Measures, interventions, and outcomes: exploring inpatient psychiatric care

Ove Sonesson

Department of Psychology

Sweden 2014

UNIVERSITY OF GOTHENBURG
DEPT OF PSYCHOLOGY

_____

ABSTRACT

Sonesson, O. (2014). *Measures, interventions, and outcomes: exploring inpatient psychiatric care.* Department of Psychology, University of Gothenburg, Sweden.

The general aim of this thesis was to investigate interventions and outcomes in psychiatric inpatient care through the use of assessment scales and database information. Another aim was to contribute to the knowledge of the Global Assessment of Functioning (GAF) scale in regard to reliability, validity and as a measure of the outcome of treatment.

Data in Study I were gathered from assessment sessions concerning the reliability of the GAF, and data in the following three studies were collected from the ELVIS healthcare information system used within Sahlgrenska University Hospital.

The reliability of the GAF scale was investigated in Study I using the GAF-ratings of six vignettes by 101 participants from an inpatient psychiatric clinic. The results demonstrated good reliability with an intra-class coefficient of 0.79. Background variables such as the number of years of experience in using the GAF and attitudes towards the GAF were entered into multiple linear regression analyses showing no statistically significant effect.

Study II investigated the outcome of inpatient psychiatric care in which the GAF was used as a measure of outcome. The sample consisted of 816 care episodes that were GAF-rated both at admission and at discharge. The difference between the patient's GAF value at discharge and admission was used as a measure of improvement in the global level of functioning. The overall GAF change was 20.7 points and represented a shift from a low to a moderate level of

functioning. The effect size measure of Cohen's *d* showed an overall effect size of 1.67, corresponding to a high effect. Within the diagnostic categories, substance-related disorders showed the lowest effect size (1.03) and other mood disorders showed the highest (2.33). Of all of the patients in the study group, 75% had a GAF change ≥ 10 points and were considered improved.

Study III investigated the influence of clinical and socio-demographic factors on psychosocial functioning as measured by the GAF scale. Statistically significant predictors of GAF scores at admission were age, schizophrenia, other psychotic disorders, and no registered diagnosis. GAF scores at admission, most diagnoses, and being a patient at a specific ward were able to significantly predict the GAF scores at discharge. It was also found that specialised wards did not necessarily deliver the highest treatment results in spite of their diagnostic specialisation.

Study IV focused on interventions in inpatient psychiatric care as described by the Swedish Classification of Health Interventions (KVÅ). A KVÅ-code list elaborated within Region Västra Götaland was used, which consisted of 76 specific codes for psychiatric interventions. Staff at the wards registered these codes when specific interventions were performed. At least one KVÅ code was registered in 83% of all episodes of care, and five codes covered 50% of all registrations. Patients with a diagnosis of schizophrenia showed the highest share of coordinating interventions, and patients with a diagnosis within substance-related disorders showed the lowest share of psychological treatments. Medical technical and coordinating interventions were related to psychosocial functioning at discharge. It was concluded that with adequate registration of the quantity and quality of interventions, the KVÅ classification system could have the potential to describe the interventions used in inpatient psychiatric care. The four studies in this dissertation support the conclusion that a central database system could be useful to investigate interventions and outcomes in psychiatric inpatient care.

Ove Sonesson, Department of Psychology, University of Gothenburg, Box 500, SE-405 30 Göteborg, Sweden. Phone +46707367179.

E-mail: ove.sonesson@gu.se

SAMMANFATTNING PÅ SVENSKA

Det övergripande syftet med denna avhandling var att undersöka interventioner och behandlingsresultat inom psykiatrisk heldygnsvård genom att använda bedömningsinstrument och information från en central databas. Ett annat syfte var att bidra med ytterligare kunskap om Global Assessment of Functioning (GAF) skalan när det gäller reliabilitet, validitet och som ett mått för att mäta behandlingsresultat.

Data i studie I hämtades från bedömningssessioner där GAF skalans reliabilitet undersöktes och data för de tre följande studierna hämtades från det patientadministrativa systemet ELVIS, som används inom Sahlgrenska Universitetssjukhus. GAF skalans reliabilitet (mätsäkerhet) undersöktes genom att vårdpersonal från sex psykiatriska avdelningar för heldygnsvård fick i uppgift att skatta sex olika patientfall. Tre av fallen presenterades i text och tre presenterades genom video. Det visade sig att reliabiliteten i skattningarna var god, med ett mätvärde på 0.79 (Intra Class Coefficient). Inga av de studerade bakgrundsfaktorerna såsom antal år med erfarenhet av GAF skattningar och attityd till GAF skalan, uppvisade något statistiskt säkerställt samband med reliabiliteten.

I studie II undersöktes behandlingsresultatet inom psykiatrisk heldygnsvård. GAF skalan användes som ett resultatmått. Data från 816 vårdtillfällen användes där patienternas globala funktionsnivå hade skattats med hjälp av GAF skalan vid både in- och utskrivning. Skillnaden mellan patientens GAF värde vid inskrivning och utskrivning användes som ett mått på behandlingseffekt. Den genomsnittliga förändringen blev 20.7 poäng, vilket också kan uttryckas som en förändring från en låg funktionsnivå till en moderat nivå. Cohen's d nådde ett övergripande värde på 1.67, vilket motsvarar en hög effektstorlek. Inom diagnosgrupperna, uppvisade substansrelaterade diagnoser den lägsta effektstorleken (1.03) och gruppen andra förstämningssyndrom den högsta

(2.33). Det framkom vidare att 75 % av patienterna hade en GAF förändring på ≥ 10 poäng och de bedömdes som förbättrade.

I studie III undersöktes inflytandet från några kliniska och socio-demografiska faktorer på den psykosociala funktionsnivån, mätt med GAF skalan. Statistiskt säkerställda prediktorer för GAF-värde vid inskrivning var ålder, schizofreni, andra psykotiska störningar och ingen registrerad diagnos. GAF-värde vid inskrivning, flertalet diagnosgrupper, och att vara patient på en specifik avdelning var statistiskt säkerställda prediktorer av GAF-värde vid utskrivning. Det visade sig också att de avdelningar som specialiserat sig på vissa diagnoser, inte nödvändigtvis var de avdelningar som hade högst behandlingsresultat för de specifika diagnosgrupperna.

Studie IV fokuserade på interventioner utförda inom psykiatrisk heldygnsvård och som registrerats utifrån Klassifikation av Vårdåtgärder (KVÅ). En KVÅ-kodlista som utvecklats inom Västra Götalandsregionen med 76 specifika koder för psykiatriska insatser användes. Avdelningspersonalen registrerade dessa koder när specifika insatser hade utförts. Vid 83 % av alla vårdtillfällen fanns det minst en KVÅ kod registrerad och fem koder täckte 50 % av alla registreringar. Patienter med diagnosen schizofreni uppvisade den högsta andelen av samordnande insatser och patienter med en substansrelaterad diagnos hade den lägsta andelen av psykologiska behandlingsinsatser. Medicintekniska och samordnande insatser hade samband med psykosocial funktionsnivå vid utskrivning. En slutsats som drogs var att KVÅ har potential för att kunna vara ett hjälpmedel att beskriva vad som utförs inom psykiatrisk heldygnsvård.

De fyra studierna i den här avhandlingen ger stöd för antagandet att en central databas kan vara användbar för att undersöka interventioner och behandlingsresultat inom psykiatrisk heldygnsvård.

ACKNOWLEDGEMENTS

I am deeply grateful to many people and circumstances that have contributed to the fulfilment of this thesis, and some of them are mentioned here.

Thank you Eva Osvald-Gustafsson, my former operations manager within the psychiatric clinic at Östra Hospital, for your courage and enthusiasm. You laid the groundwork for this thesis.

Thank you Professor Tomas Tjus, my supervisor, for guiding me with profound scientific knowledge and Wermlandian confidence, for an excellent communication capacity, for letting me walk my own paths when necessary and waiting for me when I returned, and for constituting a secure base for exploration and scientific growth.

Thank you Associate professor Hans Arvidsson, my assistant supervisor, for always focusing on finding new sustainable scientific elements in order to build solid research. With truly Gothenburg humour, you have illustrated the importance of thinking before and after writing.

Thank you late Professor emeritus Ingvar Lundberg for your gentle way of making me understand complex issues and making me feel good. Your commitment and scientific integrity will always twinkle in the sky.

Thank you Professor emerita Annika Dahlgren-Sandberg, my examiner, for integrating me into the former Health, Handicap and Aging section (HHÅ) and for encouraging and guiding me with scientific sharpness and human warmth.

Thank you to all of the members in the research group, Research in Clinical Psychology, for all of your sharing and daring, and for creating hope and curiosity.

Thank you to all of the members of the former HHÅ section for travels with joy and enthusiasm.

Thank you, University of Gothenburg, for your travel grants to conferences in Cape Town, Honolulu and Madrid, stimulating scientific exchange with other researchers and extending my knowledge. You widened my scientific scope, strengthened my identity as a researcher and encouraged me to continue.

Thank you, Forte, for your travel grants to conferences in Buenos Aires and San Sebastian - Donostia, stimulating scientific exchange with other researchers and extending my knowledge. You widened my scientific scope, strengthened my identity as a researcher and encouraged me to continue.

Dear Lord, thank you for your wind, your sea, your beaches, your mountains, your trees, your sun, your moon, your stars - filling me with energy, persistence and confidence. Thank you for your guidance.

LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals:

I. Sonesson, O., Tjus, T., & Arvidsson, H. (2010). Reliability of a functioning scale (GAF) among psychiatric ward staff. *Nordic Psychology, 62,* 53-64.

II. Sonesson, O., Arvidsson, H., & Tjus, T. (2013). Effectiveness of psychiatric inpatient care. *Scandinavian Journal of Caring Sciences, 27,* 319-326.

III. Sonesson, O., Arvidsson, H., & Tjus, T. (2014). Exploring outcome and validity of the GAF in psychiatric inpatient care. *European Journal of Psychological Assessment*. Advance online publication. doi:10.1027/1015-5759/a000225.

IV. Sonesson, O., Arvidsson, H., & Tjus, T. (2014). Interventions in psychiatric inpatient care as described by the Swedish Classification of Health Interventions (KVÅ). Submitted manuscript.

CONTENT

INTRODUCTION

Inpatient psychiatric care is a part of mental health care services and plays an important role for many patients with psychiatric disorders in severe phases of mental illness (Glick, Carter, & Tandon, 2003). An inpatient treatment episode can be seen as a phase in a continuum of care in which outpatient psychiatric treatment is another important phase (Glick et al., 2003). This means that inpatient treatment can only contribute to a limited change in the patient's problems and that the patient at discharge usually is in need of outpatient care. Glick et al. (2003) argue that there is a tendency within inpatient treatment that the personnel unsuccessfully try to do everything for the patient instead of limiting the interventions to specific, formulated problems and to the limited time period. The overall theme in inpatient psychiatric treatment is the patient's need for crisis stabilisation, where crisis in relation to a psychiatric disorder can be defined as the threat of suicide or homicide, harmful acts to self or others, and impaired self-care (Sharfstein, 2009). According to Sharfstein (2009), there are specific activities that should be performed within the inpatient treatment episode: defining the focal problem; diagnostic assessment; formulating specific goals for hospitalisation; determining and performing treatments; working with the patient's family and other support systems; coordinating care with outpatient providers and establishing an outpatient treatment plan. Important functions for the inpatient psychiatric unit are also to keep the patient safe, to provide psychoeducation and to establish a therapeutic alliance (Glick, Sharfstein, & Schwartz, 2011).

Hopkins, Loeb and Fick (2009) performed a literature review focusing on what service users expect from inpatient mental health care. They found, among other results, that service users expect treatment in a safe environment and development of relationships with staff. These interpersonal relations could

involve one-on-one counselling, educational sessions and informal communication. The service users valued good staff communication skills, which led to feelings of being understood and respected.

In this thesis, inpatient psychiatric care is explored concerning used measures, types of registered interventions and outcomes. The development of psychiatric care from institutionalisation to deinstitutionalisation and a section on psychiatric nosology will be presented. Research concerning judgments and decisions in clinical practice and a section about quality development will follow. The area of outcomes research and the psychometric concepts of reliability and validity will be described. A paragraph concerning the classification of health interventions will follow. An overview of functioning and functioning scales will take place, followed by a paragraph focusing on the Global Assessment of Functioning Scale.

There will be a summary of the four studies concerning the reliability of the GAF scale, the use of the GAF scale as a measure of outcome, the influence of some socio-demographic and clinical factors in relation to psychosocial functioning, and interventions in psychiatric inpatient care as described through the Swedish Classification of Health Interventions (KVÅ). Finally, there will be a discussion related to reliability, validity, classification of health interventions, outcomes, clinical judgment and decision making, strengths and limitations, and a conclusion.

Institutionalisation and deinstitutionalisation of psychiatric care

Confinement of the poor, unemployed, criminals and the insane into institutions (asylums) was established during the eighteenth century all across Europe and continued to develop during the nineteenth century (Foucault, 1988). There are different approaches for explaining this development. One approach is related to

the progress of medical treatment, humanism and to the increase in the incidence of insanity (Grob, 1991; Hare, 1983). Hare (1983) argued that a considerable proportion of persons with a diagnosis today named schizophrenia caused increased admissions to the asylums, and this could be related to a changing social and cultural environment. Another approach states that the emergence of asylums was a response to growing social problems in western society (Focault, 1988). According to Foucault (1988), confinement was a solution to an economic crisis in the Western world where unemployment was widespread. To work was a moral requirement of both the government and church, and citizens not working were seen as idle. The asylums absorbed the idle and formed a social protection against uprisings. Mentally insane persons were often associated with animality, and it was not uncommon that they were chained. During the late eighteenth century, a process started that separated the mentally ill from the criminals. The asylums for the insane during the nineteenth century were, according to Foucault (1988), not based on a science of mental disease but on authority in which the physician was connected to juridical and moral domains.

The view of Focault can be contrasted by the achievements of Philippe Pinel, a French physician serving at the hospitals of Bicêtre and Salpêtrière in Paris from the late eighteenth century to about the middle of the nineteenth century (Weiner, 1992). Pinel could be said to be the founder of psychiatry in France and he advocated for "traitment moral", which can be translated as a psychologically oriented treatment in which it was important to interview the patient and to make careful observations of the patient in order to make a psychiatric diagnosis, it was also a concern to place patients with similar impairments in the same units. There was an aim for individually adapted treatment and to establish a personal relationship between the staff and the patient. Privileges were used as patient incentives and coercive actions were

used only under specific circumstances (Weiner, 1992; Stephanie, 2013). Pinel emphasised that it was important to understand the natural history of the patients' disturbances, including precipitating events, and he postulated a potential for recovery. He was also engaged in diagnostic classification and noted two types of madness: a continuous or chronic category and an intermittent category (Weiner, 1992). Stephanie (2013) concludes that Pinel has influenced modern psychiatry including the psychiatric diagnostic systems such as the Diagnostic and Statistical Manual of Mental Disorders (DSM).

There was a great expansion of asylums and psychiatric patients in the USA and Europe until around the middle of the twentieth century (with variations among the countries) followed by a process of deinstitutionalisation (Bülow, 2004). One definition of deinstitutionalisation is provided by Ramon (1996), who states that it concerns getting care outside of hospital settings and obtaining community support for people with severe mental illness (Ramon, 1996). The closing or downsizing of mental hospitals required access to mental health services outside the hospital. Community-based services for the care of severely mentally ill persons developed and were an important part of the psychiatric reform process (Arvidsson, 2004). Possible factors that influenced the process of deinstitutionalisation included medical factors, with the introduction of neuroleptics; economic factors, or an impending or actual fiscal crisis (asylums were expensive to run and required repair); and psychiatric practice factors, such as a change from the physical process of the brain to the psyche and social and familial networks (Prior, 1991).

The process of deinstitutionalisation has spread all over Europe, though there are differences within and between countries (Becker & Vázquez-Barquero, 2001). In Sweden, mental health care reform was established in 1995 and aimed to improve social integration and the quality of life for persons with long-standing and serious mental disorders (Regeringens proposition, 1993). Different actions

were formulated to create more efficient and coordinated community-based services for these persons. The reform implied a reduction of inpatient psychiatric care.

