Tools for Outcome-informed management of mental illness

Psychometric properties of instruments of the Swedish clinical multicenter Quality Star cohort
Tools for Outcome-informed management of mental illness

Psychometric properties of instruments of the Swedish clinical multicenter Quality Star cohort
Bo Ivarsson. *Tools for Outcome-informed management of mental illness — Psychometric properties of instruments of the Swedish clinical multicenter Quality Star cohort*

**DISSERTATION**

Karlstad University Studies 2011:34  
ISSN 1403-8099  

© The author

Distribution:  
Karlstad university  
Faculty of Economic Sciences, Communication and IT  
Psychology with an emphasis on medical psychology  
S-651 88 Karlstad  
Sweden  
+46 54 700 10 00

www.kau.se

Print: Universitetstryckeriet, Karlstad 2011
To All Participants and Collaborators in the Quality Star network.

Bo Ivarsson, Department of Psychology, Karlstad University, Sweden.

Abstract

The purpose of this thesis was to investigate the psychometric properties of three global user subjective measures of the "The Quality Star" clinical review model: Consumer Satisfaction Scale, Global Quality of Life scale, and Perceived Global Distress scale. The mental health implementation context of this review model emphasizes the client as an agent of change, taking part in shared decision making in an empowered role as collaborative partner to the professional clinicians. In Paper I study the patient self-rating Consumer Satisfaction Scale gave results comparable to those obtained by independent interviewer assessors. Out of cost-effective perspective professional time is saved and logistics simplified. In Paper II the visual analogue self-rating Global Quality of Life scale was shown to have satisfactory test-retest reliability, and concurrent validity with the “Life as a whole” item of Manchester Short Assessment of Quality of Life (MANSA). The patients’ conceptualizations of the scale based on associative findings with a number of validating instruments were consistent with expected areas of concern for Serious Mentally Ill persons. Similarly, in Paper III the visual analogue scale the Perceived Global Distress scale, showed acceptable clinical test-retest reliability, and concurrent validity with the MANSA item, “How satisfied are you with your mental health”. In associative analyses it was found that depressive, anxiety, interpersonal and existential elements contributed to the patient’s conceptualization of the construct. In Paper IV, a previous finding suggesting that women were more satisfied with the health care and had better social functioning compared to men was further elaborated investigating the discriminative properties of the subjective instruments. In the multi-centre cohort of 2552 patients it was possible to detect differences between genders and functional levels professionally assessed with the split version of Global Assessment of Functioning rating scale. The General discussion underlines that although subjective measures tend to have strong interrelations, supporting earlier findings, one has to use multiple measures for an optimal management of mental illness as the subjective outcome ratings have to be
individually interpreted in a feed-back dialogue with the patient and be compared to observational assessments.

**Keywords:** Consumer satisfaction, Quality of Life, Perceived distress, Severe mental illness, Schizophrenia, Social function, Global Assessment of Functioning, MANSA, Gender
Doktorsavhandling: Verktyg för resultatstyrda hantering av psykisk sjukdom, vägleder av brukarsjälvskattningar. Psykometriska egenskaper för instrument i den svenska kliniska multicenter-kohorten Kvalitetsstjärnan.

Bo Ivarsson, Avdelningen för psykologi, Karlstad Universitet, Sverige.

Sammanfattning

Syftet med denna avhandling var att undersöka de psykometriska egenskaperna hos tre globala patientsubjektiva instrument ur den kliniska uppföljningsmodellen "Kvalitetsstjärnan": Brukartillfredsställelseskalan, Globala livskvalitetsskalan och Globala skalan för upplevda besvär. Vid uppföljningsmodellens användning i psykiatrisk vård betonas synen på brukaren som förändringsaktör, som deltagare i gemensamt beslutsfattande i en ”empowered” roll och som samarbetspartner med de professionella i vården. Studien i Paper I visade att självskattningsversionen av Brukartillfredsställelseskalan gav resultat jämförbara med de som erhålls genom oberoende bedömare. Ur kostnads-effektivitetssynpunkt sparas tid och förenklas genomförandet. I Paper II visades att den visuella analogskalan Global livskvalitet hade tillfredsställande test-retest reliabilitet och samtidig validitet med item ”Hur tillfredsställd är du i dag med ditt liv som helhet?” från Manchester Livskvalitetsskala (MANSA). Patienternas konceptuella förståelse av skalan var i överensstämmelse med förväntade angelägna områden för allvarligt psykiskt sjuka, indikerat av funna associationer med ett flertal valideringsinstrument. På liknande sätt, visades i Paper III att den visuella analogskalan Upplevda globala besvär hade acceptabel test-retest reliabilitet och samtidig validitet med item ”Hur tillfredsställd är du med din psykiska hälsa?” ur MANSA. Genom associativa analyser befanns att depressivitet, ångest, interpersonella och existentiella element bidrog till patienten konceptualisering av konstruktet. I Paper IV genomfördes en fördjupad studie beträffande tidigare fynd som tytt på att kvinnor var nöjdare med sin vård och att de hade bättre social funktion jämfört med män där de diskriminerande egenskaperna hos de subjektiva instrumenten undersökes. I kohorten med 2552 patienter var det möjligt att finna genderskillnader och skillnader mellan professionellt skattade funktionsnivåer baserat på split-versionen av GAF. I den sammanfattande diskussionen underströks att även om subjektiva mått tenderade att ha starka interrelationer, som bekräftar tidigare fynd, måste man ändå använda flera mått vid handläggning av psykisk sjukdom på ett optimalt sätt, eftersom subjektiva skattnings av
resultat måste tolkas på ett individuellt sätt i återkopplingsdialog med brukare och jämföras med skattningar baserade på observation.

**Nyckelord:** Brukartillfredsställelse, Livskvalitet, Upplevda besvär, Allvarlig psykisk sjukdom, Schizofreni, Social funktion, Global skattning av funktionsförmåga, MANSA, Gender
The thesis is based on the following four research papers, which will be referred to in the text by their Roman numerals:


Reprints were made with kind permission from the publishers.
Acknowledgements

My Supervisor and expert advisors. I would like to express my deep gratitude and great admiration to my examiner and supervisor Professor Torsten Norlander, Karlstad University for the way you have guided me with kindness, optimism, an extraordinary talent to recharge my batteries when needed, with nice psychological and pedagogical tricks. Assistant professor Ulf Malm, Gothenburg University, this thesis had never been written without the support of your extraordinary ability to make people believe they can, always with a warm heart. Professor Leif Lindström, Uppsala university, your expertise guidance along the road has been a solid support and assurance. To all three of you, thanks for cooperation with the original papers for this thesis. Professor Merete Nordentoft, Faculty of Health Sciences, University of Copenhagen and Ph. Dr., Psychologist Stefan Tungström, Psychiatric services, Dalarna County, thanks for constructive criticism of thesis drafts.

In memoriam.
Professor Ian Falloon, my thoughts goes to you with thanks for cooperation during the Optimal Treatment Project and lessons learnt from you that have been essential for the my work. Psychiatric Specialist nurse Lars Kärräng, Svenljunga district psychiatric team, who in practice did the hard work to start implementation of outcome informed praxis in my own clinical world and helped me understand what happens when theory meets reality.

Collegues in the Quality Star network and Your participating Clients.
Thanks for many years of collective enthusiastic and creative cooperation in development of the follow-up method.
My wish is that you all would regard this thesis as Yours, only written by me.

Support from Employers. Thanks goes to Assistant hospital director Suzanne Guregård, to Lise-Lotte Risö-Bergerlind and Bengt-Arne Andersson, successive Heads of the Psychiatric Department of South Älvsborg hospital, Borås for your always inspiring attitude and generous support.

Dear Monika, my wife. Thanks for patience with my neglect of you in periods of work.
## CONTENTS

1 INTRODUCTION 1

1.1 Background to the main source of data – The Quality Star 1

1.1.1 Development of the Quality Star concept 1

1.1.2 Implementation of the Quality Star 6

1.2 Theoretical perspectives 9

1.2.1 Outcome domains in psychiatry 9

1.2.2 Consumer satisfaction within mental health service 15

1.2.2.1 Conceptual considerations 15

1.2.2.2 Consumer Satisfaction measure rationale in present work 17

1.2.3 Quality of life in psychiatry 17

1.2.3.1 Conceptual considerations 17

1.2.3.2 Quality of Life measure rationale in present work 20

1.2.4 Mental distress in serious mental illness 21

1.2.4.1 Conceptual considerations 21

1.2.4.2 “Mental distress” measure rationale in present work 22

1.2.5 Associations between The Subjective and The Objective 23

1.2.5.1 Choices within the Quality Star concept 25

2 THE PRESENT INVESTIGATION 27

2.1 Introduction 27

2.1.1 The purpose of the current thesis 27

2.1.2 Procedures for the Quality Star instrument 27

2.1.3 The four articles contained in the thesis 28

2.2 Paper I. Psychometric properties of the Consumer Satisfaction Rating Scale 29

2.2.1 Aim 29

2.2.2 Design 30

2.2.3 Instruments 30

2.2.4 Procedure 32

2.2.5 Statistics 34

2.2.6 Results 34

2.2.7 Addendum 35

2.3 Paper II. Psychometric properties of the Global Quality of Life Scale 36

2.3.1 Aim 36

2.3.2 Design 36

2.3.3 Instruments 38

2.3.4 Procedure 41

2.3.5 Statistics 42

2.3.6 Results 43

2.4 Paper III. Psychometric properties of the Perceived Global Distress Scale 45

2.4.1 Aim 45

2.4.2 Design 45

2.4.3 Instruments 46

2.4.4 Procedure 49

2.4.5 Statistics 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.6 Results</td>
<td>50</td>
</tr>
<tr>
<td>2.5 Paper IV. Levels of Social Function and patients subjective Measures</td>
<td>51</td>
</tr>
<tr>
<td>2.5.1 Aim</td>
<td>51</td>
</tr>
<tr>
<td>2.5.2 Design</td>
<td>51</td>
</tr>
<tr>
<td>2.5.3 Instruments</td>
<td>52</td>
</tr>
<tr>
<td>2.5.4 Procedure</td>
<td>53</td>
</tr>
<tr>
<td>2.5.5 Statistics</td>
<td>53</td>
</tr>
<tr>
<td>2.5.6 Results</td>
<td>54</td>
</tr>
<tr>
<td>3 GENERAL DISCUSSION</td>
<td>55</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>55</td>
</tr>
<tr>
<td>3.2 Validity of patient subjective instruments of Quality Star supported</td>
<td>55</td>
</tr>
<tr>
<td>3.2.1 The self-rating version of the Consumer Satisfaction Rating Scale (Pat-UKU-ConSat) (Paper I)</td>
<td>57</td>
</tr>
<tr>
<td>3.2.2 Subjective Global Quality of Life Scale (GQL) (Paper II)</td>
<td>59</td>
</tr>
<tr>
<td>3.2.3 The Perceived Global Distress Scale (PGD) (Paper III)</td>
<td>61</td>
</tr>
<tr>
<td>3.3 Subjective ratings, differences by gender and function (Paper IV)</td>
<td>63</td>
</tr>
<tr>
<td>3.4 Working together using structured instruments</td>
<td>66</td>
</tr>
<tr>
<td>3.5 Future research and development</td>
<td>68</td>
</tr>
<tr>
<td>4 CONCLUSIONS</td>
<td>69</td>
</tr>
<tr>
<td>5 REFERENCES</td>
<td>71</td>
</tr>
<tr>
<td>6 APPENDIX</td>
<td>85</td>
</tr>
</tbody>
</table>

PAPER I
PAPER II
PAPER III
PAPER IV
1. INTRODUCTION

This thesis is devoted to investigating the psychometric properties of the three main global patient subjective instruments used in a clinical follow-up model used in Sweden, "The Quality Star. The introduction describe the development of the follow-up model (section 1.1.) and the theoretical perspectives to the patient subjective instruments in focus (section 1.2).

1.1. Background to the main source of data – The Quality Star

In this section the development of the of the Quality Star review model is described, in order to give a background to the conditions under which the main data for the present naturalistic studies were generated.

1.1.1 Development of the Quality Star concept

The Swedish National Patient Register (NPR) at the National Board of Health and Welfare in Sweden so far does not have patient individual level outcome measures (National Board of Health and Welfare, 2011), such as for instance the Health of the National Outcome Scales used in the UK (Wing et al., 1998) and Australia (Burgess, Trauer, Coombs, McKay, & Pirkis, 2008), or the Global Assessment of Functioning (GAF) in the US (Greenberg, & Rosenheck, 2005).

Over the years, models for follow-up on reviewing outcomes and treatment monitoring, which could inform the future care provided to individual patients as well as providing aggregated data for benchmark have been developed in Sweden but no model has attained general usage throughout the country or for any longer time.

The community based mental health services (“district psychiatry”) established during the reform of the hospital based psychiatry during the 80ties was at several places accompanied by the use of local patient registers following
international the examples (for instance in Mannheim and Verona). The most prominent example in Sweden was created at Nacka, Stockholm (Spri 1988), though structure and process data was dominating and patient information did not include severity ratings or subjective measures. The Diagnostic and Statistical Manual (DSM) system, with its GAF scale, was introduced to clinicians in about the same time, and created awareness that GAF might serve on patient level as well as for group comparisons, but nation-wide use was not established. In the early 90ties a comprehensive model for description of contents, resource use, effects and quality of psychiatric care was elaborated in a national project (Blom et al., 1994). The model included GAF and paid attention to the integration of service and care, but did not include user measures in the data set. The Swedish Psychiatric Association together with major national stakeholders tried to support the development though a quality indicator project in the mid 90ties, where again GAF was suggested as an indicator. In addition time was ripe to suggest that amongst the suggested 6 indicators participation of relatives in treatment and written treatment plan were recommended, indicating that user participation and dialogue were recognized as important areas for follow-up (Spri 1997). By the turn of the century a county Key Figure Network (sv.: Nyckeltalsamarbetet) performed yearly period studies where GAF as a measure of severity was the central measure to which most other patient and service items were related (Candefjord & Larsson, 2005). All these historical efforts were basically generic in nature and mainly driven by a focus to monitor implementation at large of community based psychiatry.

Eventually, insight grew that such monitoring did have weaknesses and needed to be geared also by evidence-based principles, and for instance include elements to follow effects of specified interventions, or monitor implementation of state-of-the-art guidelines for specified need groups.

One such development was the process that ended in the construction of the Quality Star. In about 1990, after a decade of implementation of community base district psychiatry it was noted that cooperation and integration between service and caregivers often was not optimal and that state-of-the art strategies and treatment was not fully implemented. This was the case also in the small country side area,
Svenljunga, where the author was at that time working. In looking for answers to overcome the problems, the integrated family and community based service in Buckingham, a the small town area outside London, was seen as a suitable example to learn from (Falloon & Fadden, 1993). Staff was educated and a three year project was set-up. The model included the systematic use of problem inventories and follow-up forms, that were interactively used together with the users. Modifications to suit the Swedish context were done, in collaboration with university in Gothenburg and other specialists that had before studied the Buckingham model were consulted. The change accomplished was positive (Kärräng 1995) and confirmed by experience from Dr. Malm and his Gothenburg team who were developing their service in similar ways.

During these years the head of the Buckingham project, professor Ian Falloon, organized an international study, “Optimal treatment project” (Falloon et al., 2004) to further confirm the model and validate it in different countries, including quarterly ratings of outcomes and monitoring progress by life charts (Falloon & Fadden p 132). The Gothenburg and Svenljunga services become involved together with a newly started service at Lysekil, north of Gothenburg.

The total quality assurance package derived from the Integrated Care concept was 1996 presented by Ivarsson, Malm, and Swärd (1995), after which the regional political and administrative level expressed the wish that the professionals would try to implement its outcome elements, i. e the Quality Star (see graphic illustration in Fig. 1).

The positioning of the Star rays has didactic meanings. Consumer satisfaction is on top to remind that as a rule a satisfied consumer will benefit from treatment. At the horizontal axis to the left the professional assessment of symptom severity has a “closeness” to perceived distress in the NW direction. The professional assessment of psychosocial functioning on the horizontal axis to the right of has relation to subjective quality of life in the NE direction. The importance of considering important/significant others is represented by the SE axis. Particularly important domain(s) in various patient groups has got its place on the axis to the SW.
The outcome measures in the Star are “carried” by the downward pointing resource use axis.

