Screening and diagnostics in child and adolescent psychiatry

Andersson, Markus

2019

Document Version:
Förlagets slutgiltiga version

Screening and diagnostics in child and adolescent psychiatry

MARKUS ANDERSSON
DEPARTMENT OF CLINICAL SCIENCES, LUND | LUND UNIVERSITY

Screening and diagnostics in child and adolescent psychiatry

Screening and diagnostics in child and adolescent psychiatry

Markus Andersson

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Baravägen 1 Lund on 25th May 2019 at 9.00

Faculty opponent
Jan-Olov Larsson
Department of Women's and Children's Health
Karolinska Institutet Stockholm, Sweden

Organization LUND UNIVERSITY Faculty of Medicine Department of Clinical Sciences, Lund, Sweden Author: Markus Andersson	Document name Doctoral Dissertation Series 2019:54 Date of issue 2019-05-25 Sponsoring organization

Title and subtitle: Screening and diagnostics in child and adolescent psychiatry

Abstract

Background

At least ten percent of children and adolescents suffer from mental disorders with significant impairment in their daily life. There are reasonably effective treatments for many disorders but Child and Adolescent Mental Health Services (CAMHS) needs an effective screening procedure for referrals followed by a reliable diagnostic procedure in order to offer appropriate treatment. The Brief Child and Family Phone Interview (BCFPI) is used for intake screening, the Screen for Child Anxiety Related and Emotional Disorders Revised (SCARED-R) for further screening focused on anxiety, and the Kiddie-Schedule for Affective Disorders and Schizophrenia for School Aged Children-Previous and Lifetime (K-SADS-PL) is a semi-structured diagnostic interview used in the diagnostic procedure.

Objective

The intent of this thesis was to evaluate the BCFPI, the SCARED-R, and the K-SADS-PL in an outpatient setting. The primary aim was to evaluate the factor structure and the validity of the six symptom subscales of the BCFPI. Secondary aims were to evaluate the validity of the K-SADS-PL and the validity of the SCARED-R.

Method

The factor structure of the Swedish version of BCFPI was evaluated by examining BCFPI interviews collected at four CAMHS as part of their standard intake procedure and comparing to the factor structure of the original English version. The validity of the BCFPI, the SCARED-R and the K-SADS-PL was evaluated by comparing them to diagnoses elicited from a LEAD (Longitudinal Expert All Data) procedure in newly admitted unselected outpatients.

Results

The Swedish version of the BCFPI had a factor structure that was almost the same as the original English version. There was no major variability in factor structure or item intercepts between boys and girls, between children and adolescents, or between native Swedish children and those with parents born abroad. The predictive validity of diagnoses elicited by K-SADS-PL compared to LEAD diagnoses was good to very good for most child psychiatric diagnoses except for autism spectrum disorder. The criterion validity for the six symptom subscales of the BCFPI was fair to good compared to the corresponding LEAD diagnoses. The criterion validity of SCARED-R parent report was fair to good for most anxiety disorders compared to LEAD diagnoses. The parent report was overall more valid than the self-report.

Conclusions

The Swedish version of the BCFPI is reliable as a screening measure for the major child psychiatric disorders in both genders, in the age span 6-17 years and even with parents who speak Swedish but have a different native language. The K-SADS-PL diagnoses elicited at one visit with a well-trained and supervised clinician have good to very good validity for most child psychiatric diagnoses. The SCARED-R is reliable as a screening measure for anxiety disorders in the parent version while the patient version adds less value in a clinical environment with significant comorbidity.

Key words Screening, child psychiatric diagnoses, K-SADS-PL, BCFPI, SCARED-R, LEAD		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN and key title 1652-8220 Screening and diagnostics in child and adolescent psychiatry		ISBN 978-91-7619-783-7
Recipient's notes	Number of pages 72	Price
	Security classification	

Screening and diagnostics in child and adolescent psychiatry

Markus Andersson

LUND
UNIVERSITY

Coverphoto by Rouna

Department of Clinical Sciences, Lund
Faculty of Medicine
Lund University

ISBN 978-91-7619-783-7
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2019

Table of Contents

Svensk sammanfattning

Bakgrund

Ungefär 10 % av barn och ungdomar lider av psykisk ohälsa som på ett påtagligt sätt påverkar hur de fungerar i skolan, hemmet, socialt och på fritiden. Det finns idag effektiva behandlingsalternativ för flertalet barnpsykiatriska problemområden. För att barnen ska få tillgång till behandling krävs det att vården har en effektiv screeningprocedur så att de familjer som söker hjälp kan lotsas till rätt insats. Det behövs också en tillförlitlig diagnostisk procedur för att adekvat behandling ska kunna sättas in. Brief Child and Family Phone Interview (BCFPI) är en screeningintervju som används då familjerna kommer i kontakt med vården. Screen for Child Anxiety Related and Emotional Disorders Revised (SCARED-R) är en skala som screenar för ångestproblematik hos barn. Kiddie-Schedule for Affective Disorders and Schizophrenia for School Aged Children-Previous and Lifetime Version (K-SADS-PL) är en semi-strukturerad diagnostisk intervju som täcker flertalet barnpsykiatriska diagnoser.

Syfte

Det övergripande syftet för denna avhandling har varit att utvärdera BCFPI, K-SADS-PL och SCARED-R i en Barn- och ungdomspsykiatrisk öppenvårdsmiljö. Primärt mål var att utvärdera faktorstrukturen och validiteten hos de sex symptomskalorna i BCFPI. Sekundära mål var att utvärdera validiteten hos K-SADS-PL och SCARED-R.

Metod

Faktorstrukturen i BCFPI utvärderades genom analys av intervjuer gjorda inom fyra Barn- och ungdomspsykiatriska kliniker som en del i deras intagningsprocedur. Dessa jämfördes med faktorstrukturen i den engelska originalversionen. Validiteten hos BCFPI, K-SADS-PL och SCARED-R utvärderades genom att de jämfördes med diagnoser erhållna från en LEAD (Longitudinal Expert All Data) procedur på nya och icke-selekterade patienter inom barnpsykiatrisk öppenvård.

Resultat

Faktorstrukturen i den svenska versionen av BCFPI visade sig vara god och lik faktorstrukturen i den engelska originalversionen. Faktorstrukturen var överlag densamma för pojkar och flickor, barn 6-12 år respektive tonåringar 13-17 år och för barn till föräldrar med svenska eller annat språk som modersmål. Den prediktiva validiteten hos diagnoser erhållna med K-SADS-PL jämfört med LEAD diagnoser var mycket god för flertalet diagnoser, god för ADHD och mindre god för autism spektrum störning. Kriterievaliditeten hos symptomskalorna i BCFPI jämfört med

LEAD diagnoser var rimlig till god. Kriterievaliditeten för SCARED-R jämfört med LEAD diagnoser var rimlig till god för flertalet ångestdiagnoser. Föräldrarnas skattningar var överlag mer valida än barnens egna skattningar.

Diskussion och slutsats

Den svenska versionen av BCFPI kan med tillförlitlighet användas för att screena för de vanliga barnpsykiatriska diagnoserna hos pojkar och flickor i ålder 6-17 år, och oberoende av föräldrarnas modersmål. K-SADS-PL har god till mycket god validitet för flertalet barnpsykiatriska diagnoser om den utförs av en tränad läkare, med möjlighet till konsultation hos erfaren kollega. Föräldraversionen av SCARED-R kan användas för screening av flertalet ångeststörningar hos barn medan barnversionen tillför något mindre i en psykiatrisk öppenvårdsmiljö där många barn lider av samsjuklighet.

Acknowledgements

I first want to thank all families that have participated in these studies and making my research possible. I am also grateful for all the clinicians that have helped me throughout the project.

Special thanks to my friend, colleague and co-supervisor Håkan Jarbin whom I deeply admire both as a professional and as person.

I would like to thank Peik Gustafsson, my main supervisors in the project, and Maria Råstam who was the main supervisor for most of the time that the project was running. Both of you combines knowledge, engagement, friendliness, and humour that have been very helpful for me and sets a good tone for the child psychiatric section in the Faculty of Medicine in Lund.

Further thanks go to Martin Bäckström, Tord Ivarsson, and Gudmundur Skarphedinsson who all were a part of the research group.

I also would like to thank all my friends and colleagues in the child psychiatry in Halland. I am proud to be a part of a workplace with such an engagement in providing good health care to children with mental health problems.

Through the years I have had the privilege to meet and work with a lot of people that have been committed to improving the intake procedure and the diagnostic assessment in child mental health services. For my own work and research in these areas, Lars-Henry Gustle, Annika Nilsson Wendel, and Pia Tallberg, who all works in the child psychiatry in Skåne, have been especially important.

Most of all I want to thank my wife Josefine and my children Olle and Ester for giving me energy and love throughout the years of studies.

This project was supported by grants from Söderström-Königska Foundation, and FOU grants from Region Halland and Region Skåne. The department of child and adolescent within Psychiatry Halland offered generous support to conduct research within my clinical position throughout the years of these studies.

List of papers

This thesis is based on the following papers:

Paper 1 Andersson M, Jarbin H, Rastam M, Backstrom M Construct Validity of the Swedish version of the Brief Child and Family Phone Interview (BCFPI). (Manuscript unpublished)

Paper 2 Jarbin H, Andersson M, Rastam M, Ivarsson T (2017) Predictive validity of the K-SADS-PL 2009 version in school-aged and adolescent outpatients. Nord J Psychiatry 71:270-276

Paper 3 Andersson M, Backstrom M, Ivarsson T, Rastam M, Jarbin H (2018) Validity of the Brief Child and Family Phone Interview by comparison with Longitudinal Expert All Data diagnoses in outpatients. Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, Vol 6(2): 83-90

Paper 4 Ivarsson T, Skarphedinsson G, Andersson M, Jarbin H (2017) The validity of the Screen for Child Anxiety Related Emotional Disorders Revised (SCARED-R) Scale and Sub-Scales in Swedish youth. Child Psychiatry and Human Development, 49(2), 234-243

Abbreviations

ADHD	Attention Deficit Hyperactivity Disorder
AUC	Area Under the Curve
BCFPI	Brief Child and Family Phone Interview
CAMHS	Child and Adolescent Mental Health Services
CAPA	Child and Adolescent Psychiatric Assessment
CBCL	Child and Behaviour Check List
CD	Conduct Disorder
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
DICA	Diagnostic Interview for Children and Adolescents
DISC-IV	Diagnostic Interview Schedule for Children-IV
FSSC-R	Fear Schedule Survey for Children Revised
GAD	Generalized Anxiety Disorder
ISCA	Interview Schedule for Children and Adolescents
KID-SCID	Structured Clinical Interview for DSM-IV Childhood Disorders
K-SADS-PL	Kiddie-Schedule for Affective Disorders and Schizophrenia- Previous and Lifetime version
LEAD	Longitudinal Expert All Data
MDD	Major Depressive Disorder
MINI-KID	Mini International Neuropsychiatric Interview for Children and Adolescents
ML	Maximum Likelihood
NPV	Negative Predictive Value
OCD	Obsessive Compulsive Disorder
OCHS-R	Ontario Child Health Study Revised
ODD	Oppositional Defiant Disorder
PD	Panic Disorder
PDQ-4+	Personality Diagnostic Questionnaire-4+

PPV	Positive Predictive Value
PTSD	Post-Traumatic Stress Disorder
ROC	Receiver Operating Characteristic
RMSEA	Root Mean Square Error of Approximation
SAD	Separation Anxiety Disorder
SCARED-R	Screen for Child and Anxiety Related Emotional Disorder Revised
SDQ	Strengths and Difficulties Questionnaire
SoP	Social Phobia
SP	Specific Phobia
WLSMV	Weighted Least Square corrected for Means and Variances

Introduction

At least ten percent of children and adolescents suffer from mental disorders with significant functional impairment in important domains of everyday life such as family, school, and socializing with peers (Polanczyk, Salum et al. 2015). This leads to extensive suffering and costs not only for the child and her/his family but also for society. In 2010 an estimate of the total cost of all youth mental disorders in Europe was 21.3 billion € (Olesen, Gustavsson et al. 2012). Half of all lifetime cases of mental illness have begun by age 14. Untreated mental disorders often lead to more severe and difficult-to-treat illnesses, along with the development of comorbid disorders (Kessler, Berglund et al. 2005). There are reasonably effective treatments for children with many of the psychiatric disorders such as depression (Goodyer, Reynolds et al. 2017, Stevanovic, Tadic et al. 2014) anxiety disorders (Wehry, Beesdo-Baum et al. 2015, Walkup, Albano et al. 2008), attention deficit hyperactivity disorder (ADHD) (Catala-Lopez, Hutton et al. 2017, Fernandez de la Cruz, L., Simonoff et al. 2015), and disruptive behaviour disorders (oppositional defiant disorder (ODD) and conduct disorder (CD)) (Connor, Glatt et al. 2002, Hood, Elrod et al. 2015). The individual, social and economic cost for mental health problems may be reduced with early treatment. Therefore methods that target early detection, assessment and treatment of psychiatric illness are important. This, together with the need to manage the increasing demand for mental health services (Socialstyrelsen 2017) calls first, for an effective screening procedure for referrals at intake to CAMHS and second, a reliable diagnostic assessment.

Screening procedure at intake to improve triage to CAMHS

Triage can be defined as the process of determining clinical need, the level of urgency, and the likely response to an intervention (Parkin, Frake et al. 2003). In primary healthcare and emergency care structured triage procedures have led to improved care in terms of a reduction in workload for physicians and nurses, improved accessibility to services, improved patient satisfaction, and better use of employees' skills and expertise (Lake, Georgiou et al. 2017, Harding, Taylor et al. 2011, Martinez-

Gonzalez, Djalali et al. 2014, Martinez-Gonzalez, Rosemann et al. 2015, Buchan, Dal Poz 2002, Cariello 2003)

A standardized screening instrument may improve intake services and triage to CAMHS. At referral to CAMHS children often present a complex pattern of symptoms, functional impairment and risk factors. The intake clinician must decide whether the problems are in accordance with the organization's service mandate or if the child should be directed to another service. The clinician has to establish risk, establish priorities, triage the child/family to a specialized treatment, or suggest interim alternatives while the family waits for services (Cunningham, Boyle et al. 2009). Still, even though standardized measures seem to improve decision making (Galanter, Patel 2005a, Parkin et al. 2003) and make the subsequent assessment more efficient (Hughes, Emslie et al. 2005) most intake decisions are made based on subjective and unstructured interviews (Cunningham et al. 2009).

Intake screening instrument to CAMHS

Broadband measures are suitable for intake screening, briefly covering different problem areas, the daily functioning of the child, and family factors, (Rutter 2008). Broadband screening instruments that are used to improve intake services to CAMHS are the Child Behaviour Check List (CBCL) (Achenbach, Rescorla 2001) the Strength and Difficulties Questionnaire (SDQ) (Deighton, Croudace et al. 2014, Goodman 1999), and the Brief Child and Family Phone Interview (BCFPI) (Cunningham, Pettingill et al. 2006) Broadband measures as well as other scales usually include parent versions as well as teacher reports and self-reports. Although there is often an additive effect of getting information from different sources there are also practical restraints. At least one study suggests that the parent report may be more reliable (Kuhn, Aebi et al. 2017).

The Child Behaviour Check List (CBCL)

The CBCL is a widespread scale in both child psychiatric research and clinical practice. The scale is based on empirical studies of psychiatric or behavioural problems in children and adolescents (Achenbach, Rescorla 2001) and has normative data from large population samples for children aged 6-17 years. The CBCL includes 118 items with three response options, not true, somehow or somewhat true and often true (Achenbach, Dumenci 2001). The items are divided into 8 empirically-based subscales: aggressive behaviour, anxious depressed, attention problems, rule-breaking behaviour, somatic complaints, social problems, thought problems, and withdrawn/depressed. The CBCL also includes two composite subscales: internalizing

problems and externalizing problems. Further, there are three additional subscales "social", "activities" and "school" developed by comparing children referred to mental health services to controls. There are versions of the CBCL for parent, teacher, and self-reports. CBCL was not originally designed for diagnostic purposes (Sadock, B. Sadock, V. Kaplan, H. 2005) but there have been studies supporting the convergence between subscale for attention problems and ADHD (Chen, Faraone et al. 1994) for the subscale for rule-breaking behaviour and CD (Biederman, Monuteaux et al. 2005), for the subscale for aggressive behaviour and ODD (Biederman, Ball et al. 2008). Further, the CBCL includes DSM-oriented scales that were constructed by experts who rated and selected questions in the CBCL that were consistent with DSM disorders and formed the subscales; affective problems, anxiety problems, somatic problems, ADHD, ODD, and CD. The parent report of CBCL has been shown to discriminate between psychiatric patients and non-patients (Achenbach, Rescorla 2001).

Summing up, the CBCL is a valid screening measure. However, it is both long and time consuming but fails to provide information of the family situation that may be useful for triage.

