Kendall, P. C., & Weisz, J. R. (2017). Does the delivery of CBT for youth anxiety differ across research and practice settings? *Behavior Therapy*, *48*(4), 501-516.

Solem, K. M., & Moen, C. M. (2015). *Validation of the Norwegian version of the 10-item version of Clinical Outcomes in Routine Evaluations Outcome Measure (CORE-10) in a Norwegian adolescent population (aged 14-18)* UiT Norges arktiske universitet].

Steele, S. J., Farchione, T. J., Cassiello-Robbins, C., Ametaj, A., Sbi, S., Sauer-Zavala, S., & Barlow, D. H. (2018). Efficacy of the Unified Protocol for transdiagnostic treatment of comorbid psychopathology accompanying emotional disorders compared to treatments targeting single disorders. *Journal of psychiatric research*, *104*, 211-216.

Steer, R. A., Kumar, G., Ranieri, W. F., & Beck, A. T. (1998). Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of psychopathology and behavioral assessment*, *20*(2), 127-137.

Stice, E., Rohde, P., Gau, J., & Wade, E. (2010). Efficacy trial of a brief cognitive-behavioral depression prevention program for high-risk adolescents: effects at 1-and 2-year follow-up. *J Consult Clin Psychol*, *78*(6), 856-867.

Suárez-Álvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30*(2), 149-158.

Sørheim Nilsen, T. (2017). *Change and predictors of rate of change in the treatment of children and adolescents with emotional disorders:*

A naturalistic observational study in two child and adolescent mental health services in Norway [Doctoral thesis, UiT The Arctic University of Norway]. UiT The Arctic University of Norway. https://hdl.handle.net/10037/11076

Sørheim Nilsen, T., Handegård, B.-H., Eisemann, M., & Kvernmo, S. (2015). Evaluating change in symptomatic and functional level of children and youth with emotional disorders: a naturalistic observation study. *European child & adolescent psychiatry*, *24*(10), 1219-1231.

Sørheim Nilsen, T., Handegård, B. H., Eisemann, M., & Kvernmo, S. (2016). Predictors of rate of change for children and youth with emotional disorders: a naturalistic observational study. *Child and adolescent psychiatry and mental health*, *10*(1), 1-11.
Teri, L. (1982). The use of the Beck Depression Inventory with adolescents. *Journal of abnormal child psychology*, *10*(2), 277-284.
Tomas, J. M., Oliver, A., Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural equation modeling: a multidisciplinary journal*, *20*(2), 299-313.
Treatment for Adolescents with Depression Study, T. (2005). The Treatment for Adolescents With Depression Study (TADS): demographic and clinical characteristics. *J Am Acad Child Adolesc Psychiatry*, *44*(1), 28-40. https://doi.org/10.1097/01.chi.0000145807.09027.82
Trosper, S. E., Whitton, S. W., Brown, T. A., & Pincus, D. B. (2012). Understanding the latent structure of the emotional disorders in children and adolescents. *Journal of abnormal child psychology*, *40*(4), 621-632.
Uji, M., Sakamoto, A., Adachi, K., & Kitamura, T. (2012). Psychometric properties of the Japanese version of the Clinical Outcomes in Routine Evaluation–Outcome Measure. *Comprehensive Psychiatry*, *53*(5), 600-608.
Ulvestad, A., Henriksen, A., Tuseth, A., & Fjellstad, T. (2007). Klienten-den glemte terapeut [The client-the forgotten therapist]. *Oslo: Gyldendal*.
Van Ameringen, M., Mancini, C., & Farvolden, P. (2003). The impact of anxiety disorders on educational achievement. *J Anxiety Disord*, *17*(5), 561-571. https://doi.org/10.1016/s0887-6185(02)00228-1
Van Buuren, S., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in RJ Stat. Software, 45, 1–68. In.
Van Dam, N. T., Hobkirk, A. L., Danoff-Burg, S., & Earleywine, M. (2012). Mind your words: positive and negative items create method effects on the Five Facet Mindfulness Questionnaire. *Assessment*, *19*(2), 198-204.
Van Droogenbroeck, F., Spruyt, B., & Keppens, G. (2018). Gender differences in mental health problems among adolescents and the role of social support: results from the Belgian health interview surveys 2008 and 2013. *BMC psychiatry*, *18*(1), 1-9.
van Loo, H. M., & Romeijn, J. W. (2015). Psychiatric comorbidity: fact or artifact? *Theor Med Bioeth*, *36*(1), 41-60. https://doi.org/10.1007/s11017-015-9321-0
Van Roy, B., Grøholt, B., Heyerdahl, S., & Clench-Aas, J. (2006). Self-reported strengths and difficulties in a large Norwegian population 10–19 years. *European child & adolescent psychiatry*, *15*(4), 189-198.
Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one*, *8*(7), e68967.
Villabø, M. A., Narayanan, M., Compton, S. N., Kendall, P. C., & Neumer, S.-P. (2018). Cognitive–behavioral therapy for youth anxiety: An effectiveness evaluation in community practice. *Journal of consulting and clinical psychology*, *86*(9), 751.
Villabø, M. A., & Neumer, S. P. (2017). Psychometric properties of the Norwegian version of Multidimensional Anxiety Scale for Children (MASC). *PsykTestBarn*, *1:3*.
Waite, P., & Creswell, C. (2014). Children and adolescents referred for treatment of anxiety disorders: differences in clinical characteristics. *J Affect Disord*, *167*, 326-332. https://doi.org/10.1016/j.jad.2014.06.028
Walfish, S., McAlister, B., O'Donnell, P., & Lambert, M. J. (2012). An investigation of self-assessment bias in mental health providers. *Psychological reports*, *110*(2), 639-644.

Walkup, J. T., Albano, A. M., Piacentini, J., Birmaher, B., Compton, S. N., Sherrill, J. T., Ginsburg, G. S., Rynn, M. A., McCracken, J., & Waslick, B. (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *New England Journal of Medicine*, *359*(26), 2753-2766.

Wampold, B. (2015). Routine outcome monitoring: Coming of age--With the usual developmental challenges. *Psychotherapy (Chicago, Ill.)*, *52*(4), 458-462.

Wang, C., Berge, K., Arnesen, Y., Breivik, C., Johnsen, L., Bruun-Hansen, T., Jentoft, G., & Skre, I. (2007). Psykhjelpen på Tvibit: En psykologisk rådgivningstjeneste på ungdommens kulturhus midt i Tromsø sentrum.[Easily accessible psychological counselling for adolescents: A Norwegian example.-English abstract]. *J Norw Psychol Assoc*, *44*, 247-253.

Warren, J. S., Nelson, P. L., & Burlingame, G. M. (2009). Identifying youth at risk for treatment failure in outpatient community mental health services. *Journal of Child and Family Studies*, *18*(6), 690-701.

Warren, J. S., Nelson, P. L., Mondragon, S. A., Baldwin, S. A., & Burlingame, G. M. (2010). Youth psychotherapy change trajectories and outcomes in usual care: Community mental health versus managed care settings. *Journal of consulting and clinical psychology*, *78*(2), 144.

Weems, C. F., Zakem, A. H., Costa, N. M., Cannon, M. F., & Watts, S. E. (2005). Physiological response and childhood anxiety: Association with symptoms of anxiety disorders and cognitive bias. *Journal of Clinical Child and Adolescent Psychology*, *34*(4), 712-723.

Weersing, V. R., Jeffreys, M., Do, M.-C. T., Schwartz, K. T., & Bolano, C. (2017). Evidence base update of psychosocial treatments for child and adolescent depression. *Journal of Clinical Child & Adolescent Psychology*, *46*(1), 11-43.

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of marketing research*, *49*(5), 737-747.

Weiss, B., & Weisz, J. R. (1995). Relative effectiveness of behavioral versus nonbehavioral child psychotherapy. *J Consult Clin Psychol*, *63*(2), 317-320. https://doi.org/10.1037//0022-006x.63.2.317

Weissman, M. M., Wolk, S., Goldstein, R. B., Moreau, D., Adams, P., Greenwald, S., Klier, C. M., Ryan, N. D., Dahl, R. E., & Wickramaratne, P. (1999). Depressed adolescents grown up. *JAMA*, *281*(18), 1707-1713. https://doi.org/10.1001/jama.281.18.1707

Weisz, J. R., Donenberg, G. R., Han, S. S., & Kauneckis, D. (1995). Child and adolescent psychotherapy outcomes in experiments versus clinics: why the disparity? *J Abnorm Child Psychol*, *23*(1), 83-106. https://doi.org/10.1007/BF01447046

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annu. Rev. Psychol.*, *56*, 337-363.

Weisz, J. R., Jensen, A. L., & McLeod, B. D. (2005). Development and Dissemination of Child and Adolescent Psychotherapies: Milestones, Methods, and a New Deployment-Focused Model.

Weisz, J. R., Jensen-Doss, A., & Hawley, K. M. (2006). Evidence-based youth psychotherapies versus usual clinical care: a meta-analysis of direct comparisons. *American Psychologist*, *61*(7), 671.

Weisz, J. R., Krumholz, L. S., Santucci, L., Thomassin, K., & Ng, M. Y. (2015). Shrinking the gap between research and practice: Tailoring and testing youth psychotherapies in clinical care contexts. *Annual review of clinical psychology*, *11*(1), 139-163.

Weisz, J. R., Kuppens, S., Eckshtain, D., Ugueto, A. M., Hawley, K. M., & Jensen-Doss, A. (2013). Performance of evidence-based youth psychotherapies compared with usual clinical care: A multilevel meta-analysis. *JAMA psychiatry*, *70*(7), 750-761.

Weisz, J. R., Kuppens, S., Ng, M. Y., Eckshtain, D., Ugueto, A. M., Vaughn-Coaxum, R., Jensen-Doss, A., Hawley, K. M., Krumholz Marchette, L. S., Chu, B. C., Weersing, V. R., & Fordwood, S. R. (2017). What five decades of research tells us about the effects of youth psychological therapy: A multilevel meta-analysis and implications for science and practice. *Am Psychol*, *72*(2), 79-117. https://doi.org/10.1037/a0040360

Weisz, J. R., McCarty, C. A., & Valeri, S. M. (2006). Effects of psychotherapy for depression in children and adolescents: a meta-analysis. *Psychological bulletin*, *132*(1), 132.

Weisz, J. R., Ng, M. L., Rutt, C., Lau, N., & Masland, S. (2013). Psychotherapy for children and adolescents. In M. Lambert (Ed.), *Handbook of psychotherapy and behavior change, 6th Edition* (pp. 541-586). John Wiley & Sons.

Weisz, J. R., Ugueto, A. M., Cheron, D. M., & Herren, J. (2013). Evidence-based youth psychotherapy in the mental health ecosystem. *J Clin Child Adolesc Psychol*, *42*(2), 274-286. https://doi.org/10.1080/15374416.2013.764824

Wergeland, G. J. H., Riise, E. N., & Ost, L. G. (2021). Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clin Psychol Rev*, *83*, 101918. https://doi.org/10.1016/j.cpr.2020.101918

WHO. (2021). *Depression*. Retrieved April 2, 2022 from https://www.who.int/news-room/fact-sheets/detail/depression

Wittchen, H. U., Nelson, C. B., & Lachner, G. (1998). Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. *Psychol Med*, *28*(1), 109-126. https://doi.org/10.1017/s0033291797005928

Wolpert, M. (2014). Uses and abuses of patient reported outcome measures (PROMs): Potential iatrogenic impact of PROMs implementation and how it can be mitigated. *Administration and Policy in Mental Health and Mental Health Services Research*, *41*(2), 141-145.

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of consumer research*, *30*(1), 72-91.

World Health Organization. (1992). *International statistical classification of diseases and related health problems* (10th revision. ed.). World Health Organization.

World Medical Association. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191-2194.

Ye, F., & Wallace, T. L. (2014). Psychological sense of school membership scale: Method effects associated with negatively worded items. *Journal of Psychoeducational Assessment*, *32*(3), 202-215.

Yoon, Y., Eisenstadt, M., Lereya, S. T., & Deighton, J. (2022). Gender difference in the change of adolescents' mental health and subjective wellbeing trajectories. *European child & adolescent psychiatry*, 1-10.

Aasen, H. (2001). *An empirical investigation of depression symptoms: norms, psychometric characteristics and factor structure of the Beck Depression Inventory-II* The University of Bergen].

Paper 1

Lorentzen, Veronica; Fagermo, Kenneth; Handegård, Bjørn Helge; Skre, Ingunn; Neumer, Simon-Peter. A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14–17 years old in child and adolescent mental health services (CAMHS). BMC Psychology 2020; Volume 8 (1). ISSN 2050-7283.s 1 - 12.s
doi: 10.1186/s40359-020-0393-x.

Lorentzen et al. BMC Psychology (2020) 8:25
https://doi.org/10.1186/s40359-020-0393-x

BMC Psychology

RESEARCH ARTICLE Open Access

A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14–17 years old in child and adolescent mental health services (CAMHS)

Veronica Lorentzen[1,2*], Kenneth Fagermo[2], Bjørn Helge Handegård[3], Ingunn Skre[1,4] and Simon-Peter Neumer[5,3]

Abstract

Background: This study aims to investigate effectiveness of a 6-week, transdiagnostic cognitive behavioral therapy (CBT) for anxiety and depression in adolescents, the Structured Material for Therapy (SMART), in naturalistic settings of child and adolescent mental health outpatient services (CAMHS).

Methods: A randomized controlled trial with waiting list control (WLC) was performed at three community CAMHS in Norway. Referred adolescents ($N = 163$, age = 15.72, 90.3% girls) scoring 6 or more on the emotional disorders subscale of the Strengths and Difficulties Questionnaire (SDQ) were randomly assigned to SMART or to WLC.

Results: In the treatment group (CBT), 32.9% improved in the main outcome measure (SDQ), compared to 11.6% in the WLC. Clinically significant and reliable change was experienced by 17.7% in the CBT condition, compared to 5.8% in the WLC. No patients deteriorated. Statistically significant treatment effects were achieved for internalization symptoms, anxiety symptoms and general functioning.

Conclusions: These promising findings indicate that SMART may be considered as a first step in a stepped care model for anxiety and/or depression treatment in CAMHS. The recovery rates imply that further investigations into the effectiveness of brief treatments should be made. Furthermore, there is a need for more comprehensive second-stage treatments for some of these patients.

Trial registration: ClinicalTrials.gov Identifier: NCT02150265. First registered May 292,014.

Keywords: Cognitive behavioural therapy, Adolescence, Emotional disorders, Treatment, Effectiveness, Transdiagnostic

* Correspondence: veronica.lorentzen@uit.no
[1]Department of Psychology, Faculty of Health Sciences, UIT The Arctic University of Norway, 9037 Tromsø, Norway
[2]Department of Child and Adolescent Psychiatry, Divisions of Child and Adolescent Health, University Hospital of North Norway, P.O. Box 19, 9038 Tromsø, Norway
Full list of author information is available at the end of the article

Background

Anxiety and depression are the most frequently diagnosed mental health disorders, both in the general population, and consequently also in child and adolescent mental health outpatient services (CAMHS) [1–3]. In the general population up to 10% of children and 20% of adolescents will meet the criteria of an anxiety disorder at any point in time [4]. Adolescents are at high risk for the development of depression. The percentage of adolescents with major depressive disorder range from 8 to 20% before the age of 18 [5–8]. Emotional disorders interfere negatively with various aspects of functioning and quality of life [9–13]. The prevalence of both anxiety and depressive disorders increase during adolescence [14, 15]. Comorbidity and co-occurrence of anxiety and depression is high [16] and studies show that anxiety and depression both have shared and separate features and etiology [16]. Hence, combined treatments for emotional disorders could offer effective treatments for these complex disorders.

Cognitive behavioral therapy (CBT) and interpersonal treatment (IPT) are *well-established* interventions for adolescent depression [17], and numerous studies have also demonstrated that CBT relieves anxiety symptoms in youths [18, 19]. In a comprehensive multilevel meta-analysis [20], integrating the results of 140 studies from the past five decades (1963–2013), youth psychotherapies showed a significant post-treatment effect size (ES) of 0.46. For the separate disorders, the largest ES was reported for anxiety (0.61), while treatments of depression in youths have yielded weaker ES (0.29) [20]. Other systematic reviews and meta-analyses examining the effect of youth CBT, show moderate to large treatment effects on anxiety and depression in youths [21, 22]. The majority of the evidence-based protocols for youths target symptoms of single disorders [23, 24], or symptom domains [25]. So far, the most well-known and well-studied combined treatment for the comorbid features of emotional disorders is the unified protocol for treatment of emotional disorders in adolescents [26]. Although not all studies have found a relationship between treatment outcome and comorbidity [22], some have found that comorbidity predicts poorer response to interventions in youth with both primary anxiety [27–29] and primary depression [16]. According to the previously mentioned multilevel meta-analysis [20], treatments of concurrent multiple problems, as opposed to any single targeted problem, showed an effect that was not significantly different from zero at post-treatment or follow-up [20]. Some argue that this could suggest that efforts made to concurrently treat multiple problems have been less effective than focusing more narrowly [30], suggesting new ways to address comorbidity in youths [31–33]. In an earlier review of trials of 461 youth psychotherapies, spanning from the 1960-ies and 50 years onward, Weisz and colleagues [20] found that the interventions were usually delivered in settings outside regular clinical practice, i.e. in research settings. Across the trials, only 2.1% of all study groups were described as involving clinically referred clients treated by practitioners in regular clinical practice settings [20]. When delivered in regular clinical practice, evidence-based treatments (EBT), compared to treatment as usual (TAU), has modest outcome (ES, d = 0.29) [29]. Furthermore, in several instances TAU delivered in regular clinical practice, outperformed standard EBT, usually delivered as single-disorder interventions. Even studies using exclusively diagnosed samples (d = 0.09) and studies on clinically referred youths (d = 0.17) showed low and non-significant ES values [34]. Despite the importance of quality assurance in routine practice, most CAMHS do not evaluate patients clinical change systematically [35]. A report from the Child and Outcomes Research Consortium (CORC) 2013–2016 with patients receiving treatment over six months in one of our participating CAMHS, showed improvement in many patients, however as many as 27% deteriorated [36].

In a CAMHS setting, there are high production requirements for staff, so the treatments need to be short and effective. Clinicians in a managed care setting reportedly emphasize short-term cognitive behavioral strategies [37]. Transdiagnostic treatment focuses on treatment strategies that may be generic across diverse conditions and can be defined as a therapy made available to individuals with a wide range of disorders [38]. Transdiagnostic treatment is characterized by a focus on cognitive, behavioral, and physiological processes that are shared or common across diverse disorders [38]. Although focal EBT are excellent in many ways (see 20), there may be challenges associated with implementation of several disorder-specific CBTs in regular clinical practice, and hence reasons for advocating training in one transdiagnostic CBT intervention that spans over several disorders or symptom clusters. In the framework of regular clinical practice, transdiagnostic CBT could be more applicable, time-saving, realistic to learn and cost-efficient for therapists in terms of training and application, and last but not least, it addresses the comorbid states we encounter in regular practice (e.g. 20,33).

There is a growing body of evidence demonstrating that transdiagnostic treatments could be effective in the reduction of symptoms of anxiety and depression [39], furthermore that transdiagnostic CBT has similar effects as disorder-specific interventions [40], and finally that effect sizes range from medium to large for these types of interventions [41]. The Unified Protocol for the Treatment of Emotional Disorders in Adolescents (UP-A) showed a significant effect compared to waiting list controls on all outcome measures [26]. However, as highlighted in Weisz and colleagues [20] extensive meta- analysis, the vast majority of the 1160 treatment and control groups included therapy that was not delivered in regular clinical care settings.

To the best of our knowledge, the present study is the first RCT performed with short-term transdiagnostic CBT for adolescents, the SMART protocol, with combined emotional disorders in regular clinical settings in CAMHS.

Objectives

The objective of the present study was to examine the effectiveness of a short-term, transdiagnostic CBT (SMART) in adolescents with clinically significant emotional symptoms referred to community clinics. The effectiveness is investigated both with regard to

a) emotional problems as defined by the SDQ,
b) symptoms of depression,
c) symptoms of anxiety
d) general functioning
e) and general clinical status.

Methods

The study is a randomized controlled study of the effects after 6 weeks of Structured Material for Therapy (SMART) treatment, compared with a waiting list control (WLC).

Participants

The analyzed sample comprised of 145 adolescents 14–17 years old ($M = 15.72$, $SD = 1.14$, 90.3% females), recruited from referrals to three Norwegian public child and adolescent mental health outpatient clinics (CAMHS) between January 2012 and November 2016. Participants were informed about the study during the routine intake procedure of the clinic. All adolescents, parents of children under 16 years, and adolescents over 16 years signed informed consent and received the Strengths and Difficulties Questionnaire (SDQ). Inclusion criteria were [1] age between 14 and 17 years [2]; a probable diagnosis of emotional disorder as indicated by a score of at least 6 on SDQ emotional problems subscale; and [3] maintenance of a maximum waiting time for necessary medical care of 6 weeks given by Norwegian health authorities. Exclusion criteria were [1] a diagnosis of pervasive developmental disorder (PDD) [2]; psychotic symptoms [3]; Use of anxiolytic or anti-depressant medication effects during the treatment period; and [4] patients who did not speak the Norwegian language. A total of 199 adolescents were assessed for eligibility and were asked for informed consent. Of these, 36 did not consent, 7 were excluded due to exclusion criteria [1–4], 11 withdrew from the study. A total of $N = 163$ were block randomized into direct treatment, or six-week waiting list. In the current study, 19 patients did not complete the treatment. Of these, we had no information on the reason for non-completion for 11 patients, while 2 cited lack of motivation, 3 were referred to other treatment (2 received trauma treatment, 1 regular cognitive behavioral therapy), and finally 3 withdrew because of geographical distance (2 moved to another location, 1 had a long distance to travel to get to the CAMHS). (See CONSORT flow diagram in Fig. 1).

Ethics, consent, permissions

The study was performed in compliance with the Helsinki Declaration for research on humans and was approved by the Regional Committees for Medical and Health Research Ethics (REC North, Reference number 2011/1937).

All participants participated and consented according to the regulations in the research project, also with written parental consent for those under age 16 (REC North, Reference number 2011/1937). Consent to publish was given from every participant and parents when warranted. The study adheres to CONSORT guidelines.

Measures

Diagnostic instruments for study entry

Development and Well Being Assessment (DAWBA) [42], Norwegian language version [43], was a part of intake procedures for all patients at the participating clinics. DAWBA is a web-based diagnostic interview that is multi-informant with both open- and closed-ended questions. In this study, only information from the patients was used to assess diagnosis. When completed online, DAWBA uses computer algorithms to suggest the likelihood of diagnoses. DAWBA covers diagnoses in band levels corresponding to the prevalence of the disorder [44]. The bands range from levels 0–5 and are dichotomously combined to either 'absent' (levels 0–3; < 0.1 to 15% probability of disorder) or 'present' (levels 4–5; ~ 50 to > 70% probability of disorder). Goodman and colleagues [42] found that DAWBA could discriminate between community and clinic samples of youth. Goodman et al. [44] found that the DAWBA bands were well suited to find an approximate prevalence of disorders. Comparing the computer-generated DAWBA bands to clinician-rated diagnoses, Goodman et al. [44] found that DAWBA underestimates the actual prevalence on a group level. Agreement on an individual level showed kappa values that were usually between 0.4–0.7, sensitivity 0.4–0.8, specificity 0.98–0.99, positive predictive values 0.5–0.8 and negative predictive values 0.96–0.99.

The Strengths and Difficulties Questionnaire (SDQ) [45], Norwegian language version [46]. To measure emotional problems the SDQ was completed as part of the DAWBA package at the times of enrollment and end of therapy. The version used was the self-rated SDQ for 11 to 17-year-olds with five subscales. The emotional problems subscale was the main inclusion criteria and the primary outcome measure for emotional symptoms. SDQ is an emotional and behavioral screening questionnaire,

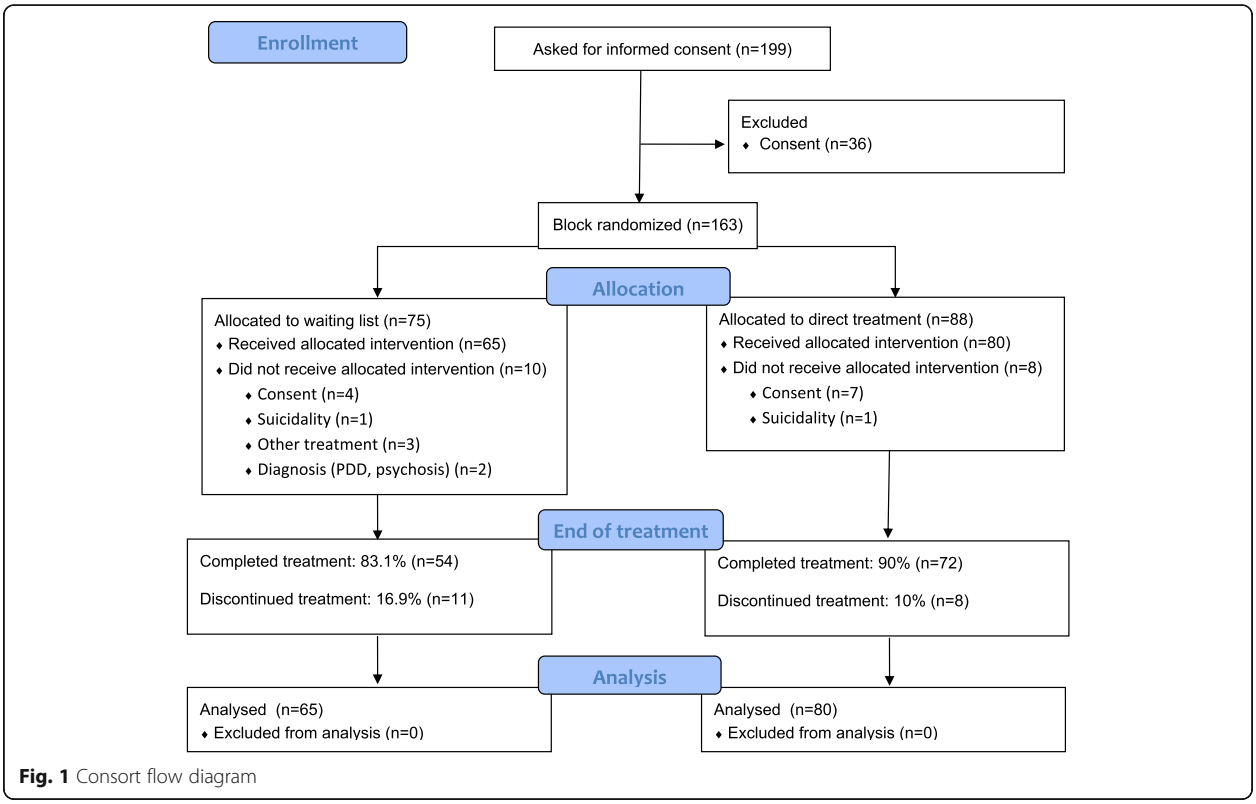

Fig. 1 Consort flow diagram

using a 3-point Likert scale, from 0 (not true) to 1 (somewhat true) to 2 (certainly true), giving a maximum score of 10 on the emotional symptom subscale. Goodman [45] suggested a cutoff on the emotional problem subscale of 6/7. SDQ is a frequently used screening instrument and has satisfactory psychometric properties [45–47]. In this study, we only used the emotional symptoms subscale, which has shown acceptable reliability and adequate internal consistency [46]. Internal consistency in our sample was acceptable (Cronbach's α = .70).