* Group specific dimension
  (Generic: SERS)
  (Schizophrenia: RS-S)

---

* This axis is intended for group specific extra measures when suitable.
  Health screening (SERS) is recommended for all users.
  For psychosis groups, Remission index (RS-S) is now being introduced.

**Figure 1.** Graphic representation of global instruments in “The Quality Star” concept. Measures used for each dimension in parentheses. In addition to chosen quality dimensions, a representation of resource use is included in the bottom of the “star” by number of “days in care”, later to be specified by national treatment content codes. (Adapted from Paper IV, Ivarsson, Malm, Lindström & Norlander, 2011).

The conceptual relationship between contents in the Quality Star with conceptualizations from other authors will be briefly elaborated in Sections 1.2.1. “Outcome domains in psychiatry” and in Sections 1.2.5.1 “The broad picture” and 1.2.5.2 “Choices within the Quality Star concept”.

In summary, the conceptual development of The Quality Star aiming at empowering client, the framework of family and social network in mental health is in accordance with the disability ideological model described by Maycraft Kall (2011) and summarized in Table 1. This approach is suitable, as earlier gloomy prognostic views on serious mental illness, especially schizophrenia, have turned more optimistic with long term follow up findings in the Vermont study and elsewhere (Bleuler, 1974; Ciompi, 1988; DeSisto, Harding, McCormick, Ashikaga & Brooks, 1995; Harding, Brooks, Ashikaga, Strauss, & Breier, 1987; Huber, Gross, &
Schuttler, 1975; Tsuang, Woolson, & Fleming, 1979), and works to further clarify the nature of recovery and remission (see for instance Andreasen et al., 2005; Helldin, 2009; Liberman, and Kopelowicz, 2005; Van Os, et al., 2006). This has been accompanied by improved integrated treatment and care service strategies, were in addition to medication, components like tailored specific psychological interventions, social skills training, family involvement and education have contributed. This has been supported by, on the structural level, development of community based multi-professionalism and improved case management, which in turn demands practical information handling with continuous monitoring. These findings and developments during the last decades explain emerging interest and usage of user-friendly holistic follow-up methods, where the Quality Star can serve as an example of such efforts.

Table 1. Three mental health care ideologies (Based on Wendy Maycraft Kall, 2011)

<table>
<thead>
<tr>
<th></th>
<th>The medical model</th>
<th>The disability model</th>
<th>Risk-management model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td>Emphasizing mental health as an illness requiring treatment</td>
<td>Based on normalization concepts where users are perceived as citizens making own decisions</td>
<td>Perception of users as a danger and treat to public safety</td>
</tr>
<tr>
<td><strong>Policy and service</strong></td>
<td>Based on hospitals, evidence-based drug treatments and psychological therapies with doctors as key decision makers</td>
<td>Arguing for user empowerment, autonomy and enforcing citizen rights with a long-term objective of the user controlling her/his illness</td>
<td>Perception of social problems in terms of risk and reduced solidarity</td>
</tr>
<tr>
<td><strong>Frames of users</strong></td>
<td>In terms of their diagnosed mental illness. Accords users a passive role as 'patients'</td>
<td>Individualised and tailored with a coordinated support network of public, voluntary and private actors to meet users’ individual priorities and preferences. User-generated support and services</td>
<td>Emphasises steering strategies and increased control strategies. Perception that risks can be managed and focuses on ideas that risks are caused and failure to control means someone is to blame.</td>
</tr>
</tbody>
</table>

The Quality Star concept also have similarities with views advocated by Duncan, Miller, Wampold, and Hubble (2010), with strong emphasis on the client as the agent of change taking part in shared decision making in an empowered role (‘the heroic
client' role) (Duncan, Miller, & Sparks, 2007) as collaborative partner to the professional clinicians. To promote empowerment of the role of the patient and her/his social network, systematic procedures for shared decision-making by the “assessment-planning-intervening-reviewing” revolving cycle take place in the Quality Star service user-provider feedback teamwork.

The heads of the five psychiatric departments in Western Sweden decided to evaluate the instruments and make field test, which was carried out 1997-1999.

1.1.2 Implementation of the Quality Star

After developing the “The Quality Star concept” for clinical follow-up and field testing it as described in section 1.1.1 a decision was taken by the council of psychiatry in West Sweden to implement the model in routine care from the year 2000. All participating centers had a one-day education in the use of the rating scales. About every six months additional education days were staged to which new staff could be sent for their education. Emphasis was put on the two aspects of the concept:

In the first place, the (at least) yearly ratings, should serve to have an annual review of the treatment received together with the patient. The recommendation was always to find a local format for using the review as a means in treatment and care planning.

Secondly, the ratings were to be used to establish a local database and subsequently to aggregate these for all participating centers to allow benchmark. Software was developed in Access.

Already in the first year other parts of Sweden become interested. St Göran hospital, Stockholm decided to join as an initiative partner. Others followed suite. Regular meetings to which all centers were invited were arranged, initially about every second month were local responsible persons, and above all the local enthusiasts, had an opportunity to share experiences. This was soon established as a forum for common decisions for the use of the model and provided a forum for
sharing experience. It got the name “GGG” (short acronym for Common Implementation Group in Swedish).

In the first years, focus was almost solely on establishing the model as a means to include patients in dialogue about their situation and needs, and to suggest its use in connection with care planning, giving it a supportive function for shared decision making in the ordinary work. It should be remembered that the overhead intention of “the Quality Star” concept was to promote that patient subjective perspective was to be strengthened in care and treatment praxis and to, in a tangible way, demonstrate that global user perspective instruments were possible to use and record. At this time no such routine standard documentation was at all used in Sweden. This was a main mission, to start a process in every-day praxis aimed to create a minimal generic platform for follow-up of psychiatric care in a multi-dimensional, holistic perspective.

Ethics and integrity was frequently discussed and experience shared regarding the use of the model in dialogue with users. The GGG meetings usually have had an important initial point on the agendas, “Around the country”, where centers have the possibility to lift problem encountered for collegial discussion – apart from reporting advances. The general advise have been that if user or personnel in a specific case have doubts about using an instrument in a particular situation, for ethical reasons, the instrument should not be used in a forced way. Instead the question should be raised if and how the dialogue with the user can be developed so that user subjective perceptions trustfully can be reviewed together. A typical example has been intimate relations and sex. Though this is repeatedly found to be an area of outmost concern for users, there have been frequent discussions in workplaces about if and how to bring this up with patients. The use of the Quality Star has had the effect to be a lever to bring such discussions to the surface.

Only later, emphasis was put on the second purpose of the Quality Star, to establish a database. Initially most centers were reluctant to admit open identification of their sites in the accumulated reports. It took a few years until trust in this respect grew. Subsequently the time got ripe for a consensus decision to henceforth appear identified by names of the centers in reports. This made yet other
centers interested to use the concept. In this way, step by step, the purpose of the
Quality Star to demonstrate and provide psychiatry with easy, nationally applicable
outcome measures, subjective and professional, in order to give a brief overview of
the patient’s situation and to describe the effects, was perceived as meaningful.

In addition to the global outcome measures and forms, the basic
instruments also included brief indicators of patient history and patient background
data. These were chosen – not to give a full picture – but because they were possible
independent or mediating factors to the dependent Quality Star subjective and
professional measures. These items were kept to a minimum, but were somewhat
expanded at two occasions on demand from participating centers. It has all the time
been the policy not to establish a parallel recording system. Eventually such data
should be drawn from the ordinary electronic patient record. This is also the reason
why treatment details were not included in the initial dataset. The Quality Star
outcome measures were, once experience was gained, envisaged to constitute an
“add-on” module to an integrated electronic patient record and administrative patient
register. As, during the last years, it has become obvious that this so far two decades
process is not going to result in a nationwide use for foreseeable time, the GGG has
also decided to start parallel recordings of days in care (hospital days and days with
ambulatory visits) as an indicator of cost accountability. The GGG further started to
describe the content of the encounters using the national codes of proceedings.
Cooperation with the emerging national psychiatric registers has been established,
where instruments and items are harmonized and the Quality Star will be an option in
the webb-portal, which is about to be launched.

In this way, the model has been developed and a database accumulated
where psychosis patients have been the major part. About yearly summary reports
have been produced. Also, the network has to a limited extent made some
compilations in order to confirm the psychometric and statistical properties of the
instruments. Some of the instruments measures have however not yet been properly
investigated regarding their psychometric properties. The collected data are now
considered sufficient for conducting in-depth research with a higher degree of
accuracy. It was therefore decided that a research program in two steps was to be
carried out at Karlstad University. Step one includes the analyzing of the material collected during the first ten years with the purpose of (a) assessing the psychometric properties of the Quality Star and (b) investigating group differences within the cohort between patients with different background characteristics and with different intervention patterns.

1.2  Theoretical perspectives

The final data instrument set for the Quality Star was developed and decided through a consensus procedure during the pilot project and followed by minor revisions during the first years in use where experts as well as experienced clinicians participated. The broad competence used was felt as a guarantee for valid choices. However, systematic documentation of this was not done. The adequacy of these choices in relation to literature is therefore accounted for in this section and, as this thesis is focused on the patient subjective instruments, these are given special attention.

1.2.1.  Outcome domains in psychiatry

In a systematic review regarding outcomes to measure in routine mental health services, and how to assess them, Slade (2002a) identified 16 outcome domain proposals, which were synthesized into seven emergent categories: wellbeing, cognition/emotion, behavior, physical health, interpersonal, society and services (see Table 2). It was further noted that two themes emerged from the review: 1) early proposals did not include assessment of the experience of receiving services, which only came to prominence in the mid-1990s, and 2), two distinct perspectives were identified, where publications within the medical literature used a more psychiatric language, emphasized the staff perspective, and had a focus on the amelioration of disability, whereas one publication from outside the medical literature (see Table 2, Campbell) used a more phenomenological language emphasizing the patient’s experience of care, and focused on increasing the patient’s wellbeing and avoiding
harm from and dependence on mental health services. Slade cautiously mentions that in the broader social science or user movement literature there are probably other well-developed proposals for outcome domains that were not identified in his review. Later works, citing Slade, was therefore searched to learn if new suggestions for additional outcome domains had appeared.

In the psychiatric perspective few additions were noted. There was no study found that suggested entirely new outcome domains or categories in the psychiatric perspective. A rather confirmative result was presented by Montgomery, Rose, and Carter, (2009), who in a literature review examined evidence concerning the relationship between psychiatric mental health nursing interventions and patient-focused outcomes. Most frequently used outcome instruments assessed psychiatric symptom severity, especially altered thoughts/perceptions and altered mood. Other outcome instruments were categorized in the domains of self-care, functioning, quality of life and satisfaction. The most important finding of this review is the lack of consistently strong evidence to support decisions concerning which outcome instrument or combination of instruments to recommend for routine use, sensitive to psychiatric mental health nursing interventions. Another review, focusing adverse outcomes for schizophrenia (Kooyma, Dean, Harvey, & Walsh (2007) found that the negative outcomes of violence, victimization, suicide and self-harm, substance use, homelessness, and unemployment are important domains, where the Slade review had identified few studies examining these outcome areas. Kooyma et al suggests that reducing these outcomes require implementation of a combination of strategies at national, local and patient levels, e.g. the matrix model of healthcare provision (Tansella & Thornicroft, 1998), but essentially these outcomes are in general terms covered by the categories in Table 2.

Most other studies, with a mainly psychiatric perspective are seemingly in principle d’accord with Slade’s overview of outcome areas, and deal with the problem of reliable measurement. From Australia, Eagar, Trauer, and Mellsop (2005) notes that the importance of routine outcome measurement in mental health care is increasingly recognized and mention Slade’s review as a central reference when examining the performance of two widely used outcome measures Health of the
Nation Outcome Scales (HoNOS) (Wing et al., 1998) and the Life Skills Profile (LSP) (Buckingham & Eagar, 1999; Rosen, Hadzi-Pavlovic, & Parker, 1989), finding them to appear suitable for routine collection, but also that these clinician-rated measures need to be complemented by both consumer-rated and family/carer measures. In addition, Burgess, Pirkis, and Coombs (2009) studied the reliable measurement of change with the HoNOS scales using three approaches to measuring effectiveness, effect size (ES); the reliable change index (RCI); and standard error of measurement (SEM).

In the United Kingdom Jacobs (2007) elaborate on the question of the possibility of a generic measure for psychiatry, essentially using HoNOS and the Clinical Outcomes in Routine Evaluation - Outcome Measure (CORE-OM) (Evans et al., 2002). and mentioning the generic use of GAF by the US Veteran Administration and the already mentioned instruments from Australia. Again in the UK Slade, Leese, Cahill, and Thornicroft (2005) showed that is was possible to use Manchester Short Assessment of Quality of Life, (MANSA) (Priebe, Huxley, Knight, & Evans, 1999) and Camberwell Assessment of Need Short Appraisal Schedule (CANSAS) (Slade, Thornicroft, Loftus, Phelan, & Wykes, 1999) assessments to determine that reduction in patient-rated unmet need precedes improvement in quality of life. From the USA, Eisen et al. (2011) states that mental health outcomes user-centered assessments should cover symptom reduction for remission of a disorder, quality of life and functional impairment. Given the evidence that self-report measures are valid with respect to functional status a study was conducted to determine whether a self-report measures of mental health status, the Behavior and Symptom Identification Scale (BASIS-24) (Eisen, Normand, Belanger, Spiro, & Esch, 2004), the Brief Symptom Inventory (BSI) (Derogatis, 1993), and the Veterans/Rand Short Form-36 (Ware & Sherbourne, 1992; Kazis et al., 2004) provide additional explanatory power relative to clinician-assessed GAF in predicting functional outcomes with regard to re-hospitalization and paid employment, which was found possible.

Regarding further identified possible user defined outcome domains in the broader humanistic, nursing, and user perspective, a number of studies was identified, which are briefly described. In a qualitative study Perry and Gilbody
(2009) found six categories within emerging concepts of Personal change (Changes in Myself; Believing in Change; Satisfaction with Treatment) and Social changes (Managing Independently; Social Contact and Connection with Society). Goal attainment as a means of promoting recovery was studied by Clarke (2008) and discussed with regard to mental health outcomes for consumers with enduring mental illness. In early stages, physical health, employment and maintaining relationships goals was prominent, and later other aspects of recovery, such as hope, self-confidence, sense of purpose and positive identity emerged. Examining consumers’ experiential definition of psychological recovery Andresen (2007) identified four key processes: 1) finding hope, 2) re-establishment of identity, 3) finding meaning in life, and taking responsibility for wellness and life generally. Four stages in these processes were synthesized from qualitative research: 1) moratorium, 2) awareness, 3) preparation, 4) rebuilding, and 5) growth. A brief measure, the Self-identified stage of recovery (SISR) and a longer, The Stages of Recovery Instrument (STORI) were developed to monitor recovery and refined version is envisaged as an outcome tool meaningful to consumers and an adjunct to conventional clinical measurement.

Newnham and Page (2010), reviewing of problems with outcome measurements mentions, remind of three phases throughout the course of therapy noted by Howard, Lueger, Maling, and Martinovich (1993): 1) remoralization; involving the alleviation of hopelessness and the boosting of wellbeing, 2) remediation; when a therapeutic relationship has been established and a reduction in symptoms becomes evident; 3) phase of rehabilitation, characterized by the modification of previous (and perhaps long-standing) patterns of relating to the world and to others, and an enhancement of life functioning. That is, Newnham marks, clinical practice uses a broad focus that concerns not only symptoms, but also life functioning, coping with stressors, and quality of life, and These variables are more vague and difficult to measure as an outcome, though Newnham mentions the development of the CORE-OM (Barkham et al., 2001; Barkham, Gilbert, Connell, Marshall, & Twigg, 2005) which covers four domains: Subjective Wellbeing, Problems, Functioning, and Risk (to self and others), and short forms used routinely in psychological services of the United Kingdom National Health Service Trust, as
well as the simultaneous development work by Lambert and colleagues in the United States with a similar system of patient response monitoring, with the Outcome Questionnaire (OQ-45). (Lutz et al., 2006).

