Thesis for doctoral degree (Ph.D.)
2012

On the Children's Global Assessment Scale (CGAS)

Anna Lundh

Karolinska Institutet

Thesis for doctoral degree (Ph.D.) 2012
On the Children's Global Assessment Scale (CGAS)
Anna Lundh

DEPARTMENT OF CLINICAL NEUROSCIENCE
Karolinska Institutet, Stockholm, Sweden

ON THE CHILDREN'S GLOBAL ASSESSMENT SCALE (CGAS)

Anna Lundh

Karolinska Institutet

Stockholm 2012

Illustrations by Olle Nilsson.

Published by Karolinska Institutet. Printed by Larserics Digital Print AB.

ISBN 978-91-7457-619-1

ABSTRACT

Rating scales and diagnostic instruments have become increasingly important tools in psychiatric care over the past several decades. Using these standardized tools to collect information and evaluate patients enables streamlined evidence-based diagnosis and assessments of functioning. This thesis revolves around the Children's Global Assessment Scale (CGAS), a widely used rating scale designed to measure how a child functions psychosocially in daily life.

In Paper I, the inter-rater reliability (IRR) and accuracy of CGAS ratings among untrained raters (n=703) were assessed in a large clinical setting. The untrained raters scored case vignettes significantly higher than the gold standard established by experts. The IRR in terms of intra-class correlation coefficient (ICC) was 0.73. Social workers and psychologists were significantly more likely to have overall aberrant ratings than medical doctors. The results suggest that reliability and accuracy is moderate when CGAS is used in a clinical setting with untrained raters.

In Paper II, two training methods to improve CGAS ratings were evaluated. Untrained raters (n=648) were randomised to training either by a CD-ROM or in a seminar. In addition, 55 raters formed a non-randomised comparison group. There was no significant difference between the two training groups at the 12-month follow-up. The untrained comparison group improved at the same order of magnitude as the training groups. The ICCs at baseline and at end-of-study were 0.71/0.78 (seminar), 0.76/0.78 (CD-ROM), and 0.67/0.79 (comparison). These results speak in favour of using the less resource-demanding computer-based training. However, the overall training effect was too small to be clinically relevant. Future evaluations of training methods should include a control group to control for unspecific learning effects.

Registration of CGAS ratings in the clinical database Pastill was initiated at the completion of the training activity carried out for Paper II. This enabled a study on the effectiveness of child psychiatric treatment by examining the change in psychosocial functioning as measured by CGAS described in *Paper III*. The change in CGAS ratings between intake and case closure was investigated for 12,613 patients. CGAS improved during the course of treatment across all diagnostic groups. In the mood disorder group, several psychotherapies were associated with improved outcome whereas medication was not. In the Attention-Deficit Hyperactivity Disorder (ADHD) group, medication with central stimulants was not associated with improvement. Treatment-as-usual was found to be less effective than clinical trials have indicated, particularly for the ADHD group, suggesting that results from clinical trials cannot be extrapolated to routine child psychiatric care. Hence, more studies of ADHD and mood disorders are needed to investigate the effectiveness of medication/psychotherapy in regular treatment.

In Paper IV, the Pastill data were linked to Swedish national registers to see whether CGAS ratings at end-of-treatment predict long-term negative outcomes in young adults. To do this, 4,876 patients were followed up prospectively. Patients with CGAS≤60 at end-of-treatment had a moderately increased risk of a criminal conviction and a substantially increased risk for bipolar disorder and borderline personality disorder during follow-up compared to patients with CGAS>60. Low CGAS ratings were not associated with depression, suicide attempt, or substance misuse. Hence, CGAS ratings provide specific long-term prognostic information, and adolescents with CGAS scores below 60 at end-of-treatment should be considered for intensified follow-up.

SAMMANFATTNING

Skattningsskalor och diagnostiska instrument har fått en allt större utbredning inom barn- och ungdomspsykiatrin de senaste 30 åren. Dessa instrument hjälper behandlare och forskare att beskriva psykiatriska diagnoser, symtom och funktionsnivåer på ett mer standardiserat och likvärdigt sätt. Denna avhandling beskriver Children's Global Assessment Scale (CGAS), en spridd och användbar skala för att mäta hur barn och ungdomar fungerar i vardagen (hemma, i skolan och med kamrater).

Delarbete I utvärderar reliabiliteten mellan olika bedömare och hur skattningarna överensstämmer med expertskattningar. Fem erfarna kliniker skattade fem fallvinjetter för att skapa en gold standard. Otränade kliniker (n=703) skattade samma vinjetter och resultatet visade att de skattade högre än gold standard. Intra-klass korrelationskoefficienten (ICC) var 0,73. Socionomer och psykologer hade en högre risk jämfört med läkare för att skatta alla fall tydligt avvikande från gold standard. Resultaten visar en måttlig reliabilitet mellan bedömare och en måttlig överensstämmelse i jämförelse med experter när otränade kliniker använder CGAS.

Delarbete II undersöker om två olika utbildningsmetoder, seminarier och CD-skiva, kan påverka skattningarnas kvalitet. Otränade kliniker (n=648) lottades mellan de bägge träningsmetoderna, och 55 kliniker deltog i den otränade jämförelsegruppen. Det fanns ingen skillnad mellan grupperna 12 månader efter träning. Jämförelsegruppen förbättrades också i samma storleksordning under uppföljningstiden. ICC vid början och vid slutpunkten var 0,71/0,78 (seminarium), 0,76/0,78 (CD), och 0,67/0,79 (jämförelsegrupp). Resultaten talar för att använda den datorbaserade träningen i ökad utsträckning, men den faktiska förbättringen var liten och inte kliniskt relevant.

I samband med den omfattande utbildningsaktiviteten ovan påbörjade BUP i Stockholm rutinmässiga CGAS-bedömningar som lagras i en klinisk databas, Pastill. *Delarbete III* studerade "verklighetens" barnpsykiatriska vård utifrån skillnader i psykosocial funktion mätt med CGAS. Skillnaden mellan CGAS vid nybesök och avslutad behandling undersöktes hos 12,613 patienter. Samtliga diagnosgrupper visade förbättring av CGAS efter behandling. Olika typer av psykoterapier visade på samband med förbättring av CGAS i depressionsgruppen, men däremot ej farmakologisk behandling. För Attention-Deficit/Hyperactivity Disorder (ADHD) var inte heller medicinering med centralstimulantia associerat med förbättring av CGAS. Resultaten visade sämre utfall i jämförelse med kliniska prövningar, framför allt för ADHD, vilket gör att man inte självklart kan översätta kliniska studier till verklighetens vård. Både farmakologisk behandling och psykoterapi behöver utvärderas ytterligare för ADHD och depression i naturalistiska kliniska sammanhang.

Slutligen länkades Pastilldata till nationella register för att undersöka om CGAS vid avslutad behandling kan predicera framtida psykisk sjukdom och kriminalitet hos unga vuxna. I *delarbete IV* följdes 4,876 patienter i 1-1½ år. Ungdomar med CGAS≤60 vid avslutad behandling visade en ökad risk för kriminalitet, bipolär sjukdom och borderline personlighetsstörning jämfört med gruppen CGAS>60, men däremot ingen riskökning för depression, suicidförsök och missbruk. CGAS skattningar vid avslutad behandling kan ge information om framtida utfall. Ungdomar med CGAS skattningar som är 60 och lägre vid avslutad behandling bör uppmärksammas med tanke på behov av intensifierad uppföljning

LIST OF PUBLICATIONS

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I. Lundh A, Kowalski J, Sundberg CJ, Gumpert C, Landén M.

Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings.

Psychiatry Research 2010, 177, 206-210

II. Lundh A, Kowalski J, Sundberg CJ, Landén M.

A comparison of seminar and computer-based training on the accuracy and reliability of raters using the Children's Global Assessment Scale (CGAS).

Administration and Policy in Mental Health and Mental Health Services Research 2011, DOI:10.1007/s10488-011-0369-5

III. Lundh A, Forsman M, Serlachius E, Lichtenstein P, Landén M.

The effectiveness of child psychiatric treatment.

Submitted

IV. Lundh A, Forsman M, Serlachius E, Långström N, Lichtenstein P, Landén M.

The Children's Global Assessment Scale (CGAS) predicts negative outcomes in early adulthood.

Submitted

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AAU	Assessment As Usual
ADHD	Attention-Deficit Hyperactivity Disorder
ADL	Activities in Daily Life
CAMHS	Child and Mental Health Services
CAP	Child and Adolescent Psychiatry
CBCL	Child Behaviour Checklist
CBT	Cognitive Behavioural Therapy
CD	Conduct disorder
CD-ROM	Compact disc-Read Only Memory
CGAS	Children's Global Assessment Scale
ESQ	Experience of Service Questionnaire
CORC	Child and Mental Health Services Outcome Research Consortium
DSM-III-R	Diagnostic and Statistical Manual of Mental Disorders, third version, revised
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, fourth version, text revision
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, fifth version, planned 2013
FLX	Fluoxetine
GAF	Global Assessment of Functioning
GAPD	Global Assessment of Psychosocial Disability
GAS	Global Assessment Scale
HSRS	Health Sickness Rating Scale
IBD	Inflammatory Bowel Disease
ICC	Intra-class Correlation Coefficient
IMI	Imipramine
IRR	Inter-Rater Reliability
K-SADS	Kiddie-Schedule for Affective Disorders and Schizophrenia
MPH	Methylphenidate
NST	Nondirective Supportive Therapy
OCD	Obsessive-Compulsive Disorder
OROS	Osmotic-controlled Release Oral delivery System
PLC	Placebo

PXT	Paroxetine
RCT	Randomised Controlled Trial
SBFT	Systemic Behavioural Family Therapy
SDQ	Strengths and Difficulties Questionnaire
SSRI	Selective Serotonin Reuptake Inhibitors
TADS	Treatment for Adolescent Depression Study
TAU	Treatment-As-Usual

1 PROLOGUE

Does child psychiatric treatment make a difference? As a young child psychiatrist I often wondered if it did, and I was frustrated over the lack of a clear answer. As I thought about undertaking Ph.D. research, I realized this question was too comprehensive to be addressed in one Ph.D. project. If I wanted to contribute to the evidence base for child psychiatry's effectiveness I needed a more specific research question.

During my residency in child psychiatry I concentrated on learning as much as possible about diagnoses, diagnostic criteria, and diagnostic systems. However, at the same time I started to find out more about functional impairment, not least due to the growing body of knowledge on neuropsychiatric disorders where the level of functioning is the key to treatment planning. I realised that although symptom reduction usually precedes functional improvement in child psychiatric practise, treatment cannot be considered truly successful until the patient also reaches a near-normal or normal functional level. This means that the ultimate goal of child psychiatry is to improve children's psychosocial function. Hence, it does not suffice that a depressed adolescent no longer meets the diagnostic criteria for depression. He or she should also be able to return to school, see friends again and get along at home.

This line of reasoning led me to the question: how can child psychiatry measure psychosocial functioning?

While I pondered this question, the inpatient unit at Child and Adolescent Mental Health Services (CAMHS) in Stockholm decided to start using the Children's Global Assessment Scale (CGAS) in clinical work. This gave me the opportunity to plan a training activity and compare the effectiveness of two different training methods. This thesis presents the results from these training activities. The CGAS ratings that CAMHS in Stockholm began to routinely collect were then used to conduct two large-scale studies to answer the following questions: i) Is child psychiatric treatment-as-usual effective in terms of improved psychosocial functioning measured by CGAS, and ii) Do CGAS ratings predict future mental health disorders and criminality?

2 INTRODUCTION

2.1 RATING SCALES AND DIAGNOSTIC INSTRUMENTS

Rating scales and diagnostic instruments are attempts to create objective measures of psychiatric symptoms. These tools can consist of a battery of questions, a series of statements requiring a yes or no answer, numbered scales corresponding to emotional states, functional states, or behavioural patterns, checklists, self-reports, or other standardized ways to assess a patient's symptoms and needs (Myers and Winters 2002; Myers and Winters 2002; Ohan et al. 2002; Winters et al. 2002; Collett et al. 2003; Collett et al. 2003; Winters et al. 2005). Some tools are clinician-rated, some are self-rated by the patient, and some are rated by parents or others.

By allowing for a high degree of standardization, rating scales and diagnostic instruments can contribute to many aspects of psychiatry (Shaffer et al. 1999; Fitzpatrick et al. 2011). First, standardized instruments provide measureable data that help clinicians make evidence-based diagnoses and assess function (Hodges 1993; Zanarini et al. 2000; Gilbody et al. 2002; Mazade and Glover 2007; Wolpert et al. 2007; Garland et al. 2010; Stein et al. 2010; Bickman et al. 2011; Follan et al. 2011). Second, these tools are key to obtain replicable data in clinical trials and epidemiological surveys (Hoagwood et al. 1995; Hoagwood et al. 1996; Jensen et al. 1996; Leaf et al. 1996; Fonagy 1997; Myers and Winters 2002; Kim-Cohen et al. 2003; Galanter and Patel 2005). Third, health care providers can use rating instruments to prioritise health care resources, and to compare different clinical units in quality assurance and improvement programmes (Wolpert et al. 2007). A fourth potential application is to predict future risk for adverse events, such as mental disorders, criminality, substance misuse, accidents, or other factors associated with low quality of life. Identifying children at risk for these outcomes might lead to a more prudent use of healthcare resources by allowing help to be allocated the most vulnerable.

Rating scales and diagnostic instruments have become increasingly widespread in psychiatric care over the past decades (Myers and Winters 2002), and together with evidence-based guidelines they increasingly assist decision-making for clinicians and patients (Kendall et al. 2005; Wolpert et al. 2006; National Institute for Health and Clinical Excellence 2008; BUP divisionen 2010; Stein et al. 2010; Wolpert et al. 2011). This development is in accordance with the fifth edition of the Diagnostic Statistical Manual of Mental Disorders (DSM-5) draft, which suggests that assessment and diagnosis should be based at least in part on rating scales (American Psychiatric Association 2010).

2.2 THE DEVELOPMENT OF SCALES TO RATE MENTAL HEALTH AND FUNCTIONAL IMPAIRMENT

2.2.1 Historical background

Prior to 1960, psychiatry used "improvement scales" that measured how much a patient had improved with treatment, but did not assess the patient's overall level of mental health or functioning. Since all patients have different baselines, a "much improved" patient could have more severe symptoms than a "slightly improved" patient. In 1962, the Psychotherapy Research Project of The Menninger Foundation tried to figure out how clinicians could quantify *mental health* over time in such a way that the changes would be comparable between patients. The members of the project agreed upon seven criteria for judging *mental health:* 1) the ability to function autonomously, 2) the seriousness of symptoms, 3) the degree of discomfort, 4) the effect upon the environment, 5) the utilization of abilities, 6) the quality of interpersonal relationships, and 7) the breadth and depth of interests. The project resulted in a 100-point scale where standard patients were graded depending on their degree of overall *mental health* (Luborsky and Bachrach 1974). The score was obtained by comparing the observed patient with a series of short vignettes. Launched as the Health-Sickness Rating Scale (HSRS), it became one of the first rating scales that included a measure of psychosocial functioning in its assessment of *mental health* (Luborsky 1962). A limitation with the HSRS global assessment, however, was that diagnostic terms were included in the descriptions. For example, one rating corresponded to *"Most clear-cut, overt psychoses, psychotic characters, severe addictions (which require hospital care)"*. Hence, even though the scale included psychosocial functioning, it was essentially diagnosis-based, limiting its usefulness for measuring changes in symptoms and functioning (Endicott et al. 1976).

This encouraged Endicott and Spitzer at Columbia University to develop a new rating scale that was launched in 1975: the Global Assessment Scale (GAS) (1976). The diagnostic terms were removed and examples with behavioural descriptions were added. Also, the GAS had a narrower scope than HSRS's assessment of the overall level of mental health, and aimed only to evaluate the overall level of functioning.

2.3 THE CHILDREN'S GLOBAL ASSESSMENT SCALE (CGAS)

In the early 1980s, Shaffer and colleagues - also at Columbia University - used GAS as a basis to develop a new scale for use with children and adolescents, 4-16 years old (Figure 1) (Shaffer et al. 1983). The scale was coined The Children's Global Assessment Scale (CGAS). CGAS is a clinician-rated tool that aims to assess the child's lowest level of global psychosocial functioning (at home, at school, and with peers) during the last month, taking into account all available information. The time period may be adjusted depending on the purpose of the ratings. However, it is reasonable to let at least one time period pass between repeated ratings to avoid overlap of the assessed time period. The scoring ranges from 1 (the most impaired level) to 100 (superior level of functioning). The scale is divided into 10-point intervals that are headed with a description of the level of functioning. The anchor points also contain examples of behaviours and life situations adequate for children and adolescents matching the level of functioning for each interval.

CGAS is currently routinely used in many countries to measure psychiatric treatment outcomes in spite of the limited amount of research evaluating CGAS in large-scale clinical settings. Usually CGAS is administered as one scale among a battery of instruments (Gold et al. 2009; National Board of Health and Welfare 2009). In the UK, for example, the influential CAMHS Outcome Research Consortium (CORC) chose to use CGAS along with the Strengths and Difficulties Questionnaire (SDQ) and the Experience of Service Questionnaire (ESQ) as a basis for treatment evaluation (Wolpert et al. 2007).

CGAS has also been used in several longitudinal and epidemiological studies (Bird et al. 1993; Milne et al. 1995; Leaf et al. 1996; Steinhausen and Metzke 2001; Canino et al. 2004; Petersen et al. 2006; Ayton et al. 2009; Bella et al. 2011). In addition, CGAS has been used as one of several outcome measures in both randomised controlled trials and observational studies (Table 1).

Figure 1. Children's Global Assessment Scale, original version 1983 (Shaffer et al. 1983).

CHILDREN'S GLOBAL ASSESSMENT SCALE

For children 4–16 years of age

David Shaffer, M.D., Madelyn S. Gould, Ph.D.
Hector Bird, M.D., Prudence Fisher, B.A.

Adaptation of the Adult Global Assessment Scale

(Robert L. Spitzer, M.D., Miriam Gibson, M.S.W., Jean Endicott, Ph.D.)

Rate the subject's most impaired level of general functioning for the specified time period by selecting the *lowest* level which describes his/her functioning on a hypothetical continuum of health-illness. Use intermediary levels (e.g., 35, 58, 62).

Rate actual functioning regardless of treatment or prognosis. The examples of behavior provided are only illustrative and are not required for a particular rating.

Specified time period: 1 month

100–91 **Superior functioning** in all areas (at home, at school, and with peers), involved in a range of activities and has many interests (e.g., has hobbies or participates in extracurricular activities or belongs to an organized group such as Scouts, etc.). Likeable, confident, "everyday" worries never get out of hand. Doing well in school. No symptoms.

90–81 **Good functioning in all areas.** Secure in family, school, and with peers. There may be transient difficulties and "everyday" worries that occasionally get out of hand (e.g., mild anxiety associated with an important exam, occasional "blow-ups" with siblings, parents, or peers).

80–71 **No more than slight impairment in functioning** at home, at school, or with peers. Some disturbance of behavior or emotional distress may be present in response to life stresses (e.g., parental separations, deaths, birth of a sib) but these are brief and interference with functioning is transient. Such children are only minimally disturbing to others and are not considered deviant by those who know them.

70–61 **Some difficulty in a single area, but generally functioning pretty well**, (e.g., sporadic or isolated antisocial acts, such as occasionally playing hooky or petty theft; consistent minor difficulties with school work, mood changes of brief duration; fears and anxieties which do not lead to gross avoidance behavior; self-doubts). Has some meaningful interpersonal relationships. Most people who do not know the child well would not consider him/her deviant but those who do know him/her well might express concern.

60–51 **Variable functioning with sporadic difficulties or symptoms in several but not all social areas.** Disturbance would be apparent to those who encounter the child in a dysfunctional setting or time but not to those who see the child in other settings.

50–41 **Moderate degree of interference in functioning in most social areas or severe impairment of functioning in one area**, such as might result from, for example, suicidal preoccupations and ruminations, school refusal and other forms of anxiety, obsessive rituals, major conversion symptoms, frequent anxiety attacks, frequent episodes of aggressive or other antisocial behavior with some preservation of meaningful social relationships.

40–31 **Major impairment in functioning in several areas and unable to function in one of these areas**, i.e. disturbed at home, at school, with peers, or in the society at large, e.g., persistent aggression without clear instigation; markedly withdrawn and isolated behavior due to either mood or thought disturbance, suicidal attempts with clear lethal intent. Such children are likely to require special schooling and/or hospitalization or withdrawal from school (but this is not sufficient criterion for inclusion in this category).

30–21 **Unable to function in almost all areas**, e.g., stays at home, in ward or in bed all day without taking part in social activities OR severe impairment in reality testing OR serious impairment in communication (e.g., sometimes incoherent or inappropriate).

20–11 **Needs considerable supervision** to prevent hurting others or self, e.g. frequently violent, repeated suicide attempts OR to maintain personal hygiene OR gross impairment in all forms of communication, e.g. severe abnormalities in verbal and gestural communication, marked social aloofness, stupor, etc.

10–1 **Needs constant supervision** (24-hour care) due to severely aggressive or self-destructive behavior or gross impairment in reality testing, communication, cognition, affect, or personal hygiene.

Table 1. Randomised controlled trials and observational studies with CGAS as one of several outcome measures.