In Sweden, there were approximately 35 000 psychiatric beds in the late 1960s, and in 2010, there were 4514 beds. The amount of hospital days was reduced from six million in 1987 to approximately 1.6 million in 2008 (Socialstyrelsen, 2003; Sveriges Kommuner och Landsting 2010). In the past, inpatient psychiatric units were often located in the outskirts of a city and based on long-term treatment; but today they are often acute units located within general hospital areas (Curtis, Gesler, Priebe, & Francis, 2009). During the last 50 years, there have been considerable changes in psychiatric inpatient care concerning structure and content. However, there is a lack of knowledge of how the efforts in inpatient care should be performed to be effective, and research is needed in this area (Sveriges Kommuner och Landsting, 2010).

Psychiatric nosology

As mentioned in the section on institutionalisation and deinstitutionalisation, during the eighteenth and nineteenth centuries, there was an increase in admissions to asylums, and according to Hare (1983), a great part of admissions could be ascribed to persons with a diagnosis of what is today named schizophrenia. Emil Kraepelin was a German psychiatrist who, over a century ago, contributed to the classification of mental illness by organising functional psychotic disorders into the categories of dementia praecox, manic-depressive illness and paranoia (Decker, 2007). He was a clinician as well as a researcher and was devoted to empirical research as a major way to acquire medical scientific knowledge. To understand the mental illness of the patient and to make a diagnosis, Kraepelin considered it very important to obtain information

from different sources (Decker, 2007). Kraepelin was oriented towards the biological aspects of mental illness and regarded the medical and somatic areas as a starting point for psychiatric research (Jablensky, 2007). He postulated the possibility of degeneration as a result of mental illness (Engstrom, 2007; Kraepelin, 2007). However, he also recognised environmental influences on the course of mental illness, such as the movement of people from rural areas to large cities (Kraepelin, 2007).

In 1911, the Swiss psychiatrist Eugen Bleuler published "Dementia praecox or the group of schizophrenias" (Carpenter, 2011). His classification was a comprehensive development from the classification of Kraepelin. Bleuler argued that the Kraepelin construct of dementia praecox was misleading because the symptoms of these patients did not necessarily arise in adolescence and were not necessarily characterised by severe memory deficits.(McGlashan, 2011). Bleuler introduced the concept of schizophrenia centred on distorted and disorganised mental functions, followed by different subcategories (Carpenter, 2011; McGlashan, 2011). Bleuler assumed a neural basis of schizophrenia but he was not oriented towards neuro-scientific research but rather concerned about psychological processes to obtain knowledge about the disorder (Heckers, 2007).

The works by Bleuler and particularly Kraepelin have contributed to the development of the International Classification of Diseases and Related Health Problems (ICD) governed by the World Health Organisation and to the Diagnostic and Statistical Manual of Mental Disorders (DSM) published by the American Psychiatric Association (Compton & Guze, 1995). The DSM-I and DSM-II encompassed a theory-oriented and an environmental and psychological approach. Beginning with the DSM-III, there was a radical change and the diagnostic categories were data-oriented and defined by operationalised

descriptive criteria in accordance with the diagnostic approach of Kraepelin, an approach that has also been used in the ICD-10, though to a somewhat lesser degree (Compton et al. 1995; Decker, 2007). Spitzer and Sheehy (1976) argued that the operationalised criteria of the DSM-III would strengthen the reliability and validity of the diagnoses. Both the ICD and the DSM are currently based on a medical model in which the scientific study of the relationship of specific brain structures and brain processes to functional mental impairment is of primary concern, which is in line with the Kraepelinian model of psychiatry (Compton et al., 1995, Jablensky, 2007).

Clinical judgment and decision making

In psychiatry, there are many various judgments and decisions to be made by staff in everyday practice. Knowledge of the strengths and limitations of these cognitive processes is important (Crumlish & Kelly, 2009). Research within cognitive psychology has made important contributions in this area, and in the last decade, there has been an accompanying clinical interest (Crumlish et al., 2009). To better understand the strengths and biases in clinical judgment and decision making, paragraphs on pseudoscientific strategies, conscious and unconscious processing, intuition, and heuristics will follow.

Pseudoscientific strategies

A lot of research has centred on the accuracy of clinicians' diagnostic efforts (Kim, 2002; Miller, Dasher, Collins, Griffiths, & Brown, 2001). Kim (2002) found that clinicians use a theory-based strategy when diagnosing patients according to the DSM-IV system. It turned out that clinicians used personally constructed symptom-based theories when deciding on a diagnosis. The theories consisted of peripheral and central symptoms, and assumed that the central

symptoms were causally related to each other. The a-theoretic and criteria-oriented manual of the DSM-IV was not applied. Miller et al. (2001) also studied the effect of interview structuring on the diagnostic assessment of patients in inpatient care. It was discovered that the structured interview surpassed the unstructured interview in diagnostic accuracy.

Garb and Boyle (2003) have presented results from research on the use of scientific and pseudoscientific methods. They proposed that in many cases concerning clinical judgment, experienced clinicians have not performed better than less experienced clinicians and clinicians have seldom been more accurate than graduate students. Garb and Boyle attribute these findings particularly to the clinicians' use of pseudoscientific methods but also to the difficulties in getting valid feedback from clinical experiences and to heuristics and other biases.

As mentioned above, the DSM system has evolved through the years from a more subjective and theory-based approach to an empirically and criteria-based approach (Miller et al. 2001; Broberg, Almqvist & Tjus, 2003). The Global Assessment of Functioning (GAF) scale in the DSM-IV was an important measure in the four studies in this thesis and encompasses psychological, social, and occupational functioning on a hypothetical continuum of mental health – mental illness, and constitutes a global measure of psychosocial functioning, with a range from 1 to 100 points (American Psychiatric Association, 2000). The GAF can be seen as a semi-structured and standardised measure.

Conscious and unconscious processing

Human information can be processed at conscious and non-conscious levels (Wilson, 2002). According to Wilson, there are conscious and non-conscious

types of thinking, feeling and motivation. Wilson (2002) terms the non-conscious operating as the "adaptive unconscious".

The adaptive unconscious is seen as a fundamental and necessary resource in most aspects of human life. Consciousness alone is too limited in its information processing capacity. Because non-conscious processing is not reached by consciousness, it is hard to obtain direct knowledge of the details in the unconscious processing. Wilson (2002) states that introspection is of limited use in acquiring information from the adaptive unconscious. However, the results of non-conscious operating can, to some degree, become known at the conscious level. As humans, we consciously construct reason and meaning for decisions and actions that we believe are true, when in fact we might not really know the causation chain. In this way, we can sometimes incorrectly experience the performance of an act as arising from our thoughts and our conscious willing (Wegner & Wheatley, 1999). According to Wilson (2002), it is difficult to know the right answer in regard to decisions. It is possible to make a list of pros and cons and exclusively decide from that. Wilson argues that too much conscious effort might disturb the holistic adaptive unconscious processing and result in an inferior decision. He recommends the use of gut feelings as a decision guide. To strengthen the processing of the adaptive unconscious and the accuracy of the gut feeling, it is necessary to first gather a foundation of reliable information.

Intuition

Klein (2004) also focuses on unconscious processing under the name of intuition. He defines intuition as "the way we translate our experiences into judgments and decisions" (Klein, 2004, p. 23). Intuition is considered a natural consequence of experience and is essential in judgment and decision making. The intuitive effort is made quickly and unconsciously. Klein argues that intuitive processing in most cases is superior to deliberate analytical processing.

The classical model of decision making, encompassing a lot of gathered information, with options and conscious evaluation, is, according to Klein, a logical model that is not very useful in regard to practical use. However, intuition is not always a reliable skill and therefore needs to be developed. One way to develop intuition is to obtain more purposive experience. It is also possible to strengthen intuitive power through specially arranged exercises. In some steps in these exercises, participants make use of deliberate analytical thinking. In Klein's (2004) Recognition-Primed Decision Model, there is a mental simulation loop in which conscious and deliberate information processing play a part. Klein does not exclude the analytical process from successful judgments and decisions but gives it a balancing function in relation to intuition.

Heuristics

Heuristics are rules of thumb concerning judgment and decision making, primarily processed on an unconscious level (Gigerenzer, Brighton, 2008). According to Kahneman (2011), there is an association between heuristics and intuition. Some intuitions are based on skill and expertise stemming from repeated experience with appropriate feedback for validation. Other intuitions are based on heuristics.

One branch of research in the heuristic area is the heuristics and biases approach (Tversky & Kahneman, 1974). The biases perspective is related to the biases that sometimes follow the use of heuristics. To understand the workings of heuristics, it is useful to have a model of cognitive processing. Kahneman (2011) advocates a dual-process theory composed of System 1 and System 2. System 1 is characterised by fast, automatic, associative, and effortless processing. This system neglects ambiguity and supports clear expressions about

causes and intentions. Heuristics emanate from System 1. The processing of System 2 is slow, reflective and effortful. System 2 has the capacity to monitor and control the results of System 1 processing, which gives a person a sense of being in charge and knowing the reason for his or her decisions and actions. According to Kahneman (2011), System 2 has restricted attentional capacity and sometimes does not prioritise controlling the judgments and decisions of System 1. According to the heuristics and biases approach, this means that the biases produced by System 1 will not be corrected. Furthermore, System 2 does not always have the capacity to analyse and correct the biases produced by heuristics. A clarification of the heuristic and biases approach is to claim that there is an original target question that is substituted by a heuristic question, which is easier to answer (Kahneman & Frederick, 2002).

In the beginning of the formulation of the heuristics and biases approach, the anchoring heuristic was formulated (Tversky et al., 1974). Anchoring is a term that connotes the use of a standpoint or initial value from which the judgment or decision starts. The result of the judgment or decision will be influenced in a biased way by this starting point. In an experiment Tversky et al. (1974) performed using a spinning wheel of fortune for producing starting values, participants' judgments concerning a following question of frequency estimation were greatly influenced by the number where the spinning wheel stopped. In clinical diagnostic praxis, this could correspond to a situation of prominent information presented by the patient in the beginning of an interview making a disproportionate influence on the assessment in regard to the following information. The clinician tends, in this case, to hold on too much to a perspective developed early (Croskerry, 2003).

Another heuristic is the availability heuristic (Kahneman, 2011). The judgment or decision of a person is dependent on the ease with which information is

retrieved from memory. Information that easily comes to mind will make up the answer. There are various reasons for what becomes available; it could, for example, be recently occurred incidents that in some way can be related to the current issue. This availability approach is in contrast to the view that judgment begins when a lot of adequate information has been collected and thoroughly reflected on (Waddington & Morley, 2000). Tversky and Kahneman (1974) mention an issue of assessing the risk of heart attack among middle-aged persons, where respondents remember such incidences among familiar persons and answer in relation to that memory. An example from clinical praxis could be when a specific intervention is suggested for a patient because the same intervention was recently chosen for another patient (Crumlish et al., 2009).

When using representative heuristic, the judging person utilises the dimension of similarity when the target attribute is composed of probability (Kahneman, 2003). An example from Kahneman et al., 2002, p. 55: "Are more deaths caused by rattlesnakes or bees? The respondents might make up an impression of the "dangerousness" of the typical snake or bee, an application of representativeness." The heuristic answer is rattlesnakes, based on an associative similarity between rattlesnakes and danger, without considering the frequency of the object, which is related to the frequencies of deaths. Making judgments and decisions through the use of stereotypes is also an expression of the representative heuristic. This form of processing information can, for example, be seen in the process of assigning a psychiatric diagnosis to the psychiatric symptoms of a person (Cantor, deSales French, Smith & Mezzich, 1980; Garb, 2005). The clinician matches the patient's symptoms with the clinician's mental prototype of the diagnostic category, and thereby confirms or rejects the diagnosis. This pattern-recognition approach could result in missing atypical variants of a diagnosis (Croskerry, 2003).

The last heuristic to be presented within the biases approach is the affect heuristic (Slovic, Finucane, Peters, & MacGregor, 2002) in which persons are strongly influenced and guided by their emotions when making their decisions or judgments. According to Slovic et al. (2002), affects embrace the dimensions of goodness or badness and are experienced as an unconscious or conscious feeling. The affect heuristic has its origin from the previously mentioned System 1 and is related to the affective charge of objects and incidents. There is an interplay between affect and cognition. Affect can act both directly on judgments and as an associated reaction to a made decision. The affect heuristic can also be associated to other heuristics, and in that way can be perceived as a validation of the performed judgment or decision. The use of emotions in decision making could be a strength as well as a disadvantage (Garb, 2013). Some research on clinical judgment has shown overconfidence associated to the influence of emotions (Garb, 2013). A clinical example when the affect heuristic is in use could be a clinician's positive feelings towards a patient leading to a more benign diagnosis than would be justified by the gathered information (Crumlish et al., 2009).

One heuristic mentioned by Gigerenzer et al. (2011) is the tallying heuristic. This heuristic favours frequencies of elements related to an issue but ignores the strength of each of them. An example could be the assessment of suicide-risk; the more risk factors the greater the assessed risk of suicide. Gigerenzer and Brighton (2009) state that heuristics are valuable tools in judgement and decision making. According to them, heuristics can, by ignoring information, make decisions faster and more accurately than complex and resource-intensive processing procedures; a less-is-more effect. Humans are equipped with a "toolbox" (the adaptive toolbox) of different heuristics to be used under different circumstances, and with individual variations (Gigerenzer &

Gaissmaier, 2011). Heuristics and the principles of using them are hardwired by evolution, individually learned, and learned through social processes.

Quality development

Competence centres for the National Quality Registries in Sweden concerning health and medical services have been established (Nationella kvalitetsregister, 2014). The Västra Götaland Registercentrum includes eleven psychiatric quality registries (Portal för psykiatriska kvalitetsregister, 2014). Among the psychiatric registries, there is a registry concerning persons with a diagnosis of ADHD (BUSA) and a registry concerning persons with a bipolar diagnosis (BipoläR). These registries encompass data related to patient problems, performed interventions and outcomes. The GAF scale is one of the measures used in these registries. The main aim of the registries is to follow-up on the content of care and to continuously develop the quality of care. Furthermore, the competence centres should actively support research related to the registries.

The registries publish annual reports. For example, the BipoläR registry has published results related to the process and outcomes of treatment (Kvalitetsregister BipoläR, 2013). Among other outcome measures, the GAF scale and the proportion of patients with relapses have been used. An association between the GAF score and relapse has been found. During the period of 2008 – 2013, persons belonging to the group with the lowest GAF scores had the highest proportion of relapse (approximately 70%), and persons belonging to the group with the highest GAF scores had a lower proportion of relapses (approximately 33%).

In this thesis, the reliability of the GAF has been investigated, the GAF has been used as a measure of psychosocial functioning at admission and at discharge, as a measure of outcomes, and some socio-demographic and clinical variables have

been used as predictors of the GAF score at admission and discharge. At Östra Hospital psychiatric clinic in Gothenburg, the GAF has been used for priority reasons to obtain information about patients' levels of global functioning at different stages in the care process, and for assessing the outcomes of treatment. However, the different clinics of psychiatric care in the Region Västra Götaland have all in some way been involved in quality development related to the GAF measurement.

In February 2004, the medical sector council of psychiatry in the Region Västra Götaland set forth a document concerning vertical priorities within psychiatry (Västra Götalandsregionen, 2006). The main aim of that document was to draw the border of responsibility concerning interventions of assessment and treatment regarding mental health between the specialised psychiatric county council care, primary care and municipality interventions. The tools used to prioritise patients concerning the appropriate level of care were mainly their current GAF-levels and psychiatric diagnoses.

Indicators of quality have been developed in the RegionVästra Götaland to follow-up the health and medical care services in the purchaser-provider model (Västra Götalandsregionen, 2009). The criteria for the indicators were, among others, that they should be reliable and easy to provide. There were 15 indicators listed for psychiatry in the year 2009. One of the indicators was related to the GAF scale and was defined as the proportion of patients that were assessed by the GAF scale at some occasion during the last year.

