Things change: The early identification of patients with an unfavourable prognosis

Boer, S.

Chapter 3

Prediction of prolonged treatment course for depressive and anxiety disorders in an outpatient setting: the Leiden routine outcome monitoring study.

S. Boer[1,2,4]
O.M. Dekkers[1]
S. le Cessie[1,3]
I.V.E. Carlier[2]
A.M. van Hemert[2]

[1] Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden

[2] Department of Psychiatry, Leiden University Medical Centre, Leiden

[3]. Department of Medical Statistics and Bio-informatics, Leiden University Medical Centre, Leiden

[4] Department of Biomedical Data Sciences (section Medical Decision Making), Leiden University Medical Centre, Leiden

Journal of affective disorders. 2019; 247: 81-87

ABSTRACT

Objective

The aim of this study was to improve clinical identification of patients with a prolonged treatment course for depressive and anxiety disorders early in treatment.

Method

We conducted a cohort study in 1.225 adult patients with a depressive or anxiety disorders in psychiatric specialty care setting between 2007 and 2011, with at least two Brief Symptom Inventory (BSI) assessments within 6 months. With logistic regression, we modelled baseline age, gender, ethnicity, education, marital status, housing situation, employment status, psychiatric comorbidity and both baseline and 1st follow-up BSI scores to predict prolonged treatment course (> 2 years). Based on the regression coefficients, we present an easy to use risk prediction score.

Results

BSI at 1st follow-up proved to be a strong predictor for both depressive and anxiety disorders (OR = 2.17 (CI95% 1.73-2.74); OR = 2.52 (CI95% 1.86-3.23)). The final risk prediction score included BSI 1st follow-up and comorbid axis II disorder for depressive disorder, for anxiety disorders BSI 1st follow-up and age were included. For depressive disorders, for 28% of the patients with the highest scores, the positive predictive value for a prolonged treatment course was60% (sensitivity 0.38, specificity 0.81). For anxiety disorders, for 35% of the patients with the highest scores, the positive predictive value for a prolonged treatment course was 52% (sensitivity 0.55, specificity 0.75).

Conclusions

A high level of symptoms at 2-6 months of follow-up is a strong predictor for prolonged treatment course. This facilitates early identification of patients at risk of a prolonged course of treatment; in a relatively easy way by a self-assessed symptom severity.

INTRODUCTION

Depressive and anxiety disorders are the most common mental disorders (Vos et al. 2012), with an estimated prevalence of respectively 298 and 273 million people worldwide. These disorders are associated with a high burden of disease (Wittchen et al. 2011) and high impact on society (Gustavsson et al. 2011), translating into substantial direct and indirect costs. Direct costs are related to treatment and the use of other health care services, and indirect costs to reduced quality of life, loss of productivity, absenteeism and functional impairment in many other personal and interpersonal areas of life (Donohue & Pincus, 2007; Combs & Markman, 2014).

The course of depressive disorders is variable, with approximately 60% of patients recovering within the first six months after diagnosis and up to 80% within two years (Steinert et al. 2014). Recurrence risk is 15-40% in two years. A persistent course with no major improvement despite treatment over two years or more, has been reported for 5 to 20% of patients, although slow improvements tend to continue over time (Hardeveld et al. 2010; Stegenga et al. 2012; Riihimaki et al. 2014; Steinert et al. 2014). For anxiety disorders the initial course is less favourable, with only 46% of patients recovering within two years and a similar recurrence risk of 15-40%, depending on type of anxiety disorder (Steinert et al. 2005; Penninc et al. 2011; Bruce et al. 2013).

In general, slow and incomplete recovery is associated with longer treatment duration (Riihimaki et al. 2014) and a longer treatment duration is associated with higher healthcare resource utilization (Haller et al. 2014); as for example more (severe) symptoms for patients with a prolonged treatment course, comorbidities, or treatment resistance in patients with a prolonged treatment course (Von Korff et al. 1992, Crown et al. 2002, Richards 2011, Dennehy et al. 2015). The majority of healthcare resources are consumed by a relatively small group of patients with a prolonged treatment course (Rais et al. 2013, Robinson et al. 2016).

Several studies have found that early response to treatment within two to eight weeks partially predicts further recovery (Van et al. 2008; van Calker et al. 2009; Tadic et al. 2010; Kim et al. 2011; Baldwin et al. 2012). Identification of patients with an unfavourable initial course of treatment could provide opportunities to target this subgroup with higher intensity treatment and potentially reduce chronicity early in the course of treatment (Trivedi & Baker, 2001; Lutz et al. 2009; Kendrick et al. 2016). Given that only limited data are published to support this, further research is implicated.

The implementation of Routine Outcome Monitoring (ROM) in mental health care provides an opportunity to study treatment course and symptom change, measured by general symptom inventories, such as the Brief Symptom Inventory (BSI) (Lutz et al. 2009, Katon et al. 2010, de Beurs et al. 2011). In the current study, we aimed to improve the clinical prediction of treatment duration for depressive and anxiety disorders in a routine care outpatient setting, and to identify patients with an unfavourable prognosis early in treatment course. Especially, we aimed to assess the role of the BSI, as an indicator of composite symptom severity, to predict prolonged treatment course and to develop an easy to use prediction model.

METHODS

This is a naturalistic cohort study with routine outcome monitoring (ROM), being collected in routine care by GGZ Rivierduinen, a Regional Mental Health Care Provider in the Western part of The Netherlands.

