On the effectiveness of psychotherapy in personality disorders

Anna Bartak
On the effectiveness of psychotherapy in personality disorders

Anna Bartak
© Anna Bartak, 2010

Cover Photo: Ben Melis
Layout: Simone Vinke, Ridderprint, Ridderkerk, The Netherlands
Printed by: Ridderprint, Ridderkerk, The Netherlands

ISBN 978-90-5335-311-0

The publication of this dissertation was financially supported by Psychotherapeutisch Centrum De Viersprong, Erasmus MC Rotterdam, Altrecht Cura, Arkin, de Gelderse Roos Centrum voor Psychotherapie, and Vereniging voor Klinische Psychotherapie.
On the effectiveness of psychotherapy
in personality disorders

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties
ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 15 oktober 2010, te 13:00 uur

door

Anna Bartak

geboren te Lich, Duitsland
Promotiecommissie

Promotores:  Prof. dr. P.M.G. Emmelkamp
             Prof. dr. T. Stijnen

Co-promotor:  Prof. dr. J.J. Busschbach

Overige Leden:  Prof. dr. J.H. Kamphuis
                Prof. dr. A. Lange
                Prof. dr. W. van den Brink
                Prof. dr. P.M.M. Bossuyt
                Prof. dr. A.H. Schene
                Prof. dr. A.R. Arntz
                Prof. dr. W.J. Livesley

Faculteit der Maatschappij- en Gedragswetenschappen
# Contents

**Chapter 1**  
Introduction  
7

**Chapter 2**  
Strengthening the status of psychotherapy for personality disorders: An integrated perspective on effects and costs  
17

**Chapter 3**  
The use of propensity score methods in psychotherapy research. A practical application  
31

**Chapter 4**  
Effectiveness of different dosages of psychotherapeutic treatment for patients with cluster C personality disorders: Results of a large prospective multicentre study  
47

**Chapter 5**  
Effectiveness of outpatient, day hospital, and inpatient psychotherapeutic treatment for patients with cluster B personality disorders  
67

**Chapter 6**  
Patients with cluster A personality disorders in psychotherapy: An effectiveness study  
87

**Chapter 7**  
General discussion  
109

References  
121

Summary  
141

Samenvatting (Summary in Dutch)  
149

Acknowledgements  
157

Publications  
163
Chapter 1

Introduction
Introduction

This thesis is on the effectiveness of different “dosages” of psychotherapy in patients with personality disorder (PD). It is the first large research project elucidating the impact of dosage, i.e., the amount of therapy, and how this affects treatment outcome in psychotherapy for PDs. In three quasi-experimental effectiveness studies, the effects of different dosages of psychotherapy for cluster A, B, and C PDs are evaluated. These studies aim to contribute to the understanding of psychotherapy for PDs and to refine the toolkit for psychotherapists working with this patient group.

Definition of PD

PD nowadays is defined as “an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual’s culture, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment” (DSM-IV-TR; American Psychiatric Association, 2000).

The roots of the current definition of PD go back to historic documents like the “characters” of Theophrastus (371-287 BC; Pirckeymherus, 1527), the “insanity without delirium” of Pinel (1801), or the “psychopathic personalities” of Schneider (1923), with a first attempt of categorisation. Doctors, researchers, and philosophers have been interested in PDs for centuries, trying to explain this phenomenon and discussing the question of changeability. PDs were first established formally in the third version of the Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association (DSM-III; American Psychiatric Association, 1980). The two important steps achieved in DSM-III were (a) the placement of PDs on a separate axis, which recognises their clinical significance and high prevalence; and (b) the description of explicit diagnostic criteria for each PD encouraging empirical research (Livesley, 1995). This classification system was also adopted by the World Health Organisation (WHO) in their International Classification of Diseases and Related Health Problems (ICD-10; World Health Organization, 1992).

Today in DSM-IV-TR the ten official PDs are classified in three clusters (see Figure 1.1): cluster A — the odd cluster, cluster B — the dramatic cluster, and cluster C — the anxious cluster. Additionally, the classification comprises a category of PD not otherwise specified (PD NOS), including two provisional diagnoses, namely depressive PD and passive-aggressive PD.
**Figure 1.1. Classification of PDs according to DSM-IV-TR**

This classification was designed to simplify professional communication and is currently used by scientists and practitioners throughout the world. However, it has also been subject to criticism. A major revision of the DSM will be published in May 2013 (DSM-5; American Psychiatric Association, 2010), taking this criticism into account. The current PD classification will be largely reformulated with the aim of greater clinical utility (Tyrer et al., 2010). It will most likely be replaced by a general definition of PD, accompanied by (a) five severity levels of personality functioning, (b) five PD types (i.e., antisocial/psychopathic type, avoidant type, borderline type, obsessive-compulsive type, schizotypal type), and (c) a dimensional description of the patient in terms of personality trait domains. The present research, conducted between 2003 and 2006, is still based on the three traditional clusters.

**Impairment of PD patients**

Research has shown that patients with PD have largely impaired well-being and functioning, and that this impairment is indeed due to the presence of PD (Cramer, Torgersen, & Kringlen, 2006; Soeteman, Verheul, & Busschbach, 2008). What distinguishes PDs from (ego-dystonic) clinical syndromes such as depression or anxiety is their rootedness in the character of the person (Emmelkamp & Kamphuis, 2007). PDs are ego-syntonic: Maladaptive personality traits are the cause of ongoing struggle in a patient’s life. The disorder is much more something they “are” rather than they “have”. In conclusion, patients with PD are a group in need of treatment, but treatment is far from easy as complaints are strongly interwoven with the patient’s character.
Assessment of PD

The most reliable way to assess PD is to conduct a semi-structured interview, such as the Structured Interview for DSM-IV Personality (SIDP-IV; Pfahl, Blum, & Zimmerman, 1997), used in present studies. Self-report questionnaires at most should be screening instruments and cannot replace a diagnosis made by a clinician based on a semi-structured interview (McDermut & Zimmerman, 2008).

Prevalence, aetiology, and course of PD

PDs are highly prevalent both in the general community and in psychiatric patients. In the community, the point prevalence of any PD lies between 7.3 and 15.7% (Crawford et al., 2005; Klein et al., 1995; Lenzweger, Lane, Loranger, & Kessler, 2007; Maier, Lichtermann, Klingler, Heun, & Hallmayer, 1992; Moldin, Rice, Erlenmeyer-Kimling, & Squires-Wheeler, 1994; Samuels et al., 2002; Torgersen, Kringlen, & Cramer, 2001; Zimmerman & Coryell, 1989). The lifetime risk of having a PD is probably much higher: an estimated 30 to 40% (Johnson, Cohen, Kasen, Skodol, & Oldham, 2008; Torgersen, 2009). In (outpatient) populations of psychiatric patients, the prevalence of PD is estimated at 45.5% (Zimmerman, Rothschild, & Chelminski, 2005). Overall, the most common diagnosis in the general community is obsessive-compulsive PD, while in psychiatric populations it is mostly borderline PD (Torgersen, 2009; Zimmerman, et al., 2005).

Researchers have tried to disentangle the influence of genetics and environmental factors on the development of PD. Torgersen, who conducted several twin studies in Scandinavia, found that genetic factors contribute to around 40 to 50% of the variation, with a very small effect of shared family environment (Torgersen, 2009). This implies that PDs seem to be more strongly influenced by genetic effects than almost any Axis I disorder (Torgersen et al., 2000). But so far, as this kind of research still is in its infancy, there are no final conclusions. The answer to the question “How do PDs develop?” most likely lies in the interplay between genetic and environmental factors, together with gene-environment, gene-experiences, and gene-gene interactions (Torgersen, 2009). If we wish to draw further conclusions, longitudinal multivariate genetic studies are needed (Emmelkamp & Kamphuis, 2007).
Contrary to a widely held belief, PD characteristics do change over time (Johnson et al., 2000; Seivewright, Tyrer, & Johnson, 2002). Skodol et al. (2005) showed that some characteristics are more stable and others change more easily. The more impulsive aspects of borderline PD seem to vanish slowly with age, while cluster A and C PD traits seem to become more pronounced (Paris & Zweig-Frank, 2001; Seivewright, et al., 2002). There remains considerable debate about the natural course of PDs and their changeability. However, treatment can significantly accelerate the improvement of harmful characteristics (Perry, Banon, & Ianni, 1999). Research into the merits and methods of PD treatment is therefore important.

Research on PD and treatment

In the last 30 years scientific literature on PD has grown steadily (see Figure 1.2), which can largely be attributed to the merits of DSM-III. This new classification considerably stimulated research on the subject of PD.

**Figure 1.2.** Number of publications on PDs (source: http://www.ncbi.nlm.nih.gov/pubmed/)

![Number of publications over time](image)

*Search terms: “XXXX”[Publication Date] AND (“personality disorders”[ti] OR “personality disorder”[ti])

The increase in studies on PDs over the last 30 years is paralleled by a steadily rising number of papers on psychotherapeutic treatment of PDs, revitalising the discussion about the changeability of personality and PDs. Particularly in the last decade, many therapies tailored to a PD patient population were designed and validated empirically (see Figure 1.3).
Figure 1.3. Number of publications on psychotherapeutic treatment of PDs (source: http://www.ncbi.nlm.nih.gov/pubmed/)

PD & Psychotherapy

*Search terms: “XXXX”[Publication Date] AND (“personality disorders”[ti] OR “personality disorder”[ti]) AND (“psychotherapy”[ti] OR “psychotherapeutic”[ti])

From the increasing numbers of studies on psychotherapeutic treatment of PDs, the consensus has grown that psychotherapy is the treatment of choice for patients with PD (e.g., Landelijke Stuurgroep Richtlijnontwikkeling in de GGZ, 2008; National Institute for Health and Clinical Excellence, 2009a, 2009b). However, when asking how much psychotherapy is best for certain groups of PD patients, we have no answer yet. Unfortunately, the concept of dosage, as widely used in pharmacological research, has thus far received little attention. It still plays a negligible role in both research into the most effective treatment for PDs and psychotherapy research in general. This lack of attention is arguable. When it comes to questions of effectiveness and cost-effectiveness, dosage is one of the main determining factors. In a classic study, Howard et al. (1986) showed in a large patient group covering multiple diagnostic groups that after one year of outpatient treatment, 85% of the patients were measurably improved in terms of “global” outcome measures. Different symptomatic classes showed different patterns of recovery, as the authors demonstrated in a later study (Kopta, Howard, Lowry, & Beutler, 1994). Taken together, their research efforts should have cleared the path for the discipline of dose-effect research in psychotherapy (Kopta, 2003), but unfortunately this area of research is still small and—with few exceptions—limited to outpatient settings. For patients with PD, Bateman and Fonagy (2000) argued that effective treatments should be “relatively long-term”, but detailed information about the most effective duration is lacking. Like in general psychotherapy research, researchers and health institutions in the PD field were mainly engaged in the establishment of specifically tailored therapies and the comparison of therapies from different therapeutic schools within the same dosage (e.g., 40 sessions of outpatient treatment with therapy A versus 40 sessions of outpatient treatment with therapy B). What we know from these studies is that
almost all specifically tailored therapies show a higher effectiveness than treatment-as-usual or waiting list controls. When different specifically tailored therapies are compared, though, mostly no difference is found (Bartak, Soeteman, Verheul, & Busschbach, 2007). In other words, specialised psychotherapy for PD has effect, but different contents of therapy seem equally effective. This has been shown in most existing effectiveness studies, Emmelkamp (2006) and Giesen-Bloo (2006) being the exceptions. The fact that most of these studies compare equal dosages of therapy is understandable given the difficulty of randomly assigning patients to widely differing treatments (e.g., outpatient vs. inpatient treatment). Randomising patients to treatments of different theoretical schools while keeping dosage constant causes less ethical problems. A consequence is, unfortunately, that dosage as a determining factor has been overlooked. If research wants to be meaningful and applicable to clinical practice, dosage has to be taken into account. The problem of selection bias due to non-randomisation can be tackled by strong statistical control of possible confounders in a so-called “quasi-experimental” study design. The present thesis is based on the first large-scale quasi-experimental study on dosage conducted in the field of PD.

SCEPTRE – Study on Cost-Effectiveness of Personality Disorder TREatment

This thesis is based on and a result of the project SCEPTRE, a large prospective psychotherapy study in the Netherlands (Viersprong Institute for Studies on Personality Disorders, 2010). SCEPTRE was initiated by the research department of “De Viersprong”, a psychotherapeutic centre in Halsteren and involved five other mental health sites throughout the Netherlands: Mentrum/Arkin (Amsterdam), Zaans Medical Centre (Zaandam), Altrecht (Utrecht), De Gelderse Roos (Lunteren), and GGZWNb (Bergen op Zoom/Roosendaal). More than 800 patients with personality pathology were included in the study between 2003 and 2006, and were followed for three years. Research was conducted on effectiveness and cost-effectiveness of different dosages of psychotherapy for patients with PD, as well as on treatment selection. SCEPTRE has a high follow-up response, making its results meaningful for clinical practice. Meanwhile, a five-year follow-up is being conducted, with results due in 2011.
Aims of this thesis

1. Explore what we know about psychotherapy for patients with PD, and what we still need to know.
2. Investigate the effectiveness of different dosages of psychotherapy for different groups of PD patients.
3. Examine a method for comparing the effectiveness of widely differing treatments without randomising patients.

Content of this thesis

Chapter 2 summarises the literature on psychotherapy for PDs and lines up current and future research goals. The existing evidence on effectiveness and cost-effectiveness of psychotherapy for PDs is presented and ways to overcome the present difficulties in its recognition and reimbursement in the public health field are discussed.

Chapter 3 explores a statistical method for comparing widely differing treatments when randomisation is not feasible. The propensity score method as a tool for the correction of selection bias is tested on its applicability for psychotherapy research.

Chapter 4 compares the effectiveness of five different treatment modalities for patients with cluster C PDs within a prospective non-randomised research design, using the propensity score method to correct for initial patient differences. The effects of long outpatient treatment, short and long day hospital treatment, and short and long inpatient treatment are compared at 12 months after baseline. Outcome areas in this chapter, as well as in chapters 5 and 6, are psychiatric symptoms, psychosocial functioning, and quality of life.

Chapter 5 compares the effectiveness of three different treatment settings for patients with cluster B PDs. The effects of outpatient, day hospital, and inpatient treatment are compared at 18 months after baseline, using the propensity score method to correct for initial patient differences.

Chapter 6 compares the effectiveness of three different treatment settings for patients with cluster A PDs and shows the limits of the propensity score method. The effects of outpatient, day hospital, and inpatient treatment are compared at 18 months after baseline.

Chapter 7, the general discussion, summarises the results and presents the strengths and limitations of the present studies. The findings are discussed in the light of past and future research, and the issue of clinical applicability is raised.
Chapter 2

Strengthening the status of psychotherapy for personality disorders: An integrated perspective on effects and costs

During the last decade, research has shown that psychotherapy is an effective treatment for patients with PD (e.g., Leichsenring & Leibing, 2003). Despite scientific evidence of effectiveness, psychotherapy is not fully deployed in PDs, nor is its reimbursement self-evident. We argue that psychotherapy has the potential to be the treatment of choice for people with PD and the potential to be valued highly by society if research and practice work together to present more convincing evidence to the medical field and the outside world.

Today, both clinicians and health care policy-makers increasingly rely on evidence-based medicine and health economics when determining a treatment of choice and reimbursement (Russell, Gold, Siegel, Daniels, & Weinstein, 1996; Rutten, Brouwer, & Niessen, 2005). A possible explanation for the leeway of psychotherapy in PDs could therefore be a mismatch of the presented scientific evidence and modern standards of evidence-based medicine and health economics. These modern standards not only focus on efficacy (Does this treatment work in a well-controlled environment?) but also on the added value and costs for patients and society. The status of psychotherapy as a valuable treatment for patients with PD will be endorsed if the psychotherapy field adopts these modern standards.

Understanding of this reasoning by clinicians working with patients with PD is warranted for several reasons. The first is that clinicians should be inspired to enhance their level of clinical practice according to the modern demands of evidence-based practice. A second reason is that clinicians are held more and more responsible for managing the scarce resources in health care as efficiently as possible to deliver beneficial interventions to as many patients as possible (Rutten, et al., 2005). Notably in the development of practice guidelines, it is important for clinicians to adopt modern quality standards, otherwise the increasing number of treatment options in PDs would be associated with increasing medical costs to be paid by society. One more reason to plea for the adoption of new clinical quality norms is the strategic argument of professional autonomy: If clinicians ignore the recent developments of evidence-based medicine and health economics, the risk is that mental health care decisions are taken by policy-makers alone and clinicians lose influence on developments in their own professional field.

The reasons mentioned above provide the rationale to accumulate and integrate empirical evidence and provide convincing arguments for the benefits of psychotherapy for PDs. In pharmacy, such integration already exists. In many countries the pharmaceutical industry has to convince physicians and reimbursement
authorities that their medication is safe, effective, and cost-effective. Moreover, reimbursement authorities may ask for evidence concerning the necessity of the treatment. This quest for comprehensive evidence in reimbursement issues is no longer limited to pharmaceutics. It is becoming more and more accepted as the preferred route for implementation of all medical interventions (Rutten, et al., 2005). If we apply this line of reasoning to psychotherapy for PDs, we have to create new standards of evaluation to strengthen the position of psychotherapy. This article aims to contribute to that understanding, by applying the criteria of evidence-based medicine on psychotherapy as a treatment for patients with PD. We will do this by critically analysing the available empirical evidence on effectiveness and cost-effectiveness, and combining this in an integrated model with necessity of treatment as a moderating factor.

Empirical evidence

Effectiveness
In evidence-based medicine, the highest level of evidence is achieved when empirical studies, preferably randomised clinical trials, can be combined in systematic literature reviews and meta-analyses. The evidence on the effectiveness of psychotherapy for PDs is evolving in that direction. In the last decade, two meta-analyses, six reviews, and one Cochrane review have been published (Bateman & Fonagy, 2000; Binks et al., 2006; Gabbard, 2000; Leichsenring & Leibing, 2003; Ogrodniczuk & Piper, 2001; Perry, Banon, & Ianni, 1999; Perry & Bond, 2000; Sanislow & McGlashan, 1998; Shea, 1993).

The available evidence clearly presents favourable results for the effectiveness of psychotherapy for PDs, notably borderline and avoidant PD. The first meta-analysis, published by Perry et al. (1999), showed that the effect size Cohen’s $d$ of psychotherapy for PDs was 1.1 to 1.3, against 0.25 to 0.5 for various control conditions, such as waiting lists or treatment as usual. This result is encouraging, as an effect size of 0.8 or higher is considered large (Cohen, 1988). The authors also analysed the relation between treatment duration and recovery in four outpatient studies and reported a strong dose-effect relation. After 1.3 years of outpatient psychotherapy (one or two sessions per week), an average of 52% of patients no longer met criteria for the diagnosis of PD. By modelling both recovery with treatment and natural recovery, they estimated that treatment is associated with up to seven times faster recovery than the natural course of PD (Perry, et al., 1999).
The second meta-analysis (Leichsenring & Leibing, 2003) described the effects of cognitive-behavioural therapy and psychodynamic therapy for PDs and showed that both therapies led to a significant decrease of symptoms. The average effect size for different outcome parameters was 1.5 for psychodynamic therapy and 1.0 for cognitive-behavioural therapy. Of the patients in psychodynamic therapy, 59% no longer met criteria for the diagnosis of PD after treatment. In the cognitive-behavioural therapy group, this figure was 47%. The difference between these two results must be interpreted with caution because the recovery figures are based on a small number of studies and the two therapies are not easily comparable, owing to different treatment durations. Nevertheless, an important conclusion from this meta-analysis is that psychotherapy not only reduces psychiatric symptoms, it also has a strong effect on personality pathology as well (Leichsenring & Leibing, 2003).

Recent evidence seems to confirm the results of the meta-analysis, showing that psychotherapy is a valuable treatment for patients with PD (Brown, Newman, Charlesworth, Crites-Christoph, & Beck, 2004; Chiesa, Fonagy, Holmes, & Drahorad, 2004; Emmelkamp et al., 2006; Giesen-Bloo et al., 2006; Huband, McMurry, Evans, & Duggan, 2007; Koons et al., 2001; Svanberg, Stiles, & Seltzer, 2004; Turner, 2000; Verheul et al., 2003). Further, Binks et al. (2006) published a Cochrane review on psychosocial interventions for patients with borderline PD, establishing the value of psychotherapy for this patient group on the highest scientific level. Their evidence suggested that with the help of psychosocial therapies patients show improvement in self-harm and parasuicidal behaviour, which are specific problem areas of borderline PD. The investigated therapies are still experimental and the number of studies is still too small; therefore, the authors concluded that their findings should be replicated in larger real-world studies (Binks, et al., 2006). Nevertheless, these results are an important step in establishing a firm base of knowledge for the effects of psychotherapy for PDs.

**Cost-effectiveness**

Although the evidence on cost-effectiveness of psychotherapy for PDs is still limited, we can draw some preliminary conclusions from the existing data. For instance, several cost-benefit studies provided arguments in favour of reimbursing psychotherapy for patients with PD. For that particular patient population, these studies indicated that psychotherapy can lead to significant reductions in the use of other (mental) health care services and, therefore, has the potential to reduce health care costs (Bateman & Fonagy, 2003; Dolan, Warren, Menzies, & Norton, 1996; Gabbard, Lazar, Hornberger, & Spiegel, 1997; Hall, Caleo, Stevenson, & Meares,
2001; Stevenson & Meares, 1999). For example, Stevenson and Meares (1999) showed that the costs for hospitalisation in 30 patients with borderline PD were reduced significantly following outpatient psychotherapy for 12 months. Their calculations indicated that psychotherapy rendered savings of $4833 per patient in the first year after treatment. For 24 patients with PD, Dolan et al. (1996) showed that costs on psychiatric care and imprisonment decreased after treatment. They even argued that, if the patients’ recovery continue, the treatment costs of psychotherapy would not only be paid back by the savings achieved through therapy but also lead to additional savings after two years. Gabbard et al. (1997) conducted a review of the published evidence on costs and reductions through psychotherapeutic treatment for PDs. They concluded that the total direct medical costs of psychotherapy were negative; accordingly, psychotherapy would not lead to expenditures, but to savings.

The studies mentioned above mainly focused on direct (medical) costs. But psychotherapy for PDs can also lead to a reduction in indirect costs, such as productivity losses caused by absenteeism, for instance. There is evidence to support this. Stevenson and Meares (1992) found that psychotherapy reduced absenteeism from work among patients with PD from an average of 4.7 to 1.4 months per year. In a follow-up after five years, this reduction was still evident.

A shortcoming of the studies published so far is that they cannot be classified as formal cost-effectiveness analyses. A lot of studies used tariffs as a proxy for costs instead of estimates of the true (direct and indirect) costs. Moreover, costs were usually presented out of context and were not explicitly related to the effects in a standardised cost-effectiveness ratio. However, this shortcoming does not necessarily jeopardise the evidence that for severe PDs, especially borderline PD, psychotherapy seems to save medical as well as work-related costs.

**Discussion**

Given the evidence regarding effectiveness and the preliminary—but nevertheless favourable—cost estimates of psychotherapy for PDs, the question arises: Why the reluctance to fully deploy psychotherapy as a treatment of choice for PDs and encourage its reimbursement? We will discuss three important obstacles for that deployment and ways to overcome them. The first obstacle is the interaction between (cost-)effectiveness and necessity, which is not yet fully recognised by the field. The second obstacle is the ongoing discussion about necessity of treatment.
The third obstacle consists of still existing gaps in the evidence of effectiveness, cost-effectiveness, and the assessment of necessity.

**Necessity—the missing link between cost-effectiveness and reimbursement**

(Cost-)Effectiveness is often proclaimed as the ultimate criterion for the value of a certain treatment, but it is not the only important factor in the reimbursement discussion. It is a fact that not all cost-effective interventions are reimbursed and some very expensive treatments with a low effectiveness are nevertheless reimbursed. A stereotypical example of this is the reluctance to reimburse Viagra with its eminent cost-effectiveness ratio, while lung transplantation is usually reimbursed in spite of high costs and low effectiveness (Stolk, Brouwer, & Busschbach, 2002). Obviously, factors other than cost-effectiveness play an important role in reimbursement policy. One of the identified factors is necessity of treatment. That is, the high need of patients assigned to lung transplantation actually gives rise to favourable sentiments in the reimbursement decision process, while the burden of erectile dysfunction in elderly men is not considered decisive. This means we should appreciate evidence about cost-effectiveness and necessity of treatment in a broader perspective. This integrated perspective is well recognised in health economics and is called the “equity debate”. In this equity debate, health economists discuss how efficiency should be traded off with solidarity toward the patients most in need, meaning a trade-off between cost-effectiveness and necessity of treatment. One could take an egalitarian point of view and argue that all resources should be allocated to patients most in need. One could also take a utilitarian point of view, arguing that health care resources should be spent efficiently to do as much good as possible, meaning we should spend the limited budget on interventions proven to be most effective. In practice, most people take an in-between position: We feel solidarity with patients most in need while at the same time we feel that interventions should be distributed efficiently. Consequently, if one proves that a treatment option represents an efficient remedy for patients high in need, chances for reimbursement increase. If our field were to present such evidence convincingly, the status of psychotherapy in PDs would be strengthened and reimbursement would be facilitated.

**Burden of disease—the missing proof for necessity of treatment**

The necessity of treatment for patients in psychotherapy still is a matter of debate. This is not just a popular belief. Even health policy-makers and clinicians tend to refer to patients in psychotherapy as YAVIS-patients. YAVIS is an acronym for young, attractive, verbal, intelligent, and successful, labelling patients who would benefit the most from psychotherapy. It is a common idea that psychotherapists prefer to help
YAVIS-patients (Schofield, 1964), that is, patients with a low burden of disease and thus a low necessity of treatment. Normally in health care, a high burden of disease is associated with a greater need for treatment—and more willingness to allocate financial resources (Stolk, Pickee, Ament, & Busschbach, 2005). Consequently, proving a high burden of disease gives access to resources.

In the field of psychotherapy, efforts to contradict the YAVIS assumption so far have failed. In general, studies use indicators of suffering that are most often only meaningful inside their own scientific community. Researchers in PDs are tempted to choose outcome measures as ego strength, defence style, and borderline PD severity. This may be meaningful within the field of PD, but these concepts do not present a reference point for comparing the suffering of PD patients to the suffering of somatic patients, for instance. Whenever treatments for PDs are competing for resources with treatments for somatic illnesses, the undetermined necessity of treatment pushes psychotherapy into a defensive position.