Regarding Outcome measures in the specific case of schizophrenia disorders (which constitutes the majority of participants in the papers of this thesis), (Burns, 2007) summarized (slightly changed citation) that effective drugs and service developments such as community mental health teams and assertive community treatment, mean that the outcome differences must be measured more sensitively but also broader. The whole range of outcome perspectives necessary to consider includes ‘softer’ outcome measures such as quality of life, social functioning and personal well-being, cognitive functioning and neurobiological parameters, greater attention to adverse effects of drug treatment, and assessment of met needs. Patient satisfaction has become important patient-centered measure of how services are functioning. As systems of care become more consistent and predictable they themselves affect the outcomes to be measured, their costs and the degree to which they reduce the wider social consequences associated with schizophrenia. Hospitalization can be used as a decent proxy outcome measure for local variations in services. The emergence of patient-reported outcomes reflects not only this increase in interest in softer outcomes, but, perhaps a philosophical shift with a greater recognition of the importance of the patient’s view. People with schizophrenia are more treated as agents of their management, and not only can report on how they are doing but can influence what should be the measure of how they are doing – hence the appearance of self-empowerment and general wellbeing measures in the schizophrenia literature. This partnership approach has resulted in greater prominence for non-clinical hard outcomes such as employment or adequacy of accommodation, etc. From the opposite direction (from society and healthcare funders) there is increased sensitivity to the social burden of the disorder, attention paid to risk and the wider societal outcomes of violence, victimization, suicide, etc.

In summary there seems to be general agreement of which main outcome domains and categories are relevant in a psychiatric perspective in the present era and that the overview by Slade (Table 2) can be used as guidance, taking in
Table 2. Outcome domain proposals from 16 studies for use in mental health services, grouped into 7 emergent categories. (Adapted from Slade, 2002a).

<table>
<thead>
<tr>
<th>Study</th>
<th>Wellbeing</th>
<th>Cognition/ emotion</th>
<th>Behaviour</th>
<th>Physical health</th>
<th>Interpersonal</th>
<th>Society</th>
<th>Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wenger [22]</td>
<td>Perceptions of health status &amp; wellbeing</td>
<td>Intellectual Emotional Symptoms</td>
<td>Daily routine Symptoms – other illnesses</td>
<td>Social</td>
<td>Economic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Berger [23]</td>
<td>Health perceptions General life satisfaction</td>
<td>Symptoms Emotional status Cognition</td>
<td>Functional status Sleep and rest Energy and vitality</td>
<td>Role activities Social functioning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hargreaves [24]</td>
<td>Humanistic</td>
<td>Clinical</td>
<td>Rehabilitation Public safety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ware [25]</td>
<td>xentai neam</td>
<td>Physical</td>
<td>Social functioning Role functioning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nagi [27]</td>
<td>Quality of life Needs for care</td>
<td>Psychopathology</td>
<td>Social functioning and support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sederer [28]</td>
<td>Symptom</td>
<td>Functional</td>
<td>Satisfaction with treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cook [29]</td>
<td>Quality of life</td>
<td>Vocational Educational Residential</td>
<td>Hospitalization Consumer satisfaction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McGlynn [30]</td>
<td>Quality of life</td>
<td>Clinical</td>
<td>Functional</td>
<td>Adverse events Satisfaction with medical care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schieter [31]</td>
<td>Life satisfaction Life direction</td>
<td>Emotional Mental</td>
<td>Physical</td>
<td>Social</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Campbell [34]</td>
<td>Wellbeing Personhood</td>
<td>Recovery</td>
<td></td>
<td>Golf-help Empowerment Iatrogenic effects and negative outcomes Satisfaction and Disatisfaction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clifford [32]</td>
<td>Psychological Psychological wellbeing</td>
<td>Activities of daily living</td>
<td>Physical wellbeing Interpersonal relationships Social circumstances</td>
<td>Response to care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitzpatrick [33]</td>
<td>Psychological wellbeing</td>
<td>Cognitive functioning</td>
<td>Physical function Social well-being Role activities</td>
<td>Satisfaction with care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings [34]</td>
<td>Health status/ health-related quality of life</td>
<td>Diagnosis Psychological function Symptom management</td>
<td>Behavioural Activities of daily living Comfort/discomfort Physical function Mobility Disability</td>
<td>Social function</td>
<td>Patient satisfaction Appropriateness of treatment Sentinel events Technical proficiency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thornicroft [35]</td>
<td>Quality of life</td>
<td>Disabilities Needs</td>
<td>Impact of caring Satisfaction with services</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thornicroft, 2000 [36]</td>
<td>Quality of life</td>
<td>Needs</td>
<td>Care burden</td>
<td>Quality of care Satisfaction</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

consideration the additional suggestions found from works out of a humanistic, nursing and user perspectives when choosing suitable instruments for measurements.
In the following passages the rationale for instrument choices in main patient subjective domains dealt with in this thesis are accounted for.

1.2.2 Consumer satisfaction within mental health service

1.2.2.1. Conceptual considerations

Satisfaction with psychiatric service is a complex multidimensional construct. Patient satisfaction includes two subjective sections: the expectation of care and the evaluation of care. Satisfaction is a subjective measure, and it is not a matter of being right or wrong when evaluating quality of care (Vuori, 1991).

Theories of patient satisfaction indicate that the patient construct encompasses all aspects of care: processes and outcomes of treatment as well as structural aspects (Mahon, 1996; Ahlfors et al., 2001; Howard, Rayens, El-Mallakh, & Clark, 2007).

Further elaborated, research has shown that in varying extent, and as a rule, central aspects of the concept are patients’ expectations of treatment, their own view of the care provided, the attitude of the personnel, and the treatment outcome. Regarding the content of care construct involved this is multi-factorial comprising not only quantity and quality of various treatment modalities but also the availability and usefulness of treatment and care, information given, shared decision making, environmental aspects, and the degree to which patients perceive that their specific needs are met by the care with competence in specific as well as in terms of improving general well-being (Ahlfors et al., 2001; Mahon, 1996; Siponen & Välimäki, 2003).

There are variations in emphasis of these different elements, in their detailed descriptions and definitions, but most such additions in detailed descriptions encountered in literature would likely be captured within the frames outlined above or be substantially overlapping.

What elements of quality of care that are perceived as important to patients in particular situations vary. As there is broad recognition that improvement
and favorable outcome, as a rule, is related to services being given in a way that is perceived with satisfaction by the users (Priebe, Kaiser, Huxley, Röder-Wanner, & Rudolf, 1998; Tattan & Tarrier, 2000; Sörgaard et al., 2002; Mattsson et al., 2005; Ruggieri et al., 2000; Ruggeri et al., 2003; Helldin, Kane, Karilampi, Norlander, & Archer, 2008), there is a plethora of studies published on how patients’ evaluations of aspects of elements of care, and their outcome evaluations are influenced by patient characteristics and by circumstantial factors for the patient. As an example (Taylor et al., 2009) contributes with a literature review regarding quality of institutional care for people with long term mental health problems. It is not necessary for the present work to give a full review of such studies, but examples of relevant research questions could be:

For which patient groups under which conditions is age or gender as such influencing patients’ satisfaction of particular aspects of care?

Are there differences in satisfaction with particular care elements between patients with different diagnostic entities (and their duration and severity), other circumstantial aspects being equal?

How do patients social circumstances such as living conditions, occupation, economy and social activities influence patient satisfaction with elements of care for patients with similar characteristics?

The list of relevant questions is lengthy (see for instance Blenkiron & Hammill, 2003; Holloway & Carson, 1999; Eklund & Hansson, 2001; Ahlfors et al., 2001; Ruggieri et al., 2003).

The measures of the various care satisfaction elements, as well as of factors exercising influence on satisfaction need special attention, regarding methodological principles and in validity aspects (Noble, Douglas, & Newman, 2001; Barker & Orrell, 1999; Lebow 1983; Vouri, 1987; Ware & Hays, 1988). The construction of the instruments and the principles for performance of the interviews differ between scales. How, by which method, at which phase of the treatment, at what time-point, and by who should the measurement/ rating be done? (Allebeck & Hansagi, 1995; Ahlfors et al., 2001).
1.2.2.2 Consumer Satisfaction measure rationale in present work

The guiding principle was to, if possible, use an instrument that had established broad use. The only contestant was Swedish Institute for Health Services Development (SPRI) instruments (Hansson & Höglund, 1995) which had a broad use, but were considered too detailed for the purpose. At that time the Committee for Clinical Trials, a working group within the Scandinavian Society for Psychopharmacology (SSP) had completed the development of an instrument that addressed limitations of currently available rating scales and questionnaires of satisfaction with care, the interview-based Consumer Satisfaction Scale (ConSat). The instrument was intended to be administered by a person not directly involved in the treatment and care of the patient should do the interview (Ahlfors et al., 2001). During the Quality Star tests it became clear that the logistics needed to arrange for an independent rater to do the interview was causing too much practical problems in the routine care. After discussion and consensus it was decided to transform the scale for self-administration by the patient.

1.2.3 Quality of life in psychiatry

1.2.3.1 Conceptual considerations
A number of overviews (Awad, Varuganti & Heslegrave, 1997; Brekke & Long, 2000; Fakhoury & Pribe, 2002; Holloway 1999; Holloway & Carson, 1999; Holloway & Carson, 2002; Katschnig, 2000; Katschnig, Freeman, & Sartorius, 1997) clarify that, although there is yet no single accepted model of quality of life in mental illness or a universal instrument to measure it, most attempts to conceptualize or measure the total life experience of individuals propose some constellation of objective and subjective dimensions.

Objective indicators of the quality of life has been suggested to refer to resource availability and objective life conditions though associations with subjective quality of life has been found complex and varying (Prince & Prince, 2001;
The social indicator line of research is the major exponent of this concept and covers a vast array of activities, life events, and characteristics pertaining to individuals, and groups of individuals, typically monitored by governmental and social institutions that “do not depend on the individual’s description of his own life” (Prince & Prince, 2001; Campbell, Converse, & Rodgers, 1976, p. 118). Among these domains, demographic variables, such as income, education level, and rates of unemployment provide a basis for developing or altering social policy. More sophisticated indicators might be the size and nature of one’s social network, opportunities for leisure and recreation and access to certain commonly accepted social goods, such as a car, a telephone, a TV and video and perhaps a holiday. This approach to conceptualizing quality of life, can establish “standard needs” common to all individuals, which can be determined through the consensus of experts, or from normative data collected in large population surveys (Browne, O'Boyle, McGee, McDonald & Joyce, 1997). This logic can also be applied to patient groups.

Holloway and Carson, (2002) provide a mapping of models from quality of life literature in severe mental illness with respect to objective and subjective components. Studies on associations between indicators of the two components have shown varying results according to literature (Fakhoury & Priebe, 2002; Oliver, Huxley, Priebe, & Kaiser, 1997).

Subjective’ quality of life refers to the individual’s subjective evaluation of his/her life (Carpiniello et al.,1997; Malmgren, 1997). This may refer to, as compiled by Fakhoury and Priebe (2002), a perception, cognitive and/or emotional, of being somewhere in the gap, at a given period in time, between the aspirations of a patient and his/her perceived reality. It can also be a function of one’s comparisons with previous best, or experiencing one’s life.in comparisons of own situation to the situation of others. Issues included in the quality of life conceptualization may be all facets of life such as health, job, family, environment, and many other life situations. (Awad et al., 1997). It can also in essence relate to a person’s sense of wellbeing (Anger, 2011), his satisfaction with life, or his happiness or unhappiness, that may either be broken down into satisfaction within life domains, which might be defined a
priori or empirically. Common range of life domains include work, leisure, social relationships, finances, health, social and physical environment and opportunities for self-fulfillment.

Research typically look at related objective and subjective quality of life domains. (Prince & Prince, 2001) notes that, whereas the objective conditions in which individuals reside may influence the experience of life, they do not necessarily reflect that experience. Objective data on quality of life and patients' satisfaction with life have been found to be weakly associated in a number of studies (Fakhoury & Priebe, 2002; Oliver, Huxley, Priebe, & Kaiser, 1997; Priebe et al., 1998; Skantze, Malm, Dencker, May, & Corrigan, 1992).

Whereas correlates between defined elements of objective and subjective quality of life, as already mentioned, are often found relatively weak, there is also the question of independent factors influencing or predicting elements of quality of life. Holloway and Carson (2002) list examples of studies that have found relationship between objective quality of life and for instance pre-morbid educational level, pre-morbid social functioning and duration of untreated psychosis, and negative symptoms at onset of schizophrenia. Regarding global subjective quality of life numerous associations are found. Associations have been reported for temperament, hope, autonomy and positive self-concept, intelligence, size of the support network, employment status, and more. Mood state of often found to have substantial influence, as well as cognitive functioning, and can clearly be aspects of objective assessments of disorders, as such belonging to health aspects. Obviously it is necessary when studying quality of life using such elements that one conceptually clarify if “health aspects” are to be treated as part of “objective quality of life” parameters or as independent factors in relations to a used quality of life construct.

When it comes to measurements of quality of life it may be wise to bare in mind when taking part of works in the literature that Holloway and Carson (2002) has concluded that “a bewildering variety of instruments purporting to measure quality of life are in use, often without clarification of what aspect(s) of the construct are being assessed.” The conceptual discussion seems to take the position of for instance Awad et al., (1997) and Fakhoury and Pribe (2002) that any approach to its
measurement should include a subjective component assessing patients’ self reports about their quality of life. It also seems correct, given the complex interrelationships between objective, subjective elements of quality of life, and associated independent elements to follow the advise of (Awad et al., 1997; Simeoni et al., 2000): “The choice of measure must rely on the particular purpose and needs” and that “for comparison purposes generic instruments of quality of life are better than illness-specific ones, at the same time as there is also a priority also to develop illness-specific quality of life measures.

1.2.3.2 Quality of Life measure rationale in present work

The choice of quality of life instrument for use in the Quality Star was based on the knowledge that quality of life for mentally ill persons is often based on health related quality of life, social circumstances, existential aspects, and human relations aspects and at the same time being an individual phenomenon. (e.g., Katschnig 2000); Bigelow, Gareau, & Young, 1990; Hansson & Björkman, 2007; Lehman, 1983; Malmgren, 1997; Oliver 1992; Ruggeri, Gater, Bisoffi, Barbui, & Tansella, 2002). The construct should be able to, depending on individuals' present situation, let the rating be based on emotional as well as cognitive perceptions and regardless if ratings are based on hopes, expectations, comparisons with previous situation, with reference to other's situation, in relation to common views, professional or other values, objective social or health norms. The QLS-100 scale (Skantze & Malm, 1994) was fairly established at the time (for details see section 2.3.3 Instruments) and did meet the requirements to invite patients to report in a broad array of areas in open-question form. The scale was also specifically developed not only to get scores but also to support dialogue, in line with a prime purpose of the Quality Star. However, during the tests it was obvious that not all sites could manage to take care of all information gained using QLS-100 and wanted a quality of life instrument less demanding. In search after an alternative among established instruments, it was deemed that the item General Well Being (GWB) (Oliver 1992), used in the beginning of the Lancashire Quality of Life Profile (LQLP) (Oliver et al 1997),
similar to the first item, "Life as a whole" in the abbreviated version Manchester Short Assessment of Quality of Life (MANSA) (Priebe, Huxley, Knight, & Evans, 1999), was providing a suitable construct, and a stand-alone version was developed, the Global Quality of Life scale (GQL) (see 2.3.3 Instruments). The QLS-100 scale was remained an optional add-on scale to be used when clinically appropriate.

1.2.4 Mental distress in serious mental illness

1.2.4.1 Conceptual considerations

Amongst main outcome domains and categories described in the review by Slade (Slade, 2002a) a number of suggested categories are relating to what is in this work treated as a concept of “mental distress”. Included in the general category of Well-being (Table 2) are mentioned phenomena proposed by several authors, such as: Perceptions of health status, Health perceptions, Mental health. As noted in Section 1.2.1. a number of user defined outcome domains in the broader humanistic, nursing, and user perspective were identified, which in patient conceptualizations are likely to be perceived as elements of, or overlapping, “mental distress”, or its absence. This included concepts as Personal change/Changes in Myself; Believing in Change, Managing Independently Hope, Self-confidence, Sense of purpose, Positive identity, Wellbeing. Brekke and Long (2000) reminds that until at that time, whereas some attention had been paid to client satisfaction with services, the subjective experience of clients improvement had been less considered. In the literature review by McCabe, Saidi, and Priebe (2007) it was found that instruments and research regarding resilience of the self was less common. A development was noted for areas like patient reported symptoms, insight, attitude towards medication, and clinical communication, and also including empowerment, self-esteem, sense of coherence and recovery. It was further noted that scores of different patient-related outcomes overlap and are influenced by a general tendency, largely influenced by mood, for more or less positive appraisal.
Looking at the perceived mental distress issue from the desire of psychiatric care, aiming to minimize the degree of self-rated symptoms and the number of needs, to maximize subjective quality of life, and to make (or keep) patients' assessment of treatment as positive as possible, in the evaluation of psychotherapeutic and psychopharmacological interventions, self-ratings of specific or nonspecific symptoms are becoming widely used, and numerous scales have been developed for self-rating of symptoms, mainly of anxiety and depression but also of schizophrenia and other psychiatric disorders (Priebe et al., 1998). In one such work, Lindström, Jedenius, and Levander (2009), highlighted that: “patients appear to construct their appreciation of degree of illness differently than clinicians. Much of the feelings of being ill seem to be channeled via affective symptoms.”