Outcomes research

Efficacy studies

The randomised controlled trial (RCT) design is generally seen as the "gold standard" of research (Dunn, 1994) and is characterised by random sampling of

the participants to an experimental condition and to a control condition. Intervention procedures are rigorously guided and controlled. The experimental setting is specifically arranged for in the study. Randomised clinical trials accomplish efficacy studies and provide results on treatment efficacy under best-practice conditions.

Effectiveness studies

In outcomes research, effectiveness studies are performed, and the use of scientific methods for the analysis and interpretation of data that are routinely collected in clinical practice is achieved. The aim is to evaluate the effectiveness of the accomplished interventions (Gilbody, House & Sheldon, 2002). The sample of patients, the interventions that are used, and the assessment procedures are all part of the ordinary health care environment.

Comparisons between efficacy and effectiveness research

Efficacy studies performed within the psychiatric domain are usually oriented towards short-term outcomes while effectiveness studies evaluate long-term outcomes. The interventions in effectiveness studies are more of a "black-box-type" compared to efficacy studies. They are not specified and controlled, making it difficult to obtain information on what interventions were used and how they were used. Efficacy studies are usually characterised by more frequent follow-up occasions than effectiveness studies (Wells, 1999).

Outcomes research in mental health has its strength in relation to its natural base in which the investigated patients are all patients receiving ordinary care, including important subgroups. Studied interventions occur in daily practice and encompass different interventions with diversified combinations, and clinicians make efforts to match the patients and interventions (Essock, Drake, Frank &

McGuire, 2003). Data in effectiveness studies are gathered without heavy expenses and too much effort, and the database is generally large, which strengthens the statistical power. This leads to an easier application of research results to the ordinary treatment settings (Gilbody et al., 2002) and strengthens the external validity of the studies.

Randomised controlled trials have methodological advantages due to the established control conditions. The random allocation of patients to an intervention or to a control group makes the two groups similar and makes it more probable that changes in the outcome measures can be attributed to differences in the interventions (Essock et al., 2003). The control design strengthens the internal validity and supports interpretations about factors contributing to the observed effect. Randomised controlled trials require a lot of time and money. The addressed questions and used interventions are often simplified, and the experimental conditions tend to be artificial. The sampled participants are often highly selected through specific inclusion and exclusion criteria (Essock et al., 2003; Gilbody, House & Sheldon, 2003).

The weaknesses of outcomes research can be assigned to the choice of collected data and to selection bias (Gilbody et al., 2002; Iezzoni, 1997). The collected data could be more related to an administrative process than to clinically important questions. There is also a risk of poor quality data. The selection bias makes it hard to sort out patient related factors from other factors that might have contributed to the obtained results. Unmeasured characteristics might affect the outcome but are not available for analyses (Wells, 1999). It is complicated to compare results from different divisions within a project owing to the different composition of participating patients. To reduce this influence of case-mix, correcting statistical methods have to be used (Davies & Crombie, 1997).

Gilbody et al. (2003) conducted a survey concerning the use of outcomes research in psychiatry. The inclusion criteria consisted of research conducted in an ordinary care setting and the outcome data were collected routinely. Studies that only investigated the association between patient characteristics and the outcomes were excluded. Nine studies were identified. The research questions that were addressed encompassed the evaluation of mental health policy and the evaluation of new technologies. The sample size in these studies was generally larger compared to randomised controlled trials. All studies used methods to statistically adjust for case-mix and confounding variables.

A constructive way to look at these seemingly contradictory standpoints between experimental efforts and investigations in clinical practice is that both are needed to increase knowledge. Marks (1998) noted that "The results of RCTs and of routine care are two sides of the same gold coin. Each deserves equal scientific status and funding to yield its own kind of essential information" (p. 281).

Outcome assessment

Outcome measures can be divided into two types, unstandardised and standardised. A standardised measure has known psychometric properties in terms of validity, reliability and sensitivity. However, unstandardised measures are usually used in routine mental health care and are relied upon by staff. Slade, Thornicroft and Glover (1999) put forward the term "feasibility" as an important characteristic of a useful standardised measure. Feasibility relates to the usefulness of an instrument in typical clinical settings. Many instruments lack feasibility. According to Slade et al. (1999), it is important that the instrument is easy to administer and is not too time consuming. Another requirement is that it should be possible to use the measure with minor formal training. The obtained

results should be easy to understand and in accordance with clinical judgement. It is essential that the management explicitly expresses the utility of the measures and also provides proper feedback to the staff.

A questionnaire survey of consultant psychiatrists in the UK, with a focus on the use of outcome measures, was conducted by Gilbody et al. (2003). The disorders where outcome measures were most commonly used for assessing the severity of specific psychiatric problems were depression/anxiety (44.6%) and cognitive impairment (55.3%). Very few respondents reported the use of measures for identifying deficits in social functioning or quality of life. Among the clinicians, 11% stated that they used a measure for measuring clinical change over time. As the main result, the authors stated that the majority of clinicians did not routinely use outcomes measures. Gilbody et al. (2003) suggest different ways to improve the use of outcome measures and outcome research. Measures used should adequately assess the well-being of the patient and add clinically useful information, and they need to be valid, reliable and sensitive to change. The used measures should also be able to answer questions about the effectiveness of interventions and services. It is important to use adequate information technology to record, store and retrieve information, and clinicians need feedback about the patient outcomes at an aggregated level. Methodological research should at best be characterised by using control groups and statistical methods to control for the influence of confounding variables.

Walter, Cleary, and Rey (1998) conducted a survey related to mental health staff attitudes about using outcomes measures. The respondents had all been part of a Common-wealth-funded project concerned with rating patient outcomes. Few respondents (9%) believed that using outcome measures improved patient management and 67% were reluctant to use outcome measures in the future. A positive attitude to routinely measure outcomes was associated with having

experienced the fact that the measures were not too time-consuming. A negative attitude was associated with an increased workload for already overloaded staff, reducing time for contact with patients and not leading to improved care. Walter et al. (1998) concluded that their study indicates that outcome measures should be short and few. The measures should also express a patient's clinical state, inform about the treatment course and be useful in resource allocation.

The multidisciplinary group of the Outcomes Roundtable has set up guiding principles of outcomes assessment (Smith, Manderscheid, Flynn, & Steinwachs, 1997). Outcomes assessment should: be appropriate to the application or question being answered; include generic and disorder-specific information; place a minimal burden on the respondent and have the ability to be adapted to different health care systems. Tools for assessing outcomes should quantify the type and extent of treatment the patient receives, have demonstrated validity and reliability and must be sensitive to clinically important change over time. Outcomes should also be initially assessed and reassessed at clinically meaningful time points.

Psychometric properties of measures

As mentioned in the previous section on outcome assessment, the use of standardised measures to assess the outcomes of psychiatric treatment is recommended. According to Slade et al. (1999), a standardised assessment procedure is characterised by measuring the intended outcome (validity), the measure should produce the same result independent of time and the user of the measure (reliability), and the measure should have the ability to capture clinically significant changes. The two following paragraphs will focus on the constructs of reliability and validity.

Reliability

The concept of reliability concerns error in measurements and is related to the consistency of the measure. Reliability refers to the extent to which measurement error is absent from the obtained data (Suen, 1988). An estimate of reliability encompasses the consistency of results from repeated measurements or the consistency of results among different users of the measure (Carmines & Zeller, 1979). According to Shrout and Fleiss (1979), measurement error is common in the behavioural sciences.

Referring to classical test theory, objects of measurement have true scores on the dimension being measured. A true score of an ability is the true capacity of the ability. The obtained measurement score consists of the true score and the error score, and the error score encompasses systematic and random error processes (Nunnally & Bernstein, 1994). Systematic error concerns factors that affect all observations equally or systematically affect certain types of observations. Random error is related to factors that randomly affect the measurement of the attribute. Measurement error originates from an interaction between the object of the measure, the user of the measure, the actual measure and the surrounding context (Fhanér, 1974). The error variance of a measure adds to the standard error of the estimates. This reduces the effect size when inferring from sample to population and makes it harder to draw reliable conclusions from the sample to the population (Kazdin, 2002). Reliability can be defined as freedom from random error, and the ideal state is a measurement having the capacity to only measure the true score of the attribute. We can operationalise the concept of reliability and establish a reliability coefficient. The reliability coefficient is an estimate of the ratio of variance in true scores to the variance in observed scores (Nunnally et al., 1994).

There are different methods to estimate the reliability of a measurement that can be divided into five different approaches: internal consistency, alternative forms, retest, split-half and analysis of variance (Nunnally et al., 1994; Carmines et al., 1979).

The internal consistency approach concerns estimating the reliability of an instrument administered to a group of people on one occasion. Cronbach's alpha (coefficient α), and KR-20 for dichotomous items, are the most common estimates. These estimates are based on the average correlation among the test items. In alternative forms, there are two testing situations with the same people. In the second test, an alternative to the first used measure is administered. The correlation between these measures makes up the estimate of the reliability. In the retest method, the same test is administered twice to the same persons, after a period of time. The reliability of the measure is composed of the correlation between the scores on the two administrations. The split-half approach uses a split of the measure into two parallel halves, which is administered to the same people on one occasion. The correlation between the halves results in a measure of estimated reliability. The analysis of variance approach utilises the variance components in data to estimate a reliability coefficient. Its major use is in assessing the reliability of raters using an instrument to evaluate dimensions of specific targets. In this context, the analysis of variance generates an intra-class correlation coefficient (ICC), and the ICC is a correlation among measures constituting a class sharing the same set of variance components (McGraw & Wong, 1996). It is an estimation of reliability that takes into account both the inter-observer and intra-observer dimensions. An ICC considers true variance, random error variance and systematic variance. The ICC is computed in somewhat different ways according to the design and aim of the study. The intra-class coefficient can be conceptualised as the ratio of between-groups

variance to total variance and can be interpreted as the percentage of variance due to agreement among the raters (Bartko, 1976).

To estimate the true score of an obtained score from a measurement, the standard error of measurement can be used. The standard error of measurement can be depicted as the estimated standard deviation of obtained scores when the individual is given a large number of parallel tests. Through the use of the standard error of measurement, it is possible to estimate a confidence interval around the observed value, corresponding to the range of the true score (Nunnally et al., 1994).

Validity

The validity of a measuring instrument is concerned with how well it measures what it is intended to measure (Nunnally & Bernstein, 1994). To understand the current usage and meaning of validity it is useful to undertake a historical overview.

The concept of validity has evolved over the years. An early definition was related to a criterion-based model (Kane, 2001). The accuracy of the measure (the test) was associated to a criterion. The criterion measure was judged as having the ability to reflect the true values of the variable that the test was supposed to measure. According to this, validity was defined in terms of the degree of correspondence between the test values and the criterion values. Guilford (1946) stated that a test is valid due to its correlation with an appropriate external criterion measure. This criterion validity was seen as a property of the test (Goodwin & Leech, 2003).

In 1955, Cronbach and Meehl (1955) presented a paper focusing on a new validity dimension, namely construct validity. They also presented three other types of validity: concurrent validity, predictive validity and content validity.

Concurrent and predictive validity are specific types of criterion validity. Concurrent validity applies to the situation when the test score and criterion score are being measured at the same time. In predictive validity, the criterion score is obtained after the test score. Creating a relevant criterion is of great importance and may be harder than developing a predictive measure (Nunnally & Bernstein, 1994). Content validity is concerned with whether the test items constitute a representative sample of the domain meant to be measured. Construct validity is related to the attribute or quality of what is measured. It is concerned with whether the test measures a specific theoretical construct or trait. According to Nunnaly et al. (1994), a construct is an abstract and constructed variable and does not exist as an observable component of behaviour. The construct is explicated through observable and measurable variables. The investigator generates specific testable hypotheses to obtain a deeper understanding of the constructs related to test performance. There is a movement between the hypothesis and the obtained data.

The division presented by Cronbach and Meehl has been referred to as the trinity, or tripartite, view (Goodwin & Leech, 2003). According to Kane (2001), Cronbach and Meehl made a very important contribution with their focus on construct validity and the hypothetico-deductive model. The hypothetico-deductive model was a general scientific approach extended to measurement research. In Kane's´ model (2001), there is a second stage, termed the construct model. In the course of time, the construct view became gradually more central and comprehensive and the trinity view with its different validities was challenged. In the 1985 edition of the Standards for Educational and Psychological Testing (APA, AERA & NCME, 1985), it was stated that the use of different validity labels does not imply that there are distinct types of validity. Construct validity was put forward as the unifying concept. There was a shift from validity to validation, from intrinsic qualities of a test to supporting

evidence from many sources for using test scores. The Standards of 1999 (APA, AERA & NCME, 1999) presented a clearer picture of the change from the validity of a measure to the validity of an interpretation, "Validity is a unitary concept. It is the degree to which all of the accumulated evidence supports the intended interpretation of test scores for the intended purposes." (AERA, APA, & NCME, 1999, p. 11).

This emphasis on accumulation of research results and theoretical analysis is in line with the general definition provided by Messick (1989): "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment." (p. 13).

There are some further valuable contributions to the concept of validity. The information related to the analysis of validity can have different sources. According to the Standards of 1999 (AERA, APA, & NCME, 1999), there is evidence based on test content, response processes, internal structure, relations to other variables, and the consequences of testing. This means that it is possible to obtain support for the interpretation of the test score from many directions. Evidence based on relations to other variables is the most extensive source. In this group, it is common that the scores from the used measure are compared to scores from other measures. We can obtain construct-related information about whether the test scores converge to a measure of a closely related construct (convergent validity) or whether it diverges from a measure of a disparate construct (discriminant validity). In this way, we can obtain confirmatory or disconfirmatory support for the proposed interpretation of the test scores (Campbell & Fiske, 1959). Messick (1980) extends the meaning of validity by including ethical considerations. Is it valuable and appropriate to use the test in

the proposed application? He supports that these types of data are necessary to form a valid interpretation.

Validity also applies to the characteristics of research studies and research design. There are four types of experimental validity: internal, external, construct and statistical conclusion (Kazdin, 2002).

Internal validity is related to the intervention in the experiment. It focuses on the extent it is reasonable to suggest that the intervention (independent variable) accounts for the obtained results. External validity is concerned with the generalisability of the achieved results beyond the conditions of the study. To what degree is it possible to generalise the results to other settings, other groups of persons or to other geographical areas? Construct validity in research design is, as in the case with tests, related to the attribute or quality of the component in focus. Here, it is related to the quality of the intervention. What does the intervention consist of? What dimension caused the results? How are the findings to be explained? Statistical conclusion validity refers to the ability to make correct conclusions on statistical grounds. It is concerned with the ability of the investigation to detect effects if they exist. It relates to the size of the sample used, to the heterogeneity of the samples and the strictness of the procedures.

Classification of health interventions

The Swedish Classification of Health Interventions (KVÅ) is a national classification system of health care interventions (Socialstyrelsen, 2009; Socialstyrelsen, 2013). The historical background of KVÅ is related to a collaboration within the Nordic countries concerning the classification of interventions. The Nordic Medico-Statistical Committee (NOMESCO) published in 1996, a common Nordic Classification of Surgical Procedures

(NCSP) (Smedby & Schioler, 2006). All Nordic countries have translated national modifications of the NCSP, and the Swedish version was named the KKÅ. In 1995, the KKÅ was complemented by non-surgical procedures (Klassifikation av medicinska åtgärder, KMÅ), and this resulted in a new list of classifications: the KVÅ (Socialstyrelsen, 2009).

The KVÅ encompasses approximately 10 000 codes covering different medical specialisations. The main purposes of the KVÅ are to be an instrument for describing performed interventions and for following up on the content of care. It is maintained by the Swedish National Board of Health and Welfare and it is mandatory to report KVÅ codes in the health data registry of the National Board of Health and Welfare (Socialstyrelsen, 2009). The guidelines from the National Board of Health and Welfare for coding stipulate that routine interventions normally performed in relation to a specific problem should not be coded (Socialstyrelsen, 2006). There is also a recommendation that only the most important interventions should be coded, and in most cases, it is adequate to use less than the 12 possible registrations.