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
ADHD	(Berek et al. 2011)	Prospective non-interventional open-label study	Methylphenidate OROS n=822, 6-18 y 12 weeks	No info/ No info	Treating physician parent interview/ No info	58.5	69.6	11.1
ADHD	(Findling et al. 2008)	Open-label pilot trial	Aripiprazole n=36, 8-12 y 6 weeks	No info/ No info	No info/ No info	62.8	71.0	8.2
ADHD	(Kratochvil et al. 2007)	Retrospective study Clinical records	Atomoxetine n=22, 5-6 y 8 weeks	No info/ No info	3 centers, conference call/ No info	53.2	Not reported	18.9
ADHD	(Preuss et al. 2006)	Prospective observational study	ADORE study n=1,478, 6-18 y 2 years	No info/ No info	MD 10 European countries/ No info	55.2	Only baseline data	
ADHD	(Steinhausen et al. 2006)	Cohort study	ADORE study Impact of ADHD n=1,478, 6-18 y 2 years	No info/ No info	MD 10 European countries/ No info	58.5 55.8 53.3 54.8 57.2 50.7	ADHD only ADHD + Depr/Anxiety ADHD + ODD/CD ADHD + Tics/Tourette's ADHD + Coord. probls ADHD + ≥2 conditions	
ADHD	(Szobot et al. 2004)	RCT, double-blind, placebo-controlled	Methylphenidate n=36, 8-17y 4 days	No info/ No info	CAP MD/ No info	52.1 (MPH) 54.7 (PLC)	69.1 - 4 days 59.7	17.0 5.0

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Anxiety disorder	(McShane et al. 2007)	Uncontrolled intervention study	Multimodal treatment n=24, mean 15 y 1 year?	No info/ No info	No info/ No info	58	61	3
Anxiety disorder	(Monga et al. 2009)	Pilot study, uncontrolled	CBT n=32, 5-7 y 12 weeks	Previous month/No info	No info/ No info	46.2 45.6	61.2 55.7	15.0 10.1
Anxiety and depressive symptoms	(Muratori et al. 2002)	Non-randomised controlled study	Brief dynamic PT n=30, 6-11 y 11 sessions, (5 individual)	No info/ No info	Independent blind observer/ No info	61.7 (BDPT) 59.0 (PLC)	75.3 (6m FU) 66.3 (18m FU) 66.3 (6m FU) 69.7 (18m FU)	
Bipolar disorder	(Barzman et al. 2004)	Uncontrolled, retrospective chart review	Aripiprazole n=30, 5-19 y 1-9 months	No info/ No info	2 CAP MD/ No info	48.0	65.0	17.0
Bipolar disorder, mania	(Pavuluri et al. 2005)	Prospective open trial	Divalproex sodium n=34, 5-18 y 6 months	No info/ No info	Clinicians master level/ 6 months training, IRR 0.90-0.98	43.8	56.0	12.2
Bipolar mania, Schizo-phrenia	(Stewart et al. 2009)	Open-label follow-up study	Ziprasidone, low Ziprasidone, high dose n=63, 10-17 y 6 months	One month/ Lowest level	No info/ Instructions to all raters prior study	41.7 (low) 39.0 (high)	Reported only in diagram	14.4 (3w) 17.4 (3w)
Bipolar disorder	(Tillman and Geller 2007)	Descriptive study, comparing two populations	RCT group Consecutive study group n=243, 7-15 y	No info/ No info	Research nurses/consens us conferences	38.7 43.7		

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Conduct disorder	(Masi et al. 2006)	Naturalistic follow-up study	Olanzapine n=23, 11-17 y 6-12 months	No info/ lowest level	Experienced CAP MD/Consensus on KSADS	38.2	50.0	11.8
Depressive disorder	(Fitzpatrick et al. 2011)	Comparative study research and clinical assessment	n=100, 12-15 y	No info/ No info	Research psychiatrist and psychologist/No info	44.9 (depr disorder) 53.7 (no depr disorder)		
Depressive disorder	(Keller et al. 2001)	RCT	Paroxetine, Imipramine, Placebo n=275, 12-18 y 8 weeks	No info/ No info	12 centers/ No info	42.7 (PXT) 42.5 (IMI) 42.8 (PLC)	Not reported	
Depressive disorder	(Mufson et al. 2004)	RCT	Interpersonal Psychotherapy Treatment as usual n=53, 12-18 y 12 weeks	No info/ No info	Psycholo-gist or social worker, blinded/ Trained, 21 clinicians rated 20 vignettes ICC 0.83	52.6 (IPT) 52.7 (TAU)	66.7 (IPT) 59.5 (TAU)	14.1 6.8
Depressive disorder	(Abeles et al. 2009)	Uncontrolled intervention study	Computerized CBT n=23, 12-16 y 1-8 sessions	No info/ No info	No info/ No info	47.0	59.6 70.1 (12w)	12.6 23.1
Depressive disorder	(Brent et al. 1997)	RCT	CBT/SBFT/NST n=107, 13-18 y 12-16 weeks	No info/ No info	No info/ No info	58.8 (CBT) 54.5 (SBFT) 56.3 (NST)	65.4 63.5 63.3	6.6 9.0 7.0

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Depressive disorder	(Emslie et al. 1997)	RCT	Fluoxetine (FLX) n=96, 7-17 y 8 weeks	No info/ No info	3 CAP MD/ Experienced	47.9 (FLX) 48.4 (PLC)	63.9 60.1	16.0 11.7 ns diff
Depressive disorder	(Mufson et al. 1999)	RCT	Interpersonal Psychotherapy/Control n=48, 12-18 y 12 weeks	No info/ No info	Independent evaluator/ No info	52	No sign group diff at week 12	Not reported
Depressive disorder + IBD	(Szigethy et al. 2007)	RCT	CBT Comparison treatment n=41, 11-17 y 12-14 weeks	No info/ Current level of impairment	Independent evaluators, psychologists/ Training. IRR 0.90	61.8 (CBT) 62.4 (TAU)	69.9 (CBT) 62.8 (TAU)	7.8 0.9
Depressive disorder	(Vitiello et al. 2006)	RCT	Fluoxetine + CBT Fluoxetine CBT Placebo n=439, 12-17 y 12 weeks	Past week/ Level of functioning	Independent evaluator, blinded/ No info	50.0 (COMB) 49.5 (FLX) 50.0 (CBT) 49.1 (PLC)	66.6 62.1 60.0 59.3	16.6 12.6 10.0 10.2
Depressive disorder	(Wagner et al. 2003)	RCT	Sertraline Placebo n=376, 6-17 y 10 weeks	No info/ No info	53 units. Patient-rated/ No info	50.2 49.7 (PLC)	66.0 64.7	16.0 14.7 ns
Depressive disorder	(Wagner et al. 2006)	RCT, double-blind, placebo-controlled	Escitalopram Placebo n=261, 6-17 y 8 weeks	No info/ No info	No info/ No info	52.9 51.9	68.5 64.6	15.6 12.7

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Depressive disorder Outcomes	(Wiggins et al. 2010)	Observational study	Treatment as usual n=76, 12-18 y 3 months	One month/ Lowest level	Psychologists, social workers, nurses, occ therapists, psychiatrists/No info	52.3 (Depr) 57.1 (Other)	62.8 65.9	10.5 (Depr) 8.8 (Other)
Mood disorder	(Cummings and Fristad 2011)	Randomised clinical trial	MF-PEP Waiting list (Ctrl) n=165, 8-11 y 18 months follow-up	No info/ No info	2 psycholo-gists reviewed reports, consensus/ No info	43.0 44.4	More baseline anxiety symptoms assoc w greater CGAS improvement	
OCD	(Rosenberg et al. 1999)	Uncontrolled intervention study	Paroxetine n=20, 8-17 y 12 weeks	No info/ No info	Treating MD, review by CAP MD?/ No info	46.8	57.5	10.7
OCD	(Valderhaug and Ivarsson 2005)	Observational study, Norway and Sweden	n=68, 8-17 y	No info/ No info	CAP MD, psychologist or nurse/ No info	49.6 Swe 55.6 Nor		
Outcomes	(Garralda et al. 2000)	Naturalistic follow-up study	Treatment as usual (TAU) 6 months n=248, 3-18 y (n=191, 4-16 y CGAS)	One month/ Level of functioning	31 clinicians, 7 disciplines/ 30-60 min training of HoNOSCA	53.9	60.9	7.0

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Outcomes	(Gold et al. 2009)	Naturalistic follow-up study	Inpatient n=398 Acute residential n=350 Partial hospital n=203 Outpatient n=202 Mean age 13.4 90 days/discharge	One month/ Lowest level	Clinicians 20 sites/ No info	32.8 (Inpat) 44.1 (AR) 45.7 (PH) 55.7 (Outpat)	47.8 49.8 51.6 58.8	15.0 5.7 5.9 3.1
Outcomes	(Setoya et al. 2011)	Prospective study	Inpatient treatment as usual n=126, <15 y 11 months (mean)	One month/ Lowest level	Attending psychiatrist/ No info	38.1	57.9	19.8
Psychosis	(Castro-Fornieles et al. 2011)	Non-interventional follow-up	Pat with psychotic symptoms n=83, 9-17 y 2 years follow-up, stability of diagnosis	No info/ No info	Research psychiatrists psychologist/ Training	33.2 (mean) 22.9 32.7 50.0 31.7	65.8 (mean) Schizophrenia at endpoint No schizophrenia at endp. No diagnosis at endpoint Any disorder at endpoint	
Schizo-phrenia	(David et al. 2011)	Descriptive study	Document non-auditory hallucinations n=117, 6-17 y	No info/ No info	No info/ No info	38.7 admission / No visual hallucination 33.1 drug-free / No visual hallucination 31.8 admission / Visual hallucinations 22.7 drug-free / Visual hallucinations		
Schizo-phrenia	(Findling et al. 2003)	Prospective open-label trial	Olanzapine n=16, 12-17 y 8 weeks	No info/ No info	No info/ No info	41.9	52.7	10.8

Diagnosis/ Theme	Author	Study design	Intervention Number, Age Duration	C G A S				
				Time period/ Level	Rater background/ Training	Baseline	Endpoint	ΔCGAS
Schizo-phrenia	(Sporn et al. 2007)	Long-term follow-up study	Clozapine n=54, 7-19 y n=33 5 y follow-up	No info/ No info	No info/ No info	24.1	40.2 after 6 w 41.5 after 2-5 y	16.1
Self-harm, suicidal and non-suicidal	(Ougrin et al. 2011)	RCT	Therapeutic assessment AAU n=70, 12-18 y 3 month follow-up	No info/ No info	3 higher spec trainees in psychiatry/ No info	53.3 suicidal 54.8 non-suicidal	Nonsuicidal higher CGAS after 3 months than suicidal	
Tourette	(Gorman et al. 2010)	Descriptive study	n=65, 18 y Tourette n=65, 18 y Controls	No info/ No info	2 CAP MD, consensus procedure/ No info	56.4 (T) 70.4 (C)		

2.3.1 Swedish translation

P. Gustafsson & M. Helgesson translated CGAS into Swedish in 2001. In 2005, a translator and I revised the translation, which was then back-translated from Swedish into English. This work was done in collaboration with the research group at Columbia University that developed CGAS. The age span was increased from 4-16 years to 4-20 years. The revisions and change of the age span were discussed and approved of by the research group at Columbia University. CGAS has since been used in numerous Swedish research and clinical settings (Figure 2) (National Board of Health and Welfare 2009).

2.3.2 Reliability

Prior to the launch of CGAS, Shaffer and his colleagues tested the scale's inter-rater reliability and stability. In order to minimize variation due to clinical background, a group of five medical residents participated as raters (Shaffer et al. 1983). The intraclass correlation coefficient (ICC) was 0.84, which lies in an interval considered to correspond to "substantial" inter-rater reliability according to Shrout (1998). All but one rater were consistent over time when rating the vignettes after 6 months, yielding an ICC of 0.85.

A few years later Steinhausen evaluated the properties of the scale both in a research setting with vignettes (14 raters) and in a clinical setting (number of raters not accounted for) (1987). The inter-rater reliability, expressed as ICC, was 0.93. Interestingly, the stability coefficient (the intra-rater reliability, measuring agreement at two different time points for the same rater) varied between 0.22 and 0.85 and showed less agreement in cases with less severe emotional disorders than in cases with severe or more clear-cut single symptom disorders. The diagnostic agreement was also lower between the clinicians when assessing patients with less severe emotional disorders.

The ICC in other clinical settings ranges from 0.53 to 0.90, with the number of raters spanning from 2 to 20 (Rey et al. 1995; Dyrborg et al. 2000). Importantly, however, the results from these reliability studies comprising few raters with homogeneous background cannot be generalized to nationwide clinical settings involving large groups of heterogeneous raters. A recent study investigated the inter-rater reliability of CGAS in a somewhat larger clinical setting (n=78) (Hanssen-Bauer et al. 2007). All participants received general information about CGAS followed by a discussion on how to rate five short written vignettes before rating the written case vignettes. The inter-rater reliability was fair to moderate as indicated by an ICC of 0.61 (Shrout 1998), considerably lower than the ICC of 0.84 in Shaffer's original study (Shaffer et al. 1983).

Figure 2. Children's Global Assessment Scale, Swedish version 2005.

CHILDREN´S GLOBAL ASSESSMENT SCALE – C-GAS

För barn och ungdomar i åldrarna 4-20 år.

Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, & Aluwahlia S. *Psychopharmacology Bulletin 1985, 1:747-8.*
Anpassning av "the Global Assessment Scale for Adults" Spitzer RL, Gibbon M & Endicott i *Archives of General Psychiatry 1983, 40:1228-1231.*

Svensk översättning 2001-04-06 M Helgesson, fil lic leg psykolog och P Gustafsson, MD PhD Barn-och ungdomspsykiatriska kliniken, Linköping

Svensk nyöversättning och bearbetning 2005-08-14, i samarbete med Shaffer D, Gould MS, Fisher P, Bird H Columbia University, New York.
Anna Lundh, MD, Barn- och Ungdomspsykiatriska kliniken, Stockholm. E-post Anna.Lundh@sll.se

Skatta personens mest nedsatta generella funktionsnivå under den specificerade tidsperioden genom att välja den lägsta nivå som beskriver hans/hennes fungerande på ett hypotetiskt kontinuum av hälsa/sjukdom. Använd även de intermediära nivåerna (t.ex. 35, 58, 62).

Skatta aktuell funktionsförmåga utan hänsyn till behandling eller prognos. De tillhandahållna exemplen på beteenden är enbart illustrativa och erfordras inte för en speciell skattning.

Specificerad tidsperiod: 1 månad

100-91 **Synnerligen god funktionsförmåga inom alla områden** (hemma, i skolan och med kamrater), involverad i flera olika aktiviteter och har många intressen (t.ex. har hobbies eller deltar i aktiviteter utanför skolan eller tillhör en organiserad grupp såsom scouter, etc.). Sympatisk, gott självförtroende, vardagliga bekymmer blir aldrig ohanterliga. Klarar sig bra i skolan. Inga symtom.

90-81 **God funktionsförmåga inom alla områden**. Trygg i familjen, skolan och med kamrater. Det kan förekomma tillfälliga svårigheter och vardagsbekymmer som ibland blir ohanterliga (t.ex. oro i anslutning till ett viktigt prov, sporadiska vredesutbrott mot syskon, föräldrar eller kamrater).

80-71 **Endast lindriga funktionssvårigheter** hemma, i skolan eller bland kamrater. Viss beteendestörning eller vissa känslomässiga problem kan förekomma som reaktion på stressframkallande livshändelser (t.ex. föräldrars separation, dödsfall eller ett syskons födelse), men dessa är kortvariga och funktionssvårigheterna övergående. Dessa barn är ytterst lite störande för andra och anses inte avvikande av personer som känner dem.

70-61 **En del svårigheter inom ett enstaka område, men fungerar allmänt sett ganska väl** (t.ex. sporadiska eller isolerade antisociala handlingar som tillfälligt skolk eller snatteri; genomgående smärre svårigheter med skolarbete, kortvariga växlingar i stämningsläge; rädslor och ängslan som inte leder till undvikande beteende; tvivel på sig själv). Har meningsfulla relationer. De flesta personer som inte känner barnet väl skulle betrakta honom/henne som normal, men de som känner honom/henne väl skulle kunna uttrycka oro.

60-51 **Varierande funktionsförmåga med sporadiska svårigheter eller symtom inom flera, men inte alla, sociala områden**. Störningen skulle vara uppenbar för dem som träffar barnet i ett dysfunktionellt sammanhang eller vid en dysfunktionell tidpunkt, men inte för dem som ser barnet i andra sammanhang.

50-41 **Måttlig störning av funktionsförmågan inom de flesta sociala områden eller allvarlig störning av funktionsförmågan inom ett område**, vilket kan orsakas av t.ex. suicidal upptagenhet eller suicidala grubblerier, skolvägran och andra former av ångest, tvångsmässiga ritualer, allvarliga konversionssymtom, täta ångestattacker, ofta förekommande aggressivt eller annat antisocialt beteende med visst bibehållande av meningsfulla sociala relationer.

40-31 **Betydande nedsättning av funktionsförmågan inom flera områden och oförmögen att fungera inom ett av dessa områden**, dvs. störd hemma, i skolan, med kamrater eller i samhället i stort (t.ex. ihållande aggression utan uppenbar anledning; påtagligt tillbakadraget och isolerat beteende beroende på antingen stämnings- eller tankestörning, suicidförsök med tydlig dödlig avsikt). Dessa barn behöver sannolikt särskild skolgång och/eller intensifierad öppenvård/inläggning på sjukhus.

30-21 **Oförmögen att fungera inom nästan alla områden**, t.ex. stannar hemma eller i säng hela dagen utan att delta i sociala aktiviteter **eller** allvarlig störning av realitetsprövning **eller** allvarlig kommunikationsstörning (t.ex. ibland osammanhängande eller inadekvat).

20-11 **Kräver ansenlig tillsyn och övervakning** för att förhindras att skada andra eller sig själv, t.ex. ofta våldsam, upprepade suicidförsök **eller** för att sköta personlig hygien **eller** grav störning av all kommunikation, t.ex. allvarlig avvikelse i verbal kommunikation och kroppsspråk, markant socialt otillgänglig, stupor, etc.

10-1 **Kräver ständig tillsyn och övervakning** (24-tim vård) på grund av allvarligt aggressivt eller självdestruktivt beteende eller grav störning av realitetsprövning, kommunikation, kognition, affekt eller personlig hygien.

2.3.3 Validity

To validate CGAS, one needs a reference CGAS score or similar measures against which ratings can be compared. The most commonly used approach to define a "true" CGAS rating is to analyse the *post hoc* mean rating of a group of raters. An alternative approach is to use experienced clinicians to reach a best estimate that is then defined as a "gold standard" (Wu et al. 2007). The latter method has been employed in a study involving a total of 30 raters from 5 countries (but only 15 of them did the CGAS ratings, 3 from each country), where each national group made a consensus rating of case vignettes (Hanssen-Bauer et al. 2007). The expert rating made by consensus was 2 points lower than the mean rating of the 15 participants. This difference was not considered clinically relevant.

Each method of establishing a "true" CGAS value for a patient has advantages and disadvantages. Using the mean group rating as a reference is straightforward and easy. However, this method does not control for the raters' professional training, clinical experience, familiarity with rating tools and diagnostic instruments, or attitudes towards such instruments, all of which are likely to affect their ratings (Söderberg et al. 2005). Using expert raters is more complicated and resource demanding. The advantage of using experienced clinicians, however, is that these expert raters have sufficient training in child psychiatry, broad experience with rating instruments, and generally a positive attitude towards them, all of which are likely to make their ratings more valid. In my opinion, this speaks in favour of using expert raters. This strategy is also in accordance with the way in which a gold standard is established in diagnostic assessments, where experienced clinicians rely on semi-structured instruments as well as the opinions of other experienced clinicians.

As to external validity, research has found that CGAS ratings correlate fairly well with a child's intellectual functioning, school competence as rated by parents, and social skills as rated by clinical staff in clinical settings (Weissman et al. 1990; Green et al. 1994). Furthermore, CGAS ratings have been found to agree with ratings from the total Behaviour Problem Score and the Social Adaptation Score of the Child Behaviour Check List (CBCL) in epidemiological settings (Bird et al. 1987). Moreover, discriminant validity has been established in a study that found significantly lower CGAS scores in a group of children referred to mental health services than in a group of children that was not referred, and in which a group of clinical "cases" scored lower on CGAS than did "non-cases" (Bird et al. 1987).

2.3.4 CGAS, a forerunner to the Global Assessment of Functioning (GAF)

The Global Assessment of Functioning (GAF) is a rating scale similar to CGAS. GAF is a widespread tool in adult psychiatry and in some child and adolescent psychiatric settings. A common misconception is that CGAS is a children's version of GAF, but GAF is in fact a hybrid of GAS and CGAS that was developed in conjunction with the DSM-III-R (American Psychiatric Association 1987).

GAF was designed for use with both children and adults. There are particular advantages to using a scale appropriate for all ages, especially in longitudinal studies. The disadvantage, however, is that the descriptions of different levels of functioning must be general enough to fit the whole life span, which may degrade face validity and

hamper specificity. For child and adolescent use, it is therefore an advantage that the CGAS descriptions are adapted specifically for children and adolescents.

In the original introduction of CGAS (Shaffer et al. 1983) it is clear that the scale should be used to assess the most impaired level of psychosocial functioning during the last month. (There are, however, many examples of studies in which these instructions are not followed, as elaborated below in the *General Discussion*.) The GAF assessment, by contrast, reflects the lowest level of either symptom severity, level of functioning or a combination during a time period defined by the clinician (American Psychiatric Association. 2000). These loose instructions for GAF users inevitably lead to a diverse array of rating strategies (highest, lowest and average) and hamper comparisons of GAF ratings between study populations (Bates et al. 2002). A final difference between CGAS and GAF is that the range that is defined as serious impairment of psychosocial functioning differs between the two scales (CGAS 1 to 40, GAF 1 to 50).

2.3.5 Global functioning and DSM

Ever since the DSM-III was published in 1980, the definition of mental disorders has included both symptomatic criteria and functional impairment (American Psychiatric Association 1980; 1987; 1994; 2000). However, in clinical practice diagnostic criteria are not primarily used to decide if a child needs treatment; defining clinical caseness is instead an issue of clinical judgement. This is a problem in research where the definition of caseness needs to be operationalized. Several studies have shown that the number of cases - expressed as the overall prevalence of a psychiatric disorder - is markedly reduced when a functional impairment rating with CGAS is added to the diagnostic procedure. This is true whether using DSM-III-R (American Psychiatric Association 1987; Bird et al. 1993) or DSM-IV-TR (American Psychiatric Association 2000; Canino et al. 2004). This suggests that combining diagnostic criteria with a measure of psychosocial functioning may be the best way to classify a disorder.

2.4 CLINICAL EDUCATION IN THE USE OF RATINGS SCALES AND GUIDELINES

It cannot be assumed that the studies summarised above, that investigate CGAS reliability and validity in small groups of clinicians, can be extrapolated to naturalistic circumstances that involve large heterogeneous groups of raters. Before using CGAS in large clinical settings, we therefore need to find out whether practitioners use CGAS consistently enough to make it a useful tool in evidence-based child psychiatry. This information on the quality of the ratings is important to professionals, researchers, and managers relying on CGAS for their work.

Paper I in this thesis examined the inter-rater reliability when a large group of untrained practitioners used CGAS.

2.4.1 GAF and CGAS rater training

When CGAS was introduced in 1983, there were no instructions as to whether training was required prior to employing the tool in research or clinical settings (Shaffer et al.

1983). At Columbia University, however, it has been common practice to introduce CGAS by gathering researchers/clinicians to discuss some cases together to ensure that everyone rates cases within the desired range of ±5 points (personal communication Prudence Fisher). Several researchers, however, have suggested that the lack of standardized and evaluated training material may jeopardize the quality of ratings (Schorre and Vandvik 2004; Winters et al. 2005; Rush et al. 2008), and numerous researchers whose studies are based on the similar GAF scale have suggested that users should undergo training before using GAF (Fernando et al. 1986; Goldman et al. 1992; Hilsenroth et al. 2000). There are, however, no studies to date on the actual effects of training on the reliability and validity of CGAS ratings. In one of the few studies that looked at training's effect on the use of GAF, brief one-hour training sessions improved participants' understanding of how to use the scale, but did not bring ratings closer to the expert rating enough to be clinically relevant (n=31) (Bates et al. 2002).

Even though there are no studies evaluating the effect of CGAS training, there are studies that have performed training prior to the use of the scale. In one small-scale study, five raters read the available literature about CGAS and the Global Assessment of Psychosocial Disability scale (GAPD) for two months before the study was initiated (Dyrborg et al. 2000). Three of the raters in the study, all experienced clinicians, also met weekly and agreed upon some ratings each time. The results showed that the agreement between raters was higher among those who were clinically experienced and had performed more than 50 CGAS ratings. This observation indicates that practice can improve the quality of ratings. In another recent study, there was no positive effect on inter-rater reliability from feedback given to a group of raters about 1) how they rated vignettes themselves and 2) how three experienced clinicians had rated vignettes 6 months before they received new vignettes (Hanssen-Bauer et al. 2007).

2.4.2 Designing and evaluating successful training

The lack of evaluated training programmes for rating scales like CGAS raises questions about to how a training model can be constructed so that it can be used by large groups of practitioners. Training methods that only include dissemination of information to passive participants have not been shown to have an impact on professional skills (Davis et al. 1995; Bero et al. 1998). This finding has lead to an increased focus on active learning in medical universities (Biggs 2003).