Since 2002, all patients referred to GGZ Rivierduinen for treatment of mood, anxiety and somatoform disorders are routinely assessed with a psychometric test battery. Data on diagnosis and severity of psychiatric symptoms are collected at intake, after treatment is initiated, and subsequently every 3-4 months. ROM includes self-reported and observer-rated measures, as well as generic and disorder-specific questionnaires. Completion of ROM questionnaires is supervised by trained psychiatric research nurses (or psychologists), not involved in treatment. ROM data are primarily used for diagnosis and to inform clinicians and patients about treatment progress. A detailed description of ROM can be found elsewhere (de Beurs et al. 2011).

For the purpose of research, patient-identifiable data were removed from the database to secure patients' confidentiality and to comply to Dutch law on research with clinical data. The Medical Ethical Committee of the LUMC approved the general study protocol regarding ROM, in which ROM is considered as an integral part of the treatment process (no written informed consent is required and the use of anonymized data for research is approved). A comprehensive protocol (titled "Psychiatric Academic Registration Leiden database") was used, to safeguard the anonymity of participants and ensure proper handling of the data. None of the participants objected to the anonymized use of their data for scientific purposes.

Patients

For the current study, we selected patients aged 18-65 years, who were referred to GGZ Rivierduinen between January 2007 and June 2011, with a primary clinical diagnosis of a depressive or anxiety disorder according to the attending physician. In the administrative system of GGZ Rivierduinen, the primary clinical diagnosis represents the primary focus of clinical care. The diagnostic classification was based on the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) and included depressive disorders (coded as 296.20 - 296.24, 296.30 - 296.34, 296.90, 300.4 or 311) and anxiety disorders (coded as 300.00, 300.01, 300.02, 300.21-300.23, 300.29, 300.3, 308.3 or 309.81). Bipolar and cyclothymic disorders were not included. For every patient, only the first treatment episode in the study period was considered. The selection was further restricted to patients with at least two ROM assessments within the first six months of treatment, one at baseline and a second ROM at least eight weeks after treatment initiation. As patients with only one ROM assessment were most likely not treated in the outpatient clinics, and no early treatment response can be assessed with only one assessment. We included only patients who had a continuation of treatment for at least six weeks after the second ROM assessment, continuation of treatment was defined as the absence of administrative termination of care. As the database closed June 2013, we accounted for the problem of right truncation by only including incident patients with a diagnosis up to June 2011. Therefore, each patient had a potential follow-up period of two year. The final sample included 1,225 patients, with minimum treatment duration of 14 weeks as a result of the specified inclusion criteria. See flow chart (figure 1) for details.

Psychiatrists and clinical psychologists or psychotherapists provided outpatient treatment in accordance with national evidence-based guidelines, consisting of pharmacotherapy, psychotherapy, or a combination of both. Psychotherapy for anxiety and depressive disorders, according to national guidelines, is mostly time limited cognitive behavioural therapy or interpersonal therapy. In our naturalistic cohort data, we did not have sufficient detail to accurately capture provision of care for individual patients and these data were not included in the analyses. From previous studies in GGZ Rivierduinen, we know that MDD is more frequently treated with pharmacotherapy than psychotherapy (55% and 24% respectively), while this is the reverse for anxiety disorders (23% and 59%). For both conditions, the remaining minority is treated with combinations or with other treatments. Guideline adherence in general was good (van Fenema et al. 2012). As this is an observational study, we did not in any way influence treatment modalities.

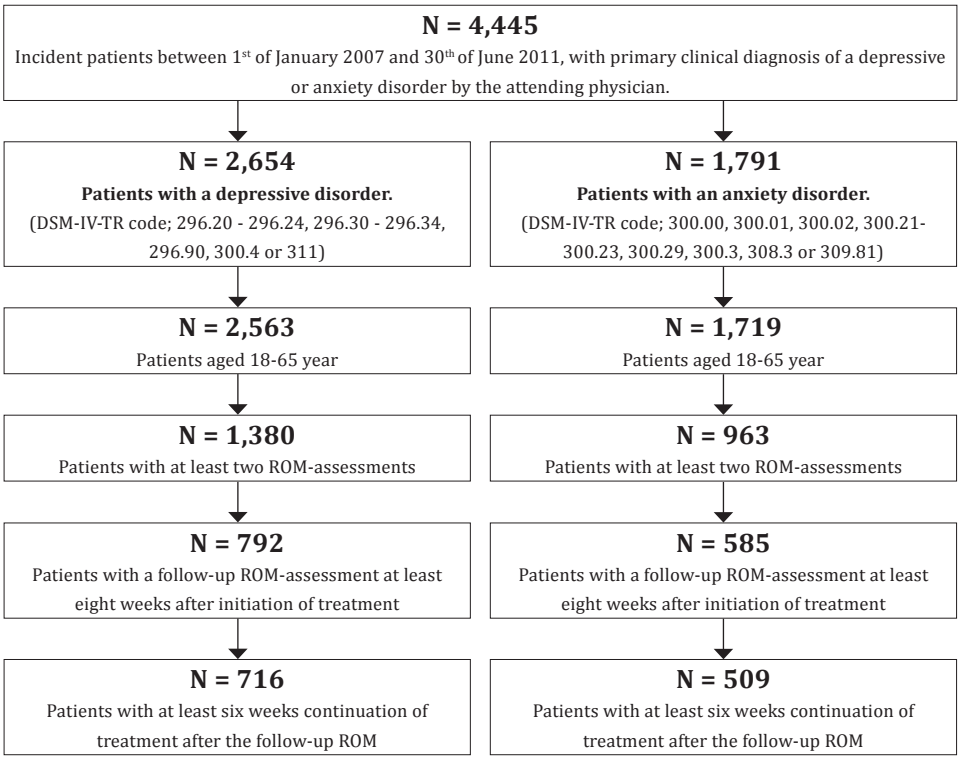

FIGURE 1. *Flowchart of patient selection*[1].