What is needed are indicators of suffering in PDs that are widely accepted in health policy. Such an unequivocal estimation of the necessity of treatment for PDs is only possible when measuring the burden of disease with generic measures, focusing on quality of life. Only then it is possible to relate the burden of disease of PDs to that of other mental and somatic disorders. Among the first to choose this approach were the investigators of the Dutch Standard Evaluation Project, who used the generic EuroQol EQ-5D (EQ-5D; Brooks, Rabin, & de Charro, 2003) in a large sample of patients admitted to specialised units providing psychotherapy. This study showed that those patients having severe personality problems and disorders experienced a high burden of disease (Soeteman, Timman, Trijsburg, Verheul, & Busschbach, 2005). A limitation of this study was the lack of standardised diagnoses. However, the findings were replicated in a large multicentre trial showing that the quality of life in patients with standardised diagnoses of PDs can be compared with the quality of life in patients with chronic diseases such as rheumatic disease, Parkinson disease, or even lung cancer. The burden of having a PD was found to be even higher than in patients with type II diabetes and HIV-infected patients (Soeteman, Verheul, & Busschbach, 2008). These results are also in line with other studies showing that patients with PD have a low global level of functioning (Abrams, Alexopoulos, Spielman, Klausner, & Kakuma, 2001; Hueston, Mainous, & Schilling, 1996; Nakao, Gunderson, & Phillips, 1992; Skodol et al., 2002). In one of these studies, Skodol et al. (2002) compared psychosocial functioning in patients with schizotypal, borderline, avoidant, and obsessive compulsive PD to that of patients with mood disorder who
have a global level of functioning comparable to patients with chronic diseases such as diabetes or arthritis (Hays, Wells, Sherbourne, Rogers, & Spritzer, 1995; Wells et al., 1989). The results indicated that patients with schizotypal or borderline PD had even lower psychosocial functioning than patients with mood disorder. In a general psychiatric population, Nakao et al. (1992) reported a strong association between the number of criteria from DSM-IV Axis II (American Psychiatric Association, 1994) and the degree of functional impairment ($r = .60, p < .01$). Moreover, Verheul et al. (2000) showed that the relation between personality pathology and global functioning was not (fully) accounted for by Axis I comorbidity. Although these studies used intermediate outcomes (such as psychological variables instead of quality of life) to measure the burden of disease, they suggest that PDs are indeed specifically associated with a high burden of disease and, thus, a high necessity of treatment.

**Gaps in the evidence—the missing research**
The third obstacle is the existence of gaps in the integrated evidence of effectiveness, cost-effectiveness, and necessity.

*Increasing the quality standards of psychotherapy research*
The favourable results presented by the reviews and meta-analyses so far did not improve the deployment of psychotherapy in PDs and its reimbursement. One explanation is that the reviews discussed do not live up to modern scientific standards, most notably Cochrane reviews. Binks et al. (2006) recently conducted such a Cochrane review and found preliminary but encouraging results for psychotherapy in PDs. It should be noted that most Cochrane reviews are extremely critical toward accepted standards in medicine, as they rely heavily on high-quality randomised trials, which are still scarce in long-term psychotherapy. Nevertheless, by introducing Cochrane standards in the treatment of PDs, the authors set an important trend. The investigation of Binks should be considered a sign of a maturing science. Such maturation will strengthen the empirical base of psychotherapy and will enhance its chances in guideline discussions and reimbursement policy. It would be naive to assume that psychotherapy will ever be fully deployed if the scientific and clinical community does not adopt modern scientific standards. To keep the field of psychotherapy in line with the rest of the medical world, high-quality effectiveness studies, covering the broad spectrum of PDs, should be introduced urgently.
Chapter 2

From efficacy to effectiveness

When reporting results on the effectiveness of psychotherapy, a critical remark has to be made concerning the distinction between efficacy and effectiveness (Haynes, 1999). Many studies pretending to prove effectiveness of psychotherapy were conducted with highly selected patients and treatments in academic treatment settings and well trained and supervised therapists, making a generalisation of the results to the general patient population difficult. In fact, these studies are efficacy studies, answering the question: Does this treatment work in a well-controlled environment? However, in health care policy, the most important question to be answered is: Does this treatment work in everyday practice (Wells, 1999)? This question can be answered by true effectiveness studies, investigating the effect of interventions done by ordinary practitioners, without extensive training and supervision, given to ordinary patients usually seen in clinical practice (Emmelkamp, 2007; Emmelkamp, et al., 2006). In the aforementioned meta-analyses, for example, this distinction was not clearly made. While both efficacy and effectiveness studies are important to strengthen the status of psychotherapy for patients with PD, there is a clear need of well-designed effectiveness studies to demonstrate the value of psychotherapy in regular clinical practice.

Dose-effect relations

In effectiveness research, much effort is put into proving the superiority of one theoretical orientation over another. Despite all the effort and enthusiasm involved, most of the time little difference has been found between psychotherapies from different theoretical orientations (e.g., Svartberg, et al., 2004). Typically, these results stem from research in which treatment dosage (number of sessions or days of treatment) was kept constant. But dosage in fact matters. Several researchers found a positive relation between treatment duration (number of therapy sessions) and health improvements or recovery of personality pathology (Høgland, 1993; Howard, Kopta, Krause, & Orlinsky, 1986; Perry, et al., 1999). This was confirmed by a randomised trial on the effectiveness of day hospital treatment of borderline patients. Treatment results take time. A clear reduction of symptoms and maladaptive behaviour, as well as improved social functioning, did not appear before six months of treatment (Bateman & Fonagy, 1999). A significant reduction of care requirement only appeared after 12 months of treatment and the improvements grew over the course of an 18-month follow-up care period (Bateman & Fonagy, 2001). These findings suggest that more progress will be found in dose-effect studies than in comparing rivalling theoretical orientations.
Sophisticated evidence about dose-effect relations in psychotherapy will give psychotherapists the evidence they need to counter the modern trend of short therapies. Of course new evidence has to be firm and has to include cost-effectiveness research. High doses (and thus high costs) are not necessarily a problem if a high dosage has a stronger effect. An example of this approach is the study by Beecham et al. (2006), showing a clear advantage of a step-down treatment program compared to a fixed long-term inpatient stay. Nevertheless, there is still much to be learned about dose-effect relations, especially in inpatient and day hospital settings. This is important because sound dose-effect data might serve as a powerful argument to endorse psychotherapy for PDs.

**Effective ingredients of psychotherapy**

There are many forms of therapy, all with their merits. Hence, the question is: What makes each therapy work? This quest is comparable with the search for the active substance in pharmacology. Despite the success of the search for the active substance in our neighbouring field of science, this knowledge gap still exists in psychotherapy. One exception is the relationship between therapist and patient, which is generally considered an established, major determinant of the effect of psychotherapy (Martin, Garske, & Davis, 2000). But next to relationship factors, there are other important ingredients of psychotherapy that might be crucial. That is why researchers get more and more interested in therapy factors such as degree of structure and clear focus of treatment, coherence of therapeutic framework, and integration with other patient services (Bateman & Fonagy, 2000), as well as global principles of change (Castonguay & Beutler, 2006). If it would indeed be possible to identify the active substances in psychotherapy, it might be possible to isolate them from other — possibly expensive — components of therapy. As such, the search for the effective factors in psychotherapy represents an effort to increase both effectiveness and cost-effectiveness.

**Formal cost-effectiveness research**

Formal cost-effectiveness research explicitly studies the relation between costs and effects. High costs of a treatment are not necessarily a problem as long as the effects are substantial. In a literature search, we identified promising book titles, such as “Efficacy and Cost-Effectiveness of Psychotherapy” (Spiegel, 1999) and “Cost-Effectiveness of Psychotherapy” (Miller & Magruder, 1999). However, these studies presented cost data but failed to establish a meaningful relation between cost parameters and effects. It would be more justified to classify these studies as cost studies or cost-minimisation studies.
Recently, an extensive report was published in which the available evidence on cost-effectiveness of psychological therapies for borderline PD was reviewed systematically (Brazier et al., 2006). The review team did an excellent job in performing separate economic evaluations for the six randomised controlled trials identified in their review of published studies. Cost-effectiveness was assessed in terms of costs per avoided parasuicide event in all six trials, and costs per Quality Adjusted Life Year (QUALY; Drummond, Sculpher, Torrance, O’Brien, & Stoddart, 2005) in four of the six trials. The outcome could not provide a convincing conclusion, owing to the poor quality of the original studies, a mixture of methods to assess outcome, and a doubtful generalisability. Nevertheless, the results suggested that such interventions have the potential to be cost-effective. The authors used the results of their study to stress the need for high-quality cost-effectiveness research in which a meaningful cost-effectiveness ratio can be calculated, preferably in general terms such as costs per QALY.

In addition, future cost-effectiveness research has to include both direct and indirect costs caused by the illness and saved by certain treatments. This is especially true for work-related costs caused by educational delay, absenteeism, and presenteeism, the latter describing the behaviour of people who, despite serious complaints and ill health, still turn up at their jobs (Aronsson, Gustafsson, & Dallner, 2000). If formal cost-effectiveness studies of psychotherapy indeed show results comparable to already reimbursed treatments for somatic disorders, psychotherapists would have a strong and formal argument to plea for reimbursement of their therapy.

Proof of the necessity of treatment
We argued that generic instruments measuring quality of life, such as the EuroQol EQ-5D (Brooks, et al., 2003) are good choices as these instruments can compare the suffering of patients with PD with the suffering of patients with well-known (somatic) illnesses. The findings of Soeteman et al. (2005) must be seen as a first effort to contradict the persistent belief that psychotherapy patients experience a low burden of disease. We argue that psychotherapists should challenge the YAVIS belief in a convincing empirical way, otherwise the YAVIS sentiment will jeopardise any claim for reimbursement of psychotherapy in patients with PD. Additional evidence using quality of life assessments and standardised diagnoses are needed to provide the decisive evidence.
Conclusion

Psychotherapy has the potential to develop into an evidence-based field, broadly accepted and widely reimbursed health discipline if innovative and comprehensive research on effectiveness, cost-effectiveness, and necessity of treatment is initiated. By working closely together, research and practice can provide efficient and equitable mental health care for patients in need.
Chapter 3

The use of propensity score methods in psychotherapy research. A practical application

The first randomised study in medicine was conducted by Amberson in 1931 by flipping a coin (Amberson, McMahon, & Pinner, 1931). Now, randomised controlled trials are considered the gold standard for comparing the effectiveness of psychotherapeutic treatment methods. Randomisation assumes that all known and unknown characteristics of the participants are balanced between the experimental groups, except for the treatment condition. With randomisation, treatment effects can theoretically be estimated by merely subtracting the mean responses of the treatment groups (Rubin, 1997). In many cases, though, randomisation may be difficult, unethical, or impossible (Black, 1996), especially in psychotherapy research (Castonguay & Beutler, 2006; de Maat, Dekker, Schoevers, & de Jonghe, 2007; Leichsenring, 2004; Westen, Novotny, & Thompson-Brenner, 2004). Here patients’ and clinicians’ personal preferences regarding treatment allocation may work against randomisation. The resulting high number of excluded subjects makes the generalisation of such results difficult (Brewin & Bradley, 1989). Hence, research on treatment effects in various (para)medical fields often requires well-designed and carefully conducted non-randomised studies (e.g., Chiesa & Fonagy, 2007; Forstmeier & Rueddel, 2007). Shadish et al. (2002) called these studies quasi-experimental, based on their resemblance to true experiments, except for the random assignment of participants to treatments. In these quasi-experimental designs, the researcher has some influence on the manipulation of treatment and measurement. This is in contrast to pure observational studies, where the size and direction of a relationship among variables are simply observed (Shadish, et al., 2002). In case of non-random allocation to treatment, persons with different treatments can differ on pre-treatment characteristics. This selection bias affects the estimates of the treatment effect. Rosenbaum (2002) distinguishes two types of bias: (a) hidden bias, due to unobserved differences in baseline characteristics; and (b) overt bias, due to observed differences in baseline characteristics. Hidden bias is the most difficult to deal with. Overt bias can be corrected with various statistical methods, by incorporating known initial differences into the statistical analysis. The most widely used methods are matching, stratification, and regression adjustment (Frangakis & Rubin, 2002; Rosenbaum & Rubin, 1984; Rubin & Thomas, 1996). In matching, each individual in the treatment group is paired with the most similar individual in the control group. After matching, the groups as a whole are assumed to be as similar as possible on the matched characteristics. In stratification, subgroups of patients are formed based on baseline variables. In psychotherapy research, however, there is usually a large number of variables to match or stratify on, making it almost impossible to find patients or groups similar on all these variables. This is called the “dimensionality problem”. Regression analysis with covariates, a third tool
to compensate for overt bias, has limitations as well: When many pre-treatment variables are used as covariates, statistical-modelling problems and a loss of power arise. A promising alternative method to correct for overt bias is the propensity score method (Rosenbaum, 2002; Rosenbaum & Rubin, 1983).

**Propensity score**

Rosenbaum and Rubin (1983) suggested using the propensity score method to reduce the dimensionality problem. The propensity score method reduces the entire collection of observed pre-treatment variables (X) to a single score. The estimated propensity score is defined as the conditional probability of assignment to a particular treatment, given a set of observed pre-treatment characteristics. Let Z denote treatment group membership, where Z = 0 denotes the control condition and Z = 1 denotes the treatment condition. Then, propensity score is defined as:

\[ PS = P(Z = 1 \mid X) \]

Rosenbaum and Rubin (1983) proved that, given the value of the propensity score, assignment to treatment no longer depends on baseline variables. The propensity score is a score balancing all observed pre-treatment variables among patients with the same value of the propensity score. In this way, the propensity score method can put overt bias under statistical control. Different from the conventional approach, i.e., controlling for or matching on many baseline variables, the propensity score enables researchers to deal with one single, composite variable which is much easier and, in regression analysis, preserves power. The propensity score has so far been used in medicine (e.g., Chan et al., 2002; Connors et al., 1996; Lieberman et al., 1996; Lytle et al., 1999; Mehta, Pascual, Soroko, & Chertow, 2002; Potosky et al., 2000; Stenestrand & Wallentin, 2001; Wolfe & Michaud, 2004), social sciences (e.g., Gibson, 2003; Guo, Barth, & Gibbons, 2006; Leow, Marcus, Zanutto, & Boruch, 2004; Yoshikawa, Magnuson, Bos, & Hsueh, 2003), and economics (e.g., Dranove & Lindrooth, 2003; Jalan & Ravallion, 2003; Lechner, 1999). The United States Food and Drug Administration recommended the propensity score as a tool to overcome selection bias in treatment studies (Jung, Chow, & Chi, 2007). In psychotherapy research, however, the propensity score is not widely known. To the best of our knowledge, only a handful of pioneering studies have used this instrument for selection bias control in non-randomised studies (Golkaramnay, Bauer, Haug, Wolf, & Kordy, 2007; Hill, Waldfogel, Brooks-Gunn, & Han, 2005; Kächele, Kordy, & Richard, 2001; Robinson, Harper, & Schoeny, 2003).
Aim

The aims of this paper are (a) to investigate if the propensity score method is applicable in psychotherapy research, and (b) to outline a step-by-step protocol for the psychotherapy researcher to facilitate use of the propensity score in comparative outcome studies when randomisation is unfeasible. We applied the propensity score method to a case study, the research project SCEPTR ("Study on Cost-Effectiveness of Personality Disorder TREATment"); Viersprong Institute for Studies on Personality Disorders, 2010). We compared two treatment groups from SCEPTR, using the propensity score to correct for known baseline differences. The two treatment groups selected for comparison are short versus long psychotherapy duration, as this distinction is straightforward and simple to understand. Results should only be interpreted as an illustration, not as a relevant clinical message. All statistical techniques presented in this paper are easily done in common statistical packages, such as SPSS.

Method

Participants
Patients were recruited from six mental health care centres in the Netherlands offering outpatient, day hospital and/or inpatient psychotherapy for patients with personality pathology. Out of 2,540 patients who were admitted to the centres from March 2003 to March 2006, 1,047 were selected for treatment, i.e., short- or long-duration psychotherapy in various settings. Before treatment allocation, all patients were assessed with a routinely distributed assessment battery including self-report questionnaires. Of the 1,047 patients selected for treatment, 298 patients had not yet completed a follow-up measure, so no outcome score could be calculated. These were excluded from the analyses, leaving 749 patients. Of these, 507 (67.7%) were female. The mean age was 34.24 years (SD = 9.93). We divided this sample into two groups: one group allocated to short-term therapy (up to six months), the other group allocated to long-term therapy (more than six months).

Measures
The baseline assessment measured a long list of social, economic, and diagnostic variables carefully selected by both clinicians and researchers, based on literature and clinical knowledge (see Tables 3.1 and 3.2).
Table 3.1. Difference scores (unstandardised $B$) in continuous variables between short-term and long-term treatment group

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>Unstandardised $B$ treatment duration (short/long)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short ($n = 331$)</td>
<td>Long ($n = 328$)</td>
</tr>
<tr>
<td>Age</td>
<td>36.83 (9.63)</td>
<td>31.86 (9.49)</td>
</tr>
<tr>
<td>Personality pathology (DAPP-BQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional dysregulation</td>
<td>21.93 (4.02)</td>
<td>22.87 (3.66)</td>
</tr>
<tr>
<td>Dissocial behaviour</td>
<td>17.35 (4.10)</td>
<td>18.01 (4.40)</td>
</tr>
<tr>
<td>Inhibitedness</td>
<td>22.11 (5.06)</td>
<td>22.51 (4.97)</td>
</tr>
<tr>
<td>Compulsivity</td>
<td>24.29 (6.84)</td>
<td>23.87 (7.29)</td>
</tr>
<tr>
<td>Motivation (MTQ-8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need for help</td>
<td>28.87 (5.23)</td>
<td>28.46 (5.22)</td>
</tr>
<tr>
<td>Readiness to change</td>
<td>30.70 (5.04)</td>
<td>29.96 (5.16)</td>
</tr>
<tr>
<td>Quality of life (EQ-5D)</td>
<td>0.59 (0.26)</td>
<td>0.55 (0.26)</td>
</tr>
<tr>
<td>Psychological capacities (SIPP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-control</td>
<td>4.65 (0.91)</td>
<td>4.48 (0.90)</td>
</tr>
<tr>
<td>Social concordance</td>
<td>5.72 (0.78)</td>
<td>5.63 (0.81)</td>
</tr>
<tr>
<td>Identity integration</td>
<td>3.54 (0.71)</td>
<td>3.38 (0.65)</td>
</tr>
<tr>
<td>Relational capacities</td>
<td>3.97 (0.84)</td>
<td>3.79 (0.78)</td>
</tr>
<tr>
<td>Responsibility</td>
<td>4.67 (0.84)</td>
<td>4.52 (0.88)</td>
</tr>
<tr>
<td>Psychiatric symptomatology (GSI)</td>
<td>2.39 (0.62)</td>
<td>2.55 (0.65)</td>
</tr>
<tr>
<td>Functioning (OQ-45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpersonal functioning</td>
<td>20.07 (6.29)</td>
<td>21.60 (6.01)</td>
</tr>
<tr>
<td>Social role functioning</td>
<td>15.28 (4.86)</td>
<td>15.59 (4.58)</td>
</tr>
</tbody>
</table>
Table 3.1. (continued) Difference scores (unstandardised B) in continuous variables between short-term and long-term treatment group

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean (SD)</th>
<th>Unstandardised B treatment duration (short/long) Before PS correction</th>
<th>After PS correction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short $(n = 331)$</td>
<td>Long $(n = 328)$</td>
<td></td>
</tr>
<tr>
<td>Axis II diagnosis (SIDP-IV)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Axis II cluster A disorders</td>
<td>0.04 (0.19)</td>
<td>0.09 (0.29)</td>
<td>0.05*</td>
</tr>
<tr>
<td>Number of Axis II cluster B disorders</td>
<td>0.19 (0.48)</td>
<td>0.34 (0.58)</td>
<td>0.15***</td>
</tr>
<tr>
<td>Number of Axis II cluster C disorders</td>
<td>0.65 (0.78)</td>
<td>0.70 (0.79)</td>
<td>0.05</td>
</tr>
<tr>
<td>Duration of psychological problems</td>
<td>3.59 (0.81)</td>
<td>3.59 (0.79)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*p < .05, **p < .01, ***p < .001

Psychiatric symptomatology was measured with the Symptom Checklist 90-Revised, Dutch version (SCL-90-R; Arrindell & Ettema, 2003; Derogatis, 1992). In this study, we used the mean score of the 53 items of the Brief Symptom Inventory, i.e., the Global Severity Index (GSI), ranging from 0 to 4. To measure the type and degree of personality pathology we used the four higher-order factors of the Dimensional Assessment of Personality Pathology Basic Questionnaire, Dutch version (DAPP-BQ; Livesley & Jackson, 2002; van Kampen, 2002): (a) emotional dysregulation, (b) dissocial behaviour, (c) inhibition, and (d) compulsivity. Psychosocial functioning was measured with the Outcome Questionnaire 45, Dutch version (OQ-45; Lambert et al., 1996). Of this self-report measure, we used two subscales: interpersonal relations and social role functioning. Health-related quality of life was assessed with the Dutch version of the EuroQoL EQ-5D (EQ-5D; Brooks, Rabin, & de Charro, 2003). PDs were assessed with the Structured Interview of DSM-IV Personality, Dutch version (SIDP-IV; de Jong, Derks, van Oel, & Rinne, 1996; Pfohl, Blum, & Zimmerman, 1997). The severity of personality pathology was measured with five higher-order domains of the Severity Indices of Personality Problems, Dutch version (SIPP; Verheul et al., 2008): (a) self-control, (b) social concordance, (c) identity integration, (d) relational capacities, and (e) responsibility. To measure patients’ motivation for treatment, we used the two scales of the Motivation for Treatment Questionnaire (MTQ-8; van Beek & Verheul, 2008): need for help and readiness to change.
Table 3.2. Differences in categorical variables between short-term and long-term treatment group

<table>
<thead>
<tr>
<th>Variable</th>
<th>%</th>
<th>Odds Ratio (short/long)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short</td>
<td>Long</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>65.3</td>
<td>68.9</td>
</tr>
<tr>
<td>Male</td>
<td>34.7</td>
<td>31.1</td>
</tr>
<tr>
<td>Civil status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>27.5</td>
<td>18.0</td>
</tr>
<tr>
<td>Widowed or divorced</td>
<td>13.3</td>
<td>10.1</td>
</tr>
<tr>
<td>Never married</td>
<td>59.2</td>
<td>72.0</td>
</tr>
<tr>
<td>Living situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>39.0</td>
<td>38.4</td>
</tr>
<tr>
<td>With partner (with/without child)</td>
<td>44.4</td>
<td>29.3</td>
</tr>
<tr>
<td>With child without partner</td>
<td>5.7</td>
<td>6.4</td>
</tr>
<tr>
<td>With parent(s)</td>
<td>4.2</td>
<td>17.7</td>
</tr>
<tr>
<td>With other people</td>
<td>6.6</td>
<td>8.2</td>
</tr>
<tr>
<td>Childcare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No care for children</td>
<td>72.5</td>
<td>80.5</td>
</tr>
<tr>
<td>Care for children</td>
<td>27.5</td>
<td>19.5</td>
</tr>
<tr>
<td>Work situation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>33.2</td>
<td>36.3</td>
</tr>
<tr>
<td>Study or paid work</td>
<td>66.8</td>
<td>63.7</td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>19.3</td>
<td>28.0</td>
</tr>
<tr>
<td>Middle</td>
<td>22.7</td>
<td>17.7</td>
</tr>
<tr>
<td>High</td>
<td>58.0</td>
<td>54.3</td>
</tr>
<tr>
<td>Previous outpatient treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>17.2</td>
<td>22.6</td>
</tr>
<tr>
<td>Yes</td>
<td>82.8</td>
<td>77.4</td>
</tr>
<tr>
<td>Previous inpatient treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>83.4</td>
<td>79.9</td>
</tr>
<tr>
<td>Yes</td>
<td>16.6</td>
<td>20.1</td>
</tr>
<tr>
<td>Previous medication treatment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>53.8</td>
<td>52.7</td>
</tr>
<tr>
<td>Yes</td>
<td>46.2</td>
<td>47.3</td>
</tr>
</tbody>
</table>
Table 3.2. (continued) Differences in categorical variables between short-term and long-term treatment group

<table>
<thead>
<tr>
<th>Variable</th>
<th>%</th>
<th>Odds Ratio (short/long)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short (n = 331)</td>
<td>Long (n = 328)</td>
</tr>
<tr>
<td>Alcohol abuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>84.5</td>
<td>87.2</td>
</tr>
<tr>
<td>Yes</td>
<td>15.5</td>
<td>12.8</td>
</tr>
<tr>
<td>Drug abuse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>86.1</td>
<td>77.4</td>
</tr>
<tr>
<td>Yes</td>
<td>13.9</td>
<td>22.6</td>
</tr>
<tr>
<td>Preference for treatment setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>12.1</td>
<td>22.9</td>
</tr>
<tr>
<td>Day hospital</td>
<td>30.9</td>
<td>24.8</td>
</tr>
<tr>
<td>Inpatient</td>
<td>35.5</td>
<td>29.4</td>
</tr>
<tr>
<td>Do not know</td>
<td>21.5</td>
<td>22.9</td>
</tr>
<tr>
<td>Preference for treatment duration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Up to six months</td>
<td>43.5</td>
<td>25.3</td>
</tr>
<tr>
<td>Longer than six months</td>
<td>26.9</td>
<td>37.2</td>
</tr>
<tr>
<td>Do not know</td>
<td>29.6</td>
<td>37.5</td>
</tr>
<tr>
<td>Treatment setting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient</td>
<td>18.7</td>
<td>34.1</td>
</tr>
<tr>
<td>Day hospital</td>
<td>31.7</td>
<td>30.2</td>
</tr>
<tr>
<td>Inpatient</td>
<td>49.5</td>
<td>35.7</td>
</tr>
</tbody>
</table>

\(^a\) Category is reference category; for regression purposes all categorical variables were translated into dummy-variables, whereby the first category always serves as reference category with an odds ratio of 1.00

\(^*p < .05, \,**p < .01, \,***p < .001\)

Results

To avoid bias in the estimation of the treatment effect, we corrected for the influence of known pre-treatment differences. We did this by stratification of the sample based on the propensity score. This process took nine steps, described below.

Stratification

Step 1: Effect estimation before correction

Before correction for known pre-treatment differences, we estimated the treatment
effect by conducting a linear regression analysis. In this “naïve” estimate the only independent variable was ‘group membership’ (short vs. long), the dependent variable was outcome, being defined here as the level of psychiatric symptomatology (GSI) at the first measurement following baseline. The uncorrected treatment effect $B$ was $0.20 (SE = 0.05; p < .001)$.

