THERAPY AMOUNT AND OUTCOME OF INPATIENT PSYCHODYNAMIC TREATMENT OF EATING DISORDERS IN GERMANY: DATA FROM A MULTICENTER STUDY

Horst Kächele
Hans Kordy
Matthias Richard
Research Group TR-EAT
Center for Psychotherapy Research, Stuttgart, Germany

The present study investigates (a) factors that determine length of treatment and (b) the effect of treatment duration and other factors on outcome for patients with eating disorders. In this observational study, the symptomatic status of 1,171 patients was observed for 2.5 years after admission to 1 of 43 participating hospitals. Treatment and outcome were modeled using hierarchical linear models and logistic regression. To control for possible confounding factors, propensity score adjustment was used. Treatment modalities, especially length and intensity, varied considerably between and within hospitals and were related to patient characteristics to a very small degree. At 2.5-year follow-up, 33% of anorexia patients and 25% of bulimia patients were symptom free. Length of treatment showed a weak effect on outcome and only in interaction with other patient characteristics of relevance, whereas treatment intensity was not related to outcome. Implications for treatment planning are discussed.

Anorexia nervosa (AN) and bulimia nervosa (BN) are predominantly female mental disorders arising in adolescence and early adulthood. The frequency of these illnesses, severity of their symptoms, risk of chronicity and seriousness of physical complications (Treasure & Smukler, 1995), and related social costs emphasize the clinical interest and political importance of these disorders.

Eating disorders—AN or BN—have been estimated to have a relatively high prevalence (1–4%). AN has the highest mortality risk of all mental disorders; 20% of patients die within 20 years of onset (Crisp, Callendar, Halek, & Hsu, 1992). When patients who have eating disorder symptoms not otherwise specified are considered, although they do not fulfill diagnostic criteria for AN or BN, the prevalence increases to 5 to 15% (Herzog, Keller, Lavori, & Sacks, 1991).

Since the mid-1970s, the amount of research conducted on eating disorders has increased. Standardization has been achieved in defining both symptoms of illness and diagnostic categories. Several treatment programs have been developed, and their efficacy has been investigated (e.g., Roth & Fonagy, 1996). Although the efficacy of some treatment programs has been supported empirically, especially treatment programs for BN (e.g., Fairburn et al., 1995), these studies also show that much research remains to be done. Empirically validated treatments for BN result only in 50% of cases in long-term recovery (Mitchell, Hoberman, Peterson, Mussell, & Pyle, 1996). The chances for full recovery in AN patients is much lower (Crisp, Joughlin, Halek, & Bowyer, 1996).

Only a few predictors of outcome have been consistently identified (Bulik, Sullivan, Joyce, Carter, & McIntosh, 1998; Herzog, Deter, Fiehn, & Petzold, 1997; Keel, Mitchell, Miller, Davis, & Crow, 1999; Wilson et al., 1999). Empirical findings have not been significant enough to justify the use of one treatment over another (Herzog & Hartmann, 1997). Most studies were carried out under unique research conditions in specialty treatment centers, often with selected samples (Mitchell, Maki, Adson, Ruskin, & Crow, 1997). Studies concerned with the effectiveness of treatment in a naturalistic setting are sparse, as are studies that investigate the relations between amount and outcome of treatment (Herzog, Hartmann, & Sandholz, 1996; Treasure & Schmidt, 1999). Mitchell, Pyle, Pomeroy, et al. (1993) reported a higher rate of success at the end of treatment when treatment intensity was increased (two sessions a week instead of one), holding constant the treatment duration. We are not aware of comparative studies of alternative inpatient treatment programs such as those that AN patients often receive (McKenzie & Joyce, 1992; Treasure, Todd, & Smukler, 1995). Although treatment programs are usually very intensive and involve comprehensive therapeutic resources, data concerning the effect of the composition of the inpatient treatment program, its duration, and intensity on outcome are practically nonexistent.

We found only one study in the German language literature that investigated the effect of treatment intensity on success of outcome (Herzog et al., 1996). Treatment duration (with the same number of therapy sessions) was not a significant variable. However, once again, treatment intensity was an important factor. An increase in treatment intensity allowed for a comparable outcome to be reached within 3 months as opposed to 12 months.

Most studies concerned with the effectiveness of psychotherapy on eating disorders have investigated standardized treatment programs with predefined treatment duration and intensity. However, this does not reflect the naturalistic process of treatment, in which the treatment is not defined at admission and it is possible for treatment to be expanded or shortened. Conclusive findings that identify variables predictive of the length or intensity of treatment are largely not available. Thus, the questions of whether patients with severe symptoms receive the longer or more intensive treatments and whether other factors, and which ones, are involved cannot be answered yet.

The Center for Psychotherapy Research in Stuttgart initiated a multicenter study on the effectiveness of inpatient psychodynamic treatment of eating disorders in the early 1990s. Besides this overall objective, the study aimed to estimate the amount of therapeutic resources that were applied within the various treatment programs in everyday clinical practice and the impact on the short- and medium-term course of eating disorders. After the pilot phase, the main study began in 1993. It was completed at the end of 1998 with a 2.5-year follow-up assessment. Forty-three specialty

hospitals and departments for psychosomatic medicine and psychotherapy in Germany participated in the data collection.

We address the main questions investigated by the study (see Kächele for the research group TR-EAT, 1999): What factors determine the length of inpatient treatment? What is the effectiveness of inpatient psychodynamic therapy (i.e., what rates of recovery or improvement, measured from therapist's as well as patient's perspective, were reached and for which patients)? How do treatment duration and intensity contribute to the effectiveness? Can such effects be attributed to specific patient characteristics?

Method

Study Design

Project TR-EAT was a naturalistic, longitudinal, and observational study. Treatment duration and intensity were not standardized to observe the naturalistic clinical course of treatment. Data were collected from 1993 to 1998. Patients included in this study were at least 18 years old and fulfilled diagnostic criteria for AN or BN, or both, at screening.

All patients were assessed over a 2.5-year period (Figure 1). At therapy admission and discharge, as well as 1 year and 2.5 years after index admission, patients were questioned as to physical condition, mental state, and level of psychosocial functioning using a comprehensive battery of inventories. At the same time, the condition of each patient was clinically evaluated at admission and discharge by their primary therapist. The 2.5-year assessment was conducted by clinical experts. The 1-year assessment was conducted by mail and thus was limited to self-evaluation. The questionnaires covered sociodemographic and historical variables; the battery of psychological inventories included the Symptom Checklist-90-R (SCL-90-R), the Eating Disorder Inventory (EDI), the Freiburg Personality Inventory (Freiburger Persönlichkeitsinventar [FPI-R]), the Narcissistic Personality Inventory (Narzissmus
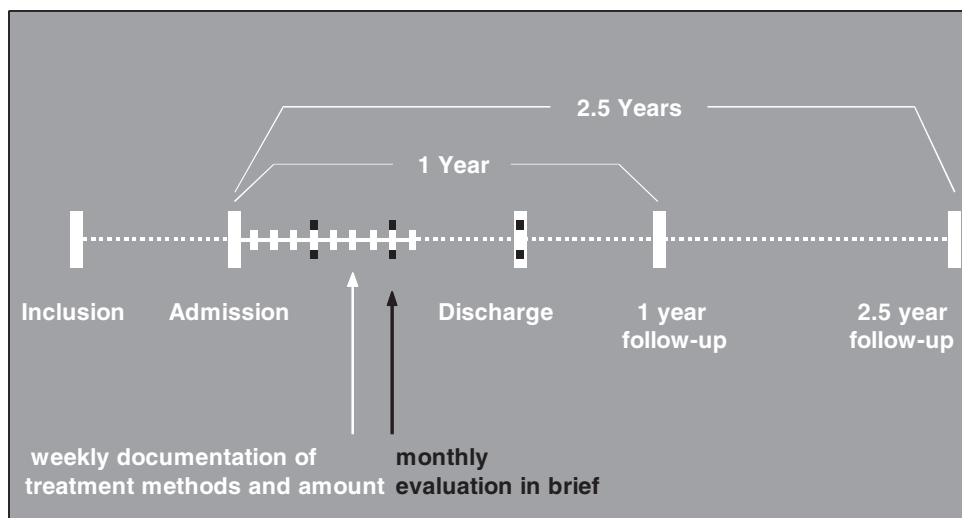

FIGURE 1. Study design.

