Bjørnar Torske Antonsen

Long-term clinical outcome of psychotherapeutic treatment for patients with personality disorders: findings from a randomized study
# Contents

Acknowledgements 4  
Summary 5  
List of papers 7  
Abbreviations 8  

1. Introduction 9  

2. Background  
2.1 Personality pathology 10  
2.1.1 Personality disorder, definition and current diagnostic classification 11  
2.1.2 Personality functioning 12  
2.1.3 The course of personality pathology and PDs 14  
2.2 Psychotherapy for personality disorders 17  
2.2.1 Levels of care for psychotherapy 19  
2.2.2 Long-term outcome of psychotherapy 23  
2.3 Pre-treatment variables and long-term psychotherapeutic outcome 26  
2.3.1 Predictors and moderators 26  
2.3.2 The theory of mentalization 28  

3. Aims of the present study 30  
3.1 Ullevål Personality Project and overall aim of this dissertation 31  
3.2 paper I 31  
3.3 paper II 32  
3.4 paper III 32  
3.5 paper IV 32  

4. Methods 33  
4.1 Treatments 33  
4.1.1 Step-down treatment 33  
4.1.2 Outpatient treatment 34  
4.2 Therapists 35  
4.2.1 Individual therapists 35  
4.2.2 Group therapists 36  
4.3 Diagnostic assessments 37  
4.3.1 Personality disorders 37  
4.3.2 Symptom disorders 38  
4.4 Clinical measures 39  
4.4.1 Symptom distress 39  
4.4.2 Psychosocial functioning 39  
4.4.3 Personality functioning 40  
4.4.4 Interpersonal problems 41
Acknowledgements

Many people have contributed to the realization of this project. First and foremost I would like to thank all the patients participating in Ullevål Personality Project (UPP) for sharing so much of their private lives over so many years. I would also like to thank the therapists participating in the project for having the courage to let us research their practice. Moreover, a big thanks to all employees at the Department of Personality Psychiatry involved in the collecting and organizing of UPP data. I was extremely privileged to receive an organized data file containing 6 years of hard effort.

I would like to thank my first supervisor Theresa Wilberg for her support, thoroughness, and for giving me the time and space to grow into this field in my own pace. I could not have asked for a better supervisor. My second supervisor was Sigmund Karterud whose dedication, working capacity and not least personality has truly been inspirational. Thanks to all the co-authors who contributed with the publication of the articles. A special thanks to my UPP Ph.D. precursors Espen Arnevik and Frida Gullestad Rø, their solid work truly made my past three years more enjoyable. Also thanks to the Research Group at the Department of Personality Psychiatry, especially Geir Pedersen and Siri Lundquist for answering all sorts of crazy questions. A warm thanks to all employees at the Department of Personality Psychiatry for your kindness and a lot of great lunch breaks.

Thanks to my friends and family, especially my parents for always being supportive! Thanks to the Medical Student Research Program at the University of Oslo and Trygve Leergaard for inspiring me to pursue a degree as a researcher. Also I want to thank Christian Schlüter and Bendik Høygård (the smokin bandits) for always making me look forward to another workday. Believe it or not, my limited clinical experience was truly helped by these smoking brakes (outside the walls of the hospital of course). Thanks to the heavy mental band “tung-psykiatri-bandet” Pretend Mode for friendship and good vibrations. Thanks to Else Margrethe Berg for excellent career advice and for conducting outstanding psychotherapeutic treatment for a confused and anxious medical student in over his head. Psychotherapy works!

Last but not least, thanks to “Extrastiftelsen” for financing this thesis.

Oslo, August 15, 2015
Bjørnar Antonsen,
Summary

Promising reports from studies of coherent and structured treatments for patients with personality disorders (PDs) have turned treatment pessimism to optimism for this group of patients, especially for patients with borderline PD. However, several important questions remain insufficiency addressed. Although various treatments are provided in different settings (e.g., outpatient, day-hospital, and inpatient treatments), there is still no consensus on the optimal treatment format or levels of care for patients with PDs. Also, there is great variety in the severity and types of personality pathology, both between the PDs and within specific PDs; perhaps different subgroups of patients with PD benefit from different treatment formats. Moreover, most studies of the treatment of patients with PDs only include short-term follow-up. Given the entrenched and chronic nature of PDs, long-term follow-up is central for establishing the significance of treatment programs.

In recent decades, particularly in Europe, traditional inpatient units have often been replaced by various day-hospital treatments that also serve as step-down programs. Compared to inpatient treatment and outpatient individual psychotherapy, it has been suggested that day-hospital treatment offers an optimal level of care with suitable levels of intensiveness and containment. This dissertation utilizes data from the Ullevål Personality Project (UPP), a randomized clinical trial to compare a day-hospital step-down treatment program with outpatient individual psychotherapy for 113 patients with various PDs. The patients in step-down treatment in the UPP were initially offered short-term intensive day-hospital treatment (consisting of several types of group therapy for 18 weeks) followed by long-term outpatient psychotherapy (group and individual therapy), with a corresponding reduction in treatment intensity. In contrast, the outpatient psychotherapy condition consisted of a variety of individual psychotherapies; the therapists in this treatment condition were instructed to treat the patients according to their own preferred method and practice. Patients were assessed at several time points during the 6-year follow-up period.

The findings from this dissertation support both treatment formats as viable options for patients with PDs. There were no statistically significant differences in outcome at the 6-year follow-up in the total sample of mixed PDs. However, as a group, patients in the outpatient condition experienced a marked decline in psychosocial functioning during the period between the 3- and 6-year follow-ups; in contrast, psychosocial functioning continued to improve in the step-down condition during the same period, indicating that longer-term changes were stimulated during treatment. For the subsample of patients with borderline PD, findings indicate that a more-intensive treatment format combining group and individual psychotherapy is particularly helpful for patients struggling with
emotional dysregulation, stormy relationships, and destructive acting-out behaviors, at least more than outpatient individual therapy alone. Moreover, this thesis indicates that patients’ capacity for understanding their own and others’ mental states (patients’ capacity for mentalization) may impact the outcome of treatment. Patients with a low capacity for mentalization had better outcomes in outpatient individual therapy than in the step-down program. In contrast, patients with a greater capacity for mentalization achieved better results in the step-down program. Also, findings in this dissertation indicate that mentalization capacity is associated with core aspects of personality pathology and captures clinically relevant phenomena in adult patients with PDs. Last, consistent with previous prospective naturalistic studies and treatment trials, in the UPP, patients with PD frequently continued to function at suboptimal levels of psychosocial functioning, long after achieving diagnostic remission. This impact was also evident in the UPP at the 6-year follow-up. Using statistical models for causal inference, we found strong and reciprocal positive associations between personality functioning and global psychosocial functioning. These results indicate that a successful intervention on one of the two constructs positively affects the results of the other construct.

Various explanations and implications for the above findings are considered in this dissertation, as are possible explanations for differences in long-term clinical course between treatment conditions. This dissertation also discusses how different treatment formats and different levels of care may be suitable for some patients but not for others. A limitation of this study is that it did not control for other types of treatment that patients may have received outside of the UPP.

The UPP primarily compared treatment provided in different formats and at different levels of care. However, one of the main challenges in the field of psychotherapy for PDs is to better understand how and why change occurs. To meet this challenge, theoretically promising concepts must be identified. The results from this dissertation suggest that both mentalization and personality functioning have future roles in the treatment and assessment of personality pathology. Both concepts are also interesting as potential processes that lead to change during psychotherapy. While this dissertation establishes a causal inference with regard to change in personality functioning and psychosocial functioning, such analysis is lacking for the concept of mentalization. Future research should investigate the causal inference linking improvement in mentalization capacity and psychotherapeutic outcome.
List of papers

Paper I

Step-down versus outpatient psychotherapeutic treatment for personality disorders: 6-year follow-up of the Ullevål personality project.

Published in BMC psychiatry 2014 Apr 23;14:119.

Paper II

Favorable outcome of long-term combined psychotherapy for patients with borderline personality disorder:  Six-year follow-up of a randomized study.

Published in Psychotherapy Research 2015 Aug 11:1-13

Paper III

Adaptive personality and psychosocial functioning in patients with personality disorders - A causal feedback mechanism illustrated by a repeated-measures marginal structural model.
Klungsøyr, O., Antonsen, B.T. & Wilberg, T.

Submitted for publication.

Paper IV

Is reflective functioning associated with clinical symptoms and long-term course in patients with personality disorders?

Published in Comprehensive Psychiatry Jun 3. pii: S0010-440X(15)00091-7
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAI</td>
<td>Adult attachment interview</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike’s information criteria</td>
</tr>
<tr>
<td>APA</td>
<td>American Psychiatric Association</td>
</tr>
<tr>
<td>BDI</td>
<td>Beck depression inventory</td>
</tr>
<tr>
<td>CIP</td>
<td>Circumplex of personality problems</td>
</tr>
<tr>
<td>CP</td>
<td>Combined program (different term of the step-down treatment condition)</td>
</tr>
<tr>
<td>DBT</td>
<td>Dialectical behavior therapy</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic and Statistical Manual of Mental Disorders</td>
</tr>
<tr>
<td>FFM</td>
<td>Five factor model</td>
</tr>
<tr>
<td>GAF</td>
<td>Global assessment of functioning</td>
</tr>
<tr>
<td>GLMM</td>
<td>General linear mixed modelling</td>
</tr>
<tr>
<td>GSI</td>
<td>General Symptom Index (SCL-90-R)</td>
</tr>
<tr>
<td>ISE</td>
<td>Index of self-esteem</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>MBT</td>
<td>Mentalization based treatment</td>
</tr>
<tr>
<td>MDD</td>
<td>Major Depressive Disorder</td>
</tr>
<tr>
<td>MINI</td>
<td>MINI-international Neuropsychiatric Interview for axis I diagnoses</td>
</tr>
<tr>
<td>MSM</td>
<td>Marginal Structured modeling</td>
</tr>
<tr>
<td>Neo-Pi-R</td>
<td>Revised NEO Personality Inventory</td>
</tr>
<tr>
<td>OIP</td>
<td>Outpatient individual psychotherapy (outpatient treatment condition in UPP)</td>
</tr>
<tr>
<td>OPD</td>
<td>Operationalized psychodynamic diagnostics</td>
</tr>
<tr>
<td>PD</td>
<td>Personality disorder</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>RF</td>
<td>Reflective functioning</td>
</tr>
<tr>
<td>SCID-II</td>
<td>Structured Clinical Interview for DSM-IV, axis II</td>
</tr>
<tr>
<td>SCL-90-R</td>
<td>Revised symptom checklist - 90 items version</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SFT</td>
<td>Schema focused therapy</td>
</tr>
<tr>
<td>SIPP-118</td>
<td>Severity Indices of personality problems - 118 items</td>
</tr>
<tr>
<td>SIPP-118 SF</td>
<td>Severity Indices of personality problems – short form</td>
</tr>
<tr>
<td>TFP</td>
<td>Transference focused psychotherapy</td>
</tr>
<tr>
<td>UPP</td>
<td>Ullevål Personality Project</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
</tbody>
</table>
1. Introduction

Patients with personality disorders (PDs) exhibit a wide range of clinical manifestations, from very noticeable and treatment-seeking behavior to avoidance and suspicion, from extreme self-harm to very dependent and careful behavior. However, one thing PD patients have in common is their marked dysfunction in numerous life areas. A Norwegian population study showed that the general impact of PD on quality of life was stronger than the influence of sociodemographic factors, physical-health factors, and Axis I disorders (Cramer, Torgersen, & Kringlen, 2006). Similarly, a USA-based community population study, the McLean Study of Adult Development (MSAD), reported severely impaired psychosocial functioning in patients with PDs (Zanarini, Frankenburg, Hennen, Reich, & Silk, 2005b). Poor social functioning was also demonstrated among American patients in the Collaborative Longitudinal Personality Disorders Study (CLPS) (Skodol et al., 2002), in British clinical samples of PD patients in secondary health care (Newton-Howes, Tyrer, & Weaver, 2008), and among PD patients admitted to treatment within the Norwegian Network of Personality Focused Treatment Programs (Wilberg, Karterud, Pedersen, & Urnes, 2009). In both the CLPS and the Norwegian Network, evaluations of functioning included assessments of employment, education, and domestic situations. Moreover, with regard to symptomatic distress, some studies indicate that PD patients experience more psychological distress than psychiatric patients without PDs (Noren et al., 2007). Also, personality pathology in PDs negatively affects both the recurrence and the duration of several other psychiatric disorders (Cyranowski et al., 2004; C. M. Grilo et al., 2010; Hart, Craighead, & Craighead, 2001; Newton-Howes, Tyrer, & Johnson, 2006; Shea et al., 1990).

Several studies have demonstrated that PDs are associated with high social costs in the form of disablement pension and extensive use of health-care services, as well as with indirect costs associated with failing parental functioning and crime rate (Bender et al., 2001; Cramer, Torgersen, & Kringlen, 2007; Noren et al., 2007; A. E. Skodol et al., 2005). In a British national cross-sectional study (Yang, Coid, & Tyrer, 2010) and an Australian national cross-sectional survey (Jackson & Burgess, 2002), an increased likelihood of being part-time employed, unemployed, without independent income, or disabled was associated with the diagnosis of PD; this likelihood increased with increasing severity of PD. In the MSAD, Zanarini et al. (2009) reported that ~50% of borderline PD patients were unable to support themselves. In the conclusion of a large multicenter study in the Netherlands, Soeteman et al. (2008) stated that the diagnosis of PD represented a considerable economic burden when compared to other mental and physical disorders.
The overall prevalence of PDs in the general population of western cultures is ~10% (Lenzenweger, Lane, Loranger, & Kessler, 2007; Reich, Yates, & Nduaguba, 1989; Samuels et al., 2002), while the prevalence in an urban Norwegian population (Oslo) was estimated to be as high as 13.4% (Torgersen, Kringlen, & Cramer, 2001). However, the prevalence of PDs in clinical settings is much higher, with typical estimates ranging from 30-50% in community mental-health teams (Keown, Holloway, & Kuipers, 2002; Merson et al., 2002; Newton-Howes et al., 2010) to 70-90% among more severely disturbed inpatients (Marinangeli et al., 2000; Oldham et al., 1995). Given the total burden of PD for the patients themselves, their families, and society, the development of effective treatments for this group of patients is important.

2. Background

2.1 Personality pathology

Everyone has a personality, a characteristic manner of thinking, feeling, behaving, and relating to others (Matthews, 2009; Mischel & Shoda, 1995; Westen, 1995). The characteristics of personality are clearly associated with a wide range of important life outcomes, such as subjective wellbeing, social acceptance, relationship conflict, marital status, academic success, crime, unemployment, and physical and mental health (John, 2008; Lahey, 2009; Ozer & Benet-Martinez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). However, when the characteristics of personality are reflected in behavior associated with considerable personal and social disruption, that personality is termed pathological. To date, personality pathology is one of the most debated aspects of psychiatry.

Several current theories offer distinct conceptualizations of what personality pathology is, how it should be assessed, and how it should be treated (Livesley, 2012a; Tyrer, Crawford, & Mulder, 2011; Widiger & Simonsen, 2005). In clinical practice, this debate addresses whether patients should be treated according to psychodynamic or cognitive-behavioral approaches (Feist, 2012). As in the rest of psychiatry, it is also an ongoing debate concerning the assessment of PDs. The latter debate has traditionally been dominated by the so-called medical model, which erects distinct boundaries between normality and pathology. This is reflected in the most-recent diagnostic manuals, were PDs are classified according to a categorical approach that views PDs as discrete entities that are distinct from each other (Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition; DSM-5; APA 2013 and the International Statistical Classifications of Diseases and Related Health Problems, Tenth Edition; ICD-10; WHO 2010). However, the various approaches to PD assessment are evident in the
recent DSM-5; which also contains a dimensional assessment of PDs in the section Emerging Measures and Models (DSM-5, section III; APA, 2013) that allows a continuous gradient between normality and abnormality. This dissertation includes both dimensional and categorical assessment of PDs.

2.1.1 PD definition and current diagnostic classification

In Norway, clinical assessments of PDs are made on the basis of the ICD-10 (WHO, 2010), a diagnostic classification similar to the DSM-IV and to DSM-5 section II. However, for research purposes, the DSM classification is still the most frequently used in Norway, which is also the case in this dissertation.

The general diagnostic criterion for a PD is “an enduring pattern of inner experience and behavior that deviates markedly from the individuals’ culture” (Table 1). The DSM includes 10 specific types of PDs and one category for PDs that are not otherwise specified. PDs are divided into three clusters: cluster A, the odd or eccentric cluster, which includes paranoid PD, schizotypal PD, and schizoid PD; cluster B, the dramatic or emotional cluster, which consists of borderline PD, histrionic PD, antisocial PD, and narcissistic PD; and cluster C, the anxiety cluster, which includes avoidant PD, dependent PD, and obsessive-compulsive PD.

In the DSM-IV and DSM-5, the diagnosis of PD is based on a polythetic-categorical system. “Polythetic” refers to the fact that specific PDs are defined by multiple criteria, and not all listed criteria are necessary to diagnose a PD; any combination of fulfilled criteria is sufficient to consider a PD diagnosis. The DSM is a categorical system because a PD is binary: either the patient has a PD or the patient does not. One argument for continuing to diagnose PD in a categorical manner is that clinical decisions are often categorical as well (e.g. to diagnose or not, to treat or not).

However, extensive criticism has been levied against categorical DSM classifications. Among the issues of greatest concern is the extensive co-occurrence of PD diagnoses: most patients who receive a PD diagnosis meet the criteria for more than one PD (Grant, Stinson, Dawson, Chou, & Ruan, 2005; Oldham et al., 1992). Another concern is the relatively poor convergent validity of PD criteria sets; PDs diagnosed with different methods may be only weakly related to one another (L. A. Clark, 2007; Pilkonis et al., 1995). Moreover, there is an increasing empirically based consensus that there is a dimensional continuum between normal and pathological personality (Livesley, Schroeder, Jackson, & Jang, 1994; Widiger, 1993).
Table 1. The general diagnostic criteria for Personality Disorders (DSM-IV; APA).

A. An enduring pattern of inner experience and behavior that deviates markedly from the experience of the individuals’ culture. This pattern is manifested in two (or more) of the following areas:

(1) Cognition (i.e., ways of perceiving and interpreting self, other people and events).