There has also been little discussion about how "improvement" in CGAS rating skills should be defined and how the effect of training should be evaluated. A possible explanation is that the original work presupposed that clinicians could use the scale without training. Kirkpatrick & Kirkpatrick provides a Four Level Model to evaluate training programmes generally in corporate, government, and academic worlds (2005). The first level, *reaction*, is evaluated in most training settings. This can be done with "happy sheets" or surveys collecting information about the participants' experiences from the training. Even though the reaction level says nothing about what the participants learned, a positive reaction promotes successful learning (Kirkpatrick and Kirkpatrick 2005). *Learning* is the second level evaluated and consists of assessments of knowledge increase, skills improvement, and changed attitudes. The second level can be fulfilled without discernible *behavioural changes*, which are evaluated at level three. The last and fourth level, *results*, evaluates changes that take place in a whole

organization as a result of training, for example, if clinicians were to more regularly assess the level of psychosocial functioning with CGAS and use the information in treatment planning.

2.4.3 Computer-based training

Gathering all clinicians for scheduled seminars consisting of both skills practice training and basic theoretical background education would be time-consuming and costly in large organisations. New employees arrive, some people work part-time, and some are out of work due to sickness or parental leave. A computer-based method with interactive training available through a Compact Disc (CD-ROM) or the Internet would be a more flexible way to offer CGAS training, as it is available whenever clinicians have time. It could also be a cost-efficient way to train raters in large clinical settings. The use of such methods has, however, not previously been evaluated.

Paper II in this thesis evaluates two types of training methods, a seminar and a computer-based training module, with respect to their effect on rating improvements as defined by agreement with expert raters. Both training methods combined theoretical background information about CGAS with practical training in its use, and were thereby more extensive than any previously described training or introduction to CGAS or GAF (Bates et al. 2002; Hanssen-Bauer et al. 2007).

2.5 EFFICACY VERSUS EFFECTIVENESS

Clinical guidelines within CAMHS are based mainly on Randomised Controlled Trials (RCTs), since these efficacy trials are the gold standard for evidence-based medicine (Weisz et al. 1995; Fonagy 1997; Gilbody et al. 2002). Such trials differ, however, from routine care in many respects, which limits the generalisability of their findings. First, clinical efficacy trials typically enrol patients with a clear-cut diagnosis without co-morbidity in order to produce a homogenous patient group. Exclusion criteria typically prevent enrolment of children/adolescents (and/or parents) with alcohol or drug abuse, reported physical abuse, mental retardation, pervasive developmental disorders, major neurological or medical illness, eating disorders, psychosis, bipolar disorder, significant suicidal ideation, long periods of absence from school, and other co-morbid axis I or II conditions. Applying these exclusion criteria effectively excludes large numbers of patients who actually receive treatment within CAMHS (Fonagy 1997; Bridge et al. 2009). Second, research has shown that not only patients' but also clinicians' behaviour differ between a research setting and a naturalistic setting (Kendall and Southam-Gerow 1995). Third, high expectations are more common in research patients and have shown to correlate with better outcomes (Lambert et al. 2004, p. 205). Lastly, it is likely that treatment compliance is better in clinical trials than in naturalistic settings.

Because of these limitations with RCTs' applicability to real world psychiatric settings, clinical efficacy trials need to be complemented with effectiveness studies that take place in real world settings with unselected heterogeneous groups of both patients and clinicians (Weisz et al. 1995; Weisz et al. 1995; Gilbody et al. 2002; March et al. 2004; Mufson et al. 2004). Unfortunately, such effectiveness studies are scarce compared to

RCTs evaluating efficacy (Hoagwood et al. 1995; Weisz et al. 1995; Weisz et al. 1995; Jensen et al. 1996; Garland et al. 2010).

An important source of information when conducting effectiveness research is routinely collected clinical data. This is also the basis of *outcomes research*, the field of research that studies the ultimate effect of medical care on the health and well-being of patients and populations. The results from outcomes research are used to compare existing treatments, improve the quality of care, and evaluate new treatments without the expense of clinical trials and the loss of generalisability (Ellwood 1988; Busch and Sederer 2000; Gilbody et al. 2002). Outcomes research can also investigate cost-effectiveness under routine conditions.

The management of CAMHS in Stockholm has long endorsed the importance of outcomes research, and the clinical database Pastill – described in detail under *Methods* – is an effort to facilitate such work. Soon after the training activity evaluated in Paper II, CGAS ratings were being registered in Pastill from 1 July 2006. Adding CGAS ratings to the database Pastill made it possible to design an effectiveness study, in which change in CGAS ratings served as the outcome measure of child psychiatric treatment. This work is described in Paper III.

2.6 THE IMPORTANCE OF PROGNOSTIC TOOLS

Longitudinal studies have shown an increased rate of negative adulthood outcomes in child psychiatric populations compared to healthy children (Engqvist and Rydelius 2006; Engqvist and Rydelius 2007; Mordre et al. 2011). However, there is a considerable spread of risk for long-term adverse outcomes also *within* the child psychiatric population, and longitudinal studies following children with mental disorders into adulthood have revealed various patterns of developmental trajectories with respect to diagnoses, behaviours, and functional outcomes (Colman and Jones 2004; Thompson et al. 2010). If patients who are at higher risk for psychiatric disorders, substance misuse, or criminality could be identified early, intervention programs could be targeted to those who need it the most. This would facilitate a more prudent use of healthcare resources and ultimately increase psychiatric services' ability to prevent suffering (Leaf et al. 1996; Steinhausen and Metzke 2001; Canino et al. 2004; Sourander et al. 2004). To this end, clinicians need more knowledge about how mental disorders develop as well as valid tools for risk assessment.

Most longitudinal studies of adverse outcomes in child psychiatric patients are based on the association with categorical childhood psychiatric diagnoses rather than psychosocial functioning (Rutter et al. 1976; Kim-Cohen et al. 2003; Engqvist and Rydelius 2007; Sourander et al. 2007; Pickles et al. 2010; Mordre et al. 2011). For example, the risk of recurrent depression in adulthood has shown to be higher among those who suffer from depression during adolescence (Lewinsohn et al. 2000). According to another study, depressed adolescents also report more impaired familial and social relationships, and negative effects on work and school performance than normal controls (Geller et al. 2001). Furthermore, childhood conduct problems with or without attention-deficit hyperactivity disorder (ADHD) have repeatedly been associated with future criminality (Fergusson et al. 2005; Olsson et al. 2006; Engqvist and Rydelius 2007; Satterfield et al. 2007; Sourander et al. 2007; Forsman et al. 2010;

Mordre et al. 2011). Conduct disorder and oppositional defiant disorder have also been associated with psychiatric disorders later in life (Rutter et al. 1976; Kim-Cohen et al. 2003; Fergusson et al. 2005), especially in girls (Olsson et al. 2006).

However, using categorical diagnoses as prognostic tools has several limitations. For one, psychiatric diagnoses in younger children have been shown to have a low predictive value (Bennett and Offord 2001; Kim-Cohen et al. 2003; Kendler et al. 2008), partly due to the high rate of comorbidity (Bird et al. 1993; Angold et al. 1999). Also, a categorical diagnosis does not inform us about the severity or consequences of a disorder (Bird et al. 1990; Canino et al. 2004), which is information that presumably adds prognostic information. One way to improve prognostic precision might therefore be to combine diagnostic information with ratings of psychosocial functioning, for example CGAS ratings. However, it is as yet unknown if CGAS predicts future negative outcomes in young adults. In the only study to date on the subject, retrospective CGAS ratings at intake based on hospital records did not differ between later convicted and non-convicted adults in a 30 years follow-up study of 541 child psychiatric inpatients (Mordre et al. 2011). If CGAS ratings could be shown to contain prognostic information, they could help clinicians identify early those individuals who are at risk for psychiatric disorders, substance misuse, or criminality.

The large number of registered patients with CGAS ratings in the clinical database Pastill enabled a prospective follow-up study to examine the predictive properties of CGAS. In Paper IV, CGAS scores were linked to national registers and the risk of negative outcomes was analysed.

3 AIMS

Despite the widespread use of CGAS to assess global functioning, little is known about the accuracy of the instrument in routine clinical care, and whether training programmes to improve precision are worthwhile. Moreover, there is a dearth of studies assessing the effectiveness of child and adolescent psychiatry treatment by means of psychosocial functioning. Finally, the use of CGAS for long-term treatment planning is limited because the prognostic value of the scale is unknown.

Against this background, the specific aims of this thesis were to:

1. investigate the inter-rater reliability and the accuracy in terms of agreement with expert ratings when CGAS is used in a large-scale naturalistic clinical setting.
2. compare the effectiveness of two different training methods, live seminars and computer-based training, and to estimate the overall effect of training to improve CGAS ratings.
3. investigate the effectiveness of child psychiatric outpatient treatment as measured by change in CGAS ratings, and to identify predictive factors for CGAS change.
4. investigate whether CGAS ratings can predict future mental health disorders, suicide attempts, criminal conviction, substance misuse, and accidents.

4 METHODS

Herein, the methods used in each study are briefly presented. The full account of the methods used are given in the respective paper.

4.1 PAPERS I AND II

The first two studies in this thesis evaluated inter-rater reliability and the accuracy of CGAS ratings among untrained raters in a large clinical setting. Different methods to improve CGAS ratings were also evaluated. I capitalized on the fact that the implementation of CGAS in Stockholm CAMHS was accompanied by a training activity for which I was responsible. Baseline data were collected between October 2005 and June 2006, and end-of-study data between October 2006 and June 2007.

4.1.1 Introduction of CGAS in Stockholm

In 2000, Stockholm County Council decided to implement the use of GAF for assessment of all psychiatric patients, including children and adolescents. GAF patient ratings were to be recorded and followed over time as part of a quality improvement programme throughout the county (which comprises the greater Stockholm region).

In 2006, all CAMHS in Stockholm switched to using CGAS instead of GAF. The ratings were registered in the clinical database Pastill (described below, Paper III and IV). It has been difficult to find any documentation of discussions between health administrators and management at CAMHS regarding the decision to change scales, and to understand how the ratings were intended to be used.

In 2008, the Stockholm County Council launched a remunerative incentive scheme to increase the use of CGAS (Personal communication, Yvonne Björklund, Department of Finance, CAMHS 2011). Health administrators decided that 3% of the total budget (590 million SEK) for CAMHS in Stockholm County Council would be available only if

1. >85% of patients with three visits or more had CGAS ratings at intake and at end-of-treatment
2. >90% of patients had an ICD-10 or DSM-IV-TR diagnosis
3. >90% of patients had an individual treatment plan

There were no quality requirements in terms of reliability measures (for example inter-rater reliability and accuracy in ratings compared to expert ratings) or formal training coupled to the reimbursement programme. There was no explanation of how the collected ratings would be used. Previous research has found that the quality of GAF ratings is related to the raters' attitude towards the scale (Söderberg et al. 2005), suggesting that introducing mandatory use of CGAS in this way may not have been optimal. I was asked to be responsible for training clinicians throughout the county in the use of CGAS.

4.1.2 Raters and units

A total of 703 health care professionals participated in Papers I (Lundh et al. 2010) and II (Lundh et al. 2011), of which a majority (n=624) were from CAMHS in Stockholm County. CAMHS health care professionals in the counties of Östergötland (n=24), Norrbotten (n=21) and Småland (n=34) also participated.

A total of 33 CAMHS (29 outpatient units, 4 inpatient units) were represented, of which 25 outpatient and all 4 inpatient units were located in Stockholm County.

A total of 648 raters comprised the intervention groups, and 55 raters were included to form a non-randomised comparison group.

There were approximately 770 CAMHS employees in Stockholm during the years the study was conducted (Personal communication, Maria Norrbin, Human Resources CAMHS 2011). Hence, more than 80% (624/770) of the health care professionals in Stockholm participated in the study.

Demography and background characteristics of the raters were recorded anonymously on coded forms. The raters were psychologists, social workers, medical doctors and other staff members (nurses, psychiatric technicians, special education teachers, occupational therapists etcetera). The vast majority had no experience using CGAS, but a majority had used GAF.

4.1.3 Development of written case vignettes and expert ratings

The vignettes used in the study were based on actual patients' first visits at CAMHS in order to provide a more realistic and richer presentation than the cases on a website providing training vignettes (WIMHRT 2009). Changes were made in the histories in order to make it impossible to trace the case vignette to a specific patient, clinician or unit.

Four experienced clinicians were asked to participate as expert raters together with me. This number was chosen to create a group that could reach a consensus on 50 cases during 1 day. The group consisted of three child and adolescent psychiatrists and two psychologists, with a mean experience in child and adolescent psychiatry of 12 (range 6-20) years. They all had expertise and skills in the use of structured psychiatric interviews and rating scales. They also had a positive attitude towards using different tools in their clinical work. These five experts first rated the cases individually. Thereafter, a joint discussion resulted in a consensus rating for each case, henceforth referred to as the "expert rating" or "gold standard".

A sample of ten vignettes was selected by me to provide cases that varied in age, sex, level of functioning, and problem area. Only five vignettes were used for baseline and end-of-study ratings to ensure that all raters would have time to finish the ratings within the one-hour that was at our disposal. The other five vignettes were used in the training activity.

4.1.4 Rating procedures

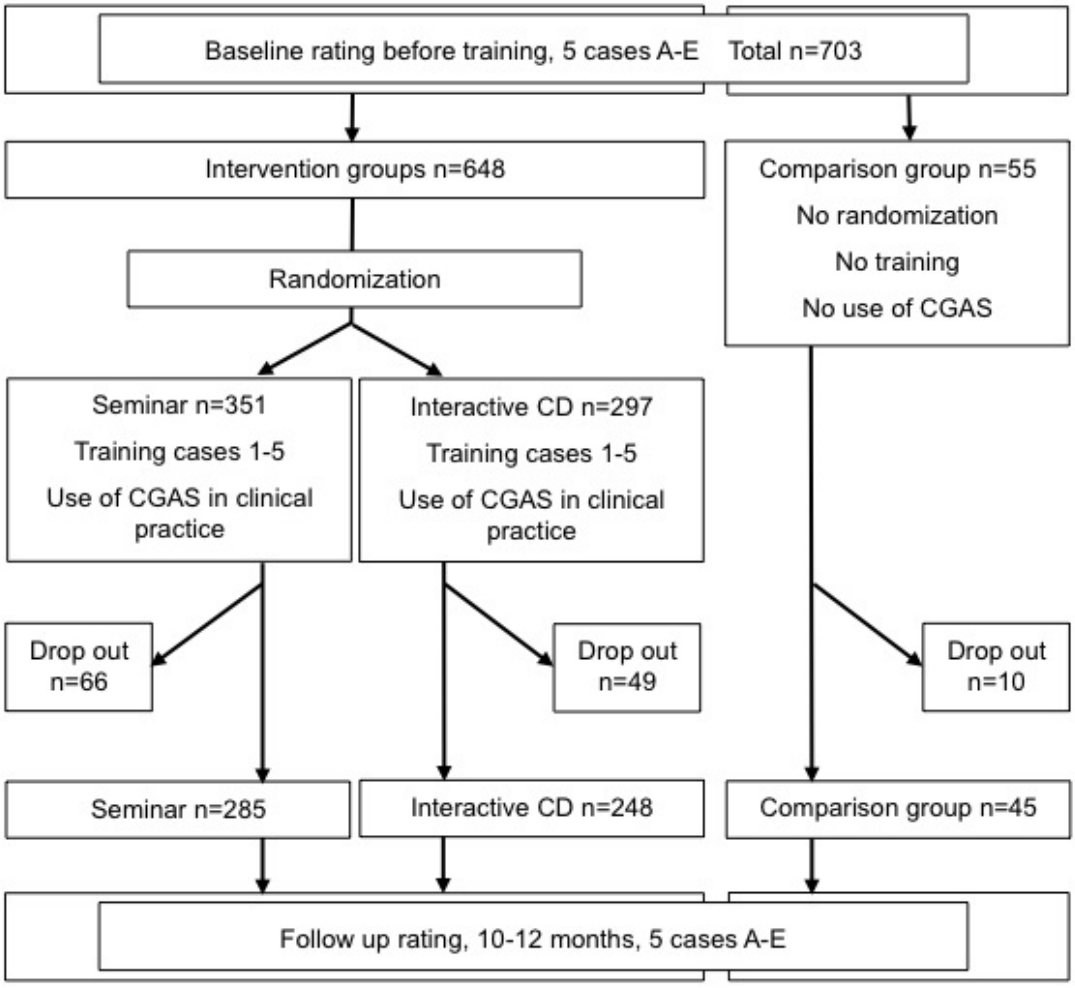
The flow chart below (Figure 3) displays the time schedule and the number of participants for each step in the rating procedure.

All participants rated the same five written vignettes individually at baseline. Raters were randomised to receive training with either a seminar or an interactive CD-ROM. Every CAMHS unit head was instructed to randomise the participants in advance.

After completion of the baseline ratings, the raters in the computer-based intervention group were given a CD-ROM with instructions to complete the training during the following week. Those who were randomised to the seminar group continued with their training after a break. Within a week after the training, participants completed a questionnaire via the Internet assessing their satisfaction with the training.

The follow-up rating was scheduled to occur after 6 months, but due to the great interest in the training programme, the follow-up had to be postponed to make room for additional training sessions in the county. During the 10-12 month follow-up period, all raters in the intervention groups were requested by the CAMHS administration to do clinical CGAS ratings in their daily practice. At the follow-up, the five baseline vignettes were rated again.

Figure 3. Flow chart Papers I and II.



4.1.5 Training programmes

The training programmes were developed based on knowledge about clinical training methods (Davis et al. 1995; Bero et al. 1998), combining a theoretical and a practical part. The seminar group also engaged in a discussion. It should be pointed out that the training method employed in this thesis is a more thorough procedure than prior, non-evaluated, training activities of CGAS (Dyrborg et al. 2000; Hanssen-Bauer et al. 2007)

Each seminar lasted about two hours and I conducted all of them (31 seminars in total). The seminars were highly structured in order to be as similar as possible. The theoretical part consisted of information on the history of CGAS, its psychometric properties, and rating techniques. The raters were instructed to identify the CGAS score that best matched the lowest level of global functioning during the last month. Skills training followed the theoretical part of the seminar in groups of two or three. Five training vignettes were rated, one at a time. The small groups reported their scores on a whiteboard that was viewed by the whole group and thereafter I led a discussion. At the end the "expert rating" was revealed, followed by a short explanation of how they motivated that rating.

The computer-based training covered the same topics as the seminar. After reading the vignette on the screen, the rater identified a score on the CGAS scale that was then entered into the computer. The rating was compared with the expert rating. If the rating was more than ±5 points but less than ±25 points off from the expert rating, the rater received a comment that the rating was too low or too high and was asked to try again. If the rating was more than ±25 points off from the expert rating, the rater received a comment included that his or her rating was not reasonable. The rater was encouraged to reread the case and perform a new rating.

4.1.6 Definition of aberrant rating

Ratings five points higher or lower than the expert values were defined as *aberrant* ratings for the purposes of this study. This range of ±5 points was decided based on the fact that the common range of CGAS improvement in clinical trials is from five to fifteen points (see Table 1), and based on informal guidelines from Columbia which consider ten points around a mean rating to be desirable and acceptable. In Paper I, when a participant rated all five cases outside the range of ±5 points, this was defined as an *overall aberrant rating.* The raters could hence be divided into two groups: *overall aberrant raters* and *not overall aberrant raters.* The latter category included those who were aberrant raters on some but not all cases, as well as the single participant with no aberrant ratings at all. The risk of aberrant ratings was analysed in Paper I for different groups i.e. age, sex and professional background.

4.1.7 Statistical methods

The ICC was used to determine the inter-rater reliability (Bland and Altman 1996). The analysis of overall aberrant raters was conducted using a logistic regression model to explore which factors could discriminate between overall and not overall aberrant raters. All tests were two-sided and $p<0.05$ was regarded as a statistically significant result. In Paper II, results are presented as complete cases analyses, i.e., only subjects who had eligible observations at both baseline and end-of-study were included in the analysis.

4.2 PAPERS III AND IV

CGAS ratings and other clinical data from the database Pastill (described below) were used in Papers III and IV. The registration of CGAS in Pastill offered a unique

possibility to use a large number of routinely collected CGAS ratings by a group of trained raters from real world child psychiatry.

4.2.1 Setting

In Papers III and IV, all subjects were patients within CAMHS in Stockholm County. This service consists of both inpatient and outpatient clinics and provides mental healthcare in a catchment area comprising 420,000 children and adolescents (age 0-17 years). There are 14 regular outpatient clinics and four clinics working with intensive, often home-based, outpatient treatment. Seven outpatient clinics provide specialized care either for particular diagnostic groups or employ specific treatment models.

4.2.2 The clinical database Pastill

During the 1990s, CAMHS sought to develop and expand the collection of data on patients that began in the 1950s, for use in quality assurance programs and in future research. No specific research questions, however, guided what variables should be collect, a well-known phenomenon in outcomes research (Gilbody et al. 2002). The clinical database Pastill was launched in 1998. After one year, all outpatient clinics were equipped with computers and data were registered online. In 2003, the inpatient clinic joined the project. Between 2001 and 2010, 165,000 cases were registered, representing 109,000 unique children and adolescents. Since the inception of Pastill, variables have been added and changed to meet needs and interests of clinicians, administrators, and managers. Today, epidemiological researchers meet regularly with representatives from CAMHS to discuss how to improve future data collection and how data can be used for research.

To enter data into Pastill, the clinician managing the case either fills in a paper form or registers the required information online. The information needed for the registration is collected both during the intake interview of the patient and family (and school if possible), and at the end of the treatment period. Table 2 presents some of the many variables recorded in Pastill.

Table 2. Variables in the clinical database Pastill.

Age at intake
Status as asylum seeker
Cause of referral
Global Assessment of Functioning (GAF) (until 30 June 2006)
Children's Global Assessment Scale (CGAS) (from 1 July 2006)
Clinical assessment report
Clinical issues/ problems addressed
Nationality
Diagnosis according to the DSM-IV-TR and/or ICD10
Family situation, special circumstances, i.e., death of a parent or a sibling, adoption, sibling with severe medical illness
Family situation, who the child lives with
Given treatment
Initiator of the contact with CAMHS / Referral source
Legal guardian

Provided care levels
Medication other than central stimulants
Medication with central stimulants
Neighbourhood
Interventions aiming at patient's social network, collaboration with other caregivers
Number of appointments/telephone calls
Overall assessment of treatment outcome, according to clinician
Overall assessment of treatment outcome, according to family/adolescent
Parent visitation
Prior contact with CAMHS and at what age
Psychosocial stressors
Sex
Siblings
Strengths and Difficulties Questionnaire (SDQ)
The Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA)
Time period from contact to first appointment
Time period from contact to first offered appointment
Treatment performed outside CAMHS
Established treatment plan
Type of daily activities i.e. day care, school, work
Variables possible to add, general or local for different units

Out of 29 diagnostic alternatives available in Pastill, 26 adhere to the headings in DSM-IV-TR and/or ICD 10. The other three alternatives include "lack enough information to perform a diagnostic assessment", "diagnostic criteria not fulfilled", or "other". It is possible to choose more than one diagnosis. Diagnostic categories are listed in Table 8.

The variable describing treatment given has evolved since the first version of Pastill due to changes in treatment options available within CAMHS. Treatment options available for Paper III included counselling and psychotherapy with different time frames and settings (individual child or adolescent, group, parents, family, network). There are, however, no data in Pastill about which specific psychotherapeutic method has been used. Medication information is restricted to whether the drug is a central stimulant or not.

Axis V, global functioning, is rated with CGAS before and after given treatment. Clinicians are instructed to rate the lowest level of functioning during the last month.

In addition to entering CGAS ratings at intake, the clinicians make an overall assessment of the treatment outcome and also report the family's/adolescent's overall opinion of how the care has worked. This is rated on a 1-5 Likert scale where 1=worsening of symptoms, 2=no change in symptoms, 3=symptoms are the same but easier to cope with, 4=improved symptoms, 5=no symptoms. The score -1 was chosen when the overall assessment was not possible to complete.