Outcome

The primary outcome of the study was prolonged treatment course. For depressive and anxiety disorders as primary treatment focus, it is common practice in the Netherlands to limit specialist care, as far as possible, to the symptomatic episodes. After an episode is successfully treated, it is common practice to refer patients back to primary care, where pharmacotherapy may be continued. Coverage of primary care is close to 100% in the Netherlands. We defined prolonged treatment course as a consecutive treatment duration of two years or more. For depressive and anxiety disorders, prolonged treatment courses, of two year or more, in general not only indicate difficult stabilization of the patients, but also constitute a high impact on service of costs and are preferably avoided. Treatment duration was defined as time since the first ROM-assessment to the administrative date of termination of care. Reasons for termination care are routinely coded in the administrative system. Around 80% of termination of care was mostly in

1 ROM =Routine Outcome Monitoring

mutual consensus, usually suggesting that treatment goals were obtained. One-sided termination by the patient can be considered as dropout, which amounted to 11%. Other reasons for termination include one-sided by the therapist (6%), moving home (<1%) or death (<1%).

Potential predictor variables

Demographic variables age and gender were obtained from the administrative data at baseline. Data on ethnicity, education, marital status, housing situation, and employment status were based on a ROM self-report questionnaire. Ethnicity was assumed to be Dutch, if both parents were born in the Netherlands. Education was divided into three levels 'low' (no education, primary school until approximately 10th grade), 'medium' (ranging from 11th grade through high school and community college) and 'high' (college undergraduate/graduate and higher). Marital status was categorized as 'married or cohabiting', 'divorced', 'widowed' or 'never in a relationship'. Housing situation was categorized as 'living alone', 'living with a partner without children', 'living without a partner, but with children', 'living with a partner and children, or living with family' and 'other'. Housing situation category 'other' was treated as missing as it contained fewer than 5 patients. For employment status, we distinguished between five categories 'housewife or -man, or retired', 'working', 'unemployed', 'on sick leave' or 'other'. Diagnostic information was extracted from the Dutch so-called Diagnosis-Treatment combination administrative data. Data on DSM-IV-TR diagnosis on Axis I and Axis II, other than the primary diagnosis, were extracted and considered as psychiatric comorbidity.

Ratings of psychiatric symptoms over the past seven days were obtained from the Brief Symptom Inventory (BSI). The BSI is a short form of the Symptom CheckList-90, consisting of 53 items rated on a 5-point Likert scale, and covering nine dimensions: somatization, obsessive-compulsive, internal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. Total BSI score, and scores on subscales, are computed as the sum of item-scores divided by the number of items. Higher scores indicate more severe psychiatric symptoms. The BSI has an acceptable internal consistency (Cronbach's alphas 0.71 to 0.85), and test-retest reliability (0.68 to 0.91) (Derogatis & Melisaratos, 1983). The total score and subscale scores are sensitive indicators of therapy effect (de Beurs et al. 2012). Here, we use the BSI-total score as an indicator of composite symptom severity, capturing not only symptoms of a particular disorder, but also symptoms of any comorbid symptoms. The normal range of the score may vary somewhat between populations. For the population of GGZ Rivierduinen we found in a previous study that 95% of a normal reference group will have BSI-total

score below 0.68, while more than 75% of a treated group will have a rating above this cut-off (Schulte-van Maaren et al. 2012).

General health status was assessed with the Short Form-36 (SF-36). The SF-36 is a 36-item self-report questionnaire that assesses health status in eight domains: physical functioning, social functioning, physical problems, emotional problems, mental health, vitality, bodily pain and general health (Ware & Sherbourne, 1992). For the purpose of this study, we decided to only use the subscale "general health" (Cronbach's alpha 0.70-0.81), based on previous literature (van Noorden et al. 2012).

Statistical analysis

Analyses were performed separately for patients with a primary depressive disorder and patients with a primary anxiety disorder. Baseline characteristics are summarized as number and percentage for categorical variables, or as mean and standard deviation (SD) for continuous variables. We used logistic regression analysis to predict prolonged treatment course, defined as treatment duration of two years or more since the first ROM-assessment. Baseline characteristics that were univariably associated with prolonged treatment course were selected for prognostic modelling. We used backward selection with criteria for variable removal of $p < 0.10$. Model performance was assessed and compared with the area under the receiver operating curve (AUROC) and internally validated with bootstrap resampling (n = 750). Model fit was assessed and compared by measuring Bayesian information criterion (BIC). Three models were compared:

1. Baseline characteristics, excluding baseline BSI (model 1)
2. Baseline characteristics, including baseline BSI (model 2)
3. Baseline characteristics, including baseline BSI plus BSI 1st follow-up (model 3)

Based on the regression coefficients of our final model a risk prediction score was developed as extensively described by Sullivan et al. (2004), in order to facilitate clinical application of the model. The risk prediction score was divided in quartiles and tested for sensitivity and specificity in relation to the outcome of prolonged treatment course (*see the online supplement for more details, appendix 1)*.