(2) Affectively (i.e., the range, intensity, lability and appropriateness of emotional response)

(3) Interpersonal functioning

(4) Impulse control

B. The enduring pattern is inflexible and pervasive across a broad range of personal and social situations.

C. The enduring pattern leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.

D. The pattern is stable and of long duration and its onset can be tracked back to adolescence or early childhood.

E. The enduring pattern is not better accounted for as manifestations or consequences of another mental disorder.

F. The enduring pattern is not due to direct psychological effects of a substance (i.e., a drug of abuse, a medication) or a general condition (i.e., head trauma).

2.1.2 Personality functioning

The new classification of PDs in the DSM-5 was the subject of intense debate. The American Psychiatric Association decided to retain the DSM-IV system, but as already mentioned, the DSM-5 includes a dimensional framework for the assessment of PDs. Section III contains two general criteria for PDs: criterion A requires moderate to greater impairment in personality functioning, while criterion B requires the presence of pathological personality traits. The evaluation of personality functioning in criterion A is based on assessments of impairments in self and interpersonal functioning that are consistent with multiple theories of PD, including cognitive-behavioral, interpersonal, psychodynamic, attachment, developmental, and evolution theories. Impairments in
self and interpersonal functioning have been viewed as key aspects of personality pathology that merit clinical attention (Hopwood, Schade, Krueger, Wright, & Markon, 2013; Pincus, 2011).

The importance of dysfunction in self and others is also recognized in the general criteria of the DSM-IV. The cognition area under criterion A states that “ways of perceiving and interpreting self, other people, and events” (DSM-5; APA, 2013) should be evaluated. However, the significance of self and interpersonal issues in PDs was not represented systematically or consistently (Skodol et al., 2014). In work with the Level of Personality Functioning Scale, Bender et al. (2011) reviewed several reliable and valid clinician-administered measures for assessing personality functioning and psychopathology. This review demonstrated that a self-other dimensional perspective has an empirical basis and significant clinical utility. Skodol, Bender and Oldham (2014) list numerous studies that have used measures of self and interpersonal functioning to show that a self-other approach is informative in determining the existence, type, and severity of personality pathology: Salvatore et al. (2005) reported that patients with paranoid PD typically see themselves as weak and inadequate and others as hostile and deceitful. Patients with narcissistic PD have been found to have dominant states of mind pervaded by distrust toward others and feelings of being excluded or being harmed (G. Dimaggio et al., 2008). Jovev and Jackson (2004) demonstrated that individuals with avoidant PD utilize maladaptive schemas centering on a self that is defective and shame ridden, expecting to be abandoned because of their shortcomings; persons with obsessive-compulsive PD are burdened by a schema of self-imposed, unrelenting standards. Eikenæs et al. (2013) found that patients with avoidant PD could be distinguished from patients with social phobia because they had more problems with self-esteem, identity, and relationships. Dimensions based on self-interpersonal problems have discriminated types of PD pathology and predicted various areas of psychosocial functioning (DeFife, Goldberg, & Westen, 2013; Feenstra, Hutsebaut, Verheul, & Busschbach, 2011; Peters, 2006). The severity of impairment in self and interpersonal functioning was also shown to predict treatment course and outcome (Ackerman, Hilsenroth, Clemence, Weatherill, & Fowler, 2000; Harpak-Rotem & Blatt, 2009; Vermote et al., 2010).

Verheul et al. (2008) assessed core components of personality functioning in patients and community members in the Netherlands using the self-reporting questionnaire Severity Indices of Personality Problems (SiPP-118). Twelve out of 16 facets of personality functioning distinguished patients with PDs from subjects who were psychiatrically healthy and patients with other mental disorders (Verheul et al., 2008). The 16 facets were factored into five higher-order domains: Self-control, Identity Integration, Relation Capacity, Social Concordance, and Responsibility. Each of the five domains distinguished patients with no PDs from those with one PD and those with one PD from
those with two or more PDs (Verheul et al., 2008). When this study was replicated in a sample of adolescent patients and comparison subjects, all 16 SIPP-118 personality functioning facets reflected greater impairment in patients with PDs (Feenstra et al., 2011). Patients that fulfilled more PD criteria had the most impairment in the five domains of SIPP-118, with Self-control and Identity Integration showing the largest differences (Feenstra et al., 2011). Berghuis et al. (2012) assessed personality functioning with the General Assessment of Personality Disorder and the SIPP-118, PDs with the SCID-II, and personality traits with the NEO-PI-R. Principle component analysis clearly distinguished general personality dysfunction from personality traits (Berghuis et al., 2012).

According to Berghuis (2012) the general personality dysfunction model consisted of three factors: Self-identity Dysfunction, Relational Dysfunction, and Prosocial Functioning. Taken together, the studies mentioned in this paragraph involved >4000 patients and control subjects and lend strong support for the emphasis of personality functioning in the evaluation of personality pathology.

In a sample of university students, the test-retest reliability of the SIPP-118 domains over 14-21 days was very good to excellent (Verheul et al., 2008). Further, in 60 patients treated for an average of 11 months as outpatients or in a day hospital and followed for 2 years, the SIPP-118 domains of Self-control, Identity Integration, and Responsibility gradually improved over time, Relational Capacity improved over the first year, and Social Concordance improved during the second year (Verheul et al., 2008). In a subsample adolescents treated as inpatients, 14/16 facets of the SIPP-118 showed significant improvement after 1 year, with effect sizes ranging from 0.37 to 1.24, indicating small to very large effects (Feenstra et al., 2011). These studies indicate that the self-other dimension does not merely reflect short-term changes in the clinical state, but can reflect more profound adaptive change. However, in order to investigate adaptive change during psychotherapy, treatment studies need to include personality functioning as an outcome. Moreover, the question of how potential changes in self and interpersonal functioning relate to other clinical variables is rarely addressed in the literature. This dissertation aims to contribute more knowledge about these topics; personality functioning (SIPP-118) is included as an outcome in papers II, III, and IV.

2.1.3 The course of personality pathology and PDs

Empirical knowledge of the course of personality pathology stems mostly from large clinical and non-clinical cohorts of subjects with PD, for example the CLPS and the MSAD. It is important to state that these studies do not describe the natural course of PD in terms of untreated PD; on the contrary, a multitude of psychotherapeutic outpatient and inpatient interventions and pharmacological
treatments were received over the years (Bender et al., 2006; Zanarini, Frankenburg, Khera, & Bleichmar, 2001). Given the burden of PD, it is ethically unacceptable to follow a cohort of untreated patients with severe PDs.

The CLPS (Gunderson et al., 2000; Skodol et al., 2005) is an American prospective repeated-measure study designed to examine the course of patients with one of four (DSM-IV) PD diagnoses: schizotypal PD, borderline PD, avoidant PD, or obsessive-compulsive PD. The CLPS also includes a comparison group of patients with major depressive disorder (MDD) without any PD. Various concepts of categorical and dimensional stability were investigated over 12 months (Shea & Yen, 2003), 24 months (Grilo et al., 2004), and 10 years (Gunderson et al., 2011) using data from 668 patients recruited from diverse settings at four university clinics across the USA. Shea et al. (2003) reported that compared with patients in the MDD group, a significantly greater proportion of patients in each of the four PD groups remained at the diagnostic threshold throughout the first 12 months of follow-up. Grilo et al. (2004) found that compared with the four PD groups, the MDD group had significantly shorter time to, and higher rates of, remission. However, although PDs were more stable than MDD, a substantial number of remissions occurred during the 24 months of follow-up (Grilo et al., 2004). CLPS operates with different sets of remission norms. When using a remission definition of two consecutive months with two or fewer criteria, rates were 33% for schizotypal PD, 42% for borderline PD, 47% for avoidant PD, and 55% for obsessive-compulsive PD (Grilo et al., 2004). Even with a more stringent definition of 12 consecutive months with <3 criteria, remission rates were 23% for schizotypal PD, 28% for borderline PD, 31% for avoidant PD, and 38% for obsessive-compulsive PD (Grilo et al., 2004). Gunderson et al. (2011) reported 10-year outcomes in terms of diagnostic stability and psychosocial functioning. Although remission of borderline PD was significantly slower than that of other PDs, 85% of patients with borderline PD attained remission using the 12-month definition and 91% attained it using the 2-month definition (Gunderson et al., 2011). Only 12% of patients with borderline PD experienced a relapse; this rate was lower, and the time to relapse slower, than for MDD and for the other PDs (Gunderson et al., 2011). In contrast to the substantial and durable reduction in borderline PD-specific psychopathology over time, social functioning measures continued to show severe impairment with only modest, although statistically significant, improvements over time (Gunderson et al., 2011). Social functioning in patients with borderline PD remained persistently more impaired than in the MDD group and in the group of other PDs (Gunderson et al., 2011). Taken together, these findings, which are based on 10 years of prospective follow-up, indicate that the course of borderline PD is characterized by high rates of
diagnostic remission and low rates of relapse, but also with severe and enduring impairments in psychosocial functioning.

The MSAD (Zanarini, Frankenburg, Hennen, Reich, & Silk, 2005a; Zanarini, Frankenburg, Hennen, & Silk, 2003) is an ongoing American study comparing the course and outcome of hospitalized patients with borderline PD (n=290) with a group of patients with other PDs (n=72). MSAD performs assessments every 2 years and has reported outcomes for 6 years (Zanarini et al., 2003, 2005a), 10 years (Zanarini, Frankenburg, Reich, & Fitzmaurice, 2010b), and 16 years (Zanarini, Frankenburg, Reich, & Fitzmaurice, 2012) of follow-up. As in the CLPS, the MSAD operates with different sets remission norms; in the MSAD, remission is defined as no longer meeting the criteria for borderline PD or for another PD for 24 months, 48 months, etc. MSAD uses the term “recovery,” which requires decent measures on social and vocational functioning as well as symptomatic remission. Remission rates for borderline PD were 35%, 49%, 74%, 91%, and 99% by years 2, 4, 6, 10, and 16, respectively. Reporting on findings consistent with those of early CLPS reports, Zanarini et al. (2003) concluded that “symptomatic improvement is both common and stable, even among the most disturbed borderline patients, and the symptomatic prognosis for most, but not all, severely ill borderline patients is better than previously recognized.” The MSAD also demonstrated that borderline PD is associated with significant psychosocial impairment (Zanarini, Frankenburg, Reich, & Fitzmaurice, 2010a). In contrast to findings from the CLPS, much of this impairment was associated with vocational rather than social measures (Zanarini et al., 2010a; Zanarini et al., 2009). Accordingly, in the 16-year report, Zanarini et al. (2012) showed that all patients with PDs struggled more to achieve recovery than to achieve remission. Patients with borderline PD were significantly slower to achieve recovery (40-60%) than the comparison group with other PDs (75-85%), and relapses occurred significantly faster and at a higher rate among patients with borderline PD than among patients with other PDs (Zanarini et al., 2012). In conclusion Zanarini et al. (2012) stated, “Patients with borderline PD should continue with psychotherapy after symptomatic remission to guard against relapse and to help improvement in psychosocial functioning.”

The Longitudinal Study of Personality Disorders (Lenzenweger, 1999; Lenzenweger, Loranger, Korfine, & Neff, 1997) assessed 250 students from an American university at three time points over a 4-year period. One hundred and twenty-nine students met the criteria for at least one PD, and 121 did not have any PD diagnosis. This study utilized a semi-structured interview (International Personality Disorders Examination; Loranger et al., 1994) and a self-reported measure (Millon Clinical Multiaxial Inventory-II; Millon 1987) to obtain information about personality. Dimensional scores for the PDs were characterized by significant stability for both the interview and self-reported measures. Stability
coefficients for the total set of PD features ranged from 0.61 to 0.70; cluster B PDs had the highest stability coefficients and cluster A PDs had the lowest (Lenzenweger 1999). The Longitudinal Study of Personality Disorders was limited by its relatively homogeneous study group of college students, its narrow time frame, and, most importantly, the insufficient frequency of any PD diagnosis at a categorical (diagnostic) level, which prohibited analysis of entity (Grilo & McGlashan, 2005).

The Children in the Community Study (Brook, Whiteman, Cohen, Shapiro, & Balka, 1995; Cohen, Crawford, & Johnson, 2000) is an ongoing prospective study of nearly 1000 families with children aged 1-10 years that were recruited in 1975 in New York state using a random sampling procedure. Important findings include age-related changes in PD symptoms, including moderate levels of stability throughout adolescence and early adulthood (Crawford et al., 2008). The Children in the Community Study also detected an association between PD psychopathology in adolescence and impairments in educational achievements (P. Cohen, Crawford, Johnson, & Kasen, 2005). Early forms of behavior disturbances predicted PDs during adolescence; PDs during adolescence displayed significant levels of continuity into adulthood and predicted other mental disorders and suicidality (Johnson et al., 1996) as well as violent and criminal behavior (Johnson et al., 2000). Collectively, these findings show that although many children and young people with personality psychopathology may be expected to improve, the most severely affected are likely to have problems in later life and should be followed closely. These patients may require long-term treatment to prevent the development of later impairments in functioning (Grilo & McGlashan, 2005).

In sum, prospective naturalistic studies have found that PD diagnoses are less stable than previously assumed, with high rates of diagnostic remission. These studies also showed that many patients continue to function at suboptimal levels of work and social functioning. Few of these studies described the type of treatment that these patients received, preventing us from drawing conclusions about the impact of psychotherapeutic treatment on the long-term outcomes of patients with PD. This dissertation aims to contribute to our knowledge of the long-term outcomes of psychotherapeutic treatment.

2.2 Psychotherapy for PDs

In recent decades, the empirical basis for psychotherapeutic treatment of patients with PDs has strengthened. This change can be concluded from two formal meta-analyses (Leichsenring & Leibing, 2003; Perry, Banon, & Ianni, 1999), one Cochrane review (Binks et al., 2006b), three clinical
guidelines APA (2001); National Institute for Mental Health in England (2009); National Health and Medical Research Council in Australia (2012), and several systematic or critical literature reviews (Bateman & Fonagy, 2000; Gabbard, 2000; Piper & Ogrodniczuk, 2001; Sanislow & McGlashan, 1998). Other treatments, such as pharmacological interventions, have received less empirical support (Binks et al., 2006a; Paris, 2008, 2009; Schulz et al., 2008).

Traditionally, patients with PD have been offered intensive inpatient treatment, with very little empirical evidence to support this strategy. Much of the progress in the psychotherapeutic treatment of and research into PDs has come through the development of manualized treatments for patients with borderline PD. The cognitive-behavior tradition Dialectical Behavior Therapy (DBT) (Linehan, Heard, & Armstrong, 1993) emerged in the early 1990s and is currently the most investigated psychotherapeutic treatment for PDs (Kliem, Kroger, & Kosfelder, 2010). Although integrating techniques from different traditions, Schema-Focused Therapy developed with a strong basis in cognitive-behavioral therapy (Young, 1999). Mentalization-Based Treatment (MBT) (Bateman & Fonagy, 2004) and Transference-Focused Psychotherapy (Levy et al., 2006) emerged from the psychodynamic tradition.

Although early randomized clinical trials (RCTs) reported favorable outcomes for all manualized treatments (Linehan, Armstrong, Suarez, Allmon, & Heard, 1991; Verheul et al., 2003), there is still limited evidence for the supposed superior efficacy of these manualized treatments for borderline PD (Stoffers et al., 2012). A few recent studies described similarities in the effects of specialized borderline PD treatments and so-called structured clinical care practiced by specialist clinicians (Bateman & Fonagy, 2009; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; McMain et al., 2009). This finding indicates that adequate treatment response can arise from non-manualized psychotherapy if the strategy is based on structured treatment principles (Livesley, 2012b).

Most of the existing research literature on psychotherapeutic outcome for patients with PD concerns patients with borderline PD. There is a remarkable lack of empirical evidence regarding psychotherapy and treatment for other PDs, despite the large sizes of these patient populations. Thus, more studies emphasizing the outcome of psychotherapeutic treatment for a variety of personality pathologies are needed. Moreover, although psychotherapy as a general term encompasses various types of treatment approaches, there are also large variations in treatment formats, settings, and duration (Verheul & Herbrink, 2007), the impacts of which remain mostly unexplored.
2.2.1 Levels of care for psychotherapy

According to Gunderson et al. (2005), “level of care” is a multi-dimensional construct that considers containment, intensity, structure, costs per day, and duration. Clinically speaking, levels of care can be viewed as the organization of patients into inpatient hospitalization, partial hospitalization/day treatment, and outpatient treatment, an organization that often reflects the intensity of treatment (Gunderson et al., 2005). Another aspect linked to the levels-of-care construct is the arrangement of treatment into formats such as group, individual, or conjoint/combined treatment. Since the psychotherapeutic treatment for PD patients has been and still is characterized by significant heterogeneity (Verheul & Herbrink, 2007), research on treatment provided at different levels of care is important for progress in the field. To date, few studies have sought to characterize the optimal levels and formats of care for patients with PDs.

In recent decades, particularly in Europe, traditional inpatient units have often been replaced by day-hospital treatment and later step-down treatment (Kallert et al., 2004). In step-down programs, patients are initially offered more-intensive treatment (e.g. inpatient or day-hospital treatment) followed by outpatient psychotherapy, with a corresponding reduction in treatment intensity. This step-down arrangement has been based on the assumption that the initial intensive phase of treatment enhances the patient’s ability to benefit from psychotherapy by establishing a solid treatment alliance and attachment relationships to the therapists and fellow patients, thereby increasing patient motivation (Karterud & Wilberg, 2007). The second phase, which consists of long-term continuation psychotherapy, is thought to result in more profound changes, including increased self-reflection, improved interpersonal and psychosocial functioning, personality development and integration, and fewer interpersonal problems. Taken together, these step-down treatment programs provide long-term psychotherapy in accordance with the common assumption that many patients with PD require long-term treatment (e.g. Perry et al., 1999).

The treatment offered in the intensive phase typically consists of a mix of diverse types of group therapies, including psychodynamic, cognitive-behavioral, and art-therapeutic approaches, which are sometimes combined with individual therapy (Karterud & Wilberg, 2007). The prominence given to group therapy in day-hospital treatment is thought to be beneficial for patients with PDs. Patients are expected to benefit from multiple interactions with staff and group members by obtaining feedback from several people and by having the opportunity to learn from other patients. New adaptive behavior can be practiced in the groups and maladaptive ways of relating to others can be challenged and modified through group interactions and interpersonal understanding (Bateman &
Tyrer, 2004). Compared to individual therapy, group therapy may dilute intense transference reactions toward therapists, thereby preventing acting-out behaviors (Ogrodniczuk & Piper, 2001).