The distinction between patient reports of mental distress arising from a unique individual conceptualization and patients being invited to self-rate symptoms (or other similar professionally defined concepts) is crucial. The advice from McCabe et al. (2007) seems very appropriate: “For the future development of patient-reported outcomes, the most pressing issues are conceptual clarity and the consideration of empirical findings, particularly the association with other outcomes.”

1.2.4.2 “Mental distress” measure rationale in present work

The phenomenon sought to capture by the measure is the degree of feeling mentally distressed and/or having complaints regarding mental health, irrespective of how the respondent arrive to an opinion. The instrument should also be suitable for dialogue with the user.

In the initial development of the Quality Star model, the distress rating included in the Community Health Record instrument set used in the Optimal Treatment Project was tested where distress (pain and mental suffering) is rated as 0=none, 1=mild, 2=moderate, and 3=severe, followed by free text recording of described suffering (Falloon et al., 2004). In the field test for the Quality Star, feedback was that the four levels were too restrictive for recording progress/deterioration.
Some centers were suggesting more elaborate instruments, but the majority of participating centers declined this. While it was accepted that more detailed instruments should be used is clinically appropriate situations, like using the Symptoms Check List (SCL90) (Derogatis, 1994; Derogatis, 2000; Fridell, Cesarec, Johansson, & Malling Andersen, 2002). Other, shorter alternatives with similar aims, for instance the WHO Self Reporting Questionnaire 20 (Harpham et al., 2003), General Health Questionnaire (Goldberg & Williams, 1988), Kessler Psychological Distress Scale (Andrews & Slade, 2001), Psycho-Social Well-Being Scale (O’Hare et al., 2003) were discussed.

However, the network finally choose to use a single visual analogue scale (Bech, 1993; Everitt & Wykes, 1999), the Perceived Global Distress scale (PGD), constructed for the purpose by the network (GGG-group, 2009). It is an adaptation for stand-alone use of the last item “How satisfied are you with your mental health” in Manchester Short Assessment of Quality of Life (MANSA) (Priebe et al., 1999). However, the focus was narrowed by changing the introductory question to "How much have you been bothered by your psychiatric problems during the last month?”. Degrees of positive mental health are not focused, which is a difference to the MANSA. The PGD scale thus assumes that most psychiatric patients experience degrees of mental health distress, and if not, the scale does not give room for expressing degrees of satisfaction with mental health. The intention with this restriction was to guide the respondent to focus if there are mental health problems at all. In this sense, the PGD is intended to support problem identification in the mental health area together with the patient. A following dialogue should clarify the precise nature of the problems that the patient experiences.

1.2.5 Associations between The Subjective and The Objective

Brief mentions have been made in the previous sections of associations between user satisfaction, quality of life and mental distress and other subjective as well as to associated objective concepts.
The general view presented by (Priebe et al., 1998), that several subjective criteria are used in evaluative research in psychiatry, and that particularly common criteria are: self-rated symptoms, subjective quality of life, self-rated needs, and assessment of or satisfaction with treatment, and that within these there are distinct theoretical foundations and a research tradition of their own, seems to be a valid description also today. Priebe further notes that for each of the four subjective evaluation criteria, there is an objective or observer-rated counterpart: observer-rated symptoms, objective quality of life or standard of living, observer-rated needs, and clinical audits, as illustrated in Table 3.

<table>
<thead>
<tr>
<th>SUBJECTIVE</th>
<th>OBSERVED/OBJECTIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-rated Symptoms</td>
<td>Observer-rated Symptoms</td>
</tr>
<tr>
<td>Subjective Quality of Life</td>
<td>Objective Quality of Life or Standard of Living</td>
</tr>
<tr>
<td>Self-rated Needs</td>
<td>Observer-rated Needs</td>
</tr>
<tr>
<td>Assessment of or Satisfaction with Treatment</td>
<td>Clinical Audits</td>
</tr>
</tbody>
</table>

Table 3. Frequent subjective outcome domains and their broad objective counterparts. (Based on Priebe et al., 1998)

The correlations between objective data or observer ratings on the one hand and self-ratings of the same criterion on the other are often surprisingly low, and varies depending on which sample is investigated and which measures are used within the domains, i.e. the specific construct. The complexity of overlaps between constructs can be exemplified by comparing Table 3 above and Table 2 in chapter 1.2.1. “Outcome domains in psychiatry”.
1.2.5.1 Choices within the Quality Star concept

As focus for this thesis is subjective measures there are discussed in some detail, whereas their “objective counterparts” (Table 3) are accounted for only as far as needed for the main purpose. However, as short background overview short notes will be given here to the instrument choices regarding “objective counterparts” within the Quality Star model.

Self-rated Symptoms and Observer-rated Symptoms.

The visual analogue scale PGD (see chapter 1.2.4.2) is mirrored in the Quality Star with the professional global rating of symptom severity by the split-GAF-symptoms (GAF-S) (Pedersen, Hagtvet & Karterud, 2007). During the test period Clinical Global Impression scale (CGI) adopted by Falloon (Falloon & Fadden, 1993; Guy, 1976), and Health of the Nation Outcome scale (HoNOS) (Wing et al., 1998) were tried as alternatives. The CGI was not chosen because, though proven robust for research purposes, the interval levels of the scale were not found suitable to inform dialogue with patients. The HoNOS was not chosen because the proportion of points for the nonspecific item 8 (“Other mental and behavioural problems”) was found too large in for instance eating disorders. The main reason for choosing GAF-S was however that the scale was becoming widely used in Sweden.

Subjective Quality of Life and Objective Quality of Life/Standard of Living.

The subjective GQL visual analogue scale (see chapter 1.2.3.2) has a broad content. The “objective” counterpart in the Quality Star outcome set, the split-GAF-function rating cannot be expected to have but partial overlap. During the test period the Short Form Health Survey (SF-36) (Ware & Sherbourne, 1992) was suggested, but this construct was not appreciated being too medically oriented. Standard of living aspects and conditions of life and relations are part of the background data set (i.e. education, work, living, relations, children), not included as primary outcome measures, but available for analyses if wanted.
Assessment/Satisfaction of/with Treatment and Clinical Audits.
The consumer satisfaction scale (see chapter 1.2.2.2) does not have an obligatory audit rating. General audit protocols have been developed by the Swedish Psychiatric Association and also a version with special focus on Integrated mental care for use during the Optimial Treatment Project, but they have not been successful to reach broad use and data were therefore so far not available.

Self-rated Needs and Observer-rated Needs.
The Needs aspect is not part of the basic Quality Star set, though it was considered. The QLS-100 inventory of problems and solutions (Skantze et al., 1992) and Camberwell Assessment of Needs (CAN) (Phelan et al., 1995) in particular. However, the Quality Star was to be purely outcome related. The first part of the two instruments meet this prerequisite, but the second parts, judgments of needs were deemed to relate to process rather than outcome. The first, descriptive, part of QLS-100 was included as an optional detailed quality of life instrument, whereas the corresponding use of CAN was not possible due to copyright.
2. THE PRESENT INVESTIGATION

2.1 Introduction

2.1.1 The purpose of the current thesis

The aims of this thesis were to a) investigate the psychometric properties of the patient subjective measures used in the Quality Star concept, and b) to describe their relationship with regard to social function and gender for seriously mental ill persons. The research plan has been evaluated and approved by the Regional Ethical Vetting Board in Uppsala and the study followed the ethical standards of the World Medical Association declaration of Helsinki concerning Ethical Principles of Medical Research Involving Human Subjects.

2.1.2 Procedures for the Quality Star instrument

In order to briefly describe the use of the method, and as three of the four studies (II-IV) are partly based on data from the routine use of the Quality Star instruments, a full description of the procedure for its use is given here and referred to in the Procedure sections in the following.

Decision to participate in the Quality Star network by psychiatric departments include ethical considerations regarding clinical follow up by using data from routine care and quality systems. The data software was not delivered to any site unless such declaration was given. The personal were trained in use of the instruments following the manual (GGG-group, 2009). Instructions include that subjective instruments, should be used in an as neutral situation as possible, for instance not directly after focus on topics that may influence rating. Subjective instruments should not be used immediately following each other, for the same reason.
Participants were introduced to the Quality Star at routine visits by their case manager and given written information. Right to withdraw without further motivation, and right to get extracts from the database was part of the information. If they chose to participate, this is noted in the clinical case notes. Case managers go through the instruments with the patient at one or two ordinary appointments with prime purpose to review the situation and further to document findings in the patient record and in the Quality Star database. Case managers prepare the first contact by reviewing notes to enter background data regarding history and socio demographic questions. Recommended order to use the instrument is to start with GQL and then ConSat-P. Next the case manager turn to the question if the participant would be willing to let a next of kin or important other tell if he/she is troubled by burden in order to – if so – discuss how the situation might be improved. Written information of the Quality Star method and the instrument PGB is presented. The next instrument presented is the PGD, followed by the Health screening using the UKU-Side Effect Rating Scale (SERS). Finally the CM presents her/his rating of the GAF.

Additional instruments used locally to penetrate further the areas covered by the global measures mentioned above are presented after the basic measures. Concluding, a summary presentation of all ratings is done using the “Star” (Fig. 1, Section 1.1. p. 3), where “good” ratings are marked centrifugally and “worst” to the middle of the figure. In a following dialogue it should be established what needs to be taken care of in following review of the personal treatment plan.

Entries to the local Quality Star software are transferred to the national database at intervals where participant identification is replaced by random identification.

2.1.3 The four articles contained in the thesis

The initial study (Paper I) was designed to establish the psychometric properties of the self-rating version of the Consumer Satisfaction scale, which had been developed in an iterative process in the Quality Star network praxis from the original interviewer version. It was found in the tests that the interviewer version was too
resource demanding and not suited for routine clinical use. It was also anticipated that given a proper administration by the case managers a self-rating version would be supportive for the open dialogue with the patient, promoting the development of treatment alliances. The study was done when items had been finally phrased and when it had been used in practice for about a year.

Subjective quality of life is treated by the Quality Star concept as a major outcome dimension. The construct of the visual analogue scale Global Quality of Life (GQL) was therefore a particularly important area for considerations when the instruments were decided. The validation study (Paper II) was designed with several (8) different validating measures in three separate sub-studies, apart from a test-retest study.

The next visual analogue scale, the Perceived Global (mental) Distress scale (PGD) was designed in order to facilitate the expression of patients’ own perception. The philosophy behind it was that dialogue with patient, and treatment alliance, would be supported if the point of departure is the patient perspective. Conceptually there is a radical difference from self-report scales that are designed to help patients to express signs and symptoms from a basically professional concept.

The validations study (Paper III) used 10 validating measures to clarify the psychometrics of the scale.

With the properties of the three patient subjective instruments having been described in Papers I - III a cross-sectional study, using the Quality Star accumulated database, the discriminative properties of the subjective instruments was studied with gender and social function as independent variables (Paper IV).

The specific aims, designs, instruments, procedures, statistic, and results regarding each of the study is presented in the respective sections.

2.2 Paper I. Psychometric properties of the Consumer Satisfaction Rating Scale

2.2.1 Aim
The aim of the study was to validate a patient self-rating version of the UKU Consumer Satisfaction scale (Pat-UKU-ConSat) in relation to the original independent interviewer version (UKU-ConSat) (Ahlfors et al., 2001), and to analyse its internal consistency.

2.2.2 Design

As a preparatory step, the independent raters, 10 clinically experienced doctors, nurses or psychologists, were trained with 7 patients who were not subjects in the main study. Sufficient inter-rater reliability was established before the main study was conducted.

In the main study, Thirty-seven patients from five psychiatric services in western Sweden participated, each service contributing with 5-10 subjects. Twenty-five patients were aged between 31 and 50, five between 21 and 30 years of age, and seven between 51 and 60. The male/female ratio was 7:3. The diagnostic groupings of the 37 patients were psychotic disorders (35%), anxiety disorders (35%), affective disorders (16%) and addictions (14%). At the time of the study, 78% were inpatients (in one case with compulsory care).

The self-assessment version UKU-ConSat-Pat was first completed by the patients. The instructions to the participating patients were given by the respective raters at the five sites. They had no previous treatment relation to the patients. They also remained ignorant of the results.

Later the same day, the professional assessors performed the interview according to the original instrument UKU-ConSat.

The two protocols were then sent to the study authors for analysis.

2.2.3 Instruments

The UKU-Consumer Satisfaction Rating Scale (UKU-ConSat)

This scale for assessment of consumers’ satisfaction with psychiatric care was constructed for administration by an independent assessor (Ahlfors et al. 2001). It
consists of six items related to care structure and process (availability, atmosphere, treatment modalities, information, drug treatment and psychosocial interventions), and two items related to outcome (usefulness of treatment/care and general well being).

Each rating is scored from +3 to −3. +3 points represents a fully positive attitude to an item, +2 represents a mainly positive attitude and +1 a somewhat positive attitude. A zero score is given whenever the answer is difficult to interpret. Raw scores are added to a Total score. Subscales for the six first items constitute a Process quality subscore, and an Outcome subscore is formed by adding the two last items.

The inter-rater reliability has been established with an ICC coefficient 0.76-0.95. Item analyses for different patient groups (psychotic, affective, alcoholic) had indicated that the scale permits discrimination between various elements of consumer satisfaction. Correlations with measures for Symptoms were found in the range of $\rho=0.4$, and somewhat lower for Social functioning, Distress and Carer distress.

**UKU-Consumer Satisfaction Rating Scale, self-rating version (UKU-ConSat-Pat)**

The self-rating measure UKU-ConSat-Pat (Paper I, Appendix 1 & 2) was constructed to reach as close a correspondence to the original scale as possible. In order to make the scale easily understandable and consumer-friendly, some items of the original instrument containing more than one element were divided into separate items. The original item “Environment and services” (A2) was divided into two, as was the item “Availability of various treatment modalities” (A3). The Item “Drug treatment” (A5) was divided into three in the self-assessment version in order to clearly permit patients without prescribed medication to express their views. Corresponding to the item “General well being” (A8), the visual analogue scale for subjective quality of life (GQL) of the Quality Star algorithm was used (Paper I, Appendix 2). The formulations of the items were adjusted to final forms in an iterative process, where preliminary versions were used in ordinary clinical practice and experiences were regularly discussed and successive improvements made.
The 11 first items are rated on a seven point scale, as the original interviewer version, with the format in principle +3 full satisfaction, +2 satisfied but with minor dissatisfaction, +1 More satisfaction than dissatisfaction, 0 equally satisfaction/dissatisfaction or indecisive, -1 to –3 formulated in a reciprocal fashion. The final 12th item regarding “General well being” is a visual analogue scale ranging 1-100. The endpoints are marked “Best possible life situation” (point 100) and “Worst possible life situation” (point 0). The algorithms for calculation of analogous to the original Con-Sat items A1-A8 from the ConSat-Pat items are as follows: Items S1-S10 and the VAS scale are used. VAS scale raw points are transferred to scaling –3 to +3, where 0=–3 and 100=+3. The 11th item of the self-report version was created to capture the reference to service satisfaction in original items A7 and A8, but was omitted in the analysis as core content was judged to be corresponding with the self-report items S10 and the VAS scale respectively.

A1 Availability of treatment/care = S1
A2 Environment and services = (S2-S3)/2
A3 Availability of various treatment modalities = (S4-S5)/2
A4 Information = (S6-S7)/2
A5 Drug treatment = (S8b or S8c)
A6 Psychosocial interventions =S9
A7 Usefulness of treatment/care =S10
A8 General well being =transformed VAS scale to –3 to +3

From this two subscores are calculated, A1-A6 Subscore “Structure and process” A7-A8 Subcore “Outcome”, and finally the Total score A1-A8.