Fragebogen), and the Parental Care Index (Familien-Klima-Skalen). Treatment exposure was measured using weekly documentation of the frequency and duration of all psychotherapeutic contact. In addition, the occupation of the participant was considered (to estimate therapy cost), and the number of participants in group or family therapy was monitored. The 2.5-year follow-up assessment was completed using a semistructured interview, the Longitudinal Interval Follow-up Evaluation (LIFE; Keller, Lavori, Friedman, et al., 1987), adapted for use in the present study. Retrospective, longitudinal information on symptomatic disturbance (the "change points") during the postdischarge course of the illness was obtained using LIFE. This information was used to track the course of recovery and relapse between the point of discharge and the 2.5-year follow-up interview (Kordy et al., in press).

Participants

Treatment of 1,171 patients from 43 sites in Germany was examined; 355 patients fulfilled diagnostic criteria for AN, 647 patients fulfilled criteria for BN, and 169 patients fulfilled criteria for both disorders. The majority were female. Less than 4% of the sample were male (AN = 3.3%, BN = 2.3%). The mean age of the participants was 24.8 years for AN (SD = 5.6) and 25.9 years for BN (SD = 6.3). The mean duration of illness before admission was 8.2 years (SD = 6.1) for BN and 5.7 years (SD = 5.3) for AN. Anorexic patients had a body mass index (BMI) 72.1% of expectation (SD = 8.4).[1]

At the 2.5-year follow-up, 879 patients (75.1%) could be contacted. A comprehensive interview was conducted with 781 patients (66.7%). Limited information given by family or doctors was available for 98 patients (8.4%). Of the 292 patients who could not be reached, only 64 declined participation. It was not possible to establish contact with the other 207 patients despite repeated attempts to contact them via mail and telephone. For various other reasons, no data were available for 11 patients. In the course of our study, 10 patients died: 6 through suicide, 1 as a result of medical complications related to the illness, and 3 of unknown causes.

Few differences were found between patients who participated in this 2.5-year study and those who did not. Anorexic patients who participated received 2 more weeks of therapy on average, had a slightly higher BMI at discharge, and were not as motivated to partake in treatment at admission. Patients diagnosed with BN who participated in this study had a slightly higher BMI at index admission and had a longer history of illness on average. They also scored higher on the SCL-90-R Phobic Anxiety subscale and slightly lower on the Morgan-Russell Socioeconomic State (E) scale. Overall, there were few indications of possible bias resulting from nonrandomization; one eventual consideration is that the BN sample at the 2.5-year follow-up assessment tended toward chronicity.

Definitions of Recovery

The diagnostic criteria used to measure outcome were largely based on data concerning symptoms specifically related to eating disorders. These data were obtained through self-report questionnaires and therapist assessments. The three required characteristics for AN (underweight, fear of gaining weight, and body perception distortion) and BN (recurrent episodes of binge eating, recurrent compensatory

---

[1]Detailed tables are available from the authors.

behavior to prevent weight gain, and overconcerns with body and shape) were operationalized according to specific criteria. Outcome was evaluated as successful when at least two of these diagnostic criteria (within each disorder), including underweight for AN and binge episodes for BN, were no longer present at the end of treatment or at the follow-up assessment (Table 1). This operationalization relies on diagnostic criteria used in other international studies (Fichter, Quadflieg, & Reif, 1994; Mitchell et al., 1996; Strober, Freeman, & Morrell, 1997).

Whereas in most studies underweight is defined by a fixed threshold, usually 17.5 BMI, we used German norms, which take age into account (Hebebrand, Heseker, Himmelmann, Schäfer, & Remschmidt, 1994; Oehlschlägel, Malewski, & Mahon, 1999). Thus, the age-specific expected weight is set to 100%. A 25-year-old woman with a BMI of 17.5 would be 84% of expected weight for her age group.

Predictor Variables

The variables included in our study were chosen to cover several central domains and tended to correlate only moderately (Table 2). These variables were investigated in association with treatment duration and outcome. Covered domains included the history of the illness, pretreatment syndrome status, and treatment parameters such as duration and intensity. Hospitals were classified according to the type: specialty eating disorder hospitals and hospitals with a separate eating disorders ward, those with a specialized eating disorders treatment program, and those without treatment settings specifically for patients with eating disorders. Treatment intensity was measured using the number of hours of therapy from Weeks 3 and 4. Hours of therapy were separately measured within two subsets: standard therapy hours (individual or group sessions) and extratherapeutic activity (e.g., ergotherapy, sports therapy, dance therapy, art therapy). Weeks 3 and 4 were chosen because treatment intensity at this stage was expected to be most representative of the treatment intensity for the entire program.

Analyses

The analysis was divided into two main parts. The first was concerned with variables that predict treatment duration, and the second with variables that predict

TABLE 1. Criteria for Definition of Therapy Success

Symptom domain	Symptom no longer significant at assessment points
Anorexia nervosa	
Low body weight	Weight ≥ 85% of expected BMI
Fear of weight gain	P: EDI "drive for thinness" ≤ 4.5
	T: ≤ "light" ("not at all" to "very strong" in 5 steps)
Body image distortion	P: estimated vs. real %BMI ≤ 10%;
	T: ≤ "light" ("not at all" to "very strong" in 5 steps)
Bulimia nervosa	
Binge eating[a]	Fewer than 2 episodes/week
Weight-reducing methods	No vomiting, no use of laxatives, no strict dieting
Weight and shape concern	≤ "Sometimes" ("never" to "always" in 5 steps)[b]

Note. P = patient evaluation; T = therapist evaluation.
[a]Main symptom.
[b]The question was "Do you pay a lot of attention to your figure?"