**Step 2: Balance check before correction**

We compared the two treatment groups on pre-treatment variables before stratification. Note that this step is neither relevant for variable selection for the propensity score, nor for further analyses. It is only important here to be able to demonstrate the influence of propensity correction on the balance between groups. This demonstration can be done in several ways. For illustration purposes, we chose to show a comparison of overall regression coefficients. We conducted a number of regression analyses with ‘group membership’ as an independent variable, and pre-treatment characteristics as dependent variables (linear regression analyses for continuous variables, see Table 3.1; and multinomial logistic regression analyses for categorical variables, see Table 3.2). The two patient groups (short- vs. long-term treatment) differed significantly on 19 of the 34 baseline variables. This implies that, without correction for these baseline differences, the two groups were not readily comparable—a problem that may be dealt with using the propensity score.

**Step 3: Variable selection for propensity score estimation**

To estimate the propensity score, we used all baseline variables related to outcome (GSI). To identify related variables, we conducted a number of linear regression analyses with the GSI as the dependent variable, and each potential confounder as an independent variable. The following variables emerged as primary candidates for the estimation of the propensity score: level of personality pathology (i.e., emotional dysregulation, dissocial behaviour, and inhibitedness), motivation for treatment (i.e., need for help), quality of life, psychological capacities (i.e., self-control, social concordance, identity integration, relational capacities, and responsibility), level of psychiatric symptomatology, functioning (i.e., interpersonal and social role functioning), number of cluster A, B and C PDs, working situation, level of education, previous inpatient treatment, and patient preferences for treatment duration and setting of treatment. Sociodemographic variables were added to the propensity score model as well, because they are considered highly relevant in psychotherapy research: age, gender, marital status, living situation, and responsibility for the care of children.
Step 4: Exclusion of incomplete cases
In this example, only patients with no missing values on the selected potential confounders (see “Step 3”) were included in the propensity score analysis. The final sample therefore consisted of 659 patients. Alternatively, imputation techniques might be used to fill in the missing values in estimation variables.

Step 5: Propensity score estimation
The propensity score was estimated in a logistic regression analysis. All selected potential confounders were used as independent variables, and ‘group membership’ as the dependent variable. One can estimate and save these probabilities for each subject, for example by using the option “save predicted probability” in SPSS.

Step 6: Inspection of overlap and exclusion of non-overlapping cases
For the short-term treatment group \((n = 331)\), the propensity score ranged between .03 and .98; for the long-term treatment group \((n = 328)\), the propensity score ranged between .10 and .99 (see Figure 3.1). The propensity score range that both groups cover is between .10 and .98. Patients with a propensity score outside this common range \((n = 24)\) were excluded from the stratification, leaving a sample of 635 patients.

Figure 3.1. Overlap of propensity scores in the two treatment groups (short/long)
Step 7: Stratification of the sample based on the propensity score
The sample of 635 patients was divided into five equal subgroups with similar propensity score (so-called “strata”; Cochran, 1968), which can be seen in Table 3.3. We then created four dummy variables based on these five groups.

Table 3.3. Distribution of patients across the five strata

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Short</th>
<th>Long</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>104</td>
<td>23</td>
<td>127</td>
</tr>
<tr>
<td>2</td>
<td>78</td>
<td>49</td>
<td>127</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>65</td>
<td>127</td>
</tr>
<tr>
<td>4</td>
<td>48</td>
<td>79</td>
<td>127</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>110</td>
<td>127</td>
</tr>
<tr>
<td>Total</td>
<td>309</td>
<td>326</td>
<td>635</td>
</tr>
</tbody>
</table>

Step 8: Balance check after correction
We needed to know if the stratification of the sample based on the propensity score resulted in a balance of pre-treatment variables between the two treatment groups. Therefore, we checked again for differences in pre-treatment variables. This might be done for instance by comparing groups per stratum, but to keep in line with the illustrative analyses of step 2, we calculated the corrected differences between treatment groups by performing a number of regression analyses: this time with ‘group membership’ and the four dummy variables indicating stratum membership as independent variables, and pre-treatment characteristics as dependent variables. The regression coefficients in Tables 3.1 and 3.2 (with stratum membership as covariate) indicate that—on average across all strata—there were no longer significant differences in pre-treatment variables. Our estimated propensity score seemed to balance, in a satisfactory way, the observed significant pre-treatment differences between the short-term and the long-term group. In case differences in pre-treatment variables between groups are more persistent, one can try to re-estimate the propensity score, for instance by including interaction terms or non-linear relationships and restart at step 5.
Step 9: Effect estimation after correction
After taking into account the influence of known pre-treatment characteristics using the propensity score, a corrected estimate of the treatment effect can be calculated. This can be done in different statistical ways, for instance by weighting the five treatment effects of the different strata. To keep in line with our analysis in step 1, we used a linear regression analysis with the GSI as the dependent variable, but this time ‘group membership’ and the four dummy variables indicating stratum membership were the independent variables. The effect of the treatment group on outcome was reduced from $B = 0.20$ ($SE = 0.05; p < .001$) before propensity score correction to $B = 0.15$ ($SE = 0.06; p < .05$) after propensity score correction. This shows that, when observed pre-treatment differences were not taken into account, the treatment effect was overestimated. Stratification of the sample based on the propensity score reduced this bias.

Alternatives to stratification: Propensity score in regression analysis and matching
We present the results of two alternative methods for adjusting a treatment effect estimation using the propensity score.

Regression analysis
We performed a linear regression analysis with the GSI as the dependent variable, and the propensity score (as a continuous covariate) and the variable ‘treatment group’ as independent variables. After controlling for the propensity score by including it as a covariate in the regression analysis, the effect of treatment group membership was reduced from $B = 0.20$ ($SE = 0.05; p < .001$) before the correction to $B = 0.14$ ($SE = 0.06; p < .05$) after the propensity score correction. This is similar to the result of adjustment by stratification.

Matching
We matched each subject from the long-term group (this was the smallest group) with a subject from the short-term group, based on nearest available propensity score. Each subject from the short-term group only served once as matching partner for a subject from the long-term group (“sampling without replacement”). To ensure similarity in the matched pairs we used “caliper matching” (Quade, 1982), i.e., all pairs with a propensity score difference larger than .10 were removed from the analysis. This meant only 179 matched pairs (358 individuals) remained in the analysis. After matching, the two groups showed no difference on any of the observed pre-treatment variables. To keep in line with our previous analyses, a regression analysis was conducted in the matched sample, with the GSI as the dependent
variable, and the variable ‘group membership’ as the independent variable. The effect of treatment group membership was reduced from $B = 0.20$ ($SE = 0.05; p < .001$) before matching to $B = 0.15$ ($SE = 0.07; p < .05$) after matching (alternatively, a paired $t$-test might be conducted in the matched sample). Though our matching procedure was successful in balancing and correcting for observed pre-treatment differences, we lost a substantial amount of information due to a reduced sample size. In other (bigger) samples, matching might still be a useful strategy to correct for overt bias, especially when the control pool is large.

Discussion

Randomisation in general and its application in psychotherapy research have been criticised by different authors for various reasons. Non-randomised studies, however, face the serious problem of selection bias. As a result, a need is felt for alternative and complementary research designs in the field of psychotherapy, like quasi-experimental designs. The propensity score method offers a solution to one part of the problem, overt bias, by balancing the treatment groups with regard to observed pre-treatment differences. To overcome selection bias, the propensity score method offers advantages compared to traditional methods. First, the propensity score provides better insight in the selection process. Modelling treatment selection in a logistic regression analysis clarifies which variables affect selection and to what degree. Second, it is easier to match or stratify on a single score (like the propensity score) than on a range of pre-treatment characteristics. The same holds true for regression adjustment techniques. Use of the single score propensity score enhances statistical power, as compared to many covariates in a regression analysis. Third, both the overlap in the distribution of the propensity score and balance of baseline variables after correction can be investigated and used as a descriptive tool (Rosenbaum & Rubin, 1983). The propensity score method, like any statistical correction method for selection bias, is only helpful given a considerable balance of baseline characteristics. After all, comparing very different subject groups in an outcome study is irrelevant, both scientifically and clinically. The propensity score helps to identify subjects differing widely on their pre-treatment characteristics (and, as a consequence, on their propensity score). Determining the (essential) overlap of the distributions and balance with classical covariate regression analysis is cumbersome and therefore probably rarely done. As a last advantage, we would mention that the propensity score method can be applied in different ways (stratification, matching, in a regression analysis). Therefore, it can be tailored to sample characteristics and researchers’ insights and decisions. Obviously, the
Propensity score method

The propensity score method is not without limitations and has to be used responsibly (Yue, 2007). A researcher using the propensity score should take into account the following recommendations. First, the propensity score only corrects for observed pre-treatment characteristics, not for unobserved (unknown) variables, hampering true cause-effect analysis. This is called the “ignorability” or “no unobserved confounders” assumption. Even when using the propensity score carefully, results may still be biased due to unobserved variables. This is why, before starting a study, as many confounders as possible should be identified and measured in a reliable way. This reduces the risk that important variables are overlooked. It is recommended to consult several experts from both the clinical and statistical field to gain insight into the most relevant pre-treatment variables. Experts’ consensus and statistical relevance should guide the choice for potential confounders. Interestingly, when prognostic factors are well understood and controlled for, and inclusion/exclusion criteria are the same, randomised and non-randomised studies can have similar outcomes (Benson & Hartz, 2000; Concato, Shah, & Horwitz, 2000; Mc Kee et al., 1999). Second, be careful when selecting variables to estimate the propensity score. Brookhart et al. (2006) tested several ways of selecting relevant variables in a simulation study. Their findings suggest that all variables related to study outcome should be included in the propensity score model, whether or not these variables influence treatment assignment. In this study, we followed their advice. However, in the field there is still discussion on which is the best method for selecting the variables for the propensity score model (e.g., Austin, Grootendorst, & Anderson, 2007). Third, the sample size of a study has to be sufficiently large, especially for stratification purposes, to allow for a meaningful correction of bias by means of the propensity score. Otherwise, several strata might be populated exclusively by patients with the same treatment condition, making comparison impossible. A high number of missing values on baseline variables causes problems as well. As the propensity score method uses a combination of many variables, just one missing variable leads to a missing propensity score, excluding this patient from all further analysis. Well-chosen imputation methods can be used to fill in missing values and guarantee a sufficient sample size without losing statistical precision. The availability of all essential data is the first condition for a meaningful application of the propensity score method, just as for any other statistical correction method. We conclude that the propensity score method is a powerful way of simultaneously adjusting for many observed confounders in non-randomised studies, thereby most probably reducing bias in treatment comparisons. If used in a responsible and thoughtful way, the propensity score method used in quasi-experimentation offers a strong research design in situations where randomisation is not possible. Therefore, the propensity score method is a promising tool for future psychotherapy research.
Chapter 4

Effectiveness of different dosages of psychotherapeutic treatment for patients with cluster C personality disorders: Results of a large prospective multicentre study

An estimated 2.6% of the general population is affected by cluster C PDs: avoidant, dependent, and obsessive-compulsive PD (Coid, Yang, Tyrer, Roberts, & Ullrich, 2006). This cluster of PDs is associated with significant functional impairment (Grant et al., 2004; Skodol et al., 2002; Skodol, Johnson, Cohen, Sneed, & Crawford, 2007) and a high economic burden (Soeteman, Hakkart-van Roijen, Verheul, & Busschbach, 2008), yet studies investigating treatment effectiveness in this patient population are scarce (Duggan, Huband, Smailagic, Ferriter, & Adams, 2007). As in research on other psychological disorders (e.g., van Emmerik, Kamphuis, & Emmelkamp, 2008), the available studies on cluster C PDs typically compare treatments that are identical in treatment setting and duration. Investigators have compared different outpatient treatments (Alden, 1989; Emmelkamp et al., 2006; Hellerstein et al., 1998; Stravynski, Belisle, Marcouiller, Lavallee, & Elie, 1994; Svartberg, Stiles, & Seltzer, 2004; Winston et al., 1994), different day hospital treatments (Karterud et al., 2003; Wilberg et al., 1999), and different inpatient treatments (Gude & Vaglum, 2001; Teusch, Böhme, Finke, & Gastpar, 2001). One recent study in Norway (Arnevik et al., 2009) compared outpatient and day hospital treatment for patients with all forms of PDs, and found no significant superiority of one treatment over another at eight months after the start of treatment. However, so far, no study has compared the effectiveness of treatments across widely differing settings and durations. In this article, treatment modality was specified as a combination of treatment setting (i.e., outpatient, day hospital, or inpatient) and duration (i.e., short term or long term), as these are the most important aspects regarding treatment costs, a crucial aspect in times of restricted health care budgets.

It is likely that one of the reasons this comparison has not been undertaken previously is the difficulty of random assignment to different treatment modalities in clinical samples due to practical or ethical constraints (Black, 1996). Furthermore, even if researchers were successful in setting up and starting a randomised treatment modality study, its external validity would be doubtful because a high number of patients would refuse to participate (e.g., Zeeck et al., 2009). Therefore, quasi-experimental studies using statistical correction models to counter selection bias are increasingly being found in the literature (e.g., Facchinetti, Ottolini, Fazio, Rigatelli, & Volpe, 2007; Forstmeier & Rueddel, 2007; Golkaramnay, Bauer, Haug, Wolf, & Kordy, 2007; Grossman, Tiefenthaler-Gilmer, Raysz, & Kesper, 2007).

The aim of the present quasi-experimental study was to compare the effectiveness of different treatment modalities for patients with cluster C PDs in a naturalistic setting, thereby insuring high external validity. In fact, treatment modality might be an overlooked factor in psychotherapy effectiveness research.
Method

Participants
Participants (N = 371) were recruited from consecutive admissions to six mental health care centres in the Netherlands (Centre of Psychotherapy De Viersprong, Halsteren; Altrecht, Utrecht; Zaans Medical Centre, Zaandam; Centre of Psychotherapy De Gelderse Roos, Lunteren; GGZWNB, Bergen op Zoom and Roosendaal; Centre of Psychotherapy Mentrum, Amsterdam). These institutions offer outpatient, day hospital, and/or inpatient psychotherapeutic treatment for patients with personality pathology. From March 2003 to March 2006, 1,379 patients completed the intake procedure and were selected for treatment (see Figure 4.1). Of these, 146 patients (10.6%) were excluded from the study for not meeting one of the following inclusion criteria: age between 18 and 70 years (n = 13), significant personality pathology (n = 34), and referral for psychotherapeutic treatment aimed at personality problems (n = 99). Nine patients (0.7%) met one of the following exclusion criteria: insufficient command of the Dutch language (n = 6), organic cerebral impairment (n = 1), mental retardation (n = 1), and schizophrenia (n = 1). This left 1,224 participants, of whom 100 (8.2%) refused to participate. Another 31 patients (2.5%) could not participate due to logistic reasons (e.g., no appointment could be made to provide informed consent), and 133 patients (10.9%) were excluded due to missing or unreliable baseline data. Thirty-eight patients (3.1%) received less than two treatment sessions or less than two days of inpatient or day hospital therapy, and were therefore excluded. The remaining 922 patients were informed about the study and its procedure, provided written informed consent, and entered the study. Of those, 466 patients (50.5%) had one or more cluster C PDs.

In the absence of explicit guidelines for treatment assignment in PDs (Vervaeke & Emmelkamp, 1998), the selection procedure was based on the expert opinion of clinicians who used their clinical experience combined with patient data from standardised instruments. To elucidate the criteria used for the assignment process, our research group recently conducted a study with intake clinicians from the participating treatment centres. We found evidence of substantial consensus among clinicians concerning the criteria used for treatment decision-making. For example, focality of problems (focal or broad spectrum of problems) and ego strength were found to be related to decisions about a short or long treatment duration for a substantial number of intake clinicians (van Manen et al., 2008).
Figure 4.1. Patient flow

Refused to participate  
$n = 100$

Excluded due to other reasons  
$n = 164$
(Logistic reasons  $n = 31$
Missing/unreliable data  $n = 133$)

Assessed for eligibility  
$N = 1379$

Did not meet inclusion criteria  
$n = 146$
Met exclusion criteria  
$n = 9$

Enrolled  
$n = 960$

No DSM-IV cluster C diagnosis  
$n = 456$

Dropped out prematurely*  
$n = 38$

Allocated  
$n = 466$

Short outpatient treatment  
$n = 18$

Long outpatient treatment  
$n = 86$

Short day hospital treatment  
$n = 85$

Long day hospital treatment  
$n = 103$

Short inpatient treatment  
$n = 63$

Long inpatient treatment  
$n = 101$

Not included  
$n = 18$

No follow-up  
$n = 28$

No follow-up  
$n = 8$

No follow-up  
$n = 29$

No follow-up  
$n = 4$

No follow-up  
$n = 8$

Analyzed  
$n = 68$

Analyzed  
$n = 77$

Analyzed  
$n = 74$

Analyzed  
$n = 59$

Analyzed  
$n = 93$

* These patients did not receive a “minimal effective dose” of treatment, defined as two sessions for outpatients and two treatment days for day hospital patients or inpatients, and were therefore not included in the initial study sample.
Patients were assigned to one of six treatment modality groups: 18 to short-term outpatient (up to six months), 96 to long-term outpatient (more than six months), 85 to short-term day hospital, 103 to long-term day hospital, 63 to short-term inpatient, and 101 to long-term inpatient treatment. The short-term outpatient group was excluded from the analyses for two reasons: (a) only a minority of patients (3.9%) were assigned to this short and low-frequent treatment modality, as could be expected in a PD patient population; and (b) these patients differed significantly from patients in the other treatment groups on a high number of pre-treatment variables, indicating a dissimilar and — most importantly — a structurally less “sick” patient population, incomparable with the rest of the sample. A comparison with this treatment modality would most probably also fail when trying to design a randomised trial, as short-term outpatient therapy differs most from all other modalities in terms of its relatively low impact on patients’ lives compared to other treatment modalities. In the end, 448 participants were included in the study. Follow-up data were not available for 77 patients (17.2%; patients who did not respond to any follow-up assessment or patients where follow-up measurements were not yet available). There was no difference in psychiatric symptoms at baseline between patients with follow-up data and those without (this holds true for both the comparison in the total sample and the comparisons within the five treatment groups). The final sample consisted of 371 patients to be included in the analyses.

Treatment
The six mental health care centres offered a variety of psychotherapeutic treatments tailored to a PD patient population. Their treatments differed according to several features. As this study focused on different treatment modalities in terms of setting and duration, the following five treatment groups were compared:

- Patients in long-term outpatient treatment ($n = 68, 18.3\%$ of the study sample). These patients came for individual ($76.5\%$) or group ($23.5\%$) psychotherapy sessions, for up to two sessions per week ($\text{mean} = 0.8$ sessions/week, $SD = 0.51$, median = 0.5) for more than six months ($\text{mean duration} = 15.4$ months, $SD = 6.36$, median = 12.0).
- Patients in short-term day hospital treatment ($n = 77, 20.8\%$ of the study sample). These patients came to the institutions at least one morning/afternoon per week ($\text{mean} = 3.2$ days/week, $SD = 1.51$, median = 3.0) for up to six months ($\text{mean duration} = 5.4$ months, $SD = 1.32$, median = 6.0) and received different forms of psychotherapeutic and psychosocial treatment, but slept at home.
Patients in long-term day hospital treatment \((n = 74, 19.9\% \text{ of the study sample})\). These patients came to the institutions at least one morning/afternoon per week (mean = 3.3 days/week, \(SD = 1.42\), median = 3.0) for more than six months (mean duration = 12.1 months, \(SD = 2.41\), median = 12.0) and received different forms of psychotherapeutic and psychosocial treatment, but slept at home.

Patients in short-term inpatient treatment \((n = 59, 15.9\% \text{ of the study sample})\). These patients stayed at the institutions five days a week for up to six months (mean duration = 4.2 months, \(SD = 1.48\), median = 3.0) and received different forms of psychotherapeutic and psychosocial treatment.

Patients in long-term inpatient treatment \((n = 93, 25.1\% \text{ of the study sample})\). These patients stayed at the institutions five days a week for more than six months (mean duration = 10.2 months, \(SD = 1.98\), median = 10.0) and received different forms of psychotherapeutic and psychosocial treatment.

Day hospital and inpatient programs typically consisted of group psychotherapy as a core element, mostly in combination with one or more non-verbal or expressive group therapies, individual psychotherapy, sociotherapy within the therapeutic community, coaching for social problems, community meetings, and/or pharmacological treatment. The psychotherapists were all licensed psychiatrists or psychologists. On average, they had 14.9 years \((SD = 10.1)\) of postgraduate clinical experience. The treatments under study can be considered highly representative of regular clinical practice in the Netherlands, as therapists did not receive specific training for this study and treatment integrity was not monitored.

**Assessments**

*Baseline measures*

An extensive standard assessment battery of instruments was administered to the patients before treatment assignment. PDs were measured using the Dutch version of the Structured Interview for DSM-IV Personality (SIDP-IV; de Jong, Derks, van Oel, & Rinne, 1996; Pfohl, Blum, & Zimmerman, 1997). This interview covers the 11 formal DSM-IV-TR axis II diagnoses including PD NOS, two appendix diagnoses (i.e., depressive and passive-aggressive PD), and self-defeating PD. Interviewers were Master’s level psychologists, who were trained thoroughly by one of the authors (R.V.), and who received monthly booster sessions to avoid deviation from the interviewer guidelines. Inter-rater reliability was evaluated in 25 video-taped interviews, which were rated by three observer-raters. Percentage of agreement between observer-raters ranged from 84 (avoidant PD) to 100% (schizoid)
(median = 95%). Intraclass correlation coefficients for the sum of DSM-IV PD traits present (i.e., scores 2 or 3) ranged from .60 (schizotypal) through .92 (antisocial) (median = .74).

To measure patient characteristics at baseline, the assessment battery also included three self-report instruments. The first of those was the Dutch version of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ; Livesley & Jackson, 2002; van Kampen, 2002), for measuring the type and degree of personality pathology. We used patients’ scores on this questionnaire for the four higher-order factors: (a) emotional dysregulation, (b) dissocial behaviour, (c) inhibition, and (d) compulsivity. To measure the severity of personality pathology we used the five higher-order domains of the Severity Indices of Personality Problems, Dutch version (SIPP; Verheul et al., 2008): (a) self-control, (b) social concordance, (c) identity integration, (d) relational capacities, and (e) responsibility. To measure patients’ motivation for treatment, we used the two scales of the Motivation for Treatment Questionnaire (MTQ-8; van Beek & Verheul, 2008): need for help and readiness to change.

**Outcome measures**

The primary outcome measure was general psychiatric symptomatology. This was measured using the Dutch version of the Brief Symptom Inventory (BSI; de Beurs & Zitman, 2006; Derogatis & Melisaratos, 1983), a validated self-report scale derived from the Symptom Checklist 90-Revised (SCL-90-R; Arrindell & Ettema, 2003; Derogatis, 1992). In this study, we used the mean score of the 53 items of the Brief Symptom Inventory, i.e., the Global Severity Index (GSI), ranging from 0 to 4. Psychosocial functioning was measured with two subscales of the Outcome Questionnaire-45, Dutch version (OQ-45; Lambert et al., 1996): interpersonal relations and social role functioning. Health-related quality of life was measured using the Dutch version of the EuroQol EQ-5D (EQ-5D; Brooks, Rabin, & de Charro, 2003). All four outcome measures, the GSI, OQ-45 interpersonal relations, OQ-45 social role, and EQ-5D, were assessed at baseline and several follow-up points. Three of the six treatment centres conducted their follow-up at approximately 12, 24, and 36 months after baseline; the other three treatment centres conducted their follow-up at the end of treatment, approximately six and 12 months afterwards, and again at 36 months after baseline. The use of different assessment points was due to logistic reasons, and was taken into account by choosing multi-level modelling as the statistical method for the analyses.
Statistical analyses
We first examined the uncorrected results on all four outcome measures estimated at 12 months after baseline. We used multi-level modelling to deal with: (a) the dependency of repeated measures on the same subject in time, and (b) longitudinal data with observations unequally spaced in time (see “Outcome measures”). To estimate the uncorrected treatment effect at 12 months after baseline, we used a random intercept and random slope model with time as level I and patient number as level II. This resulted in a final best-fitting model with the following independent variables: dummy variables indicating group membership, time, and interaction between group membership and time. Subsequently, we calculated within-group effect sizes Cohen’s $d$ (Cohen, 1988) to describe change from baseline to 12 months in each group.