The WHO Family of International Classifications (WHO-FIC) (Madden, Sykes, & Ustun, 2007) encompasses the International Classification of Diseases (ICD), the International Classification of Functioning, Disability and Health (ICF), and the International Classification of Health Interventions (ICHI), which is under development. The development of the ICHI began in 2007 and an alpha 2 version was presented at the annual WHO-FIC meeting in Beijing, China, October 2013 (Madden, Napel, & Cumerlato, 2011; National Centre for Classification, 2013; Rodrigues, 2012). The ICHI is composed of seven sections, and the Interventions on Mental Functions section is related to the mental health domain.

International classifications of interventions in health care are mainly related to nursing. The International Classification for Nursing Practice (ICNP) was developed by the International Council of Nurses and encompasses codes for diagnoses, interventions, and outcomes (International Council, 2008). Efforts are being made to harmonise the ICNP coding with the ICHI coding (WHO Collaborating, 2012). Research related to the ICNP has mainly been concerned with its usefulness in clinical practice and its use as an instrument to clarify and enhance nursing (Conric, 2005; Dykes, Kim, Goldsmith, Choi, Esumi & Goldberg, 2009).

The North American Nursing Diagnosis Association-International (NANDA-I) has created the Nursing Interventions Classification (NIC), a classification system of interventions that nurses perform (The University of Iowa, 2013). Research has shown that the NIC could be useful for describing the interventions nurses use in different care settings and for different diagnoses (Wallace, O'Connel, & Frisch, 2005; Jones, 2003).

Functioning and functioning scales

Functioning is, according to Phelan, Wykes and Goldman (1994), an abstract concept encompassing a wide range of human abilities, with a close relationship to mental health and mental illness. Global functioning denotes all functioning taken together, and Phelan et al. (1994) argue that there is limited agreement about the precise definition of this concept. In their review of 14 global function scales, they focused on three areas: social attainment, social role performance and instrumental behaviour. The authors concluded that functioning is an important domain when measuring outcomes, but the relevant measure will depend on the requirements. Simple measures of functioning were recommended within routine clinical settings.

Johnson (2010) put together a compendium of psychosocial measures for researchers as an aid to investigate general and specific research questions concerning persons with serious mental illness. Johnson (2010) states that there has been a shift in recent years concerning the assessment of treatment outcomes from a reduction in symptoms to an improvement in social and vocational functioning. Johnson (2010) categorises measures in the functional domain into functional assessment, global assessment, social functioning and community living. The definition of social functioning includes the wide spectrum of social relationships in social interactions.

In their review of measures of social functioning as models for axis V in, at that time forthcoming, the DSM-IV, Goldman, Skodol and Lave (1992) investigated 20 different measures. The measures were divided into different groups dependent on whether symptoms were included, the depth and breadth of functioning and the focus of the clinical group. The definition of measures of social functioning by Goldman et al. (1992) was agreed upon by Phelan et al. (1994). Goldman et al. found that none of the alternative measures surpassed the Global Assessment of Functioning (GAF) scale, but it was recommended that the GAF scale should be moderately modified (to be described below).

The International Classification of Functioning, Disability and Health (ICF) is included in the World Health Organization (WHO) classification systems (World Health Organization, 2001). The ICF is a classification system for assessing health and health-related states, and has a holistic approach to human health and functioning. The construct of functioning is defined as the positive dimensions of the individual's interaction with the environment. Disability is accordingly defined as the negative aspects of this interaction. Body structures, body functions, activities and participation are the core components related to the individual's functioning or disability. Environmental factors consist of

physical, social and attitudinal components. Functioning and disability are made up of different structures and processes interacting with environmental requisites.

The WHO has also developed the WHO Disability Assessment Schedule 2.0 (WHODAS 2.0) (World Health Organisation, 2014). The WHODAS is based on the ICF and encompasses assessment across six activity areas: understanding and communicating; moving and getting around; hygiene, dressing, eating and staying alone; interacting with other people; domestic responsibilities, leisure, work and school; and joining in community activities.
There are three versions; the 12-item version, the 12+24 item version, and the 36-item version. The instrument can be self-administered and is assessed through an interview or by a third-party person. The WHODAS is included in the DSM-5 section of assessment measures, as a tool to assess disability (American Psychiatric Association, 2013).

The Global Assessment of Functioning scale

The Global Assessment of Functioning (GAF) scale constitutes a global measure of psychosocial functioning due to mental impairment and comprises AXIS V in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV-TR) (American Psychiatric Association, 2000). The other axes in the DSM-IV are Axis I, clinical disorders; Axis II, personality disorders; Axis III, general medical conditions; and Axis IV, psychosocial and environmental problems. The GAF encompasses psychological, social, and occupational functioning on a hypothetical continuum of mental health-mental illness from 1 (lowest score) to 100 (highest score). The measure is divided into 10 sections in which descriptions of specific symptoms and functioning capacities are presented (Appendix I).

The predecessor of the GAF scale from approximately 30 years ago was the GAS scale (Endicott, Spitzer, Fleiss, & Cohen, 1976), which was influenced by the Health-Sickness Rating Scale (HSRS) (Luborsky & Bachrach, 1974). In 1992, Skodol et al. (1992) recommended that the GAF should be modified by separating the measures of social and occupational functioning from the measures of psychological functioning and the instructions should permit ratings due not only to mental impairments but also to physical impairments. Despite this, the GAF scale has been mainly unchanged since its debut in the DSM-III-R in 1987 (American Psychiatric Association, 1987). The DSM-IV-TR was replaced by the DSM-5 in May 2013. The DSM-IV-TR structure was formed upon a categorical system but the DSM-5 is oriented towards a dimensional approach (American Psychiatric Association, 2013). Accordingly, there is no axis structure in the DSM-5; however, documentation of diagnosis is complemented by important psychosocial and disability information. This means that the GAF has been excluded from the DSM-5.

When using the GAF scale, the rater is instructed to consider psychological, social and occupational functioning along a hypothetical line from mental health to mental illness. Impairments in functioning related to physical or environmental limitations should be excluded. The rating procedure results in one global value concerning global psychosocial functioning (American Psychiatric Association, 2000). In the Structured Clinical Interview manual (SCID), there are four guiding steps depicted for determining a GAF rating (First, Spitzer, Gibbon, & Williams, 2002). At the first step, the rater should start at the top range of the scale and decide whether either the subject's symptom severity or the subject's level of functioning is worse than what is shown in the interval. The rater should thereafter continue until the interval is met, where either symptom severity or level of functioning is worse. The interval below is too severe for both symptom severity and level of functioning.

In the last step, the rater determines whether the asserted value is found in the upper or lower half of the range, and then decides a specific value. A split version of the GAF has also been developed and is comprised of the psychological dimension (GAF-S, symptom) and the social and occupational dimensions (GAF-F, functioning), resulting in two different values (Pedersen et al. 2012).

The majority of research concerning the GAF scale has centred on the psychometric properties regarding reliability and validity (Hilsenroth et al., 2000; Jones et al., 1995; Skodol, Link, Shrout & Horwath, 1988; Söderberg, Tungström & Armelius, 2005, Pedersen & Karterud, 2012).

Studies focusing on reliability have shown differences in regard to the fundamental variables included in the studies, such as the range of diagnoses of patients, outpatient or inpatient units and trained or untrained raters (Rey, Starling, Wever, Dossetor, & Plapp, 1995; Loevdahl & Friis, 1996; Hildebrand, McCann, Nelson, & Wass, 2003). It has been found that intra class coefficients (ICC) are generally higher for experts in GAF rating than for clinicians (Rey et al., 1995; Hilsenroth et al., 2000) and for trained clinicians compared to untrained clinicians (Söderberg et al., 2005; Vatnaland, Vatnaland, Friis, & Opjordsmoen, 2007). Studies in outpatient and inpatient settings have similar levels of reliability (Hildebrand et al., 2003; Jovanović, Gaŝić, Ivković, Milovanović, & Damjanović, 2008). The ICC for the reviewed GAF studies on reliability in this thesis ranged from 0.33 to 0.91, with an ICC generally larger than 0.60, which corresponds to a good to excellent level (Fleiss, Levin, & Paik, 2003).

Hilsenroth et al. (2000) investigated the validity of the GAF in a psychiatric outpatient sample. The results showed that the GAF was correlated with the Social and Occupational Functioning Assessment Scale (SOFAS), $r = .60$; and

with the Global Assessment of Relational Functioning Scale (GARF), $r = .60$. Jovanović et al. (2008) studied a sample of veterans in an outpatient clinic and computed a factor analysis on the GAF, the SOFAS and the GARF. The results showed a common single factor accounted for 77% of the variance. Concurrent validity of the GAF was investigated by Startup, Jackson and Bendix in a sample of in-patients with a diagnosis of schizophrenia. At six-month follow-up, the GAF and the Scale for the Assessment of Negative Symptoms (SANS) had a correlation of $r = -.49$; and the GAF and the psychotic dimension in the Scale for the Assessment of Positive Symptoms had a correlation of $r = -.48$.

In recent years, there has been an increase in studies using the GAF as an outcome measure. Greenberg and Rosenheck (2005b) used a nationwide mental health database to study the outcome of treatment for new outpatients, continuing outpatients and inpatients. The inpatient group had the highest change in GAF value from the start to the end of the study period (7.4 points) and continuing outpatients had the lowest (0.10 points). The expert-rated GAF scale and a self-report version were utilised in an outpatient study performed by Ramirez et al. (2008). The results confirmed generally good agreement concerning expert and self-report ratings. The overall change in the GAF value reached 14.7 points for the expert version and 17.0 for the self-report version. Wallsten et al. (2006) used the GAF scale to investigate the outcome of brief psychiatric inpatient care. A change in the GAF value from admission to discharge of 10 points or more was considered an improvement, and 58% of the patients were included in this measure. A psychogeriatric inpatient sample was studied by Kennedy, Madra and Reddon (1999). The mean GAF change from admission to discharge was 24.8 points, which corresponded to an effect size of 1.83. Piersma and Boes (1997) investigated the usefulness of the GAF scale in three samples consisting of adult inpatients, adult day hospital patients and

adolescent inpatients. Adolescent inpatients had the highest improvement with a change from admission to discharge of 11.6 GAF points.

GENERAL AND SPECIFIC AIMS

The general aim of this thesis was to investigate interventions and outcomes in psychiatric inpatient care through the use of assessment scales and database information. Another aim was to contribute to the knowledge of the Global Assessment of Functioning (GAF) scale in regard to its reliability and validity and as a measure of outcome.

Study I

The main aim of Study I was to investigate the reliability of the GAF scale when used by psychiatric staff in inpatient care. A further aim was to study factors associated with reliability and to evaluate a model of factors predicting the reliability of the GAF scale.

Study II

The overall aim of Study II was to investigate the outcome of inpatient psychiatric care. The GAF scale was used as a measure of outcome. Another objective was to obtain knowledge of patient-related characteristics concerning the GAF level at admission, at discharge and in relation to the change in the GAF level from admission to discharge. A further aim was to study the usefulness of routinely collected data stored in a large database.

Study III

The aim of Study III was to explore the influence of some clinical (diagnostic group, length of hospital stay, and ward affiliation) and socio-demographic

factors (sex and age) in relation to psychosocial functioning in the context of inpatient psychiatric care.

Study IV

The main aim of Study IV was to investigate the use of the Swedish classification system of health interventions (KVÅ) within inpatient psychiatric care. Further aims were to explore the KVÅ code groups in relation to the GAF scores at discharge and the socio-demographic and clinical variables as predictors of KVÅ code groups. Another aim was to explore the usefulness of the KVÅ in psychiatric inpatient care.

SUMMARY OF STUDIES

The introduction section has addressed different areas related to the four studies in this thesis. Some areas were more important to some of the studies and less important to other studies. All studies were performed within the context of inpatient psychiatric care and the process of deinstitutionalisation and involved psychiatric staff and psychiatric patients. In the introduction section, a paragraph about clinical judgment and decision making was included, which highlights the possibilities and problems that arise when personnel make judgements and decisions, such as when rating a patient's psychosocial functioning and deciding on a patient´s diagnosis. The issue of reliability was depicted in the introduction section, and reliability is a major concern in Study I but is also included in the other studies. The validity area was also described, especially in relation to the GAF validity, which is a concern for all of the studies, but in Study III, the validity of the GAF is of special concern. The components of effectiveness and efficacy studies were described, and comparisons were drawn. Especially Study II was an efficacy study, but this also applies to Studies III and IV. All of the

studies except Study I were related to outcome assessment. In the outcome assessment section, the importance of measuring treatment outcomes was addressed and the components of a feasible standardised outcome measure were described. A feasible measure is easy to administer, is not too time consuming, requires minor formal training, and the results are easy to understand. The GAF has been used as a standardised measure of psychosocial functioning in all four studies, and in Study II, the aim was especially focused on the outcome of inpatient psychiatric care as measured by the GAF scale. Different classifications of health interventions are described in the introduction section, and in Study IV, the use of the Swedish Classification of Health Interventions (KVÅ) was investigated within inpatient psychiatric care. Efforts within the area of quality development were mentioned, and hopefully, the four studies in this thesis will incite developmental efforts within psychiatric inpatient care.

Methods

Participants

Study I

Staff from six emergency psychiatric wards at the psychiatric clinic at Östra Hospital in Gothenburg, Sweden, participated in Study I during the autumn of 2008. The total sample consisted of 101 persons, including 93 psychiatric nurses and psychiatric technicians, seven paramedics and one resident physician. The participating psychiatric nursing staff represented 75% of all nursing staff in the wards, the paramedical staff represented 44%, and the doctors represented 14%. Fifty-seven per cent of the participants were women and 43% were men.

Study II

Study II gathered data from the healthcare information system used within Sahlgrenska University Hospital, named "ELVIS". The original sample of Study II consisted of patients admitted to and discharged from inpatient care at the psychiatric clinic at Östra Hospital, Gothenburg, Sweden, in 2009. This corresponded to a total of 1505 care episodes. Of the total care episodes, 689 care episodes were either GAF-rated only at admission or only at discharge, or were not GAF-rated at all. The study group consisted of 816 care episodes that were GAF-rated both at admission and discharge. This yielded a drop-out rate of 45.8%.

Study III

The sample of Study III was derived from a total of 1505 episodes of care, consisting of 1219 patients treated during 2009 at the psychiatric clinic of Östra Hospital in Gothenburg, Sweden. Episodes of care without GAF ratings at both admission and discharge were excluded. This resulted in a sample of 816 episodes of care with GAF ratings made both at admission and at discharge for 648 patients.

Study IV

The sample of Study IV consisted of the same participants included in Study II and Study III. That is, the sample included 816 episodes of care for 648 patients treated during 2009 at the psychiatric clinic of Östra Hospital in Gothenburg, Sweden.

Procedures and measures

Study I

The participants took part in an assessment session in which six case vignettes were to be rated according to the GAF scale. The case vignettes consisted of three text vignettes and three video vignettes and were selected to cover the patients' range of GAF values in the psychiatric wards. Fourteen sessions in total were administered to reach all participants, and each session lasted about two and a half hours.

All participants answered a questionnaire regarding gender, age, ward affiliation, occupation, completed GAF-training in the spring of 2008, experience in psychiatry, experience using the GAF, frequency of determining GAF ratings and the degree of effort required to determine a GAF rating. The questionnaire also consisted of six GAF-related attitude questions in the format of five-level Likert scales: 'Do you exert yourself to make good GAF ratings? How do you perceive the use of GAF in your work? How do you perceive the use of GAF in psychiatry in general? How do you perceive the use of rating scales in general in your work? Do you think that the GAF value is a good measure of the global functioning of the patient? Do you think that the GAF value is useful in your work with patients?' These items were used to compose an index named "attitude towards the GAF", which was the mean of the sum of the values of the five items.