TABLE 2. Overview of Investigated Predictors

Predictor	Operationalization
Patient data at index admission	
Motivation	Not motivated–motivated
Age	In years
Impairment caused by eating disorders	Patient rating: 1 (low) to 5 (high)
Low body weight	Weight (% mean expected BMI)
Diagnosis	AN, BN
EDI	Sum of subscales 1, 2, and 3
FPI	QS[a] deviations from norm (all scales)
AN and/or BN symptoms	Criteria fulfilled by none, one, both, all
Sick leave	No. days in past year
Multi-impulsivity	Substance abuse, aggression, impulsivity (ratings)
Narcissism Inventory	Total score
Duration of illness	Years (logarithmic)
Sexual abuse	None known, suspected, present
Psychological distress	GSI score
Social functioning	Morgan-Russell E scale
Number of previous treatments	None, one, two, three, and more
Health care insurer	Private–public: BfA, DAK, AOK, others
Treatment parameters	
Duration	Short (≤ 11.5 weeks), long (> 11.5 weeks)
No. psychotherapy sessions (individual or group)	As documented from Weeks 3 and 4
No. sessions of adjunctive therapy (e.g., ergo- or sports therapy)	As documented from Weeks 3 and 4
Type of hospital	
Classification of hospitals	Degree of specialization (see text)

Note. AN = anorexia nervosa; BN = bulimia nervosa; BMI = body mass index; EDI = Eating Disorder Inventory; FPI = Freiburg Personality Inventory; GSI = Global Severity Index of the Symptoms Checklist-90-R; E = Morgan-Russell Socioeconomic State scale; BfA = Bundesversicherungsanstalt fuer Angestellte; DAK = Deutsche Angestellten Krankenkasse; AOK = Allgemeine Ortskrankenkasse.
[a]Square sum.

outcome. Associations between the selected predictor variables and inpatient treatment duration (see Table 2) and the effect of the main treatment parameters duration and intensity on outcome rates were investigated. For analyses of outcome, the rate of positive outcome was the dependent variable. To control for possible bias as a result of confounding variables, estimates were adjusted using the propensity score (Rosenbaum & Rubin, 1983, 1984).

Analysis of treatment duration. Treatment duration varies within and between clinics. To test for possible predictive variables, a hierarchical linear modeling was used (Bryk & Raudenbush, 1992; Goldstein, 1995). This method uses multilevel regression analysis to investigate treatment assignment and treatment outcome A hierarchy exists between patient and clinic with patients nested within one and only one clinic. On the first level, differences within hospitals were modeled; on the second level, differences between hospitals were investigated. Variance in treatment

duration within and between hospitals was estimated in the first model, which did not include any additional predictor. A sequence of models was fitted by the gradual refinement of this first model. At a second step, patient characteristics were added as covariates. Finally, at the third step, characteristics of the treatment centers were added.

Analysis of rate of recovery and relapse. The main purpose of this analysis was to test for the effect of treatment duration and intensity on outcome success rates after 2.5 years. Because of the naturalistic design of the study, confounding variables could bias the estimated effect of these predictors on outcome.[2] To control for confounding variables, a method developed by Rosenbaum and Rubin (1983) was used. The propensity score was developed in the 1980s to remove possible bias in studies conducted without the benefit of randomization and to compare two treatment alternatives. The procedure can be divided into two steps. In the first step, each patient is given a propensity score using logistic regression or a discriminant analysis. The propensity score is the conditional probability of assignment to treatment A or B (here shorter or longer treatment) given the observed covariate information. In the second step, the sample is subclassified into, for example, quintiles of the distribution of the propensity score to form subgroups with rather similar propensity scores. As Rosenbaum and Rubin (1983, 1984) showed, the two treatment alternatives within the quintiles balanced with regard to the confounding variables (given that there are no predictor variables that clearly distinguish between the two treatments). Thus, an unbiased estimate (or with relatively small bias) of the effect of both treatment alternatives is possible.

Logistic regression was used for the statistical analysis of predictors of outcome (Hosmer & Lemeshow, 1989), in which the propensity score method was again applied for removing bias (D'Agostino, 1998). Because of the large number of possible predictor variables, there is a risk of overspecification. Therefore, the proportion of variance explained was corrected by using a bootstrap method. The same regression model was constructed with a random sample of two thirds of the initial sample, and differences between the sampled R^2 and the R^2 of the entire model were estimated. This procedure was repeated 200 times. The mean of these differences yields an adjustment for the full-model R^2 that corrects for bias (Efron & Tibshirani, 1993).

Results

The following results are based on a sample of 1,112 patients (AN: n = 338, BN: n = 605, AN and BN: n = 169) from 43 hospitals, of which 733 (AN: n = 225, BN: n = 399, AN and BN: n = 109) participated in the 2.5-year follow-up assessment. To allow for a consistent interpretation, data from 59 patients were excluded: 20 of the 59 patients received primarily outpatient treatment, and the other 39 received fewer than 4 weeks of inpatient treatment. A treatment program of fewer than 4 weeks was not offered by any of the clinics; therefore, the treatment of these 59 patients cannot be considered representative for any of the programs and these patients are dropouts.

---

[2]Chronically ill patients could have received longer rather than shorter treatments. On the assumption that these patients have a poorer prognosis, bias in the estimates of success of shorter and longer treatment programs would have resulted.

Which Patients Received Which Length of Treatment?

The mean treatment duration for all three subclasses of eating disorders was roughly 11 weeks (median weeks: AN = 11.1; BN = 11.4; AN and BN = 10.6). Twenty-five percent of patients were treated for 4 to 8 weeks; a further 50% were treated for 9 to 13 weeks. Fifteen-week treatments were rare, and those that continue for more than 6 months were exceptionally rare.

Figure 2 depicts both sources of variance: The length of each individual box plot indicates the variance within each clinic, and the differences in length between box plots illustrate variance between the various hospitals. Only three hospitals had a median treatment duration of 20 weeks or longer (Hospitals 53, 23, and 19). A large majority of the hospitals had a median treatment duration of 8 to 12 weeks. A few hospitals showed high variability in treatment duration (Hospitals 53, 19, and 5), whereas in others the length of treatment varied only slightly (Hospitals 6, 38, 11, and 26).

To investigate which variables defined treatment duration, a stepwise linear model was used. The construction of the model required three steps. The results are presented in Tables 3 and 4. Variances within and between hospitals were estimated using a first model without predictors[3] (Model 1 in Table 3), also known as an unconditional model (Bryk & Raudenbush, 1992). The effect of the patient-related variables on treatment duration is presented in the second column (Model 2). Finally, in
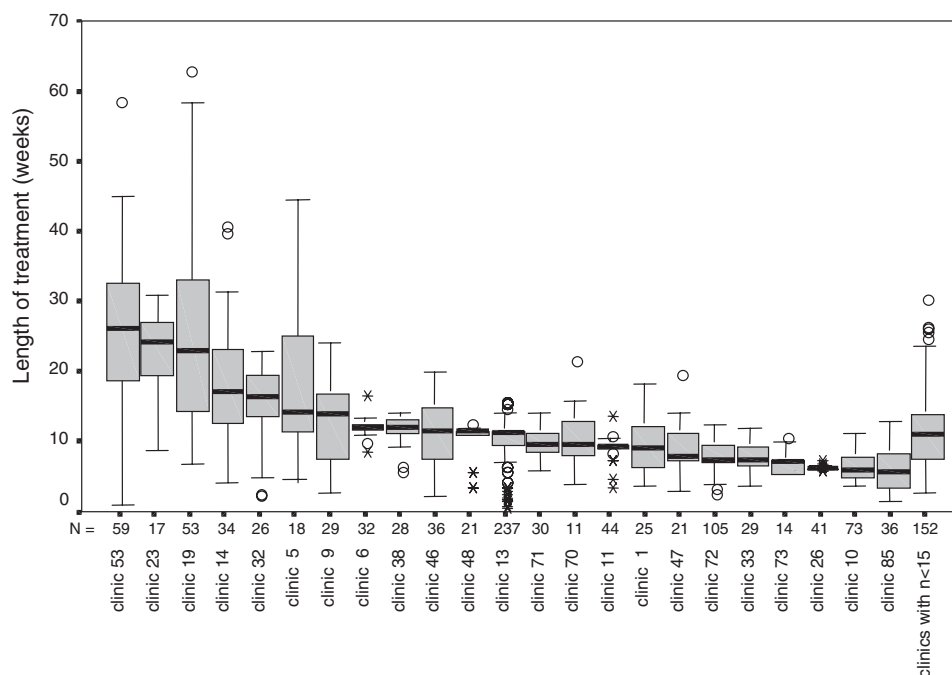

FIGURE 2. Distributions of therapy duration by the hospitals.