The Children's Global Assessment Scale (CGAS) [48], Norwegian language version [49], was used as a secondary outcome measure for general level of function and was scored at enrollment and end of therapy. CGAS is a therapist-scored rating scale of global functioning ranging from 0 to 100, with higher scores indicating a higher level of function. To ensure the stability of the CGAS scores, each child's clinical profile was scored blindly by a group of at least 3 trained clinicians, and the average score was employed. The clinicians had extensive experience with CGAS, having used it routinely in clinical practice. CGAS has shown good psychometric properties [50]. In the present sample, there was a high degree of reliability between CGAS raters (ICC = .97).

Clinical Outcome in Routine Evaluation-Outcome Measure (CORE-OM) [51], Norwegian language version [52], was used as a secondary outcome measure for general symptom pressure and risk of suicide and self-harm and was distributed at enrollment and end of therapy. CORE-OM is a 34-item questionnaire with items using a 5-point Likert scale from 0 to 4, with higher scores indicating an increased symptom pressure.

Skre et al. [52] suggested a cutoff point of 1 for discriminating between clinical and non-clinical populations. CORE has a reader interface age of 14 years in Norwegian adolescents [52] and can thus be applied in this sample. CORE has shown good psychometric properties [52, 53]. A validation study on a Norwegian sample concluded that CORE-OM has the same psychometric properties as the English version [52, 53]. The internal consistency for the CORE-OM total score in the present sample was excellent (Cronbach's α = .92).

Beck Depression Inventory, second edition (BDI-II) [54], Norwegian language version (not yet validated on a Norwegian youth sample), was used as a secondary outcome measure for extent and depth of depression and were distributed at enrollment and end of therapy. BDI-II is a 21-item self-report questionnaire from ages 13 to 80 years. Items are rated on a 4-point Likert scale from 0 to 3, giving a maximum score of 63 [54]. suggest cutoff ranges between 14 and 19 for mild depression, 20–28 for moderate depression and 29–63 for severe depression.

BDI-II has shown good psychometric properties [55–57]. The internal consistency for the BDI-II in the present sample was excellent (Cronbach's α = .91).

Multidimensional Anxiety Scale for Children (MASC) [58], Norwegian language version [59], was used as a secondary outcome measure for degree of anxiety and was distributed at enrollment and end of therapy. MASC is a 39-item self-report questionnaire for children and adolescents between ages 8 and 17 years. Items are grouped in 6 subscales and rated on a 4-point Likert scale from 0 to 3, with higher scores indicating a higher degree of anxiety. In the present study, we used the total score, converted to a t-distribution centered at approximately 50. MASC has shown good psychometric properties [58, 60]. The internal consistency of the MASC total score in the present sample was good (Cronbach's α = .88).

Evaluation of sessions [61] was employed as a measure of treatment integrity, alliance and user satisfaction and was distributed after each session and at the end of therapy. In the SMART manual, there is an evaluation after each session, where the patient rates aspects of each session on topics of specific content and satisfaction with the session on a Likert scale from 1 "very unsatisfied" to 5 "excellent". The running aim is that the therapist adjusts therapy, alliance, content and relevance in collaboration with the patient.

Procedure

Assessments were completed by youth pretreatment, post-waitlist and post-treatment. A block randomization was used in which groups of 5 youths meeting the inclusion criteria were randomized with a 1/1 chance to either waiting list control (WLC) or direct treatment. A random number generator in SPSS was used in the randomization procedure drawing numbers from the Bernoulli distribution. The participants were enrolled and assigned to treatment or WLC by administrative staff so that both the therapist, researchers and the participants were blinded to the allocation process. The mean duration of WLC was 6.8 weeks, and the mean duration of the treatment condition was 10.3 weeks. The most common explanations for prolonged treatment time were summer vacations, therapist sick leave and patients not showing up for treatment.

Treatment

The adolescents were treated with the SMART program [61], a Norwegian version of the GO! program, originally developed and evaluated in Germany [62]. SMART is developed for 14–25 years old adolescents and younger adults. SMART is an 8-week manual-based modularized CBT program, based on well supported methods for treating anxiety and depression with a strong emphasis on cognitive restructuring, exposure and activation. The special features of the program are as follows: definition of individual treatment goals, activation of personal resources, behavioral experiments, information about emotional problems and related coping strategies. The materials are organized in five modules (introduction, depression, anxiety, assertiveness training, and summary, in a total of eight sessions). The modularized organization of the materials allows for the program to be shortened to four or six sessions by selecting modules.

In this study, all modules except the assertiveness module (2 sessions) at the end of the program were employed as a standard brief therapy in the outpatient clinics. The reason for not employing the assertiveness module was that the WLC and the treatment group should initially be of the same duration, and we chose the modules that were targeting depression and anxiety symptoms. Four modules were delivered over 6 sessions, each with a duration of 1.5 h (see Table 1).

Treatment integrity

The therapists had a 2-day training course in the use of the SMART manual. The training consisted of lectures, hands-on training, and role play. When therapists started using the SMART manual, they had bi-weekly supervision based on the Cognitive Therapy Adherence and Competence Scale (CTACS) [63]. The supervision was mainly performed on Skype by the first and the last author who are trained and experienced CBT therapists and supervisors. CTACS is

Table 1 Questions used as measures on treatment integrity

Question	Satsifaction [a]	Adherence[b]
Introduction		
Introductory exercise	4.36 (0.19)	35.3%
Homework	4.07 (0.78)	90.4%
Total	**4.13 (0.74)**	**62.6%**
Module 1 (depression)		
Repetition/Homework	4.16 (0.73)	100%
Convolute exercise	3.35 (1.28)	65.4%
Cognitive distortion	4.12 (0.74)	97.8%
Looking for proof	4.15 (0.76)	94.9%
Information about depression	4.43 (0.60)	98.5%
Attributional error	4.17 (0.72)	95.6%
Total	**4.12 (0.77)**	**92.0%**
Module 2 (anxiety)		
Repetition/Homework	4.15 (0.70)	97.1%
Information about anxiety	4.47 (0.56)	100%
Anxiety circle	4.28 (0.64)	97.8%
Information about anxiety disorders	4.44 (0.58)	96.3%
Experiment with panic disorder	3.92 (1.00)	47.1%
Relaxation exercise	4.31 (0.88)	65.4%
Total	**4.31 (0.54)**	**83.9%**

Note a. Satisfaction ratings (*M* and *SD*), b. Adherence in %, 100 being full adherence

a widely used 21-item scale that measures therapist competence and adherence, and was here a part of the training and supervision of treatment integrity. In advance of each supervision session, the supervisor and therapists scored a video session with the CTACS. The session was then used to discuss the scores. To further strengthen the treatment integrity over time, periodic booster sessions where all therapists gathered for a full day meeting were arranged. The booster sessions were split between therapists presenting their way of working with the materials and sessions with supervision based on video records. As a measure of treatment integrity, scores from the *Evaluation of sessions* questionnaire where patients rated specific elements of each session were calculated (Likert scale from 1 "very unsatisfied" to 5 "excellent"). From this, an adherence score was computed where present versus missing ratings reflected the degree (in percentage) to which each element in the SMART manual was followed in the sessions (see Table 1).

Alliance and user satisfaction

The following items from the Evaluation of sessions questionnaire were used as a measure of therapeutic alliance: "I liked today's session" and "I felt understood by the therapist" (Likert scale from 1 to 5, ranging from negative to positive). Three items from the end of the therapy questionnaire were used as measures of alliance: "Therapist's competence and presentation were", "Therapist's understanding was" and "Therapist's openness was" (Likert scale from 1 to 5, ranging from negative to positive).

As a measure of user satisfaction, two questions from the *Evaluation of therapy* questionnaire: "Overall the course was" (Likert scale from 1 to 6, ranging from negative to positive) and "I would recommend this course to others" (Yes/No).

Setting, therapists and assessors

The study was conducted in three public child and adolescent mental health outpatient clinics, covering both urban and rural parts of northern Norway. Adolescents are usually referred by general practitioners. Teams are multidisciplinary and work with a variety of disorders. Twenty therapists participated (M age = 39.18 years, SD = 10.93, range 24–57, 100% females). The therapists had 6.8 years of clinical experience on average (SD = 8.23, range 0–32 years). Of the 20 therapists, 11 were psychologists, 2 psychology students, 4 pedagogues, 2 social educators and 1 was a public health nurse. Two therapists had a two-year specific education and training in CBT.

Data analysis

Power calculations

Initially, a necessary sample size of 160 patients was calculated based on two considerations: calculation of statistical power and expected attendance rate. The minimum required sample size for the comparison of group averages in two groups (two-tailed t-test with a 0.01 significance level, statistical power 0.80 and expected effect size on difference in mean scores between groups d = 0.60) was calculated to be 58–67 patients in each group [64]. The choice of the estimated effect size was based on the study by Weisz and colleagues [65], who in a summary of existing research on CBT with children and adolescents found an average effect size of 0.67. An expected attendance rate of 80% was based on data from a counselling service for young people in the same geographic area that had a no-show rate of 12–17% [66]. To treat the necessary 126 patients, we thus needed to recruit 160.

The results were reported as an intention-to-treat analysis [67] as suggested in the CONSORT 2010 statement [68]. Missing data for outcome variables were imputed using the multiple imputation (MI) procedure imputing 50 different datasets. MI has been suggested to be the recommended imputation technique when it is reasonable to assume that data are missing at random [69]. Imputations of missing data were based on predictive mean matching using the MICE package for R [70]. Each imputation was selected from a random draw among the 5 observations that were closest to the value predicted by the imputation model. Both demographic and outcome variables were used to predict (other) outcome variables. The linear mixed models procedure analyzed each of the 50 imputed datasets separately, and the results were pooled using standard procedures.

To test for the effects of the SMART treatment, linear mixed model analysis [71] was used. The data are hierarchical with measurement occasions (level 1; pre and post) nested within individuals (level 2). A random intercept was included in the model, but no random slope was included because of only two measurement occasions. A test of the significance of the time by group interaction is then a test of whether the SMART group and the control group change differently from pre- to post-treatment. A separate analysis was done which adjusted for the individual probability of being randomized directly into treatment (the SMART group). This analysis showed no difference from the main analysis.

Effect sizes with 95% confidence intervals were computed as a standardized difference between the group's gain scores (Hedges' g), using the pooled standard deviation of the pre-measurement for the standardization [72].

Pretreatment differences between the groups were tested using linear mixed models, one-way ANOVA, or chi-square tests, depending on the situation. The same methods were used to compare dropouts after pretreatment with non-dropouts on demographic variables and pretreatment outcome variables.

The Reliable Change Index (RCI) [73] was used to assess clinical and significant change on the SDQ.

We used IBM SPSS v24 for all analyses, and .05 was generally set as the significance level.

Results

Treatment integrity and user satisfaction

The adherence scores in Table 1 indicate a general high level of adherence to the manual, with a high completion percentage of the elements in the SMART manual. However, for four elements, the scores indicate that they were used to a lesser degree: "Introductory exercise", "Convolute exercise", "Experiment with panic disorder" and "Relaxation exercise".

Table 2 shows demographics, diagnoses and comorbidity. The majority of patients were female, and there were too few boys in the sample to compare between genders.

A quarter of participants had a probable pure anxiety disorder, one fifth had a probable pure depressive disorder and one third had a probable diagnosis of both anxiety and depression. One fifth had other disorders, and 10 % did not reach the probability level of any diagnosis. Severe depression (ICD-10) was diagnosed in more than half the participants with a diagnosis of depression.

Alliance

The scores on questions of alliance indicate that the patients liked the sessions, felt understood by the therapist both after sessions and at the end of therapy (see Table 3).

Pre-treatment differences between conditions and change from pre to post therapy

Group differences between the two treatment conditions were compared at baseline on the outcome variables. Differences between the groups on all variables at baseline were non-significant (see Table 4).

Self-reported emotional problems (SDQ emotional problems subscale)

There was a significant time by group interaction on the SDQ emotional scale (see Table 4). While the treatment group had a mean decrease of 1.67 points, the wait-list group had a smaller change (0.81 points), and the effect can be classified as medium ($g = 0.65$ $p = .039$).

Table 2 Demographics and diagnoses

	Total		WLC		Direct treatment	
Dawba prediction	*n*	% (of 145)	*n*	% (of 65)	*n*	% (of 80)
Only anxiety	34	23.4	18	27.7	16	20.0
Only depression	29	20.0	12	18.5	17	21.3
Depression and anxiety	46	31.7	18	27.7	28	35.0
Depression and GAD	30	20.7	10	15.4	20	25.0
Depression and Social phobia	27	18.6	10	15.4	17	21.3
Depression and specific phobia	7	4.8	0	0.0	7	8.8
Depression and agoraphobia	9	6.2	4	6.2	5	6.3
Depression and panic disorder	6	4.1	4	6.2	2	2.5
Neither anxiety nor depression	36	24.8	17	26.2	19	23.8
No diagnosis	9	6.2	5	7.7	4	5.0
Depression	78	53.8	24	36.9	54	67.5
Mild (ICD-10)	2	1.4	2	3.1	0	0.0
Moderate (ICD-10)	27	18.6	10	15.4	17	21.3
Severe (ICD-10)	39	26.9	12	18.5	27	33.8
Unknown	10	6.9	0	0.0	10	12.5
Generalized anxiety disorder	36	24.8	14	21.5	22	27.5
Social phobia	45	31.0	20	30.8	25	31.3
Specific phobia	24	16.6	7	10.8	17	21.3
Agoraphobia	14	9.7	7	10.8	7	8.8
OCD	4	2.8	3	4.6	1	1.3
Panic disorder	9	6.2	6	9.2	3	3.8
Total n	145	100	65	100.0	80	100.0

Notes.
Diagnoses in both ICD-10 and DSM-IV (same algorithm or same number of diagnoses)

Table 3 Questions used as measures on alliance

Question	*n*	Average score: M (SD)
End of session questionnaire		
I liked today's session	117	4.60 (0.60)
I felt understood by the therapist	120	4.43 (0.69)
End of therapy questionnaire		
Therapist understanding was	119	4.61 (0.69)
Therapist openness was	120	4.60 (0.76)

Note. Scores on a Likert scale from 1 to 5, where 5 is most satisfied
Scores based on means across all sessions

General functioning (CGAS)

The two conditions showed significantly different pre-post changes on the CGAS (see Table 4). While the treatment group had a mean increase of 8.6 points, the wait-list group had a smaller change (3.9 points), and the effect can be classified as medium ($g = 0.56$, $p = .019$).

Anxiety (MASC)

For the MASC total score, there was a significant time by group interaction, and the treatment condition had a better development (decrease of 7 points) than the wait-list condition (decrease of 3.3 points), and the effect can be classified as small ($g = 0.34$, $p = .035$) (see Table 4).

Depression (BDI-II)

There was no significant group difference on the BDI-II total score between the treatment (decrease of 8.46 points) and the wait-list condition (decrease of 4.96 points), $g = 0.30$, $p = .066$ (see Table 4).

General symptom pressure and risk of suicide and self-harm (CORE-OM)

There was no significant group difference on the CORE-OM total score between the treatment (decrease of 0.44 points) and the wait-list condition (decrease of 0.33 points), $g = 0.19$, $p = .29$.

Reliable change index and clinically significant change

Of 62 patients eligible for this analysis in the treatment group (SMART), 17.7% ($n = 11$) experienced clinical and reliable change (see Table 5). Of 52 patients eligible for this analysis in the waiting-list condition, 5.8% ($n = 3$) experienced clinical and reliable change. Furthermore, 16.1% ($n = 10$) of the patients in the treatment condition experienced either clinical, or reliable change, as compared to 5.8% (n = 3) of the waiting-list patients. No patients in either group showed deterioration.

Discussion

To the best of our knowledge, this is the first RCT testing the effect of a 6-session transdiagnostic CBT treatment in an adolescent sample suffering mainly from combined emotional disorders, receiving the intervention in a CAMHS naturalistic setting. Despite that this brief manualized treatment was delivered over only six sessions, the treatment condition showed statistically significant treatment effects for internalization symptoms, anxiety symptoms and general functioning compared to waiting-list controls. Clinically significant change in emotional problems (SDQ) was observed significantly more frequently in the treatment condition.

Table 4 Summary results for main outcome variables in each treatment condition

	SMART		Wait list		Group effect[a]	Effect size
	n	mean (sd)	*n*	mean (sd)		g (95% CI)
SDQ emotion (youth)						
Pre	80	7.89 (1.45)	65	7.99 (1.19)	t = −0.30, n.s.	
Post	80	6.22 (2.65)	65	7.18 (2.00)	t = 2.06, p = .039	0.65 (0.31, 0.98)
CGAS						
Pre	80	52.08 (8.97)	65	49.58 (7.69)	t = 1.35, n.s.	
Post	80	60.68 (12.33)	65	53.48 (11.81)	t = 2.35, p = .019	0.56 (0.23, 0.89)
MASC total						
Pre	80	61.10 (11.57)	65	62.72 (9.76)	t = −0.84, n.s.	
Post	80	54.09 (13.26)	65	59.42 (10.99)	t = 2.10, p = .035	0.34 (0.01, 0.67)
BDI						
Pre	80	28.98 (12.48)	65	29.19 (10.91)	t = −0.10, n.s.	
Post	80	20.52 (14.37)	65	24.23 (10.84)	t = −1.84, p = .066	0.30 (−0.03, 0.63)
CORE total						
Pre	80	1.93 (0.65)	65	2.04 (0.47)	t = −1.07, n.s.	
Post	80	1.49 (0.68)	65	1.71 (0.62)	t = −1.06, $p = .289$	0.19 (−0.13, 0.52)

Note a. At pretreatment, main effect of group (SMART/Wait-list); At post-treatment, group interaction effect (time * group)

Table 5 Summary of reliable change and clinical change

	SMART[a, b, c]		Wait list	
	n	%	*n*	%
Clinical and reliable change	11	17.7%	3	5.8%
Only clinical change	9	14.5%	3	5.8%
Only reliable change	1	1.6%	0	0%
No change	41	66.1%	46	88.5%
Deterioration	0	0%	0	0%
Total	62		52	

a. *Group difference on clinical change* vs. *non-clinical change: Fisher's exact test p = 0.013. b. Group difference on no change* vs. *clinical change or reliable change: Fisher's exact test p = 0.007. c. Group difference on clinical and reliable change* vs. *not both clinical and reliable change: Fisher's exact test p = 0.084*
Note
Reliable change = 4 point improvement on the SDQ emotional scale
Clinical change = from above to below cutoff
Clinical and reliable change = both
Deterioration = 4 point increase on the SDQ emotional scale

The effect sizes of the intervention on the measures of emotional problems, general functioning and anxiety symptoms were moderate to small. Furthermore, both depressive symptoms and general clinical outcome changed in the desired direction, albeit showing no statistically significant differences between the treatment and wait-list conditions. Finally, the users report of satisfaction and alliance indicate that the treatment was well received for the adolescents.

The recovery rates in this study were lower than what is expected for CBT efficacy trials for single-disorder treatments [20] where the majority of studies show moderate to large effects targeting anxiety and depression [21, 22]. However, with SDQ emotion (g = 0.65) and MASC (g = 0.34) for anxiety this trial shows comparable effects to [20] and lays slightly above the effects shown by Weisz and colleagues [34] in real-world settings. Compared to effectiveness studies and the effect sizes in the context of studies performed in ordinary clinical practice [34], the present effect sizes are promising, keeping in mind the brief duration of the intervention. Considering the rate of change on the SDQ and CGAS, compared to the total of patients receiving treatment in the participating CAMHS after six month of treatment [36], our sample shows comparable rates of change after six session of treatment, 10 weeks with a decrease of 2.6 points on the SDQ total score and an increase of 9.6 points on the CGAS.

Using the RCI [73] as a measure of clinical significant change, nearly one-sixth of the youths receiving treatment in the present study obtained clinical meaningful change, none deteriorated, while nearly two-thirds showed no reliable change. Although the proportion who experienced partial or full recovery may seem modest, there are indications that the majority of children and adolescents in regular community mental health care do not experience clinical improvement applying these conservative criteria [74, 75]. In addition, the adolescents in the treatment condition improved nearly one category on the CGAS indicating clinically meaningful change after a brief intervention. Some studies from ordinary care report poorer outcome following treatment [76, 77]. In the CORC report mentioned earlier 27% of the CAMHS patients showed deterioration [36]. No patients in either group in this study showed worsening.

The treatment program and the inclusion of both therapists and patients have shown good feasibility and transportability to ordinary clinical practice. Firstly, the therapists in this study were representative of clinical practice with their diverse educations and years of occupational experience, where most of them had limited experience with CBT beforehand. Secondly, the treatment is of short duration and adherence to the manual components was satisfying. In addition, the patients rated the alliance and their satisfaction with the program as good.

Keeping these characteristics in mind, the SMART program could be considered as a first step in a clinical stepped care delivery followed by more intensive evidence-based treatments for single disorders, e.g., the C.A.T. program [78] for anxiety and more intensive programs for depression.

Strengths and limitations

The strengths of the present study were that it was performed in an ordinary clinical setting, by therapist's representative for regular CAMHS practice, with regular referred patients. Furthermore, the sample size is fairly large, and the study has good statistical power to detect moderate to large effects. The data quality is good, with a nearly complete data-set from pre- to post-therapy and measures for treatment integrity. The study included a number of outcome measures, including adherence and acceptability.

A possible limitation of the study was that the mean duration for treatment condition was 3.5 weeks longer than for the WLC, which could be in favor of the intervention. Another possible limitation regarding representativeness is that the governmental restrictions on waiting time prevented us to enroll a part of referred patients in the study. However, in the mentioned report from the Child and Outcomes Research Consortium (CORC) concerning the total of patients receiving treatment in one of our participating CAMHS, from 2013 to 2016 [36], the SDQ total score and the CGAS score before treatment was lower in the total population of CAMHS than in our sample [36].These scores indicate that more acute and serious disorders was not underrepresented in our sample. With this in mind, the results should be interpreted with caution. Regarding representativity, there were too few boys ($n = 14$) to perform analysis of gender differences. Recent studies show evidence

of partial gender non-equivalence with a tendency for girls to more often endorse items measuring symptoms of emotional problems and prosocial behavior on the SDQ [46]. This could have implications for the SDQ as a screening instrument and could have affected the recruitment of boys in the study, indicating a lower cut-off for inclusion of boys. Another limitation is the use of a waiting list as a control condition as opposed to an active control condition. However, the intention was that the two experimental conditions have equivalent length. To create equal length between the conditions the assertiveness module in SMART was removed. This was warranted to accommodate the ethical and legal issues concerning that the patients could not wait more than 6 weeks, however this conflates time and treatment and limits the ability to ascertain the comparative effect of full-scale SMART treatment. SMART is both transdiagnostic and modularized, however in this study SMART was delivered in a linear order to all patients. Some results show that delivering in a modular, flexible fashion gives better results [34]. Although SMART shows promising results, the effectiveness should be qualified as preliminary, requiring future evaluation of the full-scale program in a modularized format to assess "active ingredients" as well as predictors of treatment response and assessment of long-term effects. The treatment should also be compared to an active control group.

Conclusions

Results from this RCT are promising and indicate support for the effectiveness of a transdiagnostic short-term CBT compared to no intervention for youth with emotional problems in community clinics with only 6 sessions of treatment. The recovery rates highlight the need for further improvement for some of the patients. Considering that the treatment is very short, only two sessions of CBT for depression and two for anxiety transdiagnostic treatment, SMART can with further investigations be considered as first step in a stepped care model of treatment of anxiety and/or depression in CAMHS. Further evaluation is needed of the full-scale program and to find the optimal combination with additional interventions.

Abbreviations

BDI II: Beck Depression Inventory, second edition; CAMHS: Child and Adolescent Mental Health Services; CBT: Cognitive Behavior Therapy; CGAS: The Children's Global Assessment Scale; CORC: Child and Outcomes Research Consortium; CORE-OM: Clinical Outcome in Routine Evaluation-Outcome Measure; DAWBA: Development and Well Being Assessment; EBT: Evidence Based Treatment; ES: Effect Sizes; IPT: Interpersonal Therapy; MASC: Multidimensional Anxiety Scale for Children; PDD: Pervasive Developmental Disorder; RCI: Reliable Change Index; RCT: Randomized Controlled Trial; SDQ: Strength and Difficulties Questionnaire; SMART: Structured Material for Therapy; TAU: Treatment as Usual; UP-A: The Unified Protocol for the Treatment of Emotional Disorders in Adolescents; WLC: Waiting List Control

Acknowledgements

We want to express our deep gratitude to the adolescents and parents who, through their participation, have contributed to more knowledge supporting the treatment of anxiety and depression. Finally, we are grateful to and would like to thank the clinicians and administrative staff management at the participating community mental health clinics for all the hard work they have put into the project.