However, since this is a non-randomised study, the comparison of the groups had to be corrected for the influence of confounders, i.e., initial patient differences. To adjust for these differences and avoid bias in effect estimation, we included a “multiple propensity score” in our analyses. The classic propensity score is defined as the conditional probability of assignment to one of two treatment groups given a set of observed pre-treatment variables (Rosenbaum & Rubin, 1983). The multiple propensity score is an extension of the classic propensity score to more than two treatment groups (Imbens, 2000). Statistical inclusion of possible confounders in the outcome analyses controls selection bias due to known confounders while comparing multiple groups. To identify relevant confounders, we considered a long list of social, economic, and diagnostic variables carefully selected by both clinicians and researchers, based on the literature and clinical knowledge (Bartak et al., 2009). All variables significantly related to a specific outcome were used to estimate the multiple propensity scores in a multinomial regression analysis, with ‘group membership’ as a dependent variable (see Table 4.1 for the variables included in the GSI propensity score). One major advantage of the propensity score method, as compared to other correction techniques, is the fact that the overlap in propensity score distributions (and thus the overlap in relevant variables) between treatment groups can be easily judged and visualised (Spreeuwenberg et al., 2010). From looking at the overlap between the five treatment groups, it appeared that in spite of some differences these groups were readily comparable. For a detailed description of this method and its use in psychotherapy research, see Bartak et al. (2009) and Spreeuwenberg et al. (2010).
A more sophisticated multi-level model, now including multiple propensity scores, was used to compare change in outcome variables across treatment groups. Dependent variables were all available change scores observed during follow-up for each of the outcome measures. Independent variables were dummy variables indicating group membership, time, interaction between group membership and

Table 4.1. Variables used for propensity score estimation, outcome GSI

<table>
<thead>
<tr>
<th>Variable</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Patient’s age</td>
</tr>
<tr>
<td>DAPP-BQ Emotional dysregulation</td>
<td>Unstable affective responding, interpersonal problems</td>
</tr>
<tr>
<td>DAPP-BQ Inhibition</td>
<td>Deriving little enjoyment from intimate relationships</td>
</tr>
<tr>
<td>MTQ-8 Need for help</td>
<td>Patient’s expressed desire for external help</td>
</tr>
<tr>
<td>MTQ-8 Readiness to change</td>
<td>Willingness for treatment-seeking behaviour</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Quality of life</td>
</tr>
<tr>
<td>SIPP Self-control</td>
<td>The capacity to tolerate, use, and control one’s own emotions and impulses</td>
</tr>
<tr>
<td>SIPP Identity integration</td>
<td>Coherence of identity; the ability to see oneself and one’s own life as stable, integrated, and purposive</td>
</tr>
<tr>
<td>SIPP Relational capacities</td>
<td>The capacity to genuinely care about others as well as feeling cared about 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</td>
</tr>
<tr>
<td>SIPP Responsibility</td>
<td>The capacity to set realistic goals and to achieve these goals in line with the expectations you have generated in others</td>
</tr>
<tr>
<td>GSI</td>
<td>Level of psychiatric symptoms</td>
</tr>
<tr>
<td>OQ-45 Symptom distress</td>
<td>Level of symptom distress</td>
</tr>
<tr>
<td>OQ-45 Relational functioning</td>
<td>Level of interpersonal functioning</td>
</tr>
<tr>
<td>OQ-45 Social role functioning</td>
<td>Level of social and work functioning</td>
</tr>
<tr>
<td>Dimensional score cluster C PDs</td>
<td>Dimensional score of cluster C PD characteristics</td>
</tr>
<tr>
<td>Total dimensional score all PDs</td>
<td>Dimensional score of all PD characteristics</td>
</tr>
<tr>
<td>Avoidant PD</td>
<td>Diagnosis of avoidant PD</td>
</tr>
<tr>
<td>Dependent PD</td>
<td>Diagnosis of dependent PD</td>
</tr>
<tr>
<td>Obsessive-compulsive PD</td>
<td>Diagnosis of obsessive-compulsive PD</td>
</tr>
</tbody>
</table>

DAPP-BQ = Dimensional Assessment of Personality Pathology-Basic Questionnaire, MTQ-8 = Motivation for Treatment Questionnaire, EQ-5D = EuroQoL, EQ-5D, SIPP = Severity Indices of Personality Problems, GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45.
time, and the multiple propensity scores (with their mutual interactions). This model estimated differences in change scores at 12 months after baseline in pairwise comparisons of the five treatment groups. If significant differences in change scores were found, we calculated between-group effect sizes.

To render the outcome estimates at 12 months more reliable, we made optimum use of the potential of our data-set by including all available data collected up to 800 days after baseline. Data collected after that point was not used in order to prevent bias of the 12-month data due to changes much later in the process. The number of available follow-up measures was as follows: Up to 800 days, 30.5% of the total sample had one follow-up measure, 36.7% had two follow-up measures, and 32.9% had three follow-up measures. The analyses were performed using SPSS 15.0 for data preparation and Proc Mixed of SAS 9.1.3 for multi-level modelling (SAS Institute, Cary, N.C., USA).

Results

Sample characteristics
Of the 371 patients, 29.6% were male and 70.4% were female. The mean age was 33.5 years (SD = 9.5). The highest level of education was low for 22.9%, medium for 19.4%, and high for 57.7%. Furthermore, 70.4% were unmarried, 21.3% were married, and 8.4% were divorced or widowed. The majority, 66.6%, had pure cluster C PD (i.e., no comorbid cluster A or B PD), 23.7% had a combination of cluster C PD and cluster B PD, 4.0% had a combination of cluster C PD and cluster A PD, and 5.7% had a combination of cluster C PD and both cluster A and B PD. A majority (63.3%) had a diagnosis of avoidant PD, 49.3% had a diagnosis of obsessive-compulsive PD, and 22.6% a diagnosis of dependent PD.

Uncorrected outcome
One year after baseline, patients in all treatment groups showed improvement in terms of psychiatric symptoms (GSI), the primary outcome measure. This is shown in Table 4.2 and Figure 4.2. Within-group effect sizes of the uncorrected scores ranged from 0.62 (medium effect, short-term day hospital group) to 1.78 (huge effect, short-term inpatient group).
### Table 4.2. Uncorrected outcomes and effect sizes in five treatment groups for all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment group</th>
<th>Baseline</th>
<th>12 months</th>
<th>Within-group effect size</th>
<th>Cohen's $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Long outpatient</td>
<td>1.49 (0.69)</td>
<td>1.07 (0.65)</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>$n = 68$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>1.44 (0.63)</td>
<td>1.04 (0.67)</td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>$n = 77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>1.68 (0.61)</td>
<td>1.12 (0.94)</td>
<td></td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>$n = 74$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>1.75 (0.52)</td>
<td>0.76 (0.60)</td>
<td></td>
<td>1.78</td>
</tr>
<tr>
<td></td>
<td>$n = 59$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>1.77 (0.72)</td>
<td>1.03 (0.68)</td>
<td></td>
<td>1.06</td>
</tr>
<tr>
<td></td>
<td>$n = 93$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OQ-45 Social role</td>
<td>Long outpatient</td>
<td>15.84 (4.27)</td>
<td>12.98 (4.42)</td>
<td></td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>$n = 68$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>15.20 (4.52)</td>
<td>13.59 (4.53)</td>
<td></td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>$n = 77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>16.79 (4.75)</td>
<td>13.39 (5.29)</td>
<td></td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>$n = 74$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>17.78 (3.84)</td>
<td>12.41 (4.83)</td>
<td></td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>$n = 59$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>16.97 (4.64)</td>
<td>12.42 (5.31)</td>
<td></td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>$n = 93$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OQ-45 Interp. relations</td>
<td>Long outpatient</td>
<td>22.22 (5.98)</td>
<td>19.37 (6.43)</td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>$n = 68$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>20.93 (5.24)</td>
<td>18.17 (5.90)</td>
<td></td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>$n = 77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>22.89 (6.41)</td>
<td>18.41 (8.05)</td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>$n = 74$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>23.97 (5.63)</td>
<td>17.54 (6.77)</td>
<td></td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>$n = 59$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>24.09 (5.24)</td>
<td>18.38 (6.59)</td>
<td></td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>$n = 93$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Long outpatient</td>
<td>0.58 (0.24)</td>
<td>0.73 (0.16)</td>
<td></td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>$n = 68$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>0.60 (0.25)</td>
<td>0.69 (0.24)</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>$n = 77$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>0.50 (0.27)</td>
<td>0.72 (0.22)</td>
<td></td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>$n = 74$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>0.49 (0.27)</td>
<td>0.78 (0.21)</td>
<td></td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>$n = 59$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>0.51 (0.26)</td>
<td>0.68 (0.25)</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>$n = 93$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D = EuroQoL EQ-5D.
Improvements were also seen in terms of psychosocial functioning and quality of life (see Table 4.2). Effect sizes for these outcome measures were somewhat lower compared to psychiatric symptoms, but a positive change in psychosocial functioning and quality of life was evident.

**Corrected comparison**

After correction for all relevant pre-treatment differences, improvement between baseline and 12 months proved to be significant for patients in all treatment groups on all four outcome measures ($p < .001$).

The short-term inpatient group showed significantly more improvement in psychiatric symptoms (GSI) than the following three treatment groups: the short-term day hospital group ($B = 0.38, p = .0059, 95\% \text{ CI} [0.11, 0.65]$), the long-term day hospital group ($B = 0.43, p = .0032, 95\% \text{ CI} [0.15, 0.71]$), and the long-term inpatient group ($B = 0.31, p = .0248, 95\% \text{ CI} [0.04, 0.57]$) This can be seen in Table 4.3. Between-group effect sizes (Cohen’s $d$) were 0.54, 0.57, and 0.40, respectively. This indicates medium effect sizes for the between-group comparisons of short-term inpatient treatment versus other treatment groups.
Table 4.3. Difference scores (unstandardised $B$) of change scores from baseline to 12 months, corrected for propensity score, all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment group</th>
<th>$n$</th>
<th>Long outpatient</th>
<th>Short day hospital</th>
<th>Long day hospital</th>
<th>Short inpatient</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Long outpatient</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>77</td>
<td>-0.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>74</td>
<td>-0.128</td>
<td>-0.050</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>59</td>
<td>0.302</td>
<td>0.380 **</td>
<td>0.430 **</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>93</td>
<td>-0.004</td>
<td>0.075</td>
<td>0.124</td>
<td>-0.306 *</td>
</tr>
<tr>
<td>OQ-45</td>
<td>Long outpatient</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social role</td>
<td>Short day hospital</td>
<td>77</td>
<td>-1.632</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>74</td>
<td>-1.123</td>
<td>0.460</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>59</td>
<td>0.876</td>
<td>2.508 **</td>
<td>2.048 *</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>93</td>
<td>-0.169</td>
<td>1.463</td>
<td>1.003</td>
<td>-1.045</td>
</tr>
<tr>
<td>OQ-45</td>
<td>Interp. relations</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>77</td>
<td>-0.836</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>74</td>
<td>-0.611</td>
<td>0.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>59</td>
<td>1.704</td>
<td>2.540 *</td>
<td>2.315</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>93</td>
<td>-0.084</td>
<td>0.752</td>
<td>0.527</td>
<td>-1.788</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Long outpatient</td>
<td>68</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short day hospital</td>
<td>77</td>
<td>-0.060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long day hospital</td>
<td>74</td>
<td>0.001</td>
<td>0.061</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short inpatient</td>
<td>59</td>
<td>0.089</td>
<td>0.149 ***</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long inpatient</td>
<td>93</td>
<td>-0.021</td>
<td>0.039</td>
<td>-0.022</td>
<td>-0.110 *</td>
</tr>
</tbody>
</table>

*Positive coefficients indicate that the treatment group shown in the left column is superior, negative coefficients indicate that the treatment group in the above row is superior.

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D = EuroQoL EQ-5D.

* $p < .05$, ** $p < .01$, *** $p < .001$. 
In terms of social role functioning, the short-term inpatient group improved significantly more than two other groups—the short-term day hospital group ($B = 2.51, p = .0067, 95\% \text{ CI } [0.71, 4.31]$) and the long-term day hospital group ($B = 2.05, p = .0476, 95\% \text{ CI } [0.02, 4.07]$) — with between-group effect sizes of 0.49 and 0.38, respectively. The improvement in interpersonal functioning was significantly higher in the short-term inpatient group than in one other group—the short-term day hospital group ($B = 2.54, p = .0319, 95\% \text{ CI } [0.22, 4.86]$) — with a between-group effect size of 0.39. Quality of life improved significantly more in the short-term inpatient group than in two other groups: the short-term day hospital group ($B = 0.15, p = .0009, 95\% \text{ CI } [0.06, 0.23]$) and the long-term inpatient group ($B = 0.11, p = .0113, 95\% \text{ CI } [0.03, 0.19]$). Between-group effect sizes were 0.6 and 0.42, respectively.

All results were based on intention-to-treat analyses (ITT), whereby ITT is defined as assignment and a minimal exposure to the intended treatment modality. The analyses were repeated with the treatment completers, i.e., those who actually stayed in the intended treatment modality group during their treatment ($n = 298$, 80.3\% of the ITT sample, ranging from 66.2\% for short-term day hospital to 89.7\% for long-term outpatient treatment). These results followed the same pattern as the results from the ITT analyses: significant change within all treatment groups and a superiority of short-term inpatient treatment across all outcome measures.

**Discussion**

This is the first study comparing the effectiveness of five modalities of psychotherapeutic treatment in a large population of patients with cluster C PDs, as a contribution to the search for effective treatments for this patient group. Patients in all treatment groups had improved psychiatric symptoms, psychosocial functioning, and quality of life after 12 months. Most improvement was observed in the short-term inpatient group. This finding held when pre-treatment differences were controlled for with the propensity score.

**Strengths and limitations**

A clear strength of the present study is its external validity and clinical utility: It was conducted in regular clinical practice, not under experimental conditions (Hodgson, Bushe, & Hunter, 2007). A second strength is the rigorous statistical control of potential confounders, using the multiple propensity score methodology. Finally, a major asset of this study is its large number of patients. All this enabled
the comparison of different psychotherapeutic treatment modalities while keeping sufficient statistical power.

Despite these strengths, the present findings have to be interpreted considering several limitations. First, even though we controlled for all observed pre-treatment differences, it cannot be ruled out that results have been influenced by unobserved confounders. To diminish this constraint as much as possible, a broad range of possible confounders was carefully selected and measured, based on both clinical and empirical knowledge (Bartak, et al., 2009), including variables identified in the literature as significant predictors of therapy outcome or process, such as severity of baseline psychopathology, previous hospitalisation, and substance misuse (e.g., Gunderson et al., 2006; Links, Mitton, & Steiner, 1993; McGlashan, 1985; McMain et al., 2009; Ogrodniczuk et al., 2008; Plakun, 1991; Ryle & Golynkina, 2000). In line with these earlier findings, previous hospitalisation and substance misuse for example were significantly related to one of the secondary outcome measures, interpersonal functioning, and were therefore included in the propensity score for this measure. However, even when considerably reducing the possibility of important confounders being overlooked, not all possible variables could be covered in interviews and questionnaires at baseline, and therefore several variables, such as self-harm (Chiesa & Fonagy, 2007), were not measured.

Second, for ethical reasons, a control group receiving no treatment at all was not included. Yet, several previous studies showed that specialised psychotherapeutic treatment yields better outcomes than various control conditions (e.g., waiting list controls; Alden, 1989; Emmelkamp, et al., 2006; Winston, et al., 1994).

Third, research compliance differed between the treatment groups compared with most missing follow-up observations in the long-term treatment groups (see Figure 4.1). This might cause a problem of internal validity if non-response is not random, but related to systematic bias in effect estimation (positive or negative). However, there are two reasons why systematic bias seems unlikely: (a) responders and non-responders did not differ in psychiatric symptoms at baseline, and therefore it seems that they do not represent two structurally different groups of patients; and (b) during the frequent telephone contact the authors had with non-responding patients to remind them to send back their questionnaires, these patients reported both negative and positive outcomes as reasons why they did not respond: Some of them argued that their problems had worsened and that therefore they felt they did not have enough energy to fill in the questionnaires, others argued that their life had
changed in a positive way and that therefore they did not want to be reminded of their time in therapy by filling in the questionnaires. Keeping this in mind, it seems unlikely that non-response was related to systematic negative or positive bias.

Fourth, this study does not rule out the possibility that treatment characteristics other than setting and duration played a role in the differential effectiveness of the five treatment modalities, for example frequency of sessions or theoretical orientation of treatment. This might represent a potential threat to internal validity. This is especially true for the role of theoretical orientation as a possible factor in the superiority of short-term inpatient treatment: Most short-term inpatient programs were based on psychodynamic principles. This concern is somewhat mitigated by previous studies comparing different theoretical orientations where no differences were found (e.g., Svartberg, et al., 2004). However, to test the differential effect of modality and other treatment characteristics, a combined research design combining all these factors is needed.

**Future directions and implications**
What are the implications of the present results for future research, for practice guidelines, and for everyday clinical practice? For patients with cluster C personality pathology, short-term inpatient treatment clearly was associated with the highest improvement within 12 months. For this patient group, this modality of therapy seems to be the treatment backed up by the best available evidence—in absence of long-term follow-up data. Replication of these results in a long(er)-term follow-up study is of vital importance to draw final conclusions. There might be a bias in favour of short-term treatment because patients in the long-term treatment groups might still be in therapy at 12 months. Long-term follow-up after termination of all treatment programs is therefore warranted. Another question is whether the benefit in terms of effectiveness is worth the potential cost differences when evaluated with recently upcoming state-of-the-art cost-effectiveness analyses (e.g., Leichsenring et al., 2009; McCrone et al., 2007). From these analyses within our study sample, it appeared that the mean direct treatment costs of the five treatment modalities were EUR 10,005 (SE = 1,134) for long-term outpatient treatment, EUR 16,813 (SE = 1,361) for short-term day hospital treatment, EUR 27,648 (SE = 2,654) for long-term day hospital treatment, EUR 25,933 (SE = 859) for short-term inpatient treatment, and EUR 49,260 (SE = 2,435) for long-term inpatient treatment (Soeteman et al., in press). It would be interesting to compare the cost-effectiveness of short-term inpatient psychotherapeutic treatment with that of manualised outpatient treatments, such as cognitive-behavioural therapy (Emmelkamp, et al., 2006). A state-of-the-
Chapter 4

Art cost-effectiveness analysis would include medical costs incurred outside the treatment institution, productivity costs, and other indirect costs. This kind of analysis and its economic interpretation is beyond the range of this study and needs considerable research efforts in the future. If the superiority of short-term inpatient psychotherapeutic treatment holds at long-term follow-up, in cost-effectiveness analyses, and in comparison with other evidence-based manualised treatments, this treatment modality might be considered as the treatment of choice for this patient group. This would be a thought-provoking finding, as previous studies in cluster B PD patients have found outpatient (Chiesa, Fonagy, & Gordon, 2009; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Giesen-Bloo et al., 2006) and day hospital (Bateman & Fonagy, 2001) treatments to be very effective in this population. Even though no study compared one of these modalities directly with inpatient therapy, one might speculate that different therapy modalities are effective for different groups of patients. It could be that the success of short-term inpatient treatment in a cluster C PD sample is embedded in the combination of only short hospitalisation — thereby preventing iatrogenic effects — and a high level of therapeutic intensity and pressure. Patients with cluster C personality pathology might be able to handle the high pressure of this treatment modality better than (pure) cluster B PD patients, who probably have a lower tolerance for therapeutic pressure, resulting in more early dropouts and thus a less effective treatment. They might instead need less pressure with a longer treatment duration (Bateman & Fonagy, 2001; Lorentzen & Høglend, 2008). Future studies may verify this hypothesis. However, even when superiority of short-term inpatient treatment for cluster C PD patients will be confirmed in the literature, patients caring for children might still not be assigned to inpatient treatment. Also, patients with a high severity of psychiatric symptoms or a low level of ego strength might not be able to handle the pressure of intensive inpatient treatment. It is recommended to investigate these potential matching factors further as this would enable clinicians to make specific treatment recommendations for different subgroups of cluster C PD patients and to develop new clinical practice guidelines.
In conclusion, this study suggests that psychotherapy, especially in a short-term inpatient modality, is an effective treatment for patients with cluster C PDs. This makes short-term inpatient psychotherapeutic treatment an interesting option for patients with avoidant, dependent, and obsessive-compulsive PD. The present findings can contribute to more adequate and tailored health care for this vulnerable patient group, as implementing effective treatments may reduce the considerable burden to individuals and society as a whole.
Chapter 5

Effectiveness of outpatient, day hospital, and inpatient psychotherapeutic treatment for patients with cluster B personality disorders

Cluster B PDs affect a considerable percentage of the general population (3.1% - 4.5%; Samuels et al., 2002; Torgersen, Kringlen, & Cramer, 2001) and are highly prevalent in psychiatric patients (13.0%; Zimmerman, Rothschild, & Chelminski, 2005). Borderline, histrionic, narcissistic, and antisocial PD are not only associated with individual suffering (Duggan, 2009; Lieb, Zanarini, Schmah, Linehan, & Bohus, 2004; Nestadt et al., 1990; Ogrodniczuk, Piper, Joyce, Steinberg, & Duggal, 2009), but also with early institutional care and criminality (Coid, Yang, Tyrer, Roberts, & Ullrich, 2006). They place a heavy burden on both individuals and society as a whole (Bender et al., 2001; Cramer, Torgersen, & Kringlen, 2006; Skodol et al., 2005; Soeteman, Hakkaart-van Roijen, Verheul, & Busschbach, 2008; van Asselt, Dirksen, Arntz, & Severens, 2007). To date, a number of specific treatment protocols tailored to this patient group have been developed. This is especially the case for borderline PD (de Groot, Verheul, & Trijsburg, 2008; Zanarini, 2009). Examples are Dialectical Behavior Therapy (DBT; Linehan, 1993), Mentalization Based Treatment (MBT; Bateman & Fonagy, 2004), Schema Focused Therapy (SFT; Young, Klosko, & Weishaar, 2003), and Transference Focused Therapy (TFP; Yeomans, Clarkin, & Kernberg, 2002). As a consequence, methodologically sound research on psychotherapeutic treatments for cluster B PDs has increased during the last 10 years. However, no study has directly compared outpatient, day hospital, and inpatient treatment within one patient sample. The available published studies have focused either on the development of specific treatment manuals and their evaluation (e.g., Bateman & Fonagy, 1999, 2001, 2008, 2009; Davidson et al., 2006; Doering et al., 2010; Farrell, Shaw, & Webber, 2009; Koons et al., 2001; McMain et al., 2009; Soler et al., 2009; Teusch, Böhme, Finke, & Gastpar, 2001; Turner, 2000) or on the comparison of different theoretical schools within outpatient settings (Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Cottraux et al., 2009; Giesen-Bloo et al., 2006). Another focus of psychotherapy outcome research was the effectiveness of specific treatment ingredients, such as transference interpretations (Hoglund et al., 2006; Hoglund et al., 2008) or crisis support outside office hours (Nadort et al., 2009). So until now, hardly any study has addressed possible differences in treatment effectiveness between different settings and across theoretical schools. Exceptions are the studies by Chiesa et al. (2009; 2006; 2004) comparing inpatient treatment and step-down/outpatient treatment and the study by Arnevik et al. (2009) comparing day hospital treatment and outpatient treatment.

Comparing different treatment settings is relevant as it can be anticipated that the type of setting will affect costs for both patients and society, as well as patients’ time. One of the reasons this comparison (e.g., outpatient versus day hospital versus inpatient treatment) has not been made previously, is probably the difficulty of
randomly assigning patients to different treatment settings in clinical samples due to practical or ethical constraints (Black, 1996). Furthermore, even if researchers were successful in setting up and starting a randomised treatment setting study, external validity would be doubtful because a high number of patients would probably refuse to participate. In non-randomised studies, however, researchers often cannot draw valid conclusions due to the strong influence of selection bias. Accordingly, there is a need to design and conduct non-randomised effectiveness studies while minimising the influence of selection bias. This kind of study is increasingly being found in the literature (e.g., Facchinetti, Ottolini, Fazio, Rigatelli, & Volpe, 2007; Forstmeier & Rueddel, 2007; Golkaramnay, Bauer, Haug, Wolf, & Kordy, 2007; Grossman, Tiefenthaler-Gilmer, Raysz, & Kesper, 2007). In the present study this has been undertaken for the first time in a population of cluster B PD patients using the propensity score method to minimise the influence of selection bias.

The purpose of our study was to compare the effectiveness of different treatment settings (outpatient, day hospital, inpatient) in psychotherapy for cluster B PDs 18 months after baseline. The primary outcome measure in this study was psychiatric symptomatology. Additionally, effects on interpersonal and social functioning, as well as on quality of life were evaluated.

**Method**

**Participants**

Participants (N = 207) were recruited from consecutive admissions to six mental health care centres in the Netherlands (Centre of Psychotherapy De Viersprong, Halsteren; Altrecht, Utrecht; Zaans Medical Centre, Zaandam; Centre of Psychotherapy De Gelderse Roos, Lunteren; GGZWNB, Bergen op Zoom and Roosendaal; Arkin, Amsterdam). These institutions offer outpatient, day hospital, and/or inpatient psychotherapy for patients with personality pathology. From March 2003 to March 2006, 1,379 patients completed the intake and screening procedure, and were selected for treatment (see Figure 5.1). The intake and screening data were collected for all applicants, irrespective of study participation later on. The data obtained from this initial assessment served as baseline data for our study. As it was part of the standard screening procedure, informed consent for the baseline data collection was not mandatory under Dutch law.
Figure 5.1. Patient flow

Assessed for eligibility
N = 1379

Refused to participate
n = 100

Excluded due to other reasons
n = 164
(Logistic reasons n = 31
Missing/unreliable data n = 133)

Enrolled
n = 960

Dropped out prematurely*
n = 38

Did not meet inclusion criteria
n = 146
Met exclusion criteria
n = 9

No DSM-IV cluster B PD diagnosis
n = 677

Allocated
n = 245

Outpatient treatment
n = 59

Lost to follow-up
n = 13

Analyzed
n = 46

Day hospital treatment
n = 99

Lost to follow-up
n = 18

Analyzed
n = 81

Inpatient treatment
n = 87

Lost to follow-up
n = 7

Analyzed
n = 80

* These patients did not receive a “minimal effective dose” of treatment, defined as two sessions for outpatients and two treatment days for day hospital patients or inpatients, and were therefore not included in the initial study sample.
Of these 1,379 patients, 146 (10.6%) were excluded from the study for not meeting one of the following inclusion criteria: age between 18 and 70 years ($n = 13$), significant personality pathology ($n = 34$), and referral for psychotherapeutic treatment aimed at personality problems ($n = 99$). Nine patients (0.7%) met one of the following exclusion criteria: insufficient command of the Dutch language ($n = 6$), organic cerebral impairment ($n = 1$), mental retardation ($n = 1$), and schizophrenia ($n = 1$). This left 1,224 eligible participants, of whom 31 (2.5%) could not participate due to logistic reasons (e.g., no appointment could be made to provide informed consent). A total of 133 patients (10.9%) had to be excluded due to missing or unreliable baseline data during the intake and screening procedure. For the majority of these 133 patients ($n = 106$) no standardised Axis II diagnosis was obtained. This was mostly due to a practical problem, i.e., a shortage of interviewers at the start of the study ($n = 101$). Because of this logistic reason, it can be assumed that these data were “missing completely at random” and therefore they do not threaten internal validity (as they were unrelated to specific patient characteristics). Five patients had an unreliable Axis II diagnosis as determined by the interviewer, for example because the patient was too anxious or too depressed to obtain a reliable diagnosis with a semi-structured interview. A few patients did not return their assessment booklet during the intake procedure ($n = 27$). Thirty-eight patients (3.1%) received less than two treatment sessions of outpatient therapy or less than two days of inpatient or day hospital therapy. For example, some of the patients left treatment after one session, and some did not even show up for the first session. They were therefore excluded beforehand from the study sample in which we only included patients with a “minimal effective dose” of treatment. One hundred patients (8.2%) refused to participate. The remaining 922 patients were informed about the study and its procedure, provided written informed consent, and entered the study. Of those, 245 patients (26.6%) had one or more cluster B PDs.