Treatment in the second phase often consists of group therapy or conjoint individual and group therapy. Some programs also integrate outreach nursing or social work in this phase (Bateman & Fonagy, 2008). Conjoint-therapy formats have been regarded as valuable in the treatment of PDs. As described above, it is assumed that the interpersonal problems activated in group therapy can be more thoroughly understood and worked through in individual sessions (Karterud, Johansen, & Wilberg, 2007).

Despite the theoretical literature on day-hospital and step-down treatments for patients with PDs, the empirical data for these treatment modalities remain limited. Several naturalistic studies have reported the benefits of day-hospital treatments for patients with PDs (Karterud et al., 2003; Karterud et al., 1992; Krawitz, 1997; Wilberg, Karterud, Urnes, Pedersen, & Friis, 1998; Wilberg et al., 1999). However, with two exceptions, there is a lack of RCTs of the effects of day hospital-based treatment. The first RCT uncovered significant treatment effects after 18 weeks of psychodynamically group-oriented day-hospital treatment for patients with both affective disorder and longstanding PDs (Piper et al., 1993). Compared with wait-list controls, symptomatic distress, interpersonal functioning, self-esteem, life satisfaction, and defensive functioning improved after 4 months of treatment, and gains were maintained at the 8-month follow-up (Piper et al., 1993). The second RCT investigated the potential effects of a mentalization-based day-hospital treatment for patients with borderline PD (Bateman & Fonagy, 1999; 2001; 2008). The MBT program consisted of 18-month individual and group psychotherapy in a day-hospital setting offered within a structured and integrated program provided by a supervised team. The day-hospital treatment also included various forms of group therapy, such as art and writing groups. The initial day-hospital treatment was followed by twice weekly outpatient group psychotherapy for a further 18 months. The MBT day-hospital treatment condition was compared to treatment as usual, which consisted of community support from mental-health nurses, medication prescribed by a consultant psychiatrist, and periods of partial hospital and inpatient treatment as necessary; the patients received no specialist psychotherapy (Fonagy & Bateman, 2008). The 8-year follow-up revealed superior effects of the day hospital-based MBT program with regard to symptomatic distress and interpersonal functioning versus treatment as usual (Fonagy & Bateman, 2008). Nonetheless, these findings should be interpreted with caution due to the small sample size and low statistical power. Superior outcome was also reported in a similar MBT step-down program conducted in the Netherlands (Bales et al., 2014) that was not an RCT but had a matched control group that received specialized
psychotherapeutic treatment. Other step-down programs using inpatient treatment as the initial phase lack RCTs.

A naturalistic study by Chiesa et al. (2004) compared the effectiveness of a step-down program consisting of a 6-month residential stay followed by 12-18 months of outpatient treatment to: 1) a 12-month inpatient specialist program with no planned outpatient follow-up, and 2) a general community psychiatric model that included supportive outpatient and community contact with non-specialist caregivers. The residential treatment combined individual and group psychoanalytically oriented psychotherapy and therapeutic community meetings with activities for increasing the acquisition of interpersonal skills, resocialization, and rehabilitation. The outpatient stage of the step-down program consisted of group analytic psychotherapy, meetings with a psychosocially trained outreach nurse, and meetings with a senior psychiatrist to review progress. A sample of 143 patients with different PDs (borderline PD and paranoid PD were the most common PDs) was allocated to the three treatment conditions. At the 2-year follow-up, the authors concluded that the step-down treatment was more effective than both the residential treatment and the general psychiatric treatment in the community (Chiesa et al., 2004). Patients in the step-down program exhibited significantly higher rates of clinically relevant change in the dimensions of symptom severity, social adjustment, and global functioning relative to the other two groups. Four years after the expected end of treatment (6-year follow-up), >60% of the patients treated with the step-down approach displayed clinically relevant change, while more modest rates of improvement were evident in the inpatient condition (26%) and the community condition (13%) (Chiesa, Fonagy, & Holmes, 2006).

Although the studies described above compared treatment at different levels of care, there are other major differences between treatments in terms of theoretical background, duration of treatment, etc. These differences make it difficult to draw conclusions about the effects of different levels of care.

Few studies have directly compared treatment settings for patients with PDs. However, Bartak and colleagues recently performed three (one for each PD cluster) multi-center studies to directly compare treatments at different levels of care (Bartak, Andrea, Spreeuwenberg, Thunnissen, et al., 2011; Bartak, Andrea, Spreeuwenberg, Ziegler, et al., 2011; Bartak et al., 2010). This 3-year study was conducted in six mental health-care centers in the Netherlands and involved several hundred patients with PD diagnoses. At each participating center, patients were assigned to modalities of psychotherapeutic treatment that represented three levels of care: outpatient treatment, day-hospital treatment, and inpatient treatment. The first report to come from this study concerned patients with cluster C PDs (Bartak et al., 2010). The authors explored the effectiveness of five modalities of psychotherapeutic treatment embedded within three levels of care: long-term (16
months) outpatient treatment, short-term (up to 6 months) day treatment, long-term day treatment, short-term inpatient treatment, and long-term inpatient treatment. While data were collected from patients who received short-term outpatient treatment, these data were excluded from analysis because only 3.9% of patients received such care. After correction for all relevant baseline differences among the treatment groups, Bartak et al. determined that while improvement between baseline and 12 months after baseline was significant for patients in all groups on all outcome measures, patients in short-term inpatient treatment evidenced the most improvement; the within-group effect sizes for this treatment modality were consistently the largest across outcome domains (Bartak et al., 2010). The authors did not report findings pertaining to specific PDs in cluster C. These findings suggest that a level of care characterized by high degrees of containment, intensity, and structure for a relatively short duration may have the greatest potential for improvement for patients with cluster C PDs. The second report addressed patients with cluster B PDs (Bartak, Andrea, Spreeuwenberg, Ziegler, et al., 2011). The authors modified their presentation of results for this article by reporting on the effectiveness of the three levels of care without considering short-term and long-term variants of each, as in their first report. After correction of multiple propensity scores to control for baseline differences among groups, patients in inpatient treatment displayed the largest improvements, particularly for symptom distress (Bartak, Andrea, Spreeuwenberg, Ziegler, et al., 2011). It should be noted that this significant advantage of inpatient treatment only appeared in relation to outpatient treatment for the outcome measure of symptomatic distress; all other comparisons between levels of care yielded smaller differences. In summary, this second report suggested that all three levels of care were effective for patients with cluster B (primarily borderline) PDs (Bartak, Andrea, Spreeuwenberg, Ziegler, et al., 2011), although there seemed to be a slight advantage after inpatient treatment even after accounting for the strong influence of baseline patient characteristics. In the final paper from this study, Bartak et al. (Bartak, Andrea, Spreeuwenberg, Thunnissen, et al., 2011) presented findings for patients with cluster A PDs. As with their second report, the authors presented 18-month post-baseline data related to the three levels of care (outpatient, day treatment, inpatient) without considering variations in the duration of treatment within each level. Patients assigned to day treatment and inpatient treatment improved significantly on all outcome measures, but patients assigned to outpatient treatment improved significantly on only 1/4 measures (social role functioning) (Bartak, Andrea, Spreeuwenberg, Thunnissen, et al., 2011). After correction for baseline differences, improvements in psychiatric symptoms for patients in day treatment and inpatient treatment were significantly greater than for patients in the outpatient condition. Day-treatment patients also experienced significantly greater improvements in quality of life than did patients in the outpatient condition. No other significant
differences among levels of care were evident. The findings derived from this series of reports should be considered in the context of the absence of randomization of patients to the various levels of care.

A 2001 Cochrane review of day-hospital treatment versus outpatient treatment for patients with non-psychotic disorders concluded that there was limited evidence to justify the provision of such treatment for PD patients (Marshall, Crowther, Almaraz-Serrano, & Tyrer, 2001) and that further studies with a randomized design were needed. The Ullevål Personality Project (UPP), which began in 2004, was designed in response to the lack of evidence for day-hospital step-down treatment programs. When a randomized design was chosen for the UPP, the question of an appropriate control condition emerged.

While there have been relatively few RCTs of day-hospital and step-down treatments, there has been an increasing number of randomized studies of different forms of outpatient psychotherapy. These RCTs have typically addressed the efficacy of specialized treatment programs for borderline PD (Bateman & Fonagy, 2009; Clarkin et al., 2007). Despite the differences in theoretical orientations, all treatments have consisted of coherent and structured treatment programs; several treatments have made use of conjoint (individual and group) psychotherapy (Bateman & Fonagy, 2009; Linehan et al., 2006). Although the majority of RCTs have investigated the effect of manualized treatments, one of the most common psychotherapeutic treatments for patients with PDs is non-manualized outpatient individual psychotherapy. This kind of treatment is likely to include therapeutic principles from different theoretical directions, rendering treatment eclectic. However, there is little empirical knowledge about the outcome of such treatment. Vinnars et al. (2005) conducted an RCT to compare supportive expressive psychotherapy (a manualized treatment provided in a controlled university setting) with community-delivered individual psychodynamic psychotherapy by clinical specialists. There were no differences in effects between treatments, with improvement in psychosocial functioning and decreases in personality problems and symptom distress in both groups (Vinnars et al., 2005). The findings by Vinnars et al., (2005) suggest that outpatient individual psychotherapy is a suitable treatment option for PD patients. Concerning UPP, outpatient individual psychotherapy was considered the most realistic control condition to the day-hospital step-down programs.

### 2.2.2 Long-term outcome of psychotherapy

Most established psychotherapeutic treatments for PD argue for their capability to facilitate sustained change in patients. However, relatively few studies have addressed the long-term effects of these treatments (Bateman & Fonagy, 2008; Chiesa et al., 2006; Davidson, Tyrer, Norrie, Palmer, &
Tyrer, 2010; McMain, Guimond, Streiner, Cardish, & Links, 2012). With a few exceptions (Table 2), the duration of follow-up after treatment has been short. Given the entrenched and chronic nature of PDs, long-term follow-up is central for establishing the significance of psychotherapeutic treatments (Levy, 2008). Establishing the existence of beneficial long-term effects could have major impacts on costs for society and quality of life for patients.

Prospective naturalistic studies and treatment trials have suggested that although PD symptoms and diagnoses tend to remit over time, subjects frequently continue to function at suboptimal levels of work and social functioning (see above). Certainly, psychosocial dysfunction is central to the concept of PD diagnosis, which in the DSM-5 is defined as “an enduring pattern of inner experience and behavior that leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning” (APA, 2013). However, the low temporal stability of the PD criteria weakens the association between PD and psychosocial functioning. This discrepancy between diagnostic instability and psychosocial dysfunctional continuity suggests that there are aspects of PDs that are not fully captured by current diagnostic descriptions. The construct of “personality functioning,” which is conceptualized as a dimensional phenomenon ranging from adaptive levels of personality functioning to various degrees of maladaptive functioning, could contribute to understanding the persistence of psychosocial impairment. Indeed, increasing psychosocial functioning seems especially important, as it may have several positive impacts on patient life (Innstrand, Langballe, & Falkum, 2012).
Table 2. Overview of clinical studies of PDs with a longitudinal design.

<table>
<thead>
<tr>
<th>Study</th>
<th>Allocation</th>
<th>Participants</th>
<th>Intervention</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fonagy &amp; Bateman 2008</td>
<td>Randomization</td>
<td>41 patients</td>
<td>MBT, 18 months of conjoint therapy in a day-hospital setting, followed by 18 months of outpatient group psychotherapy or TAU.</td>
<td>Patients in the MBT program achieved better results on most variables. There was a marked reduction in symptomatic distress. However, psychosocial functioning continued to indicate deficits in both groups.</td>
</tr>
<tr>
<td>8 year follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chiesa et al., 2006</td>
<td>Naturalistic</td>
<td>111 patients</td>
<td>12 months of inpatient conjoint psychoanalytic psychotherapy and sociotherapy, or 6 months of inpatient conjoint psychoanalytic psychotherapy and sociotherapy followed by 12-18 months of outpatient group psychotherapy (step-down). Or general psychiatric service (TAU).</td>
<td>Patients in the step-down treatment had the highest percentage of patients achieving a clinically relevant change. The greatest change were found in symptomatic distress. Psychosocial functioning improved most in the step-down, but was still in a range considered markedly impaired.</td>
</tr>
<tr>
<td>6 year follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Davidson et al., 2010</td>
<td>Randomization</td>
<td>106 patients</td>
<td>CBT, 12 months of individual therapy in an outpatient setting or 12 months TAU.</td>
<td>No significant differences between groups. Half of patients lost their BPD diagnosis, and experienced marked reduction in symptomatic distress. Psychosocial functioning improved but was still regarded poor after 6 years.</td>
</tr>
<tr>
<td>6 year follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMain et al., 2012</td>
<td>Randomization</td>
<td>180 patients</td>
<td>DBT, 12 months conjoint psychotherapy in outpatient setting or 12 months GPM individual psychotherapy in an outpatient setting.</td>
<td>No significant differences between groups. Significant decrease in mean symptomatic distress and general psychopathology. Post therapeutic improvement in interpersonal problems. Considerable impairment in psychosocial functioning at 3 years follow-up.</td>
</tr>
<tr>
<td>3 year follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Personality functioning is regarded as relatively stable over time and more stable than symptoms, although the empirical basis for this temporal stability is limited (Gunderson et al., 2011). In contrast to personality traits that are relatively resistant to change by treatment (Hopwood et al., 2011; Wilberg, Karterud, Pedersen, Urnes, & Costa, 2009), a patient’s ideas and feelings regarding self and interpersonal relations could constitute more-changeable aspects of personality that may be targeted through treatment. Interestingly, a few recent studies have reported significant developments toward more-adaptive levels of identity integration and relational capacities over 1-2 years in PD patients after intensive psychotherapy (Verheul et al., 2008; Bales et al., 2012, 2014; Feenstra et al., 2014). Thus, the concept of personality functioning may capture underlying core aspects of PD that are more stable than PD diagnoses but are associated with persistent psychosocial impairment, while at the same time representing phenomena that may be targeted through psychotherapy or other interventions. These hypotheses merit investigation.

Personality function can be viewed as one component of a broadly defined psychosocial function (Ro & Clark, 2013). This view suggests that reducing impairments in self and relational functioning inherently influences psychosocial function. It also raises a causal question: if more-adaptive levels of personality functioning can be achieved by therapy, how much gain in long-term psychosocial functioning can be expected? Such a causal influence may operate as a feedback mechanism; for example, access to work or positive changes in social environments could strengthen the individual’s experience of self-esteem and agency and improve interpersonal skills. An example of a successful intervention directed at psychosocial functioning is vocational rehabilitation for individuals with psychotic disorders (Falkum et al., Submitted). In a potential reciprocal association between increase in personality functioning and long-term improvement in psychosocial functioning, it is of interest to determine which direction is stronger. In paper III, we investigated the longitudinal reciprocal associations between personality functioning and psychosocial functioning. Causal inferences about these reciprocal associations can be used to guide the distribution of resources to therapy, vocational counseling/rehabilitation, or a combination of approaches.

2.3 Impact of pre-treatment variables on psychotherapeutic outcome

2.3.1 Predictors and moderators

Studies of PD treatment typically report large variations in clinical outcome, even when patients receive the same psychotherapeutic treatment. There is currently no consensus as to the factors that
underlie these differences in outcome. Such information would be valuable because identifying the patient characteristics that influence the outcome of therapy could enable earlier identification of patients who may be at risk of poor outcomes and may therefore require altered treatment strategies. Understanding how therapy characteristics or processes can influence outcomes could also enable the identification of helpful factors. Psychotherapies for PDs could then be modified in order to enhance the presence or effect of these helpful factors (Barnicot et al., 2012).

Attempts to synthesize research on predictors of psychotherapeutic treatment for PDs have included those of Lieb et al. (2004), who summarized the results of four relevant studies, and Robins and Chapman (2004), who summarized the results of five relevant studies. Predictive factors identified in these brief reviews included affective instability, self-harm history, length of previous hospitalization, abuse history, maternal psychopathology, patient age, schizotypal symptoms, hostility, and therapist adherence to the treatment model. Although a wide range of potential predictors was identified, these reviews uncovered few consistent findings across studies.

In a recent review of 33 studies, Barnicot and colleagues (2012) highlighted two consistent predictors of psychotherapeutic outcome for patients with borderline PD: 1) a strong alliance between the patient and the therapist, and 2) severity of initial symptoms (patients with higher severity often achieved greater symptom reduction). Findings on the predictive value of sociodemographic variables were mainly non-significant; perhaps most notably, age was consistently not associated with outcome.

Barnicot et al. (2012) stated that interpretation of their findings was complicated by heterogeneities in research methods and analysis quality, and that there was still no consensus on the influences underlying the outcome of therapy for these patients. Another explanation for the difficulties in identifying predictors of outcome may be that the interactive nature of psychotherapy makes it difficult to isolate pre-treatment patient characteristics from the interactions between patients and their therapists or treatment formats, which may account for some of the reluctance to measure potential predictors of course (Clarkin & Levy 2004). In their conclusion on client contributions to general outcomes of psychotherapy, Bohart and Wade (2013) concluded that there is little evidence that demographic variables significantly moderate psychotherapeutic outcome. Their conclusion was not based on patients with PD, but rather was relevant to psychotherapeutic research in general. Bohart and Wade (2013) further stated that looking for simple relationships between individual demographics and outcome no longer seems fruitful, and that it is time for research to develop more-sophisticated hypotheses about psychological variables that may also moderate and mediate
effects (Clarkin & Levy, 2004; Bohard & Wade, 2013). In paper IV, we investigated how a pretreatment psychological variable (mentalization) affected the long-term clinical course of PD patients.

Concerning how pretreatment variables affect therapeutic outcome, this thesis includes two terms to describe such influence. While a general predictor predicts outcome across treatments, a moderator differentially influences outcome in different treatment conditions. If, for instance, men improve more than women across two treatment conditions, then gender would be a general predictor of outcome. If women experience better effects from one type of therapy, then gender would be a moderator of outcome. Thus, a moderator specifies for whom various treatments work (Kraemer, Wilson, Fairburn, & Agras, 2002). Given the massive variability among patients with PD (both within and between PD categories), we need more information about which strategies work for which patients. In paper IV, we sought clinically relevant information about this issue.