Authors contribution

VL wrote the manuscript, were active in designing the study, collecting and analyzing the data. KF participated in collecting, analyzing and writing of the method and result section. BHH conducted the analysis and contributed in the writing of the method, results and discussion section. IS was the project leader and active in designing the study, and in writing all parts of the manuscript. SPN were active in designing, and writing of all parts of the manuscript. All authors revised the manuscript. All authors approved the manuscript.

Funding

The study was financed from the Northern Norway Regional Health Authority, through project number: PFP1110–13, PFP1112–13 and HNF1383–17, The pilot project received financial support from The Regional Centre for Child and Youth Mental Health and Child Welfare-North. The funding body had no role in designing the study or collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

The datasets generated and analysed during the current study and the full study protocol are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was performed in compliance with the Helsinki Declaration for research on humans and was approved by the Regional Committees for Medical and Health Research Ethics (REC North). Reference number 2011/1937. All participants participated and consented according to the regulations in the research project, also with written parental consent for those under age 16 (REC North, Reference number 2011/1937).

Consent for publication

Not applicable.

Competing interests

The last author received royalties from the publisher of the manual.

Author details

[1]Department of Psychology, Faculty of Health Sciences, UIT The Arctic University of Norway, 9037 Tromsø, Norway. [2]Department of Child and Adolescent Psychiatry, Divisions of Child and Adolescent Health, University Hospital of North Norway, P.O. Box 19, 9038 Tromsø, Norway. [3]Regional Centre for Child and Youth Mental Health and Child Welfare, UIT The Arctic University of Norway, 9037 Tromsø, Norway. [4]Department of General Psychiatry, University Hospital of North Norway, P.O. Box 6124, 9291 Tromsø, Norway. [5]Regional Centre for Child and Adolescent Mental Health - Eastern and Southern Norway, 0484 Oslo, Norway.

Received: 30 September 2019 Accepted: 3 March 2020
Published online: 14 March 2020

References

1. Chavira DA, Stein MB, Bailey K, Stein MT. Child anxiety in primary care: prevalent but untreated. Depress Anxiety. 2004;20(4):155–64.
2. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry. 2003;60(8):837–44.
3. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10–24 years: a systematic analysis. Lancet. 2011;377(9783):2093–102.
4. Essau CA, Gabbidon J. Epidemiology, Comorbidity and Mental Health Services Utilization. In: The Wiley-Blackwell Handbook of The Treatment of Childhood and Adolescent Anxiety. Chichester: Wiley; 2012. p. 23–42.

5. Cheung AH, Dewa CS. Canadian community health survey: major depressive disorder and Suicidality in adolescents. Health Policy. 2006;2(2):76–89.
6. Kessler RC, Walters EE. Epidemiology of DSM-III-R major depression and minor depression among adolescents and young adults in the national comorbidity survey. Depress Anxiety. 1998;7(1):3–14.
7. Naicker K, Galambos NL, Zeng Y, Senthilselvan A, Colman I. Social, demographic, and health outcomes in the 10 years following adolescent depression. J Adolesc Health. 2013;52(5):533–8.
8. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: emerging gender differences in a 10-year longitudinal study. J Abnorm Psychol. 1998;107(1):128–40.
9. Copeland WE, Angold A, Shanahan L, Costello EJ. Longitudinal patterns of anxiety from childhood to adulthood: the Great Smoky Mountains study. J Am Acad Child Adolesc Psychiatry. 2014 Jan 1;53(1):21–33.
10. Mendlowicz MV, Stein MB. Quality of life in individuals with anxiety disorders. Am J Psychiatry. 2000;157(5):669–82.
11. Jaycox LH, Stein BD, Paddock S, Miles JNV, Chandra A, Meredith LS, et al. Impact of teen depression on academic, social, and physical functioning. Pediatrics. 2009;124(4):596–605.
12. Wittchen H-U, Nelson CB, Lachner G. Prevalence of mental disorders and psychosocial impairments in adolescents and young adults. Psychol Med. 1998;28(1):109–26.
13. Van Ameringen M, Mancini C, Farvolden P. The impact of anxiety disorders on educational achievement. J Anxiety Disord. 2003;17(5):561–71.
14. Silverman WK, Field AP. Anxiety Disorders in Children and Adolescents. Cambridge: Cambridge University Press; 2011. p. 447.
15. Merikangas KR, Knight E. The epidemiology of depression in adolescents. In: Handbook of depression in adolescents. New York, NY, US: Routledge/Taylor & Francis Group; 2009. p. 53–74.
16. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. Psychol Bull. 2013;140(3):816.
17. Weersing VR, Jeffreys M, Do M-CT, Schwartz KTG, Bolano C. Evidence base update of psychosocial treatments for child and adolescent depression. J Clin Child Adolesc Psychol. 2017;46(1):11–43.
18. Silverman WK, Pina AA, Viswesvaran C. Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. J Clin Child Adolesc Psychol. 2008;37(1):105–30.
19. Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. J Clin Child Adolesc Psychol. 2016;45(2):91–113.
20. Weisz JR, Kuppens S, Ng MY, Eckshtain D, Ugueto AM, Vaughn-Coaxum R, et al. What five decades of research tells us about the effects of youth psychological therapy: a multilevel meta-analysis and implications for science and practice. Am Psychol. 2017;72(2):79–117.
21. Kendall PC, Peterman JS. CBT for adolescents with anxiety: mature yet still developing. Am J Psychiatry. 2015;172(6):519–30.
22. James A, James G, Cowdrey F, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev. 2013;6.
23. Beidel DC, Turner SM, Morris TL. Behavioral treatment of childhood social phobia. J Consult Clin Psychol. 2000;68(6):1072–80.
24. Pincus DB, Ehrenreich JT, Mattis SG. Mastery of anxiety and panic for adolescents: riding the wave - therapist guide [internet]. Cary, UNITED STATES: Oxford University Press USA - OSO; 2008. Available from: http://ebookcentral.proquest.com/lib/tromsoub-ebooks/detail.action?docID=415524. [cited 4 2019 Sep].
25. Kendall PC, Hedtke KA. Cognitive-behavioral therapy for anxious children: therapist manual. Ardmore: Workbook Publishing; 2006.
26. Ehrenreich-May J, Kennedy SM, Sherman JA, Bilek EL, Buzzella BA, Bennett SM, et al. Unified protocols for Transdiagnostic treatment of emotional disorders in children and adolescents: therapist guide. New York: Oxford University Press; 2017. p. 482.
27. Berman SL, Weems CF, Silverman WK, Kurtines WM. Predictors of outcome in exposure-based cognitive and behavioral treatments for phobic and anxiety disorders in children. Behav Ther. 2000;31(4):713–31.
28. Ginsburg GS, Kendall PC, Sakolsky D, Compton SN, Piacentini J, Albano AM, et al. Remission after acute treatment in children and adolescents with anxiety disorders: findings from the CAMS. J Consult Clin Psychol. 2011;79(6):806–13.
29. Rapee RM. The influence of comorbidity on treatment outcome for children and adolescents with anxiety disorders. Behav Res Ther. 2003;41(1):105–12.
30. Craske MG, Farchione TJ, Allen LB, Barrios V, Stoyanova M, Rose R. Cognitive behavioral therapy for panic disorder and comorbidity: more of the same or less of more? Behav Res Ther. 2007;45(6):1095–109.
31. Bearman SK, Weisz JR. Review: comprehensive treatments for youth comorbidity – evidence-guided approaches to a complicated problem. Child Adolesc Ment Health. 2015;20(3):131–41.
32. Weisz JR, Krumholz LS, Santucci L, Thomassin K, Ng MY. Shrinking the gap between research and practice: tailoring and testing youth psychotherapies in clinical care contexts. Annu Rev Clin Psychol. 2015;11(1):139–63.
33. Ehrenreich-May J, Chu BC. Overview of transdiagnostic mechanisms and treatments for youth psychopathology. Transdiagnostic Treat Child Adolesc Princ Pract. 2013:3–14.
34. Weisz JR, Kuppens S, Eckshtain D, Ugueto AM, Hawley KM, Jensen-Doss A. Performance of evidence-based youth psychotherapies compared with usual clinical care: a multilevel Meta-analysis. JAMA Psychiatry. 2013;70(7):750.
35. Nilsen TS, Handegård BH, Eisemann M, Kvernmo S. Predictors of rate of change for children and youth with emotional disorders: a naturalistic observational study. Child Adolesc Psychiatry Ment Health. 2016;10(1):11.
36. University Hospital of Northern Norway CORC report. London: CORC, Anna Freud National Centre for Children and Families; 2018. Reporting period: 01 December 2013 to 31 December 2016. (Unpublished report).
37. Doss AJ. Evidence-based diagnosis: incorporating diagnostic instruments into clinical practice. J Am Acad Child Adolesc Psychiatry. 2005;44(9):947–52.
38. Mansell W, Harvey A, Watkins ER, Shafran R. Cognitive behavioral processes across psychological disorders: a review of the utility and validity of the Transdiagnostic approach. Int J Cogn Ther. 2008;1(3):181–91.
39. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev. 2015;40:91–110.
40. Newby JM, Twomey C, Yuan Li SS, Andrews G. Transdiagnostic computerised cognitive behavioural therapy for depression and anxiety: a systematic review and meta-analysis. J Affect Disord. 2016;199:30–41.
41. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther. 2017;46(1):1–28.
42. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. J Child Psychol Psychiatry. 2000;41(5):645–55.
43. Brøndbo PH, Mathiassen B, Martinussen M, Handegård BH, Kvernmo S. Agreement on diagnoses of mental health problems between an online clinical assignment and a routine clinical assignment. J Telemed Telecare. 2013 Feb;19(2):113–9.
44. Goodman A, Heiervang E, Collishaw S, Goodman R. The "DAWBA bands" as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples. Soc Psychiatry Psychiatr Epidemiol. 2011 Jun;46(6):521–32.
45. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337–45.
46. Bøe T, Hysing M, Skogen JC, Breivik K. The strengths and difficulties questionnaire (SDQ): factor structure and gender equivalence in Norwegian adolescents. PLoS One. 2016;11(5):e0152202.
47. Muris P, Meesters C, van den Berg F. The strengths and difficulties questionnaire (SDQ). Eur Child Adolesc Psychiatry. 2003;12(1):1–8.
48. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A Children's global assessment scale (CGAS). Arch Gen Psychiatry. 1983;40(11):1228–31.
49. Hanssen-Bauer K, Gowers S, Aalen OO, Bilenberg N, Brann P, Garralda E, et al. Cross-National Reliability of clinician-rated outcome measures in child and adolescent mental health services. Adm Policy Ment Health Ment Health Serv Res. 2007;34(6):513–8.
50. Schorre BEH, Vandvik IH. Global assessment of psychosocial functioning in child and adolescent psychiatry: a review of three unidimensional scales (CGAS, GAF, GAPD). Eur Child Adolesc Psychiatry. 2004;13(5):273–86.
51. Evans GM. John Mellor-Clark, frank Margison, Michael Barkham, Kerry Audin, Janice Connell Chris. CORE: clinical outcomes in routine evaluation. J Ment Health. 2000;9(3):247–55.

52. Skre I, Friborg O, Elgarøy S, Evans C, Myklebust LH, Lillevoll K, et al. The factor structure and psychometric properties of the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) in Norwegian clinical and non-clinical samples. BMC Psychiatry. 2013 Dec;13(1).
53. Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: clinical outcomes in routine evaluation. J Ment Health. 2000;9(3):247–55.
54. Beck AT, Steer RA, Brown GK. Beck depression inventory-II. San Antonio. 1996;78(2):490–8.
55. Aasen H. An Empirical Investigation of Depression Symptoms: Norms, Psychometric Characteristics and Factor Structure of the Beck Depression Inventory-II [Internet]. The University of Bergen; 2001. Available from: https://bora.uib.no/handle/1956/1773 [cited 27 2019 Aug].
56. Osman A, Downs WR, Barrios FX, Kopper BA, Gutierrez PM, Chiros CE. Factor structure and psychometric characteristics of the beck depression inventory-II. J Psychopathol Behav Assess. 1997;19(4):359–76.
57. Steer RA, Kumar G, Ranieri WF, Beck AT. Use of the Beck depression inventory-II with adolescent psychiatric outpatients. J Psychopathol Behav Assess. 1998;20(2):127–37.
58. March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The multidimensional anxiety scale for children (MASC): factor structure, reliability, and validity. J Am Acad Child Adolesc Psychiatry. 1997;36(4):554–65.
59. Martinsen KD, Holen S, Neumer S-P, Waaktaar T, Rasmussen LMP, Kendall PC, et al. The factor structure of MASC youth report in Norwegian school children. J Child Fam Stud. 2017;26(7):1808–16.
60. Rynn MA, Barber JP, Khalid-Khan S, Siqueland L, Dembiski M, McCarthy KS, et al. The psychometric properties of the MASC in a pediatric psychiatric sample. J Anxiety Disord. 2006;20(2):139–57.
61. Neumer SP, Junge-Hoffmeister J. SMART. Forebygging av emosjonelle problemer hos ungdom. Veileder. Oslo: Kommuneforlaget; 2010.
62. Junge J. Primäre Prävention von Angst und Depression bei Jugendlichen: Entwicklung und Evaluation des Programms GO!-" Gesundheit und Optimismus" [PhD Thesis]. Verlag nicht ermittelbar; 2003.
63. Barber JP. Development of the cognitive therapy adherence and competence scale. Psychother Res. 2003;13(2):205–21.
64. Machin D, Campbell M, Fayers PM, Pinol A. Sample size tables for clinical studies. 2nd ed. Wiley; 1997.
65. Weisz JR. Psychotherapy for children and adolescents: evidence-based treatments and case examples: Cambridge University Press; 2004.
66. Wang C, Berge K, Arnesen Y, Breivik C, Johnsen L, Bruun-Hansen T, et al. Psykhjelpen på Tvibit: En psykologisk rådgivningstjeneste på ungdommens kulturhus midt i Tromsø sentrum.[Easily accessible psychological counselling for adolescents: A Norwegian example.-English abstract]. J Norw Psychol Assoc. 2007;44:247–53.
67. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. Br Med J. 1999;319(7211):670–4.
68. Schulz KF, Altman DG, Moher D. The CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med. 2010 Mar 24;8(1):18.
69. Re ACD, Maisel NC, Blodgett JC, Finney JW. Intention-to-treat analyses and missing data approaches in pharmacotherapy trials for alcohol use disorders. BMJ Open. 2013;3(11):e003464.
70. Buuren, S. van, Groothuis-Oudshoorn, K. MICE: Multivariate Imputation by Chained Equations in R. J Stat Softw. 2010;.
71. Singer JD, Willett JB. Applied longitudinal data analysis: modeling change and event occurrence. New York, NY, US: Oxford University Press; 2003. 644 p.
72. Schmidt FL, Hunter J. General mental ability in the world of work: occupational attainment and job performance. J Pers Soc Psychol. 2004; 86(1):162–73.
73. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991; 59(1):12–9.
74. Garland AF, Haine-Schlagel R, Brookman-Frazee L, Baker-Ericzen M, Trask E, Fawley-King K. Improving community-based mental health Care for Children: translating knowledge into action. Adm Policy Ment Health Ment Health Serv Res. 2013;1(40):6–22.
75. Warren JS, Nelson PL, Mondragon SA, Baldwin SA, Burlingame GM. Youth psychotherapy change trajectories and outcomes in usual care: community mental health versus managed care settings. J Consult Clin Psychol. 2010; 78(2):144.
76. Manteuffel B, Stephens RL, Sondheimer DL, Fisher SK. Characteristics, service experiences, and outcomes of transition-aged youth in Systems of Care: programmatic and policy implications. J Behav Health Serv Res. 2008;35(4): 469–87.
77. Warren JS, Nelson PL, Burlingame GM. Identifying youth at risk for treatment failure in outpatient community mental health services. J Child Fam Stud. 2009;18(6):690–701.
78. Kendall P, Choudhury M, Hudson J, Webb A. The CAT project manual: manual for the individual cognitivebehavioral treatment of adolescents with anxiety disorders. Ardmore. PA: Workbook Publishing Inc; 2002.

Publisher's Note

Paper 2

Lorentzen, Veronica; Handegård, Bjørn Helge; Lillevoll, Kjersti; Solem, Kenth; Moen, Connie Malén; Skre, Ingunn. CORE-OM as a routine outcome measure for adolescents with emotional disorders: factor structure and psychometric properties. BMC Psychology 2020; Volume 8 (86). ISSN 2050-7283.s 1 - 14.s doi: 10.1186/s40359-020-00459-5.

Lorentzen et al. BMC Psychology (2020) 8:86
https://doi.org/10.1186/s40359-020-00459-5

BMC Psychology

RESEARCH ARTICLE Open Access

CORE-OM as a routine outcome measure for adolescents with emotional disorders: factor structure and psychometric properties

Veronica Lorentzen[1,2*], Bjørn Helge Handegård[3], Connie Malén Moen[1,4], Kenth Solem[5], Kjersti Lillevoll[1] and Ingunn Skre[1,6]

Abstract

Background: Instruments for monitoring the clinical status of adolescents with emotional problems are needed. The Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) according to theory measures problems/symptoms, well-being, functioning and risk. Documentation of whether the theoretical factor structure for CORE-OM is applicable for adolescents is lacking.

Methods: This study examined the factor structure and psychometric properties of the CORE-OM based on two samples of adolescents (age 14–18): youths seeking treatment for emotional problems ($N = 140$) and high school students ($N = 531$). A split half approach was chosen. An exploratory factor analysis (EFA) was performed on the first half of the stratified samples to establish the suitability of the model. A Confirmatory Factor Analysis (CFA) with the chosen model from the EFA was performed on the second half. Internal consistency and clinical cut-off scores of the CORE-OM were investigated.

Results: The best fitting model only partially confirmed the theoretical model for the CORE-OM. The model consisted of five factors: 1) General problems, 2) risk to self, 3) positive resources 4) risk to others and 5) problems with others. The clinical cut-off score based on the all-item total was higher than in an adult sample. Both the all-item total and general problems cut-off scores showed gender differences.

Conclusion: The factor analysis on CORE-OM for adolescents resulted in a five-factor solution, and opens up for new subscales concerning positive resources and problems with others. A 17-item solution for the general problems/symptoms scale is suggested. We advise developers of self-report instruments not to reverse items, if they do not intend to measure a separate factor, since these seem to affect the dimensionality of the scales. Comparing means for gender in non-clinical samples should not be done without modification of the general emotional problem and the positive resources scales. Slightly elevated CORE-OM scores (up to 1.3) in adolescents may be normal fluctuations.

Keywords: Outcome measure, CORE-OM, Validation, Adolescents, Clinical, Emotional disorders

* Correspondence: veronica.lorentzen@uit.no
[1]Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, P.O. Box 6050 Langnes, N-9037 Tromsø, Norway
[2]Department of Child and Adolescent Mental Health, Divisions of Child and Adolescent Health, University Hospital of North Norway, 9038 Tromsø, Norway
Full list of author information is available at the end of the article

Background

Emotional disorders represent the most prevalent mental health problem in adolescence, and the comorbidity among emotional disorders is high. The onset of emotional problems typically occurs in childhood and adolescence or early adulthood [1]. There is a call for screening tools that can detect mental health problems in adolescents and determine their clinical status. Valid and reliable routine outcome measures are key tools in monitoring treatment effects and for detecting and preventing treatment failure [2]. There is a need for transdiagnostic measures that address comorbidity and are sensitive to change to monitor the treatment of adolescents.

The self-report Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM) is a 34-item questionnaire using a 5-point Likert scale from 0 (not at all) to 4 (most of the time). CORE-OM is widely used measure in outpatient mental health and counselling services and in psychotherapy research with adult patients [3–5]. The CORE-OM items were chosen based on their clinical significance and their sensitivity to change in psychological status [3, 4]. The CORE-OM theoretically covers four dimensions: Well-being (4 items); Functioning (12 items); Problems/symptoms (12 items); and Risk (6 items) [6, 7]. 'Well-being' refers to a patient's sense of life quality and emotional health. 'Problems/symptoms' is associated with psychological health issues such as anxiety and depression symptoms, reactions to trauma, and physical complaints. 'Functioning' relates to interpersonal, social and general functioning in daily life. A high correlation has been found among these three domains [5, 8, 9], and combining them into a general psychological distress scale called All-items-minus Risk has been recommended [10]. 'Risk/harm' includes items covering harm to self and suicidal ideation (risk-to-self items) and violent behaviour and threats towards other people (risk-to-others items) [6, 11]. It has been recommended that risk be monitored separately to help the clinician detect a patient's thoughts and plans regarding self-harm, suicide and violence [10].

The original CORE-OM [5] has been translated into more than 20 languages and has good psychometric properties in adult samples [7, 8, 12–15]. The CORE-OM has been benchmarked in student counselling and primary care service users aged 16 to 24 [16], however, this study lacked a control group of non-service users.

One version of the CORE developed for young people aged 11–16 is the CORE-YP [17]. However, the CORE-YP includes only ten of the CORE-OM items, phrased in simplified wording, and it is not adapted for the whole age span of adolescents (up to age 18) received in Norwegian child and adolescent mental health outpatient services. The CORE-OM, with its 34 items, gives more detailed information when needed; it addresses the most common comorbid symptoms in emotional disorders, is sensitive to change, and exists in a ten-item version (the CORE-10) that can be used session by session. Measures with these qualities are scarce in child and adolescent mental health services. For researchers and clinicians, it is more convenient to use one tool for adolescents, both for longitudinal research and to monitor treatment, without having to change tools due to changing age norms. In Norway youths up to age 18 are referred to child and adolescent mental health services, while individuals aged 18 and older are referred to adult outpatient services.

Since the CORE-OM addresses anxiety, depression, the aftermath of trauma, physical complaints, daily life functioning, subjective well-being, and risk to self and others, it is a highly relevant outcome measure, not only for adults but also for adolescent service users, and is particularly relevant for those with emotional problems. However, there is no existing validation of the full CORE-OM scale in high school age adolescent samples.

There is a need for knowledge about whether the test parameters for the CORE-OM in youths are comparable with those obtained in adult populations.

Previous research on the factor structure of the CORE-OM [9] has not fully confirmed the theoretically derived sub-scales but has rather indicated a structure constituted by a g-factor of psychological distress and residualized latent theoretically derived domains (Symptoms/Problems, Functioning and Well-being). 'Risk-to-self' correlates with the g-factor, while 'risk-to-others' has a poor fit in the structural models [9]. Several factor analytic evaluations of the CORE-OM have been performed. The test developers [5] suggested a first component that explained 38% of the variance, a risk component and a positively worded component. In later factor analyses some of the same researchers [9] suggested a bifactor model with an overall g-factor, a method factor (with positively and negatively keyed items) and risk to self and others. But, their best fitting model included the well-being, psychological problems, and functioning domains, although most of the factor loadings associated with these three domains were small after the g-factor was accounted for. In the Norwegian version [7] a bifactor model was also suggested, but here the method factor did not contribute to improve model fit. The model considered best in that study was a bifactor model with a general distress factor and the four CORE-OM domains. Investigations on the British version suggested using "Mokken Scaling" [10]. This scale is unidimensional resulting in a general distress factor, with items differentiating between more or less severe levels of stress. This approach result in the well-being items giving information about lower levels of stress,

while the risk items covers higher levels of stress, with the other items in between. When the CORE-OM was constructed [5], the majority of items were statements about psychological distress or negative life situations, i.e., psychological distress. Most of the items were thus negatively keyed. However, eight items were positively keyed, allegedly to mitigate response bias [5]. However, mixing negatively and positively keyed items may add method effects that threaten the construct validity of an instrument. Whether the method effects of negatively and positively keyed items are caused by responder error [18], by response bias [19], or because positive and negative utterances actually measure different constructs are, however, unclear. Lyne et al. [9] demonstrated that the positively and negatively keyed items in the CORE-OM formed two separate method factors across the theoretically defined domains. In the validation of the CORE-OM in adult Norwegian samples, these method factors were also observed but were deemed negligible [7].

The CORE-OM has the potential to discriminate between non-clinical and clinical adult populations [5, 7, 8, 12, 13, 15]. The recommended cut-off score for the CORE-OM in the adult population is 1.0 [20].

Gender effects have been found for the All-items score, with women generally scoring higher than men, and consequently, gender-specific cut-off scores have been recommended by some authors [5, 13]. Other validations of the CORE-OM have not recommended separate cut-off scores based on gender, suggesting that gender effects were small and negligible [7, 12].

CORE-OM and Beck Depression Inventory (BDI) scores correlate in clinical samples [21]. Adolescents tend to have higher scores than adults on the BDI [22, 23]. Findings suggest that younger respondents generally score higher on the CORE-OM than older respondents [5, 7, 8, 14] and motivate the present authors to ask whether separate cut-off scores for adolescents are needed, as has been found for the BDI [22, 23].

Since no study has evaluated the factor structure of CORE-OM and the reliability of its factors in an adolescent population, there is a need for such studies. Since several studies have shown that girls score higher than boys do on emotional problems [24, 25] it is interesting to evaluate whether this also is observed when using the CORE-OM. To be able to do such a comparison, a requirement is that the factor structure for boys and girls are similar. In the factor analytic framework this can be done evaluating measurement invariance of the factors, and mean comparisons on gender require at least partial scalar invariance [26].