In the absence of explicit guidelines for treatment assignment in PDs (Vervaeke & Emmelkamp, 1998), the selection procedure was based on the expert opinion of clinicians who used their clinical experience combined with patient data from standardised instruments. To clarify the criteria used for the assignment process, our research group had recently conducted a study with intake clinicians from the participating treatment centres. We found evidence of substantial consensus among clinicians concerning the criteria used for treatment decision making (van Manen et al., 2008).
Patients were assigned to one of three setting groups: 59 to outpatient, 99 to day hospital, and 87 to inpatient psychotherapy. Follow-up data were not available for 38 patients (15.5%, patients who did not respond to any follow-up assessment or patients where follow-up measurements were as yet not available). There was no difference in psychiatric symptoms at baseline between patients with follow-up data and those without (this holds true for both the comparison in the total sample and the comparison within the three treatment groups). The final sample consisted of 207 patients who were included in the ITT analyses.

**Treatment**
The six mental health care centres offered a variety of psychotherapeutic treatments tailored to a PD patient population. Their treatments differed according to several features. As this study focused on different treatment settings, the following three treatment groups were compared:

- Patients in outpatient treatment ($n = 46$, 22.2% of the study sample). These patients came for individual (89.2%) or group (10.9%) psychotherapy sessions for up to two sessions (mean = 0.7 sessions/week, $SD = 0.4$, median = 0.5) per week (mean duration = 14.5 months, $SD = 6.6$, median = 12.0).
- Patients in day hospital treatment ($n = 81$, 39.1% of the study sample). These patients came to the institutions at least one morning / afternoon per week (mean = 3.5 days/week, $SD = 1.4$, median = 3.0) and received different forms of psychotherapeutic and psychosocial treatment, but slept at home (mean duration = 10.4 months, $SD = 4.8$, median = 12.0).
- Patients in inpatient treatment ($n = 80$, 38.6% of the study sample). These patients stayed at the institutions five days a week and received different forms of psychotherapeutic and psychosocial treatment (mean duration = 9.1 months, $SD = 3.0$, median = 10.0).

Day Hospital and inpatient programs typically consisted of group psychotherapy as a core element, mostly in combination with one or more non-verbal or expressive group therapies, individual psychotherapy, sociotherapy within the therapeutic community, coaching for social problems, community meetings, and/or pharmacological treatment. The therapists were all licensed psychiatrists or psychologists. On average, they had 14.9 years ($SD = 10.1$) of postgraduate clinical experience. The treatments under study can be considered highly representative for regular clinical practice in The Netherlands, as therapists did not receive specific training for this study and treatment integrity was not monitored. The study protocol
was approved by the Medical Ethics Committee of the Erasmus University Medical Centre in Rotterdam.

Assessments

Baseline measures

An extensive standard assessment battery of instruments was administered to the patients before treatment assignment. PDs were measured using the Dutch version of the Structured Interview for DSM-IV Personality (SIDP-IV; de Jong, Derks, van Oel, & Rinne, 1996; Pfohl, Blum, & Zimmerman, 1997). This interview covers the 11 formal DSM-IV-TR Axis II diagnoses including PD NOS, two appendix diagnoses (i.e., depressive and passive-aggressive PD), and self-defeating PD. Interviewers were Master’s level psychologists, who were trained thoroughly by one of the authors (RV). They received monthly booster sessions to avoid deviation from the interviewer guidelines. Inter-rater reliability was evaluated in 25 videotaped interviews, which were rated by three observer-raters. Percentage of agreement between observer-raters ranged from 84 (avoidant PD) to 100% (schizoid) (median = 95%). Intraclass correlation coefficients (ICC) for the sum of DSM-IV PD traits present (i.e., scores 2 or 3) ranged from .60 (schizotypal) through .92 (antisocial) (median = .74).

To measure patient characteristics at baseline, the assessment battery also included three self-report instruments. The first of those was the Dutch version of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ; Livesley & Jackson, 2002; van Kampen, 2002), for measuring type and degree of personality pathology. We used patients’ scores on this questionnaire on the four higher order factors: (a) emotional dysregulation, (b) dissocial behaviour, (c) inhibition, and (d) compulsivity. To measure the severity of personality pathology we used the five higher-order domains of the Severity Indices of Personality Problems, Dutch version (SIPP; Verheul et al., 2008): (a) self-control, (b) social concordance, (c) identity integration, (d) relational capacities, and (e) responsibility. To measure patients’ motivation for treatment, we used the two scales of the Motivation for Treatment Questionnaire (MTQ-8; van Beek & Verheul, 2008): need for help and readiness to change.

Outcome measures

The primary outcome measure was general psychiatric symptomatology. This was measured using the Dutch version of the Brief Symptom Inventory (BSI; de Beurs & Zitman, 2006; Derogatis & Melisaratos, 1983), a validated self-report scale derived from the Symptom Checklist 90-Revised (SCL-90-R; Arrindell & Ettema, 2003;
Derogatis, 1992). In this study, we used the mean score of the 53 items of the BSI, i.e.,
the Global Severity Index (GSI), ranging from 0 to 4. Psychosocial functioning was
measured with two subscales of the Outcome Questionnaire-45, Dutch version (OQ-
45; Lambert et al., 1996): interpersonal relations and social role functioning. Health-
related quality of life was measured using the Dutch version of the EuroQol EQ-5D
(EQ-5D; Brooks, Rabin, & de Charro, 2003). All four outcome measures, GSI, OQ-
45 interpersonal relations, OQ-45 social role, and EQ-5D, were assessed at baseline
and several follow-up points. Three of the six treatment centres conducted follow-
up at approximately 12, 24, and 36 months after baseline; the other three treatment
centres conducted follow-up at the end of treatment, subsequently approximately
six and 12 months afterwards, and again at 36 months after baseline. The use of
different assessment points was due to logistic reasons and was taken into account
by choosing multi-level modelling as the statistical method for the analyses.

Statistical analyses

We first examined the uncorrected results on all four outcome measures estimated
at 18 months after baseline. We used multi-level modelling to deal with (a) the
dependency of repeated measures on the same subject in time, and (b) longitudinal
data with observations unequally spaced in time (see “Outcome measures”). To
estimate the uncorrected treatment effect at 18 months after baseline we used a
random intercept and random slope model with time as level I and patient number
as level II. This resulted in a model with the following independent variables:
dummy variables indicating group membership, time, and interaction between
group membership and time. Subsequently, we calculated within-group effect sizes
Cohen’s $d$ (Cohen, 1988) to describe change from baseline to 18 months per treatment
group.

However, treatment groups cannot be compared based on the uncorrected results in
this non-randomised clinical trial as these findings might be confounded by initial
patient differences. To adjust for these differences and to avoid bias in effect estimation,
we included a “multiple propensity score” in our analyses. The classic propensity
score is defined as the conditional probability of assignment to one of two treatment
groups given a set of observed pre-treatment variables (Rosenbaum & Rubin, 1983).
The multiple propensity score is an extension of the classic propensity score to more
than two treatment groups (Imbens, 2000). To identify relevant confounders, we
considered a long list of social, economic, and diagnostic variables carefully selected
by both clinicians and researchers, based on the literature and clinical knowledge
(Bartak et al., 2009). All variables significantly related to a specific outcome were
used to estimate the multiple propensity scores in a multinomial regression analysis, with ‘group membership’ as a dependent variable (see Table 5.1 for the variables included in the GSI propensity score). A major advantage of the propensity score method, as compared to other correction techniques, is the fact that the overlap in propensity score distributions (and thus the overlap in relevant variables) between treatment groups can be easily judged and visualised (Spreeuwenberg et al., 2010). From looking at the overlap between the three treatment groups, it appeared that in spite of some differences, these groups were readily comparable. For a detailed description of this method and its use in psychotherapy research see Bartak et al. (2009) and Spreeuwenberg et al. (2010).

To compare change in outcome variables across treatment groups, a more sophisticated multi-level model, now including multiple propensity scores, was used. Dependent variables were all available change scores observed during follow-up for each of the outcome measures. The following independent variables were entered in the initial model: dummy variables indicating group membership, time, the multiple propensity scores (with their mutual interactions), the two-way-interactions between group membership and time, the two-way-interactions between propensity scores and time, the two-way-interactions between propensity scores and group membership, and the three-way-interactions between propensity scores, time, and group membership. Then variables were eliminated from the model by backward selection to obtain a final best-fit model. This model estimated differences in change scores at 18 months after baseline in pair-wise comparisons of the three treatment groups.

Follow-up response was high, enhancing the robustness of the multi-level analyses: 12.1% of the respondents completed one follow-up measurement, 6.8% completed two follow-up measurements, 39.6% completed three follow-up measurements, 39.1% completed four follow-up measurements, and 2.4% completed five follow-up measurements. The analyses were performed using SPSS 17.0 for data preparation and Proc Mixed of SAS 9.1.3 for multi-level modelling (SAS Institute Inc, Cary, North Carolina, USA).
Table 5.1. Variables used for propensity score estimation, outcome GSI

<table>
<thead>
<tr>
<th>Variable</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Patient’s age</td>
</tr>
<tr>
<td>DAPP-BQ Emotional dysregulation</td>
<td>Unstable affective responding, interpersonal problems</td>
</tr>
<tr>
<td>DAPP-BQ Dissocial behaviour</td>
<td>Lacking regard for others</td>
</tr>
<tr>
<td>DAPP-BQ Inhibition</td>
<td>Deriving little enjoyment from intimate relationships</td>
</tr>
<tr>
<td>DAPP-BQ Compulsivity</td>
<td>Compulsivity and absence of oppositional behaviour</td>
</tr>
<tr>
<td>MTQ-8 Need for help</td>
<td>Patient’s expressed desire for external help</td>
</tr>
<tr>
<td>MTQ-8 Readiness to change</td>
<td>Willingness for treatment-seeking behaviour</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Quality of life</td>
</tr>
<tr>
<td>SIPP Self-control</td>
<td>The capacity to tolerate, use, and control one’s own emotions and impulses</td>
</tr>
<tr>
<td>SIPP Social concordance</td>
<td>The ability to value someone’s identity, withhold aggressive impulses towards others and to work together with others</td>
</tr>
<tr>
<td>SIPP Identity integration</td>
<td>Coherence of identity; the ability to see oneself and one’s own life as stable, integrated, and purposive</td>
</tr>
<tr>
<td>SIPP Relational capacities</td>
<td>The capacity to genuinely care about others as well as feeling 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</td>
</tr>
<tr>
<td>SIPP Responsibility</td>
<td>The capacity to set realistic goals and to achieve these goals in line with the expectations generated in others</td>
</tr>
<tr>
<td>GSI</td>
<td>Level of psychiatric symptoms</td>
</tr>
<tr>
<td>OQ-45 Symptom distress</td>
<td>Level of symptom distress</td>
</tr>
<tr>
<td>OQ-45 Relational functioning</td>
<td>Level of interpersonal functioning</td>
</tr>
<tr>
<td>OQ-45 Social role functioning</td>
<td>Level of social and work functioning</td>
</tr>
<tr>
<td>SIDP-IV Cluster C PDs</td>
<td>Number of cluster C PDs</td>
</tr>
<tr>
<td>SIDP-IV Dimensional score cluster C PDs</td>
<td>Dimensional score of cluster C PD characteristics</td>
</tr>
<tr>
<td>Borderline PD</td>
<td>Diagnosis of borderline PD</td>
</tr>
<tr>
<td>Narcissistic PD</td>
<td>Diagnosis of narcissistic PD</td>
</tr>
</tbody>
</table>

DAPP-BQ = Dimensional Assessment of Personality Pathology-Basic Questionnaire, MTQ-8 = Motivation for Treatment Questionnaire, EQ-5D = EuroQol EQ-5D, SIPP = Severity Indices of Personality Problems, GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, SIDP-IV = Structured Interview for DSM-IV Personality.
Chapter 5

Results

Sample characteristics
The majority of patients (71%) were female. The mean age was 31.3 years (SD = 8.5). The level of education was low for 33.3%, medium for 19.3%, and high for 47.3%. Furthermore, 78.7% were unmarried, 12.6% were married, and 8.7% were divorced or widowed. Most patients (77.3%) had a diagnosis of borderline PD, 22.7% had a diagnosis of narcissistic PD, 12.6% had a diagnosis of histrionic PD, and 8.7% a diagnosis of antisocial PD. There was considerable overlap with PDs from other clusters: 40.6% had pure cluster B PD (i.e., no comorbid cluster A or C PD), 44.9% had a combination of cluster B PD and cluster C PD, 3.4% had a combination of cluster B PD and cluster A PD, and 11.1% had a combination of cluster B PD and both cluster A and C PD.

Treatment compliance
Before start of treatment, every patient received an allocation to a certain treatment dosage. According to our registration at the end of treatment, about one third of all patients (n = 62) underwent exactly the intended treatment (31.1% of the outpatient group, 23.8% of the day hospital group, and 36.7% of the inpatient group). Of the 145 patients whose received treatment deviated from their intended treatment, 60 (41.4% of the deviating patients) stayed in treatment shorter than planned (28.1% of the deviating patients in the outpatient group, 45.2% in the day hospital group, and 45.1% in the inpatient group). Of these 60 patients, 40 decided in agreement with their therapist that treatment was no longer beneficial or necessary, whereas only 20 patients dropped out of treatment prematurely or were forced to leave earlier by the staff. From these 20 dropouts, 12 patients were from the day hospital group (14.8% of the total day hospital group) and eight patients were from the inpatient group (10.0% of the total inpatient group). All dropouts were treatment dropouts (and no study dropouts) who completed follow-up measures and were included in the ITT analyses.

Treatment outcome
Eighteen months after baseline, patients in all three settings improved remarkably in terms of psychiatric symptoms (GSI), the primary outcome measure (see Figure 5.2 and Table 5.2). Within-group effect sizes of the uncorrected scores were 0.55 (medium effect) for outpatient psychotherapy, 0.97 (large effect) for day hospital psychotherapy, and 1.37 (very large effect) for inpatient psychotherapy.
Figure 5.2. GSI uncorrected mean scores at baseline and 18-month follow-up

Table 5.2. Uncorrected outcomes (mean ± SD) and effect sizes in three treatment groups for all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment group</th>
<th>Baseline</th>
<th>18 months</th>
<th>Within-group effect size, Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Outpatient (n = 46)</td>
<td>1.52 ± 0.78</td>
<td>1.10 ± 0.75</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 81)</td>
<td>1.74 ± 0.62</td>
<td>1.09 ± 0.72</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 80)</td>
<td>1.94 ± 0.66</td>
<td>1.02 ± 0.69</td>
<td>1.37</td>
</tr>
<tr>
<td>OQ-45 Social role</td>
<td>Outpatient (n = 46)</td>
<td>15.63 ± 4.35</td>
<td>12.75 ± 4.70</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 81)</td>
<td>16.40 ± 4.46</td>
<td>12.89 ± 4.68</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 80)</td>
<td>17.33 ± 5.33</td>
<td>12.83 ± 5.05</td>
<td>0.87</td>
</tr>
<tr>
<td>OQ-45 Interpersonal relations</td>
<td>Outpatient (n = 46)</td>
<td>21.30 ± 7.45</td>
<td>19.01 ± 7.89</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 81)</td>
<td>21.34 ± 5.99</td>
<td>17.65 ± 6.41</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 80)</td>
<td>23.59 ± 5.94</td>
<td>17.92 ± 6.83</td>
<td>0.89</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Outpatient (n = 46)</td>
<td>0.57 ± 0.28</td>
<td>0.67 ± 0.26</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 81)</td>
<td>0.47 ± 0.27</td>
<td>0.66 ± 0.26</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 80)</td>
<td>0.50 ± 0.27</td>
<td>0.70 ± 0.23</td>
<td>0.80</td>
</tr>
</tbody>
</table>

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D = EuroQol EQ-5D.
Chapter 5

Improvements were also observed in terms of psychosocial functioning and quality of life, as can be seen in Table 5.2. Effect sizes for these outcome measures were somewhat smaller compared to psychiatric symptoms, with effect sizes varying between 0.64 and 0.87 for social role functioning (OQ-45), between 0.30 and 0.89 for interpersonal relations (OQ-45), and between 0.37 and 0.80 for quality of life (EQ-5D).

Improvement between baseline and 18 months proved to be significant for patients within all setting groups on all four outcome measures (p < .05).

**Group comparisons**

After correction for observed pre-treatment differences by means of the multiple propensity score, the differences in improvement of psychiatric symptoms between outpatient and day hospital treatment and between day hospital and inpatient treatment were rather small, with $B = 0.11$ ($p = .44$) and $B = 0.18$ ($p = .14$), respectively (see Table 5.3). However, the difference in improvement between outpatient and inpatient treatment proved to be marginally significant in favour of inpatient treatment ($B = 0.30; p = .057$).

**Table 5.3.** Difference scores (unstandardised $B$) of change scores from baseline to 18 months, corrected for propensity score, all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment groups compared</th>
<th>$B$</th>
<th>95% CI</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Outpatient ($n = 46$) vs. day hospital ($n = 81$)</td>
<td>0.11</td>
<td>[-0.17, 0.40]</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 46$) vs. inpatient ($n = 80$)</td>
<td>0.30</td>
<td>[-0.01, 0.60]</td>
<td>.057</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 81$) vs. inpatient ($n = 80$)</td>
<td>0.18</td>
<td>[-0.06, 0.42]</td>
<td>.14</td>
</tr>
<tr>
<td>OQ-45 Social role</td>
<td>Outpatient ($n = 46$) vs. day hospital ($n = 81$)</td>
<td>-0.79</td>
<td>[-3.13, 1.54]</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 46$) vs. inpatient ($n = 80$)</td>
<td>-0.82</td>
<td>[-3.38, 1.73]</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 81$) vs. inpatient ($n = 80$)</td>
<td>-0.03</td>
<td>[-1.90, 1.84]</td>
<td>.98</td>
</tr>
<tr>
<td>OQ-45 Interpers. relations</td>
<td>Outpatient ($n = 46$) vs. day hospital ($n = 81$)</td>
<td>0.89</td>
<td>[-1.70, 3.49]</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 46$) vs. inpatient ($n = 80$)</td>
<td>2.09</td>
<td>[-0.65, 4.83]</td>
<td>.14</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 81$) vs. inpatient ($n = 80$)</td>
<td>1.19</td>
<td>[-0.96, 3.35]</td>
<td>.28</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Outpatient ($n = 46$) vs. day hospital ($n = 81$)</td>
<td>0.02</td>
<td>[-0.08, 0.11]</td>
<td>.71</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 46$) vs. inpatient ($n = 80$)</td>
<td>0.07</td>
<td>[-0.03, 0.17]</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 81$) vs. inpatient ($n = 80$)</td>
<td>0.05</td>
<td>[-0.03, 0.13]</td>
<td>.18</td>
</tr>
</tbody>
</table>

* Positive coefficients indicate that the last treatment group shown is superior, negative coefficients indicate that the first treatment group is superior.  GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D = EuroQol EQ-5D.
Group differences in the improvement of psychosocial functioning were smaller than the differences in the improvement of psychiatric symptoms (see Table 5.3). The same holds true for the differences in change scores of quality of life, with results that were far from significant.

From the inspection of the uncorrected results (i.e., without propensity score correction), it appeared that propensity score correction reduced the differences between treatment groups considerably. In other words, pre-treatment differences played an important role in the differential effectiveness observed.

**Discussion**

**Main findings**
This is the first study comparing the effectiveness of psychotherapy in three treatment settings in a large population of patients with cluster B PDs. Patients in outpatient, day hospital, and inpatient psychotherapy showed low dropout rates. They improved in all three treatment groups on psychiatric symptoms, psychosocial functioning, and quality of life at 18 months after baseline. Patients in inpatient psychotherapy showed the strongest improvement, particularly in psychiatric symptoms. This result proved to be marginally statistically significant ($p = .057$ for the comparison of inpatient and outpatient treatment) even when pre-treatment differences of patients were controlled for with the multiple propensity score.

**Strengths and limitations**
A clear strength of the present study is its external validity and clinical utility since it was conducted in regular clinical practice, not under experimental conditions (Hodgson, Bushe, & Hunter, 2007). Remarkably, dropout rates were low compared to other studies conducted in PD populations (Chiesa, Drahorad, & Longo, 2000). A second strength is the powerful statistical control of potential confounders, using the multiple propensity score methodology. Finally, a major asset of this study is its large sample size. All this enabled the comparison of outpatient, day hospital, and inpatient psychotherapeutic treatment while keeping sufficient statistical power.

Despite these strengths, the study had several limitations. First, even though we controlled for all observed pre-treatment differences, it cannot be ruled out that results were influenced by unobserved confounders. To minimise this risk as much as possible, a broad range of possible confounders was carefully selected and measured, based on both clinical and empirical knowledge (Bartak, et al., 2009), including
variables identified in the literature as significant predictors of therapy outcome or process, such as severity of baseline psychopathology, previous hospitalisation, and substance misuse (Gunderson et al., 2006; Links, Mitton, & Steiner, 1993; McGlashan, 1985; Ogrodniczuk et al., 2008; Plakun, 1991; Ryle & Golynkina, 2000). Only variables significantly related to treatment outcome were included in the propensity score. In the present sample (and contrary to previous research), this appeared not to be the case for previous hospitalisation and substance misuse. However, even when considerably reducing the possibility of important confounders being overlooked, not all possible variables could be covered in interviews and questionnaires at baseline and therefore several variables, such as self-harm (Chiesa & Fonagy, 2007), were not measured.

Second, for ethical reasons, a control group receiving no treatment at all was not included. This omission is mitigated by the fact that several previous studies have convincingly shown that specialised psychotherapeutic treatment provides better outcomes than various control conditions (e.g., waiting list controls; Bateman & Fonagy, 1999; Koons, et al., 2001).

Third, even though the study sample covers the whole range of cluster B PDs (borderline PD, narcissistic PD, histrionic PD, and antisocial PD), the majority of patients had a diagnosis of borderline PD. This makes the conclusions most applicable to this borderline group and only to a lesser extent to the other three diagnostic groups.

Fourth, the cut-off point between outpatient and day hospital treatment was arbitrary. We defined outpatient treatment as a low-intensive treatment with a maximum of two sessions of therapy per week. As soon as more therapy elements were added, we called it “day hospital” treatment. This implies that our group of day hospital treatments was a heterogeneous one, varying from one morning/afternoon per week to five days per week. Therefore, conclusions about the effectiveness of specific day hospital treatments have to be drawn carefully, bearing this cut-off score in mind.

Research implications
What are the implications of the present results for future research? First, inpatient treatment has been largely marginalised in the literature. Only a handful of studies investigated its usefulness, most of them with promising results (Bohus et al., 2004; Chiesa, et al., 2004; Fassbinder et al., 2007; Kleindienst et al., 2008; Kröger et al., 2006). Moreover, in the last two decades inpatient treatment for PDs has become infrequent in clinical practice (Chiesa, et al., 2009). In recent practice guidelines
(NICE; National Institute for Health and Clinical Excellence, 2009a, 2009b)—based on scientific knowledge—inpatient treatment is only recommended as a short-term crisis intervention. In a report of the National Institute for Mental Health in England (2003) on treatment of PDs it is also explicitly stated that there are “no plans to extend this level of residential provision”. In the light of the present results, however, specialised inpatient treatment deserves to be considered as a valuable treatment option for patients with cluster B PDs and future research should not a priori exclude this treatment setting from effectiveness studies. In a recent study on cluster C PDs, short inpatient treatment also emerged as the treatment with the best results, compared to other treatment modalities (Bartak et al., 2010). On the other hand, Chiesa et al. (2009) showed in a naturalistic study that specialised low-intensive treatment can equal (or even surpass) the results of traditional inpatient treatment. And indeed, in our study we did not find a significant difference between day hospital and inpatient treatment. Therefore, future research should focus more on the question of “ideal dosage” of treatment for patients with PD. As our study covers the whole range of cluster B PDs and therefore a heterogeneous group of patients in many aspects, results might be different across specific subgroups of patients (Digre, Reece, Johnson, & Thomas, 2009; Lorentzen & Høglend, 2008). For instance, very vulnerable patients may not be able to tolerate the therapeutic “pressure” that is often characteristic of full-time inpatient treatment or the subculture of patients’ living together (Chiesa, et al., 2000). This patient group might benefit much more from a long-term structured day hospital programme, for example MBT (Bateman & Fonagy, 2004). It is recommended to further investigate these potential matching factors. This will enable clinicians to make specific treatment recommendations for different subgroups of cluster B PD patients and to develop more sophisticated clinical practice guidelines.

Second, when randomised designs are impossible or unfeasible, important research questions have to be answered by means of non-randomised clinical trials. In these cases researchers have to control for as many patient variables as possible in a statistically sound way. This is essential in order not to draw premature and possibly misleading conclusions.

**Public health significance**

The marginally significant corrected results make it difficult to definitely determine the best treatment setting for patients with cluster B personality pathology. After propensity score correction, the favourable effect of inpatient psychotherapy on psychiatric symptoms is of marginal statistical significance, but still clinically
worthwhile. This effect cannot be explained by the influence of dropout rates, as treatment compliance was even lower in the inpatient group (10.0% dropout rate) compared to the outpatient group (0.0% dropout rate). However, the marginal statistical significance does not allow us to draw definite conclusions. Possibly, this may be caused by a lack of power, which could be solved by an even larger sample size. The effect might also become larger when the observation period is increased. A question for future research is therefore whether a certain treatment setting is superior over a longer-term perspective. It could be speculated that the outcome areas psychosocial functioning and quality of life have a slower pace of change compared to psychiatric symptoms and might also show a differential effectiveness of treatments at a later stage. Yet this remains conjecture and future research is needed urgently to shed more light on this question. Another future area of exploration should be cost-effectiveness. Expensive treatments can earn back the investment made and may even lead to cost-savings in the long run, because patients probably consume less additional forms of care after leaving an intensive (and effective) treatment. However, in the absence of long-term cost-effectiveness data an answer to this question cannot be provided here.

In conclusion, this study indicates that patients with cluster B PDs improve in outpatient, day hospital, and inpatient psychotherapeutic treatment. In addition, we observed a small trend towards larger improvements of psychiatric symptoms in the inpatient setting compared to the outpatient setting. Future studies are needed to further investigate the effect of different treatment dosages and possible differences in treatment effectiveness for different subgroups of patients. Meanwhile, inpatient therapy should still be considered a valuable option for patients with cluster B PDs and should receive as much attention as other treatment options, both in research and clinical practice.
Chapter 6

Patients with cluster A personality disorders in psychotherapy: An effectiveness study

Paranoid, schizoid, and schizotypal PD are associated with significant psychological and functional impairment and a poor long-time prognosis (Bender et al., 2001; Bornstein, Klein, Mallon, & Slater, 1988; McGlashan et al., 2005; Reich & Green, 1991; Seivewright, Tyrer, & Johnson, 2002; Skodol et al., 2005). Estimates for their prevalence in the general population range from 1.6 to 4.1% (Coid, Yang, Tyrer, Roberts, & Ulrich, 2006; Samuels et al., 2002; Torgersen, Kringlen, & Cramer, 2001), in psychiatric outpatients they are even more common (5.6% - 13.2%; Bornstein, et al., 1988; Zimmerman, Rothschild, & Chelminski, 2005). Therefore, it is important to investigate which treatment forms can produce significant and long-lasting improvement in the life of these patients.