The Strength and Difficulties Questionnaire (SDQ)

The SDQ is a well validated (Deighton et al. 2014, Goodman 1999) measure with good psychometric properties commonly used in both research and clinical practice. The Swedish version of SDQ has been evaluated (Malmberg, Rydell et al. 2003, Smedje, Broman et al. 1999) and has good internal consistency and validity. The SDQ has proven useful for intake screening and triage (Aras, Varol Tas et al. 2014, Jones, Lucey et al. 2000, Lai 2006). The questionnaire was designed with experience from other measures like the CBCL with special attention to keeping it brief. An important study (Goodman, Scott 1999) showed that the parent version of SDQ was at least as good as the more comprehensive CBCL at discriminating between psychiatric and non-psychiatric cases and was preferred by the parents. The SDQ was designed for parents, teachers and for self-reports. The self-report covered children at the ages 4-16 but later evidence has shown that it can be used up to 17 years. The parent and teacher report are better than the self-report at separating children with clinical problems from controls. The SDQ performs best in screening for behavioural syndromes such as ADHD and less well for emotional problems e.g. anxiety (Malmberg et al. 2003).

The SDQ consists of 25 items in five subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relation problems and prosocial behaviour. There is also an extended version with an additional impact supplement that facilitates determination of service assignment (Goodman 1999).

There are few, but promising, studies of using SDQ as a screening instrument to improve the triage to CAMHS (Parkin et al. 2003). In a study from Jones and colleagues (Jones et al. 2000) they evaluated a pilot project in CAMHS of a triage style assessment where families, after completing the SDQ, were invited to an appointment with clinicians to screen for the presenting problem in order to arrive at a triage decision. The possible decisions from triage were immediate allocation, further assessment, priority waiting list, routine waiting list, and closed.

The triage process had several aims: 1, to shorten the waiting time from referral to first assessment; 2, to assess more fully and accurately the reason for referral to enable better judgment of urgency, treatability and appropriateness of the referral, and by that to improve management planning; 3, to improve attendance rate at first appointments by reducing the waiting time; 4, to prevent deterioration of function due to prolonged waiting time. The triage led to improved satisfaction to patients and clinicians as well as improved the attendance rate at first appointment compared to the ordinary intake procedure.

In another study (Aras et al. 2014), SDQ plus an extra 20 minutes phone interview, followed by a short interview with a specialist in child psychiatry were used for triaging to the appropriate services. The screening procedure led to a shorter waiting list and increased access to early intervention.

In another study concerning screening at intake to CAMHS (Lai 2006) SDQ was used together with an additional phone triage interview conducted by nurses to guide families to the proper services. In this case the screening procedure also led to more efficient care.

To sum up, SDQ is a valid screening measure and can be used for intake screening. However, it needs additional information from a phone interview with a clinician or a short appointment at the clinic, to improve triage procedures to CAMHS. Another drawback is that it gives no information of basic family factors that are important for triage.

The Brief Child and Family Phone Interview (BCFPI)

The Brief Child and Family Phone Interview, BCFPI, is a semi-structured, computer-assisted, clinical intake and follow-up telephone interview with parents of children, 6 to 17 years, seeking psychiatric healthcare (Cunningham et al. 2009). The BCFPI may also be administered as a pen and paper checklist or an online version. There are also a youth version and a teacher report (Cunningham et al. 2006)

BCFPI has a broad approach targeting symptoms, child and family functioning and risk factors.

The interview begins by a narrative overview of basic problems as seen by the parent.

The BCFPI consists of six symptom subscales; regulation of attention, impulsivity and activity, cooperation, conduct, separation, managing anxiety, and managing mood, that correspond to DSM-IV diagnoses of ADHD, ODD, CD, separation anxiety disorder (SAD), generalized anxiety disorder (GAD), and major depressive disorder (MDD). All subscales consist of six questions with the response options never, sometimes or often. The subscales for ADHD, ODD, and CD form a composite subscale for externalizing problems and the subscales for SAD, GAD, and MDD form a composite subscale for internalizing problems (Cunningham et al. 2006).

Further, there are subscales on child functioning, family situation, family functioning, and informant mood. In addition there are optional questions that clinics may use in their intake service to improve triaging e.g. for tics, obsessive compulsive behaviour, autism spectrum disorder, bullying, drugs, neglect and abuse.

The BCFPI has been standardized for two age groups, 6-12 years and 13-17 years and separately for boys and girls. The results of the BCFPI are summarized as t-scores compared to population norms (Cunningham et al. 2006).

Background of BCFPI

The BCFPI was developed in Canada by adapting the Revised Ontario Child Health Study (OCHS-R) scales (Boyle, Offord et al. 1993) that were created for epidemiological and longitudinal surveys of mental health in children (Offord, Boyle et al. 1987). Questions of symptoms, child and family functioning were selected from the OCHS-R. The items within the six symptom subscales were selected to correspond to DSM-IV diagnoses (Cunningham et al. 2009).

Previous evaluations of the BCFPI

In Canada the BCFPI has been evaluated on several occasions and in different settings. Overall, it has shown good reliability and validity (Boyle, Cunningham et al. 2009, Cunningham et al. 2009, Cuthbert, St. Pierre et al. 2011, Cunningham et al. 2006, Cook, Leschied et al. 2013a). Internal consistency for subscales in field trials ranges from 0.75 to 0.85 (except for CD; Cronbach's alfa 0.68). The BCFPI has also shown good test retest reliability as well as sensitivity to change and the interview has been evaluated in different cultural groups (Bova 2006).

A confirmatory factor analysis measuring invariance across age and sex in BCFPI (Cunningham et al 2009) was performed using three different Canadian samples, a clinical sample, a community sample, and an implementation sample from a region that started up BCFPI as an intake interview to CAMHS. The BCFPI showed good internal consistency in all three samples and the item structure of the six symptom subscales was supported (Cunningham et al. 2009).

Boyle and colleagues (Boyle et al. 2009) evaluated the concurrent validity of the BCFPI in comparison to the Diagnostic Interview Schedule for Children version IV (DISC-IV). DISC-IV is a well-established structured diagnostic interview with good psychometric properties (Shaffer, Fisher et al. 2000). Parents' to 399 children and adolescents referred to outpatient mental health services were administered the BCFPI at baseline, at two months and again at 13 months. In addition they were assessed with DISC-IV at one and 12 months. Internal consistency exceeded 0.80 for all six symptom subscales. Test-retest reliability ranged from 0.45-0.62. The concurrent validity of the BCFPI subscales compared to the DISC-IV was slightly better for externalizing than for internalizing disorders. Kappa coefficients were moderate (0.40 to 0.49) for ADHD, ODD, and CD and fair (0.28 to 0.37) for SAD, GAD, and MDD. Area under the curve (AUC) was good for ADHD, ODD, CD, and SAD (0.81-0.86) and fair for GAD and MDD (0.75-0.76) (Boyle et al. 2009).

The BCFPI has also been evaluated in an inpatient setting where 227 children and adolescents were studied (Cook et al. 2013a). In that study the symptom subscales of BCFPI correlated moderate to strong to their counterparts in Conners' Rating Scales supporting the convergent validity of the BCFPI.

Further the BCFPI has been shown to be useful to evaluate treatment outcomes in a variety of settings (Moretti, Obsuth et al. 2015, Moretti, Obsuth 2009, Gordon, Antshel et al. 2006, Cuthbert et al. 2011).

The BCFPI in a Swedish setting

The BCFPI is one of the most used intake screening measures to CAMHS in Sweden. It is currently used in the regions of Skåne, Halland, Kronoberg, Örebro, Sörmland, Norrbotten, and Gävleborg. The BCFPI was translated in 2003 as part of a project to improve methodology in Swedish child and adolescent psychiatry. The interview is usually performed in CAMHS centralized intake triage units. The interview is part of a triage procedure aiming at guiding referrals to either specialized mental health care, primary care, or services beyond healthcare such as community based programs, school health, and self-help programs.

Features of the BCFPI may have facilitated the spread of the interview in Sweden. In addition to a screening instruments classification efficiency the utility of an instrument also depends on how easy it is to integrate in the service delivery, how simple it is for clinicians to acquire the information obtained from the instrument and how committed the clinicians are to the screening process (Maruish 2004).

First, the interface between specialized care and primary care is built upon diagnoses, functional impairment as well as family risk factors, areas covered by the BCFPI.

Second, the BCFPI is integrated in software that allows on-line computation of t-scores, immediate feedback to interviewers, and aggregate organizational reports, which make the information easily accessible by clinicians (Cunningham et al. 2009).

Third, the BCFPI is well liked by both clinicians and parents in a Swedish setting (Carlberg 2010) which may facilitate implementation (Grol 2013).

Summing up, BCFPI seems to be a valid measure that may be used for intake screening to improve triage. However, the psychometrics of the Swedish version has not been evaluated and the interview has not been compared to a full diagnostic workup.

The Screen for Child and Anxiety Related Emotional Disorders Revised (SCARED-R)

In CAMHS that are specialized in one specific diagnostic domain a more narrow rating scale can be useful and give additive information to the broadband measures. Concerning anxiety disorders the Screen for Child and Anxiety Related Emotional Disorders Revised (SCARED-R) (Muris, Steerneman 2001a), which is evaluated in this thesis, is commonly used.

The SCARED-R was developed to screen for anxiety disorders in children and adolescents. The first version, SCARED, consists of 38 items divided into 5 subscales where 4 correspond to DSM-IV diagnoses, panic disorder (PD), social anxiety disorder (SoP), generalized anxiety disorder (GAD), and separation anxiety disorder (SAD). In the revised version developed by Muris and colleagues (Muris, Steerneman 2001a) subscales for obsessive compulsive disorder (OCD), post-traumatic stress disorder (PTSD), and specific phobia (SP) were added. There are both a parent report and a self-report.

The SCARED-R has shown mostly good psychometric properties with adequate internal consistency and sufficient test-retest stability (Muris, Merckelbach et al. 1999). In a study by Muris and colleagues (Muris, Dreessen et al. 2004) the SCARED-R was compared to a structured diagnostic interview and demonstrated satisfactory discriminant validity both between anxiety disorders and other problems and within anxiety disorders. Further the SCARED-R scores showed correlation to the CBCL internalizing subscale and not with the externalizing subscale. The child-parent agreement for SCARED-R subscales have been high for children with anxiety disorders (Rs between 0.49-0.79) (Muris et al. 2004) and less so in normal school children (Muris et al. 1999). The Swedish version of SCARED-R was translated 2008 to be included in an OCD treatment study (Thomsen, Torp et al. 2013).

Summing up, the SCARED-R is a well-established scale, with generally good psychometric properties, for the screening of child anxiety disorders. However, the Swedish version has not yet been evaluated.

Diagnostic assessment

A classification of disorders in child and adolescent psychiatry is of paramount importance to enhance research and treatment and to enable prediction and explanations. The diagnostic system and the diagnoses need to be both valid, i.e. a clear description of a meaningful entity concerning course and treatment, and reliable, i.e. can be reproduced by different observers and in different settings (Volkmar 1996) Advances in psychiatric taxonomy took place in the 19th century particularly by Kraepelin and were later improved by the DSM-system, where experts determined valid categories and the specific criteria for making psychiatric diagnoses. DSM-III in 1980 included child specific disorders (Schwab-Stone, Shaffer et al. 1996). Psychiatric diagnoses rely on information and observation in the absence of objective measures. Thus, there is a range of fallacies which clinicians can fall victim to when determining the diagnostic categories, ranging from an unsystematic approach to the way questions are asked and answers interpreted by the clinician. The caseness is depending both on the criteria and the thresholds for impairment. Many childhood symptoms may represent transient adaptive states rather than a distinctive psychopathology (McClellan, Werry 2000).

The diagnostic assessment is typically based on a diagnostic interview. There are three approaches to clinical diagnostic interviews which, to different degrees, are vulnerable to the possible fallacies in diagnostics. First the unstructured interview, second, the highly structured or respondent based interview, third, the semi-structured or investigator-based interview (Leffler, Riebel et al. 2015).

The unstructured diagnostic interview

The unstructured interview is the interview style most used in clinical practice despite its vulnerability to diagnostic fallacies. The interviews have no guidelines for scoring and rely on the clinicians' knowledge. The quality of the interview and what areas are covered may vary due to the clinicians experience and education and thus has low interrater reliability (Leffler et al. 2015). Unstructured diagnostic interviews have been shown to have only low to moderate agreement with a diagnostic procedure based on structured diagnostic interviews (Rettew, Lynch et al. 2009).

There are a number of possible fallacies in the diagnostic assessment that are linked to the diagnostician. First, there is a tendency to determine diagnoses before all relevant

information is collected, second, the selection of diagnostic information may be influenced by confirmation biases and ignoring information that rules out a diagnosis, third, a lack of systematic approach of combining different types of information, fourth, a tendency to make diagnoses based on what is most familiar to the clinician, fifth, the tendency to see correlations that do not exists or miss real correlations (McClellan, Werry 2000).

The unstructured interview is also vulnerable to problems in diagnostic assessment that concern the informant's way of reporting. Merten and colleagues (Merten, Cwik et al. 2017) highlight problems in the diagnostic assessment with informants, teachers and parents, using heuristics. For example it has been shown that teachers rate children too high on hyperactivity when the child also exhibits symptoms of ODD. This means that a halo effect, a cognitive bias where factors that seem important for a decision influence all other information taken into the decision making process (Abikoff, Courtney et al. 1993, Jackson, King 2004). It has also been shown that there is a tendency to over report oppositional behaviour in boys compared to girls (Bruchmuller, Margraf et al. 2012). Parents with higher education tend to report more symptoms of inattention in the child than parents of lower education which is not the case for symptoms of hyperactivity (Weckerly, Aarons et al. 2005).

In summation, the unstructured interview is problematic since to a large degree it depends on the clinician that performs the assessment and therefore is vulnerable to common fallacies in the diagnostic assessment.

Respondent-based interviews, highly structured diagnostic interviews

The respondent-based diagnostic interview is a highly structured diagnostic interview. It has many advantages compared to unstructured interviews in achieving valid diagnoses since it is less vulnerable to variation in the diagnostic procedure. Structured diagnostic interviews were developed to minimize information variance and biases coupled to clinical judgment. Typically the interviewer is trained in reading the questions verbatim and the informant is asked to respond whether the symptom/behaviour is present or absent (Rutter 2008). There is strict procedure for reading the questions verbatim with specific rules for scoring and coding responses (Leffler et al. 2015). This provides a high interrater reliability which is a prerequisite for achieving valid diagnoses. Studies have shown that the diagnostic variability decreases when a structured interview is implemented in clinical practise (Hughes, Rintelmann et al. 2000, Galanter, Patel 2005b).

Another advantage with the respondent-based interview is that it does not take an experienced clinician to administer. In research it can be administered by a trained layperson which saves resources (Melzer, Tom et al. 2002). It is also less time consuming, both regarding the training of interviewers as well as the administration

of the interview, than the semi-structured interview, described below, which makes it more appropriate for some clinical practices (Rutter 2008). The structured interview is also a good training tool for inexperienced clinicians (Frick 2016).

On the negative side there is limited flexibility for the clinician to ask clarifying follow-up questions and the quality of the responses are dependent on the informants' conceptual understanding of the symptoms that are considered (Leffler et al. 2015). Further, since the respondent-based interview leaves limited room for clinical judgement there is a risk that the interview leads to inappropriate diagnoses (Shaffer et al. 2000, Boyle, Offord, Racine, Sanford, Szatmari, Fleming, and Price-Munn 1993).

One highly structured, respondent-based interview that is widely spread both in research and clinical practice is the Diagnostic Interview Schedule for Children IV (DISC-IV) (Shaffer et al. 2000).

The Diagnostic Interview Schedule for Children-IV (DISC-IV)

The DISC-IV is a fully structured, respondent-based diagnostic interview that covers 34 child psychiatric diagnoses. The DISC-IV and earlier versions of the DISC have been used in numerous studies as well as in clinical practice (Boyle et al. 2009, McGrath, Handwerk et al. 2004, Schwab-Stone et al. 1996). The interview may be administered either by a layperson, a clinician or by self-report. It takes 90-120 minutes to complete in a clinical setting. There are both parent/teacher forms for the ages 4-17 years and child forms for the ages 9-17 years. The interview was designed to elicit DSM-IV diagnoses by ascertaining the presence or absence of symptoms. The interview includes questions concerning anxiety disorders, mood disorders, disruptive disorders, alcohol and substance use disorders, and miscellaneous disorders e.g. eating disorders, tics. Diagnoses, like pervasive developmental disorders and language disorders that may need clinical observations or interpretation of psychological testing to be established are excluded. The DISC-IV also includes six domains concerning impairment (Shaffer et al. 2000).