[3]Because the assumption of homogeneous variance for treatment duration in the hospitals obviously is not adequate (see Figure 2), an estimate of variance within each hospital was calculated using a modified method by Littell, Milliken, Stroup, and Wolfinger (1996).

TABLE 3. Proportion of Variance Explained of Treatment Duration Within and Between Hospitals

| | Model 1[a] | Model 2[b] | | Model 3[c] | |
	$\sigma^2 \pm SE$	$\sigma^2 \pm SE$	σ^2 explained	$\sigma^2 \pm SE$	σ^2 explained
Between	19.19 ± 4.69	19.95 ± 4.91	—	14.75 ± 3.69	23.1%
Within	20.44 ± 0.96	19.78 ± 0.94	3.2%	19.33 ± 0.91	2.3%

[a]Without predictors.
[b]Patient characteristics as predictors.
[c]Hospital characteristics.

the third step, data regarding the specialty status of the treatment offered were included (Model 3). As can be seen in Table 3, patient characteristics accounted for only a small percentage of variance within the hospitals. These patient variables did not explain variance between the hospitals.

Detailed results for Model 3 can be found in Table 4, which lists all variables with a statistically significant effect. The strongest effect was found in hospitals with a separate eating disorders ward. When compared with a specialty eating disorders hospital, their patients received 7 additional weeks of treatment on average. A few patient characteristics were moderate predictors of outcome. Patients with low treatment motivation at baseline received half a week less treatment on average than those with higher motivation.

Although the coefficient of 4.06 for BN diagnosis appears significant, it does not indicate that BN was treated 4 weeks longer than AN on average. Three further significant interactions are involved in interpreting diagnosis coefficient. For AN patients a difference of 1 point on the SCL-90-R Global Severity Index (GSI) correlates with an extended treatment length of half a week, whereas for BN patients the same difference on the GSI leads to 0.24 weeks shorter treatment because of a GSI × BN interaction. A weight gain of 16% BMI, which equals an increase in weight from the upper diagnostic boundary of 17.5 BMI to the expected BMI of 20, results in a treatment shortened by 0.1 weeks for AN patients. This correlation is even stronger for

TABLE 4. Selected Predictors of Treatment Duration

Predictor	B	SE
Other vs. specialty hospital	0.42	3.95
Specialized program vs. specialty hospital	−0.33	3.98
Specialized ward vs. specialty hospital	7.29	4.36*
Motivation to change: no	−0.49	0.16***
Psychological distress (GSI)	0.50	0.18***
Age	0.013	0.016
Weight (% of expected BMI)	−0.006	0.004
BN diagnosis	4.06	1.32***
Weight × Diagnosis (BN)	−0.05	0.01***
Psychological Distress × Diagnosis (BN)	−0.74	0.23***
Age × Diagnosis (BN)	0.05	0.03**

Note. GSI = Global Severity Index of the Symptoms Checklist-90-R; BMI = body mass index; BN = bulimia nervosa.
*$p < .1$. **$p < .05$. ***$p < .01$.

BN patients; a weight gain of 16% BMI results in a treatment shortened by 0.75 weeks. Because BN patients have a higher weight at baseline,[4] the effect is even more pronounced. The treatment duration of very overweight BN patients (greater than 130% BMI) is approximately 2 weeks shorter than that of BN patients with expected BMI. For AN and BN patients with 90% of expected body weight, a difference of only 0.56 weeks is found. These examples underscore once more the role of interaction effects.

For Which Patients Does Inpatient Psychodynamic Therapy Provide Which Rate of Recovery, and Is Treatment Duration a Predictive Factor?

Outcome rates at the end of treatment and at the 2.5-year follow-up assessment are presented in Table 5. Essentially, the present definition of treatment success implies an almost complete lack of symptoms or only symptoms for which immediate further treatment is not necessary. Positive outcome thus defined (see Table 1) was found in a minority of AN patients and in patients who met both AN and BN criteria: 11% and 17%, respectively, at the end of treatment. Self-report evaluations did not differ substantially from therapist evaluations. Positive outcome rates for BN were markedly higher. However, outcome success rates as assessed in therapist evaluations (45%) differed greatly from self-evaluations (31%). At the 2.5-year follow-up assessment, a significantly greater proportion of patients with AN (36% based on therapists reports and 33% based on patients reports) and with AN and BN (26% vs. 21%) were to a large extent symptom free. However, the rate of positive outcome for BN dropped slightly (36% based on therapists' reports vs. 22% based on patients' reports).

The propensity score was used to estimate the effect of treatment duration on outcome, thus reducing potential bias resulting from confounding variables (see Method section). This method compares only two alternatives. In this study, the effect of short and long treatment was compared. The treatments are thus dichotomized at the median (AN: 11.1 weeks; BN: 11.4 weeks). With the exception of treatment duration and intensity, all variables listed in Table 2 are included in the propensity score adjustment. Figure 3 shows the effect using the example of low body weight in AN. As can be seen, the mean BMI of shorter and longer treated patients is not significantly different within each quintile. The same effect can be applied to the other variables. Overall, the propensity score reduces the bias of confounding variables by 72% for AN and 88% for BN (Liebermann et al., 1996). It is important to note that hospital effects cannot be controlled by this method. Because some of the hospitals used only short treatments, it is inevitable that a certain proportion of shorter treatment patients were in different hospitals than those who received longer treatment. Therefore, the potential hospital effect could not be counterbalanced for treatment duration.

The comparison of the success rate for shorter and longer treatments is demonstrated by the odds ratios[5] with propensity score adjustments in Table 5. The adjusted odds ratios indicate that a patient-reported difference in outcome at the end

---

[4]AN and BN had a different distribution for body weight at index admission: The median for anorexics was 71.5% BMI and for bulimics 102.3% BMI. The distribution overlapped between 75% and 93% BMI.
[5]An odds ratio of 1 indicates that the probability of success is equal for short and long treatments. An odds ratio of 2 indicates a success rate twice as high for long treatments, and an odds ratio of 0.5 indicates half as high a success rate compared with short treatment.