Sensitivity analyses

In a sensitivity analysis, we compared the baseline characteristics of our study population with the population of excluded patients who had only one BSI assessment, or a second BSI assessment outside our defined time range of 2-6 months. Furthermore, we analysed the final model 2 (baseline characteristics, including baseline BSI) within

the excluded population, and compared the results with the included population. Last, we performed additional prognostic modelling with BSI-delta (difference between baseline BSI and BSI 1st follow-up); baseline characteristics, including BSI 1st follow-up and BSI-delta.

Statistical calculations were performed using SPSS Version 20.0 for Windows (SPSS Inc., Chicago, III, USA) and R Statistical Software (version 3.2.3).

RESULTS

Patient characteristics

Of 1,225 patients that we included, 716 had a primary depressive disorder and 509 a primary anxiety disorder (table 1). In patients with a depressive disorder, the mean age was 41.2 years (SD 12.7) and 60.9% were female. The mean BSI was 1.38 (SD 0.70) at baseline, and symptoms significantly improved at first follow-up assessment (mean BSI 1.03; SD 0.70), on average 3.7 months after baseline. In patients with an anxiety disorder, the mean age was 35.6 (SD 12.7) and 62.3% were female. The mean BSI was 1.25 (SD 0.79) at baseline, and symptoms significantly improved at first follow-up assessment (mean BSI 0.91; SD 0.72), on average 4.1 months after baseline.

Predictors of prolonged treatment course in depressive disorders

The median treatment duration, given our selection, in depressive disorders was 16.7 months (IQR 8.0-28.1); 42.0% had a prolonged treatment course. In univariable logistic regression, prolonged treatment course was significantly predicted by BSI-baseline (OR 1.51, CI95% 1.22-1.87), BSI 1st follow-up (OR 2.09, CI95% 1.67-2.62), general health status, employment status, and comorbid Axis II disorders (*see the online supplement for more details, appendix 2)*. In multivariable logistic regression analyses model 1 (baseline variables but no BSI), the variables general health status and comorbid Axis II disorder remained in the model after backward selection. After including BSI-baseline (OR 1.37, CI95% 1.17-1.93) in model 2, no baseline variables were removed from the model. The remaining coefficients did not materially change. Adding the BSI 1st follow-up in model 3 yielded an OR of 2.17 (CI95% 1.73-2.74), while BSI-baseline was removed from the model. Model 3 included comorbid Axis II disorder and BSI 1st follow-up.

TABLE 1. *Patient characteristics of patients with a depressive disorder (n = 716) and patients with an anxiety disorders (n = 509). For continuous variables; values are stated as the mean (standard deviation). For categorical variables; values are numbers (percentages)[2].*

	Depressive disorders		Anxiety disorders	
Continuous variables	**Mean**	**SD**	**Mean**	**SD**
BSI baseline	1.38	0.70	1.25	0.79
BSI 1st follow-up	1.03	0.70	0.91	0.72
General health (SF36)	15.00	4.01	13.96	4.04
Age	41.2	12.7	35.6	12.7
Categorical variables	**N**	**%**	**N**	**%**
Female gender	436	60.9	317	62.3
non-Dutch ethnicity*	125	19.1	74	14.5
Employment status*				
Employed	276	42.1	233	49.7
Home	76	11.6	46	9.8
Unemployed	63	9.6	43	9.2
Sick leave	195	29.8	96	20.5
Other	45	6.9	51	10.9
Housing situation*				
Alone	146	22.5	92	19.6
Partner, no children	140	21.5	108	23.1
Partner and children, or with family	311	47.8	248	53.0
No partner, with children	53	8.2	20	4.3
Educational status*				
Low	279	42.6	182	38.8
Medium	261	39.8	197	42.0
High	115	17.6	90	19.2
Comorbid Axis I disorder	185	25.8	172	33.8
Comorbid mood disorder	26	3.6	95	18.7
Comorbid anxiety disorder	80	11.2	53	10.4
Comorbid somatoform disorder	30	4.2	23	4.5
Comorbid Axis II disorder	40	5.6	32	6.3

* Employment status, marital status and educational status was missing for 61 patients with a depressive disorder, and for 40 patients with an anxiety disorder. Housing situation was missing for 66 patients with a depressive disorder, and for 41 patients with an anxiety disorder. Ethnicity was missing for 62 patients with a depressive disorder, and for 40 patients with an anxiety disorder.

Model 3 had the highest internally validated AUROC of 0.65 and lowest BIC, pointing towards its better predictive performance than model 1 or 2 (respectively with an internal AUROC of 0.59 and 60). In medical diagnostic test evaluation, very high AUROCs (0.95 or higher) are desirable. In prediction studies AUROCs of 0.70 or higher would be considered strong effects (Hajian-Tilaki, 2013). Details of all three models are shown in the online supplement, *table 3A of appendix 2*.

2 BSI = Brief Symptom Inventory

Table 2A presents the risk prediction scores corresponding to the final model; higher scores indicate a higher risk of prolonged treatment course. From our final prediction model, we derived an easy to calculate score which ranged from 0 to 6; a score of 5 could not be achieved. For 27% of patients with a risk score of 4-6 points, the positive predictive value of prolonged treatment course was 60%, compared to 36% (1-negative predictive value) for patients with a score of 0-3 points. At this cut-off level sensitivity was 0.38 and specificity 0.81. See table 3A for details.