The psychometric properties of the DISC-IV, and earlier versions of the DISC, are overall good. Test-retest reliability in the parent version has been good but moderate for the youth version (Schwab-Stone et al. 1996). The reliability of the self-administered computer version is comparable to when the DISC-IV is administered with an interviewer (Lucas 2003). When an earlier version of the DISC was compared to clinical diagnoses established with supervision by senior clinicians, the parent version of the DISC showed good to excellent validity for most disorders (Fisher, Shaffer et al. 1993). In an Argentinian study (Kunst, Blidner et al. 2009) the DISC-IV showed moderate to good agreement for most child psychiatric disorders, except anxiety disorders which showed less agreement, compared to a semi-structured diagnostic interview.

The DISC-IV shares the pros and cons with other fully structured, respondent-based interviews in that it is easy and relatively inexpensive to administer and limits variance but it cannot address invalid responses from the respondent and it does not allow atypical presentations (Shaffer et al. 2000).

Semi-structured interviews, investigator-based interviews

Investigator-based diagnostic interviews are semi-structured and provide some guidelines for the interviewer but still allow flexibility. Compared to respondent-based interviews they are less strict and have fewer restrictions on how to phrase questions leaving room for the interviewer to ensure that the informants have understood the concepts of the questions. The interviewer can choose from suggested verbal probes and can ask opening questions and supplementary questions to gather sufficient information to be able to code the presence and severity of a symptom (Ambrosini 2000, McClellan, Werry 2000). The questions are asked in a flexible but still systematic way and leave room for clinical judgement which is not the case with respondent-based interviews (Leffler et al. 2015). The semi-structured interviews are considerably less vulnerable to the possible fallacies that are present in the unstructured interview. However, leaving room for clinical judgement in semi-structured interviews also leads to a somewhat lower interrater reliability than in respondent-based interviews. Further, the semi-structured interviews require more interviewer skills. The training procedure for the semi-structured interviews is therefore more comprehensive than for the highly structured interviews (McClellan, Werry 2000) making it more time consuming and costly to implement in clinical practice.

There are several comprehensive semi-structured, investigator-based, diagnostic interviews available, the Child and Adolescent Psychiatric Assessment (CAPA) (Angold, Costello 2000), the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID) (Sheehan, Sheehan et al. 2010), the Interview Schedule for Children and Adolescents (ISCA) (Sherrill, Kovacs 2000), the Diagnostic Interview for Children and Adolescents (DICA) (Reich 2000), the Structured Clinical Interview for DSM-IV Childhood Disorders (KID-SCID) (Roelofs, Muris et al. 2015) and the most widely spread namely the Kiddie-Schedule for Affective Disorders and Schizophrenia-Previous and Lifetime (KSADS-PL) (Kaufman, Birmaher et al. 1997). The KSADS-PL is the interview that is evaluated in the present thesis.

Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime (K-SADS-PL)

The K-SADS-PL is an investigator-based, semi-structured diagnostic interview, widely used in both research and clinical settings for children aged 6-17. There are

adult and child versions. The interview was designed by Kaufman et al. in 1996 to assess present and past episodes of psychopathology in children and adolescents (Kaufman et al. 1997). The K-SADS-PL is based on earlier versions of the K-SADS developed in the late seventies and through the eighties (Ambrosini, Metz et al. 1989, Spitzer, Endicott et al. 1978). The K-SADS-PL elicits 52 separate DSM-IV psychiatric diagnoses (DSM-IV). The K-SADS-PL has some advantages compared to earlier versions, first, there is a screening interview to make the interview less time consuming. When a child is subthreshold in the screen no further assessment in that diagnostic area is made. Second, the interview provides rating of both global impairment and diagnosis specific impairment. Third, both present and past diagnoses are evaluated and coded (Kaufman et al. 1997).

The K-SADS-PL interview has three sections: 1, the introductory interview 2, the screen interview and 3, eight optional diagnostic supplements.

The introductory interview begins with an open question about the major complaint followed by assessment of functional level, demographics and the family history of mental health problems.

The screening interview has 105 probes that represent 23 different diagnostic domains. If any probe reaches threshold the corresponding supplement is to be used.

The supplements, which cover all aspects, are: 1) affective disorders; 2) psychotic disorders; 3) anxiety disorders; 4) behavioural disorders; 5) substance use disorders; 6) eating disorders; 7) tic disorders and 8) autism spectrum disorders, allowing the clinicians to arrive at DSM-IV diagnoses. The supplement for autism spectrum disorder was added in the 2009 year version of K-SADS-PL and covers speech delay, motor mannerism, inflexible routines, stereotyped interests, and non-verbal behaviour but not social reciprocity (Axelson, Birmaher et al. 2009).

It is recommended to interview parent and child separately. After both have been interviewed the interviewer synthesizes any discrepancies, if necessary by undertaking further interviewing, and uses his/her clinical judgement to conclude diagnoses and functional impairment.

The K-SADS-PL has shown overall very good psychometric properties with excellent interrater agreement on audiotaped (Kaufman et al. 1997) or videotaped (Polanczyk, Eizirik et al. 2003, Ulloa, Ortiz et al. 2006) or conjoint session (Shanee, Apter et al. 1997) and good (Ghanizadeh, Mohammadi et al. 2006) to moderate (Kim, Cheon et al. 2004) agreement when the interviews were done in succession. In an evaluation from Kaufman and colleagues (Kaufman et al. 1997) the test-retest reliability was good for most disorders, ADHD, ODD, GAD, and PTSD and excellent for MDD. Similar good test-retest reliability was found for ADHD and ODD in an Iranian study (Ghanizadeh et al. 2006) as well as for ADHD in a Korean study (Kim et al. 2004).

The K-SADS-PL, or earlier versions of the K-SADS, has in general found support concerning construct validity compared to rating scales (Lauth, Arnkelsson et al. 2010, Brasil, Bordin 2010, Kim et al. 2004, Kaufman et al. 1997) Convergent validity was supported vis-à-vis the CBCL for broad categories of internalizing and externalizing disorders (Biederman, Faraone et al. 1993) as well compared to Stony Brook child Psychiatric checklist (Grayson, Carlson 1991).

There are surprisingly few studies examining concurrent validity of the earlier versions of the K-SADS, to clinical diagnoses elicited from either unstructured or structured interviews and the results are mixed (Cohen, O'Connor et al. 1987, Apter, Orvaschel et al. 1989, Hodges, McKnew et al. 1987). Later studies of the K-SADS-PL are scarce but more promising (Ghanizadeh et al. 2006, Kim et al. 2004, Shanee et al. 1997). Gahnizadeh and colleagues found good agreement for SAD, GAD, ODD and bipolar disorder and excellent for MDD, ADHD and CD, comparing diagnoses elicited by the K-SADS-PL with clinical diagnoses (Ghanizadeh et al. 2006). Further support of concurrent validity comes from Shanee and colleagues who found excellent agreement for ADHD, anxiety disorders, MDD, mania and psychoses in an inpatient sample (Shanee et al. 1997). A Korean study found less agreement to clinical diagnoses with only fair agreement for anxiety disorders, MDD and tics disorder and moderate for ODD and good for ADHD (Kim et al. 2004).

In summation, the K-SADS-PL is a widely used comprehensive semi-structured, investigator-based, diagnostic interview with overall good psychometric properties. However there are a few studies with inconsistent results, supporting its concurrent validity.

LEAD (Longitudinal Expert All Data) procedure

A LEAD procedure may be seen as a proxy to gold standard in evaluating the validity of psychiatric diagnoses (Kranzler, Tennen et al. 1997). There is a lack of laboratory tests or other objective measures to evaluate psychiatric diagnoses. Further there are often diagnostic discrepancies (Rettew et al. 2009) which led Spitzer (Spitzer 1983) to suggest a LEAD procedure to elicit valid psychiatric diagnoses to which diagnostic interviews could be compared. A LEAD procedure includes three concepts; first, *longitudinal* which refers to the diagnostic evaluation not being limited to information from one occasion but rather take into account information from several occasions over a time span. The length of the longitudinal period may vary but Pilkonis and colleagues suggested a time period of at least of six months (Pilkonis, Heape et al. 1991a). Second, *expert* means that the diagnostic evaluation should be done by experienced clinicians with expertise in diagnostics and that any diagnostic disagreements between the experts should be clarified and discussed to arrive at consensus diagnoses. Third, *all data* refers to that LEAD diagnoses should be based

on all data available from medical records, information from other professionals, information from significant others, treatment outcome and impairment (Young, O'Brien et al. 1987, Spitzer 1983). Since the quality of the LEAD diagnoses is dependent on the information that is available Kranzler and colleagues suggested that the LEAD procedure should also include a structured interview to enhance its validity (Kranzler et al. 1997). However, in most studies that have used LEAD procedure requirements have varied in terms of observation time, structure of the interview, composition of the expert team, and whether there has been additional information from significant others (Kranzler, Kadden et al. 1994, Kranzler et al. 1997, Pilkonis, Heape et al. 1991b, Skodol, Rosnick et al. 1988, Wilberg, Dammen et al. 2000, Jensen-Doss, Youngstrom et al. 2014).

Most studies using LEAD diagnoses as a standard against which to compare diagnostic instruments have been done with adults. When diagnostic interviews or scales for personality disorder have been compared to LEAD diagnoses agreement has varied. Skodol and colleagues (Skodol et al. 1988) found that the LEAD procedure elicited fewer personality disorders than the screening interview, Structured Clinical Interview for Mental Disorders II (SCID II). Pilkonis and colleagues (Pilkonis et al. 1991b) on the other hand found that the LEAD procedure elicited more diagnoses of personality disorder than the diagnostic interview, Personality Disorder Examination (PDE). Further Wilberg and colleagues (Wilberg et al. 2000) found overall poor agreement with specific personality disorders comparing Personality Diagnostic Questionnaire-4+ (PDQ-4+) to LEAD diagnoses.

In another study in adults with substance abuse disorders, the initial structured diagnostic interview was enriched by two weeks of observation, self-report forms and information from family members. The same clinician performed the initial interview and the LEAD procedure. LEAD added diagnoses within substance abuse areas while the LEAD procedure did not add comorbid diagnoses making agreement good for substance abuse and excellent for comorbidity (Kranzler et al. 1994).

To my knowledge there is only one study in a child psychiatric setting that has used a LEAD procedure as a standard to compare with clinical diagnoses. Jensen-Doss and colleagues used a LEAD procedure that included K-SADS interviews to compare with diagnoses elicited from unstructured clinical interviews (Jensen-Doss et al. 2014). The agreement between LEAD diagnoses and clinical diagnoses was low overall.

Further, a Swedish study by another research group is currently underway which uses a LEAD procedure to evaluate the diagnostic interview MINI-KID.

In sum, LEAD diagnoses where experts consider all available data, including a structured interview, gathered over a time span, may be seen as a proxy to gold standard in psychiatric diagnostics with which to compare other instruments, although there are few studies that have actually included a full LEAD procedure.

Aims

The overall aim of the thesis was to evaluate the BCFPI, the SCARED-R, and the K-SADS-PL in newly referred unselected child and adolescent psychiatric outpatients.

The primary aim of the project was to investigate the factor structure of the six symptom subscales and the criterion validity of the BCFPI compared to diagnoses elicited from a LEAD procedure. Secondary aims were to investigate the predictive validity of K-SADS-PL and the criterion validity of SCARED-R compared to LEAD diagnoses.

Specific aims

Study I (Paper 1):	To evaluate the factor structure of the six symptom subscales of the Swedish version of the BCFPI and compare it to the original English version.
Study II (Paper 2):	To compare the K-SADS-PL diagnoses against LEAD diagnoses.
Study II (Paper 3):	To investigate the criterion validity of the BCFPI against LEAD diagnoses. In addition, the CBCL was used as a comparator.
Study II (Paper 4):	To investigate the criterion validity of the anxiety scale SCARED-R against LEAD diagnoses.

Method

Participants

Study I (Paper 1)

Participants were 3753 consecutive referrals at four CAMHS in the south of Sweden that completed a BCFPI as part of the standard intake procedure during a period of five years from January 2005 to December 2009. The interviews covered 1800 (48.0%) children, aged 6 to 12 years, 1953 (52.0%) adolescents, aged 13 to 17 years, 1942 (51.7%) boys and 1811 (48.3%) girls.

Study II (Paper 2, 3, and 4)

During the study period, December 2009 through January 2013, 5908 parents of all new referred children aged 6-17 years at four CAMHS in four regions in Sweden were interviewed in the standard intake procedure with the BCFPI. The parents were informed about the study at the end of the interview. Parents and patients in need of an interpreter were excluded. If oral consent was given, a local monitor set up a K-SADS-PL interview within 6 weeks. Available time slots for K-SADS-PL interviews limited the number of interviews. The consenting parents were then subsequently and in a consecutive manner asked to participate at the available date. If the date available for K-SADS-PL was not suitable, the next consecutive patient at the site was asked. At later stages of the study 15 children (6%) were actively selected to include more girls aged 6-12 years to achieve approximately equal numbers of children and adolescents, boys and girls.

Consent to participate in the study was given by 5553 (94%) of referrals. 307 were initially included in the study and were interviewed with the K-SADS-PL. Forty of these were excluded from the study due to protocol violations in data-reporting, leaving in total 267 patients. The BCFPI and the SCARED-R were compared to the LEAD assessment. These patients constituted the study group in Paper 3 and Paper 4.

In Paper 2 additional 28 cases were excluded out of the 267 patients due to our stricter follow-up data for the LEAD procedure in this group. Thus, 239 patients constituted the study group in Paper 2. The 28 individuals had gone through the K-SADS-PL interview and the LEAD procedure, but, because of rather limited additional information beyond the K-SADS-PL interviews, these cases were excluded,

so that the agreement in the statistical analyses of the K-SADS-PL diagnoses compared to the LEAD diagnoses would not be inflated (Paper 2) (Figure 1).

Out of the 267 participants in Paper 4, 204 patients filled in the SCARED-R self-rating scale (boys n = 112; girls n = 92 respectively) and the corresponding parent-rating scale was completed by 228 parents (boys n = 122; girls n = 106). The numbers varied somewhat across analyses.

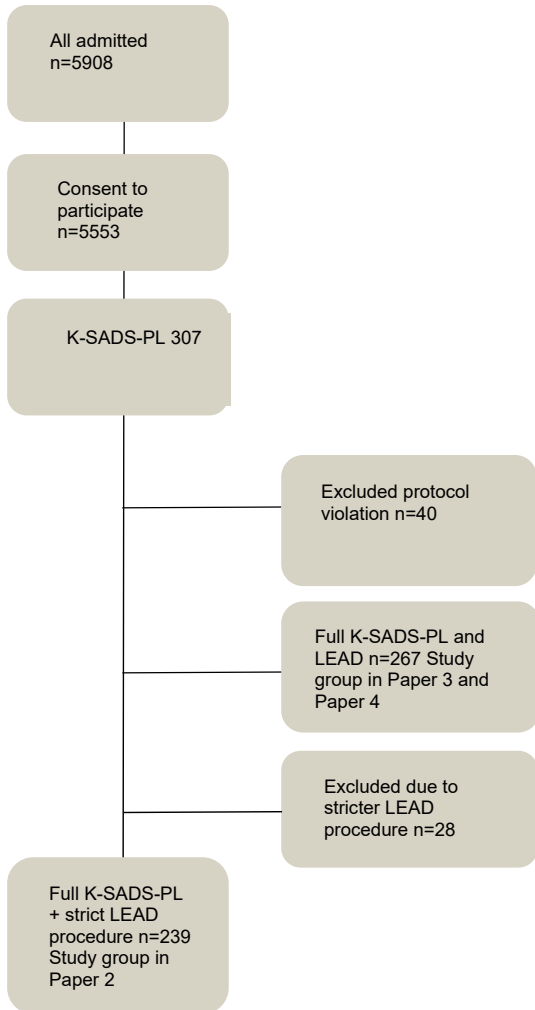

Figure 1.
Patients admitted, interviewed, excluded and chosen for Study II (Paper 2, 3, 4).

There were no significant differences in age or gender between the study groups and all admitted outpatients (Paper 2, 3, and 4). However, the study groups had significantly more externalizing and more internalizing symptoms in the BCFPI.

Procedures

BCFPI procedure

The BCFPI was performed with one parent at intake to CAMHS. The interviews were conducted over the phone by the ordinary intake team, most often nurses but also psychologists.

K-SADS-PL procedure

Patients and their parents were interviewed separately during one visit with the 2009 version of the K-SADS-PL (Axelson et al. 2009). All interviews were conducted by one of 10 residents or specialists in child psychiatry. All clinicians had participated in a course in K-SADS-PL that included scoring video interviews, performing the K-SADS-PL interview and finally passing an exam. To ensure good quality of the K-SADS-PL, each clinician aimed at performing at least 30 interviews. If the clinicians were uncertain about diagnoses they were instructed to immediately ask for guidance from a senior clinician. All diagnoses of autism spectrum disorder and bipolar disorder had to be preceded by consultations with a senior clinician. In the later stages of study the procedure included an oral report to a senior clinician in all cases since some residents at that stage were less experienced.