TABLE 5. Rate of Success (in %) at Discharge and at 2.5 Years Follow-Up Classified According to Treatment Duration

Perspective	Discharge				2.5 years follow-up			
	≤ 11 Weeks	> 11 Weeks	Total	OR	≤ 11 Weeks	> 11 Weeks	Total	O
				Anorexia nervosa				
	n = 166	n = 170			n = 112	n = 113		
Patients	7.8	13.4	10.6	1.1	34.8	31.0	32.9	0.7
Therapists	9.6	15.1	12.4	1.8	36.6	34.5	35.6	0.8
				Bulimia nervosa				
	n = 292	n = 303			n = 197	n = 202		
Patients	33.0	29.0	31.0	0.7[a]	20.6	24.0	22.2	1.0
Therapists	48.0	43.0	45.4	0.6	35.5	37.1	36.3	0.9
				Anorexia and bulimia nervosa				
	n = 73	n = 95			n = 44	n = 65		
Patients	13.7	15.8	14.9	1.2	27.3	16.9	21.1	0.5
Therapists	13.7	19.0	16.7	1.5	31.8	21.6	25.7	0.6

Note. Odds ratios after propensity score adjustment (with the exception of anorexia and bulimia nervosa).
[a]1 within the 95% confidence interval.
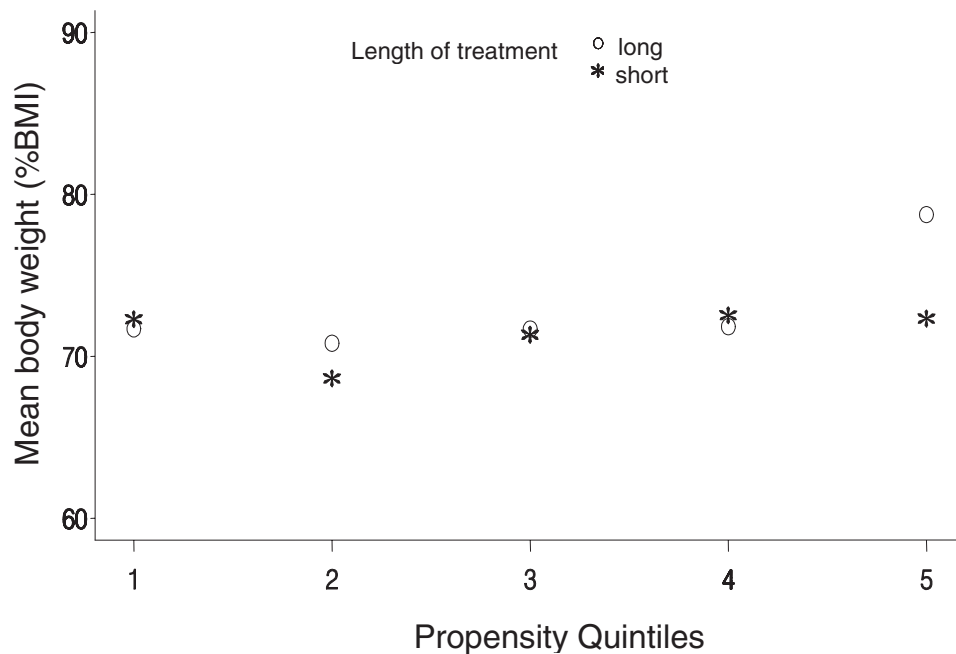
FIGURE 3. Illustration of the effect of propensity adjustment: mean body mass index (BMI) by quintiles (anorexia nervosa).

of treatment was found only in BN patients. For patients with bulimia, the patient-reported positive outcome rate for shorter treatments was 33% higher than the rate for longer treatments. The therapist-reported success rate for patients with anorexia was about 80% higher for longer term versus shorter term cases. However, this difference in success rates was not statistically reliable (despite the considerable sample sizes). Overall, there was no recognizable difference in outcome between shorter and longer treatment at the 2.5-year follow-up assessment. This analysis does not preclude the possibility that unconsidered variables may interact with treatment duration and intensity in predicting outcomes.

Correlations Between Eating Disorder Treatment Outcome and Predictors

The following analyses focus on the identification of possible predictors of treatment success at the 2.5-year follow-up.[6] The present analysis made use of a stepwise multivariate logistic regression method.[7] No substantial differences between outcome at the various hospitals could be found. The analysis of the predictive variables of treatment success used a simple logistic regression approach, excluding the hospital factor from the design.

Analyses were conducted within AN and BN groups, separately. The variables listed in Table 2 were investigated as predictors of outcome status. The propensity score and treatment duration, longer term versus shorter term, were also included. To gain an accurate picture of the possible association between treatment duration and outcome, all predictors were examined in interaction with treatment duration. Only those interactions that contributed significantly to the model were used further. The outcome criterion was based on patient reports. The resulting models are described in Table 6. The resulting models for AN and BN differ with regard to the identified predictors as well as to the goodness of fit. It is of particular interest that treatment length was found to interact with different variable sets for the AN analyses when contrasted with the BN analyses.

Overall, the goodness of fit was moderate for both models. The model for AN explained 31% of the variance. This was reduced to 25% after conducting a bootstrap adjustment and was thus relatively reliable. The model for BN showed 13% variance explained. This was reduced to 6.5% after a bootstrap adjustment, thus presenting a low reliability.

Even after controlling for other covariates, treatment intensity had no effect on the outcome for patients with AN. However, treatment length (short vs. long) in interaction with the age variable was a significant predictor. Low body weight and low desired body weight (%BMI) at treatment admission and significant deviations as measured on the FPI-R were risk factors associated with a poor outcome. The interaction between treatment duration and age is shown as an example in Figure 4: Overall, high patient age indicates a poorer outcome. The predictive value of this variable was intensified in the outcome of younger patients: Those in the 18- to 20-year age range had two to four times higher rates of positive treatment outcome with shorter treatments than age-matched counterparts with longer treatments. This cor-

---

[6]This is because the amount of treatment, the duration and intensity, can only be interpreted here as a real predictor and possible causal factor. Decisions as to the length of treatment are assumed to be associated with the outcome or the target outcome, although this association is currently not clarified and remains so within this study.

[7]We used GLIMMIX-Macros (Littell et al., 1996, Chapter 11).

TABLE 6. Predictors of Therapy Success (Patient Report)

AN[a]			BN[b]		
Predictor	B	SE	Predictor	B	SE
Propensity score	1.02	0.80	Propensity score	−0.07	1.04
Short vs. long treatment	−1.41	0.76*	Short vs. long treatment	3.90	1.52**
Age[c]	−1.98	0.68***	Age	0.03	0.03
Personality (FPI Extreme)	−0.75	0.38**	Previous treatments	−0.33	0.13**
Weight (% expected BMI)	0.07	0.02***	EDI scales 1,2,3	−0.02	0.01*
Desired body weight (self-report)	0.07	0.02***	Anorectic symptoms: yes (vs. none)	−1.27	0.44***
Treatment Duration × Age	0.71	0.41*	Multi-impulsivity	−0.53	0.26**
			Morgan-Russell E scale	0.04	0.08
			Treatment Duration × Age	0.10	0.0**
			Treatment Duration × Morgan-Russell E scale	−0.21	0.11**

Note. AN = anorexia nervosa; BN = bulimia nervosa; FPI = Freiburg Personality Inventory; BMI = body mass index; EDI = Eating Disorder Inventory; E scale = Socioeconomic State scale.
[a]R^2 (without bootstrap) = .31; R^2 (with bootstrap) = .25.
[b]R^2 (without bootstrap) = .13; R^2 (with bootstrap) = .07.
[c]Logarithmic association.
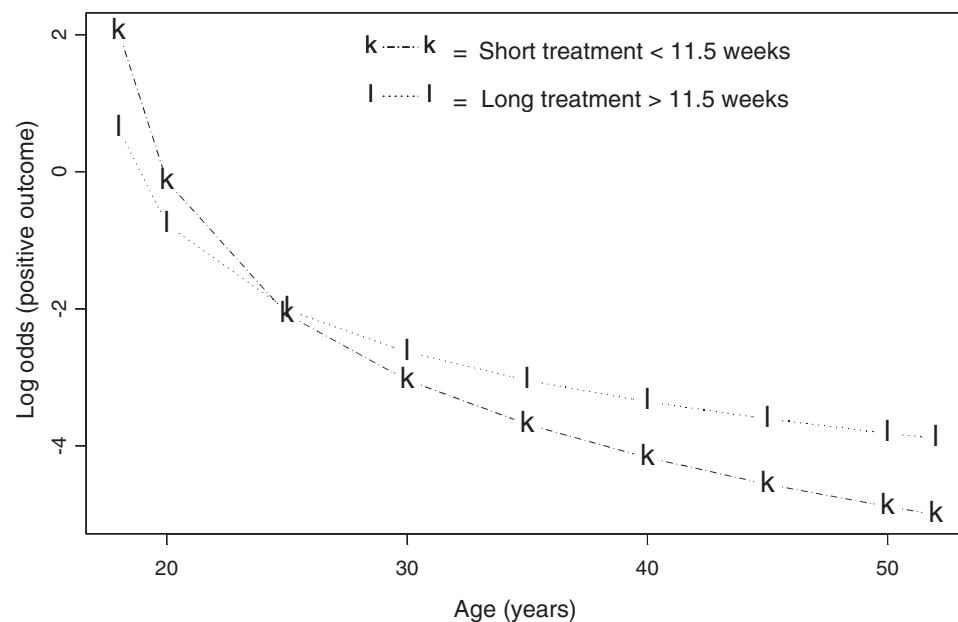*$p < .1$. **$p < .05$. ***$p < .01$.