Psychotherapy is recommended as the treatment of choice for patients with PD in general (Bartak, Soeteman, Verheul, & Busschbach, 2007). More specifically, clinical experts suggest psychotherapy as the first treatment option for cluster A PD patients (Gabbard, 2000; Stone, 1985; The Quality Assurance Project, 1990). Yet, psychotherapy effectiveness studies on the “odd” cluster A PD are rare. Most existing studies investigate the impact of cluster A PD on treatment effects of Axis I treatments or focus on the effectiveness of psychopharmacological treatment of schizotypal PD. It has been shown in these studies that the presence of cluster A PD, especially schizotypal PD, has a negative influence on the outcome of Axis I treatments of both medical and psychotherapeutic kind (Baer et al., 1992; Jenike, Baer, Minichiello, Schwartz, & Carey, 1986; Minichiello, Baer, & Jenike, 1987; Mulder, Joyce, Frampton, Luty, & Sullivan, 2006; Sato, Sakado, Sato, & Morikawa, 1994). Medication studies of schizotypal PD have shown that typical and atypical antipsychotics, and possibly antidepressants, can have a positive influence on distinguished symptoms, such as cognitive disturbance, derealisation, ideas of reference, anxiety, self-image, social and overall functioning, odd communication, obsessive-compulsive symptoms, and self-injury (Duggan, Huband, Smailagic, Ferriter, & Adams, 2008; Goldberg et al., 1986; Herpertz et al., 2007; Hymowitz, Frances, Jacobsberg, Sickles, & Hoyt, 1986; Ingenhoven, Lafay, Rinne, Passchier, & Duivenvoorden, 2010; Kapfhammer & Hilpens, 1998; Keshavan, Shad, Soloff, & Schooler, 2004; Koenigsberg et al., 2003; Markovitz, Calabrese, Schulz, & Meltzer, 1991; Serban & Siegel, 1984). The influence of medication on depressive symptoms in schizotypal patients is still controversial (Goldberg, et al., 1986; Keshavan, et al., 2004; Koenigsberg, et al., 2003; Serban & Siegel, 1984). There is no “agent of choice” (Coccaro, 1998) and medication side effects are common, especially in typical neuroleptics (Hymowitz, et al., 1986). Moreover, it is repeatedly stated that whatever the value of pharmacological treatment may be, the quality of the therapeutic relationship is one of the most crucial aspects in the treatment of cluster A PD patients (e.g., Kapfhammer & Hilpens, 1998).
Possible reasons for the paucity of research in this patient group might be the fact that cluster A PD patients seldom seek help from the psychiatric profession and the lack of consensus among clinicians about the effectiveness of psychotherapy in this patient population (Gabbard, 2000; Stone, 1993).

This lack of consensus is maintained because the few effectiveness studies of psychotherapy in cluster A PD patients show contradicting results. Plakun et al. (1985) studied former inpatients at the “Austin Riggs Center” of whom 43 had a cluster A PD. They were classified into four groups: (a) pure schizotypal PD patients ($n = 13$), (b) combined borderline and schizotypal PD patients ($n = 6$), (c) combined borderline with schizoid PD patients ($n = 5$), and (d) pure schizoid PD patients ($n = 19$). They found the highest improvement of general functioning in the combined borderline and schizotypal PD group. After a follow-up period of 14 years, this group functioned significantly better than the comparison group of schizophrenic patients, with a Global Assessment Scale score (GAS; Endicott, Spitzer, Fleiss, & Cohen, 1976) of 72.0. The other three cluster A PD groups also improved in their GAS score, but still functioned at a level comparable to schizophrenic patients at follow-up, with mean GAS scores between 55.6 and 66.6.

In their studies on day hospital treatment, a Norwegian research group found poor results regarding psychiatric symptoms and psychosocial functioning for patients with schizotypal PD ($n = 9$, with or without co-morbid borderline PD) (Karterud et al., 1992; Mehlum et al., 1991). Their study sample also included patients with paranoid PD ($n = 4$) and schizoid PD ($n = 5$), but their results were not reported separately. The same holds true for the small group of cluster A PD patients ($n = 4$) in a study of Winston et al. (1994). Gude and Vaglum (2001) in their study on inpatients on the other hand found no differences in improvement of psychiatric symptoms and occupational status between pure cluster A PD patients ($n = 21$) and both pure cluster B and pure cluster C PD patients. However, their sample of cluster A PD patients included only few patients with schizotypal PD.

The largest study reporting on cluster A PD patients is the study on day hospital treatment conducted by Karterud et al. (2003). Their sample contained 132 cluster A PD patients (treatment completers; $n = 16$ for schizotypal PD as “main” diagnosis, $n = 8$ for schizoid PD, and $n = 108$ for paranoid PD), including cluster A PD patients with co-morbid cluster B and/or C PD. Their findings indicated that patients with cluster A PDs had a high dropout rate in day hospital treatment (23.9%) and that patients with paranoid and schizotypal PD showed only modest treatment gains regarding psychiatric symptoms, quality of life, and overall functioning. Patients
with paranoid PD and co-morbid borderline PD had significantly worse results at 1-year follow-up than patients with borderline PD alone. As the group of patients with schizoid PD was too small, their follow-up results were not reported.

Given that psychotherapy studies in cluster A PD patients are scarce and the evidence is conflicting, it is important to investigate further if psychotherapy can bring about change for these vulnerable patients. There are substantial differences between treatment options in terms of time and money. That makes it relevant for both patients and society to know which setting of psychotherapy (outpatient, day hospital, or inpatient treatment) is sufficiently efficient to produce positive change. Yet, no study has compared the outcomes of different treatment settings for patients with cluster A PDs.

The aim of the present study is to explore the effects of psychotherapy on patients with cluster A PDs. More specifically, we want to answer the following research questions:

1. What is the change in psychiatric symptoms, social and interpersonal functioning, and quality of life in patients with cluster A PDs 18 months after baseline?
2. What is the effect of different treatment settings, i.e., outpatient, day hospital, and inpatient treatment?

Method

In this study we used a prospective naturalistic study design. To correct for the influence of selection bias due to non-randomisation, we used the propensity score method (see below). For details of the method section, we refer to our earlier studies on cluster C and B PD patients (Bartak et al., in press; Bartak et al., 2010).

Participants

Participants \((N = 57)\) were recruited from consecutive admissions to six mental health care centres in the Netherlands. The institutions offer outpatient, day hospital, and/or inpatient psychotherapeutic treatment for patients with personality pathology. From March 2003 to March 2006, 1,379 patients completed the intake and screening procedure, and were selected for treatment (see Figure 6.1). The intake and screening data were collected for all applicants, irrespective of study participation later on. The data obtained from this initial assessment served as baseline data for our study. As it was part of the standard screening procedure, informed consent for the baseline data collection was not mandatory under Dutch law.
Figure 6.1. Patient flow

Assessed for eligibility  
\( N = 1379 \)

- Refused to participate  
  \( n = 100 \)
- Excluded due to other reasons  
  \( n = 164 \)  
  (Logistic reasons \( n = 31 \)  
  Missing/unreliable data \( = 133 \))

Did not meet inclusion criteria  
\( n = 146 \)

- Met exclusion criteria  
  \( n = 9 \)

Enrolled  
\( n = 960 \)

Dropped out prematurely\(^*\)  
\( n = 38 \)

No DSM-IV cluster A PD diagnosis  
\( n = 851 \)

Allocated  
\( n = 71 \)

Outpatient treatment  
\( n = 24 \)

Lost to follow-up  
\( n = 4 \)

Analyzed  
\( n = 20 \)

Day hospital treatment  
\( n = 27 \)

Lost to follow-up  
\( n = 8 \)

Analyzed  
\( n = 19 \)

Inpatient treatment  
\( n = 20 \)

Lost to follow-up  
\( n = 2 \)

Analyzed  
\( n = 18 \)

\(^*\) These patients did not receive a “minimal effective dose” of treatment, defined as two sessions for outpatients and two treatment days for day hospital patients or inpatients, and were therefore not included in the initial study sample.
Of these 1,379 patients, 146 (10.6%) were excluded from the study for not meeting one of the following inclusion criteria: age between 18 and 70 years (n = 13), significant personality pathology (n = 34), and referral for psychotherapeutic treatment aimed at personality problems (n = 99). Nine patients (0.7%) met one of the following exclusion criteria: insufficient command of the Dutch language (n = 6), organic cerebral impairment (n = 1), mental retardation (n = 1), and schizophrenia (n = 1). This left 1,224 participants, of whom 31 (2.5%) could not participate due to logistic reasons. A total of 133 patients (10.9%) had to be excluded due to missing or unreliable baseline data during the intake and screening procedure. For the majority of these 133 patients (n = 106) no standardised Axis II diagnosis was obtained. This was mostly due to a practical problem, i.e., a shortage of interviewers at the start of the study (n = 101). Because of this logistic reason, it can be assumed that these data were “missing completely at random” and therefore they do not threaten internal validity (as they were unrelated to specific patient characteristics). Five patients had an unreliable Axis II diagnosis as determined by the interviewer. A few patients did not return their assessment booklet during the intake procedure (n = 27). Thirty-eight patients (3.1%) received less than two treatment sessions of outpatient therapy or less than two days of inpatient or day hospital therapy. They were therefore excluded beforehand from the study sample in which we only included patients with a “minimal effective dose” of treatment. One hundred patients (8.2%) refused to participate. The remaining 922 patients were informed about the study and its procedure, provided written informed consent, and entered the study. Of those, 71 patients (7.7%) had one or more cluster A PDs.

In the absence of explicit guidelines for treatment assignment in PDs (Vervaeke & Emmelkamp, 1998), the treatment selection procedure was based on the expert opinion of clinicians who used their clinical experience combined with patient data from standardised instruments. To clarify the criteria used for the assignment process, our research group had recently conducted a study with intake clinicians from the participating treatment centres. We found evidence of substantial consensus among clinicians concerning the criteria used for treatment decision making. Variables guiding the treatment selection process include the foci of problems, ego strength, symptom severity, psychological mindedness, relational capacities, and quality of defence mechanisms (van Manen et al., 2008).

The 71 cluster A PD patients were assigned to one of three setting groups, based on the regular treatment assignment done by expert clinicians at the clinical sites: 24 patients to outpatient, 27 patients to day hospital, and 20 patients to inpatient
treatment. Follow-up data were not available for 14 patients (19.7%). There was no
difference in psychiatric symptoms at baseline between patients with follow-up data
and those without. The final sample consisted of 57 patients who were included in
the analyses.

**Treatment**
The six mental health care centres offer a variety of psychotherapeutic treatments
tailored to a PD patient population. Their treatments differ according to several
features. As this study focused on different treatment settings, the following three
treatment groups were compared:

- Patients in outpatient treatment ($n = 20, 35.1\%$ of the study sample). These
  patients came for individual ($75.0\%$) or group ($25.0\%$) psychotherapy
  sessions for up to two sessions (mean = 1.0 sessions/week, $SD = 0.6$, median = 0.9)
  per week (mean duration = 13.3 months, $SD = 6.2$, median = 12.0). Four health care centres offered outpatient treatment.

- Patients in day hospital treatment ($n = 19, 33.3\%$ of the study sample).
  These patients came at least one morning / afternoon per week (mean = 3.3
  days/week, $SD = 1.6$, median = 3.0) and received different forms of
  psychotherapeutic and psychosocial treatment, but slept at home (mean
  duration = 10.3 months, $SD = 4.5$, median = 9.0). Five health care centres
  offered day hospital treatment.

- Patients in inpatient treatment ($n = 18, 31.6\%$ of the study sample).
  These patients stayed at the institutions five days a week and received
  different forms of psychotherapeutic and psychosocial treatment (mean
  duration = 8.6 months, $SD = 2.4$, median = 8.5). Three health care centres
  offered inpatient treatment.

Outpatient treatments consisted of individual or group psychotherapy sessions of
various theoretical orientations (50% eclectic, 20% psychodynamic, 20% cognitive-
behavioural, 10% other). Day hospital and inpatient programs typically consisted of
group psychotherapy as a core element, mostly in combination with one or more non-
verbal or expressive group therapies, individual psychotherapy, sociotherapy within
the therapeutic community, coaching for social problems, community meetings,
and/or pharmacological treatment. The therapists were all licensed psychiatrists or
psychologists. On average, they had 14.9 years ($SD = 10.1$) of postgraduate clinical
experience. The treatments under study can be considered highly representative for
regular clinical practice in The Netherlands, as therapists did not receive specific
training for this study and treatment integrity was not monitored. The study protocol
was approved by the Medical Ethics Committee of the Erasmus University Medical Centre in Rotterdam.

Assessments
Baseline measures
An extensive standard assessment battery of instruments was administered to the patients before treatment assignment. PDs were measured using the Dutch version of the Structured Interview for DSM-IV Personality (SIDP-IV; de Jong, Derks, van Oel, & Rinne, 1996; Pfuhl, Blum, & Zimmerman, 1997).

To measure patient characteristics at baseline, the assessment battery also included three self-report instruments. The first of those was the Dutch version of the Dimensional Assessment of Personality Pathology-Basic Questionnaire (DAPP-BQ; Livesley & Jackson, 2002; van Kampen, 2002), for measuring type and degree of personality pathology. We used patients’ scores on this questionnaire on the four higher order factors: (a) emotional dysregulation, (b) dissocial behaviour, (c) inhibition, and (d) compulsivity. To measure the severity of personality pathology we used the five higher-order domains of the Severity Indices of Personality Problems, Dutch version (SIPP; Verheul et al., 2008): (a) self-control, (b) social concordance, (c) identity integration, (d) relational capacities, and (e) responsibility. To measure patients’ motivation for treatment, we used the two scales of the Motivation for Treatment Questionnaire (MTQ-8; van Beek & Verheul, 2008): need for help and readiness to change.

Outcome measures
The primary outcome measure was general psychiatric symptomatology. This was measured using the Dutch version of the Brief Symptom Inventory (BSI; de Beurs & Zitman, 2006; Derogatis & Melisaratos, 1983), a validated self-report scale derived from the Symptom Checklist 90-Revised (SCL-90-R; Arrindell & Ettema, 2003; Derogatis, 1992). In this study, we used the mean score of the 53 items of the BSI, i.e., the Global Severity Index (GSI), ranging from 0 to 4, with higher scores indicating more problems. Psychosocial functioning was measured with two subscales of the Outcome Questionnaire-45, Dutch version (OQ-45; Lambert et al., 1996): interpersonal relations and social role functioning. The total scores of these two scales range between 0 and 44, and between 0 and 36, respectively, with higher scores indicating more problems. Health related quality of life was measured using two scales of the EuroQol EQ-5D, Dutch version (EQ-5D; Brooks, Rabin, & de Charro, 2003), with higher scores indicating a higher quality of life: the EQ-5D index,
which represents the societal value (i.e., the valuation of a particular health state as rated by the society) of quality of life with a value between 0 and 1; and the EQ-Visual Analogue Scale (EQ-VAS) which represents the patients own value of quality of life on a scale from 0 to 100. All five outcome measures, CSI, OQ-45 social role, OQ-45 interpersonal relations, EQ-5D index, and EQ-VAS, were assessed at baseline and several follow-up moments. Three of the six treatment centres conducted their follow-up measures at 12, 24, and 36 months after baseline; the other three treatment centres conducted their follow-up measures at the end of treatment, subsequently six and 12 months after that, and again at 36 months after baseline. The use of different assessment points was due to logistic reasons in the participating centres. As some follow-up points were related to the end of treatment and treatment duration varied considerably, we had no common measurement point for all patients to determine mid-term outcome. This was taken into account by choosing multi-level modelling as the statistical method for the analyses. Multi-level modelling enabled us to make use of all the available data collected for every patient at multiple assessment points between baseline and 36 months and to reliably estimate change at 18 months.

Statistical analyses
We first examined the uncorrected results on all five outcome measures estimated at 18 months after baseline, thus without applying the propensity score method. We used multi-level modelling to deal with (a) the dependency of repeated measures on the same subject in time, and (b) longitudinal data with observations unequally spaced in time (see “Outcome measures”). To estimate the uncorrected treatment effect at 18 months after baseline we used a random intercept and random slope model with time as level I and patient number as level II. This resulted in a model with the following independent variables: dummy variables indicating treatment group membership, time, and interaction between group membership and time. In these analyses, we used all available follow-up data from baseline to 36 months, but since we focused on mid-term outcome, we used the estimated change scores at 18 months after baseline, based on the results of the model. Subsequently, we calculated within-group effect sizes Cohen’s $d$ (Cohen, 1988) to describe change from baseline to 18 months per treatment group (outpatient, day hospital, inpatient).

However, treatment groups cannot be compared directly based on the uncorrected results in this non-randomised clinical trial as these findings might be confounded by initial patient differences. To adjust for these differences and to avoid bias in effect estimation, we used the propensity score method. The classic propensity score is defined as the conditional probability of assignment to one of two treatment
groups given a set of observed pre-treatment variables (Rosenbaum & Rubin, 1983). Propensity scores are used to reduce selection bias by equating groups based on these variables. Since we had to compare three treatment groups, we included a “multiple propensity score” in our analysis. The multiple propensity score is an extension of the classic propensity score to more than two treatment groups (Imbens, 2000). To identify relevant confounders, we considered a long list of social, economic, and diagnostic variables carefully selected by both clinicians and researchers, based on the literature and clinical knowledge (Bartak et al., 2009). All variables significantly related to a specific outcome were used to estimate the multiple propensity scores in a multinomial regression analysis, with ‘group membership’ as a dependent variable (see Table 6.1 for the variables included in the GSI propensity score). A major advantage of the propensity score method, as compared to other correction techniques, is the fact that the overlap in propensity score distributions (and thus the overlap in relevant variables) between treatment groups can be easily judged and visualised (Spreeuwenberg et al., 2010). If the distribution of propensity scores shows large overlap in the different treatment groups, the groups are readily comparable. If, however, overlap is insufficient, groups differ too much and their direct comparison might yield results which cannot unequivocally be attributed to a treatment effect. For a detailed description of the propensity score method and its use in psychotherapy research see Bartak et al. (2009) and Spreeuwenberg et al. (2010).

To compare change in outcome variables across treatment groups adjusted for baseline patient characteristics, a more sophisticated multi-level model, now including multiple propensity scores, was used. Dependent variables were all available change scores observed during follow-up for each of the outcome measures. The following independent variables were entered in the initial model: dummy variables indicating group membership, time, the multiple propensity scores, and their interactions. Then independent variables were eliminated from the model by backward selection to obtain a final best-fit model. This model estimated differences in change scores at 18 months after baseline in pair-wise comparisons of the three treatment groups.

Follow-up response was high, enhancing the robustness of the estimations at 18 months: 77.2% of the patients had three or more follow-up measurements between baseline and 36 months. The analyses were performed using SPSS 17.0 for data preparation and Proc Mixed of SAS 9.1.3 for multi-level modelling (SAS Institute Inc, Cary, North Carolina, USA).
Table 6.1. Variables used for propensity score estimation, outcome GSI

<table>
<thead>
<tr>
<th>Variable</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAPP-BQ Emotional dysregulation</td>
<td>Unstable affective responding, interpersonal problems</td>
</tr>
<tr>
<td>DAPP-BQ Compulsivity</td>
<td>Compulsivity and absence of oppositional behaviour</td>
</tr>
<tr>
<td>MTQ-8 Need for help</td>
<td>Patient’s expressed desire for external help</td>
</tr>
<tr>
<td>SIPP Self-control</td>
<td>The capacity to tolerate, use, and control one’s own emotions and impulses</td>
</tr>
<tr>
<td>OQ-45 Symptom distress</td>
<td>Level of symptom distress</td>
</tr>
<tr>
<td>OQ-45 Relational functioning</td>
<td>Level of interpersonal functioning</td>
</tr>
<tr>
<td>SIDP-IV Cluster B PDs</td>
<td>Number of cluster B PDs</td>
</tr>
<tr>
<td>SIDP-IV Cluster C PDs</td>
<td>Number of cluster C PDs</td>
</tr>
<tr>
<td>SIDP-IV Dimensional score cluster B PDs</td>
<td>Dimensional score of cluster B PD characteristics</td>
</tr>
<tr>
<td>SIDP-IV Dimensional score cluster C PDs</td>
<td>Dimensional score of cluster C PD characteristics</td>
</tr>
<tr>
<td>SIDP-IV Total dimensional score</td>
<td>Total dimensional score of PD characteristics</td>
</tr>
<tr>
<td>Drug abuse</td>
<td>History of drug abuse</td>
</tr>
<tr>
<td>Borderline PD</td>
<td>Diagnosis of borderline PD</td>
</tr>
<tr>
<td>Narcissistic PD</td>
<td>Diagnosis of narcissistic PD</td>
</tr>
</tbody>
</table>

DAPP-BQ = Dimensional Assessment of Personality Pathology-Basic Questionnaire, MTQ-8 = Motivation for Treatment Questionnaire, SIPP = Severity Indices of Personality Problems, OQ-45 = Outcome Questionnaire-45, SIDP-IV = Structured Interview for DSM-IV Personality.

Results

Sample characteristics
Of the 57 patients, 70.2% were female (see Table 6.2). The mean age was 29.4 years (SD = 8.2) and 78.9% were unmarried. The level of education was low for 43.9%, medium for 10.5%, and high for 45.6%. The majority (86.0%) had a diagnosis of paranoid PD. Only a minority (15.8%) had pure cluster A PD, while the remaining patients had cluster A PD and co-morbid cluster B and/or C PD.
### Table 6.2. Baseline characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Outpatient (n = 20)</th>
<th>Day hospital (n = 19)</th>
<th>Inpatient (n = 18)</th>
<th>Total (n = 57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Male</td>
<td>40.0</td>
<td>10.5</td>
<td>38.9</td>
<td>29.8</td>
</tr>
<tr>
<td>% Female</td>
<td>60.0</td>
<td>89.5</td>
<td>61.1</td>
<td>70.2</td>
</tr>
<tr>
<td>Age* (Mean (SD))</td>
<td>33.6 (8.6)</td>
<td>28.2 (8.0)</td>
<td>25.9 (6.1)</td>
<td>29.4 (8.2)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Low</td>
<td>55.0</td>
<td>57.9</td>
<td>16.7</td>
<td>43.9</td>
</tr>
<tr>
<td>% Medium</td>
<td>10.0</td>
<td>10.5</td>
<td>11.1</td>
<td>10.5</td>
</tr>
<tr>
<td>% High</td>
<td>35.0</td>
<td>31.6</td>
<td>72.2</td>
<td>45.6</td>
</tr>
<tr>
<td>Marital status*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Unmarried</td>
<td>60.0</td>
<td>84.2</td>
<td>94.4</td>
<td>78.9</td>
</tr>
<tr>
<td>% Married</td>
<td>20.0</td>
<td>15.8</td>
<td>5.6</td>
<td>14.0</td>
</tr>
<tr>
<td>% Widowed/divorced</td>
<td>20.0</td>
<td>0.0</td>
<td>0.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Child care</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Yes</td>
<td>30.0</td>
<td>10.5</td>
<td>5.6</td>
<td>15.8</td>
</tr>
<tr>
<td>% No</td>
<td>70.0</td>
<td>89.5</td>
<td>94.4</td>
<td>84.2</td>
</tr>
<tr>
<td>Cluster A PD diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Paranoid PD</td>
<td>90.0</td>
<td>94.7</td>
<td>72.2</td>
<td>86.0</td>
</tr>
<tr>
<td>% Schizoid PD</td>
<td>5.0</td>
<td>5.3</td>
<td>16.7</td>
<td>8.8</td>
</tr>
<tr>
<td>% Schizotypal PD</td>
<td>5.0</td>
<td>5.3</td>
<td>11.1</td>
<td>7.0</td>
</tr>
<tr>
<td>PD diagnosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Pure cluster A</td>
<td>15.0</td>
<td>10.5</td>
<td>22.2</td>
<td>15.8</td>
</tr>
<tr>
<td>% Cluster A &amp; B</td>
<td>20.0</td>
<td>0.0</td>
<td>16.7</td>
<td>12.3</td>
</tr>
<tr>
<td>% Cluster A &amp; C</td>
<td>40.0</td>
<td>31.6</td>
<td>22.2</td>
<td>31.6</td>
</tr>
<tr>
<td>% Cluster A, B &amp; C</td>
<td>25.0</td>
<td>57.9</td>
<td>38.9</td>
<td>40.4</td>
</tr>
<tr>
<td>GSI* (Mean (SD))</td>
<td>1.4 (0.7)</td>
<td>2.0 (0.6)</td>
<td>2.0 (0.8)</td>
<td>1.8 (0.8)</td>
</tr>
<tr>
<td>OQ-45 Social role (Mean (SD))</td>
<td>15.7 (4.7)</td>
<td>18.8 (5.4)</td>
<td>17.7 (5.5)</td>
<td>17.4 (5.3)</td>
</tr>
<tr>
<td>OQ-45 Interp. relations (Mean (SD))</td>
<td>21.4 (5.5)</td>
<td>23.7 (6.5)</td>
<td>25.5 (4.7)</td>
<td>23.5 (5.8)</td>
</tr>
<tr>
<td>EQ-5D index* (Mean (SD))</td>
<td>0.7 (0.2)</td>
<td>0.4 (0.3)</td>
<td>0.5 (0.3)</td>
<td>0.6 (0.3)</td>
</tr>
<tr>
<td>EQ-VAS (Mean (SD))</td>
<td>61.9 (18.4)</td>
<td>55.4 (19.1)</td>
<td>53.2 (16.5)</td>
<td>57.0 (18.1)</td>
</tr>
<tr>
<td>Psychotropic medication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Yes</td>
<td>60.0</td>
<td>63.2</td>
<td>55.6</td>
<td>59.6</td>
</tr>
<tr>
<td>% No</td>
<td>40.0</td>
<td>36.8</td>
<td>44.4</td>
<td>40.4</td>
</tr>
</tbody>
</table>

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D index = EuroQol EQ-5D index, EQ-VAS = EuroQol Visual Analogue Scale.

* Significant group differences (p <= .05)
Figure 6.2. Overlap in propensity score distributions, outcome GSI
As can be seen in Table 6.2, patients in the three treatment groups differed substantially on a number of baseline variables. Significant differences appeared in age, marital status, baseline severity of psychiatric symptoms, and baseline severity of quality of life. These substantial differences complicate the adjustment of the results with the propensity score. Indeed, the overlap in the distribution of propensity scores was imperfect (see Figure 6.2) and the results of these three groups of patients were not readily comparable, even after applying propensity score correction. One has to keep that in mind when judging possible group differences in change scores.