The aims of this paper were to study the psychometric properties of the Norwegian version of the CORE-OM in an adolescent clinical sample selected for emotional problems and a non-clinical sample by

(I) Investigating the factor structure of the Norwegian CORE-OM in adolescents aged 14 to 18 years, by establishing the suitability of the model by Exploratory Factor Analysis (EFA), and performing Confirmatory Factor Analysis (CFA) with the chosen model of the EFA.
(II) Evaluating the internal consistency on the scales from the chosen model.
(III) Performing a measurement invariance analysis for gender based on the factor structure suggested by the factor analysis.
(IV) Comparing factor means for boys and girls in the non-clinical sample, if at least partial scalar invariance under (III) is achieved.
(V) Calculating clinical cut-off scores.

Methods

Design

A between-subjects cross-sectional survey study was used to examine the psychometric properties of the CORE-OM in samples of Norwegian adolescents aged 14–18 years.

Samples

Data were gathered from two separate samples: a non-clinical and a clinical sample. The non-clinical sample ($n = 531$) was recruited for the purpose of this paper from four junior high schools and four senior high schools in both urban and rural areas in North Norway. The schools were randomized and drawn, and data were collected until at least 65 participants from each class grade were included. In the non-clinical sample (age 14–18, M = 15.91, SD = 1.45), there were 273 (51.4%) boys and 258 (48.6%) girls. The response rate was between 71.4 and 83.3%. Although the sampling procedure was systematic, the non-clinical sample should be viewed as a convenience sample.

The clinical sample consisted of patients ($n = 140$) recruited at CAMHS (Child and Adolescent Mental Health Services) located in two North Norwegian towns and one community centre. In the CAMHS sample (age 14–17, M = 15.72, SD = 1.15), there were 13 boys (9.3%) and 127 girls (90.7%). The adolescents in this sample were enrolled as participants in a psychotherapy research project, the SMART study [27] ('Evaluation of short-term treatment for adolescents with emotional disorders in five children and adolescent CAMHS—A randomized controlled trial' (ClinicalTrials.gov Identifier: NCT02150265); Regional Committee for Medical and Health Research Ethics (REC North); Reference number 2011/1937). Data for the present study were collected at enrolment, before treatment.

Inclusion/exclusion criteria

The inclusion criterion for the non-clinical sample was being a student at a junior or senior high school, while

the exclusion criterion was that the adolescent could not read or write Norwegian fluently. The inclusion criteria in the clinical sample were (1) age between 14 and 17 years; (2) a probable diagnosis of an emotional disorder as indicated by a score of at least 6 on the Strengths and Difficulties Questionnaire (SDQ) screening tool; and (3) maintenance of a maximum waiting time for necessary medical care of 6 weeks given by Norwegian health authorities. The exclusion criteria were (1) a diagnosis of pervasive developmental disorder (PDD); (2) psychotic symptoms; (3) anxiolytic or anti-depressant medication effects during the treatment period; and (4) inability to speak the Norwegian language.

Ethics and consent

The study was performed in compliance with the Helsinki Declaration for research on humans and was approved by the REC North (Reference number 2011/1937).

All participants participated and consented according to the regulations governing the research project, with written parental consent provided for those under age 16 (REC North, Reference number 2011/1937). High scores on symptom and risk items were addressed by the responsible therapist or counsellor.

CORE-OM scoring

All participants completed the CORE-OM in Norwegian translation [7] on paper with a pen or pencil. All participants provided information regarding age and gender. The scoring procedure for both samples followed the guidelines for scoring from Barkham et al. [3].

Statistical analyses

Descriptive data for each of the 34 items, separately for the clinical and non-clinical sample, have been provided as additional material.

The following procedure in evaluating the factor structure of the CORE-OM was performed: The sample was randomly split into two equally sized halves, stratified on sample (clinical/non-clinical) giving an equal proportion of cases from the clinical and non-clinical samples in each half. One of the sample halves was a training sample where an exploratory factor analysis (EFA) was done, and the other half was a testing sample where the chosen model from the EFA was tested using confirmatory factor analysis.

In the EFA, different solutions with varying number of factors was evaluated. Model fit information for each model, difference in model fit between subsequent factor solutions, and the meaningfulness of the Geomin rotated factor loadings was used to guide the choice of the number of factors.

The CFA was carried out on the chosen model from the EFA. The WLSMV estimator and delta parameterization was used [26]. Model fit for the CFA was evaluated by the Chi-square test, the chi-square to degrees of freedom ratio, the RMSEA [28] and the CFI [29]. A significant Chi-square test indicates significant model misfit, but the chi-square test is both a function of sample size and the amount of misfit so we rely mostly on the RMSEA and CFI for model fit evaluation. Models with a chi-square/df ratio < 2 [30]. RMSEA below 0.05 and a CFI above 0.95 are typically considered to be well-fitting [31].

With the WLSMV estimator and using the default missing data handling method in Mplus, pairwise deletion is used to handle the missing observations. This means that a pair of observations is used in computing a polychoric correlation if both observations in the pair are observed. Overall, the covariance coverage percentage of data for the CFA part of the sample was between 96.1 and 100%, and with the highest proportion of missing observations for items 19 and 20 where 2.4 and 2.1%, respectively, were missing.

Outliers was evaluated by Cook's d in the CFA analysis. The Cook's d computes, for each subject, the overall influence that the subject has on the parameter estimates estimated in the analysis. Additional analyses without the individuals with the highest Cook's d values was performed, and there was noted a very small improvement in the fit indices in the CFA analysis when those cases were removed. Since the overall results and conclusions were not affected to a large degree by the outliers, our results were based on the whole sample.

Reliability was evaluated by computing McDonald's Omega [32–34] for ordinal items, reporting a 95% confidence interval for the reliability parameter [35]. Omegas if item is deleted with 95% confidence intervals are also reported [36] We computed item to total (using the rest of the items) correlations by a procedure shown in Raykov & Marcoulides [37] computing polyserial correlations.

Measurement invariance for gender is necessary for gender mean comparisons on CORE-OM scales. Measurement invariance was evaluated for a configural and a scalar model for the non-clinical sample. The low number of clinical boys in our study made it impossible to assess measurement invariance for gender using the clinical sample. With ordinal indicators Muthen & Muthen [26] recommends that factor loadings and thresholds as a unity, so metric invariance holding only factor loadings invariant between genders were not carried out. We compared the difference in model fit between the configural model and the scalar model was tested using the DIFFTEST option in Mplus. Modification indices were assessed for partial scalar invariance if full scalar invariance was not achieved through the DIFFTEST. Even

though the use of modification indices are controversial, it is often considered, and Muthen & Muthen [26] recommends that equality constraints for factor loadings and thresholds are relaxed in tandem. Following their recommendations, scale factors for items with freely estimated loadings and thresholds were fixed at one for identification purposes.

Mean gender differences on latent factors were evaluated in the final partial scalar model.

All factor analyses were performed in Mplus (version 7.4). Omega coefficients were computed within R using a procedure described by Peters [33].

The acceptability of the data was assessed by analysing missing data. Chi-squared tests were conducted to explore the relationship between the missing items and groups and missing items and gender. ANOVA was conducted to determine whether there was a significant relationship between the missing items and age.

Jacobson and Truax's [24] formula was used to calculate the cut-off score for discriminating between a clinical and non-clinical sample:

$$\frac{\text{mean}_{\text{clin}}\text{sd}_{\text{norm}} + \text{mean}_{\text{norm}}\text{sd}_{\text{clin}}}{\text{sd}_{\text{clin}} + \text{sd}_{\text{clin}}}$$

Results

Acceptability

A maximum omission of 10% of the items was used as an indication of acceptability for the CORE-OM. Nine participants (1.7%) with an omission rate greater than 10% were removed from the dataset: three from the junior high school sample and six from the senior high school sample.

Exploratory factor analysis on the CORE-OM

EFAs with from one to six factors was performed. Factor solutions with one and two factors did not show good model fit. Three factors gave significantly better model fit than two factors, four factors gave significantly better model fit than three, five had significantly better model fit than four, and six was better than five according to a Chi-square difference test (see Table 1 for model fit for the 1 to 6-factor solutions). The factor loadings for the five-factor solution are shown in Table 2.

The five-factor solution was chosen, since the six-factor solution did not seem to add anything of substance. In Table 2 the factor loadings above 0.30 for each item are shown in bold for the five-factor solution. Factor 1 is interpreted as a general problem factor. All items for this factor are negatively keyed, and is a mix of what the manual describes as symptoms of anxiety, depression, physical problems, trauma, functioning and subjective well-being. Factor 2 consists of four risk-to-self items, i.e. four of the six "risk" items load highest on this factor. In addition, items 23, 27 and 30 loaded nearly as high on this factor as on Factor 1. These items points to symptoms of depression. One option is to remove such cross-loading items, but it seems important to include questions about unhappiness and despair in a questionnaire measuring symptoms of depression or general emotional problems. The level of risk of self-harm and general emotional problems are likely correlated, and it is therefore natural to observe some cross-loadings for these items.

The eight items loading highest on the third factor were all positively keyed, so this latent variable can be interpreted as positive resources that the adolescent possesses.

Factor 4 has two high-loading items that are interpreted as risk-to-others items.

For the fifth factor, items 25, 26 and 33 are functioning items related to functioning/relations with other people. Item 31 and item 3 had a cross-loading on this factor that was nearly as high as the loading on the general factor. Both these items are related to relations to other people, but the positive framing of item 3 give a higher loading on the positive resources latent variable, while item 31 (irritability with others) can both be a result of problems with others but also something that is related to general problems.

This factor structure gave the most interpretable factors in our opinion. Even though we got a significantly better fit with six factors compared to five, the sixth factor loads highest only on item 8. This item has to do with physical problems or pain and is possibly related to somatic symptoms, which is not asked for in detail in CORE-OM. The six-factor solution is not shown here, but with very few exceptions, the items were loading on the same factors as in the five-factor solution.

The two-factor and three-factor solutions gave a general factor plus additional factors that were hard to interpret. The four-factor solution had a general factor, Risk to others, Positive items and Problems with others as possible interpretations, and the difference between this and the five-factor solution was mainly that the risk to self-items loaded on the general problems factor.

Confirmatory factor analysis

A CFA was performed on the five-factor model that was selected from the EFA, using the other half of the sample.

In this analysis, all factor loadings were significantly different from zero (see Fig. 1). Item 19 showed a loading that indicated a low association to the Positive resources latent variable.

Model fit information for this analysis showed a significant $\chi^2(517) = 956.7$ ($p < .0005$), and $\chi^2/\text{df} = 1.85$;

Table 1 Model fit information EFA

# of factors	Chi-square	df	RMSEA	CFI	SRMR	Difference X^2(df)
1	1544.3*	527	0.076	0.942	0.078	–
2	1124.1*	494	0.062	0.964	0.063	2 vs. 1: $X^2(33) = 328.8$*
3	817.8*	462	0.048	0.980	0.048	3 vs. 2: $X^2(32) = 257.0$*
4	686.2*	431	0.042	0.985	0.040	4 vs. 3: $X^2(31) = 125.6$*
5	562.3*	401	0.035	0.991	0.032	5 vs. 4: $X^2(30) = 112.9$*
6	482.8*	372	0.030	0.994	0.027	6 vs. 5: $X^2(29) = 83.1$*

* $p < .0005$

Table 2 Geomin rotated factor loadings for the five-factor solution (loadings above 0.30 are shown)

Item	1	2	Factor 3	4	5
1 I have felt terribly alone and isolated	**0.53**				
2 I have felt tense, anxious or nervous	**0.83**				
5 I have felt totally lacking in energy and enthusiasm	**0.49**				
8 I have been troubled by aches, pains or other physical problems	**0.79**			0.36	
10 Talking to people has felt too much for me	**0.58**				
11 Tension and anxiety have prevented me doing important things	**0.69**				
13 I have been disturbed by unwanted thoughts and feelings	**0.70**				
14 I have felt like crying	**0.66**				
15 I have felt panic or terror	**0.68**				
17 I have felt overwhelmed by my problems[a]	**0.67**				
18 I have difficulty getting to sleep or staying asleep	**0.75**				
20 My problems have been impossible to put to one side	**0.57**				
23 I have felt despairing or hopeless	**0.40**	0.35			
27 I have felt unhappy	**0.42**	0.42			
28 Unwanted images or memories have been distressing me	**0.46**				
29 I have been irritable when with other people	**0.35**				0.32
30 I have thought I am to blame for my problems and difficulties	**0.35**	0.33			
9 I have thought of hurting myself		**0.74**			
16 I made plans to end my life		**0.76**			
24 I have thought it would be better if I were dead		**0.85**			
34 I have hurt myself physically or taken dangerous risks with my health		**0.67**			
3 I have felt I have someone to turn to for support when needed			**0.51**		0.31
4 I have felt O.K. about myself			**0.51**		
7 I have felt able to cope when things go wrong			**0.51**		
12 I have been happy with the things I have done			**0.60**		
19 I have felt warmth and affection for someone	0.32		**0.67**		
21 I have been able to do most things I needed to			**0.58**		
31 I have felt optimistic about my future[a]			**0.71**		
32 I have achieved the things I wanted to			**0.65**		
6 I have been physically violent to others				**0.69**	
22 I have threatened or intimidated another person				**0.78**	
25 I have felt criticised by other people					**0.67**
26 I have thought I have no friends					**0.45**
33 I have felt humiliated or shamed by other people					**0.87**

[a] Items 17 and 31 have switched place in the English and Norwegian version of the CORE-OM

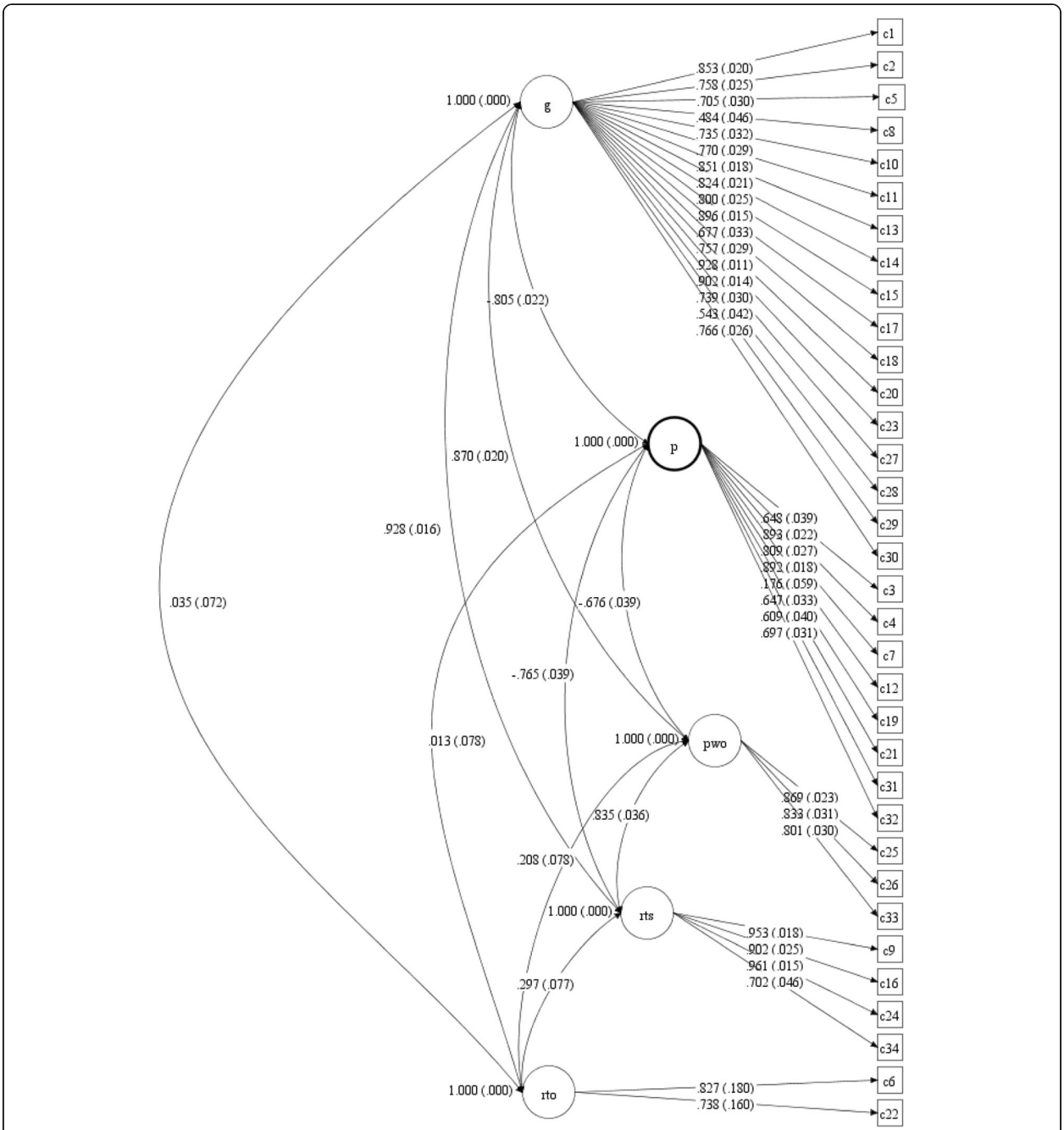

Fig. 1 Confirmatory factor analysis of the CORE-OM. Five-factor standardized solution with standard errors. Factor correlations (with SE) are shown next to curved arrows. Latent measurement errors for each item are not shown in the figure. g = General problems; p = Positive resources; pwo = Problems with others; rts = Risk to others; rto = Risk to others

RMSEA = 0.05 (90% CI: (0.045, 0.055)); CFI = 0.978. According to commonly used criteria for a "good" model, both the RMSEA and the CFI satisfy these benchmarks, as does the χ^2/df ratio (< 2).

The estimated correlations between the latent variables showed high correlations among all factors except for the Risk-to-others latent variable.

Reliability evaluation for the five factors: symptoms and problems, positive resources, risk to self, risk to others, problems with others

The CORE-OM items are measured on an ordinal (5-point Likert) scale. Therefore, we computed internal consistency reliability using the ordinal Omega coefficient [33]. Omega coefficients based on the CFA sample

are shown in Table 3. The omega if item deleted and item to total correlations are shown in Table 4.

Deleting one item from the scale, produces a small decrease in the reliability score for most of the items. For the general problem scale, reliability is not affected much by deletion of a single item, partly because of the large number of items in this scale. Reliability is maximized if item 8 ("I have been troubled by aches, pains or other physical problems ") is deleted with an increase from .958 to .960. Also, for item 29 there is an increase in Omega if the item is deleted. For the positive scale, item 19 ("I have felt warmth and affection for someone ") performs worst. Dropping this item increases the Omega reliability score from .881 to .897. For the Risk to self-scale, deleting item 34 ("I have hurt myself physically or taken dangerous risks with my health") increase the Omega coefficient slightly.

The item to total correlations gives similar results as the Omega if item deleted, with the same items as mentioned above associated with the lowest item-total correlations. Particularly item 19 may be problematic to include as an indicator of the positive resources latent variable.

Measurement invariance for gender on the non-clinical sample

The configural model that allows all parameters to be estimated freely for the genders showed good model fit ($\chi^2(1034) = 1610.4$ ($p < .0005$), RMSEA = 0.046 (90% CI: (0.041, 0.050)), CFI = 0.969). Reasonable model fit for the configural model is necessary for further measurement invariance testing. The scalar model holding factor loadings and thresholds invariant across the genders had model fit: $\chi^2(1153) = 1751.9$ ($p < .0005$), RMSEA = 0.044 (90% CI: (0.040, 0.048)), CFI = 0.968. The DIFFTEST in Mplus showed significantly worse model fit for the scalar model compared with the configural model ($\chi^2(119) = 207.1$; $p < 0.0005$), indicating that holding all loadings and thresholds to be equal across genders are not warranted, so the requirement of full scalar invariance does not hold. Similar CFI and RMSEA values for the configural and scalar model could indicate that there are problems with a relatively few loadings and thresholds. Partial scalar invariance was therefore evaluated using modification indices.

Table 3 Omega coefficients based on the CFA sample

Scale	Ordinal Omega	95% confidence interval
General problems	.958	(.952, .965)
Positive resources	.881	(.861, .901)
Problems with others	.862	(.836, .887)
Risk to self	.931	(.918, .943)
Risk to others	.576[a]	–

[a] Spearman-Brown coeff

Table 4 Omega if item deleted and item to total correlations

	Omega if item deleted	95% CI[a]	Item to rest correlation[b]
General problems			
Item 1	.955	(.947, .962)	.783
Item 2	.956	(.949, .963)	.721
Item 5	.957	(.950, .964)	.668
Item 8	.960	(.953, .966)	.489
Item 10	.957	(.950, .964)	.652
Item 11	.956	(.949, .963)	.699
Item 13	.954	(.947, .962)	.807
Item 14	.955	(.947, .962)	.780
Item 15	.955	(.948, .962)	.759
Item 17	.954	(.946, .961)	.844
Item 18	.957	(.950, .964)	.637
Item 20	.956	(.949, .963)	.723
Item 23	.954	(.946, .961)	.849
Item 27	.954	(.947, .962)	.825
Item 28	.956	(.949, .963)	.692
Item 29	.959	(.953, .966)	.518
Item 30	.956	(.949, .963)	.705
Positive resources			
Item 3	.869	(.847, .891)	.614
Item 4	.859	(.835, .882)	.692
Item 7	.856	(.832, .880)	.715
Item 12	.848	(.823, .874)	.792
Item 19	.897	(.879, .914)	.288
Item 21	.870	(.849, .892)	.605
Item 31	.867	(.845, .890)	.621
Item 32	.863	(.840, .886)	.670
Risk Self			
Item 9	.896	(.876, .915)	.698
Item 16	.901	(.883, .920)	.723
Item 24	.909	(.892, .926)	.662
Item 34	.932	(.919, .945)	.575
PWO[c]			
Item 25	–	–	.726
Item 26	–	–	.639
Item 33	–	–	.648
Risk Others			
Item 6	–	–	.610[d]
Item 22	–	–	

[a] See Dunn, Baguley & Brunsden [35]; [b]Polyserial correlations between the item and the total (without the item) of the scale; [c]Problems with others; [d] Polychoric correlation

The largest modification index for the scalar model was associated with the lowest threshold for item 14 ("I have felt like crying"). This is the item with a large difference in the distribution for boys and girls. Relaxing the loading and the threshold constraints for this item made model fit slightly better, but still the DIFFTEST was significant: $\chi^2(115) = 174.1$ ($p < 0.0005$). Next, we relaxed the loading and thresholds for item 29, and the DIFFTEST was still significant: $\chi^2(111) = 161.5$ ($p < 0.0005$).

We did modify the scalar model by relaxing the constraints one by one according to the largest modification index. Table 5 shows the steps taken in this process:

The items that the data indicate might be most problematic to establish scalar invariance are shown in Table 5. Item 14 has very different distribution for boys and girls. Maybe including an item about crying is not a good idea when assessing differences on emotional problems for boys and girls. Admitting to crying is probably very different for the genders, and a boy and a girl with the same amount of problems might answer this question very differently (gender specific differences). Item 4 has the same tendency as for item 14 with a much higher proportion of boys reporting satisfaction with themselves than girls. Items 19 and 29 had the lowest standardized loadings both for boys and girls.

Gender differences on the latent factors

Based on the final partial scalar model (E), there was a significant gender difference in the factor means for the General problems latent variable (z = − 7.52; $p < .0005$; boys scoring lower than girls), Positive resources (z = 5.46; $p < .0005$; boys scoring higher than girls), Problems with others (z = − 5.02; $p < .0005$; boys scoring lower than girls), and Risk-to-others (z = 3.11; $p = .002$; boys scoring higher than girls). There were no significant differences between boys and girls on the latent Risk-to-self variable.

Clinical cut-off score

Clinical cut-off scores were estimated by employing Jacobson and Truax' [38] formula.

Before calculating the cut-off score, participants in the non-clinical sample that reported being in treatment were excluded ($n = 23$). The estimated CORE-OM all-items cut-off score according to Jacobson and Truax' formula was 1.31 (girls: 1.44; boys: 1.02). For the 17-item general distress scale the cut-off was 1.51 (girls: 1.69; boys: 1.09).

Discussion

The main findings in this validation of the CORE-OM in a mid-adolescent sample were a new factor solution and a higher cut-off score than reported in adult samples. The EFA resulted in a five factor solution, and the factor contents were interpreted as general problems, positive resources, risk to self, risk to others, and problems with others. The CFA model fit for this model was good. The measurement invariance analysis for gender should not be performed without modification of the scale. The clinical cut-off score based on the all-item total was higher than in an adult sample. Both the all item total and general problems cut-off score showed gender difference.

Factor analysis and reliability

From the exploratory factor analysis, based on the training part of the training sample, a five-factor model was interpreted to be the best candidate for model evaluation. In the EFA, this model had improved model fit over factor solutions with less factors, and had factors that were interpreted as General problems, Positive resources, Risk to self, Risk to others and Problems with others. In the following confirmatory factor analysis, done on the testing part of the sample, model fit for this model can be characterized as good.

The developers of the CORE-OM manual describes the instrument as a four-dimensional measure with

Table 5 Steps in showing partial scalar invariance

Step	Largest modification index	Chi-square (df)	RMSEA (90% CI)	CFI	DIFFTEST (Configural vs. partial scalar)
A	Item 14[a] Threshold 1	1713.8 (1149)***	0.043 (0.039, 0,047)	0.970	$\Delta\chi^2(115) = 174.1$***
B	Item 29[b] Loading	1702.2 (1145)***	0.043 (0.038, 0,047)	0.970	$\Delta\chi^2(111) = 161.5$***
C	Item 4[c] Threshold 3	1679.3 (1141)***	0.042 (0.038, 0,046)	0.971	$\Delta\chi^2(107) = 140.7$* $p = .016$
D	Item 31[d] Loading	1665.9 (1137)***	0.042 (0.037, 0,046)	0.972	$\Delta\chi^2(103) = 127.3$, $p = .053$ n.s.
E	Item 19[e] Threshold 4	1662.0 (1133)***	0.042 (0.038, 0,046)	0.972	$\Delta\chi^2(99) = 117.8$, $p = .10$ n.s.