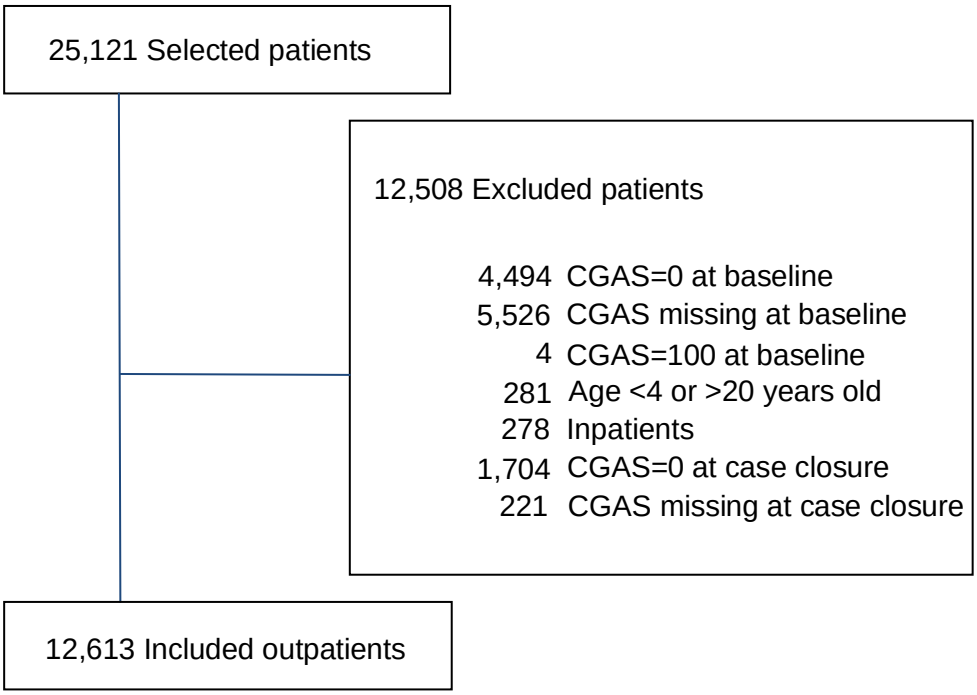
4.2.3 CGAS outcome measure, Paper III

ΔCGAS, defined as the difference between CGAS at intake and CGAS at case closure, was used to assess the change in psychosocial functioning during the course of treatment in Paper III. Mean ΔCGAS was analysed in relation to diagnosis. Four diagnostic patient groups were selected for further study of baseline factors (CGAS score at baseline, sex, age, psychosocial stressors, diagnosis) and intervention factors (treatment provided, professional background of the clinician, number of appointments) potentially predictive of ΔCGAS. The diagnostic groups were mood disorder, ADHD, obsessive-compulsive disorder (OCD), and conduct disorder.

4.2.4 Subjects in Paper III

All cases between 1 July 2006 and 31 January 2010 with a longer treatment period than one month (n=25,121) were selected (Figure 4). They constituted 57% of all the registered cases (n=44,261) during this time period. From these, the following cases were excluded: 1) Those with CGAS rating=0 or a missing CGAS rating at baseline (n=10,020); 2) CGAS ratings=100 at baseline (n=4); 3) Those younger than 4 years old (n=244) and those older than 20 years of age (n=37); 4) Those who underwent inpatient care (n=278); 5) Those with CGAS rating=0 or missing at end point (n=1,925).

Figure 4. Flow diagram, excluded and included patients.



Only outpatients were included since the majority of inpatients (n=196 of 278) had shorter treatment periods than one month and the small remaining number of subjects (n=82) was not representative of the inpatient group as a whole. Hence, the total number of cases included in this study was 12,613. The included cases were between 4 and 19 years old with a mean age of 12.0 years (SD=3.9). The group consisted of 6,012 boys (47.7%) and 6,601 girls (52.3%).

4.2.5 Statistical methods used in Paper III

All statistical analyses in Paper III were conducted using PASW Statistics 19 (SPSS Inc., Chicago, Illinois, USA) or Statistica 9.0 (Statsoft Inc, Tulsa, Oklahoma, USA). Differences between the mean group scores of CGAS before and after treatment were tested using paired t-test. All tests were two-sided and $p<0.05$ was regarded as a statistically significant result.

A stepwise multiple linear regression model was applied to estimate the contribution of each background and intervention variable on ΔCGAS. At each step, the independent variable not in the equation that had the smallest probability of F was entered, if that probability was sufficiently small ($p<0.05$). Variables already in the regression equation were removed if their probability of F became sufficiently large ($p>0.10$). The method terminated when no more variables were eligible for inclusion or removal. This method yielded a reduced set of variables from the larger set of predictors, eliminated unnecessary predictors, simplified data, and enhanced predictive accuracy.

4.2.6 National registers, Paper IV

All Swedish citizens and immigrants have a unique 10-digit identification number that makes it possible to link data across different national registers and birth or clinical

cohorts. In Paper IV, the study population from Pastill clinical database was linked to the following Swedish national registers:

i) The National Patient Register (NPR), maintained by the National Board of Health and Welfare contains main diagnosis and up to seven secondary diagnoses. The NPR contains all admissions to inpatient care since 1987 and all visits to specialized (medical doctor visits) outpatient care since 2001. The data in the register are collected once a year and cover both public and private health care since 2001. In addition to diagnoses, NPR also contains information on causes of morbidity or mortality, accidents, surgical procedure codes, and discharge dates. Diagnoses are coded according to the 9th (1987–1996), and 10th editions (1997-) of the International Classification of diseases (ICD). The validity of the NPR is considered high for most diagnoses (Ludvigsson et al. 2011) including psychiatric diagnoses (Ekholm et al. 2005; Sellgren et al. 2011). Discharges and outpatient care that occurred up to 31 December 2009 were included in the study.

ii) The Swedish Crime Register (maintained by the National Council of Crime Prevention), contains information on all criminal convictions in Sweden since 1973. The age of criminal responsibility is 15 years (i.e., individuals younger than that cannot be charged and convicted). The conviction data included all individuals who received custodial or non-custodial sentences and cases where the prosecutor decided to caution or fine. All crimes committed before 31 December 2009 were included in the study.

iii) The Cause of Death Register (CDR, Statistics Sweden) maintains data on all deaths in Sweden since 1952, including date of death and causes of death. Death events occurring before 31 December 2009 were included in the study and the information was used to set the time at risk among those who died during the follow-up period.

4.2.7 Study population and exposure assessment, Paper IV

The first treatment period (longer than one month) of all patients registered in Pastill from 1 July 2006, who were at least 18 years old on 31 December 2009 and therefore at risk for the specified outcomes, was included in Paper IV. In all, 6,525 patients fulfilled these criteria of which 5,903 had valid CGAS scores at intake and 4,876 had valid scores at end-of-treatment (not '0' or missing). Of these, the vast majority had received outpatient care, 4,661, and 215 had received inpatient care.

We dichotomised the study group into those with end-of-treatment CGAS scores up to 60 and those with scores above 60. This cut-off score was chosen based on previous research on the discriminant validity for CGAS establishing the cut-offs of 60/61 and 70/71 (Shaffer et al. 1983; Bird et al. 1987; Bird et al. 1990). According to these studies, 60 points or lower can be considered a definite case, 61-70 a probable case, and above 70 a probable non-case. This categorisation has been used in several epidemiological prevalence studies of mental disorders to define cases and separate them from non-cases (Bird et al. 1993; Milne et al. 1995; Canino et al. 2004; Geller et al. 2008).

4.2.8 Outcomes and covariates, Paper IV

We assessed whether CGAS ratings at intake and end-of-treatment were related to criminal convictions, suicide attempts, psychiatric morbidity, and accidents in adulthood (from age 18 years) during a period of up to 18 months following the end of

treatment. More specifically, we studied associations to: any criminal offence, suicide attempt (ICD-10 codes X60-X84), depression (ICD-10 codes F32-F34), anxiety disorder (ICD-10 codes F40-F41), eating disorder (ICD-10 codes F50), schizophrenia (ICD-10 codes F20, F25), bipolar disorder (ICD-10 codes F30-F31), borderline personality disorder (ICD-10 code F60.3), alcohol/substance misuse and abuse (ICD-10 codes F10-F19), and accidents (ICD-10 codes V01-X59).

Pastill contains information on psychiatric morbidity in childhood (age 0-17 years). Thus, for each psychiatric outcome we were able to adjust for the corresponding diagnosis at intake (i.e., childhood morbidity) except for i) bipolar disorder that was not specified in Pastill during the study period, and ii) borderline personality disorder, which according to common practice rarely is diagnosed before the age of 18 years. For the outcome of criminal offence, we adjusted for ADHD, oppositional defiant disorder, and conduct disorder at intake. We also used CGAS at intake as a covariate for all associations.

4.2.9 Statistical methods used in Paper IV

All participants contributed person-time from study entry (the date of the CGAS assessment at end-of-treatment) until the date of the outcome event, death, or study end (31 December 2009), whichever came first. The association between childhood psychosocial functioning as rated with CGAS and outcomes (criminal conviction, psychiatric morbidity, and accidents) was measured by hazard ratios (HRs) with 95% confidence intervals (CIs), taking follow-up time into account. HRs were estimated from Cox regression models. We present unadjusted HRs and adjusted HRs (aHRs) for all outcomes. In addition to dichotomising procedure of CGAS ratings to below 61 and above 60, we also used a 10-point scale measure of CGAS in which 1-10 was transformed to 1, 11-20 to 2 etcetera up to 91-100 that was transformed to 10.

5 ETHICS

The research plan for Papers I and II was submitted to the Regional Ethical Review Board in Stockholm. The Ethical Review Board concluded that the research did not require an ethics permit since the research did not include sensitive information about individuals. They had no objections to the research procedures (2006/286-31). Nevertheless, all participants (health care professionals employed by CAMHS) were informed orally and in writing, and they gave their consent by signing a written form.

The Regional Ethical Review Board approved the research conducted in Paper III (2010/1214-31/2). All the data used in Paper III was anonymous and neither individual patients nor health care professionals can be identified or traced in the material. Only two researchers, M.F. and myself, had access to the data.

The Regional Ethical Review Board approved the research conducted in Paper IV (2009/939-31/5). All the data used in Paper IV was anonymous and individual patients could not be identified or traced in the material.

6 RESULTS

6.1 TABLE 3. OVERVIEW OF THE FOUR PAPERS

	I	II	III	IV
Research questions	How reliable and accurate are CGAS ratings in a naturalistic setting with a heterogeneous group of clinicians?	Is training provided by computer as effective as training seminars? Has either type of training effect over and above a comparison group?	Is child psychiatric treatment-as-usual effective in terms of improved global function as measured by CGAS?	Can CGAS ratings predict future mental health disorders, and criminality?
Study design	Cross-sectional study	Randomised comparison of training programmes, non-randomised comparison group	Prospective follow-up study	Prospective follow-up study
Results	Clinicians rated vignettes significantly higher than experts. ICC= 0.73. Social workers and psychologists were significantly more likely to have overall aberrant ratings than medical doctors.	No differences were seen in training effects between computer-based training and seminars. The improvement was modest in both active groups. The comparison group improved by the same order of magnitude.	For mood disorders, several psychotherapies were associated with better outcome but not medication. For ADHD, medication with central stimulants was not associated with better outcome.	CGAS≤60 at end-of-treatment was significantly associated with higher risk of criminality, bipolar disorder and borderline personality disorder.
Conclusions	The reliability and accuracy of CGAS ratings in clinical settings is moderate without prior training.	The results speak in favour of using the less resource demanding computer-based training. However, the overall training effect was too small to be clinically relevant.	The improvement of CGAS after treatment-as-usual differs from RCTs especially in ADHD. The results from clinical trials cannot be extrapolated to routine care.	CGAS ratings at end-of-treatment provide specific long-term prognostic information. Attention should be given to patients with a CGAS rating ≤60 at end-of-treatment.

6.2 CGAS RATINGS IN A LARGE NATURALISTIC SETTING, PAPER I

Paper I surveyed the reliability and validity of CGAS ratings in a large group of clinicians.

6.2.1 Comparisons with expert ratings

Table 4 shows that the untrained raters' mean and median scores were significantly higher than the expert rating for all five cases (t-value 22.4-45.4, $p<0.001$). The range of between-rater scores (min-max) was wide for all 5 cases. The expert raters had a narrower range of scores.

Table 4. Results from the rating of five written case vignettes.

	Expert raters n=5			Untrained raters n=703[1]	
	Mean (SD)	Median (Min-Max)	Expert rating[2]	Mean (SD)	Median (Min-Max)
Case A	14.2 (9.9)	12 (3-30)	9	22.0 (13.4)	20 (1-71)
Case B	24.8 (1.8)	25 (22-27)	25	36.4 (12.5)	38 (8-80)
Case C	46.5 (5.2)	45 (42-54)	45	62.7 (10.4)	63 (31-86)
Case D	62.5 (3.0)	64 (58-64)	64	71.0 (8.2)	72 (48-94)
Case E	51.2 (2.9)	51(48-55)	51	60.5 (11.2)	60 (6-91)

[1] *One participant did not rate case E.*

[2] *The expert rating was established in a consensus discussion among the experienced raters after their individual ratings.*

6.2.2 Inter-rater reliability

The ICC was 0.73 for "rater-to-rater" agreement in the untrained group. In the group with expert raters, the ICC was 0.92. The measurement error was estimated to be 11.3 for untrained raters and 5.2 for experts.

6.2.3 Aberrant ratings

Table 5 shows that psychologists, social workers, and other staff members were twice as likely to score all cases aberrantly as medical doctors. Clinical experience or earlier experience of using CGAS had no statistically significant impact on aberrant ratings, whereas having no experience of GAF increased the risk of aberrant ratings compared to those who were very experienced GAF-users. Sex and age also affected the ratings in the present study: older raters and men had an increased likelihood of being overall aberrant raters.

Table 5. Odds ratio for the likelihood of an overall aberrant rating[1] compared to expert ratings for the respective groups.

	OR (95% CI)	Overall aberrant rating n (%)	Not overall aberrant rating n (%)
Total		197 (28)	506 (72)
Age	**1.03*** (1.01-1.05)		
Sex			
Female	1.00	152 (27)	414 (73)
Male	**1.53*** (1.00-2.32)	45 (33)	92 (67)
Work Experience			
<2 years	1.54 (0.85-2.79)	38 (30)	90 (70)
2-10 years	1.46 (0.93-2.30)	59 (29)	145 (71)
>10 years	1.00	100 (27)	271 (73)
CGAS experience			
No experience	1.26 (0.83-1.91)	151 (30)	345 (70)
Experienced	1.00	46 (22)	159 (78)
GAF experience			
No experience	**1.62*** (1.00-2.63)	67 (34)	133 (66)
Moderately experienced	1.21 (0.72-2.03)	42 (26)	123 (74)
Fairly experienced	1.28 (0.77-2.13)	40 (29)	97 (71)
Very experienced	1.00	48 (24)	151 (76)
Occupation			
Medical Doctor	1.00	15 (16)	77 (84)
Psychologist	**1.91*** (1.01-3.59)	78 (27)	207 (73)
Social worker	**2.46**** (1.26-4.80)	57 (34)	112 (66)
Other staff members	**2.04*** (1.04-4.02)	47 (30)	110 (70)

[1] *Overall aberrant rating was defined as more than ±5 points deviation from the expert ratings for all rated cases*

** significant at 0.05 level, ** significant at 0.01 level*

6.3 THE EFFECT OF RATER TRAINING, PAPER II

Paper II studied the training effect of two training methods, live seminar and CD-ROM.

6.3.1 Inter-rater reliability

With respect to inter-rater reliability, the ICC values at baseline/end-of-study were 0.71/0.78 in the seminar group, 0.76/0.78 in the CD-ROM group, and 0.67/0.79 in the comparison group. Measurement errors at baseline/end-of-study were 11.3/9.9 in the seminar group, 10.8/10.2 in the CD-ROM group, and 12.2/10.1 in the comparison group. The ICC and measurement errors after training suggest a minor improvement

after training. The comparison group's improvement was at the same order of magnitude as the groups with active training.

6.3.2 Comparison of seminar and computer-based training

There were no significant differences between the effects of the two forms of training, seminar and CD-ROM. Even though the improvement in ratings from baseline to end-of-study was statistically significant in both groups, the effect was small and not clinically relevant in either group (Table 6).

Table 6. Comparing seminar and computer-based training, cases A-E.

Case	Expert rating	SEMINAR[1]		CD-ROM[2]		Unpaired t-test		
		Mean (SD) change[4] end-of-study	*P* value[3] end-of-study	Mean (SD) change[4] end-of-study	*P* value[3] end-of-study	*P* value between groups	t-value	df
Change A1 - A2	9	-4.0 (15.0)	<0.001	-2.3 (15.2)	0.017	0.20	-1.3	529
Change B1 - B2	25	-6.0 (13.3)	<0.001	-3.9 (13.5)	<0.001	0.066	-1.8	530
Change C1 - C2	45	-1.6 (13.7)	0.049	-1.6 (11.3)	0.030	0.97	-0.04	528
Change D1 - D2	64	-2.4 (10.0)	<0.001	-2.3 (9.5)	<0.001	0.96	-0.1	531
Change E1 - E2	51	-5.3 (12.9)	<0.001	-5.4 (12.1)	<0.001	0.93	0.1	528

[1] Seminar group n= 285 (Case A, B and D), n=284 (Case C), n=283 (Case E)
[2] CD-ROM group n= 248 (Case D), n=247 (Case B, C and E), n=246 (Case A)
[3] Ho: No change between baseline and end-of-study.
[4] Mean change <0, the raters have improved and are coming closer to expert ratings

6.3.3 No overall effect of training

Data were pooled from the two training groups in order to study the effect of training as a whole. (For further details see Paper II.) Although the merged training group improved significantly on all five cases, compared to improvements on only two out of five cases in the comparison group, there were nonetheless no differences that reached statistical significance between the training group and the comparison group in end-of-study ratings (Table 7).

Table 7. Difference between (baseline rating—expert rating) and (end-of-study rating—expert rating).

	Training groups[1] Mean (SD) change[3] end-of-study	Comparison group[2] Mean (SD) change[3] end-of-study	Paired t-test *P* value between groups	t-value	df
Mean all cases	-3.5 (7.7)	-3.4 (7.6)	0.93	-0.1	572
A1-A2	-3.2 (15.1)	-5.7 (13.1)	0.29	1.1	574
B1-B2	-5.0 (13.4)	-5.4 (13.8)	0.86	0.2	575
C1-C2	-1.6 (12.6)	-0.6 (12.3)	0.63	-0.5	574
D1-D2	-2.3 (9.7)	-2.0 (11.2)	0.80	-0.3	576
E1-E2	-5.4 (12.5)	-3.4 (14.1)	0.32	-1.0	573

[1] Training groups(CD-ROM and seminar) n=533 (case D) 532 (case B) 531 (cases A, C) 530 (case E)

[2] Comparison group n=45

[3] A mean change <0 corresponds to rater improvement, that is, a change in the direction of the expert rating

6.3.4 Trainee satisfaction

Just over one third (241/648) of the participants responded to a web questionnaire regarding their satisfaction with the training they had performed. By and large, the participants were satisfied with both types of training, though those who attended the seminar were slightly more satisfied. On a scale of 1-6 where 6 corresponds to the highest level of satisfaction, the mean grade for the seminar was 5.4 and for the CD-ROM 4.9.

6.4 EFFECTIVENESS OF CHILD PSYCHIATRIC TREATMENT, PAPER III

In Paper III, the effectiveness of child psychiatric treatment was studied by evaluating the change in psychosocial functioning as measured by CGAS in a group who received outpatient care in Stockholm County between 2006 and 2010.

6.4.1 Diagnostic categories

The three most common diagnoses were anxiety disorder, mood disorders and ADHD. Table 8 lists mean CGAS ratings at baseline and ΔCGAS for all diagnoses. The mean CGAS ratings improved (mean ΔCGAS>0) during the course of care across all diagnostic groups. Patients with mental retardation showed the lowest improvement in global functioning with a mean ΔCGAS of 3.9. Those who had attempted suicide and received treatment in outpatient settings showed the highest change in CGAS with a mean ΔCGAS of 16.1.

Patients with mood disorders and ADHD had similar mean CGAS ratings at baseline (50.3 and 50.5), whereas the improvement at end-of-study differed significantly: those

with mood disorders had a mean improvement of 13.4 points, but those with ADHD a mean improvement of only 6.5 points.

Table 8. Diagnostic category, baseline CGAS rating and ΔCGAS at end-of-treatment.

Diagnostic category	n=	Baseline	ΔCGAS[1]	Paired t-test		
		Mean (SD)	Mean (SD)	t-value*	df	95% C.I.
Suicide attempt	302	43.5 (9.9)	16.1 (14.5)	19.4	301	14.5, 17.8
OCD	606	49.4 (9.7)	14.1 (13.3)	26.1	605	13.0, 15.2
Somatoform disorder	71	51.3 (10.0)	14.0 (14.6)	8.1	70	10.5, 17.4
Mood disorders	2,213	50.3 (9.8)	13.4 (12.2)	51.7	2,212	12.9, 13.9
Anxiety disorder	2,446	51.3 (10.2)	13.2 (11.9)	54.7	2,445	12.7, 13.7
Identity problem	529	51.6 (10.3)	11.9 (11.1)	24.7	528	10.9, 12.8
Eating disorder	613	48.2 (11.3)	11.8 (12.8)	22.8	612	10.8, 12.8
Sleep disorder	605	49.4 (11.1)	11.7 (11.7)	24.5	604	10.7, 12.6
PTSD	440	52.4 (9.3)	11.6 (12.5)	19.4	439	10.4, 12.7
Dissociative disorder	76	45.8 (11.8)	11.6 (13.7)	7.3	75	8.4, 14.7
Psychotic disorder	23	41.4 (9.0)	11.1 (10.9)	4.9	22	6.4, 15.8
Gender identity disorder	28	50.5 (9.6)	10.8 (11.2)	5.1	27	6.5, 15.2
Other	1,553	59.1 (8.8)	10.8 (9.3)	45.8	1,552	10.3, 11.2
Enuresis	75	51.4 (12.0)	10.5 (11.4)	7.9	74	7.9, 13.1
Drug related disorder	78	47.2 (9.3)	9.7 (12.3)	7.0	77	7.0, 12.5
No observed symptoms	169	67.1 (12.0)	9.7 (10.4)	12.2	168	8.1, 11.3
Encopresis	100	53.5 (10.8)	9.5 (10.2)	9.3	99	7.5, 11.5
Oppositional defiant disorder	643	51.4 (8.8)	9.3 (10.0)	23.5	642	8.5, 10.0
Conduct disorder	390	50.1 (10.2)	8.8 (10.5)	16.6	389	7.8, 9.9
Reactive attachment disorder	10	46.1 (6.3)	8.8 (10.4)	2.7	9	1.3, 16.3
Lack of information	1,690	56.5 (10.5)	8.2 (9.6)	34.8	1,689	7.7, 8.6
Tic disorder	151	52.9 (10.0)	8.0 (8.3)	11.9	150	6.7, 9.4
Eating disorder, infant	19	53.9 (9.6)	7.8 (12.2)	2.8	18	1.9, 13.7
Motor skills disorder	63	46.9 (11.2)	7.4 (11.0)	5.3	62	4.6, 10.2
Learning disorder	445	50.3 (9.3)	6.8 (9.4)	15.2	444	5.9, 7.6
ADHD	1,169	50.5 (9.5)	6.5 (8.9)	25.0	1,168	6.0, 7.1
Communication disorder	404	47.7 (11.2)	5.4 (8.3)	13.1	403	4.6, 6.2
Pervasive developmental disorder	1,053	45.0 (10.3)	4.3 (8.5)	16.6	1,052	3.8, 4.8
Mental retardation	224	41.2 (14.0)	3.9 (9.6)	6.2	223	2.7, 5.2

[1] *ΔCGAS=CGAS at end-of-treatment minus CGAS at baseline.*

**all differences were significant at 0.001 level.*

6.4.2 Correlation between ΔCGAS and overall assessment of treatment response

There was a moderate correlation between ΔCGAS and the assessment of treatment response made by the managing clinician after final visit ($r=0.47$, $p<0.001$).