2.2.4 Procedure

Training of Independent raters with UKU-ConSat and Interrater reliability study.
Ten clinically experienced doctors, nurses or psychologists (including the authors as being experienced users of the scale), from five participating psychiatric services, accepted to participate in the study as independent raters with the UKU-ConSat. The
authors gave thorough instructions, demonstrations and conducted role-play based training to ascertain the proper use of the instrument.

Next, all ten raters rated seven patients with the UKU-ConSat in order to verify acceptable reliability in the use of the instrument, and the data obtained were used for calculation of reliability coefficients.

The seven participating patients in this study were recruited from one of the services. The participating raters from this hospital were assisted by colleagues who invited patients, who did not have any treatment connection with the raters. The colleagues were obtaining the informed consent of the participants, with the specific assurance that results of the research was not going to be known to the colleagues in treatment contact with the patient. Arrangement was made for the patient to come to the assessor group with the support of one of the local raters. Research leaders presented all raters and described again the study and its purpose, informed the patient of the right discontinue participation at any time and the patient confirmation to participate was obtained. One of the authors conducted the interview and all assessors made their independent ratings. After having rated all the seven patients the results were reviewed and when a difference between raters were 2 points or more clarification was sough and noted by the research leaders, to be at hand in the following analyses by the authors.

Main study - validation of UKU-ConSat-Pat against the original UKU-ConSat
The study was carried out at five psychiatric services in West Sweden. At each centre the trained assessors conducted the investigations with 5-10 patients. To obtain a case-mix similar to that in the original scale article it was agreed between raters and research leaders which diagnostic categories and how many participants that each centre was going to recruit.

The recruitment, information and obtaining informed consent was done in a similar way as described under the previous heading.

The self-assessment version was first completed by the patients, the interviewers/assessors remaining ignorant of the result. Later the same day, the
professional assessors performed the interview according to the original instrument. The two protocols were then sent to the authors for analysis.

2.2.5 Statistics

The interclass correlation coefficient (ICC) was analyzed according to a variance component model (Searle, Casella, & McCulloch, 1992) for description of the inter-rater reliability of the assessors.

In the main study, correlations were analyzed between ratings on the eight items of the original instrument with corresponding items of the self-assessment version (as described in the Instruments section) using Spearman’s rank correlation method, bivariate. In the original instrument, “not applicable (NB)” answers are possible in items A5 and A6. In order to be able to include these answers in analyses, the method was used to replace a NB rating with the means of all items with answers. A corresponding operation was carried out for the self-assessment version items when data were in a few cases missing.

For analysis of internal consistency of the self-ratings scale, a Mokken analysis (Sijtsma & Molenaar, 2002; Molenaar, Debets, Systma, & Hemker (1994) was carried out as a complement to the traditional Cronbach’s alpha (Nunnally, & Bernstein, 1994).

2.2.6 Results

Inter-rater reliability was found good for the total score of UKU-ConSat (ICC= 0.91), and good to acceptable for items and subscales.

The correlations between interviewers’ assessments according to the original UKU-ConSat and the corresponding patients’ self-ratings were acceptable to good ($\rho=0.67$ to $\rho=0.82$). Correlations between sub-scores on structure and process elements and outcomes were $\rho=0.81$ and $\rho=0.75$ respectively. Total score correlation was $\rho=0.83$. 

34
Cronbach’s alpha, the standard traditional method for analyzing internal consistency (Cronbach 1951), generated a coefficient of 0.80 for the self-reported version, indicating acceptable consistency (Nunnally & Bernstein, 1994). The complementary Mokken analysis revealed acceptable or just acceptable unidimensionality of both scales (Loevinger coefficients in the range of 0.45 – 0.36).

2.2.7 Addendum

**Paper I** reports the concurrent validity of the original and the self-rating versions of the Consumer Satisfaction Scale. The self-rating Outcome item S11 was omitted for reasons described, and the Outcome items are S10 and item 12, the “Visual analogue scale for subjective quality of life”.

However, in the Quality Star, the S11 is used together with item S10 as outcome items, and the Visual analogue scale for subjective quality of life, named the Global Quality of Life scale (GQL) is used as a separate instrument (**Paper II**).

As the self-rating version of ConSat referred to in **Papers II-IV**, is the version used in the Quality Star it was necessary to supply information of the internal consistency of the actual version used. Re-analysis Paper I data, as well as analysis in this respect in the Quality Star database was therefore done.

The analyses showed that the version used in the Quality Star in fact also had acceptable correlation with the total sum of the original scale (\(\rho=0.98\)) and similar good internal consistency, indicated by Cronbach’s Alpha 0.89. However correlations between the process items were weaker and it could be concluded that the two scales total values are comparable but outcome items and subscales do not giving identical information and are different in construct. It is thus important to state what version is used in future studies to make correct comparisons.
2.3 Paper II. Psychometric properties of the Global Quality of Life Scale

2.3.1 Aim
The Global Quality of Life Scale (GQL) is a stand-alone version of the General Well Being (GWB) item (Oliver, 1992), used in the beginning of the Lancashire Quality of Life Profile (LQLP) (Oliver et al 1997), similar to the first item, "Life as a whole" in the LQLP abbreviated version, the Manchester Short Assessment of Quality of Life (MANSA) (Priebe et al., 1999). The aim of the present studies was to investigate the psychometric properties of the visual analogue scale (Everitt & Wykes, 1999) GQL for seriously mentally ill persons.

2.3.2 Design
Four different studies were conducted together with cooperating centers in order to investigate psychometric properties GQL – a test-retest study and three studies regarding concurrent and content validity.

The test-retest reliability study was carried out at ambulatory psychiatric service at three centers in different parts of Sweden. Thirty consecutive patients were recruited for participating by their main contacts, i.e. their case managers, previously trained in the use of the GQL and using the instrument in ordinary follow-up schemes. Only patients judged by their case managers having a stable and unchanged life situation for several years were considered. Patients rated themselves on the GQL on four occasions at quarterly intervals. At those occasions the participant were asked by the case manager to briefly report notable changes in health or life since last time, thus controlling for participants' subjective meaning. No participant was excluded from analysis as a result of reported events, though in some cases events were important at the moment. Their life situation was however globally judged stable rather than obviously changed. Six persons failed to fill out the GQL at one or more time due to missed appointments, not actual refusal to participate. Thus, remaining for analysis in the study were 24 participants.
In two studies the GQL was validated against two other quality of life instruments, the Inventory of Problems and Solutions (IPS) (described in Paper I) and Manchester Short Assessment Quality of Life (MANSA).

The IPS study was carried out at Växjö Rehabilitation Centre case management service, where the IPS had been used at initial contacts and at yearly follow-ups. In the study 45 clients in need for case management service because of long-standing impaired social functioning due to mental disturbance rated themselves on the GQL scale, introduced by their case manager prior to the IPS at the yearly follow-up meeting. The GQL was administered before the IPS in order to avoid influence from the detailed IPS items. Protocols were anonymized and subsequently analyzed by the first author.

The MANSA study was carried out at a Psychosis rehabilitation ambulatory department in the Swedish city of Halmstad, where MANSA was previously used, as well as the GQL scale as part of the Quality Star method. For the study, 15 patients (eight men and seven women), aged mean 41.87 years (SD= 14.27) rated themselves on both instruments. The GQL was rated first, and to avoid that responses on the GQL should not be influence answers on the MANSA, the case manager arranged a pause between the two ratings. Data were subsequently anonymized and analyzed by the first author.

In the 4th study for validation of the GQL data from the participating centers in the "Quality Star" network was used. Patients included in this study were those where patient subjective measures (GQL, consumer satisfaction and subjective distress) were all rated (N= 1919). Patients were a mixture of serious mentally ill persons with a dominance of schizophrenia spectrum disorders (83.8% ICD codes F20-F29). Core schizophrenia patients (F20) made up 56.3% and 11.3 were of the schizo-affective type (F25). The rest included for instance affective (6.6%), somatoform (3.4%), and personality (2.8%) disorders. Male/female ratio was 56.2/43.8%. Mean age was 43.99 years (SD=12.73). Reported mean duration of illness and their fist contact with psychiatric services was in most cases about 15 -20 years previously, although smaller numbers had shorter history and, for instance,
some units for early psychosis interventions have reported cases. The study was cross-sectional using data from the first time the Quality Star was used.

The personal were trained in use of the instruments following the manual (GGG-group, 2009). Instructions include that GQL, as well as the other subjective instruments, should be used in an as neutral situation as possible, for instance not directly after focus on topics that may influence rating. Subjective instrument should not be used immediately following each other, for the same reason.

Entries to the local Quality Star software were transferred to the national data-base at intervals where personal identification is automatically replaced by random identification.

2.3.3 Instruments

Global Quality of Life scale (GQL)

The GQL is a Visual Analogue Scale (VAS) (Everitt & Wykes, 1999) with the introductory question "How do you find your life just now?" and the anchor-points of the VAS line are "The worst imaginable life situation" and "The best imaginable life situation". The slight change in wording compared with the original GWB is due to the concern of the understanding in Swedish of the phrase "your life as a whole" in the originals. According to an additional instruction, staff presenting the scale can clarify, when needed, that quality of life often is understood as a global perception taking in account a persons health situation, living conditions (housing, economy, daily activities, etc.), your being with others, and how you find your life as such. The scale is a 10 cm line, thus giving a scale of 0-100 mm ranging from "Worst imaginable life situation" to "Best imaginable life situation". It is vertical, as is Cantril's ladder (Cantril, 1965), which was assumed to be clearer to respondents than a horizontal line. The GQL manual (GGG-group, 2009) suggests that the respondent finally is asked what in the life situation that made him/her put the cross at the actual place. Normal adult values for the GQL scale (Ivarsson, 2001) and variations for age and diagnostic groups and between services (Eiman & Ivarsson, 2009) have previously been shown.
Inventory of problems and Solutions (IPS) IPS is an instrument primarily intended to support working together with patients regarding problem finding and solving, but the total sum of items can be regarded as a measure of quality of life. It consists of 14 subscales with a total of 99 items. The items are a list of quality of life aspects, introduced by an instruction to encircle items which are “perceived as unsatisfactory in life right now”, which gives the value 1 and non-encircled items value 0. The IPS-total and subscales are calculated through simple averaging the scores and then multiply each with 100 i.e. giving values expressed as percentage of possible marks (0-100). The consistency of the scale is indicated by Cronbach's alpha 0.81.

Manchester Short Assessment of Quality of Life (MANSA). This scale contains 12 subjective quality of life items. The first item, "Life as a whole", is the original for the stand alone version – the GQL. The subjective satisfaction items are rated on a seven-point scale (1 = Couldn't be worse, 7 = Couldn't be better). The mean score of all satisfaction ratings is used as the total measure for subjective quality of life (MANSA-tot). The MANSA has been validated in Sweden for serious mental ill (SMI) persons, where the internal consistency was confirmed adequate (Cronbach's alpha= 0.81) (Björkman & Svensson, 2005).

Perceived Global Distress (PGD). See section 2.4.3

Perceived Global Burden (PGB) This is a visual analogue scale (Everitt & Wykes, 1999) with the introductory question "How big a burden have You experienced (felt) the last month due to your next of kin's psychiatric problems?" and the anchor-points of the VAS line are marked "No burdening feeling" and "The greatest possible feeling of burden". The scale is a 10 cm line, thus giving a scale 0-100 mm. The PGB manual (GGG-group, 2009) suggests that the respondent finally is asked what kind of burden made him/her put the cross at the actual place. The PGB has been validated in two studies (Erdner & Eiman, 2003; Hjärthag, Helldin, &
Norlander, 2008) using the Care Burden Scale for Relatives (Bergmark, Durling, Boström, & Wistedt, 1990) indicating strong associations to several subscales.

The UKU-Consumer Satisfaction Rating Scale (UKU-ConSat)
See description in section 2.2.3 and 2.2.7.

Global Assessment of Functioning (GAF), split version with separate ratings of Symptoms (GAF-S) and Functioning (GAF-F). The GAF instrument, the axis V of the DSM system, measures global mental health from the perspective of psychic, social, and functional ability (American Psychiatric Association, 2000). The scale has 10 vignettes exemplifying symptom severity and psychosocial functioning to be used as reference in rating. Each vignette give guidance to successive 10-point intervals for the semi-quantifying within the total scale range 1-100. In each vignette the first part of the text exemplifies syndrome severity and the last part psychosocial functioning. GAF is a much used scale and its psychometric properties are documented in several studies (e.g., Coffey, Jones, & Thornicroft, 1996; Patterson & Lee, 1995; Soderberg & Tungström 2007; Yamauchi, Ono, Baba, & Ikegami, 2001).

The split-GAF version is used, with separate ratings of symptom severity (GAF-S) and psychosocial functioning (GAF-F) (Jones, Thornicroft, Coffey & Dunn, 1995; Pedersen et al., 2007).

UKU-Side Effect Rating Scale (SERS). The first part of this instrument is used in the Quality Star network as a tool for health screening. The SERS comprises in its first part an inventory of 48 symptoms (45 for women and 42 for men) with well-defined items and scale steps clustered into psychic, neurological, autonomic and other symptoms areas (Lingjaerde, Ahlfors, Bech, Dencker, & Elgen, 1987). Although it was primarily constructed for monitoring side effects in drug treatment, its broad content allows it to be used as a screening devise for health problems, regardless if they are side effects, part of a parallel disorder or part of a psychiatric syndrome. The scoring pattern is: 1, Not present, 2, Possible presence or discrete, 3, Clearly present, 4, Prominent. In the second part of SERS a judgment is made if the symptom present
is or is not a medication side effect. Only this first part, the "health screening" is used in this study for validation of the GQL. The summary index used consists of the sum of ratings on all items, i.e. if no symptom is observed the sum is 45 or 42 for woman and men respectively. The rating in percentage of maximum possible rating for each sex is used for calculating a sex neutral index (SERS Index, 0-1 00).

2.3.4 Procedure

The test-retest study was approved by the heads of the three participating psychiatric departments, considered as a quality assurance project concerning the GQL already in clinical use. Patients in ordinary praxis, judged by the case manager to have a stable situation, as described in Design section, were briefly informed of the study and asked if they were willing to participate after that they received information on ethical questions according to the Helsinki declaration. It was explained that the GQL was to be filled out on four occasions at quarterly intervals. At those occasions the participant would also be asked to briefly report notable changes in health or life situation since last time. The GQL was presented by the case manager according to the manual as described in the Instruments section. In addition - in this study - participants were also specifically asked if important life events had occurred during the last three months. The answers were coded by the case managers as positive or negative events or none. When ratings had been completed, the protocols were handed over to the first author for analysis.

The validation study using the IPS scale was part of a program quality review decided by the head of the Växjö Rehabilitation Centre case management service, where the first author was invited as a consultant. The head of the Centre then introduced the study together with the researcher to all personnel and gave information on privacy rules to be observed and all case managers were trained in the use of GQL. Case managers subsequently informed clients about the study including that data files were to be kept anonymous to the researcher, i.e. coded, where legend was in the hand of the case manager only. Participants were also informed about their right not to participate and right to withdraw from the study at any time. The case
manager introduced the quality study to their clients at an early phase of contact with a new client or at the yearly follow-up meeting. The GQL was administered before the IPS in order to avoid influence from the many and detailed IPS items.

The validation study using the MANSA scale took place at a psychosis rehabilitation ambulatory department in Halmstad, Sweden by decision of the head psychiatrists in cooperation with the Quality Star network group. Participating patients were informed by their case managers that the purpose was to find out if the MANSA was suitable to use in addition to already used instruments (in particular the GQL) and that one wanted to try it with a limited number of users and make statistical comparisons. Clients were asked if willing to participate, and if not willing, this would not in any way affect their treatment. Information was also given that integrity was safeguarded as only sex and age were entered in addition to ratings and as data were to be used only by the investigating group and presentations would exclude the possibility to identify an individual's response.

Validation study using the Quality Star database. The procedure for use of the Quality Star set of instruments is described in section 2.1.2. The patient data described in the Design section was extracted for analyses.