An intra-class correlation coefficient (ICC) was used to study overall reliability. According to the guidelines of McGraw and Wong (1996), a model of the degree of absolute agreement among measurements, ICC (A.1), was chosen. The ICC is a reliability measure at the group level (Shrout & Fleiss, 1979). To compute statistics related to an individual reliability level, a residual variable was created. This variable made it possible to perform ANOVA and regression analyses based on each rater's individual residual value. The calculation method

for the residual variable originates from the work of Söderberg, Tungström and Armelius (2004).

Study II

Data in Study II were collected from the healthcare information system "ELVIS", used
within the Sahlgrenska University Hospital organisation. ELVIS includes 128 registered variables, and eight of these variables were selected from the database: episode of care, gender, age, diagnosis, length of stay, ward affiliation, GAF score at admission and GAF score at discharge.

The background categories used were gender, age, length of stay, diagnosis according to the ICD 10 (WHO, 1992) and ward affiliation. These categories were divided into subcategories. Age was divided into 18-34 years, 35-49 years and $\geq$ 50 years. Length of stay was divided into 1-7 days, 8-21 days, and $\geq$ 22 days. The diagnoses were arranged into 11 subcategories adapted from the Swedish Federation of County Councils' "Lf-lista" (Hälso- och sjukvårdens utvecklingsinstitut, 1994).

Six wards participated, where four of them were specialised in the treatment of affective disorders and two were specialised in the treatment of psychoses. The six wards were named from A to F. Wards A, C, D and F were mainly specialised in the treatment of affective disorders and Wards B and E in the treatment of psychoses.

The measure of change or improvement is central in Study II. A GAF change measure was computed as the difference between the patient's GAF value at discharge and at admission. Effect size was calculated by Cohen´s *d*, which was

used as a standardised measure to assess the effect of change in GAF points from admission to discharge.

Study III

The data in Study III were collected from the ELVIS healthcare information system used within Sahlgrenska University Hospital. Out of the 128 variables registered in ELVIS, the following variables were used in this study: the identified patient, episode of care, date of start and end of inpatient care, gender, age, ward affiliation, main psychiatric diagnosis, and GAF score at admission and at discharge. Multiple linear regressions were performed to analyse the prediction of GAF admission and GAF discharge scores. The model for predicting GAF at discharge was elaborated through the division of the total sample into different diagnostic categories.

Study IV

The data in Study IV were collected from the same source as in Study II and Study III, namely the ELVIS healthcare information system. In addition to the variables selected in Studies II and III, KVÅ code registrations based on the Swedish Classification of Health Interventions (KVÅ) were selected. The KVÅ is a national classification system of interventions within the health care sector of Sweden maintained by the Swedish National Board of Health and Welfare (Socialstyrelsen, 2009; Socialstyrelsen, 2013). Registrations were based upon a KVÅ code-list from 2009 and were elaborated within the Region Västra Götaland in Sweden, which consisted of 76 codes for psychiatric interventions. This list is annually revised (Västra Götalandsregionen, 2014). It was possible for the staff to register a maximum of 12 different codes in the registration system for each episode of care.

The KVÅ codes were used in computations and in analysis, both individually and combined into eight KVÅ code groups. The following KVÅ code groups were created, with examples of the included codes in parenthesis: clinical assessment (assessment of mental symptoms), medical treatments (administration of medication, per oral), medical-technical treatments (electroconvulsive therapy NOS), training of functions (support and training in mastering demanding social situations), psychological and social treatments (cognitive behaviour therapy and health conversations), coordinating interventions (case conferencing), coercive interventions (physical restraint less than 4 hours), and other interventions (simpler medical certificate). Logistic regression analyses were performed using the KVÅ code groups as the dependent variable and the socio-demographic and clinical variables as independent variables. One-way ANCOVA with the GAF score at admission as a covariate was computed comparing non-registered and registered KVÅ code groups in relation to the GAF score at discharge.

Main results

Study I

The overall intra-class correlation coefficient (ICC) value reached 0.79, which is in accordance with good reliability (Fleiss et al., 2003). The ICC for the text vignettes was 0.82 and 0.78 for the video vignettes. The ICC values of the different wards were centred at approximately 0.80. There were no significant differences between the different wards and occupational groups regarding reliability in terms of a comparison between their residual means.

The number of years of experience using the GAF, the frequency of ratings per month, the attitude towards the GAF, and participation in GAF training during the spring of 2008 were the independent variables and the residual variable was

the dependent variable in a multiple linear regression model. None of the predictors yielded a significant contribution.

Study II

The overall mean change in GAF value from admission to discharge was 20.7 points, with a mean value at admission of 32.6 points and 53.3 points at discharge. The overall GAF change expressed in effect size (Cohen's *d*) was 1.67, corresponding to a large effect (Cohen, 1992). Of all patients in the study group, 75% had a GAF change of ≥ 10 points and were considered improved according to the definition of Wallsten et al. (2006).

Significant differences in the GAF values at admissions were found regarding age, diagnosis, length of stay and ward. The lowest mean GAF value at admission was 29.6 and was observed for persons with schizophrenia (SD = 9.4) and for persons with other psychotic disorders (SD = 11.7). Significant differences in the GAF values at discharge were found for gender, diagnosis, length of stay and ward. The highest mean GAF value at discharge was found for persons with other mood disorders (M = 58.7, SD = 11.4) and for persons with bipolar disorders (M = 57.2, SD = 10.7).

Computations made at the category level concerning gender and age had no significant differences in GAF change scores. All of the other categories yielded a significant result.
Analysis for each subcategory of the difference between GAF at admission and at discharge yielded significant differences ($p < 0.001$) for all subcategories. The highest mean GAF change was found for persons with no registered diagnosis (M = 25.5, SD = 2.33) and for persons with other mood disorders (M = 25.2, SD = 2.33). The lowest mean GAF change was displayed for ward B (M = 11.7, SD = 13.8), persons with a diagnosis of schizophrenia (M = 12.1, SD = 14.3), and

for persons with substance related disorders (M = 12.1, SD = 12.2). When the difference between the GAF value at discharge and at admission was expressed in terms of effect size, the highest values were noted for persons with other mood disorders (Cohen's d = 2.33) and for persons with a length of stay between 8-21 days (Cohen's d = 2.09). The effect size was large for all subcategories.

Study III

Multiple linear regressions were performed to study the prediction of GAF admission and GAF discharge scores. The model for predicting GAF admission scores was statistically significant, and reached an R^2 of .06. Significant predictors were age (β = -0.15), schizophrenia (β = -0.08), other psychotic disorders (β = -0.09), and no diagnosis (β = -0.08). Multicollinearity statistics were computed according to the guidelines by Andy Field (2009). The tolerance statistic (1/VIF) for all independent variables was in the range of 0.75 – 0.91, and the Variance Inflation Factor (VIF) was in the range of 1.09 – 1.34.

The predictive model concerning GAF discharge scores was statistically significant and explained 24% of the variance in the GAF discharge variable. GAF admission, most diagnoses, and wards B, E, and F had significant results. Multicollinearity statistics were computed, and the tolerance statistic (1/VIF) for all of the independent variables was in the range of 0.51 – 0.93 and the Variance Inflation Factor (VIF) in the range of 1.08 – 1.96. The prediction of GAF at discharge was also computed separately for each diagnostic category, and some of the significant predictors were as follows: within substance-related disorders: GAF admission (β = 0.34) and Ward F (β = -0.43); schizophrenia: GAF admission (β = 0.30); bipolar disorders: Ward C (β = -0.23) and Ward D (β = 0.25); and stress-related disorders: Ward C (β = 0.33).

Study IV

For the total group of 816 episodes of care, 680 episodes of care (83.3%) had at least one registered KVÅ code. There were 3486 registered KVÅ codes, for a total mean value of 4.3 KVÅ codes ($SD = 2.9$). The 10 most frequently used codes for the total group are presented in order of descending proportion (the percentage out of total 3486 registrations): clinical assessment of mental condition (14.4%); prescription of medication (10.6%); gaining anamnesis (9%); administration of medication, per oral (8.7%); sampling NOS (7.4%); supportive conversations (6.3%); execution of prescriptions, aid cards and dietary advice (5.1%); conference about patient (4.4%); conference with patient (3.4%); and assessment of mental symptoms (3.3%). The first five codes according to frequency accounted for 50.1% of all registrations. The eight KVÅ code groups were related to ten diagnostic groups. Patients with a diagnosis of schizophrenia had the highest share of coordinating interventions (46%); bipolar disorders and other mood disorders had the highest share of medical technical treatments (10%, 13%); eating disorders had the highest share of training of functions (50%); and patients with a diagnosis within substance related disorders had the lowest share of psychological treatments (18%).

The association between a performed intervention and a patient's level of psychosocial functioning at discharge was investigated. A one-way ANCOVA with GAF score at admission as a covariate was computed to compare the non-registration of KVÅ code group to the registration of KVÅ code group in relation to the GAF score at discharge. Of the eight KVÅ code group constellations, medical-technical (mainly electroconvulsive therapy) and coordinating interventions displayed statistically significant results.

A closer analysis showed that episodes of care with a registration for medical-technical interventions had a similar GAF score at admission as episodes of care without any such registration (M = 32.60 and 32.55, respectively) but had a higher GAF score at discharge (M = 57.9 and 53.0, respectively). Episodes of care without registration of coordinating interventions had a higher GAF score at admission than those with registered coordinating interventions (M = 33.0 and 31.2, respectively) and had a higher GAF score at discharge (M = 54.4 and 49.7, respectively).

GENERAL DISCUSSION

The general aim of this thesis was to investigate interventions and outcomes in psychiatric inpatient care through the use of assessment scales and database information. Another aim was to contribute to the knowledge of the Global Assessment of Functioning (GAF) scale in regard to reliability, validity and as a measure of the outcome of treatment.

The studies in this thesis showed that a database made up of routinely registered information could be useful in following up on performed interventions and outcomes, and in contributing to scientific knowledge about the GAF measure. It also showed the utility of scientific methods when investigating these areas, although clear routines for the registration procedures to reduce data loss have to be developed.

Through the four studies in this thesis, interventions and outcomes in psychiatric inpatient care have been investigated. The interventions were captured by the KVÅ system and registered in the ELVIS database. The GAF was used as the outcome measure and was rated by personnel and registered in the ELVIS database. Of all episodes of care, 17% had no registered KVÅ-code, which means a loss of information about the performed interventions. A mean of 4.3

codes were used, which might imply an under registration of actually performed interventions. However, according to the instructions, all interventions should not be coded, only the most important. It has been discussed that the used KVÅ codes to some degree are codes on a rather general level, thereby not giving specific information about the performed interventions. Despite these limitations, this thesis has achieved to investigates interventions in psychiatric inpatient care using the KVÅ system.

GAF has been the only measure of outcome, which restricts the results to psychosocial functioning and leaves out information about symptoms and quality of life, although the GAF measure has provided valuable information about improvement in psychosocial functioning during inpatient psychiatric care. The psychometric properties of the GAF have been studied. The reliability investigation in Study I has added knowledge about the reliability dimension of the GAF, and the partial focus on the validity of the GAF in Study III has added knowledge to the validity dimension of the GAF.

In Study I, the reliability dimension of the GAF was studied. This was an important contribution in regard to the three following studies, as these used the GAF as a central measure. Personnel from the psychiatric inpatient care participated in assessment sessions, and many of these raters were the raters in the wards. As the reliability of the ratings was found to be good, the GAF measure in the following studies can be considered reliable. In accordance with the aims, a model for predicting the reliability was used, but no statistically significant results were found.

The outcome of inpatient care was investigated in Study II. In simple terms, the aim of Study II was to investigate the usefulness of inpatient psychiatric care. Usefulness was operationalised as outcomes measured by the GAF. A large improvement in psychosocial functioning could be found overall and when

divided into all of the research variables. It seems to have been useful to have used Cohen´s d as a measure of effect size. This measure has, in an overall way, the ability to depict whether results are in the range of small, medium or large effect; and it makes it also possible to compare results from other studies, where other samples or measures of outcome have been used. Study II achieved its aims to obtain knowledge about patient-related characteristics in regard to the GAF level at admission and at discharge. Study II showed that routinely collected data stored in a large database could be useful for following up on outcome.

The aim in Study III was to study the factors behind the results in psychosocial functioning found in Study II. In accordance with its aim, Study III shows the influence of some clinical and socio-demographic factors in relation to psychosocial functioning, such as age in relation to the GAF value at admission; and the GAF value at admission and most diagnoses in relation to the GAF value at discharge. The overall orientation in this thesis was to explore inpatient psychiatric care, which was also achieved in Study III, though it would have been desirable if there had been more clinical and socio-demographic factors registered in the database. It would, for example, have been fruitful to investigate the influence of voluntary or mandatory admissions, of the patients' socio-economic statuses and social networks on treatment outcomes in terms of the GAF value at discharge.

The main aim in Study IV was to investigate the use of the Swedish classification system of health interventions (KVÅ) within inpatient psychiatric care. This was investigated by describing what KVÅ codes and KVÅ group codes were registered and the amount of these codes and group of codes. Study IV met this aim through descriptive statistics. Another aim was to explore KVÅ code groups in relation to the GAF score at discharge. This aim had a link to

Study II and explored the association between the performed interventions and outcomes. This was investigated by a one-way ANCOVA computation. Episodes of care with no registered KVÅ code in the database belonging to a specific KVÅ code group were compared to episodes of care with a registered KVÅ code belonging to a KVÅ code group. The results were not altogether easy to interpret. For example, an episode of care without a registered KVÅ code belonging to the KVÅ code group for psychological treatments could mean that these types of interventions have been performed but not registered, making it hard to make a reliable comparison between the registration and non-registration of psychological treatments. To take another example, a registered intervention (for example coordinating interventions) could be associated with a lower GAF value at admission and at discharge than a non-registered intervention. This could be interpreted as a need for that specific intervention but it might be less adequate to conclude that the specific intervention had no positive impact on the treatment outcome.

Another aim was to explore the usefulness of the KVÅ in psychiatric inpatient care. It was hard to obtain an answer to this rather unspecific aim. The results of the descriptive statistics and the different statistical methods have showed some of the performed interventions and the relation between interventions and some clinical factors and outcome, and in that sense support the usefulness of the KVÅ. Study IV gave valuable information about the interventions in inpatient psychiatric care and how the KVÅ was used in clinical practice. In Study IV, there seem to have been too few codes registered per episode of care and the registered codes seem to have been too general. It could be valuable for the staff to receive clearer instructions concerning registration and to register from an elaborated KVÅ list.

Reliability

It was found in Study I that the GAF scale showed a good reliability (ICC = 0.79) when used by psychiatric inpatient care staff. Study I supports the findings of prior research where the reliability of the GAF, to a great extent, has been in the range of good to excellent. The variation of reliability scores across studies is to some degree related to different research variables, such as the range of diagnoses of the patients, outpatient or inpatient units and trained or untrained raters. There was no evidence in Study I that specific factors such as the occupational group, attitude towards the GAF scale and years of experience using the GAF had any impact on the level of reliability. Study I provides further evidence that there may not be clear cut factors associated with reliability, except for the effect of training, with respect to the GAF scale. In Study I, significant differences were not found to be related to different wards, and this is in line with the findings of Pedersen, Hagtvet and Karterud (2007), but is contrary to the findings of Loevdahl et al. (1996) and Söderberg et al. (2004). The raters in Study I belonged to different wards but were all situated in the same building. This geographic concentration might contribute to a similarity between the wards and thus one would not expect to find significant differences according to this unit.

The raters' attitudes towards the GAF were not related to the reliability measure, which differs from the results of Söderberg et al. (2005). This difference could, to some degree, be explained by the fact that the questionnaires used in the two studies were composed of different questions, although covering similar attitudinal areas.

The results from Study I did not show that the amount of years of experience using the GAF scale and the amount of GAF-ratings per week were associated with the reliability measure, which is in line with the findings of Söderberg et al.

(2005). In Study I, there might be training effects that were not considered and measured. Since 2000, the GAF has been used routinely at the psychiatric clinic in Östra Hospital. Most staff had attended a training course at least once. This means that most raters in Study I were trained, a factor that probably increased the ICC coefficient.