2.3.2 The theory of mentalization

“Mentalization” refers to how humans make sense of their social world by making inferences about the mental states of themselves and others. More specifically, it is defined as the ability to understand and interpret, implicitly and explicitly, one’s own and others’ behavior as expressions of various intentional mental states (e.g. thoughts, feelings, desires) (Fonagy et al., 2002). Poor capacity for mentalization has been proposed as a central deficit in individuals with PDs, most notably in borderline PD, but has also been applied to the broader concept of borderline personality, which is defined as a form of pathological personality organization common to all serious PDs (Fonagy et al., 2002).

Mentalization has its roots in attachment theory (Fonagy et al., 2002) and postulates that one’s understanding of others depends on whether one’s own mental states were adequately understood by caring, attentive, and non-threatening adults. Problems in affect regulation, attentional control, and self-control stemming from dysfunctional or insecure attachment relationships (Agrawal, Gunderson, Holmes, & Lyons-Ruth, 2004; Lyons-Ruth, Yellin, Melnick, & Atwood, 2005; Sroufe, 2005) are mediated by a failure to develop a robust mentalizing capacity (Fonagy & Bateman, 2010). The capacity for mentalization is thought to arise as part of an integration of two primitive modes of thinking: psychic equivalence and pretend mode (Fonagy & Target, 1996; Target & Fonagy, 1996). In the psychic equivalent mode, the child equates the internal with the external, so that there is a one-
to-one relationship between his/her inner thoughts and feelings and the real world. In pretend mode, the child’s inner experience and mental states are decoupled from the external world, and as such have no implications outside his/her own mind. In normal development, the two modes are integrated, so that the internal and external worlds will be connected, yet separate, instead of being equated with or dissociated from each other. In the mentalizing mode, thoughts and feelings can be experienced as representations (Bateman & Fonagy, 2004).

Over the last two decades, the mentalization construct has been operationalized as reflective functioning (RF), with a scoring manual for application to the Adult Attachment Interviews (AAIs) (Fonagy, Target, Steele, & Steele, 1998) that enables assessment of a patient’s mentalizing ability based on narratives of his/her childhood attachment experiences. This method measures mentalizing related to attachment relationships and expresses RF as a single global score. A recent investigation demonstrated satisfactory construct validity and psychometric properties of the RF scale applied to the AAI (Taubner et al., 2013), and the method is currently considered the gold standard for measuring RF.

Given the theoretical association between mentalizing deficits and patients with schematic, rigid, and sometimes extreme views (Fonagy et al., 2014), it seems plausible that individuals with different pre-treatment levels of mentalizing capacity may differ in their ability to engage in psychotherapy (Katznelson, 2014). However, with the exception of a publication from the current study, there are no publications on the predictive value of mentalizing capacity on treatment outcome for patients with PD. Two reports from treatment studies included patients with eating disorders (Muller, Kaufhold, Overbeck, & Grabhorn, 2006) or depression (Taubner et al., 2011). Their results were inconsistent in terms of the predictive value of mentalizing capacity for outcome of 3 months of inpatient treatment and long-term psychoanalytic treatment, respectively. Thus, the general predictive value of mentalizing capacity remains uncertain, as does the effect of mentalizing capacity on a patient’s capacity to utilize different psychotherapeutic formats, such as group and individual psychotherapy. In paper IV, we focused on RF as a potential predictor or moderator of treatment outcome.

Mentalization capacity is assumed to buffer against psychopathology in individuals exposed to trauma and abuse and to be related to more-adaptive affect regulation and interpersonal functioning (Fonagy & Target, 2006; Fonagy & Luyton, 2009; Chiesa & Fonagy, 2014). However, few studies have investigated how RF is associated with clinical manifestations of PD psychopathology in terms of symptom distress and psychosocial or personality functioning. An association among
hypermentalizing (excessive but inaccurate mentalization; assessed by the Movie for the Assessment of Social Cognition; Dziobek et al., 2006), borderline PD traits, and self-reported difficulties in emotion regulation was detected in an inpatient adolescent sample (Sharp et al., 2011). Chiesa and Fongy (2014) found that RF was negatively correlated with self-reported symptom distress in their mixed PD and non-psychiatric sample. To date, no study has investigated the relationships between RF and everyday psychosocial function in clinical PD samples. However, in a population of unemployed, low-income, and disadvantaged adults in New York who participated in a job-readiness training program, participants with higher RF had PDs less frequently and were more likely to complete the program and attain a job than other participants (Bly, Wright, & Tuber, 2012).

Two studies that examined associations between RF and severity of psychopathology in terms of the number of Axis I and Axis II disorders reported inconsistent results. Bouchard et al. (2008) included subjects from clinical and non-clinical populations and reported that RF was negatively correlated with the number of Axis I and Axis II disorders. These correlations were not confirmed in the more recent treatment study of Fischer-Kern et al. (2010). On the other hand, Fisher-Kern et al. detected an association between RF and level of personality organization among their borderline PD patients, as assessed by the Structured Interview of Personality Organization (Clarkin et al., 2004). This interview is based on Kernberg’s psychodynamic conceptualization of differentiation and integration of internal representations of self and others, and covers several domains that are central to personality functioning, such as identity consolidation and quality of object relations (kernberg, 2005). Thus, we still have scant knowledge about how RF relates to mental distress, the extent of psychopathology, and the degree of maladaptive personality functioning. Clearly, more studies are needed to evaluate the clinical implication of mentalization deficits; paper IV addresses this topic.

3. The present study

3.1 The UPP and overall aims for this dissertation

The UPP is a RCT conducted at the Department of Personality Psychiatry at Oslo University Hospital, Ullevål, Norway. The UPP began in 2004 as a response to the lack of quality trials of day-hospital step-down treatment programs. It compares two treatment modalities for patients with various PD diagnoses. With the aim of achieving initial comparability, consecutively evaluated patients were randomly assigned to 1) a step-down program consisting of initial short-term day-hospital treatment
followed by a combination of long-term outpatient group therapy and individual psychotherapy, or 2) outpatient individual treatment according to the therapists’ preferred method.

Patients were evaluated for a wide range of outcome variables at baseline and at 8 months, 18 months, 3 years, and 6 years after the initial random assignment. Baseline evaluations were performed by the clinical staff at the Department of Personality Psychiatry, while PhD students and research assistants conducted the follow-up evaluations.

The overall aims of this doctoral thesis were to examine potential differences in the long-term outcome of patients receiving either day-hospital step-down treatment or OIP and to examine the long-term processes of change in patients that participated in the UPP.

3.2. Paper I

Despite the assumption that patients with severe PDs may benefit from long-term treatment, there have been few randomized studies of therapies lasting >1 year for patients with severe PDs. While several studies have shown that treatment gains may be maintained for 1-2 years after treatment, few studies have investigated the long-term course of symptoms after treatment. The study in paper I sought to compare long-term outcomes in the two UPP treatment conditions using a wide range of clinical measures. We addressed three research questions: 1) Are there differences in overall clinical course from baseline to 6 years of follow-up in terms of the primary outcome variables (psychosocial functioning, interpersonal problems, symptom distress, and quality of life)? 2) Are there differences in the patterns of clinical course between treatment conditions between the 3-year and 6-year follow-ups? 3) Are there differences in secondary outcome variables (diagnostic status, self-harm/suicidality, vocational functioning, and use of health care services) at the 6-year follow-up?

3.3 Paper II

Borderline PD is the PD category that has received the most attention in clinical research. In order to compare the results of the UPP with the existing PD literature, we examined treatment outcomes for the subsample of UPP patients with borderline PD. The overall research question was whether the intensive combined treatment format was particularly helpful for patients with emotional dysregulation, stormy relationships, and destructive acting-out behaviors, and if so, whether this format was more effective than individual therapy. Longitudinal analyses were conducted to
investigate potential differences between treatment conditions in the following domains of clinical outcomes at the 6-year follow-up: 1) symptom distress, psychosocial functioning, interpersonal problems and quality of life (main outcomes); 2) self-reported personality functioning; 3) self-injury, suicidal thoughts, and suicide attempts; 4) diagnostic outcome and use of psychotropic medication; and 5) effects of co-occurring avoidant PD (also analyzed independent of treatment conditions).

3.4 Paper III

Prospective naturalistic studies and treatment trials have shown that although PD symptoms and diagnoses tend to remit over time, subjects frequently continue to function at suboptimal levels of work and social functioning. There is a need to identify and characterize the factors associated with persistent psychosocial dysfunction. Personality functioning is an emerging construct that may capture underlying core aspects of PD that are more stable than PD diagnoses but could be associated with more enduring psychosocial impairment. Personality functioning may also represent phenomena that may be targeted through psychotherapy or other interventions. In paper III, we investigated directional and causal inferences between personality functioning and psychosocial functioning.

3.5 Paper IV

Mentalizing deficits have been proposed to represent a core mechanism in PDs, particularly borderline PD, but few studies have explored the clinical implications of impaired mentalizing capacity. Paper IV focused on the associations between mentalizing capacity, operationalized as RF, and clinical variables in patients with borderline and/or avoidant PD. Our first aim was to investigate the relationship between mentalization capacity and clinical manifestations before entering treatment. Our second aim was to investigate the relationship between baseline levels of RF and long-term outcomes. We addressed three research questions: 1) What were the associations between RF and levels of psychosocial functioning, symptom distress, interpersonal problems, self-esteem, and personality functioning at baseline? 2) Did baseline levels of RF predict clinical outcome at 6 years of follow-up? 3) Were baseline levels of RF a moderator of long-term treatment effects, and if so, what were the magnitudes of these clinical differences?
4. Methods

4.1 Treatments

See figure 1.

4.1.1 Step-down treatment program

The step-down treatment program was a long-term psychotherapeutic treatment that combined day-hospital treatment with conjoint individual and group psychotherapy. Patients allocated to step-down treatment were initially offered 18 weeks of day-hospital treatment consisting of a combination of psychodynamic and cognitive-behavioral group therapies 3–4 days per week. This strategy adhered to relational psychodynamics with reference to group analysis, self-psychology, and MBT. Staff members received biweekly video-based supervision by a senior therapist. Although written guidelines outlining the therapeutic stance were developed, there was no formal manual that served as a standard for measuring treatment adherence. The day-treatment program had a capacity of 18 patients; all groups had a maximum of nine patients and two therapists. Due to difficult research logistics for following RCT procedures, the actual number of patients was lower during some periods. This treatment program is typical of the day-hospital treatment tradition in Norway (Karterud et al., 2003).

Psychodynamic group therapy was considered to be the core of the treatment program and was conducted according to modified group analytic principles. It was provided 1.5 h twice per week, with an emphasis on maladaptive interpersonal transactions, affect dysregulation, attachment issues, self-object needs, and object failures. Art group therapy was also provided 1.5 h twice per week and was conducted according to guidelines described by Johns and Karterud (2004). Cognitive group therapy was provided for 1 h once per week; the group worked with early maladaptive schemas identified with the self-report instrument Young 75 (Young, 1999). Problem-solving group therapy focused on strategies for solving social, economic, and interpersonal problems and lasted for 1 h once per week. Cognitive-behavioral group therapy for patients who suffered from additional anxiety disorders was optional. It lasted for 1.5 h once per week and was conducted according to generally accepted cognitive behavior group therapy guidelines that included individual exposure home lessons (Clark, 1999; Hoffart et al., 1993). During median group therapy (“community meetings”), all patients and several staff members assembled for 1 h twice per week. These sessions addressed
general issues relevant to treatment alliance (treatment ideology, group norms, adherence to the program, cooperation, and the dynamics of the unit as a whole).

Most patients (70%) were on medications by referral. Medication regimes were evaluated and eventually modified by the staff psychiatrist, who also monitored pharmacotherapy during treatment. The American Psychiatric Association’s guidelines for borderline PD (APA, 2001) were followed with a somewhat restrictive attitude.

The staff dynamic group met 1 h biweekly with an external consultant. Working with patients with PDs is well known to provoke countertransference reactions and disagreement among staff (Rossberg, Karterud, Pedersen, & Friis, 2007). The primary task of this group was to restore the vitality of the staff and to cultivate a sound, collective, reflective culture. Staff meetings were held three times per week to orchestrate ongoing events, to exchange mutual information, and to engage in ad-hoc supervision.

After the conclusion of short-term day-hospital treatment, all patients were offered continuous outpatient treatment. The mean length of the transition between day-hospital treatment and subsequent therapy was 19 days (standard deviation (SD) 19 days). Follow-up treatment consisted of once-weekly OIP (with a predefined maximum length of 2.5 years) plus 1.5 h group psychotherapy once per week (with a predefined maximum length of 4 years). Conjoint therapy did not use the same therapists for individual and group sessions. Individual therapy adhered to psychodynamic principles, while group therapy adhered to modified group analytic principles. Group therapists and individual therapists met approximately twice per year to assess process and progress issues.

The average duration of treatments in the step-down treatment condition (day-hospital and outpatient parts) was 31 months (SD 16 months), and the average number of outpatient consultations (counting both group and individual therapy) was 106 (SD 76 consultations). The mean interval between the end of study therapy and the 6-year follow-up (post-treatment phase) was 43 months (SD 18 months).

4.1.2 Outpatient treatment

Patients allocated to the OIP treatment condition received eclectic individual therapy, which was defined as therapy according to the therapists’ usual practice. Therapy was mainly conducted by clinical specialists in private practice. The researchers gave no instructions to the therapists regarding
the duration and intensity of psychotherapy, nor did they interfere with any treatment decision in the OIP condition. Thus, the length of treatment was determined according to consecutive evaluations of treatment and agreement between the therapist and the patient. The frequency of therapy ranged from once per month to three times per week, with 83% of patients attending therapy once per week. This frequency is comparable to the average treatment frequency in OIP provided by private practitioners in Norway, which has been reported as one consultation per week with a mean duration of 1.5-2 years (Gråwe et al., 2005).

The average duration of treatment in the OIP condition was 24 months (SD 20 months), and the average number of consultations was 56 (SD 56.7 consultations). The mean interval between the end of study therapy and the 6-year follow-up (post-treatment phase) was 47 months (SD 20 months) in the OIP condition.

Figure 1. Patient dispensation in a randomized clinical trial comparing a step-down treatment program with outpatient treatment.
4.2 Therapists

4.2.1 Individual therapists

To recruit individual therapists, a letter was sent to all private practitioners in Oslo who had a contract with the State Health Insurance Fund, as well as to professionals working at mental-health centers. After reading a short description of the project, the specialists were asked whether they were willing to participate, either as a step-down therapist or in the OIP. Therapists that responded were assigned to the step-down or OIP conditions according to their own preferences and signed a formal treatment contract with the project. Expenses for both treatments were covered by the State Health Insurance Fund. At the beginning of the study, therapists completed a self-report questionnaire covering theoretical preference, education, work experience, and work satisfaction. When the patient(s) terminated treatment, a second questionnaire was administered to obtain reports of frequency of treatment, duration of treatment, and the reason for ending therapy.

Thirty-two external therapists were recruited as therapists in the OIP condition (16 psychologists, 15 psychiatrists, and one resident doctor) and provided psychotherapy consistent with their usual practice style. Each therapist was allocated 1-3 patients. The group of OIP therapists included 12 females and 20 males with a mean age of 55 years (SD 7.6 years). Their mean work experience as psychotherapists was 20 years (SD 8 years). Most therapists reported adherence to psychoanalytic/psychodynamic theories, although cognitive and systemic elements were present (Wilberg, Kvarstein, & Rovik, 2014). Therapists were generally satisfied with their work as a psychiatrist, as indicated by a mean score of 4 points (SD 0.6 point) on a 6-point satisfaction scale ranging from 0 (no satisfaction) to 5 (very satisfied). Therapists were invited to attend an annual one-day motivational seminar on general aspects of PDs.

The group of individual therapists involved in the outpatient phase of the step-down treatment condition consisted of 16 psychologists, 12 psychiatrists, two psychiatric nurses, and one social worker (24 external therapists and seven therapists from the Department of Personality Psychiatry). The mean age of the individual therapists was 50 years (SD 8 years), and 58% were female. Each therapist received 1-3 patients for treatment. Therapists were generally very satisfied with their work as psychotherapists in the UPP, indicated by a mean score of 4.2 points (SD 0.6 points) on a 6-point scale from zero (no satisfaction) to 5 (very satisfied). Their main work experience as psychotherapists was 16 years (SD 8 years). Although written guidelines outlined the therapist’s adherence to self-psychology and mentalization, formal adherence was not tested, and most therapists probably delivered therapy according to their usual practice. Therapists were invited to one-day seminars.
twice per year on conjoint therapy, and they were invited to receive monthly supervision in groups, focusing on conjoint treatment.

The individual therapists in the OIP condition were somewhat more experienced than the individual therapists in the step-down treatment (p<0.05). The percentage of men was somewhat higher among therapists in the OIP condition than in the step-down condition, but this difference was not significant.

### 4.2.2 Group therapists in the step-down condition

The group therapists in the day-hospital treatment and in subsequent conjoint therapy were regular staff from the Department of Personality Psychiatry, Ullevål University Hospital. The day-hospital treatment condition included 5.5 full-time positions divided among 10 group therapists. There were three psychiatric nurses, two psychiatrists, one residential doctor specializing in psychiatry, one specialist in clinical psychology, one art therapist, one social worker, and one physiotherapist. Seven of the 10 therapists had 5 years of training in group analysis. The mean age of the therapists was 48 years (SD 9 years) and 80% were females. These therapists also conducted the outpatient group therapies.

### 4.3 Diagnostic assessments

Diagnostic assessments were conducted at baseline and at 3 years and 6 years of follow-up.

#### 4.3.1 PDs

PD diagnoses were decided according to DSM-IV criteria using the SCID-II interview (First, 1994). At baseline, 24 videotaped SCID-II interviews were rated by an independent rater. The kappa values for three PDs represented by at least five cases were 0.75 for avoidant PD, 0.66 for borderline PD, and 0.71 for paranoid PD, indicating acceptable diagnostic reliability. The intraclass correlation coefficient (ICC 2.1) for the number of fulfilled SCID-II criteria was 0.83.

At the 3-year follow-up, 27 videotaped interviews were rated by an independent rater. The only category with at least five cases was “PD not otherwise specified.” The kappa value for PD not
otherwise specified was 0.90, and it was 0.91 for any PD. The reliability coefficient (ICC 2.1) of the total number of SCID-II criteria was 0.90. At the 6-year follow-up, 30 videotapes were evaluated by an independent rater. Only avoidant PD was represented with at least five cases (10/30 patients), with full agreement of diagnosis (kappa of 1). The kappa value for the presence of any PD was 0.87. The reliability coefficient (ICC 2.1) was 0.92 for the total number of SCID-II criteria. These reliability coefficients indicate acceptable diagnostic reliability of PD diagnoses in the UPP.