* $p < .05$, *** $p < .0005$. *n.s.* non-significant. [a] I have felt like crying; [b] I have been irritable when with other people; [c] I have felt O.K. about myself; [d] I have felt overwhelmed by my problems; [e] I have felt warmth and affection for someone

dimensions of Subjective well-being, Problems, Functioning and Risk [6]. There may be many reasons why data from the present youth sample yielded a different factor structure. We believe that one main reason for this is that eight of the 34 items are positively keyed. Lyne et al. (2006) [9] showed that for CORE-OM on an adult sample, method factors related to positive and negative wording of the items played a role in achieving acceptable model fit. In our sample, all the eight positively keyed items loaded on the same factor in the five-factor EFA. It seems that when the adolescents answer these items, the positive resources in their lives are prompted, rather than just negative aspects. This highlights that assuming that a low score on a positively keyed item reflects the same as a high score on a negatively keyed item is problematic. According to the tripartite theory of anxiety and depression [39] negative affect and lack of positive affect may represent separate dimensions of internalizing problems. The current factor solution supports that negative affect and lack of positive affect are not two sides of the same coin.

Combining all the positively keyed items to a separate subscale not only solves the problem of reversed items, but also produces a substantially easier subscale to interpret than the theoretically derived Well-being scale, since it reflects resources, wellbeing and self-efficacy.

Incidentially, one of the positively keyed items had a much lower factor loading in the CFA than the other items. Although item 19 ("I have felt warmth and affection for someone») is positively keyed it differs in content from the rest of the positive items. This item may measure traits like empathy or affection directed towards other people, and not necessarily positive feelings about themselves. Also, removing this item from the scale improves the Omega reliability score by nearly 0.02, and this item had an item to rest-correlation below 0.30. One other reason for the low factor loading for item 19 may be that the Norwegian translation of the word "affection" is a word that is probably not used among Norwegian adolescents nowadays. Thus, a revision of the Norwegian translation is recommended.

The risk items split into two distinct factors. The risk to others items (item 6 and 22) correlates highly with each other but little with the other items in the questionnaire. We also see this through low factor correlations between the latent Risk-to-others variable and the other four latent variables in the CFA. In the EFA, a Risk to other scale shows up early, although it is questionable whether these two items cover a large enough range of such a dimension.

The Risk to self-dimension seems more robust, having a high reliability score for the internal consistency. The factor correlation between Risk-to-self and General problems is very high (> 0.90), and this seems natural as having many symptoms of problems may impact self-harm and suicidal ideation. Cross-loadings between Risk-to-self and General problems were evident for the items "I have felt despairing or hopeless" and "I have felt unhappy". Although such items may indirectly indicate risk of self-harm, we believe that these items are more direct indicators of the severity of of emotional problems.

The reliability analysis revealed that the Omega would increase slightly if item 34 "I have hurt myself physically or taken dangerous risks with my health" was removed from the scale. This item may or may not be related to intensions of self-mutilations or suicide. The other items within the Risk to self-scale are more directly associated with such intentions, while taking dangerous risks may be sensation-seeking behavior not directly associated with self-harm intentions.

For the General problems scale, half (17) of the CORE-OM items loaded highest on this variable in the EFA. For this scale, the Omega reliability would improve slightly if the items 8 and 29 were removed from the scale, and these two items had the lowest item to rest correlations for the items within the General problems scale. Item 8 ("I have been troubled by aches, pains or other physical problems ") may be caused by mental health issues but can also be a result of injuries, physical disease and other issues not related to emotional problems. Increased reliability removing this item from the general problem scale may be an indication of this. For item 29 ("I have been irritable when with other people") was probably the item that was most difficult to place. It loaded moderately on the General problem latent variable, and cross-loaded on the Problems with others latent variable. To be irritable when with others can be an indicator of problems with the functioning with others, but can also be an indicator of emotional problems since irritability may be associated with several traits or conditions [40].

Finally, items 25, 26 and 33 loaded on the Problems with others factor. These items have to do with relationships with others. Lyne et al. [9] pointed at the same three items as belonging to a common factor. After accounting for a general distress factor, these three items were the only items that had meaningful loadings on their residualized Functioning factor. This highlights that feelings of humiliation or critique from others and having no friends may form a separate factor in the CORE-OM instrument.

Measurement invariance for gender

We did a measurement invariance analysis for gender, to evaluate whether it is reasonable to make mean comparisons between girls and boys using CORE-OM.

Comparing the configural model and full scalar model, we found that scalar model fit significantly worse than the configural model, and this indicates that one cannot compare means for boys and girls without modifications to the scales. After 4–5 steps of relaxing constraints in the scalar model, we found a partial scalar model that did not fit significantly worse than the configural model. In comparing means for boys and girls on the CORE-OM scales, one should probably be careful in using the items 14, 29, 4, 31 and 19.

Different researchers rely on different fit statistics when evaluating measurement invariance Putnick and Bornstein [41] show that many consider that a small change in CFI or RMSEA going from a configural to a scalar model could indicate scalar invariance. The change in CFI and RMSEA shown for gender invariance in the non-clinical sample in the present study, is very small, and within the limits of full scalar invariance mentioned by Putnick and Bornstein [41]. However, it is problematic if one chooses the change in χ^2 as criterion for invariance when it is non-significant and other criteria when it is significant. We used a data driven method (modification indices) instead to establish partial scalar invariance. Partial scalar invariance can be concluded when a large majority of the items on the factors is invariant [42] The use of modification indices is also controversial [41], but can be helpful in determining items that are problematic. For example, our analysis showed that item 14 in CORE-OM ("I have felt like crying") may be an item that is problematic to include when symptoms of depression or anxiety are to be compared between the genders. Boys and girls report very differently on this item, and this difference cannot be attributed only to the amount of emotional problems on the latent scale the adolescents have but also to some gender specific traits.

Gender differences in the non-clinical sample

We compared factor means for male and female adolescents in the non-clinical group using the final partial scalar model. Boys and girls differed on four of the five latent variables. A non-significant difference between the genders was found for the risk to self-variable. For the non-clinical group few adolescents had thoughts of self-harm. The girls scored higher than boys did on the general factor, and that has also been shown in other studies [5, 13]. Boys scored significantly higher on the risk to others factor. This is consistent with other validations of the scale [7, 14] .

For the positive resources latent variable, girls scored significantly lower than boys. Finally, for the Problems with others factor the girls scored higher than boys. The items in this scale have to do with feelings of having been criticized, humiliated, made shameful or having no friends, and are as such about emotional relations with others. Girls tend to use emotional coping skills more often than boys, and help from others, while boys tend to devaluate such emotional expressions [43], hence stronger feelings related to emotional relationships can be the result. In the Japanese version of CORE-OM the female participants showed lower scores on "close relationships" subscales [14].

Factor correlations between the latent variables in the chosen factor solution were high, except for those involving risk to others. Similar gender differences for general emotional problems, positive resources and problems with others can be a sign that related concepts are being involved.

Mixing positive and negative items in a questionnaire

Lyne et al. [9] concluded their article, studying 2140 adult patients, that the most useful scoring method of the CORE-OM would be to compute a general total score based on the 28 non-risk items and a risk total based on the remaining six items. The main difference between the 17-item general problems scale from the present study and the 28-item non-risk scale is the exclusion of the positively keyed items from the 17-item version.

One of the reasons for including both positively and negatively keyed items in a questionnaire is to reduce acquiescence bias (response style bias, respondents tending to agree with statements) [44]. However, positively and negatively keyed items may involve different cognitive processes [45, 46] and this is one of the reasons that a positive item latent variable showed up in the EFA. It is a paradox that including some positively keyed items in a questionnaire consisting mostly of negatively keyed items, in order to mitigate acquiescence bias, seems to confuse the responders and therefore makes the instrument less valid and scales less reliable.

Clinical cut-off score

The original validation of the CORE suggested a clinical cut-off of 1.2 [5], and later validations have suggested a cut-off point as low as 1.0 [47] to define clinical caseness. However, in these adolescent samples, the cut-off score on the All-items CORE-OM was 1.31, 1.44 (girls) and 1.02 (boys). This finding needs to be replicated, but it corresponds well with the finding that youths also score higher than adults on the BDI [22, 23]. Consistent of the results from the present study we also recommend the 17-item factor as a measure of general problems. The positively keyed items do not interfere with this factor and the problem with others items are also excluded. In this way we have a more reliable measure on emotional problems and the cut-off scores for this factor is suggested as an alternative to the established

All items minus Risk score. The rationale for this is that the All items minus Risk 28-item score includes all reversed items, and may thus actually underestimate the level of emotional distress experienced by patients. The cut-off scores for both All-items and the 17-items general distress factor show gender differences, with girls scoring higher than boys and a higher score than in adult samples [7]. We suggest that the cut-off scores either is gender specific or that the cut-off for gender combined is set lower to accommodate for the boys lower scoring, as suggested by Connell et al. [47].

Limitations

The clinical sample may not be representative of the entire CAMHS population due to the sample being preselected based on symptoms of emotional problems. Furthermore, patients evaluated as suicidal were excluded from the sample because they could not be subjected to the 6-week waiting condition. However, since the CORE-OM was mainly developed to monitor outpatient treatment and is not the outcome measure of choice for psychosis or conduct disorder, the present clinical sample probably has a high density of the phenomena that the CORE-OM was designed to monitor.

The age span in the non-clinical sample was 14–18, while the age range in the clinical sample was 14–17. The reason for this is that Norwegian CAMHS receives only those younger than age 18 as patients, while youths 18 and older are referred to mental health services for the adult population. However, in high school, enrolment in different grades is based on the year of birth. We decided not to exclude the 18-year-olds from the non-clinical sample. Furthermore, the mean age in the two samples is similar.

Due to the low rate of males in the clinical sample, the mean and standard deviation in the male clinical sample used in the Jacobson and Truax formula have large standard errors. Therefore, the clinical cut-offs for boys are encumbered with uncertainty.

Conclusions

Although the present version of CORE-OM shows promising psychometric properties, there are some challenges with the instrument. Leaning on van Sonderen et al. [48] and Suárez-Alvarez et al. [44], we believe that using a mix of positively and negatively keyed items should be avoided, if the intention is not to measure separate dimensions. However, the five-factor solution found in this validation both had a good model fit, and not the least, yielded clinically meaningful subscales. According to the factors found in this study we recommend the 17-item factor as a more reliable measure of general problems. Comparing means for gender in non-clinical samples should not be done without modification of the general emotional problem and the positive resources scales. This should be objectives for future revisions of the scale.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40359-020-00459-5.

Additional file 1.

Abbreviations

CORE-OM: Clinical Outcome in Routine Evaluation-Outcome Measure; EFA: Exploratory Factor Analysis; CFA: Confirmatory Factor Analysis; BDI: Becks Depression Inventory; SDQ: Strengths and Difficulties Questionnaire; CAMHS: Child and Adolescent Mental Health Services; ANOVA: Analysis of Variance; WLSMV: Weighted Least Squares Means and Variance adjusted; RMSEA: Root Mean Square Residual; CFI: Comparative Fit Index

Acknowledgements

We would like to thank the clinicians and administrative staff management. We want to express our deep gratitude to all the participants.

Authors' contributions

VL wrote the manuscript, were active in designing the study for both samples, collecting and analyzing the data. BHH conducted the analysis and in writing of the analysis and result section and contributed in writing of all parts of the manuscript. KL, KS & CMM were active in designing and collection of the non-clinical data as a part of KS & CMM's master thesis [49]. IS was active in designing, collecting and analyzing the study for both samples and in the writing of all parts of the manuscript. All authors revised and approved the final manuscript.

Funding

The study was financed from the Northern Norway Regional Health Authority, through project number: PFP1110–13, PFP1112–13 and HNF1383–17, The pilot project received financial support from The Regional Centre for Child and Youth Mental Health and Child Welfare-North. The funding body had no role in designing the study or collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

The datasets generated and analysed during the current study and the full study protocol are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was performed in compliance with the Helsinki Declaration for research on humans and was approved by the Regional Committees for Medical and Health Research Ethics (REC North). Reference number 2011/1937. All participants participated and consented according to the regulations in the research project, also with written parental consent for those under age 16 (REC North, Reference number 2011/1937).

Consent for publication

Consent to publish was given from every participant and parents when warranted according with the Helsinki Declaration for research on humans and was approved by the REC North (Reference number 2011/1937).

Competing interests

The authors declare that they have no conflicts of interest.

Author details

[1]Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, P.O. Box 6050 Langnes, N-9037 Tromsø, Norway. [2]Department of Child and Adolescent Mental Health, Divisions of Child and Adolescent Health, University Hospital of North Norway, 9038 Tromsø, Norway. [3]The Regional Centre for Child and Adolescent Mental Health – North, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway. [4]The Norwegian Labour and Welfare Administration (NAV), Employment Advisory Services in Troms and Finnmark, 9811 Vadsø, Norway.

[5]Substance use and Psychiatry unit, Department of Substance Use and Addiction Medicine, Clinic for Mental Health and Substance Use, Nordland Hospital, 8076 Bodø, Norway. [6]Clinic for Mental Health and Substance Use, University Hospital of North Norway, 9291 Tromsø, Norway.

Received: 11 February 2020 Accepted: 6 August 2020
Published online: 20 August 2020

References

1. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Üstün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry. 2007;20:359–64.
2. Lambert MJ, Whipple JL, Hawkins EJ, Vermeersch DA, Nielsen SL, Smart DW. Is it time for clinicians to routinely track patient outcome? A meta-analysis. Clin Psychol Sci Pract. 2003;10:288–301.
3. Barkham M, Evans C, Margison F, McGrath G, Mellor-Clark J, Milne D, et al. The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. J Ment Health. 1998;7:35–47.
4. Barkham M, Leach C, Lucock M, Evans C, Margison F, Mellor-Clark J, et al. Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical outcomes in routine evaluation-outcome. J Consult Clin Psychol. 2001;69:184–96.
5. Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. Br J Psychiatry. 2002;180:51–60.
6. Evans C, Mellor-Clark J, Margison F, Barkham M, Audin K, Connell J, et al. CORE: clinical outcomes in routine evaluation. J Ment Health. 2000;9:247–55.
7. Skre I, Friborg O, Elgarøy S, Evans C, Myklebust LH, Lillevoll K, et al. The factor structure and psychometric properties of the clinical outcomes in routine evaluation - outcome measure (CORE-OM) in Norwegian clinical and non-clinical samples. BMC Psychiatry. 2013;13:99.
8. Elfström ML, Evans C, Lundgren J, Johansson B, Hakeberg M, Carlsson SG. Validation of the Swedish version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). Clin Psychol Psychother. 2013;20:447–55.
9. Lyne KJ, Barrett P, Evans C, Barkham M. Dimensions of variation on the CORE-OM. Br J Clin Psychol. 2006;45:185–203.
10. Bedford A, Watson R, Lyne J, Tibbles J, Davies F, Deary IJ. Mokken scaling and principal components analyses of the CORE-OM in a large clinical sample. Clin Psychol Psychother. 2010;17:51–62.
11. Barkham M, Mellor-Clark J, Connell J, Cahill J. A CORE approach to practice-based evidence: a brief history of the origins and applications of the CORE-OM and CORE system. Couns Psychother Res. 2006;6:3–15.
12. Kristjánsdóttir H, Sigursson BH, Salkovskis P, Ólason D, Sigurdsson E, Evans C, et al. Evaluation of the psychometric properties of the Icelandic version of the clinical outcomes in routine evaluation-outcome measure, its transdiagnostic utility and cross-cultural validation. Clin Psychol Psychother. 2015;22:64–74.
13. Palmieri G, Evans C, Hansen V, Brancaleoni G, Ferrari S, Porcelli P, et al. Validation of the Italian version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). Clin Psychol Psychother. 2009;16:444–9.
14. Uji M, Sakamoto A, Adachi K, Kitamura T. Psychometric properties of the Japanese version of the clinical outcomes in routine evaluation-outcome measure. Compr Psychiatry. 2012;53:600–8.
15. Viliū nienė R, Evans C, Hilbig J, Pakalniškiene V, Danilevičiute V, Laurinaitis E, et al. Translating the clinical outcomes in routine evaluation outcome measure (CORE-OM) into Lithuanian. Nord J Psychiatry. 2013;67:305–11.
16. Connell J, Barkham M, Mellor-Clark J. Mental health CORE-OM norms of students attending university counselling services benchmarked against an age-matched primary care sample. Br J Guid Couns. 2007;35:41–57.
17. Twigg E, Cooper M, Evans C, Freire E, Mellor-Clark J, McInnes B, et al. Acceptability, reliability, referential distributions and sensitivity to change in the young person's clinical outcomes in routine evaluation (YP-CORE) outcome measure: replication and refinement. Child Adolesc Ment Health. 2016;21:115–23.
18. Conrad KJ, Wright BD, McKnight P, McFall M, Fontana A, Rosenheck R. Comparing traditional and rasch analyses of the Mississippi PTSD scale: revealing limitations of reverse-scored items. J Appl Meas. 2004;5:15–30.
19. Lindwall M, Barkoukis V, Grano C, Lucidi F, Raudsepp L, Liukkonen J, et al. Method effects: the problem with negatively versus positively keyed items. J Pers Assess. 2012;94:196–204.
20. Connell J, Barkham M. CORE-10 user manual, version 1.1. Rugby: CORE System Trust & CORE Information Management Systems Ltd; 2007.
21. Leach C, Lucock M, Barkham M, Noble R, Clarke L, Iveson S. Assessing risk and emotional disturbance using the CORE-OM and HoNOS outcome measures at the interface between primary and secondary mental healthcare. Psychiatr Bull. 2005;29:419–22.
22. Albert N, Beck AT. Incidence of depression in early adolescence: a preliminary study. J Youth Adolesc. 1975;4:301–7.
23. Teri L. The use of the beck depression inventory with adolescents. J Abnorm Child Psychol. 1982;10:277–84.
24. Bøe T, Hysing M, Skogen JC, Breivik K. The strengths and difficulties questionnaire (SDQ): factor structure and gender equivalence in Norwegian adolescents. PLoS One. 2016;11(5):e0152202.
25. Khesht-Masjedi MF, Shokrgozar S, Abdollahi E, Golshahi M, Sharif-Ghaziani Z. Comparing depressive symptoms in teenage boys and girls. J Family Med Prim Care. 2017;6(4):775–9.
26. Muthén LK, Muthén BO. Version 6 Mplus user's guide. Los Angeles: Muthén & Muthén; 2010.
27. Lorentzen V, Fagermo K, Handegård BH, Skre I, Neumer SP. A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14-17 years old in child and adolescent mental health services (CAMHS). BMC Psychol. 2020;8(1):25.
28. Steiger JH. Structural model evaluation and modification: an interval estimation approach. Multivariate Behav Res. 1990;25:173–80.
29. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990;107:238–46.
30. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods Psychol Res Online. 2003;8:23–74.
31. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
32. McDonald RP. Test theory: a unified treatment. Mahwah: L. Erlbaum Associates; 1999.
33. Peters G-JY. The alpha and the omega of scale reliability and validity: why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. Eur Health Psychol. 2014;16(2):56–69.
34. Revelle W, Zinbarg R. Coefficients alpha, Beta, omega, and the GLB: comments on Sijtsma. Psychometrika. 2009;74:145–54.
35. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol. 2014;105:399–412.
36. Leppink J. Statistical methods for experimental research in education and psychology. Cham: Springer; 2019.
37. Raykov T, Marcoulides G. Introduction to psychometric theory. New York: Routledge; 2011.
38. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12–9.
39. Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. J Abnorm Psychol. 1991;100(3):316–36.
40. Leibenluft E, Stoddard J. The developmental psychopathology of irritability. Dev Psychopathol. 2013;25(4 Pt 2):1473–87.
41. Jager J, Putnick DL, Bornstein MH. II. More than just convenient: the scientific merits of homogeneous convenience samples. Monogr Soc Res Child Dev. 2017;82(2):13–30.
42. Byrne BM, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. Psychol Bull. 1989;105(3):456–66.
43. Copeland EP, Hess RS. Differences in young adolescents' coping strategies based on gender and ethnicity. J Early Adolesc. 1995;15(2):203–19.
44. Suárez-Alvarez J, Pedrosa I, Lozano LM, García-Cueto E, Cuesta M, Muñiz J. Using reversed items in Likert scales: a questionable practice. Psicothema. 2018;30(2):149–58.
45. Marsh HW. Negative item bias in ratings scales for preadolescent children: a cognitive-developmental phenomenon. Dev Psychol. 1986;22(1):37–49.

46. Marsh HW. Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? J Pers Soc Psychol. 1996;70(4):810–9.
47. Connell J, Barkham M, Stiles WB, Twigg E, Singleton N, Evans O, et al. Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. Br J Psychiatry. 2007;190:69–74.
48. Van Sonderen E, Sanderman R, Coyne JC. Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. PLoS ONE. 2013; 8(7):e68967.
49. Solem, K. M., & Moen, C. M. (2015). Validation of the Norwegian version of the 10-item version of clinical outcomes in routine evaluations outcome measure (CORE-10) in a Norwegian adolescent population (aged 14–18) (Master's thesis, The Arctic University of Norway).

Publisher's Note

Paper 3

Lorentzen, Veronica, Fagermo, Kenneth, Handegård, Bjørn Helge; Neumer, Simon-Peter; Skre, Ingunn. Long-term effectiveness and trajectories of change after treatment with SMART, a transdiagnostic CBT for adolescents with emotional problems. BMC Psychology 2022; Volume 10 (167). ISSN 2050-7283.s 1 – 18.s doi: 10.1186/s40359-022-00872-y.

Lorentzen *et al. BMC Psychology* (2022) 10:167
https://doi.org/10.1186/s40359-022-00872-y

BMC Psychology

RESEARCH Open Access

Long-term effectiveness and trajectories of change after treatment with SMART, a transdiagnostic CBT for adolescents with emotional problems

Veronica Lorentzen[1,2*], Kenneth Fagermo[2], Bjørn Helge Handegård[3], Simon-Peter Neumer[3,5] and Ingunn Skre[1,4]

Abstract

Background: There is a need for long-term effectiveness trials of transdiagnostic treatments. This study investigates the effectiveness and diagnosis-specific trajectories of change in adolescent patients attending SMART, a 6-week transdiagnostic CBT for anxiety and depression, with 6-month follow-up.

Methods: A randomized controlled trial with waiting list control (WLC) was performed at three child and adolescent mental health outpatient services (CAMHS) in Norway. Referred adolescents (N = 163, age = 15.72, 90.3% females) scoring 6 or more on the emotional disorders subscale of the Strengths and Difficulties Questionnaire (SDQ) were randomly assigned to treatment or to WLC. Long-term follow-up (N = 83, baseline age = 15.57, 94% females) was performed 6 months after treatment completion (Mean = 7.1 months, SD = 2.5). Linear mixed model analysis was used to assess time by group effects in patients with no diagnosis, probable anxiety, depressive disorder, and combined anxiety and depressive disorder.

Results: Almost one third (31%) obtained full recovery according to the inclusion criterium (SDQ emotional). There was highly significant change in all outcome variables. Effect sizes (ES) were largest for general functioning, measured with CGAS (ES: d = 2.19), and on emotional problems measured with SDQ (ES: d = 2.10), while CORE-17, BDI-II and CGAS all obtained ES's close to 1. There were no significant time by diagnostic group interactions for any outcomes, indicating similar trajectories of change, regardless of diagnostic group. Waiting 6 weeks for treatment had no significant impact on long-term treatment effects.

Limitations: Possible regression to the mean. Attrition from baseline to follow-up.

Conclusions: Six weeks of transdiagnostic treatment for adolescents with emotional problems showed highly significant change in emotional symptoms and functioning at 6-month follow-up. Patients with anxiety, depression, combined anxiety and depression, and emotional problems with no specific diagnoses, all had similar trajectories of change. Hence this transdiagnostic SMART treatment can be recommended for adolescent patients with symptoms within the broad spectrum of emotional problems.

*Correspondence: veronica.lorentzen@uit.no; veronica.lorentzen@unn.no

[1] Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway
Full list of author information is available at the end of the article

Trial registration: ClinicalTrials.gov Identifier: NCT02150265. First registered May 29, 2014.

Keywords: Cognitive behavioral therapy, Adolescence, Trajectories of change, Emotional disorders, Transdiagnostic, Psychological treatment, Long-term effectiveness, Anxiety, Depression, Child and adolescent mental health services

Introduction

Anxiety and depression are the most frequent mental disorders in both the general youth population and in those receiving treatment in child and adolescent mental health outpatient services (CAMHS) [1–3]. Anxiety and depressive disorders commonly co-occur during adolescence with rates as high as 75% in clinical samples [4, 5], presenting overlapping symptoms and emotional distress [6]. Untreated, this interferes negatively with numerous mental health outcomes, and can lead to psychological, cognitive, social, and academic impairments [7, 8]. Treatments based on cognitive behavioral therapy (CBT) for both anxiety and depressive disorders have empirical support [9, 10]. Multiple systematic reviews and meta-analyses have reported moderate to large effect sizes (ES) for psychological treatment in youth samples [11–13]. However, the evidence for the effects of these empirically supported treatments rests mainly on efficacy trials. A question raised is to what degree these results hold up when delivered in routine clinical care, as for instance in CAMHS, given the clear differences between the two settings [14–17]. Some have stated that treatments supported by efficacy trials may show reduced treatment effects when transferred to effectiveness trials in CAMHS settings [16, 17]. A recent meta-analysis on CBT for internalizing disorders concludes that CBT delivered in routine care is efficacious in reducing emotional disorders and symptoms, with outcomes comparable to results obtained in efficacy studies. However the authors stated that the quality of the included studies was fair, and the heterogeneity high. Limitations such as varying inclusion criteria of the studies was also highlighted [18]. The participants in this meta-analysis comprised of both children and adolescents, testing various types of CBT, cognitive therapy and behavior therapy. The meta-analysis did not provide information of whether the treatment comprised of transdiagnostic protocols or diagnosis specific treatment protocols. In sum, there are considerable differences between research clinics and routine clinical care. A remaining question is whether short-term transdiagnostic treatment is an effective and lasting treatment for adolescent patients with anxiety and/or depression when delivered in the setting of routine clinical care. Given the high co-occurrence of depression and anxiety in clinical samples, there is a need for effectiveness studies on transdiagnostic treatment that target both anxiety and depression.