Another reliability issue can be addressed by the use of KVÅ-codes. The selection of a KVÅ code for a specific intervention is not always clear-cut. Different individuals might use different KVÅ codes for the same intervention. Especially for performed interventions concerning assessment, it can be somewhat difficult to find the correct code. One way to handle this could be to make up a list of some problematic interventions and their corresponding KVÅ codes.

The reliability component concerning the assessment of psychiatric diagnoses according to the ICD-10 has not been articulated in this thesis. The psychiatrist at each ward was responsible for assessing the diagnosis of the patient. Assuming that the psychiatrist follows the instructions in the manual, this will have a positive influence on the reliability. On the other hand, research has found diagnostic deviations between different raters when assessing the same patients.

Validity

The GAF is intended to measure the construct of global psychosocial functioning. Some studies have pointed out that the GAF is too strongly related to psychiatric symptoms with a great overlap between Axis I and Axis V in the DSM (Skodol, Link, Shrout, & Horwath, 1988; Roy-Byrne, Dagadakis, Unutzer, & Ries, 1996), and other studies have shown a strong relation with functioning

measures (Hilsenroth et al., 2000; Jovanović, Jasović Gaŝić, Ivković, Milovanović, & Damjanović, 2008).

The prediction model in Study III regarding GAF at admission, when diagnostic groups were entered as predictors, had a low proportion of explained variance (6%). One possible explanation for this result is that there is hardly any overlap between Axis I and Axis V. The prediction model of the GAF score at discharge in Study III accounted for 24% of the variance; where six out of the 10 diagnostic groups exhibited statistically significant regression coefficients. These results imply that the GAF encompasses psychiatric diagnostic dimensions, and supports the view of overlap between Axis I and Axis V.

There has been a debate concerning the inadequacy of letting three dimensions in the GAF result in one global score. A split version of the GAF has been developed as a way to meet this criticism. The psychological dimension makes up the GAF-S (symptom) and the social and occupational dimensions make up the GAF-F (function) (Pedersen, Hagtvet, & Karterud, 2007). The face validity seems to be great concerning the difference between these two measures; however, a strong association has been observed between them. Pedersen et al. (2012) concluded that the original GAF scale seems to work well as a global measure of symptom distress and social dysfunction. The original global GAF measure can be seen as a satisfying way of depicting a person's psychological, social and occupational functioning.

It has been proposed that the observed correlation between a psychiatric diagnosis and the GAF is due to an overlap of symptoms (Goldman et al., 1992). The GAF is composed of psychological, social, and occupational functioning. A psychiatric diagnosis includes functioning components in addition to symptoms. Pedersen et al. (2012) observed a great overlap between the GAF-S (symptom) and the GAF-F (functioning), which might imply that functioning dimensions

are included even when the intent is to measure symptoms alone, and vice versa. The stated overlap between Axis I and Axis V might be attributable to the overlap of both symptoms and functioning factors from each of the two axes. Thus, the overlap of Axis I and Axis V might be characterised by the overlap of symptoms as well as the functional characteristics.

The DSM-IV-TR was formed by a categorical structure, but the new DSM-5 reflects the demand for a dimensional structure. Specific diagnoses are no longer considered to be separate from other diagnoses and mental conditions are seen as related to behavioural and psychosocial factors. As the borders are less pronounced in the DSM-5, it could be stated that the former distinction of five specific axes was artificial to some extent. This dimensional approach in the DSM-5 allows for the overlap between symptoms and psychosocial functioning to be plausible and not undesirable. Study III provides some additional support of the construct of global psychosocial functioning in regard to the GAF, containing psychological (primary psychiatric symptoms), social and occupational functioning.

In this thesis, there has not been a research focus concerning the validity of the assessed psychiatric diagnosis according to the ICD-10. There is a continuous process regarding the development of psychiatric nosology, as we can see from the nosology of Kraepelin and Bleuler to the psychiatric classification used in this thesis. The DSM-IV was followed by the DSM-5 by about one and a half years, and the ICD-10 is in a development phase into the ICD-11, which is expected to be released in 2017. We could assume that the newer versions of psychiatric diagnoses are more valid than the older ones. There is a desire to base the nosology on empirical evidence; however, research results do not always point in one direction. The nosology is still based on phenomenology,

although genome-wide association studies with a genetic orientation have been performed, which may influence the nosology of psychiatric diagnoses.

Classification of health interventions

There are many interventions performed by staff in relation to patients within inpatient psychiatric care. There is scarce knowledge about these interventions, their relations to patient characteristics and to outcomes. In Study IV, the Swedish Classification of Health Interventions (KVÅ) was explored. The KVÅ has the potential to generate statistics that can be used for different purposes, but there are also shortcomings of its use in clinical practice. One aspect relates to granulation, that is how detailed a code should be. Codes that are too specific may be very infrequently used, difficult to find in the classification list and might occur so infrequently that it is difficult to make useful statistics of them. Codes at an aggregated level, on the other hand, are easy to use for various interventions but they risk being used too frequently, and may convey little information. It is necessary to find a balance between the specific and the general. The KVÅ is a local Swedish system and most countries do not have a national classification list. The International Classification of Health Interventions (ICHI) encompasses an elaborated system of interventions and has the potential to be used in different countries, thereby making it unnecessary for countries to create their own classification list (National Centre for Classification in Health, 2013). The ICHI could be used for following up on the content of care, within and between countries. To motivate the personnel to take interest in the registrations, it is important that they have clear guidelines concerning registration procedures and that statistical material is used for following up on the content of care.

Outcomes

Study II found a high change in global functioning from admission to discharge, with a mean GAF value of 20.7 points and an overall effect size of 1.67 (Cohen's *d*), corresponding to a high effect (Cohen, 1992). Of all of the patients in the study group, 75% had a GAF change of ≥ 10 points and were considered improved, according to the definition of Wallsten et al. (2006). The mean value at admission indicates a low level of functioning and the corresponding value at admission indicates a moderate level of functioning.

The GAF scores at admission could be valuable to obtain information concerning whether the enrolled patients correspond to the organisational norms of which patients should be selected for inpatient care. The patients in the present study had a total mean at admission of 32.6 ($SD = 10.8$), which corresponds to a severe deficit in functioning and is related to a need for inpatient treatment. The results concerning the GAF level at discharge could provide valuable information about the patients' functioning abilities. The mean value in this study at discharge was 53.3, indicating a moderately disturbed capacity of functioning and a need for continued outpatient treatment.

A large improvement in terms of the GAF change in Study II was noted for persons with a diagnosis of mood disorder. The study by Ramirez et al. (2008) also showed that persons with mood disorders had a high observer-rated GAF change. In a review of studies focusing on patients with a diagnosis of bipolar disorders treated with antidepressants, Gijsman, Geddes, Rendell, Nolen and Goodwin (2004) estimated an overall rate of clinical remission of 43% within a follow-up period of 4 to 10 weeks.

The high improvement in this study among patients with mood disorders could be related to the large potential for improvement in this type of disorder. It could also indicate that treatment programs and their accomplishments are adequate. The lowest improvement in the diagnostic group expressed in GAF scores was found among persons with a diagnosis of schizophrenia and for persons with a diagnosis of substance-related disorders. The lower improvement rate for persons with schizophrenia could be ascribed to the narrower range of improvement potential in this type of diagnosis. The discharge mean value of 41.7 could indicate that within this diagnostic group there is an upper limit in the capacity of overall functioning. Another explanation could be that treatment programs in the studied clinic are not fully developed for persons with schizophrenia.

One factor related to the relatively low GAF change for persons with substance-related disorders could be the coexistence of an affective or psychosis diagnosis. This comorbidity could imply limitations in improvement. Another contributing factor could be that this diagnostic group is a primarily cared for by the specialised Substance Abuse Clinic. When persons with substance-related disorders have to be admitted to the psychiatric clinic due to full occupancy at the Substance Abuse Clinic, it is possible that they do not receive inpatient interventions that are fully adapted to their diagnosis.

Study III also had an outcome focus, exploring some socio-demographic and clinical variables as predictors of GAF scores at admission and at discharge. One finding was that there was no association between the specific wards' specialisations in diagnosis and outcomes measured by the GAF. These results question the basic assumption of a positive association between a ward´s specialisation in diagnosis and treatment effect. A requirement for improvement is treatment methods that are appropriate for the specific diagnosis. Perhaps the wards had not established diagnosis-specific treatments.

To follow-up on the content of care, it is fruitful to have a system for registering the performed interventions and to relate these interventions to outcome measures. The relation between the KVÅ code group and psychosocial functioning at discharge was explored in Study IV. Medical-technical (mainly electroconvulsive therapy) and coordinating interventions had a statistically significant relation to psychosocial functioning at discharge. Episodes of care with a registration for medical-technical interventions showed a higher GAF score at discharge compared to episodes of care without any such registration. This finding supports the use of medical-technical interventions for select patients, but it does not mean that medical-technical interventions would lead to high psychosocial functioning at discharge for all patients. Episodes of care without a registration of coordinating interventions showed a higher GAF score at admission than those with a registration and a higher GAF score at discharge. This result could be a sign of a relevant use of coordinating interventions related to patients' lower status of psychosocial functioning. When studying outcomes in relation to the KVÅ codes, it is important to undertake a thorough analysis.

Clinical judgment and decision making

The results in Study I can be considered the basis for the results in the following studies. If the result in Study I had shown low reliability, much of the results in the other studies would have been highly unreliable. Although the reliability in Study I was found to be good, there are still measurement errors in the obtained data regarding the GAF measurement. Research on the reliability of the GAF scale has yielded varied results. The measure in terms of the structure and content of the GAF scale can be considered a constant and cannot, from that perspective, be considered a cause of the variation found. The measurement error is found in the diverse processes regarding the rating activity.

Conscious and unconscious information processing occur in the measurement process (Wilson, 2002). This processing can be expressed as activations in System 1 and System 2 (Kahneman et al., 2002). When information from the patient is used to calculate a GAF-rating, the clinician will consciously try to match this information with the information given on the GAF scale sheet. In the SCID manual, there are conscious steps to take to proceed until a specific GAF value is reached (First et al., 2002). When all of the information is gathered and the clinician has interacted briefly with the GAF scale sheet, the next stage is to convert this information into a GAF value. There is a focus on conscious processing in the manual, but unconscious processing will probably also contribute to the decision of the final GAF value. The clinician knows the chosen GAF value after this processing, but has difficulties through introspection to obtain information about the content of the unconscious part of the processing (Wilson, 2002). Because many elements of the processing are unknown, it is hard to determine if there are judgment errors that need to be corrected.

The science of heuristics (Gigerenzer & Gaissmaier, 2011; Kahneman, 2003) can be used to illuminate the cognitive processes of assigning a GAF value to patient-related information. Different heuristics can be in use, and different raters can prefer different heuristics. It should be noted that the use of heuristics can be an effective way of processing information about patients as specific GAF scores, but it also entails sources of errors.

The availability heuristic could be one heuristic in use. A way to look at the availability heuristic in this perspective is to focus on what comes to the mind of the clinician, in terms of the functioning capacities of the patient. To what extent does the clinician rate the GAF level on the basis of functioning information that is easy to remember? What information is easier to remember? It could for

example be functioning behaviours that have made a strong impact or functioning behaviours that the patient recently has displayed. Through the use of the availability heuristic, there is a possibility that the GAF level will be assigned from automatically and easily remembered information, and omitting other relevant information.

Related to the availability heuristic is the take-the-best-heuristic. The rater might choose between different functioning capacities of the patient and select the capacity that the rater believes to be exclusively related to the "true" GAF value. The value associated with this capacity will be the final GAF score.

The representativeness heuristic processing could be involved when the clinician perceives a piece of functioning component in a patient that is similar to a piece of functioning component in another person, and makes that piece of similarity to be similar to the other persons overall functioning.
The process of representativeness judgment may also operate when a clinician has information of some area of functioning, but makes a judgment as if the whole functioning capacity was known. The clinician matches some functioning data with a more or less unconscious prototype. It could be assumed, for example, that the clinician has made up a prototype of the GAF score interval from 41 to 45, to which the clinician compares the limited information of the patient, and then makes a judgment.

The anchoring heuristic may also be activated in the rating process. When a patient is admitted to the ward, the clinician might unconsciously include in the judgment that this patient has a low level of overall functioning because the person will have to receive inpatient care. The opposite could be assumed in the stage of discharge when the clinician might include in the judgement that the patient has a moderate overall level of functioning because the patient is in the stage of leaving the ward. Another anchoring-related phenomenon could be

when two patients have the same rate of improvement in functioning from admission to discharge, but different GAF values at admission, and the clinician attributes a greater GAF change for the patient with the lower GAF value at admission. Anchoring may also be a concern in regard to the GAF scale sheet. Whether the rater starts at the bottom, in the middle or at the top of the scale when assigning the GAF value might influence which value is selected.

The number of functioning capacities with lower functioning levels may be summed up, and the higher the number, the lower is the GAF score. This could be attributed to the tallying heuristic.

Perhaps some types of functioning concerning the patient evoke strong positive or negative emotions in the rater. These highlighted emotions could affect the ratings and may be related to the affect heuristic. Affective reactions can also be associated with different decisions in the rating process and may guide the rater to be satisfied with the selected value or to make corrections.

Intuitive processing, beyond heuristics, is a part of the GAF-rating process. The development of intuitive GAF-rating skills requires adequate experience and appropriate feedback. Repeated experiences with GAF rating will to some degree develop intuitions. In regard to GAF-rating in psychiatric wards, it is probably difficult for the rater to obtain guiding feedback from the performed ratings. This could to some extent result in an individual feedback-loop in which the rater subjectively validates their own ratings. A good feedback source could include discussions with other employees concerning the patient's GAF scores.

One way for the clinicians to reduce their rating biases could be to thoroughly gather information of the patient in all functioning areas and to document this information for further use. It is also a positive factor if the rater has confidence in the unconscious clinical judgment process, but also is aware of possible

biases and strengthen the rating process through conscious elaborate judgment. There is research supporting the claim that clinicians make their own conclusions based on what they perceive as valid (Garb, 2005). This focuses on the necessity of providing clinicians with properly empirically based feedback on their GAF ratings. This could be reached through training programs in which clinicians receive validated feedback on their achievements.

The WHO Disability Assessment Schedule 2.0 (WHODAS 2.0) could be a complement to the globally oriented GAF scale. The WHODAS encompasses six areas of activity, and each item contains specific behaviours. There seems to be less room for subjective judgments and biases in the WHODAS than in the GAF scale.

An overall approach to reduce the influence of errors in clinical judgment and decision making could be courses in cognitive processing, focusing on unconscious processing and the use of heuristics. Metacognitive training, which is to be aware of and think about one's own thinking, seems to be a feasible way to proceed.

Strengths and limitations

There are limitations to Studies II, III and IV regarding the extent that the different aims have been captured. The reasons include a lack of randomisation of the patients to an "experimental" situation and to a control group. For example, the outcome studied in Study II might be an effect of time, the studied predictors of psychosocial functioning at discharge in Study III might be related to the specific episodes of care that might also be attributed to the KVÅ code group predictors of the GAF scores at discharge in Study IV.

The findings of Study I have limitations related to the created rating context with the use of case vignettes. The use of case vignettes creates a milieu that

differs from the rating conditions in the ward. The controlled and relatively stress-reduced experimental-like situation likely produces conditions for a higher reliability in comparison to clinical practice. In this sense, the difference between the test situation and the clinical situation contributes to a restriction of the possibility to generalise the results of Study I, limiting its external validity.

However, the vignette method and the controlled rating situation are also associated with positive factors. Many raters assessed the same patient, which provided an opportunity for comparison between the raters. The information about the patients was standardised and was the same for all raters, which reduces the sources of variance, and strengthens internal validity. Another strength of Study I is that the raters, to a great extent, were the same as the raters in the clinical practice. This supports external validity in terms of the generalisability of the results to other psychiatric inpatient clinics with psychiatric nursing staff as the primary GAF rates.