SCARED-R procedure

During the time the parents were interviewed with K-SADS-PL the patient filled in the SCARED-R and the CBCL questionnaires in separate rooms and vice versa. A local monitor was in place to assist the patients when needed.

LEAD procedure

LEAD diagnoses were established based on information from the K-SADS-PL interview, all available medical records including psychological assessment, teacher reports, treatment outcomes and reassessments by clinicians. The observation time after the K-SADS-PL for new information was 1.2 (SD 0.6, range 0.1-3.1) years. The

LEAD diagnoses were based on the KSADS-PL interview plus outcome of pharmacological treatment (47%), school information (41%), neuropsychological test results (38%), parent report scales (27%), senior psychiatrist diagnostic assessment (27%), and outcome of psychological treatment (6%). A blinded LEAD reliability test between the two senior clinicians was conducted with 30 random cases. The Kappa coefficients for spectrum diagnoses were excellent ranging from 0.92 to 1. For specific disorders the Kappa coefficients varied but were generally lower. All but one disagreement between the clinicians concerned if the medical records gave support for a specific diagnosis or if the patient should be diagnosed with a not-otherwise specified diagnosis in the concerned category. One senior clinician did all further LEAD diagnoses. In cases where diagnoses were changed from the K-SADS-PL interview diagnosis or in cases with inconclusive information, a consensus discussion with the second senior clinician was performed.

Instruments

The Brief Child and Family Interview (BCFPI)

The phone interview starts with a narrative overview of basic parental concerns.

The six subscales measuring symptoms correspond to the diagnoses ADHD, ODD, CD, SAD, GAD, and MDD. All subscales consist of six questions. However the subscale for ADHD has three questions for inattention forming a subscale of its own (ADD).

Each question comes with 3 response options: never, sometimes, and often. It has been standardized for two age groups, 6-12 years and 13-17 years, separately for boys and girls. The results of the BCFPI are summarized as t-scores.

The three subscales ADHD, ODD and CD constitute a composite scale for externalizing problems. The three subscales SAD, GAD, and MDD constitute a composite scale for internalizing problems (Cunningham et al. 2006).

Altogether the symptom subscales include 36 questions and are the ones analysed in the present study.

The BCFPI also includes three subscales concerning child functioning: “Social participation”, “Quality of relation”, “School participation and achievement”. They are summarized in the subscale “Global child functioning”.

Furthermore, there are subscales regarding the family’s situation. The subscale “Family activities” reflects to which extent the child’s problems are perceived to affect the family’s external social networks. “Family comfort” reflects the perceived impact

of the child's problems on conflicts and anxiety within the family. The two previously mentioned subscales are summarized in the scale "Global family situation". Finally, there are two subscales concerning risk factors: "Informant mood" consisting of six questions which reflect symptoms of depression of the informant and "Family function" which describes general problem solving, attachment and relationships within the family.

In addition to the subscales above, BCFPI also has optional questions that clinics may use in their intake-service to improve triaging. For example questions regarding autism spectrum disorder, obsessive compulsive behaviour, bullying, drugs, neglect, and parental physical and psychological abuse (Cunningham et al. 2006)

The Kiddie Schedule for Affective Disorders and Schizophrenia for school-aged children Present and Lifetime (K-SADS-PL)

The K-SADS-PL is a comprehensive semi-structured diagnostic interview that is extensively used in both research and clinical practice to arrive at child psychiatric diagnoses. The interview has shown good psychometric properties (Kaufman et al. 1997, Polanczyk et al. 2003, Ulloa et al. 2006). The interview has three sections: 1, the introductory interview 2, the screen interview and 3, eight optional diagnostic supplements.

The introductory interview begins with an open question about the major complaint followed by assessment of functional level, demographics and the family history of mental health problems.

The screening interview has 105 probes that represent 23 different diagnostic domains. If any probe reaches threshold the corresponding supplement is to be used.

The supplements, which cover all aspects, allowing the clinicians to arrive at reliable DSM-IV diagnoses, are the following: 1) affective disorders; 2) psychotic disorders; 3) anxiety disorders; 4) behavioural disorders; 5) substance use disorders; 6) eating disorders; 7) tic disorders and 8) autism spectrum disorders.

The K-SADS-PL supports 52 different diagnoses.

Parents and patients are interviewed separately and then the clinicians have to integrate the information and conclude a score and determine whether there are any diagnoses present, if they are subsyndromal, whether there have been previous diagnoses or if there is insufficient information.

The Screen for Child Anxiety Related Emotional Disorders Revised (SCARED-R)

The SCARED-R is a questionnaire developed to screen for anxiety disorders in children and adolescents. The first version, SCARED, consists of 38 items with three response options: not true or hardly ever true, somewhat true or sometimes true, and very true or often true. The items are divided into 5 subscales. Four of these correspond to anxiety diagnoses in the DSM-IV, SAD, GAD, PD, and SoP (Birmaher, Khetarpal et al. 1997). Muris and colleagues expanded the scale by adding questions for three further subscales SP, PTSD and OCD (Muris et al. 1999) called SCARED-R. The SCARED-R has shown adequate internal consistency as well as sufficient test-retest stability (Muris, Steerneman 2001b). In samples of school children the SCARED-R has shown acceptable discriminative validity (Muris, Merckelbach et al. 2001). In a small clinical sample it demonstrated good convergent validity to another anxiety self-report measure (FSSC-R) (Muris, Steerneman 2001b) and to CBCL it showed significant correlation to the internalizing but not to the externalizing subscales (Muris et al. 2004). The Swedish version of SCARED-R was translated in 2008 to be included in an OCD treatment Study (Thomsen et al. 2013). All DSM subscales of the SCARED-R are used in the present study, Paper 4.

The Child Behaviour Check List (CBCL)

The CBCL is a widespread scale used in both child psychiatric research and clinics. The scale is based on empirical studies of questions for psychiatric or behavioural problems in children and adolescents (Achenbach, Rescorla 2001) and it has norm data from large population samples. The CBCL includes 118 items with three response options, not true, somehow or somewhat true and often true (Achenbach, Rescorla 2001). The CBCL was not originally designed for diagnostic purposes (Sadock, B. Sadock, V. Kaplan, H. 2005) but there have been many studies supporting the convergence between subscales of the CBCL and ADHD (Chen et al. 1994), CBCL and CD (Biederman et al. 2005) and CBCL and ODD (Biederman et al. 2008). The DSM-oriented subscales of the CBCL have been developed by experts identifying questions in the CBCL that are consistent with diagnoses in the DSM-IV. The DSM-oriented subscales that are used as comparators in Paper 3 are depression, anxiety, ADHD, CD, and ODD.

Statistical analyses

All data used in this thesis was anonymized, and in Study II the individuals were coded for statistical analyses in SPSS version 24 and MPLUS was used (Muthén, L.K. and Muthén, B.O. 2017).

Statistical methods used

Study I (Paper 1)

The main method used to investigate the construct validity of the BCFPI, Paper 1, was confirmatory factor analysis (CFA). Using CFA (see Kline, 2011 for an overview) (Kline 2011) it is possible to test whether data fits the theoretical model that has been suggested for an instrument. It is especially useful when there are previous studies that have supported the model, e.g. that there is a strong theory. In BCFPI, there are six different subscales that measure six different diagnostic categories, suggesting six sub-factors (Cunningham et al. 2009). In addition, three subscales belong to an overarching externalizing factor, and three subscales to an internalizing factor. This type of hierarchical model is also possible to test within CFA. When testing the construct validity of an instrument there are possible differences between groups. To test whether this is the case it is possible to do invariance estimations within the CFA paradigm. Three hypotheses were tested, firstly configural invariance (the loadings of the factors are the same for all items included), secondly metric invariance (intercepts (e.g. item means) are the same), and thirdly we tested whether subgroups differed in factor means (e.g. that boys and girls have different mean values for the diagnostic groups).

Using CFA we first estimated suggested models from previous research and compared Swedish and English versions. In addition, to test whether sex, age and mother tongue resulted in important differences in the factor structure (configural invariance) and the intercepts (metric invariance), invariance models were estimated on item level.

Models were estimated using both Weighted Least Square corrected for Means and Variances (WLSMV) and Maximum Likelihood Estimations (ML). WLSMV is the default when observed variables and items are ordered categories. In addition to the 2 estimates of the models we added two other fit indices: The Comparative Fit Index (CFI) and the Root Mean Square Error of Approximation (RMSEA). The 2 value has the property that it can be used to test whether differences between (nested) models are significant, e.g. it has a known distribution, but since this estimate is dependent on sample size, we also added the other two fit indices that have adjustments for sample size and the number of parameters in models. The CFI represents the relative fit improvement of the estimated model with the baseline

model having zero covariation. Generally CFI above .80 represents a fair fit and above .90 an excellent fit. The RMSEA also adjusts for model complexity, but it is an estimate of lack of fit. RMSEA less than .08 have been suggested as representing fair fit and less than .05 a good fit.

Study II

The Receiver Operating Characteristic (ROC) curve and Area Under the ROC Curve (AUC)

The ROC curve is used to give a picture of the relation between sensitivity and specificity. It is helpful when deciding a cut-off score of a test.

The AUC is the measure of the overall performance of a test. It measures the ability to separate cases with diagnosis from ones without diagnosis. AUC values are considered having a poor (0.6-0.7), fair (0.7-0.8), good (0.8-0.9) or excellent (0.9-1) validity. An AUC of 0.7 indicates that 70% of the time the score on the questionnaire of a random individual with a disorder is higher than the score of a random individual without the disorder (Campbell, M. J Machin, D 1999).

Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV)

Sensitivity is the proportion of individuals with a disorder that are correctly identified as such by a test: true positives / (true positives + false negatives).

Specificity is the proportion of individuals without a disorder that are correctly identified as such by a test: true negatives/ (true negatives + false positives).

PPV is the probability that a positive prediction is correct: true positive/total positives. NPV is the probability that a negative prediction is correct: true negatives/total negatives (Campbell, M. J Machin, D 1999).

Cohen's Kappa

The agreement between classifications into different categories made by two different diagnostic tests can be determined by Cohen's Kappa. It is the proportion of cases where the two tests agree minus the proportion where they would likely agree by chance. A Kappa value of 1 is a perfect agreement and Kappa value 0 is an agreement by chance. Kappa coefficients are considered as poor (0-0.20), fair (0.21-0.40), moderate (0.41-0.60), good (0.61-0.80) or very good (0.81-1) (Landis, Koch 1977).

Study II (Paper 2)

The main method used in Paper 2 was Cohen's Kappa to evaluate the agreement between diagnoses elicited from the K-SADS-PL compared to LEAD diagnoses. We

reported sensitivity and specificity to describe the K-SADS-PL ability to correctly identify diagnoses as well as its ability to correctly identify absence of diagnosis. Further we reported PPV and NPV to describe the likelihood that a K-SADS-PL diagnosis is the same as the LEAD diagnosis, and the likelihood that the absence of a K-SADS-PL diagnosis also corresponds to the LEAD absence of diagnosis. Analyses were made both for specific diagnoses as well as for spectrum diagnoses. Comparisons with the total population of new referrals were performed with z-test for population proportion and t-test for single samples.

Study II (Paper 3)

The main method used in Paper 3 was ROC analysis to obtain the primary outcome measure AUC values. We reported AUC values to describe the overall performance of the BCFPI. Further we used Kappa statistics to compare the subscales of the BCFPI to their corresponding LEAD spectrum diagnoses and reported sensitivity, specificity, PPV and NPV (see above). The same statistics were also used for the CBCL DSM-oriented subscales for comparison. The composite subscale for internalizing problems in BCFPI, consisting of a merger of the subscales for SAD, GAD and MDD, was compared to a merger LEAD diagnosis of any depressive disorder plus any anxiety disorder except specific phobia. The composite subscale for externalizing problems in BCFPI, consisting of a merger of ADHD, ODD and CD was compared to a merger LEAD diagnosis of any ADHD disorder and any behavioural disorder.

Study II (Paper 4)

The main method in Paper 4 was ROC analysis to determine the agreement of the total score and the subscales of SCARED-R compared to LEAD diagnoses. Further we used Kappa statistics to compare the subscales to their corresponding LEAD spectrum diagnoses and reported sensitivity, specificity, PPV and NPV. We conducted series of logistic regression analyses to analyse the concurrent and discriminative validity of the subscales vis-a-vi all anxiety diagnoses. We reported odds ratios to describe the likelihood of a diagnosis for every additional score on each subscale. Further we performed sequential logistic regression analyses to investigate whether adding an informant, child or parent, would enhance the screening properties of SCARED-R. The t-test or 2 tests were used to investigate gender differences.

Ethical considerations

Study I and study II were approved by the Swedish Ethical Review Authority, Lund on 2009-05-15 and carried out in accordance with the Declaration of Helsinki.

All analyses in Study I (Paper 1) were done on non-identifiable data that had already been gathered in the standard intake procedure of the sites.

In Study II (Paper 2, 3, and 4) all parents and children were fully informed about the study and parents/guardians and children 15 years and older gave written consent. They were informed that at any time they were free to interrupt their participation in the study. Patients were given two tickets to the cinema. The compensation was considered to be small enough not to compromise the participants' free will.

All diagnostic information that was gathered in the K-SADS-PL interview was also part of the patients' medical record to be utilized in subsequent care.

For statistical analyses all data was anonymized using coded files with the code keys stored separately.

Results

Study I (Paper 1)

Construct validity and invariance of the Swedish version of the BCFPI

The main result was that the Swedish version of the BCFPI had almost the very same factor structure (Table 1) as the original English version (Cunningham et al. 2009) suggesting that items included in the scales were perceived the same way in Sweden as in Canada, where it was originally validated. This lends strong support to construct validity of the Swedish version of the BCFPI. In addition, externalizing and internalizing sub-factors were strongly supported.

The invariance tests (differences between groups) did not support major variability in factor structure or item intercepts between boys and girls, nor between the two age groups, nor between children of native Swedes and children with parents born abroad. As depicted in Table 2 there were only minor differences between the models. Regarding boys and girls, expected differences in relation to externalizing and internalizing factors were found. Boys were generally more often rated higher on externalizing symptoms, while girls where rated higher on internalizing symptoms.

Table 1. Factor loadings for the ML and DF solutions, together with the English loadings and the difference between the Swedish ML and English loadings Means and SD for scales and items

	Sw, WLSMV	Sw ML	Eng ML	∆ ML	M	SD
Regulating inattention impulsivity and activity level (ADHD-subscale)					**0.96**	**0.56**
Distractible, has trouble sticking to any activity	.70	.62	.77	-.15	1.24	0.74
Jump from one activity to another	.75	.67	.71	-.04	0.75	0.81
Fails to finish things he/she starts	.65	.59	.66	-.07	0.98	0.77
Fidgets	.68	.62	.66	-.04	0.77	0.83
Has difficulty following directions or instructions	.72	.64	.67	-.03	0.92	0.82
Impulsive, acts without stopping to think	.83	.68	.63	.05	1.08	0.80
Cooperation with others (ODD-subscale)					**1.06**	**0.58**
Defiant, talks back to adults	.82	.74	.77	-.03	1.13	0.78
Argues a lot with adults	.83	.73	.73	.00	0.73	0.77
Angry and resentful	.82	.73	.73	.00	1.18	0.76
Easily annoyed by others	.74	.67	.64	.03	1.03	0.79
Cranky	.81	.70	.62	.08	1.38	0.72
Blames others for his/her own mistakes	.67	.58	.65	-.07	0.92	0.82
Conduct (CD-subscale)					**0.19**	**0.27**
Engages in vandalism	.73	.50	.63	-.13	0.10	0.34
Destroys things belonging to others	.80	.66	.71	-.05	0.26	0.52
Uses weapons when fighting	.75	.58	.55	.03	0.12	0.37
Steal things at home	.63	.42	.56	-.14	0.23	0.51
Physically attacks people	.78	.68	.60	.08	0.39	0.60
Have broken into someone else's house, building, or car	.47	.18	.37	-.19	0.01	0.12
Separation from parents (SAD-subscale)					**0.50**	**0.47**
Overly upset when leaving loved ones	.83	.67	.76	-.09	0.23	0.52
Overly upset while away from loved ones	.84	.72	.76	-.04	0.75	0.79
Worries about being separated from loved ones	.66	.60	.70	-.10	0.58	0.80
Scared to sleep without parents nearby	.86	.72	.50	.12	0.26	0.57
Complains of feeling sick before separating from loved ones.	.73	.59	.63	-.04	0.28	0.58
Worries that something bad will happen to loved ones	.67	.49	.60	-.11	0.92	0.78
Managing anxiety (GAD-subscale)					**0.88**	**0.52**
Worries about doing the wrong thing	.69	.65	.75	-.10	0.98	0.81
Afraid of making mistakes	.77	.71	.76	-.05	0.76	0.79
Worries about doing better at things	.70	.62	.65	-.03	1.06	0.80
Worries about past behaviour	.50	.40	.59	-.19	0.71	0.73
Overly anxious to please people	.50	.46	.57	-.11	0.77	0.80
Worries about things in the future	.75	.61	.58	.03	1.01	0.78
Managing mood (MDD-subscale)					**0.83**	**0.60**
Unhappy, sad or depressed	.86	.79	.70	.09	1.03	0.75
Not as happy as other children	.88	.81	.77	.04	0.84	0.79
Feels hopeless	.82	.73	.69	.04	0.89	0.75
Has trouble enjoying him/her self	.85	.76	.74	.02	0.72	0.77
Gets no pleasure from usual activities	.90	.84	.64	.20	0.96	0.77
Has no interest in usual activities	.62	.53	.55	-.02	0.55	0.75

Note: Sw = Swedish version, Eng = English version, ∆ ML = Difference between Swedish and Engligh version, M = mean, and SD = standard deviation. ML = Maxium Likilihood estimation, WLSMV = Weighted Least Square Mean and Variance corrected estimation.