FIGURE 4. Interaction between treatment duration and age in association with rate of success (anorexia nervosa).

relation is reversed for older patient outcomes: Those patients who received a longer treatment actually had an increased rate of good outcome. However, this rate as a whole was considerably lower than that of younger patients.

The results for BN are quite complex: Multi-impulsivity, additional anorectic symptoms, and high number of previous treatments were associated with lower success rates. Treatment success was only slightly lower for patients who scored high on the first three EDI subscales (i.e., those with a more severe eating disorder syndrome). The effect of treatment duration is complicated by interaction effects with age and with the Morgan-Russell E subscale, in which relationships to family and friends are presented. The association between the three variables is shown in Figure 5: Patients with difficulty in establishing relationships (low Morgan-Russell scores) had a better prognosis when treated for a longer rather than a shorter period. If patients scored at least moderately on psychosocial functioning, the reverse effect resulted in the good outcome increasing with shorter treatment. The advantage of longer treatment for outcome was stronger for younger patients with poor social adjustment than for older patients, whereas for older patients with good psychosocial functioning the advantage of shorter treatment was stronger than for younger patients.
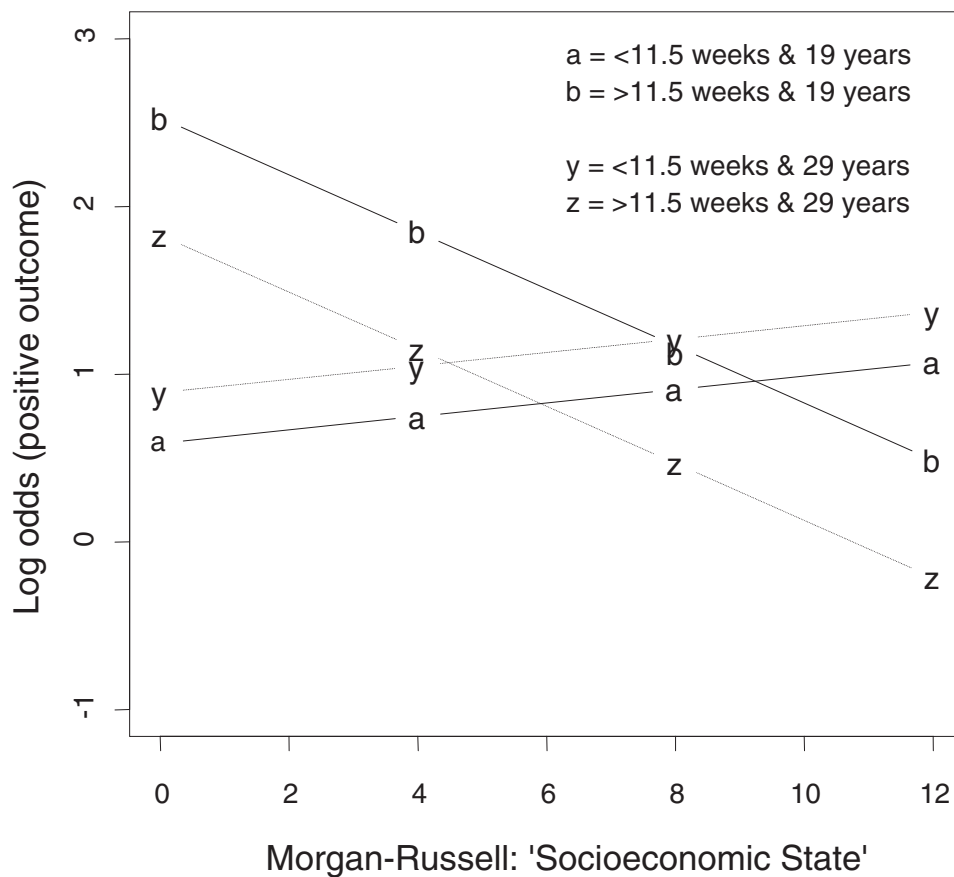

FIGURE 5. Interaction between treatment duration, age, and Morgan-Russell Socioeconomic State scale in association with rate of success (bulimia nervosa).

Discussion

The treatment of eating disorders remains challenging. Although our results are significant, some findings reflect the difficulties that are inherent in the psychodynamic treatment of eating disorders. The success rates at the end of treatment were consistent, to some degree, with those achieved under research conditions in other studies. Mitchell et al.'s (1996) overview of the effect of psychotherapy on BN estimated that 40 to 60% of patients were largely free from symptoms at discharge. Fichter et al. (1994) reported that 54% of bulimic patients who received a mean 12-week inpatient behavioral therapy were rated as having a good outcome 2 years after discharge. Our study showed a lower rate of symptom remission at the 2.5-year follow-up assessment. Furthermore, the differences of the success rates between discharge and 2.5-year follow-up indicate a high rate of relapse, characteristic of a cyclical course of illness with recurring symptomatic phases. The somewhat higher rate of recovery found by Fichter et al. could be related to the behavioral treatment's focus on symptom and relapse management; thus, patients may have been better prepared for maintaining treatment gains after hospitalization. However, this hypothesis requires further testing.

Not many treatment studies of AN have been conducted, and the results of existing studies are difficult to interpret because of various problems, especially treatment as well as study dropout (Crisp et al., 1992; van Strien, van der Ham, & van Engeland, 1992). In this respect, a solid empirical basis is lacking (Roth & Fonagy, 1996). Published results of a long-term study shift attention to outcomes over time: Strober et al. (1997) reported that 21% and 33% of adolescent anorectic patients no longer met the main diagnostic criteria of low body weight and amenorrhea 2 and 3 years after admission, respectively. One percent and 9% of patients were free of all criterion symptoms 2 and 3 years after admission, respectively. Comparisons with such studies can be problematic because definitions of successful outcome or full recovery are subject to differing criteria. The definitions of full recovery by Strober et al. (1997) and Fichter and Quadflieg (1997) are consistent with that of Project TR-EAT in that the criterion of symptom-free status was used, even when this criterion was operationalized differently.

The interpretations of findings concerning the effect of treatment duration and intensity on outcome are rather challenging. Differences in the length of treatment assigned can only partly be explained through patient characteristics or sociodemographic status at admission. Obviously, hospital-specific resource management factors (largely based on local treatment options, therapist's preferences, and experiences and only marginally based on external, empirical studies) have great effect on determining the length of treatment. The high mean treatment duration in hospitals with a separate eating disorders ward should be critically examined to determine whether higher rates of recovery truly accompany the high resource investment in such settings.