Limitations in Study II could mainly be addressed by the research design. The interpretation of the results of Study II is limited due to the absence of a randomisation procedure and a control group. The independent variables, in terms of treatment interventions, were not controlled for and regulated, which reduces the internal validity. It should also be mentioned that the analysis of data was built upon episodes of care and not upon specific patients. As the number of episodes of care exceeded the number of patients, patient characteristics might have had a higher influence on some episodes of care, and thereby the results.

In Study II, there might also have been organisational "pressure". To enrol a patient, the admission GAF should not be too high, which lowers the GAF level at admission and provides the conditions for a larger GAF value change. Another limitation of Study II is related to the performers of the assessment

because it was the providers of care who rated the GAF at admission and discharge instead of an independent expert. This could contaminate the assessment process and result in biased GAF ratings. Another source of limitation is related to the great amount of attrition and to the significant differences concerning the background variables among the studied sample and the attrition group. However, when the GAF scores at admission and discharge were investigated for the study group and the attrition group, there were no significant differences between the groups.

Study II has strengths related to external validity. The data were gathered from a naturalistic care environment without any research restrictions regarding patient recruitment, the personnel involved, or the treatments used. Another strength is the fact that all of the raters had participated in GAF training performed by the psychiatric clinic, and most of them had also participated in Study I of this thesis concerning the reliability of the GAF, which found good reliability (ICC = 0.79).

The strengths of Study III encompass the same strengths as Study II: a large database that spanned an entire year and included many patients and interventions making it possible to split data without substantial loss of statistical power; the data were gathered from a naturalistic care environment without restrictions regarding patient recruitment, the personnel involved, or the treatments used; and the personnel were trained in GAF rating and many of them participated in a GAF reliability study that found good reliability.

The limitations of Study III include the same limitations as Study II: no randomisation or control group; no control of performed interventions; the data were built upon episodes of care and not upon patients; the GAF raters were also the providers of care; and the high level of attrition. Furthermore, regression analysis was used in Study III, which examines the relationships between

variables, but it cannot determine causality. Study III suggests that there are associations between ward affiliation and outcomes, but there are no specific research variables connected to ward level that could be used to illuminate the observed between-ward differences.

The focus of research in Study IV on interventions in psychiatric inpatient care could be mentioned as a strength. Research in this field is scarce and it is important to know more about the performed interventions and their relation to patient characteristics and outcomes.

Specific limitations for Study IV concern the proposed use of too few KVÅ codes for each patient and episode of care, and the use of very general codes. This might lead to a gap between the interventions registered and the interventions performed in clinical practice, which limits the possibility of understanding which interventions are used in clinical practice and their outcomes.

Conclusion

The findings in Study I confirm that the GAF scale can be a reliable instrument to measure global psychosocial functioning within inpatient psychiatric care. GAF training seems to be an important factor that increases reliability.

The results in Study II showed a large change in patients' global social functioning from admission to discharge, indicating that the studied inpatient care units contribute to an improvement of the patients' overall psychosocial functioning. Cohen's *d* was used as a measure of effect size and it could be a valuable tool for expressing outcomes in terms of the degree of the effect.

Study III explored some socio-demographic and clinical variables in relation to the GAF score at admission and the GAF score at discharge. The study focused on factors related to the GAF improvement found in Study II. One finding was

that there was no association between the specific wards' specialisations in diagnoses and outcomes measured by the GAF. The results also suggested that, unrelated to each ward's diagnostic specialisation, some wards achieved greater improvements than others for specific diagnoses. These results imply that the basic assumption of a positive association between specialisation and treatment effect might be questioned. Furthermore, the results could be an incentive to develop the content of care.

In Study III, there was also a focus on the validity issue concerning an undesired large proportion of psychiatric symptoms in the GAF measure, leading to an overlap between Axis I and Axis V in the DSM. The results found evidence of overlap as well as of no overlap. The conclusion in Study III was that an overlap between symptoms and psychosocial functioning is plausible and not undesirable for a measure intended to encompass psychological, social and occupational functioning.

Study IV explored interventions in psychiatric inpatient care described through the Swedish Classification of Health Interventions (KVÅ). The results identified the most frequently used KVÅ codes and groups of KVÅ codes, and KVÅ code groups related to diagnostic groups and to psychosocial functioning at discharge. The study could describe performed interventions, if we assume that there is a strong correspondence between the registered and performed interventions. The results showed that the link between interventions and outcomes is important but not always easy to interpret. A classification system within the health care sector creates a foundation for following-up on the content of care, for making comparisons between different health care units, and for developing the content of care.

The four studies in this dissertation support the conclusion that routinely collected data registered in a central database system could be useful to describe

and analyse the content of care and its relation to outcomes. Scientific methods seem to be very useful when investigating these areas. Hopefully this research will contribute to the development of content in inpatient psychiatric care. Future research should benefit from focusing on the relations between performed interventions and the outcomes of these interventions within psychiatric inpatient care. It would be valuable to create clusters of interventions in the analysis, thereby creating patterns of interventions based on clinical practice. In addition to clinician-rated outcome measures, it is important to use patient-reported outcome measures and to involve relatives in the outcome measurement process. It would be valuable to use different measures of outcomes such as more comprehensive measures of psychosocial functioning, measures of psychiatric symptoms and health-related quality of life measures. The study of the long-term effects of treatment should also be a concern and should include follow-up measurements at different time intervals.

REFERENCES

Aas, I. H. M. (2011). Guidelines for rating Global Assessment of Functioning (GAF). *Annals of General Psychiatry, 10*, 1-11.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington, DC, American Psychological Association.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC, American Educational Research Association.

American Psychiatric Association (1952). *Diagnostic and Statistical Manual: Mental disorders*. Washington DC: American Psychiatric Association.

American Psychiatric Association (1968). *Diagnostic and Statistical Manual: Mental disorders, 2nd ed*. Washington DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: Author.

American Psychiatric Association (2011). *DSM-5 Development.* Retrieved April 14, 2011, from http://www.dsm5.org/Pages/Default.aspx

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.). Washington, DC: Author.

Arvidsson, H. (2004). *After the 1995 Swedish mental health care reform – A follow-up study of a group of severely mentally ill.* Doctoral thesis, University of Gothenburg, Gothenburg, Sweden. Retrieved from https://gupea.ub.gu.se/bitstream/2077/46/1/gupea_2077_46_1.pdf

Bartko, J. J. (1976). On varios intraclass correlation reliability coefficients. *Psychological Bulletin, 83,* 762-765.

Becker, T. & Vázquez-Barquero, J. L. (2001). The European perspective of psychiatric reform. *Acta Psychiatrica Scandinavica, 104* (suppl. 410):8-14.

Brennan, R.L. (1998). Misconceptions at the intersection of measurement theory and practice. *Educational Measurement: Issues and Practice, 17,* 5-9, 30.

Broberg, A., Almqvist, K., & Tjus, T. (2003). *Klinisk barnpsykologi: Utveckling på avvägar*

[*Clinical child psychology: Development gone astray*]. Stockholm: Natur och Kultur.

Bülow, P. (2004). *Efter mentalsjukhuset: en studie i spåren av avinstitutionaliseringen av den psykiatriska vården* [*After the mental hospital: a study in the wake of the deinstitutionalisation of psychiatric care*]. Linköpings universitet, Linköping. (University of Linkoping, Linkoping).

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimetod matrix. *Psychological Bulletin, 56,* 81-105.

Cantor, N., deSales French, R., Smith, E. E., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. *Journal of Abnormal Psychology, 89,* 181-193.

Carpenter, W. T. (2011). One hundred years. *Schizophrenia Bulletin, 37*, 443–444.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills: Sage Publications.

Cheniaux, E., Landeira-Fernandez, J., & Versiani, M. (2009). The Diagnoses of Schizophrenia, Schizoaffective Disorder, Bipolar Disorder and Unipolar Depression: Interrater Reliability and Congruence between DSM-IV and ICD-10. *Psychopathology, 42*, 293–298.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Compton, W. M. & Guze, B. G. (1995). The neo-Kraepelinian revolution in psychiatric diagnosis. *European Archives of Psychiatry and Clinical Neoroscience, 254*, 196-201.

Conric, M. (2005). The international classification for nursing practice: a tool to support nursing practice? *Collegian, 12,* 9-13.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281-302.

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine, 78*, 775-780.

Crumlish, N. & Kelly, B. D. (2009). How psychiatrists think. *Advances in psychiatric treatment, 15*, 72-79.

Curtis, S., Gesler, W., Priebe, S., & Francis, S. (2009). New spaces of inpatient care for people with mental illness: A complex 'rebirth' of the clinic? *Health & Place, 15,* 340-348.

Davies, H. T. O., & Crombie, I. K. (1997). Interpreting health outcomes. *Journal of Evaluation in Clinical Practice, 3,* 187-199.

Decker, H. S. (2007). How Kraepelinan was Kraepelin? How Kraepelinian are the neo-

Kraepelinians? – from Emil Kraepelin to DSM-III. *History of Psychiatry, 18*, 337-360.

Dunn, G. (1994). Statistical methods for measuring outcomes. *Social Psychiatry and Psychiatric Epidemiology, 29,* 198-204.

Dykes, P. C., Kim, H., Goldsmith, D. M., Choi, J., Esumi, K., & Goldberg, H. S. (2009). The Adequacy of ICNP Version 1.0 as a representational model for electronic nursing assessment documentation. *Journal of the American Medical Informatics Association, 16,* 238–46.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: a procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33,* 766-771.

Engstrom, E. J. (2007). On the question of degeneration by Emil Kraepelin. *History of Psychiatry.18*, 389-398.

Essock, S. M., Drake, R. E., Frank, R. G., & McGuire, T. M. (2003). Randomized controlled trials in evidence-based mental health care: getting the right answer to the right question. *Schizophrenia Bulletin, 29,* 115-123.

Fhanér, S. (1974). *Psykologisk testning, teori och tillämpningar*, 3rd ed. (Psychological testing, theories and applications). Stockholm: Aldus/Bonnier.

Field, A. (2009). *Discovering statistics using SPSS, 3rd edition*. London: Sage Publications.

First, M. B., Frances, A., & Pincus, H. A. (2004). *DSM-IV-TR guidebook*. Washington: American Psychiatric Publishing.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Users guide for the SCID-I Structured Clinical Interview for DSM-IV-TR Axis I disorders: Clinician version.* Washington: American Psychiatric Press.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions, 3rd edn.* New York: John Wiley & Sons.

Foucalt, M. (1988). *Madness and civilization: A history of insanity in the age of reason.* New York: Random House, Inc.

Garb, H. N., & Boyle, P. A. (2003). Understanding why some clinicians use pseudoscientific methods. In S. O. Lilienfeld, S. J. Lynn, & J. M. Lohr (Eds.), *Science and pseudoscience in clinical psychology* (pp. 17-38). New York: Guilford Press.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1,* 67-89.

Garb, H. N. (2013). Cognitive and social factors influencing clinical judgment in psychiatric

practice. *World Psychiatry, 12*, 108-110.

Gerolamo, A. M. (2004). State of the Science: Outcomes of Acute Inpatient Psychiatric Care. *Archives of Psychiatric Nursing, 18,* 203-214.

Geisinger, K.F. (1992). The metamorphosis of test validation. *Educational Psychologist, 27,* 197-222.

Gigerenzer, G, & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1,* 107-143.

Gigerenzer, G, & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62,* 451-482.

Gijsman, H. J., Geddes, J. R., Rendell, J. M., Nolen, W. A., & Goodwin, G. M. (2004). Antidepressants for Bipolar Depression: A systematic review of randomized, controlled trials. *American Journal of Psychiatry, 161*, 1537-1547.

Gilbody, S. M., House, A.O., & Sheldon, T. A. (2002). Outcomes research in mental health. *The British Journal of Psychiatry, 181,* 8-16.

Gilbody, S. M., House, A. O., & Sheldon, T. A. (2003). *Outcomes measurement in psychiatry: a critical review of outcomes measurement in psychiatric research and practice*. (Centre for Reviews and Dissemination, Report 24). York: The University of York.

Glick, I. D., Carter, W., & Tandon, R. (2003). A paradigm for treatment of inpatient Psychiatric disorders: From asylum to intensive care. *Journal of Psychiatric Practice, 9,* 395-402.

Glick, I. D., Sharfstein, S. S., & Schwartz, H. I. (2011). Inpatient psychiatric care in the 21st century: The need for reform. *Psychiatric Services, 62,* 206-209.

Goldman, H., Skodol, E., & Lave, T. B. (1992). Revising Axis V for DSM-IV: A review of measures of social functioning. *American Journal of Psychiatry, 149,* 1148-1156.

Goodwin, L.D., & Leech, N.L. (2003). The meaning of validity in the New Standards for Educational and Psychological Testing: implications for measurement courses. *Measurement and Evaluation in Counseling and Development, 36,* 181-191.

Greenberg, G. A., & Rosenheck, R. A. (2005a). Continuity of care and clinical outcomes in a national health system. *Psychiatric Services, 56,* 427- 433.

Greenberg, G. A., & Rosenheck, R. A. (2005b). Using the GAF as a national mental health outcome measure in the Department of Veterans Affairs. *Psychiatric Services, 56,* 420-426.

Grob, G. N. (1991). *From asylums to community: Mental health policy in modern America.* Princeton, N.J: Princeton University Press.

Guilford, J.P. (1946). New standards for test evaluation. *Educational and Psychological Measurement, 6,* 427-438.

Hare, E. H. (1983). Was insanity on the increase? *British Journal of Psychiatry, 142*, 439-455.

Heckers, S. (2011). Bleuler and the neurobiology of Schizophrenia. *Schizophrenia Bulletin, 37*, 1131-1135.

Hildebrand, R., McCann, E., Nelson, M., & Wass, P. (2003). Research note – Inter-rater reliability of the Global Assessment of Functioning Scale in a psychogeriatric population. *Canadian Journal on Aging, 22*, 229-233.

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R., Smith, S. R., … Holdwick, D. J. (2000). Reliability and validity of DSM-IV axis V. *American Journal of Psychiatry, 157,* 1858-1863.

Hopkins, J. E., Loeb, S. J., & Fick, D. M. (2009). Beyond satisfaction, what service users expect of inpatient mental health care: A literature review. *Journal of Psychiatric and Mental Health Nursing, 16,* 927-937.

Hälso- och sjukvårdens utvecklingsinstitut. (1994).*Verksamhetsberättelser för psykiatrin: En metod att beskriva och granska psykiatrisk verksamhet* [Annual reports within psychiatric care: A method to describe and audit psychiatric clinics]. (Spri rapport, 372). Stockholm: Spris förlag.

Iezzoni, L. I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine, 127,* 666-674.

International Council of Nurses (2008). Guidelines for ICNP catalogue development. Retrieved from http://www.icn.ch/images/stories/documents/publications/free_publications/icnp_guidelines_for_catalogue_dvt.pdf.

Jablensky, A. (2007). Living in a Kraepelinian world: Kraepelin´s impact on modern psychiatry. *History of Psychiatry.18*, 381-388.

Johnson, D. L. (2010). *A compendium of psychosocial measures: Assessment of people with serious mental illness in the Community*. New York: Springer Publishing Company.

Jones, E. D. (2003). Reminiscence therapy for older women with depression. Effects of nursing intervention classification in assisted-living long-term care. *Journal of Gerontological Nursing, 29*, 26-33.

Jones, S. H., Thornicroft, G., Coffey, M., & Dunn, G. (1995). A brief mental health outcome scale: Reliability and validity of the Global Assessment of Functioning (GAF). *British Journal of Psychiatry, 166,* 654-659.

Jovanović, A.A., Gaŝić, M.J., Ivković, M., Milovanović, S., & Damjanović, A. (2008). Reliability and validity of DSM-IV axis V scales in a clinical sample of veterans with posttraumatic stress disorder. *Psychiatria Danubina, 20*, 286-300.

Jörgensen, K. N., Römma, V., & Rundmo, T. (2009). Associations between ward atmosphere, patient satisfaction and outcome. *Journal of Psychiatric and Mental Health Nursing, 16,* 113–120.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment.. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge: Cambridge University Press.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist, 58,* 697-720.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38,* 319-342.

Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science.* San Francisco: Chandler, 1964.