Table 2.
Model information from the invariance estimations

Model	χ^2	df	CFI	RMSEA	∆df	$\Delta\chi^2$	∆CFI	∆ RMSEA
Sex configural	6131.8	1090	.940	.050	0			
Sex metric	6358.2	1119	.937	.050	29	277.6	.003	<.001
Sex scalar	6420.3	1148	.937	.049	29	111.7	<.001	.001
Sex latent mean	7288.1	1154	.927	.053	6	296.6	.010	-.004
Age Configural	5618.2	1090	.944	.047	0			
Age Metric	6051.1	1119	.939	.048	29	520.2	.005	-.001
Age Scalar	6284.3	1148	.937	.049	29	300.3	.002	-.001
Age Latent mean	8033.8	1154	.915	.056	6	543.0	.022	-.007
Lang Configural	5810.3	1090	.943	.048	0			
Lang Metric	5873.3	1119	.943	.048	29	141.0	<.001	<.001
Lang Scalar	5852.5	1148	.943	.048	29	73.4	<.001	<.001
Lang Latent mean	5750.2	1154	.945	.046	6	19.2	.002	.002

Note: ∆ = denote difference between models, χ^2 = chi-square model fit statistics, df = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Sqaure Error of Aproximation index, $\Delta\chi^2$ are from the MPLUS DIFFTEST.

Study II (Paper 2)

Validity of psychiatric diagnoses elicited from the K-SADS-PL compared to LEAD diagnoses

The main result was that the predictive validity of diagnoses elicited by K-SADS-PL compared to LEAD diagnoses was good to very good for most child psychiatric diagnoses except for autism spectrum disorders.

The prevalence of psychiatric disorders in the study sample were for depressive disorders (30.8%), anxiety disorders excluding specific phobia (30.1%), disruptive disorders (34.3%), ADHD spectrum disorders (68.6%) and autism spectrum disorders (11.3%).

The agreement between diagnoses elicited by the K-SADS-PL and diagnoses elicited by LEAD was very good for depressive disorders (Kappa 0.91), anxiety disorders excluding specific phobia (Kappa 0.94) and behavioural disorders (Kappa 0.91) and in the highest range of good (Kappa 0.80) for ADHD spectrum disorders. For autism spectrum disorders the agreement was in the lowest range of good (Kappa 0.62) suggesting lower predictive validity for these diagnoses (Table 3).

Table 3.
Agreement of the K-SADS-PL versus LEAD diagnoses in a Swedish sample of 239 child and adolescents 6-17 years referred for psychiatric care.

Diagnosis	LEAD		K-SADS-PL			
	n	%	n	%	T	Kappa
Any depressive disorder	74	30,8	69	28,8	14,1	0,91
Major depressive episode	45	18,8	43	18,0	13,3	0,86
Dysthymia	9	3,8	8	3,3	12,7	0,82
Depression NOS	20	8,4	18	7,5	11,9	0,37
Elimination disorders, any	19	7,9	22	9,2	14,3	0,92
Encopresis	13	5,4	13	5,4	15,5	1,00
Enuresis	13	5,4	14	5,9	14,9	0,96
Any anxiety disorder, spec phobia excluding	72	30,1	72	30,1	14,5	0,94
Panic disorder & agoraphobia	6	2,5	6	2,5	12,8	0,83
Separation anxiety	21	8,8	19	7,9	14,6	0,95
Specific phobia	41	17,2	40	16,7	15,2	0,98
Social phobia	19	7,9	18	7,5	15,0	0,97
Generalized anxiety disorder	16	6,7	16	6,7	12,4	0,80
Obsessive compulsive disorder	12	5,0	11	4,6	13,4	0,86
Anxiety disorder NOS	9	3,8	12	5,0	11,7	0,75
ADHD spectrum disorder	164	68,6	146	61,1	12,5	0,80
Combined	73	30,5	55	23,0	12,1	0,77
Inattentive	54	22,6	51	21,3	12,3	0,79
Hyperactive	7	2,9	8	3,3	10,2	0,66
ADHD NOS	30	12,6	32	13,4	9,7	0,63
Disruptive disorders, any spectrum	82	34,3	82	34,3	14,0	0,91
Conduct disorder	12	5,0	13	5,4	14,8	0,96
Oppositional defiant disorder	54	22,6	47	19,7	12,2	0,35
Behavior disorder NOS	16	6,7	24	10,0	9,8	0,62
Any tic disorder	20	8,4	21	8,8	12,6	0,81
Tourette's syndrome	8	3,3	7	2,9	12,3	0,79
Autism spectrum disorder	27	11,3	30	12,6	9,6	0,62
Asperger syndrome	7	2,9	9	3,8	9,6	0,61
Autistic disorder	7	2,9	6	2,5	9,4	0,60
PDD-NOS	13	5,4	15	6,3	4,9	0,32
All diagnoses	523		505			

For ADHD more cases were undetected (n=20) than misattributed (n=2) leading to a high specificity (0.97) but a somewhat lower sensitivity (0.88). For autism there was a similar amount of undetected (n=9) as misattributed cases (n=11) leading to a specificity of 0.95 and a sensitivity of 0.70 (Table 4).

Table 4

Table 4. Screening properties of the K-SADS versus LEAD diagnoses in a Swedish sample of 239 children and adolescents 6-17 years referred for psychiatric care. All comparisons were performed by using LEAD as the proxy gold standard.

Diagnosis	PPV %	NPV %	Sensitivity %	Specificity %
Any depressive disorder	97,1	95,9	90,5	98,8
Major depressive episode	90,7	96,9	86,7	97,9
Dysthymia	87,5	99,1	77,8	99,6
Depression NOS	83,3	97,7	75,0	98,6
Elimination disorders, any	86,4	100	100	98,6
Encopresis	100	100	100	100
Enuresis	92,9	100	100	99,6
Any anxiety disorder*	95,8	98,2	95,8	98,2
Panic disorder & agoraphobia	83,3	99,6	83,3	99,6
Separation anxiety	100	99,1	90,5	100
Specific phobia	100	99,5	97,6	100
Social phobia	100	99,5	94,7	100
Generalized anxiety disorder	81,3	98,7	81,3	98,7
Obsessive compulsive disorder	90,9	99,1	83,3	99,6
Anxiety disorder NOS	66,7	99,6	88,9	98,3
ADHD spectrum disorder	98,6	78,5	87,8	97,3
Combined	96,4	89,1	72,6	98,8
Inattentive	86,3	94,7	81,5	96,2
Hyperactive	62,5	99,1	71,4	98,7
ADHD NOS	65,6	95,7	70,0	94,7
Disruptive disorders, any	93,9	96,8	93,9	96,8
Conduct disorder	92,3	100	100	99,6
Oppositional defiant disorder	89,4	93,8	77,8	97,3
Behavior disorder NOS	54,2	98,6	81,3	95,1
Any tic disorder	81,0	98,6	85,0	98,2
Tourette's syndrome	85,7	99,1	75,0	99,6
Autism spectrum disorder	63,6	96,2	70,4	94,8
Asperger syndrome	55,6	99,1	71,4	98,3
Autistic disorder	66,7	98,7	57,1	99,1
PDD-NOS	33,3	96,4	38,5	95,6

Validity of the six symptom subscales of BCFPI compared to LEAD diagnoses

The main result was that the criterion validity for the six symptom subscales of the BCFPI was fair to good for most child psychiatric disorders compared to LEAD diagnoses.

The criterion validity was fair for ODD (AUC, 0.73), GAD (0.73) and MDD (0.78), good for ADHD (0.81) and CD (0.83) and excellent for SAD (0.90). Kappa coefficients of optimal cut-offs showed moderate agreement for ADHD (0.40) and SAD (0.45), fair agreement for MDD (0.37) and ODD (0.31) and poor agreement for GAD (0.18) and CD (0.19). The CBCL performed on a similar level as the BCFPI, but the CBCL was better regarding CD and ODD. The BCFPI was better at identifying SAD. The prevalence of disorders differed with high prevalence of ADHD and ODD and very low prevalence of GAD and CD, making results less certain for the latter two (Table 5).

Table 5.
Screening properties of the BCFPI and the CBCL versus a LEAD diagnosis

	BCFPI							CBCL		
	#Cut	AUC	κ	Sens	Spec	PPV	NPV	AUC	κ	#Diag
ADD	4	.82 (± .05)	.48	.80	.68	.82	.65			174
ADHD	7	.81 (± .05)	.40	.74	.68	.81	.58	.80	.44	174
SAD	9	.90 (± .06)	.45	.43	.97	.56	.95	.74	.11	23
GAD	11	.75 (± .12)	.18	.19	.97	.27	.95	.72	.07	16
MDD	7	.78 (± .06)	.37	.56	.81	.54	.81	.77	.36	80
CD	6	.83 (± .09)	.19	.23	.96	.23	.96	.92	.34	13
ODD	10	.73 (± .06)	.31	.53	.78	.56	.75	.80	.43	93
Int	14	.70 (± .06)	.29	.62	.67	.61	.68			120
Ext	13	.78 (± .06)	.34	.81	.54	.82	.52			193

Note: Figures in parenthesis are 95 % confidence intervals. #Cut is the cut-off raw score for the kappa analyses. AUC=Area under the curve. κ =Kappa. Sen=sensitivity. Spec=specificity. PPV=positive predictive value. NPV=negative predictive value. #Diag=number of children with a LEAD diagnosis in the spectrum. SAD=separation anxiety disorder, GAD=generalized anxiety disorder, MDD=major depressive disorder, CD=conduct disorder, ODD=oppositional defiant disorder INT=internalizing symptoms, EXT=externalizing symptoms.

There were no significant differences in criterion validity across sex and age groups for the BCFPI (Table 6).

Table 6.
Screening properties expressed as Area under the curve for BCFPI vs. LEAD divided into gender and age groups Area under the curve with 95 % confidence intervals for a subset of diagnoses

	Boys (n = 150)	Girls (n = 117)	6-12 years (n = 154)	13-17 years (n = 113)
ADD	.82 (± .07)	.80 (± .08)	.82 (± .08)	.85 (± .07)
ADHD	.84 (± .07)	.79 (± .08)	.84 (± .08)	.79 (± .08)
MDD	.76 (± .08)	.80 (± .08)	.75 (± .10)	.75 (± .09)
ODD	.73 (± .08)	.74 (± .09)	.69 (± .08)	.78 (± .09)
Int	.67 (± .09)	.72 (± .09)	.72 (± .08)	.65 (± .10)
Ext	.76 (± .09)	.79 (± .08)	.78 (± .08)	.75 (± .09)

Note: Figures in parenthesis are 95 % confidence intervals. SAD=separation anxiety disorder, GAD=generalized anxiety disorder, MDD=major depressive disorder, CD=conduct disorder, ODD=oppositional defiant disorder Int=internalizing symptoms, Ext=externalizing symptoms.

Study II (Paper 4)

Validity of the SCARED-R compared to LEAD diagnoses

The main result was that the criterion validity of SCARED-R parent report was fair to good for most anxiety disorders except for GAD and SP showing less convincing validity. The parent report was generally more valid than the self-report.

The total score on SCARED-R self-report predicted the presence of any anxiety disorder to a low degree (AUC 0.65) but the parent report to a fair degree (AUC 0.77). The disorder-specific subscales predicted their corresponding disorder with low to good accuracy in the child version (AUC SP 0.61- SoP 0.85) as well as in the parent version (AUC SP 0.68- SoP 0.89). The Kappa coefficients at optimal cut-offs were fair for all subscales in the parent version except for GAD that had poor agreement with the LEAD diagnosis. The Kappa coefficients for the self-report were poor for GAD, SP and PD and fair for SAD, SoP and OCD (Table 7).

Table 7.
Psychometric properties for the SCARED-R versus a LEAD diagnosis of any anxiety disorder and the specific anxiety disorders

SCARED scale/ subscale -> LEAD	Report	AUC	P	Cut-off	Sensitivity %	Specificity %	Kappa
SCARED total score -> Any anxiety	Child	.66	<.001	≥15	84	43	.23
	Parent	.72	<.001	≥14	68	61	.27
SCARED-R total score -> Any anxiety	Child	.65	<.001	≥25	84	43	.23
	Parent	.72	<.001	≥34	43	89	.34
SAD -> SAD	Child	.76	<.001	≥5	78	70	.21
	Parent	.84	<.001	≥5	79	80	.32
SoP -> SoP	Child	.85	<.001	≥8	77	80	.29
	Parent	.89	<.001	≥9	75	88	.39
GAD - > GAD	Child	.71	.008	≥6	79	60	.11
	Parent	.74	.002	≥8	56	80	.17
OCD -> OCD	Child	.84	<.001	≥10	60	91	.31
	Parent	.84	<.001	≥9	46	97	.42
SP -> SP	Child	.61	.043	≥7	68	57	.13
	Parent	.68	.001	≥5	63	70	.21
PD -> PD	Child	.84	<.001	≥6	92	69	.18
	Parent	.88	<.001	≥4	83	83	.28

AUC = Area Under the Curve, SAD = Separation Anxiety Disorder, SP = Specific phobia, SoP = Social phobia, GAD = Generalized anxiety disorder, PD= Panic disorder, OCD = Obsessive-compulsive disorder, APTSD = Acute or post-traumatic stress disorder

There were benefits of adding information from the self-report to the parent-report (and vice versa to a smaller degree) for most subscales. Further, there were significant benefits of adding the parent report to the child report and vice versa when predicting the presence of any anxiety disorders, from SCARED-R total score (Table 8).

Table 8.
Sequential Logistic Regression to Test the Effects of Child and Parent Report on the SCARED for the Prediction of Any anxiety disorder or specific anxiety disorders. In the multivariate model, the contribution of the child to the parent report and vice versa are added to the values in the single report from the uni-variate analysis (e.g., parent OR=4.89 are added to the child OR =3.89). The Δ19.081 is the difference between the full model χ^2 35.624 - 16.543 etc.

LEAD diagnosis	SCARED Scale - Informant	OR (95%)	Wald	Full model χ^2	Full model adding an extra report χ^2 a	R^2
Any anxiety univariate models	Revised total – child	3.98 (1.95, 8.11)	14.389***	16.543***		.11
	Revised total - parent	5.79 (2.82, 11.91)	22.844***	25.116***		.16
Any anxiety multivariate model	Revised total – child	3.23 (1.54, 6.79)	9.551**	35.624***	Δ10.508***	.22
	Revised total - parent	4.89 (2.33, 10.25)	17.687***	35.624***	Δ19.081***	
Any anxiety univariate model	Total score - Child	3.85 (1.88, 7.85)	13.687***	15.678***		.10
	Total score - Parent	2.92 (1.60, 5.32)	12.255***	12.801***		.09
Any anxiety Multivariate model	Total score - Child	3.26 (1.57, 6.77)	10.048**	23.890***	Δ11.089***	.16
	Total score - Parent	2.45 (1.32, 4.56)	8.030**	23.890***	Δ8.212**	
SAD univariate models	SAD – Child	7.96 (2.51, 25.26)	12.371***	15.367***		.16
	SAD - Parent	15.59 (4.82, 50.41)	21.042***	26.444***		.27
SAD Multivariate model	SAD – Child	3.18 (0.88, 11.54)	3.096	29.774***	Δ3.330	.31
	SAD - Parent	9.49 (2.67, 33.82)	12.057***	29.774***	Δ14.407***	
SoP Univariate models	SoP - child	11.76 (3.59, 38.56)	16.537***	19.787***		.22
	SoP - Parent	20.74 (6.17, 69.77)	23.997***	28.608***		.31
SoP Multivariate model	SoP - child	6.29 (1.74, 22.75)	7.868**	37.232***	Δ8.624**	.40
	SoP - Parent	12.72 (3.56, 45.44)	15.314***	37.232***	Δ17.445***	
GAD Univariate models	GAD – child	5.45 (1.47, 20.20)	6.435*	7.976**		.10
	GAD - parent	6.29 (1.99, 19.77)	9.854**	10.318***		.13
GAD Multivariate model	GAD – child	3.98 (1.04, 15.32)	4.039*	15.043***	Δ4.725*	.18
	GAD - parent	4.77 (1.47, 15.49)	6.759**	15.043***	Δ7.067**	
OCD Univariate models	OCD – child	14.17 (3.65, 54.93)	14.698***	14.118***		.21
	OCD - parent	36.60 (7.96, 168.20)	21.406***	19.201***		.28
OCD Multivariate model	OCD – child	18.18 (3.21, 102.83)	10.758***	31.140***	Δ11.938***	.44
	OCD - parent	47.74 (7.07, 322.47)	15.734***	31.140***	Δ17.021***	
SP Univariate models	SP – child	3.04 (1.31, 7.02)	6.743**	7.321**		.06
	SP - parent	3.96 (1.76, 8.93)	11.042***	11.558***		.10
SP Multivariate model	SP – child	2.17 (0.90, 5.26)	2.963	14.640***	Δ3.082	.12
	SP - parent	3.17 (1.36, 7.39)	7.085**	14.640***	Δ7.319**	

General discussion

The primary aim of this thesis was to investigate the factor structure of the six symptom subscales and the criterion validity of the BCFPI compared to diagnoses elicited by a LEAD procedure. Secondary aims were to investigate the predictive validity of the K-SADS-PL and the criterion validity of the SCARED-R to LEAD diagnoses.