As mentioned by Queen et al. [19], youth anxiety and depression share a number of psychological, biological and environmental risk factors (for a review see [20]). Anxiety, depression and traumatic stress have common symptom patterns such as rumination and worry [21], and behavioral avoidance [22]. Moreover, negative affect has been suggested as a latent factor underlying both depressive and anxiety disorders [23, 24]. CBT treatment trials designed to test interventions on single disorders have shown so called "spill-over effects", where similar response to treatment is shown for comorbid anxiety and depressive disorders [25]. The results of a meta-analysis of CBT for treatment of primary depression are one example demonstrating not only an effect for depressive symptoms, but also "spill-over effects" with reduction in anxiety symptoms [26].

As a consequence of the high comorbidity of emotional disorders among adolescents, and shared vulnerability factor, efforts have been directed towards a transdiagnostic approach to treatment [27–30]. Transdiagnostic treatment is built upon cognitive, behavioral, and physiological processes that are shared or common across diverse disorders. The treatment does not presuppose careful differential diagnostic assessment between disorders belonging to the targeted spectrum, and represents an adoption of an integrative approach [31]. A transdiagnostic intervention for anxiety and depression emerges as an attractive approach for community clinical practice for this target group, offering more flexible interventions compared to standard single-disorder interventions. The use of traditional single-disorder protocols can be time consuming and costly for clinical practitioners [28, 30]. It is hypothesized that transdiagnostic approaches could contribute to lowering the clinical burden in learning several manuals and allow for more flexible interventions to patients presenting with comorbid emotional disorders [27]. As a consequence, a large array of treatment protocols conceptualized as transdiagnostic have been developed and tested. The most extensive research has been performed on the Unified protocol (UP) [32]. The UP exists in many adaptations. Protocol adaptions for children (UP-C; UP Children) and adolescents (UP-A; UP Adolescents) have been developed and tested by Ehrenreich-May and colleagues [33–36] demonstrating promising effects of transdiagnostic treatment for emotional disorders in children and adolescents. There are several additional studies examining various transdiagnostic

programs for youth with emotional disorders in different research settings; in primary care [37, 38]; in school settings [39, 40]; and in parent-led teletherapy [41], all showing promising effects. Another promising transdiagnostic therapy protocol for emotional disorders is the Structured Material for Therapy (SMART) [42]. The short-term effectiveness of the SMART treatment was investigated in an RCT with a Norwegian sample of adolescents [43] and small to moderate effect sizes for the time by group interaction effect (ranging from 0.19 to 0.65) were observed for anxiety, emotional symptoms and general functioning, while the effect size for depressive symptoms did not reach significance directly after completion of treatment.

Comparable studies regarding trajectories of change from recruitment, through therapy and to follow-up, are scarce. We found one study employing the UP-A that examined the concurrent trajectories of primary anxiety and depressive symptoms across the course of treatment and at 6 month follow-up. This study showed similar rates of change on self-reported symptoms during the treatment, but whereas anxiety symptoms showed significant improvement after treatment, depressive symptoms seemed to plateau [19]. Investigating effects of transdiagnostic treatment provides the opportunity to examine whether there are separate trajectories of change when it comes to depression, anxiety, and comorbid anxiety and depression for patients in a CAMHS setting, in order to show whether they present similar or different rates of change both post-treatment and at follow-up.

Most studies measure post-treatment effects, but less frequently incorporate follow-up measurements conducted months after the end of treatment. Despite the strong evidence for youth CBT post-treatment effectiveness, relapse after treatment can be observed in as many as one third to one half of treated youths (e.g. [44, 45]). So far, our knowledge about long-term effects of CBT for emotional disorders is limited. A meta-analysis performed by Rith-Najarian and colleagues [46] provides support for stability of treatment effects of CBT for youths in a long-term follow-up with rather large within-subject effect sizes ($g = 1.23–1.82$), but also highlights the need for several improvements in research standards, with an emphasis on prioritizing assessment at long-term follow-up. Many studies conduct limited follow-up assessments only 2–3 months post-treatment (e.g. [11, 13]). Longer follow-up studies are important for several reasons and are needed to understand the persistence of treatment effects of CBTs [46]. A meta-analysis on depression demonstrated that treatment effect sizes were negatively correlated with duration from end of treatment until follow-up [26]. The same study found that treatment duration was not correlated with outcome, suggesting that some briefer treatments may have potential to be as effective as lengthier ones [26]. Number of sessions or weeks of treatment did not moderate the effect size or remission rates in a recent meta-analysis investigating CBT for children and adolescents treated for internalizing disorders in routine clinical care [18]. This could indicate that shorter treatments could be as effective as lengthier ones.

In addition to duration of treatment, there are other issues worth examining, such as effectiveness in adolescent samples and the effects of waiting for treatment. In general, there is a need for more research on adolescents, especially since we know that emotional disorders persist if left untreated [47]. A meta-analysis examining effectiveness of anxiety treatment in children and adolescent, has shown that the research is mainly based on children, despite the potential strong effects shown by the limited number of studies on adolescents [48].

A relevant question for clinical trials with a waiting list control group, is to study the effect of waiting time before receiving treatment. Waiting time before commencing treatment is frequently used as a quality benchmark for health services, also for Norwegian CAMHS. For somatic illness and for mental health conditions with high risk for harm or self-harm, waiting time can obviously imply risk of deterioration or a fatal outcome. However, little is known about whether other young patients react negatively to waiting, for instance by reduced attendance, or by developing a more treatment resistant condition while waiting [49]. Will the delayed onset of therapy after enrollment influence the short- and long-term outcome of therapy? This question has clinical implications, since most patients in CAMHS wait before starting treatment.

Aims

The aim of the present study was to examine the long-term effectiveness, and diagnostic group specific change trajectories of a six-session transdiagnostic CBT for young patients (age 14–17) with depression, anxiety and combined anxiety and depression, treated in regular CAMHS, by

1. Examining treatment effects at follow-up 6 months after treatment completion.
2. Examining the impact of waiting 6 weeks before start of therapy on the long-term treatment effect.
3. Examining change trajectories for patients with diagnoses of pure anxiety, pure depression, and combined anxiety and depression, from pre-treatment, through treatment, and at 6-month follow-up.

Methods

This is a 6-month follow-up study of adolescent patients participating in a randomized controlled clinical trial with waiting list control. The active treatment was transdiagnostic CBT, according to the treatment manual Structured Material for Therapy (SMART) [42], delivered by the clinical staff at community child and adolescent mental health service clinics (CAMHS). The pre- to post-treatment short-term evaluation of the SMART treatment protocol for the present study sample is described in Lorentzen et al. [43], including more detailed information about the SMART treatment protocol, clinical setting, training of therapists, randomization procedure, treatment integrity, user satisfaction, and therapeutic alliance.

Procedure

We asked a total of 498 patients referred to CHAMS for informed consent, from January 2012 to June 2016. Of these, 335 were excluded due to governmental restriction on waiting time prior to start of treatment. Patients in the direct intervention group were pre-evaluated and commenced treatment with SMART immediately after enrollment. After completion of the 6-week SMART treatment, they were evaluated post-treatment. The patients in the waiting list control (WLC) condition were pre-evaluated twice; first at enrollment, and secondly after a 6-week waiting period, prior to commencing treatment with SMART. The WLC group went through post-treatment evaluation immediately after completing the 6-week SMART treatment, 12 weeks after enrollment. Follow-up was scheduled 6 months post-treatment. Figure 1 shows timepoints for measurements in weeks. Figure 2 shows the consort flow diagram.

The mean duration of time between post-treatment evaluation and follow-up was 32.8 weeks (SD = 10.8, range 28–90).

Participants

The participants are described in Lorentzen et al. [43]. The full sample included 145 adolescents (aged 14–17, $M = 15.72$, $SD = 1.14$, 90.3% females), recruited from referrals to three public child and adolescent mental health outpatient clinics, from January 2012 to June 2016. During the routine intake procedure at the clinics, the participants who fulfilled the inclusion criteria described below were informed about the study and asked to participate. Adolescents over 16 years and parents of children under 16 years signed informed consent. Inclusion criteria were (a) age between 14 and 17 years; (b) a score of at least 6 on the Strength and Difficulties Questionnaire (SDQ) screening tool emotional scale; and (c) maintenance of a maximum waiting time for necessary medical care of 6 weeks given by Norwegian health authorities. Exclusion criteria were (a) a diagnosis of pervasive developmental disorder (PDD); (b) psychotic symptoms; (c) use of anxiolytic or anti-depressant medication during the treatment period; and (d) patients who did not speak the Norwegian language. A total of 199 adolescents were assessed for eligibility and were asked for informed consent. Of these, 36 did not consent. A total of 163 were randomized to either treatment or waiting list. In the current study, a total of 18 patients did not complete the treatment or waiting phase. Of these, we had no information for the reason for non-completion for 11 patients, 2 cited lack of motivation, 3 were referred to other treatment (2 received trauma treatment, 1 regular cognitive behavioral therapy), and finally 3 withdrew because of geographical distance (2 moved to another location, 1 had a long distance to travel to get to the CAMHS). The main administrative challenge to completing the follow-up was difficulty getting into contact with the patients, and in some cases the therapists forgot to perform follow-up after 6 months. Results for the final study sample are presented according to the CONSORT flow diagram in Fig. 2.

Treatment

The adolescents were treated with the SMART program [43]. The SMART program is a transdiagnostic CBT program with materials organized in five modules (introduction, depression, anxiety, assertiveness training, and summary, in a total of eight sessions). In this study, the core modules for anxiety and depression, except the module for assertiveness training (2 sessions), were employed as a standard brief therapy in the outpatient clinics. Norwegian law restricts waiting time in CAMHS, and by excluding assertiveness training both groups had similar time intervals (6 weeks waiting time and 6 weeks treatment). Four modules were given over 6 sessions, each with a duration of 90 min.

The program has proven its initial effectiveness [43] and is based on well supported methods for treating anxiety and depression with emphasis on cognitive restructuring, exposure and activation. The modules consist of a definition of individual treatment goals, activation of personal resources, behavioral experiments, information about emotional problems and related coping strategies. Although the treatment is modular and flexible, it was delivered in a linear manner for research purposes.

Measures

The following measures were employed in the present study:

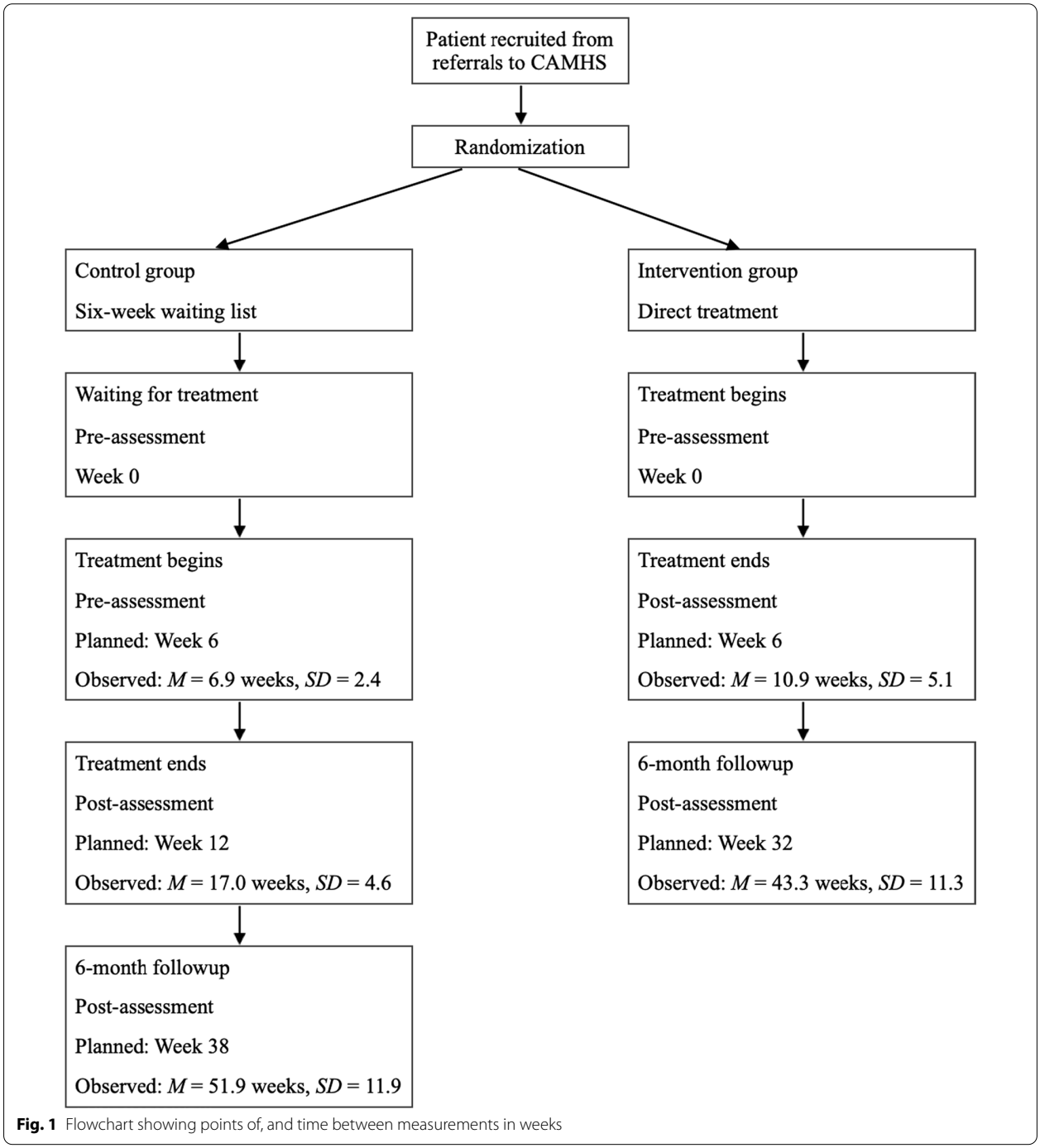

Fig. 1 Flowchart showing points of, and time between measurements in weeks

Diagnostic assessment: DAWBA

Development and Well Being Assessment (DAWBA) [50] is a multi-informant computer-administered diagnostic interview, with both open- and closed-ended questions. It is administered at intake as a digital self-report instrument in the participating CAMHS. The data is gathered through the authorized DAWBA online system and published in accordance to the copyright terms and with permission from the copyright holder. In the present study, only information from the patients was used. DAWBA uses computer algorithms to suggest the likelihood of diagnoses, ordering the probability of a diagnosis into

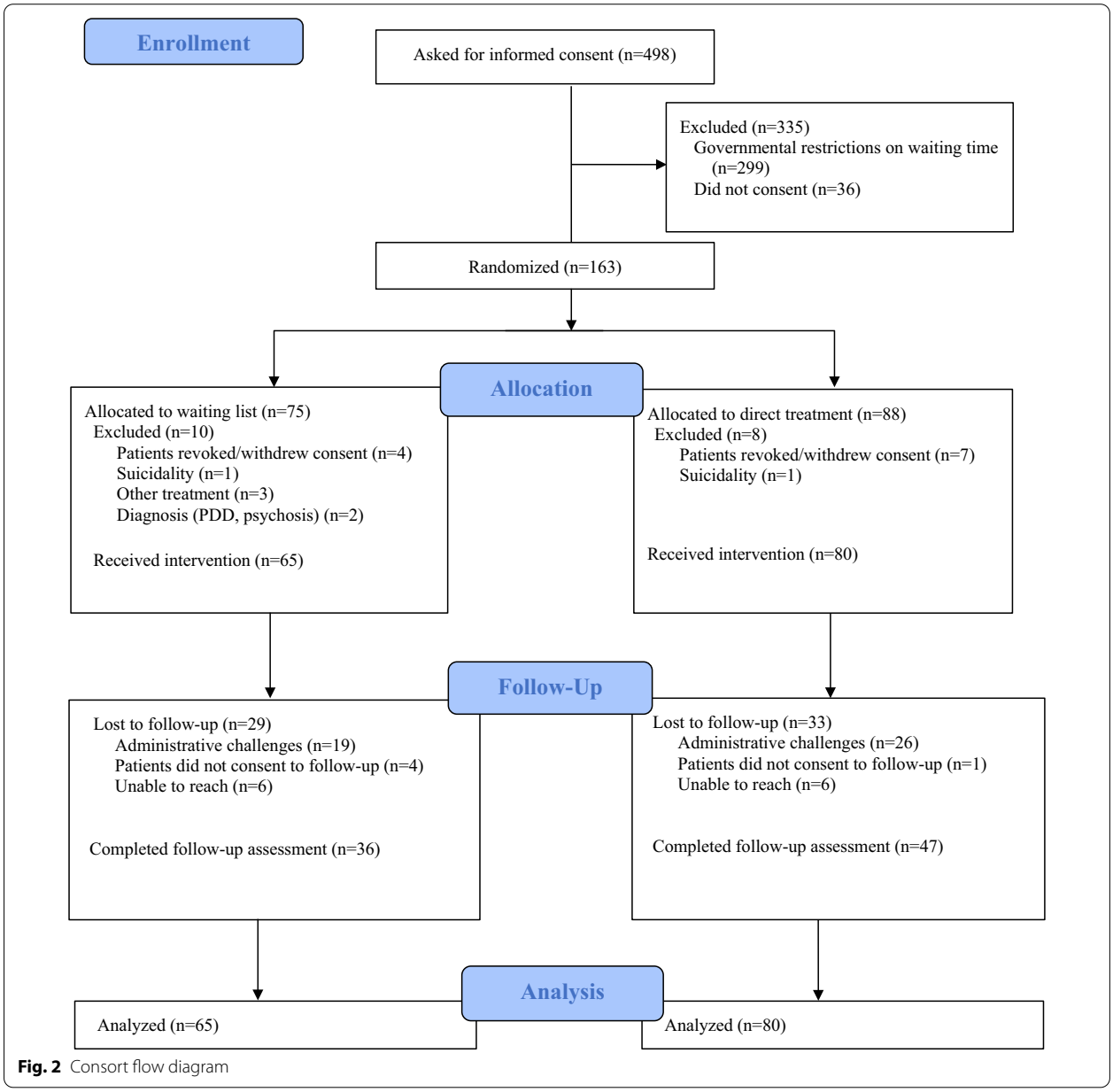

Fig. 2 Consort flow diagram

DAWBA bands ranging from 0 to 5 [51]. The top two bands (4 and 5) suggest a ≈ 50% and >70% likelihood that the patient meets criteria for the disorder. Goodman and colleagues [50] found that DAWBA could discriminate between community and clinic samples of youth, and later [51] found that DAWBA is well suited to find approximate prevalence of disorders. When comparing the computer algorithms to clinician rated diagnoses, Goodman et al. [51] found that DAWBA can underestimate the prevalence of disorders on a group level. On an individual level, kappa values showed agreement between 0.4 and 0.7, specificity 0.98–0.99, positive predictive values 0.5–0.8 and negative predictive values 0.96–0.99.

Inclusion and main outcome measure for the present study: SDQ

The Strengths and Difficulties Questionnaire (SDQ) [52] was administered as a part of the authorized DAWBA online system. Data is published in accordance with the copyright terms and with permission from the copyright holder. SDQ was administered at enrollment, end of therapy and at 6 month follow-up. The main inclusion criteria

and the primary outcome measure for emotional symptoms was the emotional problems subscale on the self-rated SDQ for 11–17 year-olds. SDQ is a brief emotional and behaviour screening questionnaire where symptom items are scored on a 3-point Likert scale, from 0 (not true) to 1 (somewhat true) to 2 (certainly true). The maximum score on the emotional subscale is 10, and based on a Norwegian sample [53] we used a cut-off of 6 or above to separate a clinical from a non-clinical population. SDQ is a frequently used screening instrument and has satisfactory psychometric properties [54, 55]. In this study, we used only the emotional symptoms subscale, which has shown acceptable reliability and adequate internal consistency [54]. Internal consistency in our sample was acceptable for the SDQ emotional scale (Cronbach's $\alpha = 0.70$).

General functioning: CGAS

The Children's Global Assessment Scale (CGAS) [56] was used as a secondary outcome measure for general level of function, and is a routine tool employed in the participating CAMHS. CGAS is a therapist-scored numeric scale ranging from 1 to 100, with high scores indicating a higher level of functioning and a score lower than 70 as the clinical cut-off point. In this study a group of at least 3 experienced clinicians scored each child's clinical profile blindly, and the scores were averaged. The clinicians had used CGAS routinely in clinical practice, and had extensive experience with the instrument. CGAS has shown good psychometric properties [57]. In the present sample, there was high inter-rater reliability between the three CGAS raters (ICC = 0.97).

General psychological distress: CORE-OM and CORE-17

Clinical Outcome in Routine Evaluation-Outcome Measure (CORE-OM) [58] was used as a secondary outcome measure for general psychological distress, and risk of suicide and self-harm, and was introduced for the purposes of the present study. CORE-OM, developed in the UK [47], is widely used as a general psychotherapy and consulting outcome measure in mental health outpatient and consulting services [59–61], and has been translated into 20 languages [62–65]. Originally CORE-OM is a 34-item questionnaire with items using a 5-point Likert scale, giving an average score between 0 and 4, where high scores indicate an increased symptom severity. CORE-OM includes items related to well-being, anxiety, depression, trauma reactions, sleep, bodily pain, daily functioning and risk of harm to self and others. CORE-OM has shown good psychometric properties [58, 62], but also has some methodological challenges [60]. In the present study a 17-item general problem scale score was used, based on the results of a validation of the CORE-OM in Norwegian adolescents; a community sample from junior and senior high-school, and the clinical sample in the present study [66]. The CORE General problem scale score reports symptoms and problems, both psychologically and in relation to others, and can be interpreted as a measure of general psychological distress [66]. In this study, the clinical cut-off point for CORE-17 was set at 1.3, based on the mentioned validation study [66]. The internal consistency of the CORE-17 subscale was excellent (Cronbach's $\alpha = 0.90$).

Depressive symptoms: BDI-II

Beck Depression Inventory, second edition (BDI-II) [67] was used as a secondary outcome measure for extent and depth of depressive symptoms. BDI-II is a frequently used instrument in the CAMHS, but is not a mandatory tool. BDI-II is a 21-item questionnaire with items using a 4-point Likert scale, giving a maximum score of 63. Cut-off scores suggested by Beck et al. [67] were between 14 and 19 for mild depression, 20–28 for moderate depression and 29–63 for severe depression. BDI-II has shown good psychometric properties [68–70]. The internal consistency for the BDI-II in the present sample was excellent (Cronbach's $\alpha = 0.91$).

Anxiety symptoms: MASC

Multidimensional Anxiety Scale for Children (MASC) [71] was used as a secondary outcome measure for the degree of anxiety. In this study we used the MASC total score. MASC is a 39-item questionnaire with items scored on a 4-point Likert scale. High scores indicate a higher degree of anxiety. MASC has shown good psychometric properties [71, 72]. The internal consistencies of the MASC subscale scores in the sample varied (Cronbach's α between 0.57 and 0.80).

Data analysis

This study was part of a randomized controlled effectiveness trial of transdiagnostic CBT for adolescents. The power calculations have been reported in Lorentzen et al. [43], and therefore, will not be presented in detail here.

The reliable change index (RCI) [73] was used as an evaluation of clinically significant and/or reliable change. RCI was evaluated on the main inclusion and outcome measure; the SDQ emotional scale. The inclusion criterium in this study was a score of at least 6 on the SDQ emotional scale, and hence, the criterium for clinically significant change was a score lower than 6. The criterium for reliable change, was that the magnitude of change was statistically significant, and in the present study reliable change was calculated to be a change of at least 4 scale points on the SDQ.

To test the time by group and time effects, we used linear mixed model analysis [74]; measurement occasion (level 1; pre, post and follow-up) is nested within individuals (level 2). In this analysis, time was treated as a continuous variable, and represented as the number of days since baseline for the different measurement occasions. A random slope model was used. We computed estimated marginal means and standard errors in a linear mixed model analysis where time was treated as categorical, in order to get model-based predictions for each treatment condition or diagnostic group on each measurement occasion.

Effect sizes for the time by group effect and the time effect were computed as the unstandardized coefficient (computed by the LMM analysis) divided by the pooled within-group standard deviation at baseline. This ratio was multiplied by the average number of days from baseline to the follow-up measurement in the total sample, since the unstandardized coefficient was given in change difference per day or change per day [75]. Follow-up was performed after on average 327 days since the baseline measurement (minimum = 151, maximum = 632).

For handling of pretreatment differences between the groups and dropout, see Lorentzen et al. [43].

We used IBM SPSS v25 for all analyses, and 0.05 was used as significance level.

Results

Additional treatment after completing SMART

Patients who were in need of further treatment after the SMART intervention, were taken care of in routine care. Data obtained from the CAMHS case records show that after completing SMART (post-treatment) and before the 6-month follow-up, 33 (22.8%) patients had zero additional sessions, 22 (15.2%) had 1 additional session, 7 (4.8%) had 2 additional sessions, 9 (6.2%) had 3 additional sessions and 62 (42.8%) had 4 or more additional sessions after the SMART intervention. We had no information about additional treatment for 12 (8.3%) patients after they received the SMART intervention. We previously analyzed the association between the number of sessions of treatment and change in general functioning and mental health from post-treatment to follow-up. We found no time by additional treatment interaction on any of the dependent variables (no association between the number of sessions of extra treatment and change in the dependent variables from post-treatment to follow-up), and therefore did not include the additional treatment variable in models used for analysis of long-term effects of treatment or diagnostic groups.