6.4.3 Outcome predictors

A complete table is displayed in Paper III of all the variables from the stepwise regression that had an F-value with a probability of less than 0.05. A higher CGAS baseline rating was predictive of lower ΔCGAS across all four diagnostic groups - mood disorder, ADHD, conduct disorder and obsessive compulsive disorder - which is in line with other studies (Wiggins et al. 2010; Setoya et al. 2011). Also, a higher number of diagnoses (comorbidity) was associated with lower ΔCGAS in all groups except conduct disorder.

Taken as a whole, the diverse psychotherapies offered within CAMHS to patients with mood disorders predicted improvement of CGAS ratings, whilst there was no significant positive effect from medication. For patients with ADHD, the number of appointments, the use of group parent counselling, and having a physician manage the cases were all predictors of improvement, whereas treatment with central stimulants was not. Two family-oriented treatments were positive predictors of outcome among those with conduct disorder (group counselling for parents and family therapy). Psychotherapy also led to improved CGAS scores for those with OCD, whereas guidance to parents predicted less improvement in CGAS.

To better understand the lack of significant positive effect of central stimulants in ADHD, those treated with central stimulants were compared post hoc with the ADHD cases where no central stimulants were prescribed (Table 9).

Table 9. Background and intervention variables in ADHD with or without central stimulants (n=1,169[1]).

	Central stimulants n=132	No central stimulants n=910	Unpaired t-test		
	Mean (SD)	Mean (SD)	*P* value	t-value	df
Age	13.3 (3.5)	11.0 (3.8)	<0.001	-6.6	1,040
Number of diagnoses	1.9 (1.1)	1.9 (1.1)	0.84	0.2	1,040
Number of psychosocial stressors	1.8 (1.3)	2.2 (1.7)	0.008	2.7	1,039
CGAS at baseline	52.2 (9.5)	50.3 (9.6)	0.038	-2.1	1,040
Number of visits	14.8 (13.8)	13.9 (13.8)	0.50	-0.7	1,040
ΔCGAS[2]	7.1 (10.0)	6.4 (8.9)	0.38	-0.9	1,040
	N (%)	N (%)	*P* Value	Pearson Chi-2	df
Sex					
Boys	95 (72.0)	658 (72.3)	0.94	0.01	1
Girls	37 (28.0)	252 (27.7)	0.94	0.01	1
Treatment intervention					
Guidance to parents	82 (62.1)	707 (77.7)	<0.001	15.2	1
Guidance to teenagers	47 (35.6)	363 (39.9)	0.35	0.9	1
Family therapy/counselling	19 (14.4)	299 (32.9)	<0.001	18.5	1
Teenage psychotherapy	5 (3.8)	32 (3.5)	0.88	0.02	1
Social network counselling	9 (6.8)	102 (11.2)	0.13	2.3	1
Cooperation counselling	0	22 (2.2)	0.086	3.0	1
Group treatment, child/adolescent	2 (1.5)	29 (3.2)	0.29	1.1	1
Child psychotherapy	1 (0.8)	4 (0.4)	0.62	0.2	1
Short term psychotherapy child/adolescent	4 (3.0)	18 (2.0)	0.43	0.6	1
Interaction treatment	1 (0.8)	12 (1.3)	0.59	0.3	1
Medication, excluding central stimulants	19 (14.4)	32 (3.5)	<0.001	29.3	1
Therapeutic summer camp	1 (0.8)	3 (0.3)	0.46	0.6	1
Environmental therapy	0	8 (0.9)	0.28	1.2	1
Unspecified treatment	9 (6.8)	42 (4.6)	0.27	1.2	1
No treatment	0	33 (3.6)	0.026	4.9	1

[1]*Missing data about treatment intervention, n=127 cases*

[2]*Difference between CGAS rating at baseline and CGAS rating at end-of-treatment*

This comparison showed that the group receiving central stimulants was older at the beginning of the treatment and had a two point higher CGAS baseline rating (that is were less functionally impaired) than those in the group that did not receive medication. There was no difference between the groups in number of diagnoses, number of visits or ΔCGAS**.** Treatment intervention differed not only with central stimulants. The group without central stimulants received more psychotherapeutic interventions such as guidance to parents and family therapy than the ADHD group who were treated with central stimulants.

6.5 CGAS RATINGS AS PREDICTORS OF FUTURE MENTAL HEALTH, PAPER IV

In Paper IV, we tested whether CGAS ratings at end-of-treatment predicts future negative outcomes in young adults. In total, 4,876 patients with valid CGAS data at the end-of-treatment were included in this prospective follow-up study (Table 10).

Table 10. Intake characteristics of individuals with CGAS≤60 at end-of-treatment and controls with CGAS>60 at end-of-treatment.

Characteristics at intake	CGAS≤60 at end-of-treatment (n = 2,260)	CGAS>60 at end-of-treatment (n = 2,616)
Sex		
Boys	868 (38.4%)	902 (34.5%)
Girls	1,392 (61.6%)	1,714 (65.5%)
Age		
Average age at end-of-treatment (SD)	16.5 (0.9)	16.5 (0.9)
Childhood psychiatric morbidity		
ADHD	116 (5.1%)	62 (2.4%)
Oppositional defiant disorder	50 (2.2%)	28 (1.1%)
Conduct disorder	45 (2.0%)	23 (0.9%)
Suicide attempt	47 (2.1%)	55 (2.1%)
Mood disorder	320 (14.2%)	524 (20.0%)
Anxiety disorder	252 (11.2%)	386 (14.8%)
Eating disorder	106 (4.7%)	126 (4.8%)
Psychotic disorder	3 (0.1%)	2 (0.1%)
Substance misuse	34 (1.5%)	15 (0.6%)
CGAS at intake		
≤60	2,173 (98.3%)	1,271 (49.9%)
>60	38 (1.7%)	1,268 (50.1%)
Mean rating (SD)	47.4 (9.5)	60.3 (10.9)

The study group was dichotomised into one group with end-of-treatment CGAS ratings of 60 and lower (n=2,260) and a second group with CGAS ratings above 60 (n=2,216).

The proportion of girls and boys, as well as the average age, did not differ between the two groups. The frequency of ADHD, conduct disorder, oppositional defiant disorder, and substance misuse was higher among those with CGAS 60 or lower, whereas mood disorder and anxiety disorder were more common in the higher than CGAS 60 group. Perhaps unsurprisingly, those who had CGAS≤60 at end-of-treatment also had lower CGAS ratings at intake. The follow-up time to first event after end-of-treatment ranged from 1 to 1½ years.

To examine whether CGAS at end-of-treatment predicted adversities, we conducted a series of Cox regression analyses. Table 11 summarizes the HRs for the outcome variables in the group with CGAS≤60 at end-of-treatment compared to the group with CGAS>60 at end-of-treatment as reference. Table 11 also shows the HRs for the outcome variables based on 10-point intervals on the CGAS.

The risk of criminal conviction was more than twice as high (HR 2.1, 95%CI 1.4-3.2) for those with CGAS≤60 at end-of-treatment compared to those with end-of-treatment CGAS scores above 60. In fact, all adult adverse outcomes except for accidents were more likely to occur in the CGAS≤60 at end-of-treatment group, with unadjusted HRs ranging from 1.9 for substance misuse to 11.7 for borderline personality disorder. The same pattern was seen in the 10-point measure of CGAS; where CGAS ratings were related to all adverse outcomes in adulthood except accidents.

HRs adjusted for age, sex, calendar year, and childhood psychiatric morbidity provide information on whether CGAS ratings add information over and beyond what can be inferred from these covariates alone. As shown in Table 11, the aHRs remained increased for all outcomes and significant with respect to criminal convictions, bipolar disorder, and borderline personality disorder, suggesting that CGAS ratings at end-of-treatment provide specific prognostic information for these conditions. The 10-point measure of CGAS yielded a similar pattern of association with the outcomes: in addition, CGAS scores also correlated with depression and schizophrenia at levels of statistical significance.

Table 11. Risk of adversities among child psychiatry patients in Sweden 1 July 2006-31 December 2009 with CGAS end-of-treatment ≤60 vs >60.

	CGAS end-of-treatment ≤60 versus CGAS end-of-treatment >60					CGAS end-of-treatm: 10-point interval		
Outcomes	Events, n= CGAS≤60 vs >60	Outcome incident rate per 1000 person-years (95% CI)		Unadjusted HR (95% CI)	Adjusted HR[a] (95% CI)	Adjusted HR[ab] (95% CI)	Unadjusted HR (95% CI)	Adjusted HR[a] (95% CI)	Adjusted HR[ab] (95% CI)
	n^1=2,260/n^2=2,616	CGAS≤60	CGAS>60						
Crime	60/36	18.8 (14.6-24.3)	9.1 (6.6-12.7)	**2.1** *** (1.4-3.2)	**1.8**[c] * (1.1-3.0)	**2.4**[c] ** (1.2-4.8)	**1.2** ** (1.1-1.4)	**1.3**[c] ** (1.1-1.5)	**1.4**[c] * (1.1-1.8)
Suicide attempt	24/8	7.5 (5.0-11.1)	2.0 (1.0-4.0)	**3.8** ** (1.7-8.4)	1.5[d] (0.5-5.0)	1.1[d] (0.3-4.0)	**1.5** *** (1.2-1.8)	1.3[d] (0.8-1.9)	1.1[d] (0.7-1.9)
Depression	59/29	18.6 (14.4-24.0)	7.3 (5.1-10.6)	2.6 *** (1.6-4.0)	1.6[e] (0.9-2.8)	1.5[e] (0.8-2.8)	**1.4** *** (1.2-1.6)	**1.3**[e] * (1.1-1.5)	1.2[e] (1.0-1.6)
Anxiety disorder	52/26	16.3 (12.4-21.4)	6.6 (4.5-9.7)	**2.6** *** (1.6-4.1)	1.5[f] (0.8-2.9)	1.3[f] (0.6-2.9)	**1.4** *** (1.2-1.6)	1.2[f] (1.0-1.5)	1.2[f] (0.9-1.5)
Eating disorder	11/4	3.4 (1.9-6.2)	1.0 (0.4-2.7)	3.4 * (1.1-10.6)	2.7[g] (0.6-11.5)	8.1[g] (0.9-74.0)	**1.4** * (1.1-1.9)	1.4[g] (0.9-2.2)	1.9[g] (0.9-3.8)
Schizophrenia	6/0	1.9 (0.8-4.1)	N/A	N/A	N/A	N/A	**2.6** *** (1.6-4.0)	**5.6**[h] * (1.2-26.2)	N/A
Bipolar disorder	20/4	6.2 (4.0-9.6)	1.0 (0.4-2.7)	**6.3** *** (2.2-18.5)	**6.0** ** (2.0-17.7)	**4.7** * (1.2-17.7)	1.6 *** (1.3-2.0)	**1.6** *** (1.2-2.0)	1.2 (0.8-1.7)
Borderline personality disorder	19/2	5.9 (3.8-9.3)	0.5 (0.1-2.0)	**11.7** *** (2.7-50.7)	**11.5** ** (2.7-49.8)	**15.7** ** (2.2-113.0)	**1.9** *** (1.5-2.4)	**1.9** *** (1.5-2.4)	1.4 (0.9-2.3)
Substance misuse	43/28	13.4 (10.0-18.1)	7.1 (4.9-10.3)	**1.9** ** (1.2-3.1)	1.4[i] (0.7-2.6)	1.2[i] (0.6-2.5)	**1.3** *** (1.1-1.5)	1.2[i] (1.0-1.5)	1.1[i] (0.9-1.5)
Accidents	41/45	12.8 (9.4-17.3)	11.4 (8.5-15.3)	1.1 (0.7-1.7)	1.0 (0.7-1.6)	0.8 (0.5-1.4)	1.1 (0.9-1.2)	1.1 (0.9-1.2)	1.0 (0.8-1.2)

Adjusted for [a] age, sex, and, calendar year [b] CGAS at intake, [c] childhood ADHD, conduct disorder and oppositional defiant disorder, [d] childhood suicide attempt, [e] childhood mood disorder, [f] childhood anxiety disorder, [g] childhood eating disorder [h] childhood psychotic disorder, [i] childhood substance misuse

** Significant at 0.05 level, ** Significant at 0.01 level, *** Significant at 0.001 level*

n^1 number of patients with CGAS at end-of-treatment ≤60, n^2 number of patients with CGAS at end-of-treatment >60

As shown in Table 10, CGAS ratings at intake were lower among those with CGAS≤60 at end-of-treatment. This first raises the question of whether the predictive value of end-of-treatment CGAS ratings can be explained by CGAS at intake. However, adjusting for CGAS at intake revealed that CGAS at end-of-treatment is an independent predictor of these adult adversities (Table 11). For example, after adjusting for all covariates including CGAS at intake the risk for criminal conviction was still increased with 2.4 (1.2-4.8) for the patients with low CGAS at end-of-treatment.

A second question is whether CGAS ratings at intake might also be an independent predictor of adult adversities. To explore this, the analyses in Table 11 were repeated using CGAS ratings at intake as the explanatory variable instead of end-of-treatment CGAS. Table 12 shows the results of these analyses. There were 5,903 patients with valid CGAS ratings at intake of which 4,402 scored 60 or lower, and 1,501 scored above 60. The follow-up time to first event after intake ranged from 1½ to 2 years.

Patients with CGAS≤60 at intake had an increased risk (unadjusted HRs) for suicide attempt, depression, anxiety disorder, and bipolar disorder compared to patients with CGAS>60 at intake. However, all these associations declined substantially when adjusted for age, sex, calendar year, and childhood psychiatric morbidity, leaving only bipolar disorder as statistically significant. However, the association between CGAS at intake and bipolar disorder also declined and was statistically non-significant when adjusting for CGAS at end-of-treatment. Similar patterns of associations were observed when we used the 10-point measure of CGAS at intake. This suggests that CGAS at intake does not predict adult adversities independent of age, sex, calendar year, childhood psychiatric morbidity and CGAS at end-of-treatment.

Table 12. Risk of adversities among child psychiatry patients in Sweden 1 July 2006-31 December 2009 with CGAS intake ≤60 vs >60.

		CGAS intake ≤60 versus CGAS intake >60					CGAS intake: 10-point interval		
Outcomes	Events, n= CGAS≤60 vs >60	Outcome incidence rate per 1000 person-years (95% CI)		Unadjusted HR (95% CI)	Adjusted HR[a] (95% CI)	Adjusted HR[ab] (95% CI)	Unadjusted HR (95% CI)	Adjusted HR[a] (95% CI)	Adjusted HR[ab] (95% CI)
	n^1=4,402/n^2=1,501	CGAS≤60	CGAS>60						
Crime	74/26	9.9 (7.9-12.4)	8.7 (5.9-12.7)	1.3 (0.8-2.0)	0.9 (0.5-1.5)	0.5[c] (0.2-0.9)	1.1 (1.0-1.3)	1.0[c] (0.8-1.3)	0.7[c] (0.6-1.0)
Suicide attempt	30/4	4.0 (2.8-5.7)	1.3 (0.5-3.5)	**3.1** * (1.1-8.9)	2.1[d] (0.5-9.8)	1.7[d] (0.3-9.0)	**1.5** ** (1.1-1.9)	1.3[d] (0.8-2.2)	1.0[d] (0.5-2.1)
Depression	78/14	10.4 (8.3-13.0)	4.7 (2.9-8.1)	**2.4** ** (1.4-4.2)	1.5[e] (0.7-2.8)	0.9[e] (0.4-2.0)	**1.4** *** (1.2-1.6)	1.2[e] (0.9-1.5)	0.9[e] (0.6-1.3)
Anxiety disorder	66/14	8.8 (6.9-11.2)	4.7 (2.8-7.9)	**2.1** * (1.2-3.7)	1.3[f] (0.6-2.7)	0.9[f] (0.4-2.2)	**1.3** *** (1.1-1.6)	1.1[f] (0.8-1.5)	0.9[f] (0.6-1.3)
Eating disorder	10/3	1.3 (0.7-2.5)	1.0 (0.3-3.1)	1.4 (0.4-5.1)	0.8[g] (0.2-4.4)	0.2[g] (0.0-2.0)	1.1 (0.7-1.7)	1.0[g] (0.5-2.0)	0.6[g] (0.2-1.3)
Schizophrenia	6/0	0.8 (0.4-1.8)	N/A	N/A	N/A	N/A	**2.4** *** (1.5-3.9)	4.0[h] (0.9-16.6)	N/A
Bipolar disorder	22/2	2.9 (1.9-4.4)	0.7 (0.2-2.7)	**4.8** ** (1.1-20.6)	**4.4** * (1.0-18.8)	1.3 (0.2-7.2)	**1.8** *** (1.4-2.4)	**1.9** *** (1.4-2.5)	1.4 (0.9-2.2)
Borderline personality disorder	17/2	2.3 (1.4-3.6)	0.7 (0.2-2.7)	3.5 (0.8-15.0)	3.2 (0.7-14.0)	0.5 (0.1-3.1)	**2.0** *** (1.5-2.7)	**2.0** *** (1.5-2.8)	1.3 (0.7-2.2)
Substance misuse	60/16	8.0 (6.2-10.3)	5.3 (3.3-8.7)	1.7 (1.0-3.0)	1.1[i] (0.6-2.2)	0.8[i] (0.4-1.7)	**1.4** *** (1.2-1.6)	1.1[i] (0.9-1.4)	0.9[i] (0.7-1.3)
Accidents	72/24	9.6 (7.6-12.1)	8.0 (5.4-12.0)	1.3 (0.8-2.1)	1.2 (0.7-1.9)	1.1 (0.6-2.0)	1.1 (0.9-1.3)	1.0 (0.9-1.2)	1.0 (0.8-1.2)

Adjusted for [a] age, sex and, calendar year [b] CGAS at end-of-treatment, [c] childhood ADHD, conduct disorder and oppositional defiant disorder, [d] childhood suicide attempt, [e] childhood mood disorder, [f] childhood anxiety disorder, [g] childhood eating disorder [h] childhood psychotic disorder, [i] childhood substance misuse

** Significant at 0.05 level, ** Significant at 0.01 level, *** Significant at 0.001 level*

n^1 number of patients with CGAS at intake ≤60, n^2 number of patients with CGAS at intake >60

7 GENERAL DISCUSSION

7.1 INTER-RATER RELIABILITY

Clinical researchers are especially interested in inter-rater reliability and I often get questions about the ICCs for CGAS ratings. High levels of inter-rater reliability suggest higher quality ratings, which minimises variability and results in less distortion of outcome measures, for example, in clinical trials (Rosen et al. 2008).

The child psychiatric literature suggests that the ICC for rating instruments should be greater than 0.80 (Myers and Winters 2002), even though literature on rating scales cites varying acceptable levels of ICC, starting as low as 0.41 (Anastasi and Urbina 1997; Shrout 1998; Renou et al. 2004). For comparison's sake, intensive training in the use of the widely used Hamilton Depression Rating Scale (HDRS) led to an ICC above 0.90.

The present study found an ICC of 0.73, which according to Shrout's standards falls into the range of *moderate* reliability (0.61-0.80). This result accords to previous smaller studies of CGAS in clinical settings that have found ICCs ranging from 0.53 to 0.90 (Green et al. 1994; Rey et al. 1995; Dyrborg et al. 2000; Hanssen-Bauer et al. 2007). In the expert group, however, the ICC reached 0.92, which corresponds to *substantial* inter-reliability. Together these findings suggest that, although the ICC would ideally be higher, CGAS is reliable enough to be used in clinical settings.

It is known that the focus and characteristics of the case influence the reliability of CGAS ratings. This includes the child's individual symptoms, diagnosis, family dynamics, or psychosocial risk factors. The inter-rater reliability in CGAS at two different time points has also been shown to show less agreement in cases with less severe emotional disorders than in cases with severe or more clear-cut single symptom disorders (Shaffer et al. 1983; Steinhausen 1987). In line with this reasoning, the range and distributions of the ratings in this study differed between the five rated cases (data not shown).

7.2 VALIDITY AND THE BEST ESTIMATE FOR A GOLD STANDARD

In contrast to the many inquiries about CGAS's reliability I receive from clinicians, I seldom get questions about the scale's validity. Establishing a rating scale's validity is difficult: it usually takes several years, and is especially challenging because there is no natural gold standard for psychiatric instruments. Widely used older scales are, furthermore, seldom revalidated after their introduction (Myers and Winters 2002).

Papers III and IV investigate different aspects of CGAS' validity. The effectiveness study in Paper III found that the CGAS ratings in clinical practise was comparable with clinical trials and follow-up studies (Table 1), and thus support the validity of the scale. Paper IV found that CGAS has predictive validity for a number of adverse outcomes.

For the purpose of Papers I and II, I chose to establish validity by allowing a group of experienced clinicians to together create a gold standard for the five cases I used. Of course these ratings are themselves contestable, but there is good reason to believe that

five experienced clinicians can reach a rating that can be considered a valid best estimate. Even though five experts may seem a small number, it is greater than the number of experts used to establish baseline ratings in previous studies of rating scales (Miller et al. 2003; Hanssen-Bauer et al. 2007).

Both Paper I and the Hanssen-Bauer study showed that when a group does ratings they tend to be lower than when individuals rate cases alone (2007). Why would group discussion lead to lower ratings of functional impairment than individual ratings? Does the group member with the lowest rating persuade the others? Or are clinicians more likely to follow the instruction that the rating should reflect the *lowest* level of functioning when in a group? Probably both of these mechanisms are at work.

Perhaps the most significant reason that raters in our studies deviated from the gold standard, virtually always setting a higher less functionally impaired score, was that they advertently or inadvertently violated the instructions. The instructions for using CGAS in our study followed Shaffer's original instructions and are clear: Score the most impaired level of global psychosocial functioning during the last month (Shaffer et al. 1983). During seminars, however, some clinicians tried to explain away a low level of functioning as a product of circumstances, saying for example, "If this child did not live in this situation then he/she would function better in school." In the same vein, some clinicians focused on the strengths and positive behaviours of the child instead of rating according to the *lowest* level of functioning.

Professionals working with crisis intervention in particular found that focusing on a patient's lowest level of functioning were at odds with the way they tend to work. Since they usually first meet children at a time of extreme functional impairment, these clinicians commonly experience a large drop in the child's symptom burden after only a few days of treatment. Yet, if they would adhere to the CGAS's instructions to rate the preceding month, the ratings at intake and at end-of-treatment a few days later would not reflect any improvement. The usage of the one-month time period for rating CGAS might be inappropriate in these settings, and that either another rating instrument, or using CGAS with a shortened time frame would be more useful.

Surprisingly, given Shaffer's original intentions for CGAS, other studies on CGAS have used numerous other time frames and various levels of functioning: for example, the highest level of functioning during a couple of months together with the average level of functioning during the last three weeks (Rey et al. 1995); the lowest level during the last three months (Dyrborg et al. 2000); the lowest level during the last week (Green et al. 1994); the average level during a whole lifetime (Weissman et al. 1990); the lowest level during the last two weeks (Hanssen-Bauer et al. 2007); the lowest level during the last four days (Szobot et al. 2004); and the current, the highest in the past, and the most severe in the past (Bella et al. 2011). It is also surprisingly common that studies fail to explain what rating strategy research subjects have been instructed to use (Table 1). Different rating instructions inevitably result in differences in the scale's reliability and validity.