Kazdin A. K. (2002). *Research design in clinical psychology*. Boston: Allyn & Bacon.

Kennedy, C. C., Madra, P., & Reddon, J. R. (1999). Assessing treatment outcome in psychogeriatric inpatients: the utility of Global Assessment of Functioning Scale. *Clinical Gerontologist, 20,* 3-11.

Kim, N. S. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology:General, 131,* 451-476

Klein, G. (2004). *The power of intuition*. New York: Doubleday.

Kraepelin, E. (2007). On the question of degeneration. *History of Psychiatry. 18*, 399-404.

Loevdahl, H., & Friis, S. (1996). Routine evaluation of mental health: reliable information or worthless "guesstimates"? *Acta Psychiatrica Scandinavica, 93*, 125-128.

Lorant, V., Kampfl, D., Seghers, A., Deliège, D., Closon, M. C., & Ansseau, M. (2003). Socio-economic differences in psychiatric in-patient care. *Acta Psychiatrica*

Scandinavica, 107, 170–177.

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

Madden, R., Napel, H.T., & Cumerlato, M. (2011). Annual report of the Family Development Committee. Retrieved from http://www.who.int/classifications/network/WHOFIC2011_C005_FDCannual_report2010_2011.pdf.

Madden, R., Sykes C., & Ustun T. B. (2007). World Health Organization Family of International Classifications: definition, scope and purpose. Retrieved from http://www.who.int/classifications/en/FamilyDocument2007.pdf?ua=1.

Marks, I. (1998). Overcoming obstacles to routine outcome measurement. *British Journal of Psychiatry, 173,* 281-286.

McGlashan, T. H. (2011). Eugen Bleuler: Centennial anniversary of his 1911 publication of dementia praecox or the group of Schizophrenias. *Schizophrenia Bulletin, 37*, 1101-1103.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35,* 1012-1027.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp13-103). New York: American Council on Education and Macmillan.

Miller, P. R., Dasher, R., Collins, R., Griffiths, P., & Brown, F. (2001). Inpatient diagnostic assessments: 1. Accuracy of structured vs. unstructured interviews. *Psychiatry Research, 105,* 255-264.

Moos, R. H., Nichol, A. C., & Moos, B. S. (2002). Global Assessment of Functioning ratings and the allocation and outcomes of mental health services. *Psychiatric Services, 53,* 730-737.

Moos, R. H., McCoy, L., & Moos, B. S. (2000). Global Assessment of Functioning (GAF) ratings: determinants and role as predictors of one-year treatment outcomes. *Journal of Clinical Psychology, 56,* 449-461.

National Centre for Classification in Health. International Classification of Health Interventions (ICHI) Alpha 2, Version 2013. Retrieved from http://sydney.edu.au/health-sciences/ncch/resources.shtml.

Nationella kvalitetsregister. [National quality registries]. (2014). Retrieved from
    www.kvalitetsregister.se.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory,* 3rd ed. New York: McGraw-
    Hill.

Pedersen, G., & Karterud, S. (2012). The symptom and function dimensions of the Global
    Assessment of Functioning (GAF) scale. *Comprehensive Psychiatry, 53,* 292-298.

Pedersen, G., Hagtvet, K. A., & Karterud, S. (2007).Generalizability studies of the Global
    Assessment of Functioning-Split version. *Comprehensive Psychiatry, 48,* 88-94.

Pedersen, G., & Karterud, S. (2007). Associations between patient characteristics and ratings
    of treatment milieu. *Nordic Journal of Psychiatry, 61*, 271-278.

Phelan, M., Wykes, T., & Goldman, H. (1994). Global function scales. *Social Psychiatry and
    Psychiatric Epidemiology, 29,* 205-211.

Piersma, H. L., & Boes, J. L. (1997). The GAF and psychiatric outcome: A descriptive report.
    Community Mental Health Journal, 33, 35-41.

Prior, L. (1991). Community versus hospital care: the crisis in psychiatric provision. *Social
    Science and Medicine, 32*:483-489.

Registercentrum Västra Götaland. (2014). Portal för psykiatriska kvalitetsregister [Portal to
    psychiatric quality registries]. Retrieved from www.psykiatriregister.se

Rabinowitz, J., Modai, I., & Inbar-Saban, N. (1994). Understanding who improves after
    psychiatric hospitalization. *Acta Psychiatrica Scandinavica, 89,* 152-158.

Ramirez, R., Ekselius, L., & Ramklint, M. (2008). Axis V – Global Assessment of
    Functioning Scale (GAF), further evaluation of the self-report version. *European
    Psychiatry, 23,* 575-579.

Ramon, S. (1996). *Mental health in Europe: Ends beginnings and rediscoveries*. Houndmills:
    Macmillan Press.

Regeringens proposition 1993/1994:218. Psykiskt stördas villkor [The conditions of severely
    mentally ill]. Retrieved from http://www.riksdagen.se/sv/Dokument-
    Lagar/Forslag/Propositioner-och-skrivelser/Psykiskt-stordas-villkor_GH03218/?text=true

Rey, J. M., Starling, J., Wever, C., Dossetor, D. R., & Plapp, J. M. (1995). Inter-Rater
    reliability of Global Assessment of Functioning in a clinical setting. *Journal of Child
    Psychology and Psychiatry, 36*, 787-792.

Rodrigues, J-M. (2012). Is the international classification of healthcare procedures (ICHI) a
    critical point for the implementation of an international Casemix grouper. *BMC Health*

Services Research, 2012, 12 (Suppl 1):I1.

Roy-Byrne, P., Dagadakis, C., Unutzer, J., & Ries, R. (1996). Evidence for limited validity of the revised Global Assessment of Functioning Scale. *Psychiatric Services, 47,* 864-866.

Sharfstein, S. S. (2009). Goals of inpatient treatment for psychiatric disorders. *Annual Review of Medicine, 60,* 393-403.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin, 86,* 420-428.

Skodol, A. E., Link, B. G., Shrout, P. E., & Horwath, E. (1988). The revision of Axis V in DSM-III-R: Should symptoms have been included? *American Journal of Psychiatry, 145,* 825-829.

Slade, M., Thornicroft, G., & Glover, G. (1999). The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology, 34,* 243-249.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397-420). Cambridge: Cambridge University Press.

Smedby, B., & Schioler, G. (2006). Nordisk Medicinalstatistisk Komité, Kobenhavn. Health classifications in the Nordic countries, historic development in a national and international perspective 2006. Retrieved from http://nowbase.org/~/media/Projekt%20sites/Nowbase/Publikationer/Andre/Health%20Cl assifications%20in%20the%20Nordic%20Countries.ashx.

Smith, G. R., Manderscheid, R. W., Flynn, L. M., & Steinwachs, D. M. (1997). Principles for assessment of patient outcomes in mental health care. *Psychiatric Services, 48,* 1033-1036.

Socialstyrelsen. (2003). Utredning av händelserna i Åkeshov och Gamla stan och dess möjliga samband med brister i bemötande och behandling inom den psykiatriska vården och socialtjänstens verksamhet [Investigation of events in Åkeshov and the Old Town and its' possible association with deficiencies in response and treatment within psychiatric care and social services]. Retrieved from http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/10593/2003-107-13_200310713.pdf

Socialstyrelsen. (2006). KVÅ – vilka åtgårder ska kodas och registreras (2006)?[KVÅ- What interventions should be coded and registered]. Retrieved from

http://www.socialstyrelsen.se/klassificeringochkoder/Documents/KVA_Vadskareg_0611
10.pdf.

Socialstyrelsen. (2009). Klassifikation av vårdåtgärder (KVÅ) – bakgrund och förklarande
information [Intervention classification in health care – background and explanatory
information]. Retrieved from
http://www.socialstyrelsen.se/klassificeringochkoder/Documents/KVA-
inledn_2007_v2.6.pdf.

Socialstyrelsen. (2013). Åtgärdskoder KVÅ: KVÅ 2013 – Klassifikation av vårdåtgärder
Inklusive Beskrivningstexter i Excel- och Access-format, KVÅ_130101 [Intervention
codes KVÅ, KVÅ 2013 – Classification of Health Interventions including description
texts in Excel- and Access-format, KVÅ_130101]. Retrieved from
http://www.socialstyrelsen.se/klassificeringochkoder/atgardskoderkva#2.

Spitzer, R. & Sheehy, M. (1976). DSM III: A classification system in development.
Psychiatric annals, 6, 102-109.

Startup, M., Jackson, M. C., & Bendix, S. (2002). The concurrent validity of the Global
Assessment of Functioning (GAF). *British Journal of Clinical Psychology, 41,* 417-422.

Stephanie, P. (2013). More than liberator of the insane. *Journal of Humanistic Psychiatry*, *1*,
12-13.

Steinert, T., & Schmid, P. (2004). Effect of Voluntariness of Participation in Treatment on
Short-Term Outcome of Inpatients with Schizophrenia. *Psychiatric Services, 55,* 786-
791.

Suen, H. K. (1988). Agreement, reliability, accuracy, and validity: toward a clarification.
Behavioral Assessment, 10, 343-366.

Sveriges Kommuner och Landsting. 2010. *Kartläggning av den psykiatriska heldygnsvården*
[*Mapping of inpatient psychiatric care*]. Stockholm: Author.

Söderberg, P., Tungström, S., & Armelius, B-Å. (2004). *GAF-skalans reliabilitet i kliniskt
arbete, forskningsrapport nr 6* [*Reliability of the GAF scale in clinical practice*].
Umeå universitet, Institutionen för psykologi.

Söderberg, P., Tungström, S., & Armelius, B.Å. (2005). Reliability of Global Assessment of
Functioning ratings made by clinical psychiatric staff. *Psychiatric Services, 56,* 434-438.

The University of Iowa, College of Nursing (2013). Center for Nursing Classification and
Clinical Effectiveness (CNC). Retrieved from

http://www.nursing.uiowa.edu/center-for-nursing-classification-and-clinical-effectiveness.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124-1131.

Waddinton, L, & Morley, S. (2000). Availability bias in clinical formulation: the first idea that comes to mind. *British Journal of Medical Psychology, 73*, 117-127.

Wallace, T., O'Connell, S., & Frisch, S. R. (2005). What do nurses do when they take to the streets? An analysis of psychiatric and mental health nursing interventions in the community. *Community Mental Health Journal, 41,* 481-96.

Wallsten, T., Kjellin, L., & Lindström, L. (2006). Short-term outcome of inpatient psychiatric care: Impact of coercion and treatment characteristics. *Social Psychiatry and Psychiatric Epidemiology,* 41, 975-80.

Walter, G., Cleary, M., & Rey, J. M. (1998). Attitudes of mental health personnel towards rating outcome. *Journal of Quality in Clinical Practice, 18,* 109-115.

Vatnaland, T., Vatnaland, J., Friis, S., & Opjordsmoen, S. (2007). Are GAF scores reliable in routine clinical use? *Acta Psychiatrica Scandinavica, 115,* 326-330.

Västra Götalandsregionen. (2014). Regionala åtgärdsförteckningar KVÅ, Psykiatri 2014 DefOrg [Regional interventions codelist KVA, Psychiatry 2014 DefOrg]. Retrieved from http://www.vgregion.se/sv/Vastra-Gotalandsregionen/startsida/Vard-och-halsa/Sa-styrs-varden/Halso--och-sjukvardsavdelningen/Klassifikationer-och-ersattningsmodeller/Atgardsforteckningar-KVA-Vastra-Gotaland/

Västra Götalandsregionen, Medicinska sektorsrådet psykiatri. (2006). *Vertikal prioritering inom psykiatrin, Västra Götalandsregionen* [*Vertical priority in psychiatry, Region Västra Götaland*]. Retrieved from http://www.vgregion.se/upload/Regionkanslierna/HSA/PrioRev%20vers%2020070104.pdf

Västra Götalandsregionen, Hälso- och sjukvårdsavdelningen. (2009). *Regional förteckning över kvalitetsindikatorer för hälso- och sjukvård* [*Regional catalogue of quality indicators regarding health and medical care*]. (Dnr RSK 769-2007). Retrieved from http://www.vgregion.se/upload/Regionkanslierna/hsskansli/Analys/Rapporter/2009/Regional%20f%C3%B6rteckning%20%C3%B6ver%20kvalitetsindikatorer%20f%C3%B6r%20h%C3%A4lso-%20och%20sjukv%C3%A5rd.pdf

Weiner, D. (1992). Philippe Pinel's "Memoir on madness" of December 11, 1794: a fundamental text of modern psychiatry. *The American Journal of Psychiatry, 149,* 725-

732.

Wells, K. B. (1999). Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. *American Journal of Psychiatry, 156,* 5-10.

Whitney, J. A., Kunik, M. E., Molinari, V., Lopez, F. G., & Karner, T. (2004). Psychological predictors of admission and discharge global assessment of functioning scale scores for geropsychiatric inpatients. *Aging and Mental Health, 8,* 505-513.

WHO Collaborating Centre for the FIC in The Netherlands. WHO Family of International Classifications (FIC), Newsletter, volume 10, number 1, 2012. Retrieved from http://www.rivm.nl/who-fic/newsletter/newsletter%202012-1.pdf.

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious.* Cambridge, Massachusetts: Harvard University Press.

World Health Organisation. (1992). *The ICD-10 Classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: Author.

World Health Organization. (2001). *ICF: International classification of functioning, disability and health.* Geneva: Author.

World Health Organisation. (2014). *WHO Disability Assessment Schedule 2.0 WHODAS 2.0.* Retrieved from http://www.who.int/classifications/icf/whodasii/en/.

Global Assessment of Functioning (GAF) Scale

Consider the client's psychological, social, and occupational functioning on a hypothetical continuum (1-100) of mental health-illness. Do not include impairment in functioning due to physical (or environmental) limitations. Assessment is of client's functioning during the previous 12 months.

CODE (Note: Use intermediate codes when appropriate, e.g. 45, 68, 72.)

91-100	**Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his or her many positive qualities. No symptoms**
81-90	**Absent or minimal symptoms** (e.g., mild anxiety before an exam), **good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more that everyday problems or concerns** (e.g., an occasional argument with family members).
71-80	**If symptoms are present, they are transient and expectable reactions to psychosocial stressors** (e.g., difficulty concentrating after family arguments); **no more than slight impairment in social, occupational, or school functioning.** (e.g., temporarily falling behind in schoolwork).
61-70	**Some mild symptoms** (e.g., depressed mood and mild insomnia) **OR some difficulty in social, occupational, or school functioning** (e.g. occasional truancy, or theft within the household), **but generally functioning pretty well, has some meaningful interpersonal relationships.**
51-60	**Moderate symptoms** (e.g., flat affect and circumstantial speech, occasional panic attacks) **OR moderate difficulty in social, occupational or school functioning** (e.g., few friends, conflicts with peers or co-workers).
41-50	**Serious symptoms** (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) **OR any serious impairment in social, occupational, or school functioning** (e.g., no friends, unable to keep a job).
31-40	**Some impairment in reality testing or communication** (e.g., speech is at times illogical, obscure, or irrelevant) **OR major impairment in several areas, such as work or school, family relations, judgement, thinking or mood** (e.g., depressed person avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).
21-30	**Behavior is considerably influenced by delusions or hallucinations OR serious impairment in communication or judgement** (e.g. sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) **OR inability to function in almost all areas** (e.g., stays in bed all day; no job, home or friends).
11-20	**Some danger of hurting self or others** (e.g., suicide attempts without clear expectation of death; frequently violent; manic excitement) **OR occasionally fails to maintain minimal personal hygiene** (e.g. smears feces) **OR gross impairment in communication** (e.g., largely incoherent or mute).
1-10	**Persistent danger of severely hurting self or others** (e.g., recurrent violence) **OR persistent inability to maintain minimal personal hygiene OR serious suicidal act with clear expectation of death.**
0	Inadequate information.

Adapted from the *Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision* (2000) American Psychiatric Association, Washington DC.

CUCS Housing Resource Center, 198 E. 121st Street, 6th Floor, New York, NY 10035, Phone: (212) 801-3333, Fax: (212) 803-5880
Revised 10/08

78