2.3.5 Statistics

Test-retest reliability of the GQL was examined using Pearson as well as Spearman correlations (5% level) between ratings at quarterly intervals and for 6-month, 9-month and 1-year intervals. Mean ratings on the GQL on the four quarterly occasions and differences between means (Wilcoxon Matched-Pair, 5% level) were calculated. In order to clarify if reports of life events between ratings (positive, negative or non) were associated with changes in GQL ratings mean differences were calculated for such ratings separately.

In the validation study using the IPS descriptive statistics for GQL, IPS-total and the 14 IPS dimensions were calculated. Correlations between GQL and the IPS (i.e. IPS-total and averaged IPS dimensions) were investigated with Spearman's _rho_ (5% level)
In the validation study using the MANSA descriptive statistics for GQL, MANSA-tot, and MANSA subjective items were calculated. Correlations between GQL and the MANSA-total and single items were investigated with Spearman's \( \rho \) (5% level). In addition Pearson correlation between GQL and its the initial item of life satisfaction scale of MANSA, "Life as a whole", the original for the stand alone version – GQL was calculated (requirements for normal distributions were met).

In the validation study using the Quality Star database descriptive statistics for the GQL and the validating measures, ConSat-P, PGD, PGB, GAF-S, GAF-F, and SERS, and Spearman’s correlations between GQL and these measures were calculated.

**2.3.6 Results**

In the GQL test-retest study mean ratings on the four quarterly occasions were 64.92 (SD=24.12, Range=9-100), 60.25 (SD=25.85, Range=16-100), 66.47 (SD=25.16, Range=7-100) and 64.75 (SD=26.40, Range=4-100), respectively, with no significant differences between means (Wilcoxon Matched-Pair, 5% level). Correlations between the three ratings 3 months apart were \( \rho = 0.80 - 0.90 \). The correlation between 1-year distant ratings was \( \rho = 0.65 \) and intermediary time distances between ratings correlated \( \rho = 0.87 \) and \( \rho = 0.70 \). Comments that positive or negative life events had occurred between ratings were reported by 18 of the 24 participants one or more times. No participant was excluded from analysis as a result of reported events, though in some cases events were important at the moment. Their life situation was globally judged stable rather than obviously changed. When a negative event was reported mean difference between quarterly ratings was -8.6 (SD = 16.24) and correlations between successive ratings \( \rho = 0.77 \) (\( p < 0.001 \)). In cases of positive or no events reported mean differences were only -0.2 (SD=8.46) and 2.6 (SD=15.83), respectively, with correlations between successive occasions \( \rho = 0.81 \) (\( p < 0.001 \)) and \( \rho = 0.78 \) (\( p < 0.001 \)) respectively. Thus, a recent negative event may be reflected in GQL change to some extent, though the main finding was that with the participants being chosen with main criterion to have a stable situation for long
time, the similarities between quarterly GQL ratings was notable. Results indicate that test-retest properties of the GQL are acceptable.

In the GQL validating study using the IPS the subscale "Anything else unsatisfactory" was excluded as it was not marked by any of the participants. Correlation between GQL and IPS total was $\rho = -0.55$ indicating an association between better global subjective quality of life and less problems declared. Correlations were also found between GQL and IPS subscales “Inner experience” ($\rho = -0.61$), and "Contacts" ($\rho = -0.49$), indicating the importance of satisfaction with inner life and relations to the perception of quality of life.

In the GQL validating study using the MANSA scale the GQL correlation (Spearman) with the MANSA-tot was $\rho = 0.66$. Single item correlation with GQL were found for items "Life as a whole", $\rho = 0.86$, "Personal safety" ($\rho = 0.72$), "Sex life satisfaction" ($\rho = 0.69$), Physical health ($\rho = 0.68$), "Leisure activities" ($\rho = 0.59$), and "Mental health" ($\rho = 0.57$). Pearson correlation between GQL and the initial item of life satisfaction scale of MANSA, "Life as a whole", was $r = 0.84$, which provides a measure for concurrent validity, bearing in mind that GQL is a VAS scale and the MANSA uses ordinal scales.

In the GQL validating study using Quality Star data Spearman's correlations between the GQL and the subjective distress measure PGD was $\rho = 0.55$. Associations, with $\rho$ in the 0.3 level, were also found with service satisfaction (ConSat), relatives' perceived burden (PGB), as well as with health problems identified together with the case manager (SERS). Lowest correlations were found with the professional ratings of syndrome severity (GAF-S) and psychosocial functioning (GAF-F), both around $\rho 0.2$. 
2.4 Paper III. Psychometric properties of the Perceived Global Distress Scale

2.4.1 Aim

The Perceived Global Distress scale (PGD) is an adaptation for stand-alone use of the last item “How satisfied are you with your mental health” in Manchester Short Assessment of Quality of Life (MANSA) (Priebe et al., 1999). The aim of the present study was to investigate the psychometric properties of the PGD scale for seriously mentally ill persons with predominantly schizophrenia disorders.

2.4.2 Design

In the validation study of the Perceived Global Distress scale (PGD) data from the participating centers in the "Quality Star" network was used. Recordings where registration was complete for all patient and professional instruments were used. They were 937 men and 733 women (total 1670), where the Quality Star had been used at one or more occasions during a ten year period at psychiatric centers in 13 areas in Sweden. At first recording used, their mean age was 43.77 years (SD = 12.15). The majority, 87.87 %, had schizophrenia spectrum disorders (ICD codes F20-F29). Remaining patients had representations in particular from depressive and anxiety states (F30 – F49). They were, by large, severely ill patients in long-term treatment, support and rehabilitation schemes. Their reported mean duration of illness was in the order of 17 years (SD about 11).

In addition, ratings on the PGB done by important others were used when completed, as well as a smaller number of ratings on optional instruments. The smaller number of patients available in some of the analyses was due to the naturalistic nature of the data, as services use the instruments when locally considered useful in the specific clinical situation only. These sub-samples were controlled regarding age, sex and diagnosis composition to support their uses in the study. These controls were
described in the Instruments section of Paper III. The personal had been trained in use of the instruments following the manual (GGG-group, 2009), as previously described in section 3.2.3. Entries to the local Quality Star software were transferred to the national database at intervals where personal identification is automatically replaced by random identification, from which the described patient data were extracted for analysis. Analyses were performed to study a) content validity was clarified by associations with a number of validating measures, b) Concurrent validity with the last item of life satisfaction scale MANSA, and c) clinical test-retest reliability.

2.4.3 Instruments

Patient self report instruments used were:

**Perceived Global Distress scale (PGD)** The PGD is a visual analogue scale (VAS) scale (Everitt & Wykes, 1999) in a similar way as the GQL, i.e. it is a stand-alone version of the last item of the life satisfaction scale MANSA, “Mental health”. The introductory question has the wording "How much have you been bothered by your psychiatric problems during the last month?" and the anchor-points of the VAS line are marked "I have not experienced any psychiatric problems at all" and "My psychiatric problems have troubled me extremely much". The scale is a 10 cm line, thus giving a scale 0-100 mm. The PGD manual (GGG-group, 2009) suggests that the respondent finally is asked what kind of distress made him/her put the cross at the actual place. Two studies have been conducted in order to shed light on the validity of the PGD (Bergman 2003; Söderberg 2007). The results indicate that the PGD correlates with scales measuring symptom severity, depression and anxiety (r about 0.50, p <0.01).

**Quality of Life-100 (QOL-100).** The Quality of Life-100 scale was constructed primarily to support working together with patients regarding problem finding and solving (Skantze et al., 1992). The inventory contains 100 items in 14 dimensions
(subscales) including Housing, Household and self-care, Leisure, Housing environment, Community services, Knowledge and education, Contacts, Physical health, Mental health, Work, Finances/Savings, Inner experience, Dependence, and Anything else unsatisfactory. The number of items in each domain differs. The patient is instructed to encircle items that are perceived as unsatisfactory in life right now. This gives the value 1 and no mark gives value 0. The subscales are calculated through simple averaging the scores and expressing the percentage of items rated unsatisfactory. Total score is obtained by expressing the percentage of total number of items encircled. Test-retest reliability has been found good. Correlations between subscales and total sum varies between \( r = 0.48 \) to 0.87. The total sum is considered as a quality of life measure (Skantze & Malm, 1994).

Symptom Check List – 90 (SCL-90). The SCL-90 is a psychiatric self-report inventory of distress with 90 items scored on a five-point Likert scale with anchor points 0: none, 1: a little bit, 2: moderately, 3: quite a bit, and 4: extremely, indicating the rate of occurrence of the symptom/distress during the last week (Derogatis, Lipman, & Covi, 1973). The SCL-90 is well-suited for measuring general mental health and changes in symptoms (Bech, et al., 1993; Derogatis, 1994; Derogatis, 2000). The SCL-90 has been used as a central outcome measure in numerous clinical trials, as an outcome measure, as a measure of mental status, and as a screening instrument. It has been shown to have a good reliability with high internal consistency. It discriminates patients from normal controls well and there is support for its validity as a measure of general symptom severity, but less support for its suggested dimensionality and the nine subscales should thus be treated with some caution. (Fridell et al., 2002; Holi, 2003). Validation of Swedish versions in use, have been done by Fridell et al. (2002).

Consumer Satisfaction Rating Scale - self-rating version (ConSat-P). See description in section 2.2.3 and 2.2.7.

Manchester Short Assessment of Quality of Life (MANSA) See section 2.3.3.
Global Quality of Life scale (GQL) See section 2.3.3.

Instruments used by important/significant others:

Care Burden Scale for Relatives (CBS-R). This comprehensive scale was developed to describe the burden of relatives and its change after interventions (Bergmark, et al., 1990). Six areas are covered with a total of 92 items. Items are constructed as 4-point scales using anchors of 1: No, 2: Sometimes, 3: Often and 4: All the time. The six areas are: emotional burden, day-to-day aspects of burden, effects on health, work, siblings, and, finally, views on medication and psychiatry. Psychometric properties of the CBS-R were studied by Hjärthag et al. (2008), who found internal consistency of the scale (Cronbach’s alpha 0.96). Concurrent validity, tested by total score correlations with the Clinical Global Impression scale, GAF, Camberwell Assessment of Needs, was plausible ($r_s = 0.28$ to 0.44). Content validity was further clarified by step-wise regression of all CBR-S items using the PGB as criterion variable.

Perceived Global Burden (PGB See section 2.3.3.

Professional instruments used:

Brief Psychiatric Rating Scale (BPRS). The 24 item BPRS, version 4.0 was constructed to follow psychotic and affective symptoms in serious mentally ill persons (Ventura et al., 1993). Ratings are based on semi-structured interview (14 items) and observations (10 items). A detailed manual contains interview questions, symptom definitions and specific anchor points for the 1 – 7 rating levels. The use requires training, and with trained raters good inter-rater reliability is reported (ICC in the range of 0.8) (Ventura, Lukoff, Nuechterlein, Subotnik, & Gilbert, 1995).
Health screening, using the Side Effect Rating Scale (SERS). See section 2.3.3.

Global Assessment of Functioning (GAF), split version with separate ratings of Symptoms (GAF-S) and Functioning (GAF-F). See section 2.3.3.

2.4.4 Procedure

The procedure for the clinical routine follow-up use of the Quality Star instruments was described in section 2.1.2. The patient data described in Design section was extracted and analyses were performed to study a) content validity clarified by associations with a number of validating measures, b) Concurrent validity with the last item of life satisfaction scale MANSa, and c) clinical test-retest reliability.

2.4.5 Statistics

Since data usually did not meet the demands for normal distribution Spearman’s correlations were used for analyses. Correlations between PGD and total sums of the used validation instruments, as well as subscales and items, was analyzed in order to detect content validating associations. Spearman’s correlation with last item of the life satisfaction scale MANSa, “How satisfied are you with your mental health” was used as a measure of concurrent validity.

As a measure for clinical test-retest reliability of the PGD scale Spearman’s correlations between a 2nd and 3rd consecutive year ratings was studied for patients in the Quality Star data base, who had previous years, i.e. 1st and 2nd recorded year, rated PGD on a equal level, not exceeding +/- 9 points between years. This was chosen as indicating a subjective near equal perception of distress the two years. As a control for that the mental situation for the patients were reasonably equal also for year 2 and 3, the professional GAF-symptom ratings was used. For the analysis only patients that did not differ more than +/- 9 points on the GAF-symptom
scale was used, assuming that this professional judgment is indicating that symptom severity was near equal both years for this subgroup.

2.4.6 Results

Content validation. The correlations between the PGD scale and the used instruments to show associations relevant for content validation are reported in three sections: Correlations with other self-report instruments (Part 1), instruments used by important/significant others (Part 2) and professionally used instruments (Part 3). Finally, the analyses regarding test-retest properties of the PGD are reported (part 4).

Part 1 - PGD correlations with other self-report instruments. Spearman’s correlation between PGD and the quality of life inventory QLS-100 total sum was \( \rho = -0.39 \) and at about the same level with items covering perceived problems with Mental health, Feeling understood by others, Shopping, and Inner harmony. The correlation with the Total index of the self-report symptom inventory SCL-90 was \( \rho = -0.54 \), and in the same order with its indices for Interpersonal difficulties and Anxiety. Correlation with the Depression index was stronger (\( \rho = -0.64 \)). The correlation with the consumer satisfaction scale ConSat-P total sum was low, though \( \rho = 0.27 \) for one of the items (Perceived results of treatment). No significant correlation was found for the MANSA mean, but two of the items, “Mental health” and “Life as a whole” had stronger correlations (\( \rho = 0.59 \) and \( \rho = 0.54 \)). The correlation with the global quality of life scale GQL was \( \rho = 0.55 \).

Part 2 - PGD correlations with instruments used by important/significant others. No significant correlation was found with the CBS-R total score, and most notable correlation was found with the item “Problem understanding what happens” (\( \rho = -0.28 \)). Correlation with the global burden scale PGB was \( \rho = -0.32 \).

Part 3 - PGD correlations with professional instruments. Correlations between PGD and the total scores of the two symptom rating scales BPRS and SERS were in the same order (\( \rho = -0.41 \) and \( \rho = -0.45 \) respectively). Several items had similar or near similar correlations: BPRS items (in falling order of correlation) “Anxiety “(\( \rho = -0.43 \)), “Depression”, ”Hallucinations”, and “Guilt” (-0.32), and
SERS items “Tension/Inner unrest” \( (\rho =-0.48) \), "Depression", "Concentration difficulties", and “Aslenia/Lassitude/ /Increased fatigability” \( (\rho =-0.34) \). (For details, see Table 1 in Paper III). The symptom severity rating with GAF-S had lower correlation \( (\rho =0.24) \).

**Concurrent validation.** Correlation between PGD and the last item of life satisfaction scale MANSA, “How satisfied are you with your mental health” was (as noted above) \( \rho =0.59 \), indicating reasonably good concurrent validity, considering the differences in construction pointed out in the Instruments section.

**Clinical test-retest reliability.** In the subgroup with inclusion critera as described in the Design section, i.e. with a stable situation, correlations between a 2nd and 3rd year PGD ratings was \( \rho =0.75 \) \((p < 0.001)\), indicating acceptable clinical test-retest reliability.

### 2.5 Paper IV. Levels of Social Function and patients subjective Measures

#### 2.5.1 Aim

The aim of the present study was to further investigate the relationships between the patient subjective measures of consumer satisfaction, perceived distress and quality of life for severely mental ill patients with regard to different functional levels and gender, and thus study the discriminative properties of the subjective instruments by exploring if there are differences regarding the patient subjective measures at different psychosocial functional levels and between genders in the multi-center cohort, using a balanced mix of subjective and clinician ratings in the outcome-informed model for a clinical management based on shared decision making, "The Quality Star”.

#### 2.5.2 Design

Data from the participating centers in the "Quality Star" network was used. First entries where registration was complete for all patient and professional instruments
were used. They were 2552 patients, 1340 men and 1212 women (52.5 and 47.5 percents respectively), accumulated during a ten year period at psychiatric centers in 13 areas in Sweden. At first recording used, their mean age was 44.23 years (SD = 13.21), men somewhat younger than women (43.60 years of age and 44.93 respectively. The majority, 83.23 %, had schizophrenia spectrum disorders (ICD codes F20-F29). Remaining patients had representations in particular from affective disorders (F3 chapter, 5.21 %), anxiety states (F4, 2.70%), eating disorders (F5, 3.17%), and Personality disorders (F6, 2.12%). They were, by large, severely ill patients (SMI) in long-term treatment and support and rehabilitation schemes. Duration of illness mean was in the order of 17 years (SD about 12), based on data available from 77.9 % of the cases.