If the syndrome status of the patient does not correlate with differences in the length of treatment assigned, to what extent can a possible increase in rate of recovery justify the implementation of a longer therapy? Success rates were almost equal for short and long treatment when they were controlled for possible confounding variables. Only a more detailed analysis allows for the identification of subclasses, in which the treatment duration is predictive of treatment outcome.

The pretreatment characteristics of anorexic patients are slightly predictive for treatment outcome. As one would expect, a severely low body weight at admission

is related to outcome. Severely underweight AN patients, although treated on average for a longer period, had a poorer rate of recovery at the 2.5-year follow-up assessment than those who had a higher BMI. The longer inpatient treatment duration could be related to the time required to reach target weight. Similar results were found when investigating the effect of age and the number of previous treatments; both variables can be used to indicate the degree of chronicity. Patients who display a higher level of chronicity benefit more from longer treatments, whereas those who are not chronically ill benefit more from shorter treatments. However, patients who were about 25 years old (the mean age of the sample) have similar predictions of outcome for both longer and shorter treatments. The higher rate of success in younger patients suggests that a long-term treatment outcome of full-symptom remission is feasible for these individuals, whereas this target is possibly too high for older and chronically ill patients. These individuals could target reaching a level of symptoms that allows them to lead a normal life with reasonable level of life satisfaction.

The model for the prediction of positive outcome with only 6.5% of the variance explained is not very impressive and calls for caution in interpreting the identified predictors, especially because of the low stability of the model. Severe multi-impulsivity, a high number of previous treatments, and additional anorectic symptoms affect the rate of success adversely. Treatment duration in interaction with psychosocial functioning seems to be an indicator of outcome: The probability of a good outcome increased for those patients with good social adjustment. This effect was intensified in older patients and suggests that long treatment (i.e., long absence) has a reduced effect for patients with good social functioning, especially when those patients are married. However, this effect reverses for young patients with difficulties establishing relationships: They do not have these problems in a longer treatment.

The participation of such a large sample of hospitals in Germany not only is an advantage but also limits possible conclusions. For example, the differences in patient mix between the hospitals limit the interpretation of the observed effect of the specialty level of the treatment. The possible bias for this factor alone or in association with the various clinical settings could not be adjusted with the propensity score method (because that would have required that all hospitals apply longer as well as shorter treatments).

Of note for the findings regarding treatment length is the range that was used in this study: The treatments investigated run from 5 to approximately 16 weeks. Treatments shorter than 5 weeks were excluded from the analyses because it was uncertain to what extent the duration was intended and what proportion could be attributed to dropouts. Treatments longer than 16 weeks were not found frequently enough to justify inclusion. Furthermore, the possible effect of many variables not included in the model remained unobserved (e.g., parameters for the course of illness not included in this study, further characteristics of the hospitals, and parameters for the treatment program). Other possible factors within the follow-up period, such as the effect of further outpatient treatment or critically decisive life events, were not included in this study and thus limit generalization.

Despite these limitations, suggestions for the treatment of eating disorders can be made. The results of this study suggest that inpatient psychodynamic treatment should have a duration of at least 8 to 12 weeks, whereas a longer treatment for older patients older than 40 years could be beneficial. Further treatment extensions should be based on individual cases as well as on the course of improvement. The decision for further inpatient or outpatient treatment should not be independent of patient status at point of discharge. Rather, new and specified treatment methods

should be developed. The results of the medium-term course of eating disorders (Kordy et al., in press) suggest an increased risk of relapse in the period immediately after hospitalization. This could be counteracted by a maintenance treatment. Booster therapies could be such an approach to reduce the rate of relapse that occurs a few months after discharge. Stepwise care provision could serve as guiding principle for the development of a comprehensive treatment strategy (Royal College of Psychiatrists, 1992). Experiences with such strategic approach are positive in those countries that do not separate inpatient and outpatient treatment as strictly as it is traditionally done in Germany.

References

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models.* Newbury Park, CA: Sage.

Bulik, C. M., Sullivan, P. F., Joyce, P. R., Carter, F. A., & McIntosh, V. V. (1998). Predictors of 1-year treatment outcome in bulimia nervosa. *Comprehensive Psychiatry, 39,* 206–214.

Crisp, A. H., Callender, J. S., Halek, C., & Hsu, L. K. (1992). Long-term mortality in anorexia nervosa. A 20-year follow-up of the St. George's and Aberdeen cohorts. *British Journal of Psychiatry, 161,* 104–107.

Crisp, A. H., Joughin, N., Halek, C., & Bowyer, C. (1996). *Anorexia nervosa: The wish to change.* Hove, East Sussex, England: Psychology Press.

D'Agostino, R. B. (1998). Tutorial in biostatistics: Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Statistics in Medicine, 17,* 2265–2281.

Efron, B., & Tibshirani, R. J. (1993). *An introduction into the bootstrap.* New York: Chapman & Hall.

Fairburn, C. G., Norman, P. A., Welch, S. L., O'Connor, M. E., Doll, H. A., & Peveler, R. C. (1995). A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments. *Archives of General Psychiatry, 52,* 304–312.

Fichter, M. M., & Quadflieg, N. (1997). Six-year course of bulimia nervosa. *International Journal of Eating Disorders, 22,* 361–384.

Fichter, M. M., Quadflieg, N., & Reif, W. (1994). Course of multi-impulsive bulimia. *Psychological Medicine, 24,* 591–604.

Goldstein, H. (1995). *Multilevel statistical models* (2nd ed.). London: Arnold.

Hebebrand, J., Heseker, H., Himmelmann, W., Schäfer, H., & Remschmidt, H. (1994). Alterspercentile für den Body-Mass-Index aus Daten der Nationalen Verzehrstudie einschließlich einer Übersicht zu relevanten Einflußfaktoren. *Aktuelle Ernährungsmedizin, 19,* 259–265.

Herzog, D. B., Keller, M. B., Lavori, P. W., & Sacks, N. R. (1991). The course and outcome of bulimia nervosa. *Journal of Clinical Psychiatry, 52,* 4–8.

Herzog, T., & Hartmann, A. (1997). Psychoanalytisch orientierte Behandlung der Anorexia nervosa. Eine methodenkritische Literaturübersicht unter Verwendung meta-analytischer Verfahren. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 47,* 263–280.

Herzog, T., Hartmann, A., & Sandholz, A. (1996). Psychotherapiedauer und Psychotherapiedosis. Die Freiburger prospektive kontrollierte Studie zur Kurz-Psychotherapie der Bulimia Nervosa. In H. Hennig, E. Fikentscher, U. Bahrke, et al. (Eds.), *Kurzzeit-Psychotherapie in Theorie und Praxis* (972–990). Lengerich, Germany: Pabst Science.

Herzog, W., Deter, H. C., Fiehn, W., & Petzold, E. (1997). Medical findings and predictors of long-term physical outcome in anorexia nervosa: A prospective, 12-year follow-up study. *Psychological Medicine, 27,* 269–279.

Hosmer, D., & Lemeshow, S. (1989). *Applied logistic regression.* New York: Wiley.

Kächele, H., für die Studiengruppe MZ-ESS. (1999). Eine multizentrische Studie zu Aufwand und Erfolg bei psychodynamischer Therapie von Eßstörungen. Studiendesign und erste Ergebnisse. *Psychotherapie, Psychosomatik, Medizinische Psychologie, 49,* 100–108.