4.3.2 Symptom disorders

Symptom-disorder diagnoses were based on the Mini International Neuropsychiatric Interview (Sheehan et al., 1998) and confirmed in accordance with the DSM-IV (APA, 1994). Assessment of inter-rater reliability of Axis I diagnoses was based on videotaped interviews with the same patients that were included in the reliability analyses of PDs, except that at baseline, one more patient interview was included (n=25). Kappa values were only calculated for diagnoses represented by at least five cases. At baseline, the kappa values were 0.51 for major depression, 0.60 for dysthymic disorder, 0.92 for social phobia, 0.52 for obsessive-compulsive disorder, 0.51 for panic with agoraphobia, 0.41 for general anxiety disorder, and 0.52 for alcohol abuse. When the diagnoses were pooled, the kappa values were 0.58 for patients with any anxiety disorder and 0.71 for patients with any substance-use disorder. There was agreement for 23/25 patients with any type of mood disorder, but the kappa value could not be computed due to empty cells in the cross-tabulation.

At the 3-year follow-up, the kappa values were 0.43 for major depression, 1 for social phobia, 1 for obsessive-compulsive disorder, 0.84 for generalized anxiety disorder, 0.90 for alcohol use, and 0.89 for drug abuse. When the diagnoses were pooled, the kappa values were 0.49, 0.72, and 0.92 for any mood disorder, any anxiety disorder, and any substance-use disorder, respectively. The reliability coefficient (ICC 2.1) for the total number of symptom disorders was 0.94.

At the 6-year follow-up, the kappa values were 0.73 for major depression, 0.92 for dysthymic disorder, 0.62 for panic disorder, 0.93 for social phobia, 0.70 for generalized anxiety disorder, and 1 for alcohol abuse. When the diagnoses were pooled, the kappa values were 0.66 for any mood disorder, 0.63 for any anxiety disorder, and 0.92 for any substance-use disorder. The reliability coefficient (ICC 2.1) of the number of symptom disorders was 0.76. The diagnostic reliability was low for some diagnoses, particularly at baseline. However, symptom disorders are not the primary focus of any of the papers in this dissertation.
4.4 Clinical measures

Clinical measures were assessed at baseline, 8 months, 18 months, 3 years, and 6 years of follow-up.

4.4.1 Symptom distress

The symptom checklist SCL-90-R (Derogatis, 1983) was used to measure symptom distress. SCL-90-R, a self-reporting questionnaire that requires responses on a 0-4 Likert scale, was designed to cover the major symptoms of psychiatric distress, represented by nine dimensions that can be meaningfully expressed by a Global Severity Index (GSI). A higher score indicates more symptomatic distress. The Beck Depression Inventory (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; Groth-Marnat, 1990) was used to measure characteristic attitudes and symptoms of depression. The Beck Depression Inventory is a 21-item inventory with a 0-3 rating scale. A sum score (0-63 points) is obtained by adding the scores for each question. Sum scores of 19-29 points are regarded as indicating moderate to severe depression, while scores ≥30 points indicate severe depression.

For the assessment of self-esteem, we applied the Index of Self-esteem (Hudson, 1982). The Index of Self-Esteem is a 25-item questionnaire measuring the degree of severity of a subject’s self-esteem problems. The scale produces scores ranging from 0-100 points, with scores >30 points indicating clinically significant problems in this area.

A 10-point scale assessed subjective quality of life. A score of 1 point represented the lowest perceived quality of life and a score of 10 points indicated the highest perceived quality of life.

4.4.2 Psychosocial functioning

Psychosocial functioning was rated using the Global Assessment of Functioning Scale (GAF) and the Work and Social Adjustment Scale. The GAF is rated on a scale from 1 to 100 points, with a higher score indicating a higher level of functioning. GAF scores were based on a short interview especially designed for the UPP and rated according to a split symptom and function version. Only the lowest GAF scores of symptoms and functions were used in analyses. The staff at the Department of Personality Psychiatry rated the GAF scores at baseline; research fellows and PhD students conducted and rated the GAF interviews at the 8-month, 18-month, 3-year, and 6-year evaluations.
The inter-rater reliability of the GAF scores at baseline was based on 18 videotaped GAF interviews scored by an external independent rater. The reliability of later GAF scores was based on 29, 30, 29, and 30 audiotaped interviews from the 8-month, 18-month, 3-year, and 6-year follow-up evaluations, respectively. Two independent raters agreed on consensus scores, which were entered in the reliability analyses. All raters were blind to treatment conditions. The reliability of the GAF scores (ICC 2.1) was 0.56 at baseline, 0.81 at 8 months, 0.85 at 18 months, 0.94 at 3 years, and 0.93 at 6 years. Unfortunately, the reliability coefficient at the 6-year follow up was erroneously reported as 0.92 in papers I, II, and IV.

The Work and Social Adjustment Scale is a self-reported five-item scale of functional impairment. It measures level of impairment on a scale from 0 to 8 points, where 0 indicates no impairment at all and 8 indicates very severe impairment. Scores for the five items are summed (range 0-40 points).

Subjective quality of life was assessed using a 10-point scale, with a score of 1 representing the lowest perceived quality of life and a score of 10 indicating the highest perceived quality of life.

4.4.3 Personality functioning

(Mal)adaptive personality functioning was measured using the 60-item short form of the SIPP-118 questionnaire. Sixty items are directly assigned to five higher-order domains: Identity Integration, Relational Capacities, Self-control, Social Concordance, and Responsibility. Scores range from 1 to 4 points, with lower scores reflecting more-maladaptive levels of personality functioning (Verheul et al., 2008). The domains are interpreted as follows: Identity Integration, coherence of identity, the ability to see oneself and one’s own life as stable, integrated, and purposeful; Relational Capacities, the ability to genuinely care about others and to feel cared for by them, to be able to communicate personal experiences, and to hear and engage with the experiences of others, often but not necessarily in the context of a long-term, intimate relationship; Self-control, the capacity to tolerate, use, and control one’s emotions and impulses; Social Concordance, the ability to value the identity of others, to withhold aggressive impulses toward others, and to work together with others; and Responsibility, the capacity to set realistic goals and to achieve these goals in line with the expectations that one has generated in others (Andrea et al., 2007). The full version of the SIPP-118 has been found to cover the self and relational domains of the alternative model in the DSM-5 (Bastiaansen, De Fruyt, Rossi, Schotte, & Hofmans, 2013). The SIPP-118 was translated from English to Norwegian by the research group at the Department of Personality Psychiatry and then translated
back into English by an independent bilingual translator. Although the psychometrics of the five
domains are not completely unambiguous, two studies have reported good psychometric properties
(Feenstra et al., 2011; Verheul et al., 2008).

4.4.4 Interpersonal problems

Self-reported interpersonal problems were assessed with the Circumplex of Interpersonal Problems
(Pedersen, 2002), a 0-4 Likert scale for a 47-item version of the Inventory of Interpersonal Problems
(Horowitz, Rosenberg, Baer, Ureno, & Villasenor, 1988). A total mean score was generated to
summarize all scores in the subscales. A higher score indicated more severe interpersonal problems.
Previously, the total Circumplex of Interpersonal Problems score had a correlation coefficient of 0.99
with the global score obtained from Alden’s Inventory of Interpersonal Problems (Alden, Wiggins, &

4.4.5 Self-injury, suicidal thoughts, and suicide attempts

Incidents of self-harm and suicide attempts were self-reported by patients and confirmed during
research interviews. At baseline and at the 6-year follow-up, patients were asked to report self-harm
and suicide attempts during the last 12 months. At the 8-month, 18-month, and 3-year follow-ups,
patients were asked to report incidents since the last follow-up, as well as suicidal thoughts during
the last 7 days. The criteria for self-injury were an intended episode of cutting, burning, etc. without
suicidal intent, whereas a suicide attempt was defined as an episode with the intention to commit
suicide, without success. Suicidal thoughts during the previous 7 days were reported as being present
or absent, and their severity was measured on a 5-point scale with a range from “passing thoughts”
to “active preparations for suicide.”

4.4.6 The AAI and the RF scale

The AAI (ref 48) was scored with the RF scale (ref 5) to measure each participant’s capacity for
mentalization. In the AAI, participants were asked to describe their childhood attachment
relationships and everyday routines (e.g., going-to-bed routines). They were also asked more-specific
questions about traumas and experiences of rejection, separation, and loss. Subjects were invited to
reflect upon their caregivers’ mental states and intentions (e.g., why they thought their parents
behaved as they did during their childhood) and how they thought their childhood experiences with their parents affected their adult personality. The RF scale evaluates a person’s capacity to think and reflect upon their own and others’ mental states and the level of complexity of this reflection. The scale ranges from -1 (negative RF: a systematic resistance to the reflective stance) to 9 (exceptional RF: fresh, complex, and unusually sophisticated reflections with original elaborations). The average RF value in the current sample was 3 points (SD 1.5 points) in the step-down treatment and 3.5 points (SD 1.7 points) in the OIP condition. The average intraclass correlation coefficient (ICC 2.1) for the overall RF scores was 0.73 after exclusion of one outlier, and 0.61 when this outlier was included (F. S. Gullestad, Johansen, Høglend, Karterud, & Wilberg, 2013). The outlier was a patient who initially received an RF score of 5 points and then a score of 0 points in the reliability test. The two raters disagreed about the level of authenticity of the patient’s capacity for mentalization. Although one of the two raters experienced the patient as genuinely mentalizing, the other rated the patient’s mentalizing efforts as self-distorted and generally poorly integrated. Because of the intention-to-treat approach in this study, all patients were included in the statistical analyses.

In the present study, RF scores were dichotomized; RF scores <3 points were categorized as low (n=26) and scores ≥3 points were categorized as medium (n=52). The term “medium RF” is used instead of “high RF” because high values in this patient group are believed to be relatively low compared to non-clinical samples. This dichotomization was based on an a-priori decision. The rationale is both empirical (Fonagy et al., 1996) and based on the clinical impression of a clear difference between patients with almost no capacity for mentalization (RF <3 points) and patients with mentalizing tendencies (RF ≥3 points).

The ratings in this study were based on a thorough reading of verbatim transcripts made from audiotapes of AAIs by three coders (all authors are co-authors of paper IV; MJ, FR, and TW). The coders had attended the course “Reflective Functioning on the Adult Attachment Interview” at the Anna Freud Centre in London, and they successfully completed the Reliability Test related to this course. All three coders were authorized to utilize the RF scale for clinical and research purposes and to publish research data obtained with the scale.

4.5 Statistical analyses

4.5.1 Mixed-model analysis
The data in papers I, II, and IV were analyzed with an intention to treat based on treatment assignment. Multilevel growth-curve analyses consisted of a linear mixed model for continuous outcomes and a generalized linear mixed model for repeated dichotomous outcomes. In multilevel models for longitudinal data, the lowest level of data is the specific measurement at a particular time. This lowest level is referred to as “Level-1” data. Each Level-1 measurement is nested within a particular research participant; thus, the individual constitutes Level-2 data. In our analyses, individuals are nested within treatment conditions, which constitute Level-3 data. This “nesting” of observations (the hierarchical structure of data) implies a strong assumption of independent observations, which underlies many standard statistical procedures such as general linear models (Student’s t tests, analyses of variance/covariance, and regression analyses). In case of dependence of residuals due to covariance between data levels, traditional statistics often underestimates the standard error and increases the risk of type-I errors. In contrast, multilevel modeling takes into account the lack of independence (the nesting of data). Moreover, inter-individual variability in assessment intervals can be tolerated by specifying individual trajectories. Thus, multilevel modeling has the advantage of allowing for irregularly spaced waves of data. Another benefit is the inclusion of cases with missing values/observations (not only patients with complete datasets), yielding a less-biased estimate of the effect of treatment within the intention-to-treat sample. When using more than two observations (test-retest), this approach permits more flexible and nuanced investigations of nonlinear individual change.

4.5.2 Marginal structured modeling

The data in paper III were analyzed using a marginal structural models (MSMs) approach that was developed in the statistical and epidemiologic literature for the purpose of causal inference. The paper that popularized MSMs in epidemiology has been cited >1500 times (Google Scholar, April 2015, but with a few notable recent exceptions (VanderWeele et al., 2012; VanderWeele et al., 2011), the method has not been used extensively in psychology. Such statistical models may be applied for exploring complex relationships between various aspects of mental health over time. MSMs are of special importance when the effect of a time-varying variable or exposure on an outcome (single or repeated measures) is of interest. When confounders for the exposure-outcome association also vary with time and are affected by prior exposure, an ordinary regression (univariate or longitudinal model) generally results in a biased estimate of the time-varying exposure effect; the MSM is unbiased. In paper III we use the repeated measures MSM (Hernan, Brumback, & Robins,
2002) to study the reciprocal associations between personality functioning and psychosocial functioning, and to determine the extent to which the effect of one on the other persists over time.

5. Results

5.1 Summary of paper I

Step-down versus OIP for PDs: 6-year follow-up of the UPP

Although psychotherapy is considered the treatment of choice for patients with PDs, there is no consensus about the optimal level of care for this group of patients. The aim of this study was to compare an intensive long-term step-down treatment program consisting of short-term day-hospital treatment followed by combined group and individual psychotherapy in an outpatient setting to OIP. The analysis included 113 patients with various PDs who were randomly assigned to one of the two treatment conditions. Outcome was evaluated after 8 months, 18 months, 3 years, and 6 years and was based on a wide range of clinical measures, such as psychosocial functioning, interpersonal problems, symptom severity, and Axis I and II diagnoses.

At the 6-year follow-up, there were no significant differences in outcome between treatment groups. Effect sizes ranged from medium to large for all outcome variables in both treatment arms. However, patients in the OIP group exhibited a marked decline in psychosocial functioning between the 3- and 6-year follow-ups; psychosocial functioning continued to improve in the step-down group during the same period. This difference between conditions was statistically significant. These findings suggest that both hospital-based long-term step-down treatment and long-term OIP may improve symptoms and psychosocial functioning in poorly functioning patients with various PDs. Continued improvement in psychosocial and interpersonal functioning in the step-down group during the post-treatment phase indicates that longer-term changes were stimulated during treatment.

5.2 Summary of paper II

Favorable outcome of long-term combined psychotherapy for patients with borderline PD: 6-year follow-up of a randomized study

Paper I of this dissertation found that both OIP and short-term day-hospital treatment followed by outpatient combined group and individual psychotherapy (the step-down treatment is referred to the “combined program” (CP) in paper II) were efficacious in a mixed PD sample. However, some PDs may receive more benefit from one of the two treatment modalities. Accordingly, paper II reports 6-
year follow-up data for the 52 patients with borderline PD who participated in the UPP. Outcomes were evaluated after 8 months, 18 months, 3 years, and 6 years based on a wide range of clinical measures, such as symptom severity, psychosocial functioning, personality functioning, and Axis I and II diagnoses.

At the 6-year follow-up, patients in the CP condition reported significantly less symptomatic distress as well as significant improvements in the personality functioning domains Identity Integration and Self-control compared with patients allocated to OIP. Patients in the CP also had a more favorable long-term course of psychosocial functioning. There were no differences between treatment conditions in terms of interpersonal functioning and self-esteem. Altogether, patients who received CP fared better on crucial parameters than patients who received OIP. Of particular importance are the positive effects on fundamental borderline problem areas like Identity Integration and Self-control. Thus, long-term psychotherapy in a CP seems favorable for borderline PD patients.

5.3 Summary of paper III
Adaptive personality and psychosocial functioning in patients with PDs: a causal feedback mechanism illustrated by a repeated-measures MSM

Prospective naturalistic studies and treatment trials have shown that although PD symptoms and diagnoses tend to remit over time, subjects frequently continue to function at suboptimal levels of work and social functioning. Longitudinal data gathered at five time points over 6 years from 113 patients with various PDs were analyzed to evaluate the reciprocal association between personality function and psychosocial function. Personality function was represented by three domains of the SIPP-118 short form: Relational Capacity, Identity Integration, and Self-control. Psychosocial function was measured with the GAF. A repeated-measure MSM was used for unbiased estimation of the causal effect (under certain assumptions) of the three domains of personality function on psychosocial function and vice versa. The results showed that a hypothetical joint intervention to increase REL by one standard deviation, both at one and two time-points prior to GAF assessment, would increase the GAF score by 15.4 points (3.5 standard deviations, 95% CI: 2.0, 4.96). Significant effects for the other domains and in the opposite direction were also found. This analysis indicates that persistent impairment in psychosocial function can be addressed through a causal pathway of personality functioning, with interventions of at least 18 months duration.
5.4 Summary of paper IV

Despite increasing interest in mentalization as a theoretically and clinically meaningful aspect of PD, few studies have investigated how mentalizing capacity is associated with clinical manifestations of personality or outcome of psychotherapy. In paper IV, we investigated the relationship between mentalization capacity operationalized as RF and clinical manifestations before entering study treatment. We next investigated the relationship between baseline levels of RF and long-term outcomes for patients in step-down and OIP treatment conditions. More precisely, we explored whether levels of RF predicted long-term clinical outcome. We also investigate whether patients with medium versus low pretreatment RF responded differently to the two treatment conditions. This study included 79 patients with borderline PD and/or avoidant PD. RF was measured before treatment; symptomatic distress, psychosocial functioning, personality functioning, and self-esteem were measured at baseline, 8 months, 18 months, 3 years, and 6 years.

RF was significantly associated with a wide range of variables at baseline. In longitudinal analyses, RF was not a predictor of long-term clinical outcome. Patients with low RF had better outcomes in the OIP than in the step-down program. In contrast, patients in the medium RF group achieved better results in the step-down program. These findings indicate that RF is associated with core aspects of personality pathology and captures clinically relevant phenomena in adult patients with PDs. Our results demonstrate that patients with different capacities for mentalization may require different therapeutic approaches.