The personal at participating centers had been trained in use of the instruments following the manual (GGG-group, 2009), as previously described in section 3.2.3. Their entries to the local Quality Star software are transferred to the national data-base at intervals where personal identification is automatically replaced by random identification, from which the described patient data were extracted for analysis according to aims of the study. An analysis model was formed where the dependent variables were the scales for satisfaction with treatment and service (ConSat-P), the subjective global quality of life scale (GQL) and the perceived global mental distress scale (PGD) scales. The independent variables were gender (man/woman) and the Global functioning scale (GAF-F) according to the Split-GAF method.

2.5.3 Instruments

The instruments used in this study are previously described:

Consumer Satisfaction Rating Scale - self-rating version (ConSat-P). See description in section 2.2.3 and 2.2.7.

Global Quality of Life scale (GQL) See section 2.3.3.
Perceived Global Distress scale (PGD) See section 2.4.3

Global Assessment of Functioning (GAF), split version with separate ratings of Symptoms (GAF-S) and Functioning (GAF-F). See section 2.3.3.

2.5.4 Procedure

The procedure for the use of instruments by the clinical routine follow-up Quality Star network was described previously in section 2.1.2. The patient data as described in Design section was extracted for analyses.

2.5.5 Statistics

As a preparatory step four GAF-F categories were constructed based on the frequency distribution of data, GAF-F 41-50, GAF-F 51-60, GAF-F 60 and above, named “Very low”, “Low, “Intermediate”, and “Higher”. Descriptive statistics, i.e. means and standard deviations for the Social Function categories and Gender in regard to consumer’s satisfaction (ConSat), global quality of life (GQL), and perceived global distress (PGD) was calculated. A Pillai’s MANOVA 4 x 2 factorial design analyse with Post hoc testing (Tukey-HSD, 5 % level) was conducted with Social Function categories and Gender (men, women) as independent variables, and consumers satisfaction (ConSat), global quality of life (GQL), and perceived global distress (PGD) as dependent variables, with subsequent trend tests for all dependent variables (Difference Custom Hypothesis Tests, 5 % level). Univariate F-tests were further used to explore effects of functional levels and gender for the dependent variables-
2.5.6 Results

The MANOVA analysis revealed significant effects for Social Function ($p < 0.001$, $Eta^2 = 0.027$, power $> 0.99$) and for Gender ($p < 0.001$, $Eta^2 = 0.008$, power $= 0.98$). However, the analysis did not show a significant main effect for the interaction Social Function x Gender ($p = 0.273$, $Eta^2 = 0.001$, power $= 0.62$).

Results from univariate F-tests regarding Social Function showed significant effects for ConSat [$F (3, 2544) = 42.19$, $p < 0.001$], GQL [$F (3, 2544) = 30.59$, $p < 0.001$], and PDG [$F (3, 2544) = 38.03$, $p < 0.001$]. Post hoc testing (Tukey-HSD, 5 % level) showed concerning ConSat significant effects between all the four groups according to a trend where the group with the higher function was the most satisfied with the care while the group with very low function was the least satisfied group. Similar pattern was found for GQL, where those with the best function scored more positively while those with the worst function score more negatively (even though there were no significant effects in regard to the low and intermediate groups) and for PGD (even though there was no significant difference between the very low function group and the low function group). Subsequently trend tests (Difference Custom Hypothesis Tests, 5 % level) confirmed significant trends for all dependent variables indicating that the higher the social function, the higher would participants score on dependent variables.

Univariate F-tests for Gender showed significant effects for ConSat [$F (1, 2544) = 7.02$, $p = 0.008$] and PGD [$F (1, 2544) = 10.79$, $p = 0.001$]. Descriptive analysis showed that women were more satisfied with the care but also more distressed as compared to men.
3. GENERAL DISCUSSION

3.1 Introduction

This thesis is part of a research program carried out at Karlstad University for analyses of the material collected in the Quality Star network during the first ten years with the purpose of (a) assessing the psychometric properties of the Quality Star and (b) investigating group differences within the cohort between patients with different background characteristics and with different intervention patterns.

The validity studies of the three patient subjective instruments for consumer satisfaction (Paper I), quality of life (Paper II), and perceived distress (Paper III) contributing to the first part of the project, are discussed in Section 3.2. The study discussed in Section 3.3 regarding differences by gender and function of the subjective ratings (Paper IV), is adding a second study to a previous (Nordén, Ivarsson, Malm, & Norlander, 2011) in a planned series of works aimed to investigating group differences within the cohort between patients with different background characteristics and with different intervention patterns.

The discussion thereafter proceeds to discuss the Quality Star method as a strategy for combined a) support for informed user dialogue, and b) simultaneous generation of follow-up data (Section 3.4). Such deliberations are appropriate, in order to, finally, (in section 3.5) discuss the need for further research regarding the subjective measures and ways to the development of the Quality Star method.

3.2 Validity of patient subjective instruments of Quality Star supported.

The main mission with the development of “the Quality Star” concept was to promote that the patient subjective perspective was to be strengthened in care and treatment praxis and to demonstrate that global user perspective instruments were
possible to use and record as no routine standard documentation in this regard was widely used in Sweden. A process aimed to create a minimal generic platform for follow-up of psychiatric care in a multi-dimensional, holistic perspective was started in the late 90ies.

A total quality assurance package derived from an international validation project of the Integrated Care concept (Falloon et al., 2004) was 1996 presented (Ivarsson et al., 1995), and field-tested at five psychiatric departments in Western Sweden 1997-1999. The concept has the main components of: program content (evidence based strategies), program fidelity monitoring, regular external audit, and outcome monitoring, i. e. the Quality Star with multi-dimensional aspects, together with the service user. This development is described in some detail in Section 1.1.

With this perspective the final choice instruments as illustrated in Figure 1 (see in Section 1.1.1, p. 3) was decided though a consensus procedure during the pilot project and followed by minor revisions during the first years in use. The domains covered are in general agreement with findings in literature (Section 1.2.1) regarding which main outcome domains and categories are relevant in a psychiatric perspective.

Regarding the patient subjective instruments of satisfaction with services, subjective quality of life, and perceived complaints, a theoretical perspective and the rational for each instrument are presented in Sections 1.2.2 – 1.2.4 and the deliberations done for choosing specific instruments are summarized in Section 1.2.5.2, where it is also explained why subjective Need assessment was not treated as an outcome domain in the Quality Star concept.

Turning to the main studies of this thesis regarding validity of the three patient subjective measures of the Quality Star (Paper I-III) a general note should be that the findings are particularly discussed in relation to the importance of clarity regarding content/construct validity, as a general dilemma in literature is the overlap of constructs in similar domains. Caution is taken to describe the Quality Star instrument constructs in order to facilitate comparisons between studies.


3.2.1 The self-rating version of the Consumer Satisfaction Rating Scale (Pat-UKU-ConSat) (Paper I)

Paper I reports the concurrent validity of the original and a self-rating version of the Consumer Satisfaction Scale.

The original interview version, the UKU-Consumer Satisfaction Rating Scale (UKU-ConsSat) is an eight-item rating scale for the assessment of consumers’ satisfaction with psychiatric care with six structure and process oriented items and two outcome oriented. The scale was demonstrated as applicable to several patient diagnostic categories (for psychoses, neuroses, organic syndromes and addictions). In conclusion, it was suggested as a useful tool for check-ups and improvement of mental health services (Ahlfors et al 2001).

The self-rating measure Pat-UKU-ConSat was constructed to reach as close a correspondence to the original as possible. In order to make the scale easily understandable and consumer-friendly, some items of the original instrument containing more than one element were divided into separate items. In addition, for the last item ‘‘General well being” (A8), a 12th item, a visual analogue scale for subjective quality of life of the Quality Star algorithm (Ivarsson, Malm, Lindström, & Norlander, 2010) was used in stead of the 11th item.

The general result of the study was that correlations between interviewers’ assessments according to the original UKU-ConSat and the corresponding patients’ self-ratings were acceptable to good ($\rho=0.67$ to $\rho=0.82$). Correlations between sub-scores on structure and process elements and outcomes were $\rho=0.81$ and $\rho=0.75$ respectively. Total score correlation was $\rho=0.83$. Cronbach’s alpha, the standard traditional method for analyzing internal consistency, generated a coefficient of 0.80 for the self-reported version, indicating acceptable consistency (Nunnally & Bernstein, 1994). A complementary Mokken analysis revealed acceptable or just acceptable unidimensionality of both scales (Loevinger coefficients in the range of 0.45 – 0.36).

It was also argued that the diagnostic mix of subjects in the study corresponds roughly to what is common in general psychiatry, and therefore, it was
suggested that the patient self-rating version should be applicable to ordinary clinical practice, provided patients are reality-oriented. The finding that a patient self-rating scale gives results comparable to those obtained by independent assessors was deemed of importance in various ways. Using the interview method takes more professional time and logistics. Another advantage is the possibility of facilitating a dialogue on the care and its quality between users and professional carers.

However, in the Quality Star, the 11th item of the self-rating scale is used together with item 10 as outcome items, and the Visual analogue scale for subjective quality of life, named the Global Quality of Life scale (GQL) is used as a separate instrument (Paper II). As the self-rating version of ConSat referred to in Papers II-IV, is the version used in the Quality Star it was necessary to supply validity information regarding the actual version used. A re-analysis of data from Paper I, as well as analysis using the Quality Star database was therefore added and reported in Appendix “Addendum to Paper I”. The analyses of the Paper I data showed that correlation between ConSat total sum of the original scale and the two self- scale versions were similar, in Paper I version rho=0.83, and in the version used in the Quality Star 0.81. Regarding internal consistency, Cronbach’s Alpha for the Paper I total scale (using item 12) was reported 0.80. Reanalysis of data for total scale in the Quality Star version (using item 11) also showed overall acceptable consistency with Alpha 0.84. This finding can probably be attributed to the fact that 6 out of the total 8 items are identical and contribute most to the total sum.

However, inconsistency was found regarding the Outcome subscale where one item in the self-rating versions are identical (Item S10), but differ regarding the second item (S12 in the Paper I version and S10 in the Quality Star version) were correlation between original ConSat items and the Quality Star version items as well as the Outcome subscores are lower, and the correlations are not strong enough to support concurrent validity with the original self-ratings outcome subscale. The scale version used in Paper II-IV should therefore properly be considered as a new construct. The findings in the re-analysis of Paper I data regarding the similarities and the differences between the two self-rating constructs was corroborated in the analyses in the larger Quality Star data base.
It can thus be concluded that the two scales total values are can be compared but outcome items and subscales are not giving identical information and are different in construct. This underlines what was noted in Section 1.2.2.1 “Conceptual Considerations” that the measures of various care satisfaction elements need special attention, regarding methodological principles and in validity aspects and exemplifies that the construction of the instruments differ between scales. Thus caution is always necessary when comparing alleged “satisfaction”. Comparisons must be done only after careful comparison of constructs.

3.2.2. Subjective Global Quality of Life Scale (GQL) (Paper II)

As noted in Section 1.2.3.1 “Conceptual Considerations” there is yet no single accepted model of quality of life in mental illness or a universal instrument to measure it, and there are numerous associations found regarding global subjective quality of life. Literature especially point to the influence of mood state as well as cognitive functioning on the subjective appraisal.

The instrument studied in this thesis is the choice of quality of life instrument for use in the Quality Star and was based on the knowledge that quality of life for mentally ill persons is often based on health related quality of life, social circumstances, existential aspects, and human relations aspects and at the same time being an individual phenomenon. The desired construct should be able to, depending on individuals' present situation, let the rating be based on emotional as well as cognitive perceptions and regardless if ratings are based on hopes, expectations, comparisons with previous situation, with reference to other's situation, in relation to common views, professional or other values, objective social or health norms. As described in Section 1.2.3.2 it was deemed that the item General Well Being (GWB) (Oliver, 1992), used in the beginning of the Lancashire Quality of Life Profile (LQLP) (Oliver et al., 1997), similar to the first item, "Life as a whole" in the abbreviated version Manchester Short Assessment of Quality of Life (MANSA) (Priebe et al., 1999), was providing a suitable construct, and a stand-alone version was developed, the Global Quality of Life scale (GQL).
In conclusion (Paper II) the GQL was found to have acceptable psychometric properties and judged valid for serious mental ill persons. Its use as easy-to-use instrument for screening of perceived global quality of life was supported. The conclusion was based on results from four studies (described in Section 2.2) where it was found that a) Test-retest reliability was satisfactory (Study I), b) Concurrent validity with the initial item of life satisfaction scale of MANSA, "Life as a whole", was demonstrated (Study III) with correlations $r=0.85$ and $p=0.86$, and c) that content validity was clarified by associations with a number of validating measures. The emerging patient concept when using the GQL was clarified by associations with IPS items covering satisfaction with Contacts and Inner experience, with MANSA items covering Life as a whole, Personal safety, Physical health, Leisure activities, Mental health, as well as Perceived Global Distress and Service satisfaction of the Quality Star. These associations are consistent with expected areas of concern for SMI persons. It is also plausible that correlations in Quality Star ratings of Relatives' perceived burden as well as with health problems identified together with the case manager are influencing the SMI persons' subjective quality of life perceptions.

The findings were strengthened by the fact that data were from several contexts in three studies (Studies II-IV), and it could be suggested that the GQL scale was valid for seriously mental ill (SMQ persons in supportive and/or rehabilitation phases in a Swedish context.

It seems as the GQL can be a successful attempt to meet proposed theoretical standpoints: “The choice of measure must rely on the particular purpose and needs” and that “for comparison purposes generic instruments of quality of life are better than illness-specific ones, at the same time as there is also a priority also to develop illness-specific quality of life measures.” (Awad et al., 1997; Simeoni et al., 2000): Also, as expressed by (Holloway & Carson, 2002) “a bewildering variety of instruments purporting to measure quality of life are in use, often without clarification of what aspect(s) of the construct are being assessed.” The GQL scale can supposedly be an alternative to have in parallel with whatever other specified quality of life scale for its own illness-specific or otherwise specified rationale. The
GQL, thus is an attempt to create a generic instrument allowing comparisons between groups. At the same time, it has been repeatedly reported during its decade of use within the Quality Star network, that the GQL can function as support to strengthen patient partnership in line with intentions described in Section 1.2.3.2.

3.2.3 The Perceived Global Distress Scale (PGD) (Paper III).

For this paper the distinction between patient reports of mental distress arising from a unique individual conceptualization and patients being invited to self-rate symptoms (or other similar professionally defined concepts) is crucial.

In Section 1.2.4.1 “Conceptual Considerations” it was noted that a review by Slade (2002a) mentioned a number of suggested conceptual categories that are relating to what is in this work treated as a concept of “mental distress”, and included in the general category of Well-being. Relevant mentioned phenomena, proposed by several authors, are for instance: Perceptions of health status, Health perceptions, Mental health. As noted in Section 1.2.1, a number of user defined outcome domains in the broader humanistic, nursing, and user perspective have been identified, which in a patient conceptualization are likely to be perceived as elements of - or overlapping - “mental distress” or its absence. This included concepts as Personal change/Changes in Myself; Believing in Change, Managing Independently Hope, Self-confidence, Sense of purpose, Positive identity, Wellbeing.

The phenomenon sought to capture by the PGD scale is the degree of feeling mentally distressed and/or having complaints regarding mental health, irrespective of how the respondent arrive to an opinion. The considerations behind the choice to use this instrument in the Quality Star are briefly described in Section 1.2.4.2. The visual analogue scale, the Perceived Global Distress scale (PGD), was constructed as an adaptation for stand-alone use of the last item “How satisfied are you with your mental health” in Manchester Short Assessment of Quality of Life (MANSA) (Priebe et al., 1999). However, the focus was narrowed by changing the introductory question to ”How much have you been bothered by your psychiatric problems during the last month?” Degrees of positive mental health are not focused,
which is a difference to the MANS. The PGD scale thus assumes that most psychiatric patients experience degrees of mental health distress, and if not, the scale does not give room for expressing degrees of satisfaction with mental health. The intention with this restriction was to guide the respondent to focus if there are mental health problems at all. Apart from obtaining a measure for degree of perceived global distress, the PGD solution was intended to support problem identification in the mental health area together with the patient. A following dialogue should clarify the precise nature of the problems that the patient experiences.