Table 1 shows diagnoses and comorbidity, and the resulting distribution of the patients into direct treatment or WLC group. More than half (n = 80 (55%)) had a probable diagnosis of one or more anxiety disorder(s), and generalized anxiety disorder and social phobia were the most frequent probable anxiety diagnoses. More than half the participants (n = 75 (52%)) had a probable diagnosis of depression. Nearly a quarter of the participants had a probable pure anxiety disorder, one fifth had a probable pure depressive disorder, and nearly one third had a probable diagnosis of both anxiety and depression. A quarter of the participants did not have probable diagnosis of either an anxiety or a depressive disorder. There were too few males in the sample to perform gender comparisons. The randomization was performed independently of the diagnostic assessment. However, the two groups had fairly equal distributions of diagnoses.

Table 1 Diagnoses (n = 145)

DAWBA prediction	Total		WLC		Direct treatment	
	n	% (of 145)	n	% (of 65)	n	% (of 80)
Pure anxiety	34	23.4	18	27.7	16	20.0
Pure depression	29	20.0	12	18.5	17	21.3
Depression and anxiety	46	31.7	18	27.7	28	35.0
Depression and GAD	30	20.7	10	15.4	20	25.0
Depression and Social phobia	27	18.6	10	15.4	17	21.3
Depression and specific phobia	7	4.8	0	0.0	7	8.8
Depression and agoraphobia	9	6.2	4	6.2	5	6.3
Depression and panic disorder	6	4.1	4	6.2	2	2.5
No diagnosis of anxiety or depression	36	24.8	17	26.2	19	23.8

Diagnoses in both ICD-10 and DSM-IV (same algorithm)

Table 2 Clinically significant and reliable change on the SDQ emotional scale from baseline to follow-up

	Reliable change[b]	
	No	**Yes**
No clinically significant change		
n	26	4
% of Total	44.8%	6.9%
Clinically significant change[a]		
n	10	18
% of Total	17.2%	31.0%

[a] Clinically significant change: The number of patients with SDQ emotional score < 6 at follow-up

[b] Reliable change: The number of patients with a change of at least 4 scale points on the SDQ emotional scale from baseline to follow-up

Clinically significant and reliable change from baseline to follow-up

The rates of clinically significant and/or reliable change on the main inclusion criterium of emotional symptoms from the SDQ is presented in Table 2.

Nearly half the patients reported clinically significant change on the SDQ emotional scale from baseline to follow-up, indicating that at follow-up they were scoring below the clinical cut-off point, and hence had a subclinical score on this instrument. Close to 40% of the patients showed statistically significant change from baseline to follow-up, indicating reliable change. Nearly a third of the patients fulfilled the criteria for both, and thus showed clinically significant and reliable change in scores from baseline to follow-up on the SDQ emotional scale.

Looking at change in general functioning from baseline to follow-up in the present sample, only 1 (0.09%) out of 109 patients were rated below the clinical cut-off point on CGAS at baseline, while nearly half the patients; 29 out of 64 (45.3%), had a CGAS score above the clinical cut-off point at follow-up.

Group differences in change and overall effect sizes from baseline to follow-up

Table 3 reports effects of time from baseline to follow-up for the outcome variables in the overall sample. There was a highly significant change in the sample for all outcome variables, and the change was in the hypothesized direction. Effect sizes were largest for general functioning, where the predicted change at follow-up corresponded to 2.19 standard deviations increase in the CGAS score, and for the SDQ emotional problems scale there was a predicted decrease equal to 2.10 standard deviations. For the secondary outcome measures of depressive symptoms (BDI-II), anxiety symptoms (MASC) and general psychological distress (CORE-17), effect sizes corresponded to a decrease of approximately 1 standard deviation.

The patients in the two treatment conditions received the SMART treatment over different time schedules. The adolescents in the treatment group started treatment immediately after baseline, while the waiting list condition waited 6 weeks before initiation of the treatment. Apart from this 6-week waiting time, the conditions for the two groups were equal. To test whether the groups differed in change from baseline to follow-up, i.e., to test whether the time schedule affected overall change rates, linear mixed models' analyses were performed. Table 3 shows results from these linear mixed model analyses for the five outcome variables.

There were no statistically significant time by group interactions for any of the outcomes [Self-reported emotional problems (SDQ); General functioning (CGAS); Anxiety (MASC); Depression (BDI-II); General psychological distress (CORE-17)]. The effects computed at follow-up were small, except for SDQ emotional problems, where the difference in change rates at follow-up corresponded to a standardized effect size of approximately 1.0 standard deviations. These results indicate that there was not sufficient evidence for different change rates in the two experimental conditions from baseline to follow-up.

Trajectories of change according to diagnostic groups

As previously shown in Table 1, each individual was classified into one of four diagnostic categories based on the DAWBA prediction levels 4–5, corresponding to at least 50% probability of the disorder. The four diagnostic categories were: No probable anxiety or depression; pure anxiety; pure depression; both anxiety and depression. We then investigated the trajectories of change for these four diagnostic groups and tested whether they changed differently on the five primary and secondary outcome measures from baseline to follow-up. The results are presented in Table 4 for overall results, Fig. 3 for SDQ emotional scale, Fig. 4 for CGAS, Fig. 5 BDI-II, Fig. 6 for CORE-17, Fig. 7 for MASC.

Table 4 and Figs. 3, 4, 5, 6 and 7 shows no significant time by diagnostic group interactions on any of the outcome measures, indicating similar patterns of change in the four diagnosis groups. All diagnostic groups had change in the clinically desired direction on all measures from baseline to follow-up. The patient group with no probable diagnosis of anxiety or depression, had a more fluctuating trajectory than the other diagnostic groups, but this group was also the smallest, and thus we believe these fluctuations may have happened by chance, due to the small sample size. Furthermore, all three groups with emotional disorders had similar change trajectories;

Table 3 Estimated marginal means (and standard errors) for measures on functioning, emotional problems, anxiety and depression

Measure	Treatment			Wait-list				Time* group F	Time*group effect size at T3 $d = b*327/SD^{ab}$	Overall change effect size $d = b*327/SD^{b}$
	Pre-treatment start 0 weeks	Post-treatment ends 6 weeks	6 months follow-up	Pre-waiting for treatment	Pre-treatment start 6 weeks	Post-treatment ends 12 weeks	6 months follow-up			
CGAS	51.77 (1.40)	61.57 (1.47)	69.00 (1.72)	49.03 (1.65)	52.13 (1.52)	59.33 (1.63)	64.78 (2.03)	0.15	− 0.19	F = 83.4**/d = 2.19
BDI total	28.98 (1.43)	19.74 (1.50)	15.59 (1.78)	29.60 (1.61)	24.12 (1.60)	18.68 (1.71)	15.76 (1.94)	0.77	0.21	F = 74.3**/d = − 1.02
MASC total	60.54 (2.02)	50.42 (2.11)	45.89 (2.53)	61.29 (2.27)	57.61 (2.26)	50.01 (2.39)	46.26 (2.73)	0.15	0.10	F = 47.0**/d = − 0.91
CORE 17	2.25 (0.09)	1.66 (0.10)	1.46 (0.11)	2.43 (0.10)	2.03 (0.10)	1.65 (0.11)	1.50 (0.12)	0.03	− 0.02	F = 60.2**/d = − 1.19
SDQ emotion	7.90 (0.22)	6.54 (0.25)	4.90 (0.33)	8.11 (0.27)	7.30 (0.26)	6.20 (0.29)	5.96 (0.37)	3.90	1.03	F = 59.0**/d = − 2.10

*< 0.05, **$p < 0.0005$

[ab] Unstandardized coefficient for the time by group effect (change difference per day), and SD is the pooled within-group standard deviation at baseline

[b] Measured on average 327 days from baseline

Table 4 Estimated marginal means (standard errors) at four measurement occasions for four diagnostic groups

Scale	Diagnose group	0 weeks EMM (SE)	6 weeks EMM (SE)	12 weeks EMM (SE)	Follow-up EMM (SE)	F[a] t*g
CGAS	No diagnosis of anxiety or depression	54.23 (2.49)	61.11 (2.49)	74.64 (4.14)	69.77 (3.14)	0.19 NS
	Pure anxiety	49.23 (2.35)	55.28 (2.27)	56.70 (2.89)	63.10 (3.10)	
	Pure depression	51.85 (2.48)	59.72 (2.44)	59.34 (3.60)	69.90 (2.59)	
	Anxiety and depression	48.91 (1.81)	55.15 (1.83)	60.42 (3.14)	66.31 (2.24)	
BDI-II	No diagnosis of anxiety or depression	22.07 (2.25)	14.83 (2.38)	4.71 (4.02)	9.59 (2.93)	0.23 NS
	Pure anxiety	23.66 (2.06)	17.29 (2.11)	14.47 (2.78)	11.68 (2.78)	
	Pure depression	32.75 (2.26)	24.96 (2.24)	22.64 (3.22)	16.36 (2.68)	
	Anxiety and depression	35.94 (1.79)	27.63 (1.85)	25.74 (2.98)	20.40 (2.19)	
MASC	No diagnosis of anxiety or depression	57.15 (3.43)	48.21 (3.62)	38.26 (5.57)	37.54 (4.53)	0.87 NS
	Pure anxiety	60.06 (3.15)	54.18 (3.22)	49.97 (4.16)	48.03 (4.16)	
	Pure depression	59.95 (3.45)	53.24 (3.41)	52.34 (4.80)	47.38 (4.04)	
	Anxiety and depression	64.54 (2.74)	55.89 (2.81)	49.09 (4.43)	46.35 (3.30)	
CORE-17	No diagnosis of anxiety or depression	1.94 (0.15)	1.34 (0.16)	0.80 (0.26)	1.21 (0.20)	0.86 NS
	Pure anxiety	2.13 (0.14)	1.74 (0.14)	1.46 (0.19)	1.35 (0.18)	
	Pure depression	2.44 (0.15)	2.03 (0.15)	1.74 (0.22)	1.36 (0.17)	
	Anxiety and depression	2.66 (0.12)	2.05 (0.12)	1.81 (0.20)	1.73 (0.14)	
SDQ emotional symptoms	No diagnosis of anxiety or depression	7.37 (0.36)	6.75 (0.42)	5.46 (0.68)	4.11 (0.54)	1.44 NS
	Pure anxiety	7.91 (0.33)	6.10 (0.35)	6.07 (0.50)	5.42 (0.50)	
	Pure depression	7.73 (0.36)	6.21 (0.38)	6.01 (0.63)	6.08 (0.50)	
	Anxiety and depression	8.61 (0.29)	8.03 (0.32)	5.74 (0.63)	5.62 (0.43)	

NS not significant, *EMM* estimated marginal means

[a] F-test for time by diagnostic group interaction, where time is treated as a continuous variable (days since baseline)

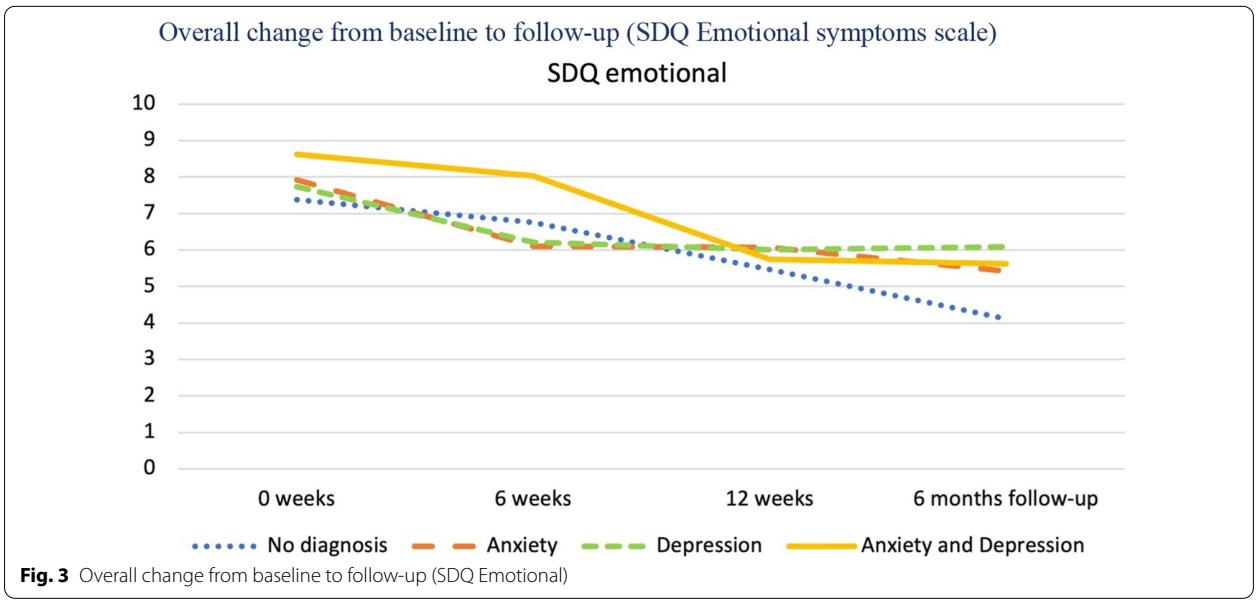

Fig. 3 Overall change from baseline to follow-up (SDQ Emotional)

and all had the same change profile from baseline, with a break point post-intervention, and a steady slope to follow-up, with a continuing change in the clinically desired direction. Albeit that there were no statistically significant differences between the trajectories, it is interesting to note that the comorbid anxiety and depression group

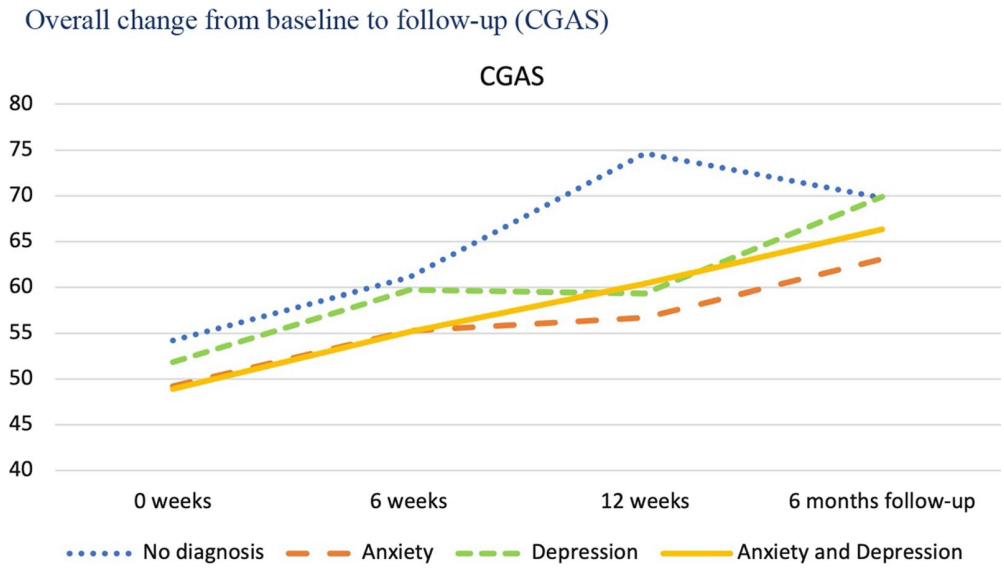

Fig. 4 Overall change from baseline to follow-up (CGAS)

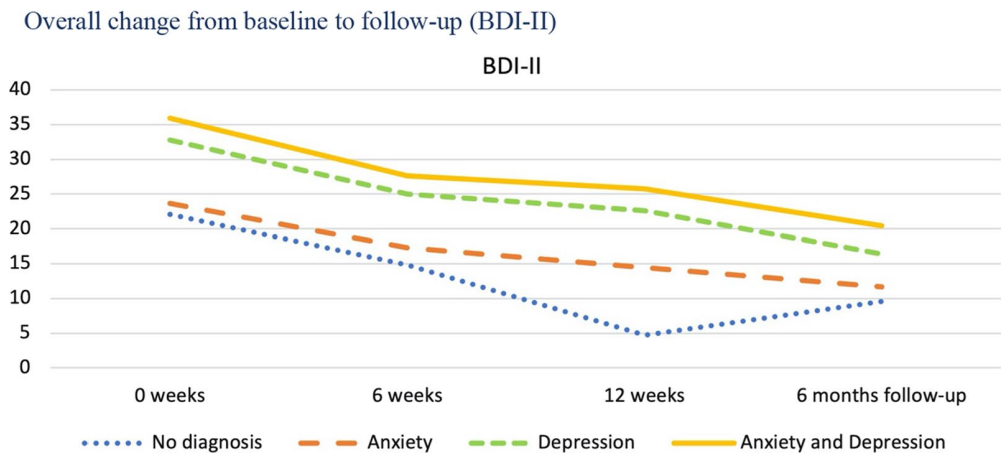

Fig. 5 Overall change from baseline to follow-up (BDI-II)

had more severe scores at almost all measurement occasions and on all outcome variables.

Discussion

The present study investigated the long-term effectiveness and trajectories of change for adolescent patients treated with the transdiagnostic CBT program SMART in a CAMHS setting. The main findings at 6-month follow-up was, firstly, that 6 weeks of transdiagnostic treatment for emotional problems yielded recovery in nearly half the patients measured at follow-up, and clinically significant and reliable change in nearly one third, according to the main inclusion criterium SDQ emotional symptoms. Secondly, nearly half the patients were rated as having normal functioning on the outcome measure CGAS at follow-up. The analysis from baseline to follow-up showed a highly significant change in the overall sample for all outcome variables (emotional symptoms, functioning, general psychological distress, depressive and anxiety symptoms), and the effect sizes were all well above what Cohen characterized as large effects [76]. However, it is important to note that these effects sizes were within-effect sizes and can only be denoted as large when compared to other within-effect sizes. Furthermore, the

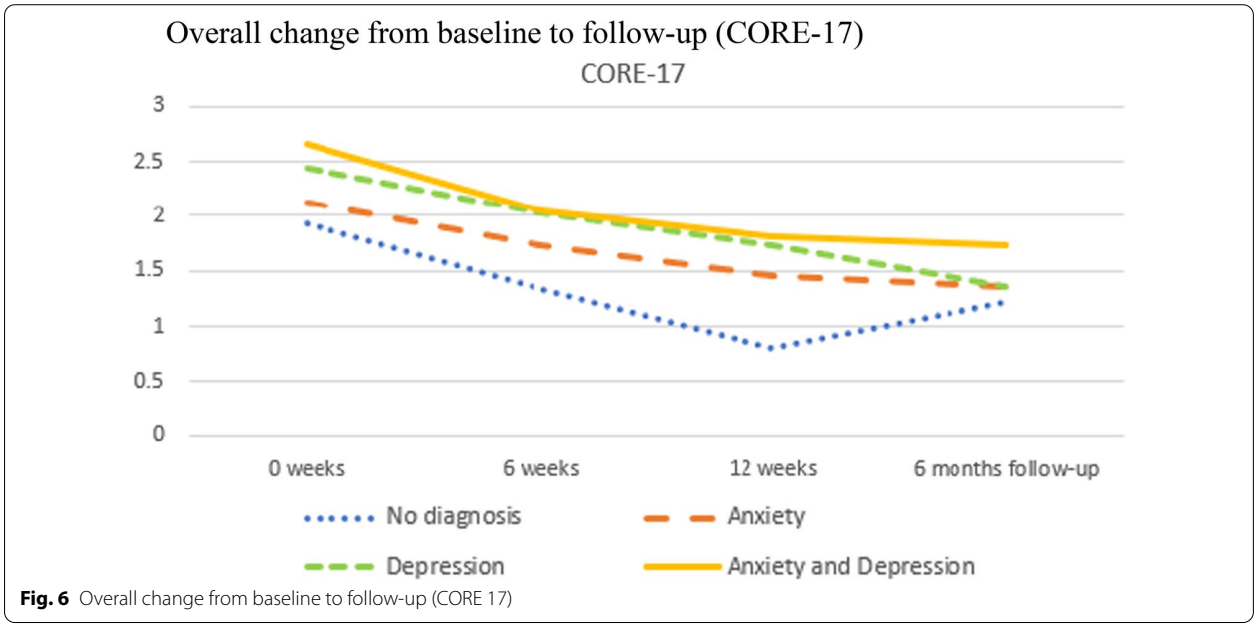

Fig. 6 Overall change from baseline to follow-up (CORE 17)

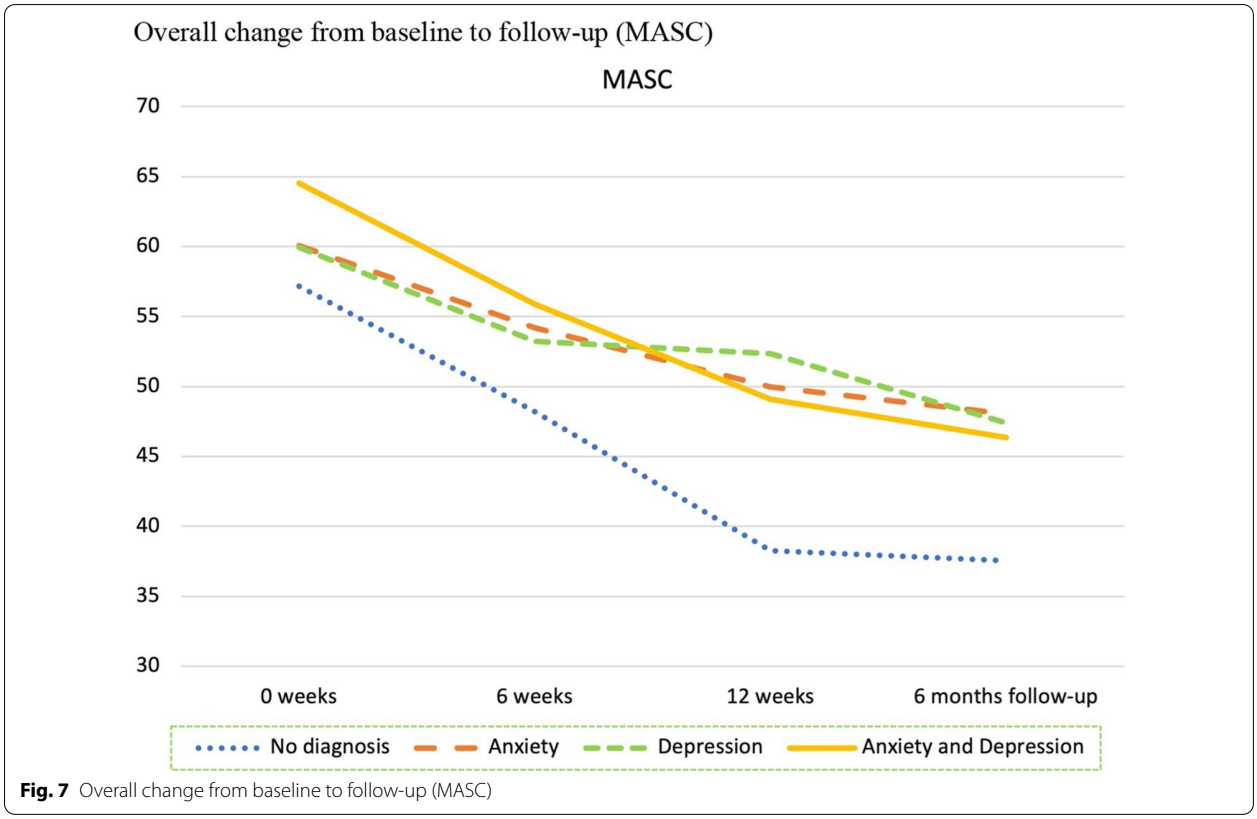

Fig. 7 Overall change from baseline to follow-up (MASC)

change trajectories for the young patients receiving this transdiagnostic treatment, were similar for pure anxiety, pure depression, anxiety and depression combined, and patients with emotional problems without a specific diagnosis. Finally, our findings indicate that waiting 6 weeks before commencing treatment with the SMART

program, seemed to have had no influence on the long-term outcome for these adolescent patients.

The magnitude of the observed changes indicates a significant drop in emotional symptoms, both anxiety and depression, and psychological distress, on all measures, and a heightening of psychological functioning. Compared to the changes observed at post-treatment, immediately after conclusion of the 6-week SMART program [26], the long-term effectiveness was even more pronounced, and the effect sizes were statistically significant on all outcome measures. Thus, the improvement seems to have continued from post-treatment to follow-up.

There are few comparable effectiveness studies with similar participants, clinical settings, treatment duration, and transdiagnostic treatment, but these changes are in line with those found in a meta-analysis of durability of effects of treatments for emotional disorders at 1-year follow-up [15]. These findings of effect durability are also consistent with the findings of a meta-analysis of long-term outcomes for youth CBTs targeting anxiety [77]; and extend to long-term effect durability for youth CBTs targeting traumatic stress and depression as well [41].

Although our treatment was only six sessions, out of which only two sessions was specific anxiety treatment, our findings are comparable to the results from the Coping Cat study, conducted in CAMHS in Norway, on the CGAS and MASC, although «Coping Cat» is a 12-session treatment solely focusing on anxiety [78]. The results for the present study are also comparable to those found at 6-month follow-up after CBT for depression, as measured with the BDI (e.g. [79]). In a recent meta-analysis of CBT for emotional disorders in routine care number of weeks or sessions did not affect ES or remission rate [18]. Hence, shorter treatments can potentially be as effective as lengthier ones in reducing symptoms of emotional disorders.