Comments on main findings

The main findings in the present thesis were:

The construct validity of the Swedish version of the BCFPI was supported and the patterns of factor loadings were closely aligned to the original English version. Statistical differences between girls and boys, age groups and whether the parents had Swedish as a native tongue or not, were in general small.

The criterion validity of the six symptom subscales of BCFPI compared against LEAD diagnoses, using the CBCL as a comparator, was supported for most major child psychiatric disorders and overall the BCFPI performed equivalent to the CBCL.

The K-SADS-PL predictive validity compared to LEAD diagnoses was strongly supported for most child psychiatric disorders except autism spectrum disorders.

The criterion validity of SCARED-R compared to LEAD diagnoses was supported for most subscales. Parent reports were more reliable than self-reports. Adding information from self-reports to the parent report gave a small increase in screening efficiency.

The validity of the BCFPI

The evaluation of the Swedish version of the BCFPI gave strong support to the construct validity in line with earlier findings in the original English version. The patterns of loadings of the six symptom subscales were closely aligned to the English version (Cunningham et al. 2009) and we found strong support for internalizing and externalizing sub-factors. We analyzed the factor loadings by different subgroups

namely girls or boys, children or adolescents and by whether the parent had Swedish as a native tongue or not. In general there were only small differences between groups and there is no call for moderation of the interview in relation to these groups. Differences in intercept (mean levels) for girls/boys and children/adolescents were in line with previous studies where internalizing symptoms are more frequent in girls and externalizing symptoms are more frequent in boys (Cunningham et al. 2009). Surprisingly, the factor MDD had lower correlation to the other five factors in the Swedish sample than in the Canadian sample. This can be due to somewhat different sample characteristics. The item "have you broken into someone else's house, building or car" is part of the CD subscale. It has low loading in the English and the Swedish version. The item is problematic since very few parents recognize this behavior in their adolescent and the behavior is very rare in younger children. This item may, in future revisions, be replaced with a more common symptom of conduct disorder. The item "worries about past behavior" had borderline loading in the Swedish version but not in the English version. This should be acknowledged in future revisions.

The evaluation of the criterion validity of the BCFPI against LEAD diagnoses, using the CBCL as a comparator, supported the criterion validity of the interview. The interview performed well according to the primary outcome measure with AUC values ranging from fair to excellent for the six symptom subscales which is in line with earlier studies (Boyle et al. 2009, Cook, Leschied et al. 2013a). The BCFPI did about equally well for girls/boys and children/adolescents which was also the case in the study by Boyle et al (Boyle et al. 2009). The BCFPI discriminated at about the same level as the more comprehensive CBCL which gives further support to the validity of the BCFPI since CBCL has shown to be a valid instrument in numerous studies (Achenbach, Howell et al. 1991, Achenbach, Rescorla 2001, Deighton et al. 2014).

Especially satisfactory was the fact that the BCFPI, in line with earlier findings (Boyle et al. 2009) could reliably screen for ADHD. Any screening instrument used for triage in CAMHS needs to identify children with ADHD since such a large portion of children seeking mental health services are afflicted with ADHD, as witnessed by a majority of the children in the sample suffering from ADHD. Somewhat surprisingly the three attention items of the ADHD subscale were equally good at identifying ADHD as the full six item ADHD subscale also covering hyperactivity and impulsivity. This has not been shown in previous studies and is an important finding.

The BCFPI performed similar to the CBCL for depressive disorder and demonstrated excellent AUC for separation anxiety. Surprisingly the parental BCFPI discriminated better for depression in our study than in the study of Boyle et al where the BCFPI was compared to the parent version of DISC-IV. This in spite of the LEAD diagnoses being based not solely on parental information as in Boyle et al (Boyle et al. 2009)

but also from interview with the child and other information. However the difference was of small magnitude.

Although the AUC values were fair for GAD and good for CD in line with Boyle et al (Boyle et al. 2009) the Kappa coefficients of the subscales were poor. The low Kappa coefficients for these subscales were not expected since they have shown markedly higher Kappa coefficients in previous research (Boyle et al. 2009). The low Kappa coefficients could partly be explained by the low prevalence of these disorders in the sample making statistical analyses more uncertain. However, the CBCL performed better than BCFPI on CD which together with the finding in Paper 1 where one item in the BCFPI subscale for CD had factor loading which was too low indicates that there is room for improvement in future revisions of the CD subscale. The BCFPI had problems identifying GAD in the present sample. This was also partly true for the K-SADS-PL, Paper 2, and the SCARED-R, Paper 4. Symptoms of generalized anxiety are present in many disorders. They can be part of an episode of depression or be an adequate reaction to real life stresses e.g. in ADHD with school tasks or in autism with relational challenges making signs of anxiety not specific just to a generalized anxiety disorder. The study sample had more symptoms than all the cases admitted to CAMHS as well as more symptoms than the sample in the study of Boyle et al (Boyle et al. 2009). Possibly GAD is more difficult to identify in more severely impaired patients with many reasons to experience adequate anxiety. This calls for precaution using the subscales GAD and CD with a fixed cut-off for triage. Scores on the subscales need to be interpreted together with the other information in the BCFPI, to allow a proper triage.

Altogether, the studies support earlier findings (Cunningham et al. 2009, Cunningham et al. 2006, Cook et al. 2013a, Boyle et al. 2009) that BCFPI is a valid scale that can discriminate most major child psychiatric disorders fair to good and overall as well as the more comprehensive CBCL.

The validity of the K-SADS-PL

The evaluation of the K-SADS-PL against LEAD diagnoses, gave strong support to the predictive validity of the K-SADS-PL for most child psychiatric disorders. This was encouraging since, although K-SADS-PL has shown excellent data regarding inter-rater reliability (Kaufman et al. 1997, Polanczyk et al. 2003, Ulloa et al. 2006), previous studies have shown varied results comparing the interview with clinical diagnoses (Ghanizadeh et al. 2006, Kim et al. 2004, Shanee et al. 1997). Kappa coefficients were very good for the spectra of depressive disorders, anxiety disorders and disruptive disorders, good for ADHD disorders and good to moderate for autism spectrum disorders.

The very good Kappa coefficients of the depressive disorders, anxiety disorders and the disruptive disorders were expected since these diagnoses are either episodic or clear cut behavioral and may therefore be easier to identify correctly. The very good Kappa coefficients for depressive disorders and anxiety disorders confirm the finding from Shane and colleagues (Shanee et al. 1997). However, the very good Kappa coefficients for disruptive disorders were in contrast to earlier findings (Kim et al. 2004). GAD that has a more continuous course had the lowest agreement to LEAD diagnosis of the anxiety disorders in line with earlier findings (Ghanizadeh et al. 2006). As mentioned above the identification of GAD was problematic also for the other screening instruments in the study, BCFPI, Paper 3, and SCARED-R, Paper 4.

The good Kappa coefficient for ADHD supports previous studies of the K-SADS-PL (Ghanizadeh et al. 2006, Shanee et al. 1997, Kim et al. 2004)). The very high PPV shows that if clinicians perform the K-SADS-PL there is a low risk of falsely attributing an ADHD diagnosis. This is an important finding since there has been a debate whether or not a proper diagnostic procedure for ADHD always should include several visits, teamwork and psychological testing which is typically the case in CAMHS. This study indicates that a proper K-SADS-PL interview in most cases is a sufficient diagnostic procedure that would save clinical resources. However, the lower NPV indicates that some children with ADHD are not identified with the K-SADS-PL. This may be partly understood from the fact that teacher reports were not present in the K-SADS-PL procedure. In some cases parents believe symptoms of ADHD are age appropriate and the functional deficit only becomes more apparent after receiving input from teachers. Therefore in CAMHS, ADHD evaluation should always include teacher reports.

The K-SADS-PL performed less well concerning autism spectrum diagnoses. The supplement for autism has recently been added (Axelson et al. 2009) and there are, to my knowledge, no prior investigations of its validity. Although, there were site differences in ability to correctly identify autism spectrum diagnoses this could not explain the finding. Diagnosing children with autism spectrum disorder from a diagnostic interview performed in one visit is far from the ordinary clinical practice which most often includes psychological test, observations, teacher reports etc. This study gives no support for diagnosing autism with a K-SADS-PL interview alone. Not even when the interviewer is a trained clinician with supervision. Rather, when the K-SADS-PL indicates autism this calls for further more comprehensive assessment. In future revision of the K-SADS-PL, a revision of the autism spectrum supplement of the K-SADS-PL should be considered. The present supplement lacks screening questions about social reciprocity (Axelson et al. 2009) which probably contributes to its difficulty identifying autism.

The validity of the K-SADS in this study is more convincing than in previous studies (Ghanizadeh et al. 2006, Kim et al. 2004, Shanee et al. 1997). This could be

understood from the fact that the K-SADS-PL was, in this study, conducted under ideal circumstances with trained specialists in child psychiatry or trained residents under immediate supervision. This guarantees the quality of the interview and therefore shows the K-SADS-PL validity at its best. Further, as Spitzer pointed out (Spitzer 1983) a LEAD procedure is a proxy to gold standard for psychiatric diagnoses, especially so when the LEAD procedure includes treatment outcome and long evaluation time (Kranzler et al. 1997) as in the present study. When validating a diagnostic interview it is a problem that the measures, questionnaires, clinical diagnoses etc. used for comparison, suffer themselves from limitations in validity and therefore result in weaker agreement (Rettew et al. 2009).These problems are minimized in the present study using the LEAD procedure.

Taken together the study gives strong support for the validity of the K-SADS-PL and its ability to identify most child psychiatric disorders in outpatients when performed by supervised trained residents or specialists in child psychiatry.

The validity of the SCARED-R

The evaluation of the questionnaire SCARED and the revised version, SCARED-R, self-report and parent report against LEAD diagnoses gave additional support to previous studies establishing the criterion validity of SCARED/SCARED-R (Birmaher et al. 1997, Muris et al. 2004). The validity of the subscales varied with AUC values ranging from fair to good for both self-report and parent reports. The subscales of SP and GAD had the least convincing AUC values while subscales for PD and SoP discriminated better. The added subscales for the revised version of SCARED, OCD and SP showed varied screening efficiency while the validity of GAD was problematic. GAD was difficult to identify also for the K-SADS-PL, Paper 2, and the BCFPI, Paper 3, (see discussion above) which may explain the lower validity for this subscale compared to earlier findings (Muris et al. 2004).

Child and parent agreement were lower than expected from previous research (Muris et al. 2004) and parent reports were more valid than self-reports. The present study sample consists of highly symptomatic and comorbid outpatients which might explain the low validity of self-reports. Though, adding information from the self-report to the parent report, and vice versa to a lower degree, increased the screening efficiency for most subscales. Although the additive effect was overall small for adding an informant, it is still preferable. The screening for anxiety disorder should however not depend solely on the self-report.

The study gives some support to the use of SCARED-R for screening purposes and most so for parent reports. However, the SCARED-R predicts LEAD anxiety diagnoses but falls far behind the semi-structured K-SADS-PL. Unexperienced clinicians run the risk of equating a screen test such as SCARED-R with a proper

diagnosis. Our data confirms that the semi-structured K-SADS-PL, but not a screening test, will be able to identify most psychiatric syndromes rather close to the gold standard LEAD diagnoses. In daily practice a clinician needs to consider the K-SADS-PL or similar semi-structured interviews in order to arrive at reasonably accurate clinical diagnoses.

Strengths and limitations

The main strength concerns the generalizability of the findings. The participants were collected from consecutive new referrals at Swedish CAMHS with no competing care givers. This makes the sample fully representative of unselected CAMHS patients with the caveat that eating disorders at some sites were treated at other facilities and that the population for studies of screening and predictive properties was somewhat more symptomatic in both externalizing and emotional symptoms compared to a very large consecutive cohort. The patients were all new referrals to CAMHS with no prior contact with child psychiatric services and they had not gone through any prior psychiatric assessment that may otherwise have influenced their responses to the instruments we were evaluating.

Our analyses of the factor structure of the BCFPI were strengthened by including raw data from the original version in our data set. This made conclusions more reliable since it enabled investigation if there were any problems with the factor structure specific to the Swedish version.

Another strength was the high quality of the LEAD procedure. The LEAD procedure included a full K-SADS-PL performed with trained and supervised residents or specialists as well as all subsequent information from medical records including teacher reports, psychological assessments, the outcome of treatments, and reassessment of senior clinicians. These are all features of a LEAD procedure that are recommended in the literature (Spitzer 1983, Young et al. 1987, Kranzler et al. 1997) to yield valid LEAD diagnoses. This, together with a high inter-rater reliability of the LEAD diagnoses between two senior clinicians, makes a strong case for the argument that the LEAD procedure in this thesis could serve as a proxy to gold standard in child psychiatric diagnostics and therefore highly suitable for comparing diagnostic instruments.

A limitation was that only BCFPI performed by phone were included in the analyses of the factor structure. This gives some uncertainty as to how well the findings generalizes to the actual use of BCFPI in clinics today with a mix of phone interviews and use of the BCFPI as a written checklist sent by mail to parents. The mail version was not in use at the time of data collection. It would have been preferable to have both phone versions and mail versions included in the study for comparison.

Another limitation concerns the analyses of the patterns of the factor structure for the subgroup of parents with language other than Swedish as a native tongue. It was found that the factor structure of BCFPI was overall good also in this subgroup which is a clinical important finding since BCFPI is used in this group and discussions in clinics have questioned how well BCFPI works for parents with less Swedish language ability. However, BCFPI could perform less well in subgroups with pronounced limitations in the Swedish language not captured by this design. For instance there was no data on how long the parent had lived in Sweden.

One limitation of the comparison of the K-SADS-PL to the LEAD diagnoses, in Paper 2, was that the K-SADS-PL interview was part of the LEAD procedure which might have led to inflated agreement. One way we tried to avoid the possibility of inflated agreement, from insufficient additional information in the LEAD procedure, was by excluding cases that had less than three further visits or had limited significant extra information. Also, it would be difficult to validate K-SADS-PL to a LEAD procedure without including the interview itself since a well performed LEAD procedure is dependent on a high quality diagnostic interview (Kranzler et al. 1997). We rejected the idea of adding an extra diagnostic interview and excluding the K-SADS-PL from the LEAD procedure since the extra interview would have been affected by the K-SADS-PL. The K-SADS-PL was included in the clinical records and the further assessments usually built on the K-SADS-PL. Thus, it was not possible to use all available information from the records while still being blinded to the K-SADS-PL. Another limitation concerns the nature of the LEAD procedure since it relies partly on medical records and they may differ in quality.

Another limitation regards the generalizability of the finding. The K-SADS-PL interviews were performed with well-trained residents and specialists all having passed a course exam. They were closely monitored by senior clinicians and prompted to discuss cases when there was an uncertainty about a diagnostic evaluation. This procedure was chosen to guarantee high quality of the K-SADS-PL diagnoses and by extension the LEAD diagnoses. However, in clinical practice this kind of procedure is rare and the K-SADS-PL interviews may be performed with less trained clinicians in a less monitored setting. Therefore precautions should be taken before concluding that the overall good predictive validity of the K-SADS-PL, found in this study, can be achieved in an ordinary clinical setting.