6. Discussion

6.1 interpretations of main findings

6.1.1 Long-term outcome for the mixed PD sample

The results of the 6-year follow-up of the UPP suggest that both long-term step-down treatment and OIP may help improve symptoms and psychosocial functioning in poorly functioning PD patients. These results differ from the findings of two previous studies that compared intensive step-down treatment to less-intensive comparison treatments; both studies reported superior outcome from intensive step-down treatment programs (Chiesa et al., 2006; Fonagy & Bateman, 2008). There are several possible explanations for these differences in outcome. First, there was considerable variety among programs with regard to duration, intensity, and composition. Chiesa et al. (2006) investigated a model consisting of a 6-month inpatient stay followed by 12-18 months of outpatient...
combined group and individual treatment; Fonagy and Bateman (2008) included 18 months of MBT-based day-hospital treatment followed by group therapy twice weekly for a maximum of 18 months. In the present study, the initial phase was restricted to 18 weeks of day-hospital treatment, followed by long-term outpatient therapy consisting of weekly combined group and individual psychotherapy. Second, there were notable differences in the comparison groups. Chiesa et al. (2006) compared their step-down model to groups that received long-term residential treatment or general psychiatric community treatment. The comparison group in the study by Fonagy and Bateman (2008) received treatment as usual in the local community. Thus, neither study included OIP as a comparison group.

Given the accumulating evidence of the benefits of OIP for patients with PDs (Bamelis, Evers, Spinhoven, & Arntz, 2014; Clarkin et al., 2007; Giesen-Bloo et al., 2006; Linehan et al., 2006; McMain et al., 2009; Vinnars et al., 2005), it is difficult to directly compare these studies. Third, there were differences in the set of PD diagnoses in these studies. The UPP study had a mixed PD sample with borderline PD and avoidant PD as the most common PD diagnoses. Chiesa et al. (2006) also studied a mixed sample, but the largest diagnostic populations were borderline PD and paranoid PD. Additionally, the study by Chiesa et al. (2006) was not randomized. The study by Fonagy and Bateman (1999, 2001, 2008) only included patients with borderline PD; based on baseline levels of psychosocial functioning, self-harm, and comorbidity, those patients were more severely impaired than participants in the UPP.

Since few studies of PD treatment have reported long-term follow-up results, it is difficult to compare our results with, for example, general effect sizes from meta-analyses of PD treatment. Direct comparisons are also complicated by the use of different assessment instruments and the lack of effect sizes reported for many treatment studies of PD. However, generally speaking, long-term outcomes in both treatment conditions of the UPP are consistent with most reports of long-term outcome of psychotherapeutic treatments on PD. Most patients seem to reduce their symptomatic distress and to obtain remission of their PD diagnosis, although most patients continue to report persistent impairment in psychosocial functioning years after treatment. These conclusions were also valid in our study, in terms of the average GAF levels after 6 years. We detected a large variance in psychosocial functioning outcome: some patients experienced great improvement and some experienced very little improvement. Diagnostic remission and continuous impairments in psychosocial functioning have led to use of the term “recovery” in the long-term assessment of patients with PDs. Zanarini and colleagues (2010b) described recovery for patients with PDs as concurrent symptomatic remission and good social and full-time vocational functioning. In accordance with this definition, many patients achieve significant remission but not recovery. This is
in line with Perry and Bond (2009) who demonstrated that recovery (as defined by Frank et al., 1991) in psychotherapy occurs in sequence, and that global measures of functioning, especially social functioning, are among the slowest to improve.

Given the equal outcome between treatment conditions in the UPP at 6 years, our findings are inconsistent with the level-of-care series of studies by Bartak et al. (2010, 2011a, 2011b). Those studies suggested that increased levels of containment, intensity, and structure provided by inpatient treatment and day treatment facilitate improvement for patients with cluster B and C PDs. For cluster A PD patients, higher levels of care were strongly preferable because outpatient treatment (representing the lowest level of care) was largely ineffective (Bartak et al., 2011a). Conceivably, a more cautious general conclusion is that the higher levels of containment, structure, intensity, and therapeutic pressure may be useful for a broad spectrum of PD patients. However, the findings derived from this series of reports should be considered in the context of some important limitations. Most obvious is the absence of randomization of patients to the different levels of care. Yet, the use of the multiple propensity score correction is a significant advance for permitting non-randomized studies to be conducted with confidence. Another limitation is that the investigators did not conduct subgroup analyses (e.g., based on specific diagnostic subgroups within each PD cluster). This limits us to making only broad conclusions from the data.

At the 3-year follow-up, UPP patients in the OIP condition had significantly better psychosocial functioning than patients in the step-down program. Given this observation, the superior psychosocial outcome for the step-down condition between the 3- and 6-year follow-ups was unexpected. The finding at the 6-year follow-up was strengthened by our demonstration that observer-rated GAF values and self-rated Work and Social Adjustment Scale scores both exhibited the same statistically significant patterns between the 3-year and 6-year follow-ups; the shift in the proportion of patients with GAF levels >60 points further supports this pattern. A GAF score <60 points describes moderate to severe psychosocial dysfunction, and a GAF score >60 points indicates a range from mild dysfunction to a high level of functioning. At the 3-year evaluation, only one-third of patients in the step-down group and two-thirds of patients in the OIP group had a GAF score >60 points (Gullestad et al., 2012). In contrast, at the 6-year follow-up, two-thirds of the step-down group and one-third of the OIP group had a GAF level >60 points. This between-group difference suggests that the differences in psychosocial functioning were not only statistically significant, but also clinically significant.
Some plausible explanations concerning the divergence in clinical course between groups should be discussed. Generally speaking, our observation of continued improvement after the end of treatment for patients with PD suggests some degree of personality integration or convalescence. Along these lines, Howard et al. (1996) suggested a three-phase dose-response model of psychotherapy in which patients initially experience remoralization (the initial boost experienced from the feeling that help is there), followed by remediation (symptom reduction) and recovery (establishing adaptive ways of living, also conceived of as personality change/integration). Remoralization is usually accomplished quickly, whereas remediation is more gradual and typically occurs between 3 months and 8 months. Recovery is quite gradual and can take years. Although there is no direct evidence that patients in the step-down program achieved recovery, the continued improvement in most outcome variables after the conclusion of treatment may suggest the presence of gradual recovery processes. Applying Howard’s model, patients in the OIP treatment condition rather quickly experienced remoralization, but the relapse from 3 to 6 years suggests that they did not reach the recovery stage.

Several specific matters within the UPP may underlie the observed between-group differences in clinical course. First, the design of the step-down program likely disrupted patient attachments to therapists and group members after the short, intensive, day-treatment phase and required that the patients form new emotional bonds with group members and therapists in outpatient treatment. The attachment difficulties typical of patients with severe PD make those patients vulnerable during transitions between therapies. This vulnerability may have affected their therapeutic process and delayed clinical improvement (Hummelen, Wilberg, & Karterud, 2007). The clinical administration, which shared this view, decided to close the step-down program in 2008 based on the results of this study. The program was changed to an outpatient MBT program without a day-hospital component. Conversely, patients in the OIP condition may have benefitted from the possibility of a stable and continuing attachment to one therapist during the entire treatment period, thereby achieving faster improvement in psychosocial adaption. Nonetheless, the advantage observed at the 3-year follow-up was not sustained after 6 years. Second, at the 6-year follow-up, most patients had been in the post-treatment phase for a long time. Perhaps the less-intensive OIP was more supportive of psychosocial problems, which facilitated social adjustment over the short term for some patients but was more easily lost in the post-treatment phase. Third, we cannot exclude the possibility that treatment received after the cessation of UPP treatment could underlie the differences in clinical course between treatment arms. However, there were no significant differences in health-care services received in the past year between groups at the 6-year follow-up. Last, group therapy was an essential element in the step-down condition. Group therapy is a demanding format for patients with
PDs because it activates and challenges their interpersonal problems and emotional difficulties. Gains may be harder to achieve and slower to appear when patients struggle with such core issues over time in therapy groups. However, if sufficiently integrated, these experiences may serve as a foundation for further development after the end of treatment. Additional individual psychotherapy offered with group treatment in the outpatient portion of the step-down program may have facilitated such integration by clarifying and encouraging the process of working with identity and relational problems in the group. The combination of group and individual treatment was previously shown to be favorable for patients with borderline PD (Fonagy & Bateman, 2008; Linehan et al., 2006), and made us curious about the outcome of borderline PD patients in the UPP.

We also wish to mention that treatment costs from baseline to the 3-year follow-up did not significantly differ between treatment conditions when costs related to study treatment and additional psychiatric treatments were included (Kvarstein et al., 2013). The present study did not include cost analyses; however, there were small between-group differences in the use of healthcare services at the 6-year follow-up, which suggests minor differences in overall costs between treatment programs.

Other limitations of this study are discussed in the method section below.

6.1.2 Is more-intensive treatment particularly useful for patients with borderline PD?

In contrast to the results from the total mixed PD sample in the UPP (paper I), which displayed no overall difference between treatment conditions after 6 years, there were several differences between treatment conditions for the subgroup of patients with borderline PD. Borderline PD patients that received initial day-hospital treatment followed by conjoint group and individual therapy achieved significantly lower levels of symptomatic distress after 6 years than borderline PD patients that received OIP. Note that paper II refers to the step-down program as the Combination Program (CP). This change of names was implemented because most clinicians evaluating the UPP results argued that the main effects of the CP/step-down program could be ascribed to the combined psychotherapy in the outpatient portion of this program (the lengthiest part of the program). However, we stress that this assumption was based on a clinical impression and not on empirical evidence. Overall, it seemed wrong to send the message that day-hospital treatment was essential to the treatment of borderline PD patients when no one believed that this message was true.
The lower level of symptom distress in patients that received more-intensive combination treatment compared with patients that received less-intensive individual treatment is consistent with earlier studies of patients with borderline PD (Bateman & Fonagy 1999, 2001, 2008; Linahan 2006). In the present study, the average GSI at the 6-year follow-up for patients in the CP was 0.87, which is close to the estimated cut-off for non-clinical values of 0.80. In comparison, patients in the OIP had a mean GSI of 1.32 after 6 years, suggesting that the difference between treatment conditions was clinically significant. This finding is further supported by the significantly lower number of Axis I diagnoses in the CP compared to the OIP at the 6-year follow-up. The present results also revealed a trend toward less use of psychotropic medications in the CP. Taken together, these finding indicate that patients with borderline PD may obtain considerable symptomatic relief from intensive long-term therapy in a combination format.

Although symptom change is central to the treatment of borderline PD, and symptom change can improve patient quality of life (Wachtel, 1994), what kind of outcome can we expect in the treatment of borderline PD (Levy, 2008)? Linehan et al. (www.linehaninstitute.org) described “a life worth living” as one laudable goal of treatment. However, in their early naturalistic follow-up noted that although the “subjects in the DBT group acted better…they were still miserable,” experiencing “moderate symptoms” and/or “generally functioning with some difficulty” and “living lives of quiet desperation” (Linehan et al., 1994). Although psychotherapeutic treatment has evolved since then, the last decade’s data on treatment outcome for patients with borderline PD suggest that we need to work harder to move beyond symptom change and to help patients improve the quality of their day-to-day lives (Levy 2008). These findings indicate that successful treatments should not only be defined by symptomatic and diagnostic remission, but also by the strengthening of certain adaptive psychological capacities that will enable patients to deal with future life stressors in more-productive ways (Feenstra, Hutsebaut, Verheul, & van Limbeek, 2014).

In the present study, we included assessments of personality functioning among the outcome measures. Patients with borderline PD in the CP achieved significantly better results in the SIPP-118 domains of Identity Integration and Self-control than patients in the OIP condition. These two domains also were associated with the largest effect sizes among the SIPP-118 domains in the CP condition. This finding is in accordance with a recent prospective cohort study of an 18-month day-hospital MBT program for patients with severe borderline PD, which also found that Identity Integration and Self-control had the largest effect sizes among the SIPP-118 domains (Bales et al., 2012). These two domains may be particularly relevant to the differentiation of severity for borderline PD pathology. In a Dutch mixed PD sample, levels of Identity Integration and Self-control
were significantly lower among patients with ≥4 PD diagnoses than in patients with <4 PD diagnoses (Verheul et al., 2008).

The presence of co-occurring avoidant PD did not influence outcome in a treatment-dependent fashion. However, patients with avoidant PD had an overall poorer 6-year course of psychosocial functioning than patients without co-occurring avoidant PD, consistent with earlier findings (Chiesa & Fonagy, 2007; Zanarini et al., 2004). These results should be interpreted with caution due to the small sample size and low statistical power of this analysis. However, they suggest that co-occurring avoidant PD could represent a challenge in the treatment of patients with borderline PD.

The findings in papers I and II suggest that the use of intensive combination treatment may be especially beneficial for patients with borderline PD. This statement is supported by Bateman and Fonagy (2009) and Jørgensen et al. (2013), who also described good results for borderline PD patients that received intensive combination therapy in an outpatient setting. Outpatient-based therapy utilizing both individual and group treatment is also the preferred treatment setting and modality used in DBT for patients with borderline PD (Linehan et al., 2006).

As with the full mixed sample of patients in the UPP, borderline PD patients were associated with long-term trajectories in psychosocial functioning that depended on the treatment condition. Patients in the CP had a significantly better course in terms of GAF from 3-6 years compared to the OIP treatment condition. Nevertheless, final GAF scores in the sample averaged ~60 points (CP, 64 points; OIP, 58 points), indicating persistent difficulties in psychosocial function in both treatment conditions. Similar GAF levels were reported in 6- and 8-year follow-up studies of other combination models (Bateman & Fonagy, 2009; Chiesa et al., 2006). In long-term follow-up investigations like in the present study, patients stayed in the post-treatment phase for a relatively long time. As with the full mixed sample of the UPP, our analysis of borderline PD patients did not control for the possible effects of life events on clinical status during or after treatment. Although the UPP has a randomized design, it is possible that clinical courses may have been influenced by factors other than gains from therapy. However, differences between treatment conditions at the 6-year follow-up could also be explained by the attainment of higher levels of personality integration in the CP, as indicated by differences in the personality functioning domains Identity Integration and Self-control. The difference in Identity Integration between treatment conditions is particularly interesting because identity disturbance is, by definition, a central part of borderline PD psychopathology (DSM-5, 2013). The Level of Personality Functioning Scale in the alternative model in DSM-5 (section III) (DSM-5, 2013) has prompted a special interest in identity as one of the core markers of PDs. In this model,
self-functioning is defined by the two concepts of identity and self-direction. More-mature levels of identity, as well as increased self-control, can potentially contribute to improved psychosocial functioning and avoidance of negative interpersonal interactions, thus making patients more tolerant of psychosocial and symptomatic distress. Further analyses will be required to determine whether the beneficial outcomes in symptom distress and post-treatment psychosocial functioning for patients in the CP were mediated by changes in personality functioning. Given other reports that PD patients continue to function at suboptimal levels of work and social functioning long after achieving symptomatic and diagnostic remission, the potential causal relationship between personality functioning and psychosocial functioning seems especially interesting.

6.1.3 The relationship between personality functioning and psychosocial functioning

In paper III, MSM analyses uncovered a strong and reciprocal positive association between personality functioning and psychosocial functioning. Our results indicated that a successful intervention on one of the two constructs also positively affects the other. If we wish to focus on improving one of these aspects (or on reducing deterioration), then this feedback mechanism suggests a route to this goal.

Both PD symptoms and personality traits have been shown to prospectively predict psychosocial functioning (Hopwood & Zanarini, 2010; Morey et al., 2007), but these studies have not investigated any causal interactions or attempted to quantify the magnitude of the association. However, a comprehensive dynamic longitudinal model confirmed the prospective prediction of psychosocial function from DSM-5 personality traits in a structural equation model framework (Wright et al., 2015). These models, which have a long tradition in psychology, constitute an alternative to MSMs for causal inference; however, structural equation models are subject to more confounding biases that are partially circumvented by the MSM approach (VanderWeele et al., 2012). We believe that the present study, with its inclusion of personality functioning, represents an extension of the findings reported by Wright et al. (2015).

The causal model applied here, quantifies the different effects separately, estimate their persistence and account for time-varying confounding. Although there was evidence for effects in both directions between personality function and psychosocial function, the direction from personality function on psychosocial function was the strongest (in terms of baseline standard deviation). Within personality function, Relational Capacity was clearly the dominating domain, which is intuitively
reasonable in light of obvious overlap between items for this domain (e.g. “... it is hard for me to feel loved by people close to me”) and the GAF score which also emphasizes social well-being. Joint hypothetical interventions that could increase Relational Capacity by one standard deviation (0.6 units) at two succeeding time — points would be expected to be rewarded with an increase in GAF of 3.5 standard deviations (15.4 points), distributed as 2.3 and 1.2 standard deviations for one and two prior time- points, respectively. Given the mean GAF value of 47.6 at baseline an increase of 15.4 points would in average result in a GAF value >60, which is often regarded a clinical cutoff (Fonagy and Bateman 2008). A GAF score <60 is descriptive of moderate to severe psychosocial dysfunction, and a GAF score >60 indicates a range from mild dysfunction to a high level of functioning. Thus, this finding indicate a substantial clinical improvement in psychosocial functioning if one is able to increase the domain of Relational Capacity. In comparison, the opposite causal effect was estimated to a total gain of 0.46 standard deviations in Relational Capacity (0.3 units) for one standard deviation improvements in GAF at two succeeding time- points, distributed as 0.21 and 0.24 standard deviations for the two preceding time- points. Such an asymmetry in strength (between directions) is in accordance with the view that personality functioning is merely one component of a more global construct of psychosocial functioning, and as such receives only a proportion of a change. With regards to persistence of effects between Relational Capacity and GAF, evidence for long term influence in both directions was found, with significant coefficients for two prior time- points, which represents a period of at least 18 months. The symmetry in persistence is less intuitive than the asymmetry in strength, in light of the large difference in temporal stability (Relational Capacity had an average autocorrelation for different lags of 0.64 versus GAF with an average autocorrelation of 0.3), but probably attributable to strong effects and domain content.

The other two domains of personality function, Self-control and Identity Integration seemed both to represent weaker causal mechanisms than Relational Capacity. Self -control, one time -point prior to assessment of GAF, had a strong effect on GAF, with short persistence (no significant effect from two time- points prior), which was also characteristic of the reverse effect. The effect of Identity Integration on GAF, on the other hand, was of slightly smaller magnitude and of longer persistence (additional effect from two time- points prior to assessment of GAF), whereas the reverse effect was of short persistence. In this case, the difference in temporal stability between Identity Integration and GAF may have contributed to the asymmetry in persistence (Identity Integration had an average autocorrelation of 0.56). Why Self-control is the only domain characterized by a short term influence on GAF (as for the reverse effect) is most likely not due to temporal stability (average autocorrelation of 0.62, which is in the middle of the three domains), but rather to specific domain content.
Relational Capacity partly overlaps with the GAF score, and Identity Integration is closely associated with the patient’s ability to make use of the therapy (Bales et al., 2012), which both seems to be of more long term nature in relation to GAF. The link to the working mechanism of therapy might be reflected in the fact that of the three domains, Identity Integration had the highest effect size for change, also recently found elsewhere (Bales et al., 2012).