**Treatment compliance**
Before start of treatment, every patient received an allocation to a certain treatment dosage in terms of setting and duration. According to our registration at the end of treatment, about one quarter of all patients (n = 14) underwent exactly the intended treatment regarding both setting and exact duration (25.0% of the outpatient group, 31.6% of the day hospital group, and 16.7% of the inpatient group) and 94.7% of the patients (n = 54) stayed in the assigned setting group (100.0% of the outpatient group, 94.7% of the day hospital group, and 88.9% of the inpatient group). Of the 43 patients whose received treatment duration deviated from their intended treatment duration, 17 (39.5% of the deviating patients) stayed in treatment shorter than planned (40.0% of the deviating patients in the outpatient group, 46.2% in the day hospital group, and 33.3% in the inpatient group). Of these 17 patients, 10 decided in agreement with their therapist that treatment was no longer beneficial or necessary, whereas seven patients dropped out of treatment prematurely or were forced to leave earlier by the staff. From these seven dropouts, one patient was from the outpatient group (5.0% of the total outpatient group), four patients were from the day hospital group (21.1% of the total day hospital group), and two patients were from the inpatient group (11.1% of the total inpatient group). All dropouts were treatment dropouts (and no study dropouts) who completed follow-up measures and were included in the ITT analyses.

**Treatment outcome**
Patients in the day hospital and inpatient group showed larger improvements than patients in the outpatient group (see Table 6.3 and Figure 6.3), when inspecting the effect sizes for the three different treatment groups without propensity score correction. More specifically, patients of the day hospital and inpatient group showed significant improvement after 18 months in terms of psychiatric symptoms (GSI), the primary outcome measure (p < .0001), whereas patients of the outpatient group did not (p = .16).
Table 6.3. Uncorrected outcomes (mean ± SD) and effect sizes in three treatment groups for all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment group</th>
<th>Baseline</th>
<th>18 months</th>
<th>Within-group effect size, Cohen’s d</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Outpatient (n = 20)</td>
<td>1.37 ± 0.68</td>
<td>1.16 ± 0.88</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 19)</td>
<td>1.99 ± 0.62</td>
<td>1.14 ± 0.69</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 18)</td>
<td>1.97 ± 0.82</td>
<td>1.01 ± 0.73</td>
<td>1.27</td>
</tr>
<tr>
<td>OQ-45 Social role</td>
<td>Outpatient (n = 20)</td>
<td>15.68 ± 4.68</td>
<td>11.97 ± 6.23</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 19)</td>
<td>18.84 ± 5.36</td>
<td>13.12 ± 5.26</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 18)</td>
<td>17.74 ± 5.54</td>
<td>11.38 ± 6.33</td>
<td>1.10</td>
</tr>
<tr>
<td>OQ-45 Interpersonal relations</td>
<td>Outpatient (n = 20)</td>
<td>21.39 ± 5.50</td>
<td>18.65 ± 9.41</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 19)</td>
<td>23.67 ± 6.50</td>
<td>17.91 ± 7.40</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 18)</td>
<td>25.52 ± 4.74</td>
<td>17.31 ± 8.12</td>
<td>1.27</td>
</tr>
<tr>
<td>EQ-5D index</td>
<td>Outpatient (n = 20)</td>
<td>0.67 ± 0.19</td>
<td>0.76 ± 0.20</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 19)</td>
<td>0.45 ± 0.28</td>
<td>0.67 ± 0.25</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 18)</td>
<td>0.53 ± 0.30</td>
<td>0.69 ± 0.26</td>
<td>0.59</td>
</tr>
<tr>
<td>EQ-VAS</td>
<td>Outpatient (n = 20)</td>
<td>61.90 ± 18.36</td>
<td>62.55 ± 17.20</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Day hospital (n = 19)</td>
<td>55.42 ± 19.09</td>
<td>72.06 ± 13.72</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>Inpatient (n = 18)</td>
<td>53.17 ± 16.48</td>
<td>66.13 ± 19.31</td>
<td>0.74</td>
</tr>
</tbody>
</table>

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D index = EuroQol EQ-5D index, EQ-VAS = EuroQol Visual Analogue Scale.

Significant improvements for patients from the day hospital and inpatient group were also observed for social and interpersonal functioning (OQ-45) and for quality of life (EQ-5D index and EQ-VAS). The single outcome measure with significant improvements after 18 months for patients of the outpatient group was OQ-45 Social role (p = .02). The difference between the outpatient group on the one hand and the day hospital and inpatient group on the other hand was especially striking for the EQ-VAS score. Patients in the outpatient group started healthier and hardly improved (effect size = 0.04), whereas patients in the day hospital and inpatient group started treatment less healthy and subsequently improved substantially with effect sizes of 1.03 and 0.74, respectively.

**Group comparisons of different settings**

The group comparisons corrected with the propensity score method confirmed the superiority of day hospital and inpatient treatment in terms of improvement.
of psychiatric symptoms, as can be seen in Table 6.4. After correction for observed pre-treatment differences by means of the multiple propensity score, the differences in improvement between outpatient and day hospital treatment and between outpatient and inpatient treatment proved to be significant with $B = 0.52$ ($p = .046$) and $B = 0.73$ ($p = .01$), respectively. The differences in improvement of psychiatric symptoms (GSI) between day hospital and inpatient treatment were minimal, with $B = 0.22$ ($p = .41$).

Group differences in the improvement of psychosocial functioning were smaller than the differences in the improvement of psychiatric symptoms, with results that were far from significant. The same holds true for the differences in change scores of the EQ-5D quality of life index. However, the difference in change scores of the EQ-VAS between outpatient and day hospital treatment was statistically significant ($B = 23.82; p = .02$), favouring day hospital treatment.

Taken together, the corrected results suggest most of all a superiority of day hospital and inpatient treatment, as compared to outpatient treatment, with regard to the improvement of psychiatric symptoms 18 months after baseline. However, as mentioned above, these results have to be interpreted cautiously as it appeared from the limited overlap of propensity score distributions that the three patient groups consisted of substantially different patients.
Table 6.4. Difference scores (unstandardised B) of change scores from baseline to 18 months, corrected for propensity score, all outcome variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Treatment groups compared</th>
<th>$B^*$</th>
<th>95% CI</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSI</td>
<td>Outpatient ($n = 20$) vs. day hospital ($n = 19$)</td>
<td>0.52</td>
<td>[0.01, 1.03]</td>
<td>.046</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 20$) vs. inpatient ($n = 18$)</td>
<td>0.73</td>
<td>[0.16, 1.30]</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 19$) vs. inpatient ($n = 18$)</td>
<td>0.22</td>
<td>[-0.31, 0.74]</td>
<td>.41</td>
</tr>
<tr>
<td>OQ-45</td>
<td>Outpatient ($n = 20$) vs. day hospital ($n = 19$)</td>
<td>-0.58</td>
<td>[-4.89, 3.73]</td>
<td>.79</td>
</tr>
<tr>
<td>Social role</td>
<td>Outpatient ($n = 20$) vs. inpatient ($n = 18$)</td>
<td>0.97</td>
<td>[-3.24, 5.18]</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 19$) vs. inpatient ($n = 18$)</td>
<td>1.55</td>
<td>[-2.42, 5.53]</td>
<td>.44</td>
</tr>
<tr>
<td>OQ-45</td>
<td>Outpatient ($n = 20$) vs. day hospital ($n = 19$)</td>
<td>0.68</td>
<td>[-4.41, 5.76]</td>
<td>.79</td>
</tr>
<tr>
<td>Interpers. relations</td>
<td>Outpatient ($n = 20$) vs. inpatient ($n = 18$)</td>
<td>2.87</td>
<td>[-2.27, 8.02]</td>
<td>.27</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 19$) vs. inpatient ($n = 18$)</td>
<td>2.20</td>
<td>[-2.50, 6.89]</td>
<td>.35</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>Outpatient ($n = 20$) vs. day hospital ($n = 19$)</td>
<td>0.01</td>
<td>[-0.12, 0.15]</td>
<td>.83</td>
</tr>
<tr>
<td>index</td>
<td>Outpatient ($n = 20$) vs. inpatient ($n = 18$)</td>
<td>0.03</td>
<td>[-0.11, 0.16]</td>
<td>.68</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 19$) vs. inpatient ($n = 18$)</td>
<td>0.01</td>
<td>[-0.11, 0.14]</td>
<td>.83</td>
</tr>
<tr>
<td>EQ-VAS</td>
<td>Outpatient ($n = 20$) vs. day hospital ($n = 19$)</td>
<td>23.82</td>
<td>[3.92, 43.72]</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>Outpatient ($n = 20$) vs. inpatient ($n = 18$)</td>
<td>-0.20</td>
<td>[-36.39, 36.00]</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>Day hospital ($n = 19$) vs. inpatient ($n = 18$)</td>
<td>-24.01</td>
<td>[-61.88, 13.85]</td>
<td>.21</td>
</tr>
</tbody>
</table>

* Positive coefficients indicate that the last treatment group shown is superior, negative coefficients indicate that the first treatment group is superior.

GSI = Global Severity Index (Brief Symptom Inventory), OQ-45 = Outcome Questionnaire-45, EQ-5D index = EuroQol EQ-5D index, EQ-VAS = EuroQol Visual Analogue Scale.

Discussion

Main findings
This study examined the effect of psychotherapy on patients with cluster A PDs. The most important conclusion is that cluster A personality pathology per se does not seem to be an impediment to benefit from psychotherapeutic treatment.

The majority of the study sample presented with a diagnosis of paranoid PD and showed high co-morbidity with the two other PD clusters. When comparing the three treatment groups, it appeared that—with regard to psychiatric symptoms—patients in the day hospital and inpatient group improved more than patients in the outpatient group.
Strengths and limitations

Strengths of the present study are its considerable sample size for a cluster A PD population and its naturalistic design, which made it possible to gain insight in treatments as given in daily clinical practice. Furthermore, the attempt of powerful statistical control of potential confounders, using the multiple propensity score methodology, is unique in the cluster A PD literature.

Nevertheless, this study also had limitations. First, due to substantial baseline differences of patients in the three treatment groups, a direct comparison of the three settings was difficult to conduct. Higher treatment gains cannot readily be attributed to a certain treatment, as variables other than treatment group might have played a role. That outpatients improve less may be due to the fact that they are older at the start of treatment or have a heavier burden of care responsibilities at home. The present study cannot rule out that patient characteristics played a role in the explanation of the results, as overlap of propensity score distributions was limited.

Second, even though we controlled for all observed pre-treatment differences, it cannot be ruled out that results were influenced by other, unobserved confounders. To minimise this last risk as much as possible, a broad range of possible confounders was carefully selected and measured, based on both clinical and empirical knowledge (Bartak, et al., 2009). If related to outcome, these variables were included in the propensity score.

Third, the majority of patients had a diagnosis of paranoid PD which makes its results mainly applicable to this diagnostic group and to a lesser extent to patients with schizotypal and schizoid PD.

Fourth, the high co-morbidity with other PD clusters makes it questionable to attribute the treatment gains to an improvement in cluster A pathology. Possibly the improvement observed in our study is due to advances made in a different area of psychiatric impairment. Nevertheless, the main conclusion stays valid: Improvement in different areas of life is possible for patients with cluster A PDs who undergo psychotherapeutic treatment.
Clinical significance

The differences in improvement between outpatients on the one hand and day hospital patients and inpatients on the other hand are striking, especially compared to the poor results of day hospital treatment in cluster A PD patients found by Karterud et al. (2003). These authors found an effect size of 0.23 (versus 1.33 in the present study), when measuring change in GSI-scores of paranoid and schizotypal patients from baseline to follow-up (one year after discharge with a mean treatment duration of +/- five months, i.e., comparable to our 18 months after baseline). Might day hospital and inpatient treatment nevertheless be the preferable treatment options for this patient group? The present data suggest this conclusion, especially because these patients start off worse compared to outpatients and seem to improve more during treatment. Moreover, this effect cannot be explained by the influence of dropout rates, as treatment compliance was even lower in the day hospital (21.1% dropout rate) and inpatient group (11.1% dropout rate) compared to the outpatient group (5.0% dropout rate). The day hospital dropout rate is comparable to the one of Karterud et al. (23.9%). One possible explanation for the differences between the two studies might be that our treatment duration was twice as long as the one of Karterud et al. Maybe, cluster A PD patients need a longer treatment duration to experience change. A complication is that regression to the mean might also explain a part of the observed effect, as the patients with the highest baseline severity showed the highest improvement. For ethical reasons we were not able to compare our findings to a control group receiving no treatment at all. Clearly, more research is needed to further elucidate these findings before drawing major conclusions. The only conclusion we can draw from our findings with confidence is that cluster A pathology does not seem to be a contra-indication for psychotherapy.

Implications for research and public health

What are the implications of the present results for future research, clinical practice, and public health policy? We found that cluster A pathology does not seem to form an impediment for psychotherapeutic treatment and that gains in different areas of life can be observed in patients with cluster A PDs. Bearing the limitations of this pioneering study in mind, this might be the first step towards a revised vision of this patient group in mental health practice. Some twenty years ago, patients with borderline PD were considered as “untreatable”, whereas now a multitude of treatment protocols have been developed for this patient group and research on their implementation showed encouraging results (e.g., Bateman & Fonagy, 2001; Clarkin, Levy, Lenzenweger, & Kernberg, 2007; Giesen-Bloo et al., 2006). The same could be true for cluster A PD patients. An important question for further treatment research
for this patient population is: Are more intensive treatments such as day hospital and inpatient treatment indeed the treatment of choice for cluster A PD patients as our results seem to suggest and, if yes, how can they be tailored to this population?

Now that we presume that cluster A pathology does not necessarily interfere with psychotherapeutic treatment, cluster A PD patients should not a priori be excluded from psychotherapy. On the contrary, if our results can be replicated, it would be important to make psychotherapy more accessible for this patient group in order to reach health gains for this vulnerable group of psychiatric patients. Concluding, clinical practice may be more optimistic regarding psychotherapeutic treatment in cluster A PD patients.
Chapter 7

General discussion
This thesis aims to contribute to the understanding of psychotherapy for PDs and to refine the toolkit for psychotherapists working with this patient group. We show the value of more intensive forms of psychotherapy for different kinds of PD patients. The strongest evidence concerns the benefits of short-term inpatient psychotherapy for patients with cluster C PDs. Our results contribute to the growing evidence base of psychotherapy for PDs, extending it by cluster-specific treatment recommendations. Furthermore, we provide the psychotherapy researcher with a powerful tool to counter selection bias in non-randomised studies.

The main aims of this thesis are to

1. Explore what we know about psychotherapy for patients with PDs, and what we still need to know.
2. Investigate the effectiveness of different dosages of psychotherapy for different groups of PD patients.
3. Examine a method for comparing the effectiveness of widely differing treatments without randomising patients.

In this final chapter, I first provide a summary of the main findings. Then I discuss the strengths and limitations of the present studies as well as the implications of the findings for future research and practice.

Summary of main findings

Chapter 2 describes the existing research base of psychotherapy for PDs in terms of effectiveness and cost-effectiveness studies. The effectiveness of psychotherapy for patients with PD is well documented but does not yet fully live up to the modern standards of evidence-based medicine. One of our criticisms is that existing research does not cover all PDs and is often limited to experimental studies with low external validity. Furthermore, costs and effects are mostly studied apart, making integration of these key aspects difficult. We make a plea for more integrated research, as well as for high-quality effectiveness studies covering the broad range of PDs. We emphasise the need for studying dose-effect relations in psychotherapy for PDs.
Chapter 3 raises the matter of selection bias in psychotherapy research and offers a solution to this problem. When studying widely differing treatments, researchers face the ethical limits of randomisation. However, in non-randomised studies researchers are confronted with the problem of selection bias. We describe the propensity score method, a sophisticated statistical tool used to adjust for selection bias in quasi-experimental studies. This method enables researchers to compare the results of different treatments in non-randomised studies. We tested the usefulness and applicability of this method for psychotherapy studies and found that—when randomisation is impossible—quasi-experimental designs using the propensity score method are a feasible alternative. Keeping several precautions in mind, the propensity score method is a promising tool not only for our studies but also for future psychotherapy research in general. The chapter provides a step-by-step protocol for the application of the propensity score method for comparing two treatment arms. In an additional publication, we also extended this method to more than two groups (Spreeuwenberg et al., 2010).

The three result chapters (chapters 4 to 6) describe the results of the large multicentre study SCEPTRE conducted in regular clinical practice, according to the recommendations from chapter 2. SCEPTRE is an integrated (cost-)effectiveness study covering the whole range of PDs. Outcome was measured in terms of psychiatric symptoms, social and interpersonal functioning, and quality of life. All results were corrected for the influence of selection bias with the propensity score method described in chapter 3.

Chapter 4 presents the results on patients with cluster C PDs (N = 371), the biggest cluster of the SCEPTRE sample. We compared five different treatment modalities:

- long (more than six months) outpatient treatment
- short (up to six months) day hospital treatment
- long day hospital treatment
- short inpatient treatment
- long inpatient treatment.

Results were reported at 12 months after baseline. Patients in all treatment groups improved significantly on all outcome measures. Short inpatient treatment showed significantly more improvement than most other treatment modalities on all outcome measures. It is argued that short-term inpatient treatment seems to be specifically beneficial for this group of cluster C PD patients.
Chapter 5 compares three treatment settings—outpatient, day hospital, and inpatient treatment—for the group of cluster B PD patients (N = 207) at 18 months after baseline. Patients in all treatment groups improved significantly on all outcome measures. Furthermore, we found a marginally significant difference between inpatient and outpatient treatment. Patients in inpatient treatment improved more than patients in outpatient treatment regarding psychiatric symptoms. Our conclusion is that inpatient treatment is a valuable treatment for patients with cluster B PDs and should be taken seriously as a treatment option both in research and clinical practice.

Chapter 6 reports on the comparison of three treatment settings—outpatient, day hospital, and inpatient treatment—for the smallest patient group of our sample: cluster A PD patients (N = 57). In the comparison of the three treatment settings in terms of psychiatric symptoms, patients in day hospital and inpatient treatments showed the largest improvements. When applying the propensity score method to this study sample, it appeared that patients in the three treatment groups were not readily comparable. Therefore, the differences found in favour of inpatient and day hospital treatment have to be interpreted carefully, and replication of the results (in larger patient samples) is needed. Nevertheless, it is an important conclusion that positive change is possible in this highly impaired patient group. Cluster A pathology does not seem to impede profiting from psychotherapeutic treatment.

Strengths and limitations

The results of the present thesis have a strong external validity. SCEPTRÉ was conducted in a large sample of PD patients followed in regular clinical practice. Therapists were doing their regular work, without special training or supervision. Exclusion criteria were minimal, meaning that the present results are on real-world patients seen in real-world mental health care. Follow-up response was high, supporting our findings in two ways: First, a high follow-up response implies that findings are applicable to a large group of patients; second, the commitment shown by the participants reflects the value psychotherapy seems to have for them: Even though patients with PDs do not represent an “easy” patient population, they had the discipline to return questionnaires faithfully, even well after they had left treatment.

In the absence of a randomisation procedure, strong statistical control was provided by the propensity score method. Given the ethical constraints when comparing treatments differing widely in terms of dosage, using a quasi-experimental design
can be considered the best alternative. In that respect, I hope to have inspired the conductors of future psychotherapy research by offering a method to study treatment effects in a less invasive way, compared to the traditional randomisation design. For the future, I recommend conducting both randomised trials and quasi-experimental studies, each for their own purpose: one measuring efficacy and the other effectiveness in clinical practice.

The present thesis also has its limitations. First, treatment institutes differed in measurement schemes because of existing research projects. Differing assessment points hampered straightforward statistical comparisons. We tackled this problem by implementing multi-level statistical modelling in our calculations. This method can handle longitudinal data with observations unequally spaced in time.

Second, the effect of possible psychotropic medication was not reported. Our aim was to measure the effectiveness of psychotherapy, so we focused on the amount of therapy patients received. We did collect patient data on medication use before, during, and after psychotherapeutic treatment but, given the differing measurement schemes mentioned above, it was difficult to disentangle the effects of medication and psychotherapy. Similarly, scientific literature does emphasise that psychotherapy should always be the first treatment option to treat PD as a whole (e.g., National Institute for Health and Clinical Excellence, 2009) and recommends pharmacotherapy only for specific symptoms (Herpertz et al., 2007; Lieb, Vollm, Rucker, Timmer, & Stoffers, 2010) and/or crisis management (Livesley, 2005). Nevertheless, possible effects of the combination of pharmacotherapy and psychotherapy should gain more attention in future (randomised) research (Bond & Perry, 2006).

Third, there was no standardised measurement of Axis I disorders, even though it is well known that PDs are significantly comorbid with a wide range of Axis I disorders (Lenzenweger, Lane, Loranger, & Kessler, 2007). While we used self-report measures of psychiatric symptoms, such as the BSI (de Beurs & Zitman, 2006; Derogatis & Melisaratos, 1983), the use of standardised Axis I diagnoses would have helped to investigate the influence of Axis I disorders on the outcome of patients with PD. The reason for this omission was the purely practical one of not burdening patients with too many assessment instruments. The baseline assessment booklet was already large and patients found the thorough Axis II interview often tiresome. In retrospect, I would still choose to assess Axis I disorders with a standardised instrument, and strongly recommend this for future research on PDs.
Fourth, PDs were assessed only before the start of treatment. To determine the percentage of patients who fell under the threshold of their (former) PD(s) after treatment, it would be helpful to have a second semi-structured interview at a follow-up point. This is certainly a suggestion for longer-term follow-up of the present patient sample, perhaps five or ten years after baseline.

Fifth, the influence of theoretical school is not taken into account in our studies. In the participating institutions the two aspects of dosage and theoretical school were interrelated as on occasion a particular dosage was administered only within one theoretical frame, making it impossible to study these two variables apart. To disentangle the effects of dosage and theoretical orientation, a different research design is necessary. Previous research has shown that different schools hardly differ in effectiveness. However, a 2 x 2 research design varying both dosage and theoretical school would yield definitive answers to the question of differential effects of these two parameters.

Sixth, there was no control for the influence of possible care consumption after the intended treatment. As we did collect patient information on care consumption after leaving the index treatment, the effects of this additional care consumption can still be determined in future manuscripts. It is therefore something I would like to recommend for future research with SCEPTR data.

Change

Notwithstanding the limitations of our research, one of the main findings of this thesis is that during psychotherapy change takes place across all three clusters of PDs, especially in the outcome area of psychiatric symptoms. But when we talk about change and changeability of personality and PDs, what kind of change do we expect? This is not always clear. To further refine the toolkit for healthcare professionals, we still need to know (a) what kind of changes can be achieved in psychotherapy for PDs, (b) how long these changes will take, and (c) how these changes are achieved.

We know that during psychotherapy psychiatric symptoms improve, and with them often quality of life and psychosocial functioning. What we don’t know from our study is if this change is accompanied, preceded or followed by a deeper personality change. Within SCEPTR we measured personality functioning with a new instrument, the Severity Indices of Personality Problems (SIPP; Verheul et al.,
2008). So far, only preliminary results of this measure are available. Nevertheless, these first findings suggest that in a small subsample of the SCEPTRÉ patients, personality traits change from less adaptive to more adaptive during treatment (Verheul, et al., 2008). Further results are awaited. In their long-term follow-up of patients treated with day hospital based MBT, Bateman and Fonagy (2008) concluded that patients improved largely in the core borderline PD symptom of suicidality, as well as in diagnostic status. Other recent studies taking personality variables into account are the studies by Vinnars (2009) and Vermote (2009). All three studies found positive change in variables related to personality pathology, but patients still showed manifest impairment in certain areas of life. Taken together, we can conclude that during psychotherapy change takes place with a prominent position for symptomatic change. Less is known about the change of underlying personality traits and it is debatable if deeper personality change is a realistic goal in psychotherapy (Ferguson, 2010). Perhaps it is much more promising to strive for “dealing with life” (given certain personality characteristics) and to help the patient achieve greater adaptive skills and—essentially—higher satisfaction in life. Further research focusing on the interplay between symptomatic, functional, and deeper structural change is clearly needed to answer this question.

We know that during psychotherapy some form of positive change takes place. What we don’t know from our results is the sequence of changes that eventually lead to recovery. A pioneering naturalistic study on the sequence of recovery in an outpatient sample with depression, anxiety, and/or PDs was conducted by Perry and Bond (2009). The authors demonstrated that different outcome variables showed different patterns of change: They observed the earliest recovery in self-destructive symptoms (one of the core symptoms of borderline PD), the ability to work, self-reported symptoms of distress, and defensive functioning. Satisfaction and observer-rated measures of symptoms and functioning took longer to show recovery. Overall, improvement was common, while true recovery proved to be a much more difficult and slower-paced process—there is no such thing as a quick fix for personality pathology (see Figure 7.1). Complaints with high proportions of recovery took between 4.1 to 6.1 years to recover, while global measures of symptoms, (Axis I) disorders, and functioning were projected to take two to three times longer. These results are consistent with the view of Clark (2007, 2009), who concluded that treatment should first focus on more acute and changeable PD manifestations (e.g., self-harm and suicidality). Subsequently, it should address more stable long-standing maladaptive PD traits that cause problems in interpersonal functioning (e.g., the preference for solitary activities). Regarding the changeability of personality and
PDs, researchers have come to the conclusion that PDs can be regarded as consisting of two different aspects: more stable trait dimensions and less stable symptomatic behaviours (Clark, 2007). The revision of Axis II in DSM-5 should and most certainly will take these findings into account (Krueger, Skodol, Livesley, Shrout, & Huang, 2007; McGlashan et al., 2005). Overall, future research should concentrate on longer-term follow-ups and more in-depth process studies, especially regarding stages of change. Then we will hopefully be able to better determine which therapy dosages are needed to produce distinctive changes in personality functioning.