TABLE 2A. *Risk prediction score for prolonged treatment course (≥ two year) in patients with a depressive disorder; assessed after approximately three months of treatment[3].*

Factor	Points
Comorbid Axis II disorder	
No	0
Yes	2
BSI 1st follow-up	
< 0.53	0
0.53-0.89	1
0.89-1.45	2
> 1.45	4
Total score (range)	**0-6**

TABLE 3A. *Risk prediction score for prolonged treatment course (≥ two year) in patients with a depressive disorder; assessed after approximately three months of treatment.**

Risk prediction score	No prolonged treatment course	Prolonged treatment course	Total
0-3	338	187	525
4-6**	77	114	191

N = 716. * For 28% (N = 191) of patients with the highest scores (4-6), the positive predictive value for a prolonged treatment course was 60% (sensitivity 0.38, specificity 0.81). ** A score of 5 could not be achieved.

Predictors of prolonged treatment course in anxiety disorders

For anxiety disorders, the median treatment duration, given our selection, was 16.6 months (IQR 10.1-28.7); 33.0% had a prolonged treatment course. In univariable analysis, prolonged treatment course was significantly predicted by BSI-baseline (OR 1.94, CI95% 1.46-2.57), BSI 1st follow-up (OR 2.64, CI95% 1.98-3.51), older age, general health status, non-Dutch ethnicity and comorbid Axis I disorder. In multivariable logistic regression analyses model 1 (baseline variables but no BSI), the variables regarding

3 ROM =Routine Outcome Monitoring

general health status, older age, and non-Dutch ethnicity remained in the model after backward selection. After including BSI-baseline (OR of 1.78, CI95% 1.31-2.43) in model 2, general health was removed from the model, with no material change for the other variables. Adding the BSI 1st follow-up in model 3, BSI 1st follow-up yielded an OR of 2.52 (CI95% 1.86-3.23), whereas the baseline BSI assessment and ethnicity were removed from the model. Model 3 included older age and BSI 1st follow-up.

Model 3 had the highest internally validated AUROC of 0.68 and lowest BIC, pointing towards its better predictive performance than model 1 or 2 (respectively with an internal AUROC of 0.60 and 0.63). In medical diagnostic test evaluation, very high AUROCs (0.95 or higher) are desirable. In prediction studies AUROCs of 0.70 or higher would be considered strong effects (Hajian-Tilaki, 2013). Details of all three models are shown in the online supplement, *table 3B of appendix 2*.

Table 2B presents the risk prediction scores corresponding to the final model; higher scores indicate a higher risk of prolonged treatment course. From our final prediction model, we have derived an easy to calculate score which ranged from 0 to 3. For 35% of patients with a risk score of 2-3 points, the positive predictive value of prolonged treatment course was 52%, compared to 23% (1-negative predictive value) for patients with a risk score of 0-1 points. At this cut-off level sensitivity was 0.55 and specificity 0.75. See table 3B for details.

TABLE 2B. *Risk prediction score for prolonged treatment course (≥ two year) in patients with an anxiety disorder; assessed after approximately three months of treatment[3].*

Factor	Points
Age	
≤ 40 year	0
> 40 year	1
BSI 1st follow-up	
< 0.38	0
0.38-0.71	0
0.72-1.34	1
> 1.34	2
Total score (range)	**0-3**

Sensitivity analyses

In sensitivity analyses, baseline characteristics and performance of final model 2 (baseline characteristics, including baseline BSI) were similar for the excluded

patients (data not shown, depressive disorder N =1.183 , anxiety disorders N = 756). Furthermore, the addition of BSI-delta did not change the choice of the final model; BSI-delta did not remain in the model.

TABLE 3B. *Risk prediction score for prolonged treatment course (≥ two year) in patients with an anxiety disorder; assessed after approximately three months of treatment.**

Risk prediction score	No prolonged treatment course	Prolonged treatment course	Total
0-1	256	76	332
2-3	85	92	177

N = 509. * For 35% (N = 177) of patients with the highest scores (2-3), the positive predictive value for a prolonged treatment course was 52% (sensitivity 0.55, specificity 0.75)

DISCUSSION

In this cohort of outpatients with depressive or anxiety disorders, we showed that higher level of symptom severity at 2-6 months is the strongest predictor for prolonged treatment course. Our prediction model showed that patients in highest risk categories had a 60% positive predictive value of prolonged treatment course in patients with depressive disorders, and 52% for patients with anxiety disorders in the highest risk categories.

Our data confirm and contribute to earlier findings. First, we showed that a higher baseline level of symptom severity is independently associated with prolonged treatment course (van Beljouw et al. 2010; Katon et al. 2010; Stegenga et al. 2012), which is in line with many previous studies predicting the course of depressive disorders (Van et al. 2008; Vuorilehto et al. 2009; Lamers et al. 2011; Boschloo et al. 2014; Riihimaki et al. 2014; Gueorguiva et al. 2017) and anxiety disorders (Andreescu et al. 2008; Penninx et al. 2011; Prins et al. 2011; Green et al. 2015; Lamers et al. 2016). Secondly, for prolonged treatment course in patients with a depressive disorder comorbid personality disorders (Beaton & Rao, 2013, Gunderson et al. 2014) was found as independent predictors. In outpatients with anxiety disorders, we found also older age (Boschloo et al. 2014; Penninx et al. 2011) to be independent predictors for prolonged treatment course. Thirdly, in our study, 1st follow-up level of symptom severity had better predictive ability for prolonged treatment course compared to baseline level of symptom severity. Taking 1st follow-up level of symptom severity into account in prediction research was, to our best knowledge, first described by Fennell and Taesdale (Fennell & Taesdale, 1987). Few other studies have used a similar approach (Trivedi & Baker, 2001; Lutz et al. 2009; van Calker et al. 2009; Tadic et al. 2010; Richardson et