Other limitations concerns the sample in Papers 2, 3, and 4. The sample was largely unselected but the exclusion of emergency referrals probably explains the lack of bipolar disorder and psychosis in the study sample. Further, children with eating disorders at most centers were directly referred to special units for eating disorders. Although the sample was representative in gender and age to a large consecutive sample it had more elevated BCFPI scores. It is not clear how this affected the agreements of the BCFPI, SCARED-R and the K-SADS-PL to LEAD diagnoses. On

one hand more severe symptoms might be easier to identify correctly, on the other hand patients with more comorbidity might be more difficult to diagnose.

Further, the study sample, in Papers 2, 3 and 4, included few children with GAD and CD which made statistical analyses and conclusions for these groups more uncertain. Although the BCFPI showed fair AUC values for GAD and good AUC value for CD the subscales had low Kappa coefficients. It would be preferable to have a larger study group to make conclusions for these two groups more solid. When we designed the study, the size of the study groups were set to ensure that proper analyses could be made for both girls/boys and children/adolescent aiming at 300 patients evenly distributed across the groups. Having residents and specialists performing a full K-SADS-PL interview is both time consuming and expensive and it was not practically feasible to get a study group size that could guarantee enough patients for the rarer diagnoses in child psychiatry.

Clinical implications

A standardized intake instrument can improve the triaging between different mental health and community services (Galanter, Patel 2005b) which may reduce costs (Martinez-Gonzalez et al. 2015) and make the clinical assessment more efficient (Hughes et al. 2005) as well as increasing patient satisfaction. Many CAMHS in Sweden and Canada have started up central triage units using the BCFPI to guide the patients to the suitable mental health service. The BCFPI is well liked by both clinicians and parents (Carlberg 2010) and its comprehensive screening for common child psychiatric disorders as well as functional impairment, family functioning, and common risk factors such as neglect and parental depression gives it an edge compared to other well validated but more narrow instruments e.g. SDQ (Goodman 1999, Malmberg et al. 2003, Smedje et al. 1999, Deighton et al. 2014) and CBCL (Deighton et al. 2014, Achenbach et al. 1991, Achenbach, Rescorla 2001)

Thus, the clinical utility of the BCFPI is known by many triage units and clinicians but there has been a lack of studies investigating the criterion validity of the BCFPI in a Swedish setting.

The major clinical implications from this study are:

First, the Swedish version of the BCFPI can be reliably used as a screening measure for the major child psychiatric disorders in both genders, in the age span 6-17 years as well as with parents who speak Swedish but have a different native language.

Second, the K-SADS-PL diagnoses elicited at one visit with a well-trained and supervised clinician have good to very good validity for most child psychiatric diagnoses.

Third, the K-SADS-PL can identify most ADHD cases at one visit with minimal risk of overdiagnosis. However, the risk for underrecognition cannot be ignored and calls for continued awareness.

Fourth, residents should be trained in conducting the K-SADS-PL interview to improve diagnostic assessment in CAMHS. However, since the interview is time consuming it is not suited for routine clinical care.

Fifth, the K-SADS-PL should not be used to diagnose autism spectrum disorders.

Sixth, the SCARED-R can be reliably used as a screening measure for anxiety disorders and OCD in the parent version while the patient version adds less value in a clinical environment with much comorbidity.

Future directions

Future research concerning the BCFPI should focus on developing a subscale for autism spectrum disorders. This would improve the triage for a large group of patients that are now screened only by the additional autism questions of the BCFPI. Further, it would be valuable to evaluate the subscales of child functioning and family functioning in a Swedish setting since these subscales is used together with the symptom subscales to improve the triage. Finally, it would be interesting to evaluate how well the BCFPI could triage patients directly to specific treatments such as different parent groups etc. The translation of the worry item needs a revision and the conduct item about stealing cars needs to be changed to an item with higher prevalence.

The future research of the K-SADS-PL should include the revision of the autism supplement. It would also be of interest to evaluate which screening items from each supplement that has the most predictive value.

Further, the K-SADS-PL should be evaluated in different populations for instance with more severely impaired inpatients as well as patients with less impairment in primary care.

It would also be valuable to validate the K-SADS-PL in a less ideal setting with clinicians conducting the interview without supervision. This would give information of how valid the K-SADS-PL is in more traditional clinical practice.

References

ABIKOFF, H., COURTNEY, M., PELHAM, W.E. and KOPLEWICZ, H.S., 1993. Teachers' ratings of disruptive behaviors: the influence of halo effects. *Journal of abnormal child psychology,* **21**(5), pp. 519-533.

ACHENBACH, T.M. and DUMENCI, L., 2001. Advances in empirically based assessment: revised cross-informant syndromes and new DSM-oriented scales for the CBCL, YSR, and TRF: comment on Lengua, Sadowksi, Friedrich, and Fischer (2001). *Journal of consulting and clinical psychology,* **69**(4), pp. 699-702.

ACHENBACH, T.M., HOWELL, C.T., QUAY, H.C. and CONNERS, C.K., 1991. National survey of problems and competencies among four- to sixteen-year-olds: parents' reports for normative and clinical samples. *Monographs of the Society for Research in Child Development,* **56**(3), pp. 1-131.

ACHENBACH, T.M. and RESCORLA, L.A., 2001. *Manual for the ASEBA School-Aged Form & Profiles.* University of Vermont Reasearch Center for Children, Youth, and Families.

AMBROSINI, P.J., METZ, C., PRABUCKI, K. and LEE, J.C., 1989. Videotape reliability of the third revised edition of the K-SADS. *Journal of the American Academy of Child and Adolescent Psychiatry,* **28**(5), pp. 723-8.

AMBROSINI, P.J., 2000. Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). *Journal of the American Academy of Child and Adolescent Psychiatry,* **39**(1), pp. 49-58.

ANGOLD, A. and COSTELLO, E.J., 2000. The Child and Adolescent Psychiatric Assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry,* **39**(1), pp. 39-48.

APTER, A., ORVASCHEL, H., LASEG, M., MOSES, T. and TYANO, S., 1989. Psychometric properties of the K-SADS-P in an Israeli adolescent inpatient population. *Journal of the American Academy of Child and Adolescent Psychiatry,* **28**(1), pp. 61-5.

ARAS, S., VAROL TAS, F. and BAYKARA, B., 2014. Triage of Patients in the Child and Adolescent Psychiatry Outpatient Clinic. *Noro psikiyatri arsivi,* **51**(3), pp. 248-252.

AXELSON, D., BIRMAHER, B., ZELAZNY, J., KAUFMAN, J. and GILL, M.K., 2009. K-SADS-PL 2009 Working Draft.

BIEDERMAN, J., BALL, S.W., MONUTEAUX, M.C., KAISER, R. and FARAONE, S.V., 2008. CBCL clinical scales discriminate ADHD youth with structured-interview derived

diagnosis of oppositional defiant disorder (ODD). *Journal of attention disorders,* **12**(1), pp. 76-82.

BIEDERMAN, J., FARAONE, S.V., DOYLE, A., LEHMAN, B.K., KRAUS, I., PERRIN, J. and TSUANG, M.T., 1993. Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *Journal of child psychology and psychiatry, and allied disciplines,* **34**(7), pp. 1241-51.

BIEDERMAN, J., MONUTEAUX, M.C., KENDRICK, E., KLEIN, K.L. and FARAONE, S.V., 2005. The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. *Archives of Disease in Childhood,* **90**(10), pp. 1010-1015.

BIRMAHER, B., KHETARPAL, S., BRENT, D., CULLY, M., BALACH, L., KAUFMAN, J. and NEER, S.M., 1997. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry,* **36**(4), pp. 545-553.

BOVA, D.:., 2006. *Differential item and test functioning of the Brief Child and Family Phone Interview in First Nations and non-First Nations Children and Adolescents.*, Lakehead Univ.

BOYLE, M.H., OFFORD, D.R., RACINE, Y., SANFORD, M., SZATMARI, P. and FLEMING, J.E., 1993. Evaluation of the original Ontario Child Health Study scales. *Canadian journal of psychiatry.Revue canadienne de psychiatrie,* **38**(6), pp. 397-405.

BOYLE, M.H., OFFORD, D.R., RACINE, Y., SANFORD, M., SZATMARI, P., FLEMING, J.E. and PRICE-MUNN, N., 1993. Evaluation of the Diagnostic Interview for Children and Adolescents for use in general population samples. *Journal of abnormal child psychology,* **21**(6), pp. 663-681.

BOYLE, M.H., CUNNINGHAM, C.E., GEORGIADES, K., CULLEN, J., RACINE, Y. and PETTINGILL, P., 2009. The Brief Child and Family Phone Interview (BCFPI): 2. Usefulness in screening for child and adolescent psychopatholog. *Journal of child psychology and psychiatry, and allied disciplines,* **50**(4), pp. 424-431.

BRASIL, H.H.A. and BORDIN, I.A., 2010. Convergent validity of K-SADS-PL by comparison with CBCL in a Portuguese speaking outpatient population. *BMC psychiatry,* **10**(1), pp. 83.

BRUCHMULLER, K., MARGRAF, J. and SCHNEIDER, S., 2012. Is ADHD diagnosed in accord with diagnostic criteria? Overdiagnosis and influence of client gender on diagnosis. *Journal of consulting and clinical psychology,* **80**(1), pp. 128-138.

BUCHAN, J. and DAL POZ, M.R., 2002. Skill mix in the health care workforce: reviewing the evidence. *Bulletin of the World Health Organization,* **80**(7), pp. 575-580.

CAMPBELL, M. J MACHIN, D, 1999. *Medical statistics: A commonsense approach.* third edition edn. Chichester, England: John Wiley & Sons.

CARIELLO, F.P., 2003. Computerized telephone nurse triage. An evaluation of service quality and cost. *The Journal of ambulatory care management,* **26**(2), pp. 124-137.

CARLBERG, M., 2010. *Initial bedömning på öppenvårdsmottagningar inom barn- och ungdomspsykiatri i Stockholms län: en jämförelse mella BCFPI, DAWBA och ett lokalt utvecklat mottagningssätt.* Stockholm: Barn- och ungdomspsykiatri, Stockholms läns landsting.

CATALA-LOPEZ, F., HUTTON, B., NUNEZ-BELTRAN, A., PAGE, M.J., RIDAO, M., MACIAS SAINT-GERONS, D., CATALA, M.A., TABARES-SEISDEDOS, R. and MOHER, D., 2017. The pharmacological and non-pharmacological treatment of attention deficit hyperactivity disorder in children and adolescents: A systematic review with network meta-analyses of randomised trials. *PloS one,* **12**(7), pp. e0180355.

CHEN, W.J., FARAONE, S.V., BIEDERMAN, J. and TSUANG, M.T., 1994. Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of consulting and clinical psychology,* **62**(5), pp. 1017-1025.

COHEN, P., O'CONNOR, P., LEWIS, S., VELEZ, C.N. and MALACHOWSKI, B., 1987. Comparison of DISC and K-SADS-P interviews of an epidemiological sample of children. *Journal of the American Academy of Child and Adolescent Psychiatry,* **26**(5), pp. 662-7.

CONNOR, D.F., GLATT, S.J., LOPEZ, I.D., JACKSON, D. and MELLONI RH JR, 2002. Psychopharmacology and aggression. I: A meta-analysis of stimulant effects on overt/covert aggression-related behaviors in ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry,* **41**(3), pp. 253-61.

COOK, S., LESCHIED, A.W., ST PIERRE, J., STEWART, S.L., DEN DUNNEN, W. and JOHNSON, A.M., 2013a. BCFPI Validation for a High-risk High-needs Sample of Children and Youth Admitted to Tertiary Care. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent,* **22**(2), pp. 147-152.

CUNNINGHAM, C.E., BOYLE, M.H., HONG, S., PETTINGILL, P. and BOHAYCHUK, D., 2009. The Brief Child and Family Phone Interview (BCFPI): 1. Rationale, development, and description of a computerized children's mental health intake and outcome assessment tool. *Journal of child psychology and psychiatry, and allied disciplines,* **50**(4), pp. 416-423.

CUNNINGHAM, C.E., PETTINGILL, P. and BOYLE, M.H., , The Brief Child and Family Phone Interview (BCFPI-3). (2006): A computerized intake and assesment tool. Interviewers manual. Available: http://www.bcfpi.ca/bcfpi/downloads.html. [0301, 2015].

CUTHBERT, R., ST. PIERRE, J., STEWART, S.L., COOK, S., JOHNSON, A.M. and LESCHIELD, A.W., 2011. Symptom Persistence in Seriously Emotionally Disordered Children: Findings of A Two-Year Follow-up after Residential Treatment. *Child youth and care forum,* **40**(4), pp. 280.

DEIGHTON, J., CROUDACE, T., FONAGY, P., BROWN, J., PATALAY, P. and WOLPERT, M., 2014. Measuring mental health and wellbeing outcomes for children

and adolescents to inform practice and policy: a review of child self-report measures. *Child and adolescent psychiatry and mental health,* **8**, pp. 14. eCollection 2014.

FERNANDEZ DE LA CRUZ, L., SIMONOFF, E., MCGOUGH, J.J., HALPERIN, J.M., ARNOLD, L.E. and STRINGARIS, A., 2015. Treatment of children with attention-deficit/hyperactivity disorder (ADHD) and irritability: results from the multimodal treatment study of children with ADHD (MTA). *Journal of the American Academy of Child and Adolescent Psychiatry,* **54**(1), pp. 70.e3.

FISHER, P.W., SHAFFER, D., PIACENTINI, J.C., LAPKIN, J., KAFANTARIS, V., LEONARD, H. and HERZOG, D.B., 1993. Sensitivity of the Diagnostic Interview Schedule for Children, 2nd edition (DISC-2.1) for specific diagnoses of children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **32**(3), pp. 666-73.

FRICK, J., 2016. *Clinical Assessment of Child and Adolescent Personality and Behavior.* third edition edn. Springer.

GALANTER, C.A. and PATEL, V.L., 2005a. Medical decision making: a selective review for child psychiatrists and psychologists. *Journal of child psychology and psychiatry, and allied disciplines,* **46**(7), pp. 675-689.

GALANTER, C.A. and PATEL, V.L., 2005b. Medical decision making: a selective review for child psychiatrists and psychologists. *Journal of child psychology and psychiatry, and allied disciplines,* **46**(7), pp. 675-689.

GHANIZADEH, A., MOHAMMADI, M.R. and YAZDANSHENAS, A., 2006. Psychometric properties of the Farsi translation of the Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version. *BMC psychiatry,* **6**, pp. 10.

GOODMAN, R. and SCOTT, S., 1999. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful? *Journal of abnormal child psychology,* **27**(1), pp. 17-24.

GOODMAN, R., 1999. The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *The Journal of Child Psychology and Psychiatry and Allied Disciplines,* **40**(5), pp. 791-799.

GOODYER, I.M., REYNOLDS, S., BARRETT, B., BYFORD, S., DUBICKA, B., HILL, J., HOLLAND, F., KELVIN, R., MIDGLEY, N., ROBERTS, C., SENIOR, R., TARGET, M., WIDMER, B., WILKINSON, P. and FONAGY, P., 2017. Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiority trial. *The lancet.Psychiatry,* **4**(2), pp. 109-119.

GORDON, M., ANTSHEL, K., FARAONE, S., BARKLEY, R., LEWANDOWSKI, L., HUDZIAK, J.J., BIEDERMAN, J. and CUNNINGHAM, C., 2006. Symptoms versus impairment: the case for respecting DSM-IV's Criterion D. *Journal of attention disorders,* **9**(3), pp. 465-475.

GRAYSON, P. and CARLSON, G.A., 1991. The utility of a DSM-III-R-based checklist in screening child psychiatric patients. *Journal of the American Academy of Child and Adolescent Psychiatry,* **30**(4), pp. 669-73.

GROL, R., 2013. *Improving patient care: the implementation of change in health care.* 2 edn. Chichester, West Sussex: Wiley Blackwell.

HARDING, K.E., TAYLOR, N.F. and LEGGAT, S.G., 2011. Do triage systems in healthcare improve patient flow? A systematic review of the literature. *Australian Health Review : A Publication of the Australian Hospital Association,* **35**(3), pp. 371-383.

HODGES, K., MCKNEW, D., BURBACH, D.J. and ROEBUCK, L., 1987. Diagnostic concordance between the Child Assessment Schedule (CAS) and the Schedule for Affective Disorders and Schizophrenia for School-age Children (K-SADS) in an outpatient sample using lay interviewers. *Journal of the American Academy of Child and Adolescent Psychiatry,* **26**(5), pp. 654-61.

HOOD, B.S., ELROD, M.G. and DEWINE, D.B., 2015. Treatment of Childhood Oppositional Defiant Disorder. *Current Treatment Options in Pediatrics,* **1**(2), pp. 155-167.

HUGHES, C.W., RINTELMANN, J., MAYES, T., EMSLIE, G.J., PEARSON, G. and RUSH, A.J., 2000. Structured interview and uniform assessment improves diagnostic reliability. *Journal of child and adolescent psychopharmacology,* **10**(2), pp. 119-131.