The study reported on the psychometric properties of the PGD scale (Paper III) in its main part, content validation of the scale though associative analyses using other instruments used in parallel by centers in the Quality Star network. The results (Section 2.4.6 Results) showed notable associations with features of depression and anxiety as well as perceptions of interpersonal aspects were indicated by notable correlations between participants’ PGD ratings and ratings on the subjective SCL90 (rhos about 0.5-0.6). Correlations with QLS-100 sum and several items (rhos between –0.36-0.39) were also indicating that the respondents’ interpretation of the PGD construct includes, apart from mental health, interpersonal as well as autonomy and satisfaction elements. There was also an association to the dynamics of improvement shown by correlation with the specific item in the satisfaction scale ConSat-P (rho=0.27). The strong association between the PGD rating and perception of mental health and a quality of life dimension were further underlined by correlations to the relevant MANSA items (rhos about 0.5-0.6) and total, and regarding quality of life, the correlation to the global quality of life scale GQL (rho=0.55). The inclusion of an interpersonal element in patients’ perception of mental health as rated with the PGD was further supported by the fact that the only item in the comprehensive next-of-kin burden scale CBS-R that showed notable correlation (rho=0.28), was an item about understanding the situation and correlation with the important other’s global burden scale PGB (rho=0.32). The correlations with professional ratings on the BPRS and SERS scales provided some external validation to the patient construct of the PGD. Items concerned with anxiety and depression features were notably correlated to the PGD (rhos about –0.4- -0.5).
Also the interpersonal aspect was supported (BPRS item “Guilt” \( \rho = -0.32 \)). Interestingly, there may also be indication that patients do also de facto consider psychosis specific elements, as there is notable correlation noted to the item “Hallucinations” in the BPRS, and there may have been a connection also to the negative symptoms of psychosis as SERS items for “Concentration difficulties” and “Asthenia” showed correlations \( \rho = -0.39 \) and \(-0.34\) respectively”. On the other hand the severity of symptoms did not seem to be similarly judged by patients and professionals as correlation to GAF-S was lower, \( \rho = 0.24 \). These findings were in support of views expressed by Lindström et al. (2009), who pointed out that: “patients appear to construct their appreciation of degree of illness differently than clinicians. Much of the feelings of being ill seems to be channeled via affective symptoms.”

### 3.3 Subjective ratings, differences by gender and function (Paper IV).

After having discussed the validity studies regarding the three patient subjective instruments of the Quality Star concept, their use in analysis will be dealt with in this section, followed by a discussion of the role of the instruments to promote patient participation in Section 3.4.

The research program started out at Karlstad University includes the analyzing of the material collected during the first ten years with the purpose of investigating group differences within the cohort between patients with different background characteristics and with different intervention patterns since sufficient data in same aspects were available.

As mentioned in section 1.1.2 the Quality Star data set includes, in addition to the global outcome measures, brief indicators of patient history and patient background data chosen on grounds of being possible independent or mediating factors to the dependent Quality Star subjective and professional measures. These items were kept to a minimum, but were somewhat expanded at two occasions on demand from participating centers. It has been the policy of the Quality Star network not to heavily
expand problems for caregivers with more parallel recordings to various systems, and it was expected that eventually data should be drawn from the ordinary structured electronic patient record. This is also the reason why treatment details were not earlier included in the initial dataset. With added contents, it will be possible to investigate complex relationships between the various dimensions of the Quality Star outcome measures and a number of associated variables. The project at Karlstad University is expected to continuously expand as the database develops.

In a first study in the Karlstad University project (Nordén et al., 2011) it could show that (a) women were more satisfied with the health care and had better functioning compared to men, (b) patients treated according to the Integrated Care Program (ICP) had better functioning and less severe symptoms as compared to other patients. In (Paper IV) this finding - that women were more satisfied with the health care and had better functioning compared to men - was further elaborated by studying the combined effect of gender and functioning on the subjective outcome measures, and thus inform about the discriminative properties of the subjective instruments by exploring if there were differences regarding the patient subjective measures at different psychosocial functional levels and between genders in the Quality Star multi-center cohort. In the analysis four functional levels were defined using the split-GAF. The four GAF-F categories were constructed based on the frequency distribution of data, GAF-F 40 and below, GAF-F 41-50, GAF-F 51-60, GAF-F 60 and above, named “Very low”, “Low”, “Intermediate”, and “Higher”.

The results (Section 2.5.6) of a MANOVA analysis revealed significant effects for Social Function and for Gender with respect to consumer satisfaction, quality of life, and perceived distress. However, the analysis did not show a significant main effect for the interaction Gender X Social Function. This is an interesting result that should be considered in gender mixed materials. The absence of combined main effect of the two independent variables indicate a need to consider separate presentations for men and women in analyses.

Univariate F-tests regarding Social Function showed significant effects for ConSat, GQL, and PDG. Post hoc testing (Tukey-HSD, 5 % level) showed concerning ConSat significant effects between all the four groups according to a trend where the
group with the higher function was the most satisfied with the care while the group with very low function was the least satisfied group. Similar pattern was found for GQL, where those with the best function scored more positively while those with the worst function score more negatively (even though there were no significant effects in regard to the low and intermediate groups) and for PGD (even though there was no significant difference between the very low function group and the low function group). Subsequently trend tests (Difference Custom Hypothesis Tests, 5 % level) confirmed significant trends for all dependent variables indicating that the higher the social function, the higher would participants score on dependent variables. Univariate F-tests for Gender showed significant effects for ConSat and PGD. Descriptive analysis showed that women were more satisfied with the care but also more distressed as compared to men.

These results allows three conclusions: First that the results confirm that the distribution ranges in the cohort Quality Star instruments used is sufficient to detect differences between genders and the constructed split-GAF based functional levels. Secondly, the results per se are in line with findings of others, and at the same time not confirming yet other studies, As has been summarized from literature in the theoretical Section 1.2 for each of the instruments there are variations in associations in this sense that has generally been understood as depending on differences in the patients studied and/or in the instruments constructs. Regarding the GQL and the PGD based results there is reasonable concurrent validity with their origins in thje MANSa scale reported in Paper II and Pater III. As an example therefore, it would seem correct to suggest that comparisons between the present findings can be made with some certainty with studies using that scale, whereas comparisons with studies using other quality of life or distress construct will be less secure. Similarly, regarding the consumer satisfaction total score comparisons can be safely done with studies using any of the three ConSat alternatives, whereas caution must be taken when and if the Outcome subscale is used as shown in Paper I and Addendum to Paper I and discussed in Section 3,2,1. The third conclusion is that the general tendency that subjective measures tend to have strong associations (as noted in Section 1.2.5.1) was again shown, also in the present study.
3.4 Working together using structured instruments

It has been repeatedly mentioned in this thesis and in Papers I-IV that the instruments for use in the Quality Star should fill the function to be a point of departure for a dialogue with users within the respective domains with an under-lying question "Where do we stand now? How do we improve?" The perspective with the dual purpose to support informed user dialogue in connection with care planning and shared decision making in the ordinary work and simultaneously generating of follow-up data has been central to the Quality Star network. It has been almost axiomatically assumed that it is advantageous with real-time feedback to patients and that collecting data with instruments of clinical relevance would generate data with better reliability as being “meaningful”. Possibly, the era in which the Quality Star developed with increased use of cognitive behavioral techniques in treatment has been has been a factor promoting this view.

Some references to literature has been mentioned supporting this perspective as a sound basis for achieving treatment alliance, shared decision making and user empowerment (Malm, Ivarsson, Allebeck, & Falloon, 2003; Priebe et al.,1998; McCabe et al.,2007). Some additional underpinning may at this point be relevant to mention. The over-riding perspective for the simultaneous handling of data for use in the clinical situation and for gaining follow-up data was discussed by (McCabe et al.,2007) in terms of “Outcome management” defined as a “technology of patient experience designed to help patients, payers and providers to make rational medical care-related choices based on better insight into the effect of these choices on the patient’s life”. Examples of developments are noted during the last years exploring ways to find practical solutions to knitting these objectives together. The randomized controlled trial of routine assessment and feedback of patient-reported outcomes in six European countries using a computer-mediated intervention to structure user–clinician dialogue (the DIALOG project) provides some experience to learn from Priebe et al (2007). Another example is the trial reported by Slade et al. (2006), from community care in Croydon (South London), where it was found that feedback of outcome where patients and staff completed standardized measures of needs, quality of life, mental health problems, and therapeutic alliance did not improve subjective outcomes, but
contributed to reducing psychiatric in-patient admissions. Regarding in-patient care, Puschner, Schöfer, Knaup, and Becker (2009) found patients willing and able to regularly provide outcome data and that they valued feedback. It was concluded that strategies need to be developed to improve active use of routinely collected treatment outcome data in mental health care. The United Kingdom National Health Service Trust “CORE Partnership” is using The Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM) for routine collection of data regarding counseling and psychological therapy across a large variety of mental health services. Experience suggest that for monitoring to truly benefit treatment outcomes, real time feedback is required (Evans, Mellor-Clark, Barkham, & Mothesole, 2006). Similarly, Lambert and colleagues in the United States has established a system of clinically based patient response monitoring, that provided clinicians with warnings prior to each session, and was further developed towards an empirically derived decision system including repeated administration of the Outcome Questionnaire (OQ-45) and mapping of each patient's progress trajectory (Lutz et al., 2006). A five-year follow up study (Malm, Ivarsson, Allebeck, 2011) of long-term effects of the “Optimal treatment project” (Falloon et al., 2004) at the Swedish site in Gotheburg (Malm, Ivarsson, Allebeck, & Falloon, 2003), confirm that the two-year significant advantages of improved social function and enhanced satisfaction remained when using an using an outcome-informed model for an optimal clinical management (using instruments included in the Quality Star set) based on shared decision making and an individualized and tailored approach with a coordinated support network (the “Resource Group”) to meet users’ priorities and preferences.

These examples provide support for pursuing in developing the Quality Star concept to amalgamate clinical decision support element with the idea of outcome management. A general problem encountered would be choosing what instruments that provide relevant information for different involved parties (Slade, 2002b). A number of aspects to consider in such trade-off have been elaborated by for instance by Newnham and Page ( 2010), from an outcome monitoring aspect, (Baars, Evers, Arntz, & van Merode, 2010) in a more managerial aspects, as well as Slade (2002a) in a pratical clinical outcome perspective, giving advise on how to arrive at instrument choices and their implementation.

For the basic aim of the Quality Star, to be a tool to support dialogue with the individual patient, it is agreed with for instance Priebe et al. (1998) that: “If one is to
make use of subjective assessments for the planning and delivery of care and treatment one has to use different instruments.” In this sense The Quality Star, with a balanced mix of user perceived and clinician ratings, is probably meeting demands for an outcome-informed model for a clinical management.

### 3.5 Future research and development

Though this thesis provides basic validation of the three patient subjective measures used in the Quality Star (Paper I-III) corroboration with further studies using other established instruments within the three domains would improve possibilities to make comparisons with literature.

Furthermore, the studies are restricted to using the Quality Star database with mainly serious mentally ill persons with schizophrenia spectrum disorders. Additional studies with other patient groups are needed to support the basic intention that the instruments should be generic.

In the analytic perspective, the Karlstad University project will be important to conduct multi-dimensional analyses in order to further elucidate the intricate relationships between outcome variables, patient and treatment data as the data base contents is enlarged to allow this.

The generating of data bottom-up, i.e. drawing from the clinical use of the Quality Star is heavily depending on practical user-friendly clinical logistic solutions. It was already in the start of the Quality Star development envisage that by now data would have been generated from a structured electronic patient record with attached clinical decision systems (as for instance a Quality Star module). However, national initiatives to support this has not yet materialized. In lack of this, cooperation with the emerging national psychiatric registers has been established, where instruments and items are harmonized and the Quality Star will be an add-on option in the webb portal. This may for the time being be the best way to improve conditions for broader participation.
Finally, so far no studies have been conducted regarding the effects of the Quality Star model as a tool to enhance patient dialogue, participation, and empowerment. Such studies should be added.

4. CONCLUSIONS

The main purpose of this thesis was to investigate the psychometric properties of the three main global patient subjective instruments used in a clinical follow-up model used in Sweden, "The Quality Star", which place emphasis on the client as an agent of change, taking part in shared decision making in an empowered role as collaborative partner to the professional clinicians, and at the same time provide data for bench-mark between participating centers.

The patient self-rating Consumer Satisfaction Scale was validated against the original interviewer based scale \( \text{(Paper I)} \), however omitting the 11\(^{th}\) item out of the 12 items, not needed for that purpose. As the version of the scale used in the Quality Star is using items 1-11, a re-analysis was done (Addendum). Concurrent validity for the total sum of three versions was established. It was also clarified that the three versions do not show strong concurrent validity regarding their subscore for outcome satisfaction. The differences in this regard was corroborated for the two self-rating versions in the large Quality Star data base. In conclusion the three instruments should be named by their versions and treated as three separate scales.

The internal consistency for the version used in the Quality Star was acceptable for both total score and subscores.

The visual analogue self-rating Global Quality of Life (GQL) scale \( \text{(Paper II)} \) was shown to have satisfactory test-retest reliability, and concurrent validity with its original, the “Life as a whole” item of MANSA. The patient conceptualization of the scale emerging from associative findings with a number of validating instruments were consistent with expected areas of concern for Serious Mentally Ill persons.

Similarly, in \( \text{Paper III} \) the visual analogue scale the Perceived Global Distress scale (PGD), showed acceptable clinical test-retest reliability, and
concurrent validity with its original MANS A item, “How satisfied are you with your mental health”. In associative analyses using a number of instruments the patients’ conceptualization was clarified and that especially depressive, anxiety, interpersonal and existential aspects contribute to the construct.

A second purpose of the thesis was to contribute to the usage of the patient subjective instruments of the Quality Star for analysis, as part of a planned series of analyses at Karlstad university.

In (Paper IV) a previous finding - that women were more satisfied with the health care and had better functioning compared to men - was further elaborated by studying the combined effect of gender and functioning on the subjective outcome measures, and thus informing about the discriminative properties of the subjective instruments Three conclusions were suggested: First that the results confirm that the distribution ranges in the cohort Quality Star instruments used was sufficient to detect differences between genders and the constructed split-GAF based functional levels. Secondly, the results per se were in line with observations from other research that subjective outcomes must be interpreted and compared with the understanding that they are depending on differences in the patients studied and/or in the instruments constructs. The third conclusion was that the general tendency that subjective measures tend to have strong associations was also noted in the present study.
5. REFERENCES


Ivarsson, B., Malm, U., & Swärd, K.-O. (1995). Kvalitetsteknologi som motor för att organisera bättre vård och service vid schizofreni - världens dyraste sjukdom [Quality technology as engine for organizing better care and service by...


78


6. APPENDIX

Paper I
Paper II
Paper III
Paper IV
Tools for Outcome-informed management of mental illness

Main purpose of this thesis was to investigate the psychometric properties of three patient subjective instruments used in a clinical outcome informed model, ”The Quality Star”, with emphasis on clients as collaborative partners, and - at the same time - providing data for bench-mark.

The Consumer Satisfaction Scale self-rating version was validated against the original interviewer and two self-rating versions (Paper I). Concurrent validity for total sum of three versions was established. Out of cost-effective perspective professional time is saved and logistics simplified using self-rating scales.

The visual analogue self-rating Global Quality of Life (GQL) scale (Paper II) was shown to have satisfactory test-retest reliability, and concurrent validity with its original item from MANSA. Similarly, in Paper III, the visual analogue scale Perceived Global Distress scale (PGD), showed acceptable clinical test-retest reliability, concurrent validity with its original MANSA item. Patients’ conceptualization included especially depressive, anxiety, interpersonal and existential aspects.

A second purpose of the thesis was to contribute to the Quality Star studies at Karlstad university. In Paper IV a previous finding that women were more satisfied with the care and had better social functioning compared to men was further elaborated studying combined effect of gender and functioning on subjective outcomes. Discriminative properties of the subjective scales were shown. Previous findings that subjective measures tend to have strong inter-associations was supported.

In conclusion, the subjective measures were found adequate for optimal management of mental illness. In the discussion it was underlined that subjective outcomes must be interpreted with notion of their dependence on patient differences and/or instruments constructs.