al. 2012). The observation that addition of the 1st follow-up assessment outperforms the baseline level of symptom severity, and thereby the rate of change deserves some further elaboration. This finding is consistent with the assumption that patients with high symptom severity at the 1st follow-up were likely to score high at the baseline BSI, or even higher. These might be the patients with poor initial recovery, which in turn might be predictive of poor outcome 2 year later. Patients with a low(er) score at the 1st follow-up could consist of a mix of patients improving from a higher score at baseline or remaining low which heralds' better outcome overall (Richardson et al. 2012). Consistent with these assumptions and our findings, the rate of change, and thereby the baseline assessment no longer added information at the moment of first follow-up. It is the level of symptom severity at the moment of assessment that determines risk of prolonged treatment course of the patient. Note that it is not our conclusion that the baseline assessment is without meaning; at baseline the level of symptom severity still predicts prolonged treatment course, albeit less discriminating. For early identification of patients with a less favourable course of treatment, the second BSI-measurement proves to be the single most informative variable.

Strengths of our study include that it was based on a large naturalistic treatment seeking population in psychiatric specialty care, with few exclusion criteria, contributing optimal external generalizability. Therefore, the external validity of our findings is likely higher than findings in samples from RCTs. Furthermore, structured data were collected with ROM as part of the normal clinical process, which was supervised by specially trained research nurses who were not involved in treatment. Based on our analyses, we could construct an easy to implement risk prediction score, which could be used in routine care.

Our results should be interpreted considering a number of limitations. First, we used the BSI for the assessment of the presence and severity of a broad range of symptoms, which may not be specific enough to fully capture clinical features or subtle changes in all depressive and anxiety disorders. The BSI has proven to be sensitive to treatment effect, but the sensitivity remains slightly behind compared to disorder-specific questionnaires (Lee et al. 2005). Second, information on specific treatment given was not available, thus the particular details of treatment for individual patients could not be taken into account in our analyses. In previous analysis in the population of GGZ Rivierduinen, we have found that treatment broadly followed guidelines, but this cannot be confirmed for individual patients (van der Lem et al. 2011; van Fenema et al. 2012). Third, data were limited to what was clinically available and we had no exhaustive set of predictors. We cannot exclude the possibility that other factors such as patient and/or

family history, childhood trauma, specific co-morbidity (e.g. pervasive developmental disorder), or personality characteristics (e.g. neuroticism) may add to the prediction (Young et al. 2006; Skodol et al. 2010; Lamers et al. 2011; Richards, 2011; Steinert et al. 2014), although in many studies the level of symptom severity outperforms these other variables as well (Spijker et al. 2002; van Beljouw et al. 2010; Stegenga et al. 2012; Boschloo et al. 2014; Riihimaki et al. 2014). The level of symptom severity could reflect to some extent the presence of a multitude of other risk factors, which could support the simplification of risk prediction to single symptom severity scores.

Despite potential limitations, the risk prediction model we developed in this study provides the clinician with an early estimate of prolonged treatment course; patients can easily be classified as having a low or a high predicted risk of poor outcome. This allows for personalized clinical risk profiling relatively early in the course of treatment. Higher risk prediction scores indicate a higher risk of a longer treatment duration, for 28% of the patients within the three highest risk categories, the positive predictive value of prolonged treatment course was 60%: corresponding to an AUROC of 0.65. Although the sensitivity of the score is not very strong the positive predictive value is sufficient to consider patients with a high-risk score for evaluation and monitoring of rational medication switches, add-on psychotherapy, application of social activation and early rehabilitation techniques or interventions to reduce adverse life circumstances. Although not studied here, patients that are initially misclassified, could perhaps be picked up later on, as it seems possible that the predictive value may continue to gain strength over time. Further studies are warranted to see how repeated monitoring could improve prediction. Evidence of effectiveness of such interventions is still needed, but our score at least allows for early risk stratification and for further research in the effectiveness of intensified treatment.

Acknowledgements
We gratefully acknowledge the essential contributions made by the participants of this study as well as the participating mental healthcare provider GGZ Rivierduinen. The authors declare no conflict of interest.