In paper II, we reported that for the subgroup of patients with borderline PD, the more intensive CP (step-down program) achieved superior results in the personality functioning domains Identity Integration and Self-control compared to results for patients in the OIP condition (Antonsen et al., 2015). These findings are consistent with reports from Bales et al. (2014) and Feenstra et al. (2014). Interestingly, patients in the CP also had significantly higher levels of psychosocial functioning from the 3-year follow-up to the 6-year follow-up compared to the outpatient treatment condition, indicating causality.

Results from Ro and Clark (2013) suggested that psychosocial functioning constructs should be subdivided into three components: Well Being, Interpersonal Functioning, and Basic Functioning. Their analyses further implied that both the SIPP-118 domains and the GAF relate to the Well Being factor, questioning the relevance of exploring causality between these constructs. As mentioned above we agree with Clark and Ro (2013) that there are some overlap between the concepts and that personality functioning can be seen as one component of a broadly defined concept of psychosocial functioning. Yet, from a clinical point of view, it seems clear that the SIPP-118 and the GAF capture different phenomena within the personality pathology spectrum. The items in the SIPP-118 questionnaire appear to be more closely related to processes central to psychotherapeutic treatment. Examples of Items in the SIPP-118 SF: “I am often confused about what kind of person I really am” (Identity Integration domain), “I often cannot help expressing my moods inappropriately” (Self-control domain), and “I believe that most people do not like to go along with me” (Relational Capacity). On the other hand, aspects central to the assessment of GAF are more closely related to classifying levels (mild, moderate, or severe) of impairment in work and social functioning and symptomatic distress. Thus, from a clinical point of view, it seems reasonable that these phenomena do not belong to the same construct and that analyses of potential causal inferences between the two are not only rational, but clinically important. Also, Ro and Clark (2013) explored these constructs from an empirical position; criticism has been levied against their application of analytical methods to complex phenomena such as personality and psychosocial functioning (Hopewood & Donnellan, 2010). Additionally, Ro and Clark’s assumptions are based partly on a non-clinical sample.
(Ro & Clark, 2009, 2013). Little is known about the relationship between these constructs in patients with PD.

Concerning the choice of statistical method, compared to an ordinary repeated-measures regression such as the linear mixed model, considerable discrepancy in estimated effects arises with the MSM approach. Apart from an expected difference due to conditional versus marginal effects measure, this discrepancy is due to time-varying confounding. Comparison of parameter estimates with and without weighting revealed considerable time-varying confounding for two of the personality dimensions, confirming the need to account for this bias. The direction of the confounding effect was not intuitive, given the positive associations among time-varying variables. During the development of the SIPP-118, the various factors underlying adaptive personality function were allowed to correlate (exploratory factor analysis with promax rotation) (Verheul et al., 2008). However, the associations between factors were often not simple and linear, but complex and non-monotone. With four time-varying variables associated with (and possibly affecting) each other in complex ways, simple intuitive rules for the direction of confounding bias do not apply (VanderWeele et al., 2008).

In this study, we did not control for the fact that the UPP contains several PDs and the possibility that the causal relationship between personality functioning and psychosocial functioning could vary according to the type of PD. Also, other psychological characteristics, such as mentalization capacity, may differentiate the degree of association between personality functioning and psychosocial functioning.

6.1.4 Clinical implications of mentalization capacity

Our finding in paper IV of an association between RF and the intensity of symptom distress and psychosocial impairment indicates that the mentalization construct captures clinically relevant phenomena. To our knowledge, this is the first clinical study to uncover a relationship between RF and everyday psychosocial functioning. However, the cross-sectional and correlational nature of our analysis limits conclusions regarding the direction of these associations (and therefore causality). However, Chiesa and Fonagy (2014) reported that RF was a significant mediator between childhood adversity and symptom distress, suggesting that mentalizing abilities may protect to some degree against the development of severe symptoms in patients with PDs.
Moreover, we detected significant associations between RF and self-esteem, interpersonal problems, and the SIPP-118 personality functioning domains Identity Integration and Relational Capacity, implying that RF is indeed related to personality pathology in terms of disturbances in self and self in relation to others. These findings are in line with Fonagy et al.’s idea of mentalization deficits as a core aspect of PDs (Fonagy et al., 2002; Fonagy & Target, 2006). A relationship between personality functioning and RF is also in accordance with the findings of Fischer-Kern et al. (2010), who detected a significant association between RF and personality organization in a sample of borderline PD patients. They performed their assessment using the Structured Interview of Personality Organization, which covers several domains of personality functioning including identity consolidation and quality of object relations. In addition, Müller et al. (2006) uncovered a significant correlation between RF and personality organization in a sample of patients with depression or eating disorders. In that study, personality organization was assessed by the structure axis (axis IV) of the operationalized psychodynamic diagnostics (OPD) (Cierpka, 2007; Force, 2001), based on an OPD interview. Like the Structured Interview of Personality Organization methodology, the structure axis of the OPD aims to cover a broad set of personality domains such as self and object recognition, regulation, communication, and attachment (Force, 2001). Recently, Doering et al. (2014) used axis IV as a measure of the level of personality functioning, and PD patients showed significantly worse results on axis IV outcomes than patients without PD.

Although the present study demonstrated a broad association between RF and clinical variables, we found no difference between groups with low and medium RF in the Self-control domain of the SIPP-118. This observation was somewhat surprising because the Self-control domain is meant to capture the characteristic borderline PD problems of emotional hyperarousal and impulsivity, the same borderline PD features that mentalizing capacity is thought to buffer against (Fonagy et al., 1996). The lack of an association between Self-control and RF may have occurred because the RF measured by AAI primarily captures mentalizing abilities related to early attachment relationships, and is neither a score of a patient’s general mentalization abilities nor a score that captures other specific deficits in mentalization (Choi-Kain & Gunderson, 2008; Semerari, 2003). This concern was also raised by Fonagy and Luyten (2009), who stated that mentalizing capacity is likely to be altered in states of emotional arousal and that the AAI may not capture fluctuating mentalizing deficits in current situations and relationships. Thus, future research should address the issue of more domain- or context-specific versions of RF. Symptom-specific RF versions already exist for panic disorder (Rudden, Milrod, Meehan, & Falkenstrom, 2009) and obsessive-compulsive disorder (Kullgard, 2013); these studies indicate that a patient’s symptom-specific RF is lower than a patient’s RF as assessed by
the AAI. On the other hand, moving to more symptom- or context-specific measures of RF may reduce the value of RF as a trans-diagnostic construct. It risks becoming a more circumscribed construct and could thus lose its ability to be compared across samples and studies. Whether it is possible to balance or combine such different perspectives in future assessments of mentalizing abilities merits further discussion and research.

The present sample included a greater proportion of patients with avoidant PD than borderline PD in the low RF group, suggesting that mentalizing problems are not limited to borderline PD but may be a feature of other types of personality pathology as well. It is unclear whether avoidant PD is associated with more mentalizing difficulties than borderline PD. Mentalizing deficits in patients with avoidant PD could reflect more limited access to mental states and poor emotional awareness in general (DiMaggio et al., 2012; Nicolo et al., 2011) rather than the context-dependent hyperarousal and impulsivity typical of borderline PD. Thus, related to the discussion above, there is a possibility that AAI-based RF is not equally able to capture particular mentalizing problems in different PDs.

The second aim of the study in paper IV was to investigate whether RF is associated with long-term clinical course. Consistent with findings at the 3-year follow-up (Gullestad et al., 2013), RF was not a significant predictor of 6-year outcome. Correspondingly, Taubner et al. (2011) reported that the level of RF did not predict outcome in terms of change in symptoms during long-term psychoanalytic treatment for patients with chronic depression. On the other hand, when Müller et al. (2006) investigated patients with depression or eating disorders who received 3 months of inpatient treatment, pretreatment RF level was significantly correlated with improvement in overall mental condition. These studies are difficult to compare with the current study and with each other due to differences in diagnoses, treatments, and duration of follow-up. Clearly, more studies are necessary in order to determine the long-term impact of pre-treatment mentalization capacity. However, given the correlations between RF and clinical severity, the lack of predictor findings in this study prompts optimism for future treatment of patients with more-severe PD.

PDs are heterogeneous in terms of severity and symptoms. To date, empirical knowledge about the interaction between treatment and patient characteristics is lacking, and the longstanding questions “what treatment, by whom, is most effective with this individual, with that specific problem, under which set of circumstances” (Paul, 1967; as cited by Levy 2008) remain mostly unanswered for the treatment of patients with borderline PD. Further, Levy (2008) makes an example of the history of DBT treatment; Linehan et al. (1991) suggested that DBT was a breakthrough strategy for the treatment of borderline PD, and DBT remains the most empirically robust treatment for borderline
PD. However as marked by Levy (2008); Bohus et al. (2000) reported that only 50% of patients in DBT had good outcomes. DBT is clearly more efficacious than treatment as usual, perhaps even more efficacious than community treatment by specialists (Levy 2008). Nonetheless, many individuals that receive DBT do not improve, and may be better served by different treatment types or settings. Such findings probably parallel results from other treatments and other PDs, but we have little knowledge of the patient characteristics that may moderate treatment outcome. Importantly, in paper IV, we identified significant and treatment-dependent moderator effects of RF. Patients with low RF had better outcomes in the outpatient treatment condition than in the step-down program. A novel finding in this 6-year follow-up was that patients in the medium RF group achieved better results in the step-down program than in the outpatient treatment. Between-group effect sizes were estimated to be large for 6/8 outcome variables at the 6-year follow-up, indicating strong clinical significance. The current results extend the findings of Gullestad et al. (2013) by detecting differences in symptomatic distress, interpersonal problems, and personality functioning. Thus, mentalization capacity seems to be a promising candidate as a characteristic that may inform the allocation of patients to different treatment settings and formats.

Although several factors may underlie the observed moderator effects, one of the most obvious structural differences between treatment conditions was the use of group psychotherapy in the step-down treatment. A possible explanation of the weaker result for patients with low RF in the step-down treatment may therefore be that group psychotherapy was too demanding for patients with a low capacity for understanding themselves and others in mental terms. This explanation may also clarify why patients with low RF achieved better results in the outpatient treatment than in the step-down program. On the other hand, the group-therapy format may not have been as influential as the type of group therapy offered in the present project. Group therapists in the step-down treatment were trained in group analytic therapy, a therapeutic style and approach that could be suitable for patients with some capacity for making sense of their own and others’ mental states but that may be too unstructured for patients with low or absent mentalizing capacity. Kvarstein et al. (2014) compared clinical outcomes for borderline PD patients in a psychodynamically oriented treatment program resembling the UPP’s step-down treatment with outcomes from an MBT in which group psychotherapy was specifically structured. Although the results were clearly in favor of the latter, Kvarstein’s study was not randomized and did not include measures of mentalization capacity. An interesting issue for further research is to determine the extent to which more-structured MBT group therapy benefits patients with low RF.
In contrast to patients with low RF, patients in the medium RF group experienced greater improvement in several clinical variables when treated in the step-down program versus outpatient treatment. Perhaps patients with an RF level ≥3 have a greater capacity to engage in a therapeutic project involving perception of and reflection on mental states and interpersonal issues. The complex group setting may not be overwhelming to these patients, and important therapeutic factors such as therapeutic alliance and group cohesion may evolve. If the patient feels reasonably secure, then the group may provide an arena for social exposure and interpersonal learning. The group format may thus challenge and stimulate mentalization more than individual therapy alone. In our study, patients with medium levels of RF may have been more capable of learning from their own emotional experiences in interactions with group members and may have been able to explore how their behavior impacts the feelings of others. When combined with individual therapy, as was the case in the step-down condition, treatment can provide guidance on understanding group interactions and prompt further reflection on how each patient understands themselves and the other group members.

It is important to keep in mind that the UPP is a study of treatment formats rather than of specific psychotherapeutic orientations and techniques. The treatments were not manualized, and there was no registration of treatment processes. Thus, interpretations of the moderator effects are highly speculative. However, our results suggest that the long-term outcome of intensive combined treatment may be favorable for patients with a minimum of mentalizing capacity. Patients with low RF may benefit from the less-intensive and more flexible OIP format, which gives the therapist a better opportunity to monitor and adjust to a patient’s ongoing or fluctuating levels of mentalizing.

In contrast to most studies on RF, we dichotomized RF scores into categories of low and medium RF, and like Fonagy et al. (1996), we set a cutoff at an RF of 3. Transcripts rated <3 on the RF scale were characterized by very restricted to complete absence of RF; subjects provided concrete explanations and referred to physical or sociological, rather than psychological, reasons for human behaviors. On the other hand, in transcripts with RF ≥3, subjects were usually capable of referring to mental state, although some of these explanations were rudimentary or cliché. Hence, our study differentiated between patients with or without mentalizing capacity and was not designed to detect finer differences between patients with, for example, medium and high mentalizing capacity.

In summary, our findings contradict the assumption that patients who function particularly poorly are in need of more-extensive treatment. Accordingly, the assumption that day-hospital treatment is more appropriate for patients with severe PD is challenged by findings in this thesis.
6.2 Methodological issues

As in all scientific fields, this dissertation relies on making valid inferences. Validity refers to whether a study is able to scientifically answer the questions it is intended to answer. The methodological considerations in this dissertation follows the structure of Gullestad (2012) who discusses aspects of validity using the typology of Cook and Campbell (1979) as further developed by Shadish, Cook & Campbell (2002); Statistical-conclusion validity, internal validity, and external validity.

6.2.1 Statistical-conclusion validity

Statistical-conclusion validity is the degree to which conclusions about the relationship among variables based on the data are correct or “reasonable.” Two types of errors can occur: type I (finding a difference or correlation when none exists) and type II (finding no difference when one exists). Statistical-conclusion validity concerns the characteristics of the study that make these types of errors more likely. There are several threats to statistical-conclusion validity, for example low statistical power, violated assumptions of test statistics, repeated testing for significant relationships (fishing), or unreliability of measures (Shadish et al., 2002).

Although we consistently sought to minimize the number of outcome measures in each paper, the number of outcome measures in papers I, II, and IV raises the possibility of type-I errors. We did not adjust for the relatively large number of outcome measures with Bonferroni corrections, and therefore we cannot exclude the possibility of false-positive findings. However, the use of a multilevel mixed modeling approach in papers I, II, and IV reduced the risk of type-I errors compared to more-traditional statistical analyses. When dependence of residuals results from covariance between data levels, traditional statistical analysis (e.g. repeated-measurement analysis of variance) often yields estimates of standard error that are too conservative. In contrast, we employed a multilevel approach that accounts for a lack of independence due to nesting of the treatment conditions.

Longitudinal studies typically allow individuals to be assessed at several time-points over several years, thereby providing important clinical information about changes over time, variability in change patterns, and systematic differences in subgroups (predictor and moderator analyses). The methods section above addresses the special features of longitudinal studies, factors that complicated our
analyses, and how these complications were solved in the UPP. Longitudinal studies also enable the answering of questions about causal inferences, as in paper III (additional information appears in the method section). Such analyses are ambitious in a field like psychiatry, but can potentially uncover new and unique information about mediators of change in complex treatment processes like psychotherapy. However, the communication of such complex analyses is challenging in a field in which most researchers have clinical backgrounds from psychiatry and psychology. We have sought to detail how the MSM-based approach was carried out in order to ensure the reproducibility of our findings; this level of detail may have affected the readability of this paper and reduced the chance of non-statisticians understanding the results.

Concerning statistical power, the sample size in the UPP varied across the studies included in this dissertation; paper II, which included analysis of the subset of patients with borderline PD, had the smallest sample size. However, we conducted a post-hoc power analysis in this study using a simplified model with constant residual variance and a linear fixed effect of time. A formula was used to approximate the observed power (Fitzmaurice, 2004). For GSI, the power was ~0.65, indicating that the same inference would result in a 65% chance of repeated identical experiments. A larger sample size would have increased the statistical power and reduced the chance of type-II errors. Overall, the statistical tests uncovered significant results and the effect-size estimates were in the moderate to large range. We also detected non-significant trends in papers I, II, and IV that could have achieved significance with a larger sample.

6.2.2 Internal validity

Internal validity refers to whether inferences about a causal relationship between an independent variable and a dependent variable can be drawn (Shadish et al., 2002), which depends on the extent to which a study minimizes systematic error (bias). In the present thesis, the internal validity of the studies dictates whether the observed results of differential treatment effects can be attributed to differences between the treatment conditions. The randomized design of the UPP strengthens the internal validity of our studies. Random assignment is essential for achieving initial comparability between treatments and reduces the plausibility of alternative explanations for observed effects (Shadish et al., 2002). Randomization of patients in the UPP was regarded as successful because the patient groups in the two treatment conditions were comparable at baseline in terms of all outcome variables, sex, sociodemographic factors, and diagnoses (there were no significant differences between groups). Nevertheless, here we consider several threats to the internal validity of these
studies: attrition, the natural course of PDs, treatment specificity, patient expectations, self-reporting, and repeated testing.

Attrition (the loss of participants to follow-up) is a serious threat to internal validity. For the full sample of UPP, the lowest response rate for the follow-up investigations occurred at 6 years, when 70% of the original sample participated. To avoid constructing statistical models of attrition bias, maximum likelihood is often chosen for estimations under the “missing at random” assumption. In the present studies, this assumption was probably violated to some degree. Although it is not possible to test this assumption, comparison with a complete-case analysis (participants without dropouts) is informative. For all papers in this dissertation, complete-case analyses revealed patterns that were similar to those of the intention-to-treat analyses. Also, papers I, II, and IV all reported small differences at baseline between patients that attended the 6-year follow-up and those that did not.