7.2.1 Variation among raters

Social workers, psychologists, and other staff members were significantly more likely to have overall aberrant ratings than medical doctors. This is in line with a previous

small study that found that experienced clinicians (n=3) had a better ICC than trainees in child and adolescent psychiatry (n=2) (Dyrborg et al. 2000), but these results contrast with other studies that have not found that clinical experience or occupational background affect ratings (Steinhausen 1987; Hanssen-Bauer et al. 2007).

A possible explanation for this variation is that medical doctors have more training in collecting and merging clinical information into a short assessment and a preliminary diagnosis than psychologists and social workers have, a process similar to that needed to reach a CGAS rating.

Sex and age also affected the ratings. Older raters and men had an increased likelihood of being overall aberrant raters. To my knowledge there are no previous studies on this issue. Some experienced clinicians explained to me before training sessions that they did not need to participate in the CGAS training because of their long clinical experience and the scale's user-friendly construction. Maybe this attitude correlates with less motivation to heed the training instructions. A Swedish study has shown that a positive attitude towards another similar unidimensional scale, GAF, minimized measurement error and led to better reliability (Söderberg et al. 2005). A possible explanation for men's tendency to have more aberrant ratings is that when the baseline ratings occurred, on average the male participants did not use the whole hour and delivered their results sooner than the female raters. From this anecdotal information one might hypothesize that spending more time and care on the ratings leads to more thoughtful and accurate ratings.

7.3 THE EFFECTS OF DIFFERENT TRAINING METHODS

Study II found that CGAS training in a "live" seminar is no more effective than computer-based training. Moreover, and perhaps more importantly, neither form of training led to clinically significant improvements in clinicians' ability to use the scale accurately. Although the merged training group improved significantly on all five cases, compared to only two out of five cases in the comparison group, there were nonetheless no significant differences between the training group and the non-randomised comparison group end-of-study ratings. As discussed above, the inter-rater reliability was also only moderate, in both the training groups and the control groups. Even though these findings are in line with a GAF training study where training did little to bring ratings closer to the established gold standard (Bates et al. 2002), this small training effect was a disappointment.

In retrospect, when planning this study I took for granted that training leads to learning. CAMHS in Stockholm devotes more than 10% of its total budget to training activities, suggesting that faith in training's effectiveness is widespread. This is despite recent research that has pointed out that educational activities are seldom evaluated for their effectiveness, only for how participants felt about them (Kirkpatrick and Kirkpatrick 2005). This is consistent with my results, which showed widespread appreciation among participants of both forms of training, even though the same participants' rating skills did not change enough to be of clinical relevance.

There are two possible explanations for why we saw no significant effect from the trainings: that we tested participants too long after the training so that any effects had

waned, and that the comparison group was neither a randomly assigned control group, nor as large as the intervention groups. At the start of the study we planned to test participants six months after training. However, the number of included units increased and this delayed the end-of-study ratings by an additional six months. In the end, one year had lapsed between the time of training and the time of testing, and this may well explain the lack of any significant training effect. This hypothesis is supported by a recently published study on training of suicide prevention gatekeepers, which showed loss of actual skills, but not loss of knowledge and attitudes, over a three-month period (Cross et al. 2011).

Because of the lack of a randomised and sufficiently large control group, the study's results cannot be interpreted in the same manner as a randomised placebo-controlled clinical trial. It is striking that none of the previous studies that have evaluated training programmes for different rating scales included a control group, which makes it difficult to determine whether the improved results were the effects of training or an effect of non-specific factors (Bates et al. 2002; Kobak et al. 2003; Kobak et al. 2006; Kobak et al. 2007; Rosen et al. 2008). One conclusion to draw from the present results is that further research on educational activities should be designed as regular RCTs, and include randomly assigned control groups to control for the placebo improvement that might occur.

Given the lack of differences in the outcomes of the two training methods, our study suggests that if future CGAS trainings are to be carried out they can use the more flexible and probably less resource-demanding computer based training method, easily adjusted for the Internet, and reachable for all mental health professionals in Stockholm and in Sweden.

7.4 THE EFFECTIVENESS OF CHILD PSYCHIATRIC TREATMENTS AS MEASURED BY CGAS

In Paper III, we used CGAS to measure treatment outcomes in child psychiatric outpatients with a range of psychiatric diagnoses. The change in CGAS ratings from intake to case closure generally reflected what a range of other studies have indicated about child psychiatric treatment. In large diagnostic groups such as mood disorders and ADHD, changes in CGAS ratings suggested that some established therapies may not be as effective as thought.

Those with mental retardation showed the lowest improvement in global functioning with a mean ΔCGAS of 3.9. This should come as no surprise given the lack of effective treatment available for this group within CAMHS. In fact, this group receives most of their care and training within the school system.

By contrast, patients who had attempted suicide and received treatment in outpatient settings showed the highest change in CGAS with a mean ΔCGAS of 16.1. This may indicate successful crisis intervention, but may also be the result of the natural reversion of the crisis and improvement of CGAS, known as regression towards the mean. This phenomenon was described already in the 19th century and showed that extreme values tend to get closer to the mean value at the second measuring point. Adding a control group would be one way to control for this possible effect, but this is

not an option in such a large naturalistic study as this one. It should be noted that the suicide attempts included in this study were among the less severe, since more severe cases receive inpatient treatment and were therefore not included.

7.4.1 Mood disorders

The group with mood disorders improved with a mean ΔCGAS of 13.4. This size of improvement, as well as the level of baseline ratings, is almost identical with the results from previous efficacy studies of depressed adolescents (Wagner et al. 2003; March et al. 2004; Mufson et al. 2004; Wagner et al. 2006). Our results of CAMHS treatment as usual of mood disorders can also be compared with a large-scale clinical trial, the Treatment of Adolescent Depression Study (TADS). Compared to adolescents who had received what proved to be the most effective intervention in TADS – a combination of selective serotonin reuptake inhibitors (SSRI) and cognitive behavioural therapy (CBT) – CAMHS patients were slightly less improved, but in the same range as the two other TADS intervention groups: the one that received only SSRIs or CBT, respectively (Vitiello et al. 2006). In a psychotherapy trial (Brent et al. 1997) comparing cognitive, family, and supportive therapy, the CGAS baseline was 4-8 points higher (less impaired) and the improvement for all groups 3-5 points lower compared to Paper III.

In contrast to the TADS study where SSRIs alone had almost the same effect as SSRIs combined with CBT, and CBT alone had almost no effect (March et al. 2006; Vitiello et al. 2006), we found that psychotherapy predicted improvement whereas medication did not. The results in Paper III are therefore at odds with the Swedish clinical guidelines on depression and anxiety disorder published (National Board of Health and Wellfare 2010), which are in part based on the results of the TADS study.

7.4.2 ADHD

Whereas the level of improvement in the mood disorder group by and large corresponded to previous efficacy studies, this was not the case in the ADHD group. First, the baseline rating in our study was 50.5, which is between 2 and 12 points lower than baseline ratings in previous ADHD studies (Szobot et al. 2004; Preuss et al. 2006; Kratochvil et al. 2007; Findling et al. 2008; Berek et al. 2011). Second, the mean CGAS improvement in the ADHD group was only 6.5 points, which can be compared with three small efficacy studies in which the mean CGAS improvement ranged from 8.2 to 18.9 (Szobot et al.

2004; Kratochvil et al. 2007; Findling et al. 2008). One of these, the MTA study, recruited a group of patients from a naturalistic population but there are unfortunately no CGAS ratings in that study to compare our results with (MTA Cooperative Group 1999). Taken together, our results and previous studies suggest that ADHD patients in clinical trials are less impaired and improve more with treatment than do ADHD patients in a naturalistic study population. This means that the results of ADHD clinical trials cannot be generalised to the real world ADHD population.

The treatment effectiveness of central stimulants has been the subject of intense discussion since the MTA 8 year prospective follow-up was published (Molina et al. 2009). We found that central stimulants did not predict improvement in the ADHD group. Instead, factors that predicted better outcomes for ADHD patients were the number of appointments, group parent counselling, and that a physician managed the case.

This finding prompted us to conduct a post hoc comparison between ADHD cases with and without central stimulants. This comparison showed that the group receiving central stimulants was older at the beginning of the treatment and had a two point higher CGAS baseline rating (less functionally impaired) compared with the group with no medication. Since the linear regression analysis revealed that higher age was positively associated with ΔCGAS, this age difference could not explain why treatment with central stimulants did not predict ΔCGAS. However, the regression analysis also revealed that a higher CGAS baseline rating was negatively associated with ΔCGAS. The group difference on CGAS at baseline might thus partially account for the fact that central stimulants did not predict ΔCGAS, even though the difference was only two points.

More importantly, however, is that the ADHD cases that continued treatment – and therefore not included in the study due to the lack of a CGAS rating at case closure – differed significantly with respect to the higher frequency of treatment by central stimulants. One possibility is therefore that cases with a positive effect from central stimulants were more likely to stay in treatment and therefore more likely to be excluded from this study. There is also the possibility that the selection worked the other way around, that is, that those with poor treatment results tended to require longer treatment periods.

The patients who received central stimulants also differed from those who did not receive medication with respect to other interventions. The non-medication group received more guidance to parents and more family therapy/counselling than the group with central stimulants. One possibility is therefore that this difference might have cancelled out any positive effect from central stimulants.

We cannot settle these issues with the current data set, and our results with respect to the effect of central stimulants should therefore be treated with caution. These results should, however, also prompt a critical discussion about whether patients receive the best available treatment within CAMHS in Stockholm and elsewhere.

Leaving the question of central stimulants' effectiveness aside, our results lend support to clinical guidelines that recommend parent counselling/training, school support, and behavioural modification for patients with ADHD (National Institute for Health and Clinical Excellence 2008).

7.4.3 Conduct disorder

Two family oriented treatments, group counselling for parents and family therapy, were positive predictors of outcome among those with conduct disorder. This is in agreement with available evidence that multi-systemic therapy is the treatment of choice for conduct disorder (Henggeler et al. 1995). Even though patients with conduct disorder represented a small fraction of the total number, the group is important to identify and treat due to the association of conduct disorder with adversities in adulthood, including criminality (Fergusson et al. 2005; Engqvist and Rydelius 2007; Sourander et al. 2007; Mordre et al. 2011) and psychiatric morbidity (Kim-Cohen et al. 2003; Olsson et al. 2006). The mean CGAS at baseline was 50.1 and the mean ΔCGAS was 8.8, leading to a mean CGAS at end-of-treatment below 61. Unfortunately, we found in Paper IV that this post-treatment rating was associated with an increased risk of negative outcomes in early adulthood, especially criminal conviction, bipolar disorder, and borderline personality disorder (Paper IV).

7.4.4 Obsessive-compulsive disorder

For OCD patients, all individual psychotherapeutic interventions were associated with improvement in CGAS ratings, but guidance to parents was less effective. Medication was not associated with improvement according to CGAS ratings. Given the apparent effectiveness of a range of psychotherapeutic interventions for alleviating functional impairment due to OCD, one can hypothesize that CGAS would show even larger changes if CAMHS followed the guidelines established in the POTS study that single out CBT combined with SSRI as the most effective available treatment for OCD (Pediatric OCD Treatment Study (POTS) 2004; Franklin et al. 2011).

7.5 THE PREDICTIVE VALUE OF CGAS RATINGS

In Paper IV, 4,876 child psychiatric patients in late adolescence were followed prospectively over 1 to 1½ years. CGAS ratings along with other clinical data from the database Pastill were linked with national Swedish registers. Hazard ratios were calculated for criminal conviction, psychiatric disorders, drug misuse, and accidents in early adulthood.

The results showed that CGAS ratings below 61 at end-of-treatment independently predicted criminal conviction, bipolar disorder, and borderline personality disorder in

early adulthood. This suggests that CGAS ratings contain prognostic information that can serve to guide clinical decision-making.

CGAS ratings are commonly performed at intake and at end-of-treatment regardless of the length of the treatment period. If the goal is to yield as valid long-term prognostic information as possible, should CGAS be rated at intake or at end-of-treatment? To explore this, the end-of-treatment results were adjusted for CGAS ratings at intake. The significantly increased risk of criminal conviction, bipolar disorder, and borderline personality disorder remained among those with CGAS ratings below 61 at end-of-treatment, suggesting that these ratings add prognostic information irrespective of CGAS ratings at intake. By contrast, further analyses found that CGAS ratings at intake were not in themselves predictive of later adversities.

This conclusion is in line with a previous study that found no relationship between CGAS ratings at hospital admission and at place of residence (home vs long-term inpatient care/institution) one year after hospital discharge (Sourander et al. 1996). Also, in a 30-year follow-up study of 541 child psychiatric inpatients, retrospective CGAS ratings at intake based on hospital records did not differ between later convicted and non-convicted adults (Mordre et al. 2011).

A possible explanation for the lack of prognostic information in CGAS ratings at intake is that they reflect temporary psychosocial functional impairment caused by the symptoms that prompt patients to seek help. By contrast, CGAS ratings at end-of-treatment are performed when the acute impairment presumably has subsided and thus more closely reflects the individual's usual level of function. This is presumably a more informative long-term prognostic marker. For example, those with the lowest mean CGAS rating at intake - suicide attempt – were also those also showed the largest mean value on CGAS improvement after end-of-treatment (Table 8). The CGAS score at intake for the suicidal group is low per definition and hence not representative of the child's level of psychosocial functioning when the risk for suicidal behaviour has subsided. Another explanation is that the clinician knows the patient less well at intake, which may increase bias and random error.

There are no previous longitudinal studies evaluating the long-term predictive properties of CGAS. However, other studies provide indirect support for the notion that CGAS contain prognostic information. Most longitudinal population-based studies (Rutter et al. 1976; Kim-Cohen et al. 2003; Olsson et al. 2006; Sourander et al. 2007; Pickles et al. 2010) as well as a longitudinal follow-up of child psychiatric patients (Engqvist and Rydelius 2007) have found that problems at school, with behaviour, and in relationships are correlated with poor outcomes in adulthood.

7.6 THE CGAS INSTRUMENT

The most striking observation made when analysing the rating distributions for each case vignette was that some CGAS intervals were rarely used, for example the interval 21-30 (Figure 1). One reason could be that the description of clinical examples in this interval was brief and the raters had difficulty recognising their case with this sparse information. Also, some intervals contain examples that reflect different levels of care intensity. Child psychiatric care has changed since 1983 when the scale was developed.

Inpatient resources have decreased and new outpatient units have developed new methods for child psychiatric patients and their families. Some examples of care levels therefore appear out of date. There were also suicidal cases in the training vignettes that created a lot of questions from the participants. This is a problem area where the scale has some weaknesses. In the interval 31-40 one of the clinical examples is "suicidal attempts with clear lethal intent". These cases are usually in need of considerable supervision, which is rated lower than 21.

7.7 METHODOLOGICAL CONSIDERATIONS

There are several methodological facets of these studies to consider. An important first strength of this thesis is the large number of research subjects. Paper I and II are based on 703 health care professionals, which is considerably more than previous studies. Paper III evaluated 12,613 child psychiatric patients. Finally, in Paper IV 4,876 child psychiatric patients were followed during 1-1½ years. Second, the randomisation procedure used in the CGAS training program is unique. This makes the findings more conclusive than those in previous studies of training's effectiveness. Also, the inclusion of the comparison group, albeit non-randomised, allowed a more rigorous evaluation of the overall training effects than previous studies. Third, Papers III and IV benefitted from the high resolution available in the clinical database Pastill, which comprises a large number of clinical variables, including axis I, IV, and V, as well as several treatment interventions and ratings of treatment response. Moreover, the high rate of registrations in the database Pastill means that almost all child psychiatric patients in Stockholm County were captured, making the results highly generalisable. Fourth, the linkage to national registers minimized patients lost to follow-up in Paper IV.

With respect to limitations, it might be argued that although written case vignettes (Paper I and II) have been shown to be a valid tool for assessing the quality of clinical work, judging written case vignettes is different from assessing real patients. For one thing, vignettes provide limited information.(Peabody et al. 2000; Peabody et al. 2004). For another, there is a risk that reliability based on vignettes inflates the ICC by minimizing information variance that would occur if two raters independently interviewed the patient.

Second, for practical reasons the number of rated cases in Paper I and II was restricted to five, which somewhat limits the statistical power. When choosing the number of vignettes it was necessary to balance how much time the research project could engage the participants and how many cases it was possible to rate without loss of concentration and motivation, on the one hand, with the number of cases needed to get enough statistical power on the other. With the available time frame it was deemed that 5 cases balanced these needs.

Third, Paper I and II examined actual CGAS ratings but did not assess health care professionals' skills in interviewing patients and their families. This is of potential importance since we have within CAMHS little systematic knowledge about how much a clinician's professional background, clinical experience, and personality affect the quality of the interview and the assessment.

Fourth, the comparison group in Paper II was not randomised and was smaller than the intervention groups. This study can therefore not be equated with an RCT. Still, the inclusion of a comparison group at all was new in terms of training studies and highlighted the important question of unspecific training effects. I have not found other training studies with randomised control groups and I look forward to conducting new training studies comparable to ordinary RCTs.

Fifth, to check overall training effect there should have been a rating directly after the training. This was not done. Instead the follow-up in Paper II occurred twelve months after the training. A recently published study suggests that the follow-up period should not be longer than three months (Cross et al. 2011). Rater drift is a known phenomenon, and in Paper II it could also have influenced the results negatively (Muller and Szegedi 2002; Yavorsky et al. 2010).

Sixth, despite the high rate of registration in Pastill there are missing data with respect to clinical and outcome variables. Treatment interventions evaluated in Paper III are broadly defined in Pastill without information about the specific psychotherapeutic method used. Also, medication information was restricted to whether the drug was a central stimulant or other than a central stimulant, and this broad level of information limits the interpretation of the results.

Seventh, formal studies of the validity of most Pastill diagnoses are lacking. However, the autism diagnoses have been evaluated and the results showed that 92% of the diagnoses in Pastill could be confirmed through an independent retrospective evaluation of medical charts (Selma Idrizbegovic, personal communication 2011). The diagnostic information in Paper IV on childhood bipolar disorder is included in the group "mood disorders" and not further specified in Pastill. Therefore, it is not possible to examine if bipolar disorder was diagnosed already during childhood.

Eight, the registering of CGAS ratings in Pastill was initiated 1 July 2006, which resulted in a limited follow-up time period of one to one-and-a-half years in Paper IV. This resulted in a relatively low number of outcome events in some rare cases (for example schizophrenia).

Ninth, the CGAS rating at case closure in Paper III and IV was made by a large group of clinicians that had access to the first CGAS rating. This might have biased the rater. However, the differences in improvement between mental retardation (lowest improvement) and suicide attempt or mood disorder (highest improvement) in Paper III are in line with clinicians' general perceptions of how the effectiveness of the different treatment options differs between the diagnostic groups. This suggests that the clinicians do not rate cases as improved by sheer routine. The validity of the CGAS ratings is also supported by the significant correlation between ΔCGAS and the overall assessment of treatment response made by the clinician at case closure also described in Paper III.

8 CONCLUSIONS

Paper I shows that the inter-rater reliability is moderate when CGAS is used in a large heterogeneous clinical setting with no prior training. This suggests that CGAS is a useful instrument, but that one should be cautious when comparing CGAS ratings of the same patient from different practitioners. The untrained raters differed substantially from the experts and tended to rate patients significantly higher, that is less functionally impaired. One cannot emphasize too much how important it is to look at the lowest level of functioning in the past month. Failure to do this is one factor that may create low inter-rater reliability using CGAS. The differences in ratings between professional groups raise questions about how and by whom CGAS ratings should be performed to be most accurate. Altogether, this stresses the importance of proper training in conjunction with the introduction of new rating scales.

In Paper II, there were no differences between the two forms of training, which speaks in favour of using the less resource-demanding CD. However, even though CGAS ratings improve with training, the effect was surprisingly small and unlikely to be clinically relevant. Intriguingly, there was a similarly positive effect in the non-randomised comparison group that received no training. These findings call into question the usefulness of this type of brief training programme. The findings also suggest that future education trials should include regular, randomly assigned control groups to control for the unspecific improvement that might occur.

The naturalistic effectiveness study in Paper III, showed that mean CGAS ratings improved after child psychiatric treatment-as-usual on the same order of magnitude as in clinical trials, but the level of improvement differed significantly depending on diagnosis. Interestingly, and at odds with results from clinical trials, medication in mood disorder treatment and central stimulants in ADHD treatment were not positively associated with improvement of CGAS ratings. By contrast, several of the different psychotherapeutic interventions were positively correlated with ΔCGAS in mood disorders. These results raise questions as to the effectiveness of medication in naturalistic child psychiatric settings. Further studies are warranted to evaluate the effectiveness of medication and psychotherapy in regular clinical settings, especially for ADHD and mood disorders.

The results in Paper IV suggest that CGAS ratings at end-of-treatment - but not at intake - provide specific information about the long-term outcome of individuals that have been subject to child psychiatric care. Particular attention regarding additional interventions or intensified follow-up might be warranted for adolescents with an end-of-treatment CGAS score of 60 or less.

9 IMPLICATIONS FOR THE FUTURE

9.1 TRAINING

It is impossible for me *not* to think that a new training study with a randomised control group and follow-up directly after training, and again after three and six months would add valuable information to CAMHS using CGAS.

I also see a potential to improve the training programme after my experience of conducting more than 70 seminars and training more than 1,000 clinicians. I would primarily emphasise that the ratings should reflect the lowest level of functioning over the past month, and emphasize the patient's impairment. Training might work best if tailored to particular professional groups.

9.2 INDIVIDUAL VS GROUP CGAS RATINGS

Clinician-rated tools like CGAS are used to operationalize the clinical evaluation. However, the process of decision-making is undisclosed; notes and charts seldom indicate how the information was collected, interpreted, and merged into the chosen index. In daily clinical practice, most CGAS ratings are done individually. It would be interesting to study how the decision-making process differs when groups do the ratings. There is a possibility that ratings done together with other team members would result in more reliable and valid results. Group decision-making is recommended in many areas (Surowiecki 2004; Wu et al. 2007), and it has been shown that groups often perform intellectual tasks better than the sum of the individuals in a group (Woolley et al. 2010).

9.3 CURRENT CGAS, REVISED CGAS OR A NEW SCALE

Despite any drawbacks with the scale, there are many advantages to keeping CGAS as is. First, it has already been heavily studied and evaluated. Second, it would allow on-going comparisons between study populations, since CGAS is and has been widely used, including in several national registers in the UK and Sweden.

An alternative is to keep CGAS but revise it in ways that would increase its reliability, validity, and clarity for clinicians. The clinical examples, for instance, need to be more thoroughly described in some of the intervals so the content is more equally distributed. Some levels of care specified in the scale are out-dated and should also be revised. Moreover, the different examples concerning suicidality need to be clarified and the degree of supervision recommended for different levels needs to be made clearer.

Another alternative is to create an entirely new rating tool for global assessment of children's psychosocial functioning. This would be demanding, but if it were done, it should be based from the start on children's levels of functioning and not be a version of an adult tool merely adapted to children. Ideally a new scale would also remedy some of the holes in CGAS, most importantly by including functioning around the basic activities of daily living (ADL) including hygiene, getting dressed, food intake,

and sleep habits, which are often impaired in children and adolescents with psychiatric disorders.