REFERENCES

1. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998; 51(11): 1055-1068.
2. Andreescu, C., Chang, C.C., Mulsant, B.H., Ganguli, M., 2008. Twelve-year depressive symptom trajectories and their predictors in a community sample of older adults. International psychogeriatrics. 20(2):221-36.
3. Baldwin DS, Schweizer E, Xu Y & Lyndon G. Does early improvement predict endpoint response in patients with generalized anxiety disorder (GAD) treated with pregabalin or venlafaxine XR? Eur Neuropsychopharmacol. 2012; 22: 137-42.
4. Beatson JA, Rao S. Depression and borderline personality disorder. Med J Aust. 2013; 199(6): 24-27.
5. Boschloo L, Schoevers RA, Beekman AT, Smit JH, van Hemert AM, Penninx BW. The four-year course of major depressive disorder: the role of staging and risk factor determination. Psychother Psychosom. 2014; 83(5): 279-288.
6. Bruce SE, Yonkers KA, Otto MW, Eisen JL, Weisberg RB, Pagano M, et al. Influence of Psychiatric Comorbidity on Recovery and Recurrence in Generalized Anxiety Disorder, Social Phobia, and Panic Disorder: A 12-Year Prospective Study. AJP. 2005; 162(6): 1179-1187.
7. Combs H, Markman J. Anxiety disorders in primary care. Med Clin North Am. 2014; 98(5): 1007-1023.
8. Crown WH, Finkelstein S, Berndt ER, Ling D, Poret AW, Rush AJ, et al. The impact of treatment-resistant depression on health care utilization and costs. J Clin Psychiatry. 2002; 63(11): 963-971.
9. de Beurs E, Barendregt M, Flens G, van Dijk E, Huijbrechts I, Meerding W. Vooruitgang in de behandeling meten - Een vergelijking van vragenlijsten voor zelfrapportage. Maandblad Geestelijke Volksgezondheid. 2012; 1-11.
10. de Beurs E, den Hollander-Gijsman ME, van Rood YR, van der Wee NJ, Giltay EJ, van Noorden MS, et al. Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. Clin Psychol Psychother. 2011; 18(1): 1-12.
11. Dennehy EB, Robinson RL, Stephenson JJ, Faries D, Grabner M, Palli SR, et al. Impact of non-remission of depression on costs and resource utilization: from the COmorbidities and symptoms of DEpression (CODE) study. Curr Med Res Opin. 2015; 31(6): 1165-1177.
12. Derogatis LR, Melisaratos N. The Brief Symptom Inventory: an introductory report. PsycholMed. 1983; 13(3): 595-605.
13. Donohue JM, Pincus HA. Reducing the societal burden of depression: a review of economic costs, quality of care and effects of treatment. Pharmacoeconomics. 2007; 25(1): 7-24.
14. Fennell MJV, Taesdale JD. Cognitive Therapy for Depression: Individual Differences and the Process of Change. Cog Ther Res. 1987; 11(2): 253-271.
15. Green SA, Honeybourne E, Chalkley SR, Poots AJ, Woodcock T, Price G, et al. A retrospective observational analysis to identify patient and treatment-related predictors of outcomes in a community mental health programme. BMJ open. 2015; 5(5).
16. Gueorguieva R, Chekroud AM, Krystal JH. Trajectories of relapse in randomised, placebo-controlled trials of

treatment discontinuation in major depressive disorder: an individual patient-level data meta-analysis. Lancet Psychiatry. 2017; 4(3): 230-237.

17. Gunderson JG1, Stout RL, Shea MT, Grilo CM, Markowitz JC, Morey LC, et al. Interactions of borderline personality disorder and mood disorders over 10 years. J Clin Psychiatry. 2014; 75(8): 829-834.
18. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, et al. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol. 2011; 21(10): 718–779.
19. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. Caspian J Intern Med. 2013; 4(2): 627-635.
20. Haller H, Cramer H, Lauche R, Gass F, Dobos GJ. The prevalence and burden of subthreshold generalized anxiety disorder: a systematic review. BMC psychiatry. 2014; 14: 128.
21. Hardeveld F, Spijker J, De Graaf R, Nolen WA, Beekman AT. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr Scand. 2010; 122(3): 184-191.
22. Katon W, Unutzer J, Russo J. Major depression: the importance of clinical characteristics and treatment response to prognosis. Depress Anxiety. 2010; 27(1): 19-26.
23. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, et al. Routine use of patient reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database Syst Rev. 2016; 7.
24. Kim JM, Kim SY, Stewart R, Yoo JA, Bae KY, Jung SW, et al. Improvement within 2 weeks and later treatment outcomes in patients with depressive disorders: the CRESCEND study. J Affect Disord. 2011; 129: 183-90.
25. Kudlow PA, Cha DS, McIntyre RS. Predicting treatment response in major depressive disorder: the impact of early symptomatic improvement. Can J Psychiatry. 2012; 57(12): 782-788.
26. Lamers F, Beekman AT, de Jonge P, Smit JH, Nolen WA, Penninx BW. One-year severity of depressive symptoms: results from the NESDA study. Psychiatry Res. 2011; 190(2-3): 226-231.
27. Lamers F, Beekman AT, van Hemert AM, Schoevers RA, Penninx BW. Six-year longitudinal course and outcomes of subtypes of depression. Br J Psychiatry. 2016; 208(1): 62-68.
28. Lee W, Jones L, Goodman R, Heyman I. Broad Outcome Measures May Underestimate Effectiveness: An Instrument Comparison Study. Child Adolesc Mental Health. 2005; 10(3): 143-4.
29. Licht-Strunk E, Van Marwijk HW, Hoekstra T, Twisk JW, De Haan M, Beekman AT. Outcome of depression in later life in primary care: longitudinal cohort study with three years' follow-up. BMJ. 2009; 338.
30. Lutz W, Stulz N, Kock K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. J Affec Disord. 2009; 118(1-3): 60-8.
31. Penninx BW, Nolen WA, Lamers F, Zitman FG, Smit JH, Spinhoven P, et al. Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). 2011. J Affect Disord. 2011; 133(1-2): 76-85.
32. Prins MA, Verhaak PF, Hilbink-Smolders M, Spreeuwenberg P, Laurant MG, van der Meer K, et al. Outcomes for depression and anxiety in primary care and details of treatment: a naturalistic longitudinal study. BMC psychiatry. 2011; 11: 180.
33. Rais S, Nazerian A, Ardal S, Chechulin Y, Bains N, Malikov K. High-cost users of Ontario's healthcare services.