Regarding the natural course of PDs, the lack of a non-treatment comparison group hindered us from concluding that the treatment was responsible for the observed improvements, thus representing a threat to the internal validity of the study. However, this problem is common to all long-term comparative studies with patient populations as severely disturbed as ours; inclusion of a non-treatment comparison group would be unethical and unrealistic. Natural observations of patients have reported an estimated recovery rate of 4-12% per year without treatment, compared to 30-74% after 1 year of treatment for various PDs (Perry, 2001, as cited by Vinnars 2005). Estimating the natural course of PDs and comparing the estimates of untreated PDs to treated PDs is tricky because it is reasonable to assume that patients who seek treatment differ from those who do not in important ways. Thus, there is considerable uncertainty attached to these estimates. In the present study, after 6 years, 70% of patients no longer met the criteria for any PD diagnosis and 49% scored above the clinically relevant cut-off score of GAF >60 points. Recent studies have indicated a greater fluctuation of PD diagnoses than previously assumed (Hopwood et al., 2011; Zanarini et al., 2012), whereas continued impairment in psychosocial functioning seems to be a more enduring component of PDs (Skodol et al., 2005). Given these estimates, it does not seem likely that the results in the present study are solely due to a “natural recovery process.” Moreover, the differential effects or courses in the two treatment conditions cannot be explained by a natural-course hypothesis.

Regarding treatment specificity, we cannot exclude the possibility that aspects other than differences in treatment modality (differences between therapists, theoretical orientation, frequency of sessions) could explain some of the differential effectiveness observed here. Although this issue represents a
possible threat to the internal validity of papers I, II, and IV, there is rarely only one difference between two groups in psychotherapy research.

UPP therapists were assigned to one of the two treatment conditions in accordance with each therapist’s individual preference. This assignment strategy may have created a selection bias, generating an interaction between the characteristics of the therapists and the treatment modality. For instance, experienced and more-skilled therapists may have chosen the OIP condition because they felt that they did not require intensive supervision; the opposite scenario may also have occurred. We determined that individual therapists in the OIP were significantly more experienced in terms of “years of practice” than therapists in the step-down treatment. Nonetheless, the therapists in both groups were highly experienced and reported approximately the same level of satisfaction with their work. Prior research on the effect of therapist experience on outcome has been mixed and inconsistent (Beutler et al., 2004; Wampold & Brown, 2005). Based on available information, there is no reason to assume that therapists in the OIP condition were more effective than therapists in the outpatient treatment in the step-down treatment. However, we speculate that therapists in the OIP group felt more confident as therapists because they could treat patients according to their preferred method and practice. In contrast, therapists in the step-down treatment were required to practice therapy according to treatment guidelines with which they had little prior experience.

Also relevant to internal validity is the relative lack of specificity with regard to treatment conditions. The step-down treatment program was multifaceted and complex, which makes it difficult to specify the effective ingredient(s) and to interpret the results. This study did not include measures of treatment adherence or competence, which affects our ability to draw conclusions about the possible impact of the quality of therapy on treatment outcome. Moreover, there was a lack of specificity with regard to the OIP condition, which consisted of different psychotherapies conducted by a heterogeneous group of therapists, mainly in private practice. Therapists in the OIP were free to refer participants to other treatments based on their clinical judgment.

Patient freedom to engage in other types of treatment without being excluded from analysis increases our ability to generalize our findings, making the therapies offered in the present study similar to therapy practiced in “real life.” At the same time, the intended treatment difference or contrast is thwarted, threatening both the internal validity and the construct validity of the treatments (Shadish et al., 2002) and making it difficult to specify the essential difference(s) between the two treatment conditions and the higher-order constructs that they represent. Between-group comparisons are said to be optimal when “each client is randomly assigned to receive one and only
one kind of therapy” (Kendell, Holmbeck, & Verduin, 2004). This was not the case for many patients in the UPP. However, we gathered extensive data on how much and what kind of additional (non-study) treatments the patients received, allowing adjustments for these variables. This strategy was employed to some degree at the 3-year follow-up by excluding patients in the OIP who had received other day-hospital treatment in control analyses (Gullestad, 2012). These outcomes did not differ from the outcomes of the full sample. No such analyses were conducted after the 6-year evaluation, and we cannot exclude the possibility that non-study treatments may have differentially impacted long-term outcome. There were no differences in the self-reported use of health-care services in the past year at the 6-year follow-up.

Another possible threat to internal validity is the fact that some patients in the OIP treatment were disappointed with not being allocated to the step-down program. Several therapists from the OIP group reported that their patients were upset about not receiving treatment at the Department of Personality Psychiatry to which they had been referred. Patients in the UPP were poorly functioning patients with severe PDs (GAF <50 points), a patient group known to represent a challenge to treatment services due to the dysfunctional nature of their disorders. Many patients with PDs have experienced repeated traumas earlier in life and are skeptical toward other people, including health personnel. Thus, many of these patients have difficulty establishing good relationships, including treatment alliance with their therapists. Patient dissatisfaction with their allocation to outpatient treatment may have impacted their motivation, making it even more difficult for therapists to establish good working relationships with them.

Research has documented that patient treatment and outcome expectations are significant contributors to the effectiveness of therapy (Greenberg, Constantino, & Bruce, 2006). We did not measure patient motivations or expectations, making it difficult to control for the possible impact of dissatisfaction. The difficulties experienced by some of the outpatient therapists may have impacted our study results. On the other hand, the OIP condition achieved greater results than the step-down group at 3-years follow-up, suggesting that this effect was minor.

A considerable amount of data in the UPP was obtained via self-reported questionnaires. Self-reported measures are popular for several reasons. First, they represent an affordable way (in terms of both time and cost) to gather data and can easily be implemented in large samples. Second, they can be used to measure constructs that would be difficult to assess through behavioral or physiological measures (e.g., personality functioning). However, the use of self-reported data also threatens the internal validity of a study for several reasons. Persons may answer questions in a
manner that will be viewed favorably by others, known as social desirability bias. Other response biases are the tendencies of some persons to respond a certain way regardless of actual evidence they are assessing (Fan et al., 2006). Also, self-reported studies are inherently biased by the person’s introspective abilities, understanding of particular questions and feelings at the time they fill out the questionnaire (Austin, Gibson, Deary, McGregor, & Dent, 1998).

When repeatedly completing the same test, patients may “learn” the tests or become exhausted. Patients in the UPP completed self-reported measures several times. However, the periods between the tests and retests were relatively long and the number of measurements was large, thereby reducing the possibility that repeated testing influenced the results. Self-reported measures were also supplemented with interview-based scoring and diagnostics. This methodological issue was equally prevalent in both treatment conditions and should therefore not have influenced the observed treatment differences between groups.

6.2.3 External validity

External validity is the extent to which the results of a study can be generalized to other situations and to other people (Aronson, Wilson, Akert, & Fehr, 2007). There is always a trade-off between the internal validity of well-controlled efficacy studies, which make causal inferences possible, and the external validity of studies of treatment effectiveness, which increase the possibility of generalizing the results.

Patient samples in research trials have been criticized for not being representative of the broader population of help-seeking patients; patients are recruited by researchers and referred by health practitioners, and they are often selected for homogeneity (Falkenstrom, 2010). Sample representativeness is an important part of this dissertation. While many RCTs for PDs focus on only one PD and include fewer comorbidities, participants in the UPP exhibited a broad range of PDs (we only excluded patients with antisocial and schizotypal PDs). Moreover, patients were referred from general and mental-health practitioners and were not recruited by researchers. The most frequent diagnoses in our patient sample were borderline PD and avoidant PD, followed by paranoid PD. PDs in the UPP were quite representative of the broader population of PD patients that undergo treatment in ordinary outpatient clinics (Narud, Mykletun, & Dahl, 2005) and in the Norwegian Network of Personality-focused Treatment Programs (Karterud et al., 2003), thus increasing the generalizability of our findings.
However, our sample may not be readily comparable to the larger group of PD patients who seek treatment at community centers. Falkenström (2010) distinguished between patients who seek help at a community center and patients who actually start therapy. In the UPP, many patients were referred to treatment but did not actually start treatment. Although most of these patients were excluded because they did not meet the inclusion criteria, some dropped out of the assessment phase before attending all evaluations. This behavior may have been due to a lack of motivation. Thus, an unintentional selection process occurred prior to the initiation of therapy.

Falkenström (2010) investigated whether patients who seek therapy at a community center do worse than those who start therapy at the same centers or those who receive therapy as participants in a research trial. They determined that patients that sought help but did not start treatment had worse outcomes than patients in an average research trial; patients who actually started therapy experienced improvements that were comparable to those of patients in research trials (Falkenström, 2010). Thus, results cannot necessarily be generalized from a research trial to the broader populations of help-seeking patients, but it may be easier to generalize to the sample of patients who actually start therapy (after an assessment phase).

Comparison with patients from other research studies may be helpful in gaining an impression of the sample in the present study. Compared to patients in the study by Fonagy and Bateman (1999), patients in the UPP were less disturbed, as indicated by the mean levels of symptom distress (GSI): 2.4 points in Fonagy and Bateman (1999) versus 1.7 points in the UPP. The sample in the present study seems to be more comparable to samples in some studies of OIP for PDs (e.g., Davidson et al., 2010; Giesen-Bloo et al., 2006). However, while the UPP included a wide range of PDs, the studies by Davidson et al. (2010), Giesen-Bloo et al. (2006) only included patients with borderline PD. Since patients in the study of Fonagy and Bateman (1999) had greater impairment, one could hypothesize that more severely disturbed patients need more intensive treatment, while better-functioning patients improve most in outpatient treatment. Contrary to our expectations, the most severely disturbed patients (as indicated by RF level) responded better to OIP than to the more-extensive step-down treatment program. Future research should address the current lack of information about the types of treatment that can be adapted for severely disturbed PD patients.

Regarding the representativeness of the treatments utilized in the UPP, day-hospital treatments (with and without “continuation treatment”) as well as outpatient individual psychotherapy were common treatment modalities for patients with PDs in Norway at the beginning of 2000. How representative were the treatments in the UPP, and how generalizable are our results? Individual
outpatient treatments were conducted by a group of therapists considered to be fairly representative of experienced and well-qualified psychotherapists (psychiatrists and psychologists) in private practice in Oslo (Wilberg et al., 2014). Most patients (83%) received individual psychotherapy once per week, which is comparable to the average treatment frequency among therapists in private practice with a contract with the State Health Insurance Fund in Norway (one consultation per week for 1.5-2 years; Gråwe et al., 2005). It is also reasonable to believe that the treatment styles of these therapists are representative for Scandinavian countries; these styles operate within welfare states with several kinds of available health care, such as emergency units, general practitioners, and other day hospitals. At the same time, there is reason to assume that the therapists in the OIP condition were of higher quality than other therapists in the Oslo area that engage in treatment as usual. Moreover, because many therapists in the OIP saw the research trial as a chance to document the effectiveness of their therapeutic practice, they may have been more flexible with their therapy strategy than normal.

The day-hospital portion of the step-down treatment program was representative of short-term day hospitals in Norway and Scandinavia at that time (Karterud et al., 2003). Thus, in contrast to many RCTs in which treatments depart markedly from clinical practice (Kazdin, 2008), the therapies offered in the present study were, in many respects, representative of treatments in ordinary settings. However, the outpatient combined group and individual psychotherapy part of the step-down program was somewhat more experimental. Although the hospital-based staff had prior experience with occasional combined group and individual psychotherapy, they were not used to using the conjoint treatment as a standard format or to including external psychotherapies that were not known to them. Thus, our results cannot be generalized to combined psychotherapy in general.

Another possible threat to the external validity of the present study concerns the interaction between patients and therapists during the initial phase. Normally, a central part of the initial evaluation involves tapping the patient’s motivations. Therapists often try to motivate patients to start group therapy, discussing in detail what group therapy is about and preparing them for the challenges that they may encounter within the group-therapy format. However, in the present study, therapists could not motivate patients as they normally would because they did not know whether the patients were going to be randomized to the day-hospital step-down treatment, consisting of several types of group therapies, or to OIP. Therefore, the initial assessment phase could be viewed as somewhat “artificial” or different from a “real-life” evaluation, with a possible reduction in the external validity of the study.
6.3 Summary of methodological strengths and limitations

The external validity of RCTs has been questioned due to their manualized treatments, close supervision, and exclusion of patients with comorbidity, thus reducing the possibility of making generalizations to therapy in the “real world.” However, this study used a population of patients seen in everyday practice, including patients with low socioeconomic status, poor health, and comorbidity, thereby generating findings that are more readily generalized to everyday practice. Moreover, the treatments offered through the UPP are fairly representative of treatments for PDs in ordinary clinical practice; they were not fully standardized or manualized, and medication was determined on an individual basis. Thus, despite its randomized design, the UPP combines elements from studies of efficacy and studies of effectiveness. Several possible threats to the internal validity of the project were difficult to control (for example, patient attrition, the natural recovery process, and other treatments received). Causal inferences about treatment effects must be drawn with caution and regarded as preliminary. The main limitations were the lack of assessments of the treatment qualities and the difficulties involved in specifying the essential differences between the two treatments, making it more difficult to draw causal inferences.

7. Concluding remarks and implications

The overall aims of this dissertation were to examine potential differences in the long-term outcome of patients receiving day-hospital step-down treatment or OIP and to examine the long-term processes of change in patients that participated in the UPP. Findings from the full mixed sample of the UPP indicate that both treatment modalities are viable treatment options for patients with PD at a group level. Thus, for PDs in general, the level of care for psychotherapeutic treatment does not seem crucial to long-term outcome.

However, analyses of subgroups of the UPP sample indicated that level of care, represented by the treatment modalities in the UPP, seems to be important for certain types of personality pathology. Patients with borderline PD who received more-intensive day-hospital treatment followed by conjoint group and individual outpatient treatment had more favorable long-term outcome than patients allocated to individual outpatient treatment alone. We determined that mentalizing capacities, which were operationalized as RF, moderated clinical outcomes in a treatment condition-dependent fashion; patients with a more-developed capacity for mentalization achieved better
results in the step-down program. In contrast, patients with low or no mentalizing capacity had better outcomes with OIP than with the step-down program.

Further, we detected associations between mentalization capacity and a wide range of clinical variables in a sample of patients with borderline PD and/or avoidant PD. Before entering study treatment, patients with a low capacity for mentalization reported higher levels of symptomatic and interpersonal distress, more self-esteem problems, lower psychosocial functioning, and lower levels of personality functioning regarding identity and relational functioning.

Consistent with previous results from prospective naturalistic studies and treatment trials, most patients in the UPP tended to achieve remission from their PD diagnoses. However, many patients continued to display long-term impairments in psychosocial functioning. Our findings revealed a causal relationship between increases in personality functioning (as measured with the SIPP-118) and later increases in psychosocial functioning (as measured with the GAF). Also, there was a causal relationship between psychosocial functioning and later improvement in personality functioning, although this effect was somewhat smaller.

7.1 Clinical implications

Due to the limitations of this study (discussed above), as well as the need to replicate our findings, it is probably premature to state definitely that these results should influence the clinical treatment of PDs. Nevertheless, the results of this study may have several clinical implications and suggestions as to the optimal treatment format for PDs.

Our results highlight the large heterogeneity within the population of patients with PDs, both between and within PD categories, underscoring the importance of a flexible approach to the treatment of PDs. Our finding suggests that a more-intensive treatment format that includes both group and individual psychotherapy could be particularly helpful for patients with the emotional dysregulation, stormy relationships, and destructive acting-out behaviors seen in borderline PD, at least more than OIP alone. This dissertation also demonstrated that differences in baseline capacity for mentalization were predictive of which patients achieved better results within each treatment modality. Associations between RF and the intensity of symptom distress and psychosocial impairment further indicate that the mentalization construct captures clinically relevant phenomena and point to the importance of evaluating levels of mentalization before a patient begins therapy.
Another possible clinical implication relates to the concept of personality functioning (self and interpersonal relations) and how aspects of personality functioning should be targeted through treatment. Our findings suggest a causal inference between increase in personality functioning and later improvement in psychosocial functioning, emphasizing the importance of increasing the adaptive levels of identity integration and relational capacities in treatment. We also determined that the causal association between psychosocial functioning and personality functioning is reciprocal, and hence it may be helpful to distribute resources to both therapy and vocational counseling/rehabilitation.

One cannot necessarily draw inferences about single cases from statistical relationships (Elster, 2007); thus, decisions about individual treatment strategies should not be made based on findings from this study. Nevertheless, our results can inform clinical thinking and decision-making in the way that a therapist approaches each unique patient (Gullestad, 2012).

7.2 Implications for future research

We hope that this dissertation will stimulate further studies on the roles of various treatment modalities and levels of care for patients with PDs. Rather than seeking one level of care that is optimal for all PD patients, the findings of this thesis suggest that future research should emphasize treatment in accordance with what works for whom.

In order to optimize treatment for each patient with PD, we must determine which information enables clinicians to identify treatment targets and to select treatment methods. Our findings concerning the clinical impact of mentalizing capacity and personality functioning suggest that traditional categorical diagnoses do not necessarily address these issues. Future research should focus on the moderator effects of concepts such as mentalization and personality functioning. Our observations of a causal inference between personality functioning and psychosocial functioning and associations between the personality-functioning domains Identity Integration and Self-control and other clinical variables support the inclusion of the Level Personality Functioning Scale in section III of the DSM-5. Gladly, this section is included in DSM-5 to stimulate further research; we hope that such research will help clinicians provide patients with the best possible treatment.

We need to know more about how and why treatment works. Thus, future research should not only seek to identify the optimal treatment or treatment format for each patient, but should also explore the effective ingredients in the various treatments (Gullestad, 2012). The causal inference between
personality functioning and psychosocial functioning suggests that better functioning in the domains Identity Integration, Relational Capacity, and Self-control may reflect processes of change during psychotherapeutic treatment of patients with PDs. Further research on causal inference will elucidate the processes that lead to change during psychotherapy. Mentalization, which has been proposed as an active ingredient in all effective psychotherapies with PDs (Allen et al., 2008), is a promising concept in this context. Further analyses should address mentalization capacity as a potential process that leads to change in PD patients. Research on mentalization capacity should lead to further investigations of the RF scale. In its current form, the RF scale has several shortcomings, and future research should investigate the reliability and validity of RF and its various sub-dimensions (Gullestad, 2012). Mentalizing capacity is likely to be altered in states of emotional arousal, and the AAI may not capture fluctuating mentalizing deficits in current situations and relationships. Future research should determine the feasibility and effectiveness of more domain- or context-specific versions of RF.

References


76
depressive disorder. Remission, recovery, relapse, and recurrence. Arch Gen Psychiatry, 48(9), 851-855.


with a personality disorder. *J Nerv Ment Dis*, 198(2), 110-115. doi: 10.1097/NMD.0b013e3181cc0d59


Zanarini, M. C., Frankenburg, F. R., Reich, D. B., & Fitzmaurice, G. (2010a). The 10-year course of psychosocial functioning among patients with borderline personality disorder and axis II