9.4 TRANSFER TO ADULT PSYCHIATRY

The increased risk of child psychiatric patients developing adult psychiatric disorders also highlights the importance of bridging the gap between child and adult psychiatry – an endeavour with well-known challenges. In the UK, for example, less than 25% of services have specific arrangements for transfer of care from child to adult psychiatry (Audit Commission 1999; Singh et al. 2005). This disrupted continuity of treatment is likely to contribute to future psychiatric illness among patients (Singh 2009). With CGAS ratings as a basis of information about the long-term health of individuals, the transition to adult psychiatry could be individualized and better planned, with more attention paid to cases with higher risk for adult psychiatric disorders. Clinical evaluations supported by rating tools may foster a better organization for the transfer of young vulnerable adolescents to adult psychiatric care.

10 EPILOGUE

Does child psychiatric treatment make a difference? This thesis suggests that child psychiatric care can be effective. Moreover, my studies indicate that CGAS can be used to help measure psychosocial functioning but that it is not a perfect instrument. Like all rating scales, CGAS has advantages and disadvantages.

I was disappointed that training did not improve the quality of ratings to make CGAS an even more valuable and reliable tool, but I have discussed several possible explanations for why training did not have the expected effects.

The most surprising results for me were the predictive properties of CGAS. That CGAS scores had such high predictive value indicates that it is a valid tool and that the scale can be highly useful for clinicians in their efforts to provide safe and effective care to children with mental illness. This finding also confirms my belief that it is important to combine clinical evaluation with rating tools and diagnostic instruments, and not to rely solely on one method. But my findings also confirm the criticisms levelled at using CGAS as the sole indicator of treatment needs or treatment outcomes. For example, performing CGAS via a telephone interview with a parent, as sometimes now occurs in Sweden, seems ill-advised. Misuse of CGAS ratings may, furthermore, increase the risk that professionals lose confidence in the instrument. Evaluation and treatment of patients must ultimately rest on clinical judgement that is based on an overall assessment of the patient's diagnosis, psychosocial functioning, level of distress, and the effect on family and network.

There is a genuine interest among CAMHS' management and health administrators to evaluate treatment continuously together with clinicians and researchers, and guarantee high quality care throughout the region and the country. In order to do this, my findings point to the necessity of more research that evaluates the effectiveness of child psychiatric care in the real world and not just within RCTs.

11 ACKNOWLEDGEMENTS

This thesis would not have been possible without the ambitious and highly skilled clinicians within CAMHS who create a caring space where children and their families can get help with psychiatric problems.

I owe a debt of gratitude to many people who have encouraged and supported me.

First of all I want to express my deepest appreciation to Mikael Landén, my main supervisor, for his never-ending optimistic belief in my CGAS project even though the project's ratings sometimes dipped quite low; for his ability to brilliantly handle my manuscripts and help me turn them into scientific papers; and for teaching me that research is joyful. I also want to express my deepest appreciation to Carl Johan Sundberg, my supervisor, for his tremendous generosity in sharing knowledge, experiences and contacts all over the world; for his inspiring way of supervising me and never hesitating to support me when I took on my PhD studies. My deepest appreciation also goes to Mai-Lis Hellenius, my mentor, for encouraging me and sharing valuable experiences from the challenging world of research.

Thanks also to Mats Forsman, for explaining complicated analyses in ways everyone can understand, and for clever solutions to obstacles with the clinical database Pastill and the linkage to national registers; to Clara Gumpert, for her inspiration to all child psychiatrists interested in research-based clinical work and for her pioneering work with the Research School for Clinicians in psychiatry, a crucial step in my PhD program; to Jan Kowalski, the most patient and cheerful of statisticians, even when faced with the same question again and again, and promoter of the unlikely idea that "statistics is a piece of cake"; to Paul Lichtenstein, for sharing creative ideas and being so truly enthusiastic about research whose main goal is to help children with psychiatric symptoms; and to Niklas Långström, for sharing his scientific expertise on register-based studies, and his great ideas for improving a manuscript.

Several people have helped me with administrative, technical and illustrative tasks of all kinds. Thanks to Lars Berglund, for excellent service and proofreading of my manuscript; to Maria Eriksson, for doing things before I have thought of them and for making my task as training director manageable; to Rolf Sjöberg, for taking care of all illustrations in any format I have come up with; to Ulf Lapidus and Yvonne Tesstor, for scheduling training seminars, and registering data with such care; to Olle Nilsson, for illustrating my fictive case vignettes with tenderness; and to the late Anders Bolin, for the successful programming of the CGAS training on CD-ROM.

I also thank Olav Bengtsson, head of CAMHS, for his solid support to the continued development of research-based child psychiatric care available to all children and families in the region; Eva-Britt Hallquist and Peter Engelsöy, former and present head of the in-patient unit, for their early initiative in implementing CGAS, and for giving me the confidence to develop the training program; Eva Serlachius, for her ability to bring the right people together and to get projects off the ground; and Olle Lindevall, for his solid management of Pastill.

Thanks also to supporting colleagues Åsa Borg; Eva Henje Blom; and to expert CGAS raters Sofia Bidö, Åsa Lundberg, Kerstin Malmberg, and Göran Parment for their

enthusiasm in creating a best estimate for the gold standard ratings used in my studies; and to Hans Hildebrand for valuable comments on my manuscript.

Thanks also to David Shaffer, Prudence Fisher, Madelyn Gould, Yanling Huo and Blake Turner at Columbia University, for support of the project and challenging questions.

I also want to say thank you to Becky Popenoe, for being an excellent research coach with brilliant ideas and questionings of my findings, and also for being a wonderful friend with humour, warmth and a great capacity to turn low ratings into top scores; Cecilia Halvorsen, for rich discussions about PhD student struggles and how to cope with refused manuscripts, and for sharing her warm and positive trust in life; Eleonore Rydén, for so many enjoyable lessons in statistics, exciting discussions about research - and about life; Louise Frisén, for touching talks on how to combine scientific thinking with life experiences that are beyond what we can explain; Anna Wennberg, for fruitful discussions about how to keep on a path of life-long learning as a medical doctor and a human being; and Louise Angenfelt, for long and supportive friendship.

Thanks to my book club, which recently celebrated its 20-year anniversary, with Agneta Zickert, Anna Gerber Ekblom, Hélena Syk, Kari Örtengren, and Ulrika Berggren. Thanks for always being there and reminding me that there is literature not published in PubMed.

Many thanks also to Eva and Erik Ståhl, my extra parents and wonderful friends, for their generosity and inclusion of me in their lives and in their home for years and years, and to their far-flung family.

My warmest thanks to Eva and Olof, my sister and brother, for being so patient with me, so available and so ready to give me a hand when I need one; and to Johanna, for the brave and inspiring choices she has made in life; and to John, Viktor and Lott for patiently accepting that living with Eva and Olof includes living with me.

Lastly, thanks to Buster and Ragnhild, for all the love and joy they share with me, and for their patient waiting for this book to be ready.

Financial support was provided through the Capio Research Foundation, the Celsing Foundation, the Fredrik and Ingrid Thuring Foundation, the Jerring Foundation, Lars Hierta Memorial Foundation, the 1.6 million Club for Women's health, the Swedish Psychiatric Foundation (Psykiatrifonden), the regional agreement on medical training and clinical research between the Stockholm County Council and Karolinska Institutet, CAMHS in Stockholm County Council, the Centre for Psychiatric Research and Education/Department of Clinical Neuroscience Karolinska Institutet, which is hereby thankfully acknowledged.

12 REFERENCES

Abeles, P., C. Verduyn, A. Robinson, P. Smith, W. Yule and J. Proudfoot (2009). "Computerized CBT for adolescent depression ("Stressbusters") and its initial evaluation through an extended case series." *Behavioural and Cognitive Psychotherapy* 37(2): 151-165.

American Psychiatric Association (1987). Diagnostic and statistical manual of mental disorders : DSM-III-R. Washington, DC, American Psychiatric Association.

American Psychiatric Association (2000). Diagnostic and statistical manual of mental disorders : DSM-IV-TR.

American Psychiatric Association. (1980). Quick reference to the diagnostic criteria from DSM-III. Washington, American Psychiatric Assn.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders : DSM-III-R. Washington, DC, American Psychiatric Association.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders : DSM-IV. Washington, DC, American Psychiatric Association.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders : DSM-IV-TR. Washington, DC, American Psychiatric Association.

American Psychiatric Association. (2010). "DSM-5 Development." Retrieved 20 February 2012, from http://www.dsm5.org.

Anastasi, A. and S. Urbina (1997). Psychological testing. Upper Saddle River, N.J., Prentice Hall.

Angold, A., E. J. Costello and A. Erkanli (1999). "Comorbidity." *Journal of child psychology and psychiatry, and allied disciplines* 40(1): 57-87.

Audit Commission. (1999). "Children in Mind: Audit Commission." from http://www.audit-commission.gov.uk/nationalstudies/health/mentalhealth/Pages/childreninmind.aspx.

Ayton, A., C. Keen and B. Lask (2009). "Pros and cons of using the Mental Health Act for severe eating disorders in adolescents." *Eur Eat Disord Rev* 17(1): 14-23.

Barzman, D. H., M. P. DelBello, R. A. Kowatch, B. Gernert, D. E. Fleck, S. Pathak, K. Rappaport, S. V. Delgado, P. Campbell and S. M. Strakowski (2004). "The effectiveness and tolerability of aripiprazole for pediatric bipolar disorders: a retrospective chart review." *Journal of Child and Adolescent Psychopharmacology* 14(4): 593-600.

Bates, L. W., J. A. Lyons and J. B. Shaw (2002). "Effects of brief training on application of the Global Assessment of Functioning Scale." *Psychological Reports* 91(3 Pt 1): 999-1006.

Bella, T., T. Goldstein, D. Axelson, M. Obreja, K. Monk, M. B. Hickey, B. Goldstein, D. Brent, R. S. Diler, D. Kupfer, D. Sakolsky and B. Birmaher (2011). "Psychosocial functioning in offspring of parents with bipolar disorder." *Journal of Affective Disorders* 133(1-2): 204-211.

Bennett, K. J. and D. R. Offord (2001). "Screening for conduct problems: does the predictive accuracy of conduct disorder symptoms improve with age?" *Journal of the American Academy of Child and Adolescent Psychiatry* 40(12): 1418-1425.

Berek, M., A. Kordon, L. Hargarter, F. Mattejat, L. Slawik, K. Rettig and B. Schauble (2011). "Improved functionality, health related quality of life and decreased burden of disease in patients with ADHD treated with OROS(R) MPH: is treatment response different between children and adolescents?" *Child and Adolescent Psychiatry and Mental Health* 5: 26.

Bero, L. A., R. Grilli, J. M. Grimshaw, E. Harvey, A. D. Oxman and M. A. Thomson (1998). "Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group." *BMJ (Clinical Research Ed.)* 317(7156): 465-468.

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

Biggs, J. B. (2003). Teaching for quality learning at university : what the student does. London, the Society for Research into Higher Education : Open University Press.

Bird, H. R., G. Canino, M. Rubiostipec and J. C. Ribera (1987). "Further Measures Of The Psychometric Properties Of The Childrens Global Assessment Scale." *Archives of General Psychiatry* 44(9): 821-824.

Bird, H. R., M. S. Gould and B. M. Staghezza (1993). "Patterns of diagnostic comorbidity in a community sample of children aged 9 through 16 years." *Journal of the American Academy of Child and Adolescent Psychiatry* 32(2): 361-368.

Bird, H. R., T. J. Yager, B. Staghezza, M. S. Gould, G. Canino and M. Rubio-Stipec (1990). "Impairment in the epidemiological measurement of childhood psychopathology in the community." *Journal of the American Academy of Child and Adolescent Psychiatry* 29(5): 796-803.

Bland, J. M. and J. D. Altman (1996). "Measurement error and correlation coefficients." *British Medical Journal* 313: 41-42.

Brent, D. A., D. Holder, D. Kolko, B. Birmaher, M. Baugher, C. Roth, S. Iyengar and B. A. Johnson (1997). "A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy." *Archives of General Psychiatry* 54(9): 877-885.

Bridge, J. A., B. Birmaher, S. Iyengar, R. P. Barbe and D. A. Brent (2009). "Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder." *American Journal of Psychiatry* 166(1): 42-49.

BUP divisionen (2010). Riktlinjer till stöd för bedömning och behandling [Guidelines to support evaluation and treatment]. CAMHS Stockholm.

Busch, A. B. and L. I. Sederer (2000). "Assessing outcomes in psychiatric practice: guidelines, challenges, and solutions." *Harvard Review of Psychiatry* 8(6): 323-327.

Canino, G., P. E. Shrout, M. Rubio-Stipec, H. R. Bird, M. Bravo, R. Ramirez, L. Chavez, M. Alegria, J. J. Bauermeister, A. Hohmann, J. Ribera, P. Garcia and A. Martinez-Taboas (2004). "The DSM-IV rates of child and adolescent disorders in Puerto Rico: prevalence, correlates, service use, and the effects of impairment." *Archives of General Psychiatry* 61(1): 85-93.

Castro-Fornieles, J., I. Baeza, E. de la Serna, A. Gonzalez-Pinto, M. Parellada, M. Graell, D. Moreno, S. Otero and C. Arango (2011). "Two-year diagnostic stability in early-onset first-episode psychosis." *Journal of child psychology and psychiatry, and allied disciplines* 52(10): 1089-1098.

Collett, B. R., J. L. Ohan and K. M. Myers (2003). "Ten-year review of rating scales. V: scales assessing attention-deficit/hyperactivity disorder." *Journal of the American Academy of Child and Adolescent Psychiatry* 42(9): 1015-1037.

Collett, B. R., J. L. Ohan and K. M. Myers (2003). "Ten-year review of rating scales. VI: scales assessing externalizing behaviors." *Journal of the American Academy of Child and Adolescent Psychiatry* 42(10): 1143-1170.

Colman, I. and P. B. Jones (2004). "Birth cohort studies in psychiatry: beginning at the beginning." *Psychological Medicine* 34(8): 1375-1383.

Cross, W. F., D. Seaburn, D. Gibbs, K. Schmeelk-Cone, A. M. White and E. D. Caine (2011). "Does practice make perfect? A randomized control trial of behavioral rehearsal on suicide prevention gatekeeper skills." *The journal of primary prevention* 32(3-4): 195-211.

Cummings, C. M. and M. A. Fristad (2011). "Anxiety in Children with Mood Disorders: A Treatment Help or Hindrance?" *Journal of Abnormal Child Psychology*.

David, C. N., D. Greenstein, L. Clasen, P. Gochman, R. Miller, J. W. Tossell, A. A. Mattai, N. Gogtay and J. L. Rapoport (2011). "Childhood onset schizophrenia: high rate of visual hallucinations." *Journal of the American Academy of Child and Adolescent Psychiatry* 50(7): 681-686 e683.

Davis, D. A., M. A. Thomson, A. D. Oxman and R. B. Haynes (1995). "Changing physician performance. A systematic review of the effect of continuing medical education strategies." *JAMA* 274(9): 700-705.

Dyrborg, J., F. W. Larsen, S. Nielsen, J. Byman, B. B. Nielsen and F. Gautre-Delay (2000). "The Children's Global Assessment Scale (CGAS) and Global Assessment of Psychosocial Disability (GAPD) in clinical practice--substance and reliability as judged by intraclass correlations." *European Child and Adolescent Psychiatry* 9(3): 195-201.

Ekholm, B., A. Ekholm, R. Adolfsson, M. Vares, U. Osby, G. C. Sedvall and E. G. Jonsson (2005). "Evaluation of diagnostic procedures in Swedish patients with schizophrenia and related psychoses." *Nordic Journal of Psychiatry* 59(6): 457-464.

Ellwood, P. M. (1988). "Shattuck lecture--outcomes management. A technology of patient experience." *New England Journal of Medicine* 318(23): 1549-1556.

Emslie, G. J., A. J. Rush, W. A. Weinberg, R. A. Kowatch, C. W. Hughes, T. Carmody and J. Rintelmann (1997). "A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression." *Archives of General Psychiatry* 54(11): 1031-1037.

Endicott, J., R. L. Spitzer, J. L. Fleiss and J. Cohen (1976). "The global assessment scale. A procedure for measuring overall severity of psychiatric disturbance." *Archives of General Psychiatry* 33(6): 766-771.

Engqvist, U. and P. A. Rydelius (2006). "Death and suicide among former child and adolescent psychiatric patients." *BMC Psychiatry* 6: 51.

Engqvist, U. and P. A. Rydelius (2007). "Child and adolescent psychiatric patients and later criminality." *BMC Public Health* 7: 221.

Fergusson, D. M., L. J. Horwood and E. M. Ridder (2005). "Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood." *Journal of child psychology and psychiatry, and allied disciplines* 46(8): 837-849.

Fernando, T., G. Mellsop, K. Nelson, K. Peace and J. Wilson (1986). "The reliability of axis V of DSM-III." *American Journal of Psychiatry* 143(6): 752-755.

Findling, R. L., N. K. McNamara, E. A. Youngstrom, L. A. Branicky, C. A. Demeter and S. C. Schulz (2003). "A prospective, open-label trial of olanzapine in adolescents with schizophrenia." *Journal of the American Academy of Child and Adolescent Psychiatry* 42(2): 170-175.

Findling, R. L., E. J. Short, T. Leskovec, L. D. Townsend, C. A. Demeter, N. K. McNamara and R. J. Stansbrey (2008). "Aripiprazole in children with attention-deficit/hyperactivity disorder." *Journal of Child and Adolescent Psychopharmacology* 18(4): 347-354.

Fitzpatrick, C., N. Nwanolue-Abayomi, A. Kehoe, N. Devlin, S. Glackin, L. Power and S. Guerin (2011). "Do we miss depressive disorders and suicidal behaviours in clinical practice?" *Clinical child psychology and psychiatry*.

Follan, M., S. Anderson, S. Huline-Dickens, E. Lidstone, D. Young, G. Brown and H. Minnis (2011). "Discrimination between attention deficit hyperactivity disorder and reactive attachment disorder in school aged children." *Research in Developmental Disabilities* 32(2): 520-526.

Fonagy, P. (1997). "Evaluating the effectiveness of interventions in child psychiatry." *Canadian Journal of Psychiatry* 42(6): 584-594.

Forsman, M., P. Lichtenstein, H. Andershed and H. Larsson (2010). "A longitudinal twin study of the direction of effects between psychopathic personality and antisocial behaviour." *Journal of child psychology and psychiatry, and allied disciplines* 51(1): 39-47.

Franklin, M. E., J. Sapyta, J. B. Freeman, M. Khanna, S. Compton, D. Almirall, P. Moore, M. Choate-Summers, A. Garcia, A. L. Edson, E. B. Foa and J. S. March (2011). "Cognitive behavior therapy augmentation of pharmacotherapy in pediatric obsessive-compulsive disorder: the Pediatric OCD Treatment Study II (POTS II) randomized controlled trial." *JAMA* 306(11): 1224-1232.

Galanter, C. A. and V. L. Patel (2005). "Medical decision making: a selective review for child psychiatrists and psychologists." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 46(7): 675-689.

Garland, A. F., L. Bickman and B. F. Chorpita (2010). "Change what? Identifying quality improvement targets by investigating usual mental health care." *Administration and Policy in Mental Health* 37(1-2): 15-26.

Garralda, M. E., P. Yates and I. Higginson (2000). "Child and adolescent mental health service use. HoNOSCA as an outcome measure." *The British journal of psychiatry : the journal of mental science* 177: 52-58.

Geller, B., R. Tillman, K. Bolhofner and B. Zimerman (2008). "Child bipolar I disorder: prospective continuity with adult bipolar I disorder; characteristics of second and third episodes; predictors of 8-year outcome." *Archives of General Psychiatry* 65(10): 1125-1133.

Geller, B., B. Zimerman, M. Williams, K. Bolhofner and J. L. Craney (2001). "Adult psychosocial outcome of prepubertal major depressive disorder." *Journal of the American Academy of Child and Adolescent Psychiatry* 40(6): 673-677.

Gilbody, S. M., A. O. House and T. A. Sheldon (2002). "Outcomes research in mental health. Systematic review." *British Journal of Psychiatry* 181: 8-16.

Gold, J., R. J. Buonopane, R. A. Caggiano, M. Picciotto, C. Vogeli, N. T. Kanner, R. L. Babcock, Z. Li, M. Jellinek, S. J. Drubner, K. J. Sklar and J. M. Murphy (2009). "Assessing outcomes in child psychiatry." *American Journal of Managed Care* 15(4): 210-216.

Goldman, H. H., A. E. Skodol and T. R. Lave (1992). "Revising axis V for DSM-IV: a review of measures of social functioning." *American Journal of Psychiatry* 149(9): 1148-1156.

Gorman, D. A., N. Thompson, K. J. Plessen, M. M. Robertson, J. F. Leckman and B. S. Peterson (2010). "Psychosocial outcome and psychiatric comorbidity in older adolescents with Tourette syndrome: controlled study." *British Journal of Psychiatry* 197(1): 36-44.

Green, B., S. Shirk, D. Hanze and J. Wanstrath (1994). "The Children's Global Assessment Scale in clinical practice: an empirical evaluation." *Journal of the American Academy of Child and Adolescent Psychiatry* 33(8): 1158-1164.

Hanssen-Bauer, K., O. O. Aalen, T. Ruud and S. Heyerdahl (2007). "Inter-rater reliability of clinician-rated outcome measures in child and adolescent mental health services." *Administration and Policy in Mental Health* 34(6): 504-512.

Hanssen-Bauer, K., S. Gowers, O. O. Aalen, N. Bilenberg, P. Brann, E. Garralda, S. Merry and S. Heyerdahl (2007). "Cross-national reliability of clinician-rated outcome measures in child and adolescent mental health services." *Administration and Policy in Mental Health* 34(6): 513-518.

Henggeler, S. W., S. K. Schoenwald and S. G. Pickrel (1995). "Multisystemic therapy: bridging the gap between university- and community-based treatment." *Journal of Consulting and Clinical Psychology* 63(5): 709-717.

Hilsenroth, M. J., S. J. Ackerman, M. D. Blagys, B. D. Baumann, M. R. Baity, S. R. Smith, J. L. Price, C. L. Smith, T. L. Heindselman, M. K. Mount and D. J. Holdwick, Jr. (2000). "Reliability and validity of DSM-IV axis V." *American Journal of Psychiatry* 157(11): 1858-1863.

Hoagwood, K., E. Hibbs, D. Brent and P. Jensen (1995). "Introduction to the special section: efficacy and effectiveness in studies of child and adolescent psychotherapy." *Journal of Consulting and Clinical Psychology* 63(5): 683-687.

Hoagwood, K., P. S. Jensen, T. Petti and B. J. Burns (1996). "Outcomes of mental health care for children and adolescents: I. A comprehensive conceptual model." *Journal of the American Academy of Child and Adolescent Psychiatry* 35(8): 1055-1063.

Hodges, K. (1993). "Structured interviews for assessing children." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 34(1): 49-68.

Jensen, P. S., K. Hoagwood and T. Petti (1996). "Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model." *Journal of the American Academy of Child and Adolescent Psychiatry* 35(8): 1064-1077.

Keller, M. B., N. D. Ryan, M. Strober, R. G. Klein, S. P. Kutcher, B. Birmaher, O. R. Hagino, H. Koplewicz, G. A. Carlson, G. N. Clarke, G. J. Emslie, D. Feinberg, B. Geller, V. Kusumakar, G. Papatheodorou, W. H. Sack, M. Sweeney, K. D. Wagner, E. B. Weller, N. C. Winters, R. Oakes and J. P. McCafferty (2001). "Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial." *Journal of the American Academy of Child and Adolescent Psychiatry* 40(7): 762-772.

Kendall, P. C. and M. A. Southam-Gerow (1995). "Issues in the transportability of treatment: the case of anxiety disorders in youths." *Journal of Consulting and Clinical Psychology* 63(5): 702-708.