Keel, P. K., Mitchell, J. E., Miller, K. B., Davis, T. L., & Crow, S. J. (1999). Long-term outcome of bulimia nervosa. *Archives of General Psychiatry, 56,* 63–69.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The Longitudinal Interval Follow-up Evaluation. A comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44,* 540–548.

Kordy, H., Krämer, B., Palmer, R. L., Papezova, H., Pellet, J., Richard, M., Treasure, J., & COST Action B6. (in press). Conceptualisation and rationale for consensus definition in eating disorders: Response, remission, recovery, relapse and recurrence. *Journal of Clinical Psychology.*

Liebermann, E., Lang, J. M., Cohen, A., D'Agostino, A., Jr., Datta, S., & Frigoletto, F. D. (1996). As-

sociation of epidural analgesia with cesareans in nulliparous women. *Obstetrics and Gynecology, 88,* 993–1000.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models.* Cary, NC: SAS Institute.

McKenzie, J., & Joyce, P. (1992). Hospitalisation for anorexia nervosa. *International Journal of Eating Disorders, 11,* 235–241.

Mitchell, J. E., Hoberman, H. N., Peterson, C. B., Mussell, M., & Pyle, R. L. (1996). Research on the psychotherapy of bulimia nervosa: Half empty or half full. *International Journal of Eating Disorders, 20,* 219–229.

Mitchell, J. E., Maki, D. D., Adson, E. E., Ruskin, B. S., & Crow, S. (1997). The selectivity of inclusion and exclusion criteria in bulimia nervosa treatment studies. *International Journal of Eating Disorders, 22,* 243–252.

Mitchell, J. E., Pyle, R. L., Pomeroy, C., et al. (1993). Cognitive behavioral group psychotherapy of bulimia nervosa: The importance of logistical variables. *International Journal of Eating Disorders, 20,* 277–287.

Oehlschlägel, J., Malewski, P., & Mahon, J. (1999). How to define anorectic weight? *European Eating Disorders Review, 7,* 321–333.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70,* 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79,* 516–524.

Roth, A., & Fonagy, P. (1996). *What works for whom? A critical review of psychotherapy research.* New York: Guilford Press.

Strober, M., Freeman, R., & Morrell, W. (1997). The long-term course of severe anorexia nervosa in adolescents: Survival analysis of recovery, relapse, and outcome predictors over 10–15 years in a prospective study. *International Journal of Eating Disorders, 22,* 339–360.

Treasure, J., & Schmidt, U. (1999). Beyond effectiveness and efficiency lies quality in services for eating disorders. *European Eating Disorders Review, 7,* 162–178.

Treasure, J., & Smukler, G. I. (1995). Medical complications of chronic anorexia nervosa. In G. I. Smukler, C. Dare, & J. Treasure (Eds.), *Handbook of eating disorders: Theory, treatment and research* (pp. 197–220). New York: Wiley.

Treasure, J., Todd, G., & Smukler, G. I. (1995). The inpatient treatment of eating disorders. In G. I. Smukler, C. Dare, & J. Treasure (Eds.), *Handbook of eating disorders: Theory, treatment and research* (pp. 275–291). New York: Wiley.

Van Strien, D. C., van der Ham, & van Engeland, H. (1992). Drop out characteristics in a follow-up study of 90 eating disorder patients. *International Journal of Eating Disorders, 12,* 341–343.

Wilson, G. T., Loeb, K. L., Walsh, B. T., Labouvie, E., Petkova, E., Liu, X., & Waternaux, C. (1999). Psychological versus pharmacological treatments of bulimia nervosa: Predictors and processes of change. *Journal of Consulting and Clinical Psychology, 67,* 451–459.

Zusammenfassung

Die vorliegende Studie untersucht (a) welche Faktoren die Dauer und Intensität der Psychotherapie für Essstörungen bestimmen und (b) wie diese evtl. in Interaktion mit anderen Faktoren die Effektivität beeinflussen. In einem naturalistischen Design wurde der Symptomstatus von 1171 Patientinnen über einen Zeitraum von 2,5 Jahren gemessen, nachdem sie in einer der 43 teilnehmenden Fachkliniken zur Behandlung aufgenommen worden waren. Die statistische Modellierung erfolgte mit Verfahren der hierarchisch linearen Regression und logistischen Regression. Mögliche Verzerrungen durch konfundierende Variablen wurden mit Hilfe der Propensity-Score-Technik korrigiert. Die Behandlungsmodalitäten, insbesondere Dauer und Intensität, variierten beträchtlich zwischen und innerhalb der Kliniken, wobei Patientenmerkmale nur eine geringe Rolle spielten. In der 2,5–Jahresuntersuchung zeigte sich für etwa ein Drittel der Anorexie-Patientinnen und etwa ein Viertel der Bulimie-Patientinnen eine klinisch bedeutsame Symptombesserung. Der Einfluß von Dauer und Intensität auf die Ergebnisse war schwach und nur in Interaktion mit weiteren Patienten- bzw. Klinikmerkmalen zu erkennen. Implikationen zur Behandlungsplanung werden diskutiert.

Résumé

Cette étude investigue (a) les facteurs déterminant la durée du traitement et (b) l'effet de la durée du traitement et d'autres facteurs sur le résultat avec des patients souffrant d'un trouble alimentaire. Dans cette étude d'observation, le status symptomatique a été suivi pendant 2.5 ans après l'admission dans l'un des 43 hôpitaux participants. Les traitements et les résultats étaient modélisés à l'aide de modèles

hiérarchiques linéaires et de régression logistique. Pour contrôler d'éventuels facteurs confondants, l'ajustement du score de propension a été appliqué. Les modalités de traitement, notamment la durée et l'intensité, ont varié considérablement entre et à l'intérieur des hôpitaux tout en étant associées à un très faible degré aux caractéristiques des patients. Lors de la catamnèse de 2.5 ans, 33% des patients anorexiques et 25% des patients boulimiques étaient libres de symptômes. La durée de traitement a montré un effet faible sur les résultats et ceci seulement en interaction avec d'autres caractéristiques pertinentes des patients, alors que l'intensité du traitement n'avait pas d'impact sur le résultat. Des implications pour la planification de traitements sont discutées.

Resumen

El presente estudio investiga (a) factores que determinan la duración del tratamiento y (b) el efecto de esta duración y de otros factores sobre el resultado de la terapia en el caso de pacientes con desórdenes alimentarios. En este estudio observacional realizado en uno de los cuarenta y tres hospitales par- ticipantes, se observó el estatus sintomático de mil ciento setenta y un pacientes durante los dos años y medio posteriores a su admisión. El tratamiento y sus resultados se modelizaron utilizando los modelos lineales jerárquicos y la regresión logística. Para controlar los posibles factores de confusión, se usó el ajuste del puntaje de propensión (*propensity score adjustment*). Las modalidades de tratamiento, especialmente su duración e intensidad, variaron considerablemente de un hospital a otro así también como en cada uno de ellos, y se relacionaron muy poco con las características del paciente. Luego de dos años y medio de seguimiento, el 33% de los pacientes anorécticos y el 25% de los bulímicos se encontraron libres de síntomas. La duración del tratamiento mostró una débil influencia sobre los resultados y esto sólo cuando se la consideró en interacción con otras características pertinentes de los pacientes; no así la intensidad del tratamiento, que fue independiente de los resultados. Se discuten las implicaciones de los planes de tratamiento.