Figure 7.1. Source http://www.bpd.solved.com

Having said this, there remains the question how change can be achieved (Lundh, 2009). In order to shed light on this issue the black box of psychotherapy must be opened. Different forms of specialistic psychotherapy yield similar amounts of change, as previously mentioned. Therefore, it is assumed that their effect is rooted in certain common factors underlying the mechanisms of change (Ahn & Wampold, 2001; Lambert & Ogles, 2004). Still, there might be specific gains attributable to certain therapy techniques yet undiscovered. Recently, one well-designed dismantling study has investigated the merit of transference interpretations in psychotherapy (Høglend et al., 2008). More studies of this kind should be done to explore further possible effective ingredients of psychotherapy.
Costs of change

What we know more about, though, is what change costs. Present results show the effectiveness of time- and cost-intensive treatments such as inpatient and day hospital psychotherapy. The favourable results for the effectiveness of short-term inpatient treatment for cluster C PD patients are supported by cost-effectiveness data (Soeteman et al., in press) from the same patient sample. However, cost-effectiveness data in cluster B PD patients shows that the effective intensive treatments are less cost-effective (Soeteman et al., 2010). It is, however, possible that for certain subgroups of patients who do not profit enough from low-intensive treatments these treatments are necessary, effective, and – possibly – also cost-effective in the long run (due to less care consumption in the future). More long-term research is needed to tailor therapies more specifically to certain patient groups (i.e., matching) in terms of effectiveness and cost-effectiveness.

Implementation of research results in public health

One of the greatest challenges facing today’s mental healthcare is to establish a research culture within clinical practice, and similarly for contemporary research to serve and inspire clinical practice, thereby improving it in a meaningful way. Delivering evidence-based and high-quality healthcare is not only in the best interest of the patients. The pressure to present evidence on (cost-)effectiveness is getting higher (Bartak, Soeteman, Verheul, & Busschbach, 2007). In times of restricted health budgets, health insurances and other third-party providers force psychotherapists to know and to show how the money is spent. Instead of justifying the status quo, mental health practitioners should be open to new findings and should work together with researchers to implement (more) effective treatment programmes in order to let more patients benefit from evidence-based mental health care (Chiesa & Healy, 2009).

So how could the present research serve and inspire clinical practice?

Once scientifically established, effective treatments have to be implemented on a large scale in clinical practice. In our case, this would imply the broad implementation of structured short-term inpatient services for patients with cluster C PDs, as we have shown this form of treatment to be both highly effective and cost-effective. Furthermore, implementation of new treatment programmes should always be
accompanied by outcome-monitoring to make early evaluation possible and to adjust treatment in the case of threatening treatment failure. However, before we make far-reaching recommendations to clinical practice, we can begin much slower. I predict that the present study will contribute to clinical practice first and foremost in the longer term by introducing the important concept of dosage to make healthcare more tailored to (subgroups of) PD patients. If the concept of dosage becomes a natural part of psychotherapy research and is included in future studies, finally leading to refined and more specific practice guidelines, it will be—at least partly—a merit of this study. In short: dosage matters.
References


References


References


References


References


References


References


References


References


Pirckeymherus, B. (1527). *Theophrastou Charaktāres*. Nürnberg: [s.n.].


References


References


References


Summary
Summary

This thesis is on the effectiveness of different “dosages” of psychotherapy in patients with personality disorder (PD). In the general population about one out of ten people suffers from PD. Patients with PD and their environment experience a high level of distress in daily life. They are faced with problems in many areas of life—work, love, friendships—and are in need of help. Psychotherapy has proven to be a useful tool in helping these patients to improve. Existing psychotherapy research in PD so far has concentrated on comparing different therapeutic schools, while keeping therapy dosage (i.e., the amount of psychotherapy) constant. Mostly, little differences in effectiveness between different schools were found. The term dosage, rooted in medical research and here defined in terms of treatment setting (outpatient, day hospital, inpatient) and treatment duration, is a neglected yet important aspect in psychotherapy research. Therapy dosage has a strong impact on patients’ lives, as well as on treatment costs. In chapter 1, we suggest that different psychotherapy dosages may have different effects and that research into dose-effect relations is a fruitful study area within psychotherapy research.

We conducted three effectiveness studies, based on data from a large prospective multi-centre study in the Netherlands, project SCEPTRE (Study on Cost-Effectiveness of Personality Disorder TREatment). SCEPTRE followed more than 800 patients from six different treatment institutes for three years. In the psychiatric classification system of DSM-IV-TR PDs are categorised in three clusters (cluster A—the odd cluster, cluster B—the dramatic cluster, and cluster C—the anxious cluster). The present studies are categorised in the same way.

Comparing the effects of different treatments confronts us with the problem of selection bias: How do we know that the observed effect is caused by the treatment in question rather than by characteristics of the patients undergoing this particular treatment? To control for selection bias in this non-randomised study, we have to find an appropriate statistical tool.

Taken together, the aims of the present thesis are:

1. Explore what we know about psychotherapy for patients with PD, and what we still need to know.
2. Investigate the effectiveness of different dosages of psychotherapy for different groups of PD patients.
3. Examine a method for comparing the effectiveness of widely differing treatments without randomising patients.
Summary

Background—the evidence base of psychotherapy in PD

Chapter 2 describes the effectiveness of different forms of psychotherapy for patients with PD, as shown by past research. Preliminary studies also demonstrated its cost-effectiveness. However, psychotherapy for PD is not yet established in the minds of policy makers and in the world of insurance and reimbursement. For full acknowledgement of the value of psychotherapy for both PD patients and society, more integrated research is needed. We have to investigate the effects and costs of treatment in high-quality and real-world (cost-)effectiveness studies, covering the broad range of PD. Especially, researchers should study dose-effect relations in psychotherapy, as dosage has a high influence on both costs and effects.

Propensity score—the statistical tool for non-randomisation

In effectiveness research, randomisation is the gold standard to identify true cause-effect relations. All possible confounders have to be eliminated or controlled for. This is most elegantly achieved by randomisation of patients to different treatment options. However, there are situations, especially in (long-term) psychotherapy research, where randomisation is impractical, unethical and/or impossible. The second best option is then to control for possible confounders statistically. Chapter 3 describes the propensity score, a method already widely used in medical and economical research. It is a sophisticated statistical tool to reduce all possible confounders to one single variable. This variable then can be entered in the analysis in various ways. We show that the implementation of this tool in psychotherapy research is possible and useful, and we present a step-by-step protocol for the interested researcher. When applied carefully, the propensity score is a powerful tool to control for selection bias.

Effectiveness studies—the effect of different dosages of psychotherapy

We examine the effectiveness of different dosages of psychotherapeutic treatment for different patient groups in three separate studies: for cluster C, cluster B, and cluster A PD patients. Outcome areas are psychiatric symptoms, social and interpersonal functioning, and quality of life. Results are corrected for the influence of selection bias with the propensity score method.
Cluster C study
In chapter 4 we compare the effects of five different dosages of psychotherapy in 371 patients with one or more cluster C PDs. The following treatment groups are compared:

- long (more than six months) outpatient treatment
- short (up to six months) day hospital treatment
- long day hospital treatment
- short inpatient treatment
- long inpatient treatment.

One year after baseline, all patients improved significantly on all outcome measures. The highest gains are observed in the patients receiving short inpatient treatment. These patients improved significantly more than patients from other treatment groups. This result was most prominent with regard to psychiatric symptoms, but also applied to the outcome areas of psychosocial functioning and quality of life. We conclude that—even when corrected for the possible influence of patient characteristics—short-term inpatient psychotherapy is an interesting option for patients with cluster C PDs and should be investigated further in terms of long-term effectiveness and cost-effectiveness.

Cluster B study
In chapter 5 we compare the effects of three different dosages of psychotherapy in 207 patients with one or more cluster B PDs, mostly borderline PD. The following treatment groups are compared:

- outpatient treatment
- day hospital treatment
- inpatient treatment.

All patients improved significantly eighteen months after baseline, with most improvement observed in the inpatient group. The difference in improvement of psychiatric symptoms between outpatient and inpatient treatment was marginally significant in favour of inpatient treatment. For the remaining outcome areas, no significant differences between treatment groups were observed. We conclude that inpatient treatment should not a priori be excluded from the package of treatment options for cluster B PD patients but—in contrary—deserves more attention in both research and clinical practice. Important future areas of research in cluster B PD patients are the question of the “ideal dosage” (as we found no significant difference between day hospital and inpatient treatment) and differential effectiveness for different subgroups of cluster B PD patients.
Cluster A study
In chapter 6 we compare the effects of three different dosages of psychotherapy in 57 patients with one or more cluster A PDs, mostly paranoid PD. The following treatment groups are compared:

- outpatient treatment
- day hospital treatment
- inpatient treatment.

Patients in day hospital and inpatient treatment showed the highest improvement eighteen months after baseline. This has to be interpreted carefully, however, as it appeared from the analysis of the overlap of propensity scores, that the three treatment groups were not readily comparable. We conclude from this cluster A study that improvement is possible in this rather vulnerable patient population that has not gained much attention yet in psychotherapy and psychotherapy research.

Conclusion—the meaning of the present results

In chapter 7 it is concluded that the present results confirm the general effectiveness of psychotherapy for patients with PD. More importantly, our research opens new perspectives on psychotherapy and psychotherapy research by emphasising the meaning of dosage when judging the effectiveness of therapy. Dosage seems to matter. In general, we can conclude that the more intensive forms of treatment (day hospital and inpatient treatment) yield better results. However, the idea of “the more, the better” does not seem to apply to the cluster C PD population, as short-term inpatient treatment shows the best results in this patient group. This short-term and at the same time high pressure format of treatment is highly effective in cluster C PD patients, even when measured several months after treatment termination.

One major asset of this study is that it is conducted in regular clinical practice following a large sample of PD patients. External validity is further enhanced by minimal exclusion criteria and a high follow-up response. A second asset is the rigorous control of possible confounders in this non-randomised study by means of the propensity score. The present research has limitations. One important limitation is missing information on Axis I diagnoses, due to the practical reason of not overloading patients with assessment instruments. A second limitation is that we did not take into account the possible influence of theoretical school. We do know from previous research that theoretical school mostly is of little importance when comparing specialised treatments. Therefore we concentrated on the comparison
between dosages, leaving the puzzle of disentangling the effects of theoretical content and dosage to future researchers.

Our general conclusion is that psychotherapy brings about positive change for patients of all clusters of PDs, most prominently in the area of psychiatric symptoms, but also in psychosocial functioning and quality of life. A remarkable finding is the high effectiveness of short-term inpatient psychotherapy for patients with cluster C PDs. This information could be a good starting point for implementing the present results in clinical practice. The main mission of this thesis is to inspire researchers and clinicians to include dosage in mental health care considerations. For policy makers, insurance experts, scientists, therapists and patients alike, the message is: Dosage matters.
Samenvatting

(Summary in Dutch)
Samenvatting

Dit proefschrift onderzoekt de effectiviteit van verschillende “doseringen” van psychotherapie bij patiënten met een persoonlijkheidsstoornis (PS). Gemiddeld heeft één op de tien mensen in de algemene bevolking last van een PS. Het dagelijks leven is zwaar voor patiënten met PSen: zij worden geconfronteerd met problemen op vele gebieden van het leven—werk, liefde, vriendschap—en hebben hulp nodig. Ook hun omgeving heeft het dikkwils zwaar. Psychotherapie is een goed middel gebleken om het leven van deze patiënten te verbeteren1. Onderzoek over psychotherapie bij PS concentreerde zich tot dusver op het vergelijken van verschillende therapeutische scholen, waarbij de dosering van psychotherapie (dat wil zeggen: de hoeveelheid) vaak gelijk bleef. In het algemeen werden weinig verschillen in resultaat gevonden tussen de verschillende therapeutisch scholen. Het begrip dosering, dat zijn oorsprong heeft in medisch onderzoek, wordt hier gedefinieerd in termen van behandeling (ambulant, dagklinisch, klinisch) en behandelduur. Dit aspect is in psychotherapie onderzoek tot dusver onderbelicht gebleven. De dosering van therapie heeft grote invloed op het leven van de patiënt, en tevens op de behandelkosten. In hoofdstuk 1 stellen wij de vraag of verschillende doseringen van psychotherapie verschillende resultaten kunnen opleveren, en nemen wij aan dat het verband van dosering en effect een beloftevol onderwerp is voor psychotherapie onderzoek.

Wij deden drie effectiviteitsstudies, gebaseerd op gegevens van een grootschalig prospectief onderzoek in meerdere Nederlandse behandelcentra: project SCEPTRE (Study on Cost-Effectiveness of Personality Disorder TREatment). SCEPTRE volgde drie jaar lang meer dan 800 patiënten, bij zes verschillende instellingen. Het psychiatrisch classificatiesysteem DSM-IV-TR onderscheidt drie clusters van PSen (cluster A—het excentrieke cluster, cluster B—het dramatische cluster, en cluster C—het angstige cluster). Deze studies zijn gebaseerd op dezelfde indeling.

Een onvermijdelijk probleem bij het vergelijken van verschillende behandelingen is “selection bias”: wordt het waargenomen effect nu veroorzaakt door de betreffende behandeling, of eerder door de karakteristieken van de betreffende patiënten? Om in een niet-gerandomiseerde studie als de onze het probleem van selection bias te controleren, is een geavanceerd statistisch instrument nodig.

---

Samenvatting

Samengevat zijn de doelen van dit proefschrift:

1. Nagaan wat al bekend is over psychotherapie voor patiënten met PSen, en aan welke nieuwe kennis behoefte is.
2. Onderzoeken wat de effectiviteit is van verschillende doseringen van psychotherapie voor verschillende groepen PS patiënten.
3. Beproeven van een methode om—zonder randomisatie—de effectiviteit van verschillende behandelingen voor verschillende patiëntengroepen te vergelijken.

Achtergrond — wat bekend is over psychotherapie voor PSen

Hoofdstuk 2 beschrijft de effectiviteit van verschillende vormen van psychotherapie voor patiënten met PSen, als aangetoond door eerder onderzoek. Verkennende studies toonden ook al de kosteneffectiviteit aan. Toch krijgt psychotherapie voor PSen nog altijd weinig aandacht van beleidsmakers, ziektekostenverzekeraars en zorg-inkopers. Voor een volledige erkenning van de waarde van psychotherapie voor PSen voor zowel patiënten als de maatschappij, is meer gericht en geïntegreerd onderzoek nodig. We hebben hoogkwalitatieve en “real world” effectiviteitsstudies nodig, die het effect en de kosten van behandeling onderzoeken voor patiënten in alle PS-clusters. Onderzoekers zouden daarbij speciale aandacht moeten geven aan de relatie van dosering en effect, aangezien dosering zowel op het effect als de kosten grote invloed heeft.

Propensity score — statistische hulp voor niet-gerandomiseerde studies

In effectiviteitsstudies is toevallige toewijzing van patiënten aan de te onderzoeken behandelingen (randomisatie) de gouden standaard om een betrouwbare relatie tussen oorzaak en gevolg te kunnen leggen. Alle verstorende invloeden (confounders) moeten worden geëlimineerd of meegewogen — en randomisatie is de elegantste manier om dat te doen. Er zijn echter situaties, en zeker in (langertermijn) psychotherapie onderzoek, waar randomisatie onpraktisch, onethisch en/of onmogelijk is. Het beste alternatief is dan om confounders onder controle te brengen met een statistisch instrument. Hoofdstuk 3 beschrijft de “propensity score”, een methode die in medisch en economisch onderzoek vaak gebruikt wordt. Het is een verfijnde methode, die alle mogelijke verstorende invloeden reduceert tot een enkele
variabele. Die variabele kan in de analyse vervolgens op allerlei manieren worden gebruikt. Wij tonen aan dat deze methode bruikbaar en nuttig is voor psychotherapie onderzoek, en wij presenteren de geïnteresseerde onderzoeker een stap-voor-stap protocol voor het gebruik ervan. Mits deskundig gebruikt, controleert de propensity score doeltreffend het verstorende effect van “selection bias”.

**Effectiviteitsstudies—effect van verschillende doseringen psychotherapie**

Wij onderzoeken de effectiviteit van verschillende doseringen van psychotherapeutische behandeling voor patiënten uit drie verschillende groepen: cluster C, cluster B en Cluster A PS patiënten. Uitkomsten worden gemeten op het gebied van psychiatrische symptomen, sociaal en interpersoonlijk functioneren en kwaliteit van leven. De propensity score corrigeert de resultaten voor de invloed van selection bias.

**Cluster C studie**

In *hoofdstuk 4* vergelijken wij de effecten van vijf verschillende doseringen van psychotherapie bij 371 patiënten met een of meer cluster C PSen. De volgende behandelgroepen worden vergeleken:

- langdurig (meer dan zes maanden) ambulant
- kort (tot zes maanden) dagklinisch
- langdurig dagklinisch
- kort klinisch
- langdurig klinisch.

Een jaar na de beginmeting (baseline) verbeterden alle patiënten significant op alle uitkomstgebieden. De grootste stap vooruit werd waargenomen in de korte klinische behandelgroep. Deze patiënten verbeterden significant meer dan patiënten in andere behandelgroepen. De grootste verbetering betrof psychiatrische symptomen, maar ook psychosociaal functioneren en kwaliteit van leven verbeterden. Wij concluderen dat—zelfs na correctie voor de mogelijke invloed van patiëntkarakteristieken—korte klinische behandeling voor cluster C PS patiënten een interessante optie is. Nader onderzoek over het lange-termijn resultaat en de kosteneffectiviteit ervan is daarom gewenst.
**Samenvatting**

**Cluster B studie**
In *hoofdstuk 5* vergelijken wij de effecten van drie verschillende doseringen van psychotherapie bij 207 patiënten met een of meer cluster B PSen, meestal borderline PS. De volgende behandelgroepen worden vergeleken:

- ambulant
- dagklinisch
- klinisch.

Achtte maanden na baseline verbeterden alle patiënten significant, met het beste resultaat in de klinische groep. Op het gebied van psychiatrische symptomen was er een (marginaal) significant verschil tussen patiënten in klinische en ambulante behandeling. Klinische behandeling laat daar iets meer verbetering zien. Op de andere uitkomstgebieden werden geen significante verschillen tussen behandelgroepen gevonden. Wij concluderen dat klinische behandeling van cluster B PS patiënten niet a priori als behandeloptie mag worden uitgesloten, maar dat deze behandelmogelijkheid even aandacht verdient binnen onderzoek en in de klinische praktijk. Belangrijke toekomstige onderzoeksgebieden zijn de ideale dosering voor cluster B PS patiënten (wij konden geen verschil vinden tussen ambulante en klinisch behandeling), en verschillen in effectiviteit voor verschillende subgroepen van cluster B PS patiënten.

**Cluster A studie**
In *hoofdstuk 6* vergelijken wij de effecten van drie verschillende doseringen van psychotherapie bij 57 patiënten met een of meer cluster A PSen, meestal paranoïde PS. De volgende behandelgroepen worden vergeleken:

- ambulant
- dagklinisch
- klinisch.

Patiënten in de dagklinische en klinische groep lieten de grootste verbetering zien, achtte maanden na baseline. Stellige conclusies over de superioriteit van een bepaalde behandeling kunnen we echter niet trekken. De drie behandelgroepen bleken, na analyse van de overlap in propensity scores, niet goed vergelijkbaar. Onze conclusie is wel dat voor deze kwetsbare groep patiënten verbetering mogelijk is, en dat de mogelijkheid van psychotherapeutische behandeling meer aandacht verdient in de behandeling van en het onderzoek naar cluster A PS patiënten.
Conclusie – betekenis van de resultaten

Onze conclusie in hoofdstuk 7 is dat de resultaten van deze studie de algemene effectiviteit van psychotherapie voor patiënten met PSen bevestigen. Bovendien biedt deze studie nieuwe uitgangspunten voor psychotherapie praktijk en onderzoek, door de aandacht te vestigen op het belang van dosering bij het beoordelen van effectiviteit.

Onze conclusie is dat, in het algemeen, de meer intensieve behandelingen (dagklinisch en klinisch) de beste resultaten opleveren. Echter, het uitgangspunt “meer is beter” schijnt niet van toepassing te zijn op cluster C PS patiënten. In die groep laat de behandelvorm kort klinisch de beste resultaten zien. Deze in tijd beperkte en tegelijkertijd intensieve behandelvorm blijkt opvallend effectief voor cluster C PS patiënten, ook enige maanden na afsluiting van de behandeling.

Een belangrijk sterk punt van deze studie is dat die werd uitgevoerd in de dagelijkse klinische praktijk, onder een groot aantal PS patiënten. De validiteit van de resultaten wordt verder verhoogd door de minimale uitsluitingscriteria, en een hoge response van patiënten bij de follow-up. Een tweede sterk punt is de rigoureuze controle op verstорrende invloeden (confounders) in deze niet-gerandomiseerde studie.

Dit onderzoek heeft enige beperkingen. Een daarvan is het ontbreken van gegevens over As I diagnoses. Het was praktisch niet haalbaar de patiënten te belasten met metingen ook op dit laatste gebied. Een tweede beperking is dat deze studie geen rekening houdt met de mogelijke invloed van de theoretische school. We weten echter uit eerder onderzoek dat bij de vergelijking van behandelresultaten de theoretische school meestal weinig invloed heeft. Daarom concentreren wij ons op de vergelijking van doseringen. Het ontwarren van de wisselwerking tussen theoretische inhoud en dosering laten wij over aan andere onderzoekers.

Onze algemene conclusie is dat psychotherapie bij PS patiënten uit alle drie clusters verbeteringen teveel brengt, vooral op het gebied van psychiatrische symptomen, maar ook wat betreft psychosociaal functioneren en kwaliteit van leven. Een opvallende ontdekking is de hoge effectiviteit van kortdurende klinische behandeling bij cluster C PS patiënten. Dit laatste gegeven is een solide vertrekpunt voor klinische implementatie en verfijning van de klinische praktijk. De opdracht die wij ons met deze studie stelden is onderzoekers en cliënten het belang van dosering te laten zien. Ook voor beleidsmakers, verzekeraars, en voor patiënten is dat de boodschap: Dosering doet er toe.
Acknowledgements
Acknowledgements

Geht nicht gibt’s nicht.¹ That is what my grandmother has taught me with her infinite energy and her endless love for me. I often had to think of that sentence during the ups and downs of writing this dissertation. Therefore I dedicate this book to her. Without her, I would not be who I am today. Oma, danke für alles.

Back to work. My daily supervisor Jan van Busschbach was one of the most important people in helping me to finish this dissertation. One of the most valuable things he taught me was not to do things. Then you can concentrate more on doing the things that really matter. Jan, het is gelukt, dank je wel.

There is somebody who shared his ambitions and ideas with me and without him this whole project wouldn’t exist: Roel Verheul. He is brilliant in starting up things and in believing that they will succeed. It is thanks to his drive and determination that project SCEPTRE turned out to be so successful.

Special thanks go out to my two PhD supervisors, Paul Emmelkamp and Theo Stijnen. You accompanied me through the whole process of setting up the project, analysing the data, and writing this dissertation. You were always there for me when I needed your advice.

All members of my PhD committee: I want to thank you for carefully reading my dissertation and for sharing your wisdom. Very special thanks go to John Livesley. John, you inspired me every time we met. I am grateful and honoured that you will be with me at my defence and I hope we can keep sharing our thoughts.

My colleagues from “De Viersprong” were often a source of inspiration and support. I learned a lot in my “Viersprong-years” and I am thankful for the opportunities offered to me there. Many people were supportive, too many to name them all. But some were special: Helene Andrea, Els Havermans, Janine van Manen, Dineke Feenstra, Eva Horn, Daniëlle Smeets, Djøra Soeteman, Hilde de Saeger, Nicole Op ‘t Veld, Johan Gudde, Dien Elshof, Joost Hutsebaut, Dawn Bales, Ab Hesselink, Mia Famaey, Stef Bouwman, Sissy Hamers, Anne-Marie Claassen, Marcia de Nijs, Greta Günther.

Sjouk Hartman, former director of “De Viersprong”: without your vision and strong will our whole research department would never have existed. I thank you for your trust in us and for your innovative thinking more than ten years ago.

¹ There is no such thing as “no can do”. 
Acknowledgements

SCEPTRE was a huge project, with six participating mental health centres. Lots of people contributed to it and I want to thank all of them. My special thanks go to: all participating patients, Uli Ziegler, Bert van Rossum, Anke Meerman, Moniek Thunnissen, Jos Delimon, Piet Rijnierse, Lot Holleman, Ellen van den Eijnden, Fleur Bouvy, Janneke Aerts, Alice Punt, and all the others who carefully took care of the growth of SCEPTRE.

My old and new colleagues at the department of Clinical Psychology of the University of Amsterdam, thank you for the support and fun in the last (and next) years. My special thanks go to: Emily Brugman, Riëtta Oberink, Jan Henk Kamphuis, Herman Vinckers, Sandra Diets, Kitty Rolf and Merel Kindt.

Marieke Spreeuwenberg, my faithful friend and intelligent co-author of most of my papers: thank you for all the hours we spent together in front of the screen, for all the work we accomplished together, for the encouraging words I needed from time to time, and most of all for your friendship. I will soon come to see your new house and your own hill.

It was not always easy to live commuting between Amsterdam and Brabant. But there were some people in Brabant who made life there warmer. First of all Denise de Weerd, my friend and dear house mate: thank you for sharing a “station” of our lives and for all our evenings. Fons Groffen and Jeannette op de Beke (who is not with us any more): thank you both for being caring neighbours and for good words and a sweet white port in difficult times. Everybody from Canoe club Zeewitoe, especially the whole family Buwalda: thank you for teaching me how to keep my head above the water and for all the fun we had together.

One of the reasons why I could concentrate on and enjoy my work as much as I did in the last few months was the fact that I knew our son was in safe hands while I was working. Bernice van Staaldjuine, Kitty van Muiswinkel, and Fatima Kraimi, thank you for being a caring friend of Kai.

Saskia Müller, of all of my friends you were the one who most strongly encouraged me to write this dissertation. Dank je wel voor alle levenswijsheid, niet alleen omtrent werk en wetenschap.
My strong and loving family - near or far - is always there for me, each of them in their own way: Mama-Kristina, Papa, and Moritz, Wolfgang, Helga and Rolf, Opa and Oma Trutzhan, Opa and Oma Bieber (who are both not with us any more), Lute and Wimie, Carmen, Heinz, Johanna, Fabian and Leo. Fijn dat jullie bestaan. Ich bin so froh, daß es Euch gibt.

When you choose the two people who will stand next to you during your defence, one of the criteria is that you feel safe with them and comforted by them. Ben, thank you for the good spirit you bring into my life and thank you for being my paranymph. Elisabeth, a mother-in-law once told me a story about what a friend is: the precious stone you find by coincidence between all these other pebbles. That is what you have been for me for a big part of my life now and I always carry you with me. Thank you for being you and for being my paranymph.

And now: Justus van Oel, my husband and soul sister. Your mental, practical, and emotional contributions to this dissertation are innumerable. Dank je dat je als cadeau in mijn leven bent gekomen. En dank je voor al het lekkere eten waarin ik elke dag jouw liefde proef.
Publications