Healthc Policy. 2013; 9(1): 44-51.

34. Richards D. Prevalence and clinical course of depression: a review. Clin Psychol Rev. 2011; 31(7): 1117-1125.
35. Richardson LP, McCauley E, McCarty CA, Grossman DC, Myaing M, Zhou C, et al. Predictors of persistence after a positive depression screen among adolescents. Pediatrics. 2012; 130(6): 1541-1548.
36. Riihimaki KA, Vuorilehto MS, Melartin TK, Isometsa ET. Five-year outcome of major depressive disorder in primary health care. Psychol Med. 2014; 44(7): 1369-1379.
37. Robinson RL, Grabner M, Palli SR, Faries D, Stephenson JJ. Covariates of depression and high utilizers of healthcare: Impact on resource use and costs. Journal of psychosomatic research. 2016; 85: 35-43.
38. Schulte-van Maaren YW, Carlier IV, Zitman FG, van Hemert AM, de Waal MW, van Noorden MS, et al. Reference values for generic instruments used in routine outcome monitoring: the Leiden Routine Outcome Monitoring Study. BMC Psychiatry. 2012; 21 (12): 203.
39. Skodol AE, Shea MT, Yen S, White CN, Gunderson JG. Personality disorders and mood disorders: perspectives on diagnosis and classification from studies of longitudinal course and familial associations. J Pers Disord. 2010; 24(1): 83-108.
40. Spijker J, de Graaf R, Bijl RV, Beekman AT, Ormel J, Nolen WA. Duration of major depressive episodes in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Br J Psychiatry. 2002; 181: 208-213.
41. Stegenga BT, Kamphuis MH, King M, Nazareth I, Geerlings MI. The natural course and outcome of major depressive disorder in primary care: the PREDICT-NL study. Soc Psychiatry Psychiatr Epidemiol. 2012; 47(1): 87-95.
42. Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community: A systematic literature review. J Affect Disord. 2014; 152-154: 65–75.
43. Steinert C, Hofmann M, Leichsenring F, Kruse J. What do we know today about the prospective long-term course of social anxiety disorder? A systematic literature review. J Anxiety Disord. 2013; 27(7): 692–702.
44. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med. 2004; 23(10): 1631-1660.
45. Tadic A, Helmreich I, Mergl R, Hautzinger M, Kohnen R, Henkel V et al. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. J Affect Disord. 2010; 120: 86-93.
46. Trivedi MH, Baker SM. Clinical significance of monitoring early symptom change to predict outcome. J Clin Psychiatry. 2001; 62(4): 37-40.
47. van Beljouw IM, Verhaak PF, Cuijpers P, van Marwijk HW, Penninx BW. The course of untreated anxiety and depression, and determinants of poor one-year outcome: a one-year cohort study. BMC Psychiatry. 2010; 10: 86.
48. van Calker D, Zobel I, Dykierek P, Deimel CM, Kech S, Lieb K, et al. Time course of response to antidepressants: predictive value of early improvement and effect of additional psychotherapy. J Affect Disorde. 2009; 114:243-53.
49. van der Lem R, van der Wee NJ, van Veen T, Zitman FG. The generalizability of antidepressant efficacy trials to routine psychiatric out-patient practice. Psychol Med. 2011; 41: 1353-1363.
50. van Fenema EM, van der Wee NJ, Giltay EJ, den Hollander-Gijsman ME, Zitman FG. Vitality predicts level of guideline-concordant care in routine treatment of mood, anxiety and somatoform disorders. J Eval Clin Pract.

2012; 18(2): 441-448.

51. Van HL, Schoevers RA, Dekker J. Predicting the outcome of antidepressants and psychotherapy for depression: a qualitative, systematic review. Harv Rev Psychiatry. 2008; 16(4): 225-234.
52. Van HL, Schoevers RA, Kool S, Hendriksen M, Peen J, Dekker J. Does early response predict outcome in psychotherapy and combined therapy for major depression? J Affect Disord. 2008; 105(1-3): 261-265.
53. van Noorden MS, van Fenema EM, van der Wee NJ, van Rood YR, Carlier IV, Zitman FG, et al. Predicting outcomes of mood, anxiety and somatoform disorders: the Leiden routine outcome monitoring study. J Affect Disord. 2012; 142(1-3): 122-31.
54. Von Korff M, Ormel J, Katon W, Lin EH. Disability and depression among high utilizers of health care. A longitudinal analysis. Arch Gen Psychiatry. 1992; 49(2): 91-100.
55. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012; 380: 2163-2196.
56. Vuorilehto MS, Melartin TK, Isometsä ET. Course and outcome of depressive disorders in primary care: a prospective 18-month study. Psychol Med. 2009; 39(10): 1697-1707.
57. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992; 30(6): 473-483.
58. Wittchen HU, Jacobi F, Rehm J, Gustavsson A, Svensson M, Jönsson B. et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur Neuropsychopharmacol. 2011; 21(9): 655-679.
59. Young JF, Mufson L, Davies M. Impact of comorbid anxiety in an effectiveness study of interpersonal psychotherapy for depressed adolescents. J Am Acad Child Adolesc Psychiatry. 2006; 45(8): 904-912.