Concerning change from baseline to follow-up for various diagnostic groups (no anxiety or depression diagnosis; anxiety; depression; anxiety and depression), the results showed that the groups shared similar patterns of change with reductions of emotional symptoms. The effects are comparable to similar studies on CBT targeting anxiety, depression and trauma [11–13] and meta-analysis of effects found at 1-year follow-up (e.g. [46]). Given that the relapse is usually as high as one third in youths [44, 45], the transdiagnostic treatment of only six sessions of SMART shows promising results also at 6-month follow-up. Although 40% of the sample had 4 or more additional sessions, there was no association between the number of sessions of additional treatment and change in the dependent variables from post-treatment to follow-up.

There were no significant differences for any of the outcome variables in longitudinal trajectories from baseline to follow-up for the two treatment conditions. The two groups received their treatment during different stages of the study, where the treatment group were treated immediately after baseline, and participants in the waiting list condition waited 6 weeks before initiation of the treatment. Hence, there were no indications for different change rate due to differing time schedules in the two groups after therapy had commenced. This implicates that waiting 6 weeks to receive treatment did not have any negative effect when measuring anxiety and depression approximately 6 months after treatment. In general, waiting for treatment has been regarded as negative and great effort has been taken in reducing waiting time [80]. Previous research on physical health services has shown that waiting time is associated with poorer functioning both socially and physical, lower quality of life and poorer health status [81–83]. However, little is known about whether other young patients react negatively to waiting, for instance by reduced attendance, or by developing a more treatment resistant condition while waiting [49]. In this study, the adolescents said yes to this particular treatment, and was told both time frame for waiting and treatment. This would make the situation not just predictable, but would make it possible for them to read about CBT. On the ethical side, it needs to be mentioned that in this study adolescents with the most serious problems had to be excluded from participation due to Governmental restrictions on waiting time for this patient group. Also, 6-week waiting time is relatively short. We have no data to show whether a longer waiting time would have negative effect on long-term outcome.

Strengths and limitations

The strengths of the present study were that it was performed in a routine clinical setting, that the therapists were representative for ordinary CAMHS, and that the patients were routinely referred cases. Furthermore, a large part of the sample showed comorbid presentations of emotional symptoms, and a variety of symptom measures with good psychometric properties were used, and the interrater reliability of the CGAS, blindly scored by at least three clinicians, was also high. Although the study results has high external validity based on routine care data, limitations concerning the internal validity must be taken into account, since there were high rates of attrition at follow-up, leading to small diagnostic subgroups. This study was not designed with the purpose of comparing the development in different diagnostic groups, which may leave the tests of time by diagnostic group interactions with low power. Therefore, the results have to be interpreted with caution. The linear mixed model

analysis applied used information from all 145 subjects when estimating effects. This analysis require that missing data are missing at random (MAR). Sixty-two cases could not be included in the follow-up assessment, and if missing observations are not missing at random (NMAR), estimates may be biased. We do not know if missing data are MAR or not. However, 42 of these cases have been classified as dropout due to "administrative challenges" that consisted of lack of monitoring and clinical routines in conducting follow-up assessment, or sick leave at the clinical administration and not by the patients, and that may reduce the probability of NMAR bias.

The primary outcome measure in this study was the SDQ emotional symptoms scale. Since the patients were screened and included into the study based on their above cut-off score on the SDQ emotional scale, all had homogenously elevated scores at inclusion. One would therefore expect more average change on this measure at follow-up, simply because one would expect that part of the change would be regression to the mean. Similarly, since all patients, baring one, had a daily functioning score (CGAS) indicating being in need of treatment at inclusion, one would also expect that the long-term improvement on this measure was partly caused by regression to the mean. However, for both SDQ emotional symptoms and CGAS the long-term effect sizes were high (d = 2.10 and d = 2.19, respectively), and double the effect sizes found for the secondary outcome measures. Hence, the changes on these two outcome measures were probably too large to solely be explained by regression to the mean. However, the standard deviation for SDQ emotional scale was low because of the screening of the patients and large effect sizes on this scale could also be explained by low standard deviations. On the other hand, the inclusion into the study was not based on the scores on measures of general psychological distress (CORE-17), depression (BDI-II) and anxiety (MASC). Thus, the distributions at inclusion for those outcome variables were more heterogeneous, and those scoring in the lower end of the distribution would have little chance for systematic improvement on these scales, due to so called floor effects. Nevertheless, the observed effect sizes at follow-up were quite large, also for these three measures.

This study supports the growing evidence that transdiagnostic treatments are effective in treating depression and anxiety [84]. Furthermore, the results support a broad and pragmatic clinical approach to treating internalization disorders in youths; with offering a short-term transdiagnostic treatment as the first choice, and a more tailored and diagnosis specific treatment as the second choice, if needed. The study should be replicated with an active control group also at follow-up to rule out possible placebo effects and to evaluate what incremental effects the SMART treatment can contribute compared to treatment already established in the CAMHS.

Conclusions

Six weeks of transdiagnostic treatment with the SMART program for emotional problems showed promising results with large significant change in overall emotional symptoms and significant improvement in daily functioning in a follow-up at 6 months post-treatment. There were no significant treatment group or diagnostic group differences in the overall rate of change from baseline to follow-up in the non-diagnosis group, depression only, anxiety only, and depression and anxiety combined group. If a six session transdiagnostic treatment can be acceptable, and have lasting impact, it is a scalable and likely cost-effective treatment to be considered as the first step in a stepped care model in CAMHS for youths with emotional disorders. The results also illuminate the need for further treatment for some of the patients.

Abbreviations

CBT: Cognitive behavioral therapy; CAMHS: Children and adolescent mental health services; WLC: Wait list control; SDQ: Strength and difficulties questionnaire; SD: Standard deviation; EST: Empirically supported treatment; ES: Effect size; UP-A: Unified protocol for the treatment of emotional disorders; RCT : Randomized controlled trial; SMART: Structured material for therapy; PDD: Pervasive developmental disorder; DAWBA: Development and well being assessment; CGAS: The children's global assessment scale; CORE-OM: Clinical outcome in routine evaluation-outcome measure; BDI-II: Beck depression inventory, second edition; MASC: Multidimensional anxiety scale for children.

Acknowledgements

We want to express our gratitude to the adolescents and parents who, through their participation, have contributed to more knowledge supporting the treatment of anxiety and depression. Finally, we are grateful to and would like to thank the clinicians and administrative staff management at the participating community mental health clinics for all the hard work they have put into the project.

Author contributions

VL wrote the manuscript, were active in designing the study, collecting and analyzing the data. KF participated in collecting, analyzing and writing of the method and result section. BHH conducted the analysis and contributed in the writing of the method, results and discussion section. SPN were active in designing, and writing of all parts of the manuscript. IS was the project leader and active in designing the study, and in writing all parts of the manuscript. All authors read and approved the manuscript.

Funding

Open access funding provided by UiT The Arctic University of Norway (incl University Hospital of North Norway). The study was financed from the Northern Norway Regional Health Authority, through project number: PFP1110–13, PFP1112–13 and HNF1383–17, The pilot project received financial support from The Regional Centre for Child and Youth Mental Health and Child Welfare-North. The funding body had no role in designing the study or collection, analysis, and interpretation of data or in writing the manuscript.

Availability of data and materials

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was performed in compliance with the Helsinki Declaration for research on humans and was approved by the Regional Committees for Medical and Health Research Ethics (REC North). Reference number 2011/1937. Adolescents over 16 years and parents and children under 16 years signed informed consent. All participants participated and consented according to the regulations in the research project, also with written parental consent for those under age 16 (REC North, Reference number 2011/1937).

Consent for publication

Not applicable.

Competing interests

The second last author SPN received royalties from the publisher of the manual. All other authors do not have any competing interests.

Author details

[1]Department of Psychology, Faculty of Health Sciences, UiT The Arctic University of Norway, 9037 Tromsø, Norway. [2]Department of Child and Adolescent Psychiatry, Divisions of Child and Adolescent Health, University Hospital of North Norway, P.O. Box 19, 9038 Tromsø, Norway. [3]Regional Centre for Child and Youth Mental Health and Child Welfare, UiT The Arctic University of Norway, 9037 Tromsø, Norway. [4]Department of General Psychiatry, University Hospital of North-Norway, P.O. Box 6124, 9291 Tromsø, Norway. [5]Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, 0484 Oslo, Norway.

Received: 19 July 2021 Accepted: 27 June 2022
Published online: 05 July 2022

References

1. Chavira DA, Stein MB, Bailey K, Stein MT. Child anxiety in primary care: prevalent but untreated. Depress Anxiety. 2004;20(4):155–64.
2. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry. 2003;60(8):837–44.
3. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10–24 years: a systematic analysis. Lancet. 2011;377(9783):2093–102.
4. Sørensen MJ, Nissen JB, Mors O, Thomsen PH. Age and gender differences in depressive symptomatology and comorbidity: an incident sample of psychiatrically admitted children. J Affect Disord. 2005;84(1):85–91.
5. Weersing VR, Gonzalez A, Campo JV, Lucas AN. Brief behavioral therapy for pediatric anxiety and depression: piloting an integrated treatment approach. Cogn Behav Pract. 2008;15(2):126–39.
6. Nigatu YT, Liu Y, Uppal M, Mckinney S, Rao S, Gillis K, et al. Interventions for enhancing return to work in individuals with a common mental illness: systematic review and meta-analysis of randomized controlled trials. Psychol Med. 2016;46(16):3263–74.
7. Asselmann E, Wittchen H-U, Lieb R, Beesdo-Baum K. Sociodemographic, clinical, and functional long-term outcomes in adolescents and young adults with mental disorders. Acta Psychiatr Scand. 2018;137(1):6–17.
8. de Lijster JM, Dieleman GC, Utens EMWJ, Dierckx B, Wierenga M, Verhulst FC, et al. Social and academic functioning in adolescents with anxiety disorders: a systematic review. J Affect Disord. 2018;230:108–17.
9. Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. J Clin Child Adolesc Psychol. 2016;45(2):91–113.
10. Weersing VR, Jeffreys M, Do M-CT, Schwartz KTG, Bolano C. Evidence base update of psychosocial treatments for child and adolescent depression. J Clin Child Adolesc Psychol. 2017;46(1):11–43.
11. Gutermann J, Schwartzkopff L, Steil R. Meta-analysis of the long-term treatment effects of psychological interventions in youth with PTSD symptoms. Clin Child Fam Psychol Rev. 2017;20(4):422–34.
12. Kendall PC, Peterman JS. CBT for adolescents with anxiety: mature yet still developing. Am J Psychiatry. 2015;172(6):519–30.
13. James AC, James G, Cowdrey FA, Soler A, Choke A. Cognitive behavioural therapy for anxiety disorders in children and adolescents. Cochrane Database Syst Rev. 2013. https://doi.org/10.1002/14651858.cd013162.
14. Weisz JR, Donenberg GR, Han SS, Weiss B. Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. J Consult Clin Psychol. 1995;63(5):688–701.
15. Weisz JR, Ugueto AM, Cheron DM, Herren J. Evidence-based youth psychotherapy in the mental health ecosystem. J Clin Child Adolesc Psychol. 2013;42(2):274–86.
16. Hunsley J, Lee CM. Research-informed benchmarks for psychological treatments: efficacy studies, effectiveness studies, and beyond. Prof Psychol Res Pract. 2007;38:21–33.
17. Santucci LC, Thomassin K, Petrovic L, Weisz JR. Building evidence-based interventions for the youth, providers, and contexts of real-world mental-health care. Child Dev Perspect. 2015;9(2):67–73.
18. Wergeland GJH, Riise EN, Öst L-G. Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: a systematic review and meta-analysis. Clin Psychol Rev. 2021;83:101918.
19. Queen AH, Barlow DH, Ehrenreich-May J. The trajectories of adolescent anxiety and depressive symptoms over the course of a transdiagnostic treatment. J Anxiety Disord. 2014;28(6):511–21.
20. Garber J, Weersing VR. Comorbidity of anxiety and depression in youth: implications for treatment and prevention. Clin Psychol Sci Pract. 2010;17(4):293–306.
21. Olatunji BO, Naragon-Gainey K, Wolitzky-Taylor KB. Specificity of rumination in anxiety and depression: a multimodal meta-analysis. Clin Psychol Sci Pract. 2013;20(3):225–57.
22. Aldao A, Nolen-Hoeksema S, Schweizer S. Emotion-regulation strategies across psychopathology: a meta-analytic review. Clin Psychol Rev. 2010;30(2):217–37.
23. Brown TA, Chorpita BF, Barlow DH. Structural relationships among dimensions of the DSM-IV anxiety and mood disorders and dimensions of negative affect, positive affect, and autonomic arousal. J Abnorm Psychol. 1998;107(2):179–92.
24. Trosper SE, Whitton SW, Brown TA, Pincus DB. Understanding the latent structure of the emotional disorders in children and adolescents. J Abnorm Child Psychol. 2012;40(4):621–32.
25. Kendall PC, Safford S, Flannery-Schroeder E, Webb A. Child anxiety treatment: outcomes in adolescence and impact on substance use and depression at 7.4-year follow-up. J Consult Clin Psychol. 2004;72(2):276–87.
26. Weisz JR, McCarty CA, Valeri SM. Effects of psychotherapy for depression in children and adolescents: a meta-analysis. Psychol Bull. 2006;132(1):132–49.
27. McHugh RK, Barlow DH. The dissemination and implementation of evidence-based psychological treatments: a review of current efforts. Am Psychol. 2010;65(2):73–84.
28. Emmelkamp PMG, David D, Beckers T, Muris P, Cuijpers P, Lutz W, et al. Advancing psychotherapy and evidence-based psychological interventions: advancing psychotherapy. Int J Methods Psychiatr Res. 2014;23(S1):58–91.
29. Barlow DH, Allen LB, Choate ML. Toward a unified treatment for emotional disorders. Behav Therapy. 2004;35(2):205–30.
30. Cook SC, Schwartz AC, Kaslow NJ. Evidence-based psychotherapy: advantages and challenges. Neurotherapeutics. 2017;14(3):537–45.
31. Mansell W, Harvey A, Watkins ER, Shafran R. Cognitive behavioral processes across psychological disorders: a review of the utility and validity of the transdiagnostic approach. Int J Cogn Ther. 2008;1(3):181–91.
32. Ellard KK, Fairholme CP, Boisseau CL, Farchione TJ, Barlow DH. Unified protocol for the transdiagnostic treatment of emotional disorders: protocol development and initial outcome data. Cogn Behav Pract. 2010;17(1):88–101.
33. Bilek EL, Ehrenreich-May J. An open trial investigation of a transdiagnostic group treatment for children with anxiety and depressive symptoms. Behav Ther. 2012;43(4):887–97.
34. Ehrenreich-May J, Rosenfield D, Queen AH, Kennedy SM, Remmes CS, Barlow DH. An initial waitlist-controlled trial of the unified protocol for the treatment of emotional disorders in adolescents. J Anxiety Disord. 2017;46:46–55.

35. Kennedy SM, Bilek EL, Ehrenreich-May J. A randomized controlled pilot trial of the unified protocol for transdiagnostic treatment of emotional disorders in children. Behav Modif. 2019;43(3):330–60.
36. Jensen-Doss A, Ehrenreich-May J, Nanda MM, Maxwell CA, LoCurto J, Shaw AM, et al. Community study of outcome monitoring for emotional disorders in teens (COMET): a comparative effectiveness trial of a transdiagnostic treatment and a measurement feedback system. Contemp Clin Trials. 2018;74:18–24.
37. Weersing VR, Brent DA, Rozenman MS, Gonzalez A, Jeffreys M, Dickerson JF, Lynch FL, Porta G, Iyengar S. Brief behavioral therapy for pediatric anxiety and depression in primary care: a randomized clinical trial. JAMA Psychiat. 2017;74(6):571–8. https://doi.org/10.1001/jamapsychiatry.2017.0429.
38. Brent DA, Porta G, Rozenman MS, Gonzalez A, Schwartz KTG, Lynch FL, Dickerson JF, Iyengar S, Weersing VR. Brief behavioral therapy for pediatric anxiety and depression in primary care: a follow-up. J Am Acad Child Adolesc Psychiatry. 2020;59(7):856–67. https://doi.org/10.1016/j.jaac.2019.06.009.
39. Chu BC, Crocco ST, Esseling P, Areizaga MJ, Lindner AM, Skriner LC. Transdiagnostic group behavioral activation and exposure therapy for youth anxiety and depression: initial randomized controlled trial. Behav Res Ther. 2016;76:65–75. https://doi.org/10.1016/j.brat.2015.11.005.
40. Martinsen KD, Rasmussen L, Wentzel-Larsen T, Holen S, Sund AM, Løvaas M, Patras J, Kendall PC, Waaktaar T, Neumer SP. Prevention of anxiety and depression in school children: effectiveness of the transdiagnostic EMOTION program. J Consult Clin Psychol. 2019;87(2):212–9. https://doi.org/10.1037/ccp0000360.
41. Guzick AG, Leong AW, Dickinson EM, Schneider SC, Zopatti K, Manis J, Meinert AC, Barth AM, Perez M, Campo DM, Weinzimmer SA, Cepeda SL, Mathai D, Shah A, Goodman WK, Salloum A, Kennedy S, Ehrenreich-May J, Storch EA. Brief, parent-led, transdiagnostic cognitive-behavioral teletherapy for youth with emotional problems related to the COVID-19 pandemic. J Affect Disord. 2022;301:130–7. https://doi.org/10.1016/j.jad.2022.01.034.
42. Neumer SP, Junge-Hoffmeister J. SMART: Forebygging av emosjonelle problemer hos ungdom: veileder (in Norwegian). Oslo: Kommuneforlaget; 2010.
43. Lorentzen V, Fagermo K, Handegård BH, Skre I, Neumer SP. A randomized controlled trial of a six-session cognitive behavioral treatment of emotional disorders in adolescents 14–17 years old in child and adolescent mental health services (CAMHS). BMC Psychol. 2020;8(1):25.
44. Ginsburg GS, Becker EM, Keeton CP, Sakolsky D, Piacentini J, Albano AM, et al. Naturalistic follow-up of youths treated for pediatric anxiety disorders. JAMA Psychiat. 2014;71(3):310–8.
45. Kennard BD, Silva SG, Tonev S, Rohde P, Hughes JL, Vitiello B, et al. Remission and recovery in the treatment for adolescents with depression study (TADS): acute and long-term outcomes. J Am Acad Child Adolesc Psychiatry. 2009;48(2):186–95.
46. Rith-Najarian LR, Mesri B, Park AL, Sun M, Chavira DA, Chorpita BF. Durability of cognitive behavioral therapy effects for youth and adolescents with anxiety, depression, or traumatic stress: a meta-analysis on long-term follow-ups. Behav Ther. 2019;50(1):225–40.
47. Copeland WE, Angold A, Shanahan L, Costello EJ. Longitudinal patterns of anxiety from childhood to adulthood: the great smoky mountains study. J Am Acad Child Adolesc Psychiatry. 2014;53:21–33.
48. Reynolds S, Wilson C, Austin J, Hooper L. Effects of psychotherapy for anxiety in children and adolescents: a meta-analytic review. Clin Psychol Rev. 2012;32:251–62.
49. Sherman ML, Barnum DD, Buhman-Wiggs A, Nyberg E. Clinical intake of child and adolescent consumers in a rural community mental health center: Does wait-time predict attendance? Community Ment Health J. 2009;45(1):78–84. https://doi.org/10.1007/s10597-008-9153-8.
50. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. J Child Psychol Psychiatry. 2000;41(5):645–55.
51. Goodman A, Heiervang E, Collishaw S, Goodman R. The, "DAWBA bands" as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples. Soc Psychiatry Psychiatr Epidemiol. 2011;46(6):521–32.
52. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337–45.
53. Rønning JA, Mørch W-T, Handegaard BH, Sourander A. The Strengths and difficulties self-report questionnaire as a screening instrument in Norwegian community samples. Eur Child Adolesc Psychiatry. 2004;13(2):73–82.
54. Bøe T, Hysing M, Skogen JC, Breivik K. The strengths and difficulties questionnaire (SDQ): factor structure and gender equivalence in Norwegian adolescents. PLoS ONE. 2016;11(5):e0152202–e0152202.
55. Muris P, Meesters C, Van den Berg F. The strengths and difficulties questionnaire (SDQ) further evidence for its reliability and validity in a community sample of Dutch children and adolescents. Eur Child Adolesc Psychiatry. 2003;12(1):1–8.
56. Shaffer D, Gould MS, Brasic J, Fisher P, Aluwahlia S, Bird H. A children's global assessment scale (CGAS). Arch Gen Psychiatry. 1983;40(11):1228–31.
57. Schorre BEH, Vandvik IH. Global assessment of psychosocial functioning in child and adolescent psychiatry: a review of three unidimensional scales (CGAS, GAF, GAPD). Eur Child Adolesc Psychiatry. 2004;13(5):273–86.
58. Evans C, Mellor-Clark J, Frank M. CORE: clinical outcomes in routine evaluation. J Mental Health. 2000;9(3):247–55.
59. Barkham M, Evans C, Margison F. The rationale for developing and outcome batteriesfor routine use in service settings and psychotherapy outcome research implementing core. J Ment Health. 1998;7(1):35–47.
60. Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C, et al. Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. J Consult Clin Psychol. 2001;69(2):184–96.
61. Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, et al. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE–OM. Br J Psychiatry. 2002;180(1):51–60.
62. Skre I, Friborg O, Elgarøy S, Evans C, Myklebust LH, Lillevoll K, et al. The factor structure and psychometric properties of the clinical outcomes in routine evaluation-outcome measure (CORE-OM) in Norwegian clinical and non-clinical samples. BMC Psychiatry. 2013;13(1):99.
63. Elfström ML, Evans C, Lundgren J, Johansson B, Hakeberg M, Carlsson SG. Validation of the Swedish version of the clinical outcomes in routine evaluation outcome measure (CORE-OM): validation of the Swedish version of CORE-OM. Clin Psychol Psychother. 2012;20:447–55.
64. Kristjánsdóttir H, Sigurðsson BH, Salkovskis P, Ólason D, Sigurdsson E, Evans C, et al. Evaluation of the psychometric properties of the Icelandic Version of the clinical outcomes in routine evaluation-outcome measure, its transdiagnostic utility and cross-cultural validation: psychometric properties and transdiagnostic utility of the Icelandic CORE-OM. Clin Psychol Psychother. 2015;22(1):64–74.
65. Palmieri G, Evans C, Hansen V, Brancaleoni G, Ferrari S, Porcelli P, et al. Validation of the Italian version of the clinical outcomes in routine evaluation outcome measure (CORE-OM). Clin Psychol Psychother. 2009;16(5):444–9.
66. Lorentzen V, Handegård BH, Moen CM, Solem K, Lillevoll K, Skre I. CORE-OM as a routine outcome measure for adolescents with emotional disorders: Factor structure and psychometric properties. BMC Psychol. 2020;8(1):86.
67. Beck AT, Steer RA, Brown GK. Beck depression inventory–II. San Antonio. 1996;78(2):490–8.
68. Osman A, Downs WR, Barrios FX, Kopper BA, Gutierrez PM, Chiros CE. Factor structure and psychometric characteristics of the Beck depression inventory-II. J Psychopathol Behav Assess. 1997;19(4):359–76.
69. Steer RA, Kumar G, Ranieri WF, Beck AT. Use of the Beck depression inventory-II with adolescent psychiatric outpatients. J Psychopathol Behav Assess. 1998;20(2):127–37.
70. Aasen H. An empirical investigation of depression symptoms: norms, psychometric characteristics and factor structure of the Beck depression inventory-II. Bergen: The University of Bergen; 2001.
71. March JS, Parker JDA, Sullivan K, Stallings P, Conners CK. The multidimensional anxiety scale for children (MASC): factor structure, reliability, and validity. J Am Acad Child Adolesc Psychiatry. 1997;36(4):554–65.
72. Rynn MA, Barber JP, Khalid-Khan S, Siqueland L, Dembiski M, McCarthy KS, et al. The psychometric properties of the MASC in a pediatric psychiatric sample. J Anxiety Disord. 2006;20(2):139–57.

73. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59(1):12–9.
74. Singer JD, Willett JB. Applied longitudinal data analysis: modeling change and event occurrence. Oxford: Oxford University Press; 2009. p. 1–644.
75. Feingold A. A regression framework for effect size assessments in longitudinal modeling of group differences. Rev Gen Psychol J Div Am Psychol Assoc. 2013;17(1):111–21.
76. Cohen J. Statistical power analysis for the behavioral sciences. New York: Routledge; 1988.
77. Gibby BA, Casline EP, Ginsburg GS. Long-term outcomes of youth treated for an anxiety disorder: a critical review. Clin Child Fam Psychol Rev. 2017;20:201–25.
78. Villabø MA, Narayanan M, Compton SN, Kendall PC, Neumer SP. Cognitive-behavioral therapy for youth anxiety: an effectiveness evaluation in community practice. J Consult Clin Psychol. 2018;86(9):751–64.
79. Stice E, Rohde P, Gau JM, Wade E. Efficacy trial of a brief cognitive-behavioral depression prevention program for high-risk adolescents: effects at 1- and 2-year follow-up. J Consult Clin Psychol. 2010;78(6):856–67.
80. Cayirli T, Veral E. Outpatient scheduling in health care: a review of literature. Prod Oper Manag. 2003;12:519–49.
81. Oudhoff JP, Timmermans DRM, Bijnen AB, Van Der Wal G. Waiting for elective general surgery: physical, psychological and social consequences. ANZ J Surg. 2004;74(5):361–7.
82. Oudhoff JP, Timmermans DRM, Knol DL, Bijnen AB, Van der Wal G. Waiting for elective surgery: effect on physical problems and postoperative recovery. ANZ J Surg. 2007;77(10):892–8.
83. Sampalis J, Boukas S, Liberman M, Reid T, Dupuis G. Impact of waiting time on the quality of life of patients awaiting coronary artery bypass grafting. CMAJ. 2001;165(4):429–33.
84. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev. 2015;40:91–110.

Publisher's Note