HUGHES, C.W., EMSLIE, G.J., WOHLFAHRT, H., WINSLOW, R., KASHNER, T.M. and RUSH, A.J., 2005. Effect of structured interviews on evaluation time in pediatric community mental health settings. *Psychiatric services (Washington, D.C.),* **56**(9), pp. 1098-1103.

JACKSON, D.A. and KING, A.R., 2004. Gender differences in the effects of oppositional behavior on teacher ratings of ADHD symptoms. *Journal of abnormal child psychology,* **32**(2), pp. 215-224.

JENSEN-DOSS, A., YOUNGSTROM, E.A., YOUNGSTROM, J.K., FEENY, N.C. and FINDLING, R.L., 2014. Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of consulting and clinical psychology,* **82**(6), pp. 1151-1162.

JONES, E., LUCEY, C. and WADLAND, L., 2000. Triage: a waiting list initiative in a child mental health service. *Psychiatric Bulletin,* **24**(2), pp. 57-59.

KAUFMAN, J., BIRMAHER, B., BRENT, D., RAO, U., FLYNN, C., MORECI, P., WILLIAMSON, D. and RYAN, N., 1997. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry,* **36**(7), pp. 980-988.

KESSLER, R.C., BERGLUND, P., DEMLER, O., JIN, R., MERIKANGAS, K.R. and WALTERS, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV

disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry,* **62**(6), pp. 593-602.

KIM, Y.S., CHEON, K.A., KIM, B.N., CHANG, S.A., YOO, H.J., KIM, J.W., CHO, S.C., SEO, D.H., BAE, M.O., SO, Y.K., NOH, J.S., KOH, Y.J., MCBURNETT, K. and LEVENTHAL, B., 2004. The reliability and validity of Kiddie-Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version- Korean version (K-SADS-PL-K). *Yonsei medical journal,* **45**(1), pp. 81-89.

KLINE, R.B., 2011. *Methodology in the Social Sciences. Principles and practice of structural equation modeling.* third edition edn. New York: Guilford Press.

KRANZLER, H.R., KADDEN, R.M., BABOR, T.F. and ROUNSAVILLE, B.J., 1994. Longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *The Journal of nervous and mental disease,* **182**(5), pp. 277-83.

KRANZLER, H.R., TENNEN, H., BABOR, T.F., KADDEN, R.M. and ROUNSAVILLE, B.J., 1997. Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *Drug and alcohol dependence,* **45**(1-2), pp. 93-104.

KUHN, C., AEBI, M., JAKOBSEN, H., BANASCHEWSKI, T., POUSTKA, L., GRIMMER, Y., GOODMAN, R. and STEINHAUSEN, H., 2017. Effective Mental Health Screening in Adolescents: Should We Collect Data from Youth, Parents or Both? *Child psychiatry and human development,* **48**(3), pp. 385-392.

KUNST, G., BLIDNER, J., ESRUBILSKY, V., LONGARELA, H. and VEGA, E., 2009. [Evaluation of the validity of Diagnostic Interview Schedule for Children IV (DISC IV) in the city of Buenos Aires]. *Vertex (Buenos Aires, Argentina),* **20**(86), pp. 245-51.

LAI, K.Y.C., 2006. The Establishment of a Triage System in a Child and Adolescent Psychiatry Clinic in Hong Kong. *Child and Adolescent Mental Health,* **11**(4), pp. 204-207.

LAKE, R., GEORGIOU, A., LI, J., LI, L., BYRNE, M., ROBINSON, M. and WESTBROOK, J.I., 2017. The quality, safety and governance of telephone triage and advice services – an overview of evidence from systematic reviews. *BMC health services research,* **17**(1), pp. 614-10.

LANDIS, J.R. and KOCH, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics,* **33**(1), pp. 159-174.

LAUTH, B., ARNKELSSON, G.B., MAGNUSSON, P., SKARPHEETHINSSON, G.A., FERRARI, P. and PETURSSON, H., 2010. Validity of K-SADS-PL (Schedule for Affective Disorders and Schizophrenia for School-Age Children--Present and Lifetime Version) depression diagnoses in an adolescent clinical population. *Nordic journal of psychiatry,* **64**(6), pp. 409-420.

LEFFLER, J.M., RIEBEL, J. and HUGHES, H.M., 2015. A Review of Child and Adolescent Diagnostic Interviews for Clinical Practitioners. *Assessment,* **22**(6), pp. 690-703.

LUCAS, C.P., 2003. *The use of structured diagnostic interviews in Clinical Child psychiatric practise.* Washington DC: American Psychiatric Publishing.

MALMBERG, M., RYDELL, A.M. and SMEDJE, H., 2003. Validity of the Swedish version of the Strengths and Difficulties Questionnaire (SDQ-Swe). *Nordic journal of psychiatry,* **57**(5), pp. 357-363.

MARTINEZ-GONZALEZ, N.A., DJALALI, S., TANDJUNG, R., HUBER-GEISMANN, F., MARKUN, S., WENSING, M. and ROSEMANN, T., 2014. Substitution of physicians by nurses in primary care: a systematic review and meta-analysis. *BMC health services research,* **14**, pp. 214.

MARTINEZ-GONZALEZ, N.A., ROSEMANN, T., DJALALI, S., HUBER-GEISMANN, F. and TANDJUNG, R., 2015. Task-Shifting From Physicians to Nurses in Primary Care and its Impact on Resource Utilization: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Medical care research and review : MCRR,* **72**(4), pp. 395-418.

MARUISH, M.E., 2004. *The use of psychological testing for treatment planning and outcomes assessment: General considerations, Volume 1, 3rd ed.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

MCCLELLAN, J.M. and WERRY, J.S., 2000. Introduction--research psychiatric diagnostic interviews for children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry,* **39**(1), pp. 19-27.

MCGRATH, A.M., HANDWERK, M.L., ARMSTRONG, K.J., LUCAS, C.P. and FRIMAN, P.C., 2004. The validity of the ADHD section of the Diagnostic Interview Schedule for Children. *Behavior modification,* **28**(3), pp. 349-74.

MELZER, D., TOM, B.D., BRUGHA, T.S., FRYERS, T. and MELTZER, H., 2002. Common mental disorder symptom counts in populations: are there distinct case groups above epidemiological cut-offs? *Psychological medicine,* **32**(7), pp. 1195-1201.

MERTEN, E.C., CWIK, J.C., MARGRAF, J. and SCHNEIDER, S., 2017. Overdiagnosis of mental disorders in children and adolescents (in developed countries). *Child and adolescent psychiatry and mental health,* **11**(1), pp. 5.

MORETTI, M.M. and OBSUTH, I., 2009. Effectiveness of an attachment-focused manualized intervention for parents of teens at risk for aggressive behaviour: The Connect Program. *Journal of adolescence,* **32**(6), pp. 1347-1357.

MORETTI, M.M., OBSUTH, I., CRAIG, S.G. and BARTOLO, T., 2015. An attachment-based intervention for parents of adolescents at risk: mechanisms of change. *Attachment & human development,* **17**(2), pp. 119-135.

MURIS, P. and STEERNEMAN, P., 2001a. The revised version of the Screen for Child Anxiety Related Emotional Disorders (SCARED--R): first evidence for its reliability and validity in a clinical sample. *The British journal of clinical psychology / the British Psychological Society,* **40**(Pt 1), pp. 35-44.

MURIS, P., DREESSEN, L., BOGELS, S., WECKX, M. and VAN MELICK, M., 2004. A questionnaire for screening a broad range of DSM-defined anxiety disorder symptoms in clinically referred children and adolescents. *Journal of child psychology and psychiatry, and allied disciplines,* **45**(4), pp. 813-820.

MURIS, P., MERCKELBACH, H., VAN BRAKEL, A. and MAYER, A.B., 1999. The revised version of the screen for child anxiety related emotional disorders (scared-r): Further evidence for its reliability and validity. *Anxiety, Stress, and Coping,* **12**(4), pp. 411-425.

MURIS, P. and STEERNEMAN, P., 2001b. The revised version of the Screen for Child Anxiety Related Emotional Disorders (SCARED--R): first evidence for its reliability and validity in a clinical sample. *The British journal of clinical psychology,* **40**(Pt 1), pp. 35-44.

MURIS, P., MERCKELBACH, H., KINDT, M., BÖGELS, S., DREESSEN, L., DORP, C.V., HABETS, A., ROSMULLER, S. and SNIEDER, N., 2001. The utility of screen for child anxiety related emotional disorders (scared) as a tool for identifying children at high risk for prevalent anxiety disorders. *Anxiety, Stress, & Coping,* **14**(3), pp. 265-283.

MUTHÉN, L.K. AND MUTHÉN, B.O., 2017. *Mplus User's Guide.* Eight Edition edn. Los Angeles: Muthén & Muthén.

OFFORD, D.R., BOYLE, M.H., SZATMARI, P., RAE-GRANT, N.I., LINKS, P.S., CADMAN, D.T., BYLES, J.A., CRAWFORD, J.W., BLUM, H.M. and BYRNE, C., 1987. Ontario Child Health Study. II. Six-month prevalence of disorder and rates of service utilization. *Archives of General Psychiatry,* **44**(9), pp. 832-836.

OLESEN, J., GUSTAVSSON, A., SVENSSON, M., WITTCHEN, H.U., JONSSON, B., CDBE2010 STUDY GROUP and EUROPEAN BRAIN COUNCIL, 2012. The economic cost of brain disorders in Europe. *European journal of neurology,* **19**(1), pp. 155-162.

PARKIN, A., FRAKE, C. and DAVISON, I., 2003. A Triage Clinic in a Child and Adolescent Mental Health Service. *Child and Adolescent Mental Health,* **8**(4), pp. 177-183.

PILKONIS, P.A., HEAPE, C.L., RUDDY, J. and SERRAO, P., 1991a. Validity in the diagnosis of personality disorders: The use of the LEAD standard. *Psychological Assessment,* **3**(1), pp. 46-54.

PILKONIS, P.A., HEAPE, C.L., RUDDY, J. and SERRAO, P., 1991b. Validity in the diagnosis of personality disorders: The use of the LEAD standard. *Psychological Assessment,* **3**(1), pp. 46-54.

POLANCZYK, G.V., EIZIRIK, M., ARANOVICH, V., DENARDIN, D., DA SILVA, T.L., DA CONCEICAO, T.V., PIANCA, T.G. and ROHDE, L.A., 2003. Interrater agreement for the schedule for affective disorders and schizophrenia epidemiological version for school-age children (K-SADS-E). *Revista brasileira de psiquiatria (Sao Paulo, Brazil : 1999),* **25**(2), pp. 87-90.

POLANCZYK, G.V., SALUM, G.A., SUGAYA, L.S., CAYE, A. and ROHDE, L.A., 2015. Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of child psychology and psychiatry, and allied disciplines,* **56**(3), pp. 345-365.

REICH, W., 2000. Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child & Adolescent Psychiatry,* **39**(1), pp. 59-66.

RETTEW, D.C., LYNCH, A.D., ACHENBACH, T.M., DUMENCI, L. and IVANOVA, M.Y., 2009. Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International journal of methods in psychiatric research,* **18**(3), pp. 169-184.

ROELOFS, J., MURIS, P., BRAET, C., ARNTZ, A. and BEELEN, I., 2015. The Structured Clinical Interview for DSM-IV Childhood Diagnoses (Kid-SCID): first psychometric evaluation in a Dutch sample of clinically referred youths. *Child psychiatry and human development,* **46**(3), pp. 367-75.

RUTTER, M., 2008. *Rutter's child and adolescent psychiatry [Elektronisk resurs].* 5. ed. edn. Malden, Mass.: Blackwell Pub.

SADOCK, B. SADOCK, V. KAPLAN, H., 2005. *Kaplan & Sadock's comprehensive textbook of psychiatry.* 8th ed. edn. Philadelphia, Pa: Lippincott Williams & Wilkins.

SCHWAB-STONE, M.E., SHAFFER, D., DULCAN, M.K., JENSEN, P.S., FISHER, P., BIRD, H.R., GOODMAN, S.H., LAHEY, B.B., LICHTMAN, J.H., CANINO, G., RUBIO-STIPEC, M. and RAE, D.S., 1996. Criterion validity of the NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3). *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**(7), pp. 878-88.

SHAFFER, D., FISHER, P., LUCAS, C.P., DULCAN, M.K. and SCHWAB-STONE, M.E., 2000. NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry,* **39**(1), pp. 28-38.

SHANEE, N., APTER, A. and WEIZMAN, A., 1997. Psychometric properties of the K-SADS-PL in an Israeli adolescent clinical population. *The Israel journal of psychiatry and related sciences,* **34**(3), pp. 179-186.

SHEEHAN, D.V., SHEEHAN, K.H., SHYTLE, R.D., JANAVS, J., BANNON, Y., ROGERS, J.E., MILO, K.M., STOCK, S.L. and WILKINSON, B., 2010. Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *The Journal of clinical psychiatry,* **71**(3), pp. 313-26.

SHERRILL, J.T. and KOVACS, M., 2000. Interview schedule for children and adolescents (ISCA). *Journal of the American Academy of Child & Adolescent Psychiatry,* **39**(1), pp. 67-75.

SKODOL, A.E., ROSNICK, L., KELLMAN, D., OLDHAM, J.M. and HYLER, S.E., 1988. Validating structured DSM-III-R personality disorder assessments with longitudinal data. *The American Journal of Psychiatry,* **145**(10), pp. 1297-9.

SMEDJE, H., BROMAN, J.E., HETTA, J. and VON KNORRING, A.L., 1999. Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *European child & adolescent psychiatry,* **8**(2), pp. 63-70.

SOCIALSTYRELSEN, 2017. *Utvecklingen av psykisk ohälsa bland barn och unga vuxna.*

SPITZER, R.L., ENDICOTT, J. and ROBINS, E., 1978. Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry,* **35**(6), pp. 773-82.

SPITZER, R.L., 1983. Psychiatric diagnosis: are clinicians still necessary? *Comprehensive psychiatry,* **24**(5), pp. 399-411.

STEVANOVIC, D., TADIC, I. and KNEZ, R., 2014. Are antidepressants effective in quality of life improvement among children and adolescents? A systematic review. *CNS spectrums,* **19**(2), pp. 134-141.

THOMSEN, P.H., TORP, N.C., DAHL, K., CHRISTENSEN, K., ENGLYST, I., MELIN, K.H., NISSEN, J.B., HYBEL, K.A., VALDERHAUG, R., WEIDLE, B., SKARPHEDINSSON, G., VON BAHR, P.L. and IVARSSON, T., 2013. The Nordic long-term OCD treatment study (NordLOTS): rationale, design, and methods. *Child and adolescent psychiatry and mental health,* 7(1), pp. 41.

ULLOA, R.E., ORTIZ, S., HIGUERA, F., NOGALES, I., FRESAN, A., APIQUIAN, R., CORTES, J., ARECHAVALETA, B., FOULLIUX, C., MARTINEZ, P., HERNANDEZ, L., DOMINGUEZ, E. and DE LA PENA, F., 2006. Interrater reliability of the Spanish version of Schedule for Affective Disorders and Schizophrenia for School-Age Children--Present and Lifetime version (K-SADS-PL). *Actas Espanolas de Psiquiatria,* **34**(1), pp. 36-40.

VOLKMAR, F.R., 1996. Childhood and adolescent psychosis: a review of the past 10 years. *Journal of the American Academy of Child and Adolescent Psychiatry,* **35**(7), pp. 843-851.

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., IYENGAR, S., MARCH, J.S. and KENDALL, P.C., 2008. Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *New England Journal of Medicine,* **359**(26), pp. 2753-66.

WECKERLY, J., AARONS, G.A., LESLIE, L.K., GARLAND, A.F., LANDSVERK, J. and HOUGH, R.L., 2005. Attention on inattention: the differential effect of caregiver education on endorsement of ADHD symptoms. *Journal of developmental and behavioral pediatrics : JDBP,* **26**(3), pp. 201-208.

WEHRY, A.M., BEESDO-BAUM, K., HENNELLY, M.M., CONNOLLY, S.D. and STRAWN, J.R., 2015. Assessment and treatment of anxiety disorders in children and adolescents. *Current psychiatry reports,* **17**(7),.

WILBERG, T., DAMMEN, T. and FRIIS, S., 2000. Comparing Personality Diagnostic Questionnaire-4+ with Longitudinal, Expert, All Data (LEAD) standard diagnoses in a sample with a high prevalence of axis I and axis II disorders. *Comprehensive Psychiatry,* **41**(4), pp. 295-302.

YOUNG, J.G., O'BRIEN, J.D., GUTTERMAN, E.M. and COHEN, P., 1987. Research on the clinical interview. *Journal of the American Academy of Child and Adolescent Psychiatry,* **26**(5), pp. 613-20.

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

FACULTY OF MEDICINE

Department of Clinical Sciences, Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2019:54
ISBN 978-91-7619-783-7
ISSN 1652-8220