Kendall, T., S. Pilling, C. Whittington, C. Pettinari and R. Burbeck (2005). "Clinical Guidelines in Mental Health II: a guide to making NICE Guidelines." *Psychiatric Bulletin* 29: 3-8.

Kendler, K. S., C. O. Gardner, P. Annas, M. C. Neale, L. J. Eaves and P. Lichtenstein (2008). "A longitudinal twin study of fears from middle childhood to early adulthood: evidence for a developmentally dynamic genome." *Archives of General Psychiatry* 65(4): 421-429.

Kim-Cohen, J., A. Caspi, T. E. Moffitt, H. Harrington, B. J. Milne and R. Poulton (2003). "Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort." *Archives of General Psychiatry* 60(7): 709-717.

Kirkpatrick, D. L. and J. D. Kirkpatrick (2005). Evaluating training programs : the four levels. San Francisco, CA, Berrett-Koehler.

Kobak, K. A., N. Engelhardt and J. D. Lipsitz (2006). "Enriched rater training using Internet based technologies: A comparison to traditional rater training in a multi-site depression trial." *Journal of Psychiatric Research* 40(3): 192-199.

Kobak, K. A., J. D. Lipsitz and A. Feiger (2003). "Development of a standardized training program for the Hamilton Depression Scale using internet-based technologies: results from a pilot study." *Journal of Psychiatric Research* 37(6): 509-515.

Kobak, K. A., M. G. A. Opler and N. Engelhardt (2007). "PANSS rater training using Internet and videoconference: Results from a pilot study." *Schizophrenia Research* 92(1-3): 63-67.

Kratochvil, C. J., B. S. Vaughan, M. L. Mayfield-Jorgensen, J. S. March, S. H. Kollins, D. W. Murray, H. Ravi, L. L. Greenhill, L. A. Kotler, N. Paykina, P. Biggins and J. Stoner (2007). "A pilot study of atomoxetine in young children with attention-deficit/hyperactivity disorder." *Journal of Child and Adolescent Psychopharmacology* 17(2): 175-185.

Lambert, M. J., A. E. Bergin and S. L. Garfield (2004). Bergin and Garfield's handbook of psychotherapy and behavior change. New York, Wiley.

Leaf, P. J., M. Alegria, P. Cohen, S. H. Goodman, S. M. Horwitz, C. W. Hoven, W. E. Narrow, M. Vaden-Kiernan and D. A. Regier (1996). "Mental health service use in the community and schools: results from the four-community MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study." *Journal of the American Academy of Child and Adolescent Psychiatry* 35(7): 889-897.

Lewinsohn, P. M., P. Rohde, J. R. Seeley, D. N. Klein and I. H. Gotlib (2000). "Natural course of adolescent major depressive disorder in a community sample: predictors of recurrence in young adults." *The American journal of psychiatry* 157(10): 1584-1591.

Luborsky, L. (1962). "Clinician's judgments of mental health." *Archives of General Psychiatry* 7: 407-417.

Luborsky, L. and H. Bachrach (1974). "Factors influencing clinician's judgments of mental health. Eighteen experiences with the Health-Sickness Rating Scale." *Archives of General Psychiatry* 31(3): 292-299.

Ludvigsson, J. F., E. Andersson, A. Ekbom, M. Feychting, J. L. Kim, C. Reuterwall, M. Heurgren and P. O. Olausson (2011). "External review and validation of the Swedish national inpatient register." *BMC Public Health* 11: 450.

Lundh, A., J. Kowalski, C. J. Sundberg, C. Gumpert and M. Landén (2010). "Children's Global Assessment Scale (CGAS) in a naturalistic clinical setting: Inter-rater reliability and comparison with expert ratings." *Psychiatry Research* 177(1-2): 206-210.

Lundh, A., J. Kowalski, C. J. Sundberg and M. Landén (2011). "A Comparison of Seminar and Computer Based Training on the Accuracy and Reliability of Raters Using the Children's Global Assessment Scale (CGAS)." *Administration and Policy in Mental Health*.

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

March, J., S. Silva and B. Vitiello (2006). "The Treatment for Adolescents with Depression Study (TADS): methods and message at 12 weeks." *Journal of the American Academy of Child and Adolescent Psychiatry* 45(12): 1393-1403.

Masi, G., A. Milone, G. Canepa, S. Millepiedi, M. Mucci and F. Muratori (2006). "Olanzapine treatment in adolescents with severe conduct disorder." *European Psychiatry* 21(1): 51-57.

Mazade, N. A. and R. W. Glover (2007). "State mental health policy: critical priorities confronting state mental health agencies." *Psychiatric Services* 58(9): 1148-1150.

McShane, G., C. Bazzano, G. Walter and G. Barton (2007). "Outcome of patients attending a specialist educational and mental health service for social anxiety disorders." *Clinical child psychology and psychiatry* 12(1): 117-124.

Miller, T. J., T. H. McGlashan, J. L. Rosen, K. Cadenhead, T. Cannon, J. Ventura, W. McFarlane, D. O. Perkins, G. D. Pearlson and S. W. Woods (2003). "Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability." *Schizophrenia Bulletin* 29(4): 703-715.

Milne, J. M., C. Z. Garrison, C. L. Addy, R. E. McKeown, K. L. Jackson, S. P. Cuffe and J. L. Waller (1995). "Frequency of phobic disorder in a community sample of young adolescents." *Journal of the American Academy of Child and Adolescent Psychiatry* 34(9): 1202-1211.

Molina, B. S., S. P. Hinshaw, J. M. Swanson, L. E. Arnold, B. Vitiello, P. S. Jensen, J. N. Epstein, B. Hoza, L. Hechtman, H. B. Abikoff, G. R. Elliott, L. L. Greenhill, J. H. Newcorn, K. C. Wells, T. Wigal, R. D. Gibbons, K. Hur and P. R. Houck (2009). "The MTA at 8 years: prospective follow-up of children treated for combined-type ADHD in a multisite study." *Journal of the American Academy of Child and Adolescent Psychiatry* 48(5): 484-500.

Monga, S., A. Young and M. Owens (2009). "Evaluating a cognitive behavioral therapy group program for anxious five to seven year old children: a pilot study." *Depression and Anxiety* 26(3): 243-250.

Mordre, M., B. Groholt, E. Kjelsberg, B. Sandstad and A. M. Myhre (2011). "The impact of ADHD and conduct disorder in childhood on adult delinquency: a 30 years follow-up study using official crime records." *BMC Psychiatry* 11: 57.

MTA Cooperative Group (1999). "A 14-month randomized clinical trial of treatment strategies for attention-deficit/hyperactivity disorder. The MTA Cooperative Group. Multimodal Treatment Study of Children with ADHD." *Archives of General Psychiatry* 56(12): 1073-1086.

Mufson, L., K. P. Dorta, P. Wickramaratne, Y. Nomura, M. Olfson and M. M. Weissman (2004). "A randomized effectiveness trial of interpersonal psychotherapy for depressed adolescents." *Archives of General Psychiatry* 61(6): 577-584.

Mufson, L., M. M. Weissman, D. Moreau and R. Garfinkel (1999). "Efficacy of interpersonal psychotherapy for depressed adolescents." *Archives of General Psychiatry* 56(6): 573-579.

Muller, M. J. and A. Szegedi (2002). "Effects of interrater reliability of psychopathologic assessment on power and sample size calculations in clinical trials." *Journal of Clinical Psychopharmacology* 22(3): 318-325.

Muratori, F., L. Picchi, C. Casella, R. Tancredi, A. Milone and M. G. Patarnello (2002). "Efficacy of brief dynamic psychotherapy for children with emotional disorders." *Psychotherapy and Psychosomatics* 71(1): 28-38.

Myers, K. and N. C. Winters (2002). "Ten-year review of rating scales. I: overview of scale functioning, psychometric properties, and selection." *Journal of the American Academy of Child and Adolescent Psychiatry* 41(2): 114-122.

Myers, K. and N. C. Winters (2002). "Ten-year review of rating scales. II: Scales for internalizing disorders." *Journal of the American Academy of Child and Adolescent Psychiatry* 41(6): 634-659.

National Board of Health and Welfare. (2009). "Barn- och ungdomspsykiatrins metoder - en nationell inventering. [Methods in child and adolescent psychiatry - a national inventory]." from http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/8367/2009-126-146_2009126146.pdf.

National Board of Health and Wellfare. (2010). "Nationella riktlinjer för vård vid depression och ångestsyndrom [National guidelines for the care of Depression and Anxiety]." Retrieved 15 January 2012, from http://www.socialstyrelsen.se/publikationer2010/2010-3-4.

National Institute for Health and Clinical Excellence. (2008). "Attention deficit hyperactivity disorder: Diagnosis and management of ADHD in children, young people and adults." Retrieved 20 February 2012, from http://www.nice.org.uk/CG72.

Ohan, J. L., K. Myers and B. R. Collett (2002). "Ten-year review of rating scales. IV: scales assessing trauma and its effects." *Journal of the American Academy of Child and Adolescent Psychiatry* 41(12): 1401-1422.

Olsson, M., K. Hansson and M. Cederblad (2006). "A long-term follow-up of conduct disorder adolescents into adulthood." *Nordic Journal of Psychiatry* 60(6): 469-479.

Ougrin, D., T. Zundel, M. Kyriakopoulos, R. Banarsee, D. Stahl and E. Taylor (2011). "Adolescents with suicidal and nonsuicidal self-harm: Clinical characteristics and response to therapeutic assessment." *Psychological Assessment*.

Pavuluri, M. N., D. B. Henry, J. A. Carbray, M. W. Naylor and P. G. Janicak (2005). "Divalproex sodium for pediatric mixed mania: a 6-month prospective trial." *Bipolar Disorders* 7(3): 266-273.

Peabody, J. W., J. Luck, P. Glassman, T. R. Dresselhaus and M. Lee (2000). "Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality." *Journal of the American Medical Association* 283(13): 1715-1722.

Peabody, J. W., J. Luck, P. Glassman, S. Jain, J. Hansen, M. Spell and M. Lee (2004). "Measuring the quality of physician practice by using clinical vignettes: a prospective validation study." *Annals of Internal Medicine* 141(10): 771-780.

Pediatric OCD Treatment Study (POTS) (2004). "Cognitive-behavior therapy, sertraline, and their combination for children and adolescents with obsessive-compulsive disorder: the Pediatric OCD Treatment Study (POTS) randomized controlled trial." *JAMA* 292(16): 1969-1976.

Petersen, D. J., N. Bilenberg, K. Hoerder and C. Gillberg (2006). "The population prevalence of child psychiatric disorders in Danish 8- to 9-year-old children." *European Child and Adolescent Psychiatry* 15(2): 71-78.

Pickles, A., A. Aglan, S. Collishaw, J. Messer, M. Rutter and B. Maughan (2010). "Predictors of suicidality across the life span: the Isle of Wight study." *Psychological Medicine* 40(9): 1453-1466.

Preuss, U., S. J. Ralston, G. Baldursson, B. Falissard, M. J. Lorenzo, R. Rodrigues Pereira, L. Vlasveld and D. Coghill (2006). "Study design, baseline patient characteristics and intervention in a cross-cultural framework: results from the ADORE study." *European Child and Adolescent Psychiatry* 15 Suppl 1: I4-14.

Renou, S., T. Hergueta, M. Flament, M. C. Mouren-Simeoni and Y. Lecrubier (2004). "[Diagnostic structured interviews in child and adolescent's psychiatry]." *L'Encephale* 30(2): 122-134.

Rey, J. M., J. Starling, C. Wever, D. R. Dossetor and J. M. Plapp (1995). "Inter-rater reliability of global assessment of functioning in a clinical setting." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 36(5): 787-792.

Rosen, J., B. H. Mulsant, P. Marino, C. Groening, R. C. Young and D. Fox (2008). "Web-based training and interrater reliability testing for scoring the Hamilton Depression Rating Scale." *Psychiatry Research* 161(1): 126-130.

Rosenberg, D. R., C. M. Stewart, K. D. Fitzgerald, V. Tawile and E. Carroll (1999). "Paroxetine open-label treatment of pediatric outpatients with obsessive-compulsive disorder." *Journal of the American Academy of Child and Adolescent Psychiatry* 38(9): 1180-1185.

Rush, A. J., M. B. First, D. Blacker and American Psychiatric Association. Task Force for the Handbook of Psychiatric Measures. (2008). Handbook of psychiatric measures. Washington, DC, American Psychiatric Pub.

Rutter, M., J. Tizard, W. Yule, P. Graham and K. Whitmore (1976). "Research report: Isle of Wight Studies, 1964-1974." *Psychological Medicine* 6(2): 313-332.

Satterfield, J. H., K. J. Faller, F. M. Crinella, A. M. Schell, J. M. Swanson and L. D. Homer (2007). "A 30-year prospective follow-up study of hyperactive boys with conduct problems: adult criminality." *Journal of the American Academy of Child and Adolescent Psychiatry* 46(5): 601-610.

Schorre, B. E. and I. H. Vandvik (2004). "Global assessment of psychosocial functioning in child and adolescent psychiatry. A review of three unidimensional scales (CGAS, GAF, GAPD)." *European Child and Adolescent Psychiatry* 13(5): 273-286.

Sellgren, C., M. Landen, P. Lichtenstein, C. M. Hultman and N. Langstrom (2011). "Validity of bipolar disorder hospital discharge diagnoses: file review and multiple register linkage in Sweden." *Acta Psychiatrica Scandinavica* 124(6): 447-453.

Setoya, Y., K. Saito, M. Kasahara, K. Watanabe, M. Kodaira and M. Usami (2011). "Evaluating outcomes of the child and adolescent psychiatric unit: A prospective study." *International Journal of Mental Health Systems* 5: 7.

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

Shaffer, D., C. P. Lucas and J. E. Richters (1999). Diagnostic assessment in child and adolescent psychopathology. New York, Guilford Press.

Shrout, P. E. (1998). "Measurement reliability and agreement in psychiatry." *Statistical Methods in Medical Research* 7(3): 301-317.

Singh, S. P. (2009). "Transition of care from child to adult mental health services: the great divide." *Current Opinion in Psychiatry* 22(4): 386-390.

Singh, S. P., N. Evans, L. Sireling and H. Stuart (2005). "Mind the gap: the interface between child and adult mental health services." *Psychiatric Bulletin* 29: 292-294.

Sourander, A., P. Jensen, M. Davies, S. Niemela, H. Elonheimo, T. Ristkari, H. Helenius, L. Sillanmaki, J. Piha, K. Kumpulainen, T. Tamminen, I. Moilanen and F. Almqvist (2007). "Who is at greatest risk of adverse long-term outcomes? The Finnish From a Boy to a Man study." *Journal of the American Academy of Child and Adolescent Psychiatry* 46(9): 1148-1161.

Sourander, A., H. Leijala, A. Lehtila, A. Kanerva, H. Helenius and J. Piha (1996). "Short-term child psychiatric inpatient treatment. Place of residence as one-year outcome measure." *European Child and Adolescent Psychiatry* 5(1): 38-43.

Sourander, A., P. MultimAki, P. Santalahti, K. Parkkola, A. Haavisto, H. Helenius, G. Nikolakaros, J. Piha, T. Tamminen, I. Moilanen, K. Kumpulainen, E. T. Aronen, S. L. Linna, K. Puura and F. Almqvist (2004). "Mental health service use among 18-year-old adolescent boys: a prospective 10-year follow-up study." *Journal of the American Academy of Child and Adolescent Psychiatry* 43(10): 1250-1258.

Sporn, A. L., A. Vermani, D. K. Greenstein, A. J. Bobb, E. P. Spencer, L. S. Clasen, J. W. Tossell, C. C. Stayer, P. A. Gochman, M. C. Lenane, J. L. Rapoport and N. Gogtay (2007). "Clozapine treatment of childhood-onset schizophrenia: evaluation of effectiveness, adverse effects, and long-term outcome." *Journal of the American Academy of Child and Adolescent Psychiatry* 46(10): 1349-1356.

Stein, B. D., J. N. Kogan, S. L. Hutchison, E. A. Magee and M. J. Sorbero (2010). "Use of outcomes information in child mental health treatment: results from a pilot study." *Psychiatric Services* 61(12): 1211-1216.

Steinhausen, H. C. (1987). "Global assessment of child psychopathology." *Journal of the American Academy of Child and Adolescent Psychiatry* 26(2): 203-206.

Steinhausen, H. C. and C. W. Metzke (2001). "Global measures of impairment in children and adolescents: results from a Swiss community survey." *Australian and New Zealand Journal of Psychiatry* 35(3): 282-286.

Steinhausen, H. C., T. S. Novik, G. Baldursson, P. Curatolo, M. J. Lorenzo, R. Rodrigues Pereira, S. J. Ralston and A. Rothenberger (2006). "Co-existing psychiatric problems in ADHD in the ADORE cohort." *European Child and Adolescent Psychiatry* 15 Suppl 1: I25-29.

Stewart, M., M. P. DelBello, M. Versavel and D. Keller (2009). "Psychosocial functioning and health-related quality of life in children and adolescents treated with open-label ziprasidone for bipolar mania, schizophrenia, or schizoaffective disorder." *Journal of Child and Adolescent Psychopharmacology* 19(6): 635-640.

Surowiecki, J. (2004). The wisdom of crowds : why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. New York, Doubleday :.

Szigethy, E., E. Kenney, J. Carpenter, D. M. Hardy, D. Fairclough, A. Bousvaros, D. Keljo, J. Weisz, W. R. Beardslee, R. Noll and D. R. DeMaso (2007). "Cognitive-behavioral therapy for adolescents with inflammatory bowel disease and subsyndromal depression." *Journal of the American Academy of Child and Adolescent Psychiatry* 46(10): 1290-1298.

Szobot, C. M., C. Ketzer, M. A. Parente, J. Biederman and L. A. Rohde (2004). "The acute effect of methylphenidate in Brazilian male children and adolescents with ADHD: a randomized clinical trial." *Journal of Attention Disorders* 8(2): 37-43.

Söderberg, P., S. Tungstrom and B. A. Armelius (2005). "Reliability of global assessment of functioning ratings made by clinical psychiatric staff." *Psychiatric Services* 56(4): 434-438.

Thompson, L., J. Kemp, P. Wilson, R. Pritchett, H. Minnis, L. Toms-Whittle, C. Puckering, J. Law and C. Gillberg (2010). "What have birth cohort studies asked about genetic, pre- and perinatal exposures and child and adolescent onset mental health outcomes? A systematic review." *European Child and Adolescent Psychiatry* 19(1): 1-15.

Tillman, R. and B. Geller (2007). "Diagnostic characteristics of child bipolar I disorder: does the "Treatment of Early Age Mania (team)" sample generalize?" *Journal of Clinical Psychiatry* 68(2): 307-314.

Valderhaug, R. and T. Ivarsson (2005). "Functional impairment in clinical samples of Norwegian and Swedish children and adolescents with obsessive-compulsive disorder." *European Child and Adolescent Psychiatry* 14(3): 164-173.

Vitiello, B., P. Rohde, S. Silva, K. Wells, C. Casat, B. Waslick, A. Simons, M. Reinecke, E. Weller, C. Kratochvil, J. Walkup, S. Pathak, M. Robins and J. March (2006). "Functioning and quality of life in the Treatment for Adolescents with Depression Study (TADS)." *Journal of the American Academy of Child and Adolescent Psychiatry* 45(12): 1419-1426.

Wagner, K. D., P. Ambrosini, M. Rynn, C. Wohlberg, R. Yang, M. S. Greenbaum, A. Childress, C. Donnelly and D. Deas (2003). "Efficacy of sertraline in the treatment of children and adolescents with major depressive disorder: two randomized controlled trials." *JAMA : the journal of the American Medical Association* 290(8): 1033-1041.

Wagner, K. D., P. Ambrosini, M. Rynn, C. Wohlberg, R. Yang, M. S. Greenbaum, A. Childress, C. Donnelly and D. Deas (2003). "Efficacy of sertraline in the treatment of children and adolescents with major depressive disorder: two randomized controlled trials." *JAMA* 290(8): 1033-1041.

Wagner, K. D., J. Jonas, R. L. Findling, D. Ventura and K. Saikali (2006). "A double-blind, randomized, placebo-controlled trial of escitalopram in the treatment of pediatric depression." *Journal of the American Academy of Child and Adolescent Psychiatry* 45(3): 280-288.

Weissman, M. M., V. Warner and M. Fendrich (1990). "Applying impairment criteria to children's psychiatric diagnosis." *Journal of the American Academy of Child and Adolescent Psychiatry* 29(5): 789-795.

Weisz, J. R., G. R. Donenberg, S. S. Han and D. Kauneckis (1995). "Child and adolescent psychotherapy outcomes in experiments versus clinics: why the disparity?" *Journal of Abnormal Child Psychology* 23(1): 83-106.

Weisz, J. R., G. R. Donenberg, S. S. Han and B. Weiss (1995). "Bridging the gap between laboratory and clinic in child and adolescent psychotherapy." *Journal of Consulting and Clinical Psychology* 63(5): 688-701.

Wiggins, A., M. Oakley Browne, C. Bearsley-Smith and E. Villanueva (2010). "Depressive disorders among adolescents managed in a child and adolescent mental health service." *Australasian Psychiatry* 18(2): 134-141.

WIMHRT. (2009). "The Washington Institute for Mental Health Research & Training " Retrieved November 4, 2009, from http://depts.washington.edu/washinst/Resources/CGAS/GCAS SCENERIO INDEX.htm.

Winters, N. C., B. R. Collett and K. M. Myers (2005). "Ten-year review of rating scales, VII: scales assessing functional impairment." *Journal of the American Academy of Child and Adolescent Psychiatry* 44(4): 309-342.

Winters, N. C., K. Myers and L. Proud (2002). "Ten-year review of rating scales. III: scales assessing suicidality, cognitive style, and self-esteem." *Journal of the American Academy of Child and Adolescent Psychiatry* 41(10): 1150-1181.

Wolpert, M., L. Cooper, K. Tingay, K. Young and E. Svanberg. (2007). "CAMHS Outcomes Research Consortium Handbook " Version 2.0 (2007). Retrieved February 2012, from http://www.corc.uk.net/media/File/CORC Resources/Handbook/HANDBOOK 2007.pdf.

Wolpert, M., P. Fuggle, D. Cottrell, P. Fonagy, J. Philips, S. Pilling, S. Stein and M. Target (2006). Drawing on the Evidence. Advice for mental health professionals working with children and adolescents. London, CAMHS Publications.

Wolpert, M., P. Fuggle, D. Cottrell, P. Fonagy, J. Philips, S. Pilling, S. Stein and M. Target. (2011). "Choosing what's best for you." 1st. Retrieved 20 February 2012, from http://www.choosing.org.uk/.

Woolley, A. W., C. F. Chabris, A. Pentland, N. Hashmi and T. W. Malone (2010). "Evidence for a collective intelligence factor in the performance of human groups." *Science* 330(6004): 686-688.

Wu, S. M., U. Whiteside and C. Neighbors (2007). "Differences in inter-rater reliability and accuracy for a treatment adherence scale." *Cognitive Behaviour Therapy* 36(4): 230-239.

Yavorsky, C., M. Opler, E. Ivanova, J. Gordon, S. Jovic and L. Yang (2010). Quantifying Rater Drift in an International Sample of Investigators Participating in Standardized Rater Training Events:

Is PANSS Reliability Maintained Over Time? ISCTM 6th Annual Scientific Meeting. Washington D.C.: 10-11.

Zanarini, M. C., A. E. Skodol, D. Bender, R. Dolan, C. Sanislow, E. Schaefer, L. C. Morey, C. M. Grilo, M. T. Shea, T. H. McGlashan and J. G. Gunderson (2000). "The Collaborative Longitudinal Personality Disorders Study: reliability of axis I and II diagnoses." *J Pers Disord* 14(4): 291-299.