Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression

Klein, Nicola S.; Holtman, Gea A.; Bockting, Claudi L. H.; Heymans, Martijn W.; Burger, Huibert

Published in:
Journal of Psychiatric Research

DOI:
10.1016/j.jpsychires.2018.06.006

Publication date:
2018

Journal of Psychiatric Research 104 (2018) 1–7

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Development and validation of a clinical prediction tool to estimate the individual risk of depressive relapse or recurrence in individuals with recurrent depression

Nicola S. Klein[a,b], Gea A. Holtman[c], Claudi L.H. Bockting[a,d,*], Martijn W. Heymans[e], Huibert Burger[c]

[a] Department of Clinical Psychology, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands
[b] Top Referent Traumacentrum, GGZ Drenthe, Altingerweg 1, 9411 PA Beilen, The Netherlands
[c] Department of General Practice, University of Groningen, University Medical Center Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands
[d] Department of Psychiatry, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
[e] Department of Epidemiology and Biostatistics, VU University Medical Center, De Boelelaan 1117, PO Box 7057, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Depression
Relapse
Recurrence
Prediction

ABSTRACT

Objectives: Many studies examined predictors of depressive relapse/recurrence but no simple tool based on well-established risk factors is available that estimates the risk within an individual. We developed and validated such a prediction tool in remitted recurrently depressed individuals.

Methods: The tool was developed using data (n = 235) from a pragmatic randomised controlled trial in remitted recurrently depressed participants and externally validated using data (n = 209) from a similar randomised controlled trial of remitted recurrently depressed participants using maintenance antidepressants. Cox regression was used with time to relapse/recurrence within 2 years as outcome and well-established risk factors as predictors. Performance measures and absolute risk scores were calculated, a practically applicable risk score was created, and the tool was externally validated.

Results: The 2-year cumulative proportion relapse/recurrence was 46.2% in the validation dataset. The tool included number of previous depressive episodes, residual depressive symptoms, severity of the last depressive episode, and treatment. The C-statistic and calibration slope were 0.56 and 0.81 respectively. The tool stratified participants into relapse/recurrence risk classes of 37%, 55%, and 72%. The C-statistic and calibration slope in the external validation were 0.59 and 0.56 respectively, and Kaplan Meier curves showed that the tool could differentiate between risk classes.

Conclusions: This is the first study that developed a simple prediction tool based on well-established risk factors of depressive relapse/recurrence, estimating the individual risk. Since the overall performance of the model was poor, more studies are needed to enhance the performance before recommending implementation into clinical practice.

1. Introduction

Major Depressive Disorder (MDD) is a prevalent disorder (Hardeveld et al., 2010) with a substantial disease burden (Whiteford et al., 2013). An important contributor to the burden of depression is its long-term clinical course. Once individuals experience a Major Depressive Episode (MDE), they are at elevated risk to develop subsequent MDEs. With every additional episode, the risk of relapse or recurrence (further referred to as recurrence) further increases and reaches up to 90% in individuals with three or more previous episodes (Solomon et al., 2000). Therefore, a personalised approach to preventing recurrence is warranted, taking into account the individual risk of recurrence and administering intensive prevention strategies for those especially at risk. Nevertheless, there are few evidence-based tools available for recurrence risk stratification in clinical practice.

Although risk factors for recurrence have been widely documented, the predictive value of these risk factors combined in a risk assessment tool has not been established. Such a tool could guide clinical decision making by generating personalised absolute risk predictions in individuals. In somatic medicine, prediction tools are used extensively

* Corresponding author. Department of Psychiatry, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.
E-mail address: c.l.bockting@amc.uva.nl (C.L.H. Bockting).

https://doi.org/10.1016/j.jpsychires.2018.06.006
Received 15 February 2018; Received in revised form 27 May 2018; Accepted 8 June 2018

(e.g., D'Agostino et al., 2008; Geersing et al., 2014; Ten Haaf et al., 2017) and only recently researchers started to develop them for mental health disorders (e.g., Fusar-Poli and Schultze-Lutter, 2016; King et al., 2013; Spijker et al., 2006; Tran et al., 2014; Wang et al., 2014). Regarding depressive recurrence, two studies developed prediction algorithms, including a wide range of risk factors (Van Loo et al., 2015; Wang et al., 2014). However, these studies seem to have limited clinical applicability. First, their risk algorithm is extensive and implementation in clinical practice is likely infeasible due to time constraints. Second, both studies included a sample from the general population rather than a clinical sample which may limit generalisability to mental health care settings. Thus far, only one study developed a simple tool, using specific symptoms of the Symptom Checklist-90 (SCL-90) to differentiate depressive relapse from nonrelapse after 6 months (Judd et al., 2016). However, the tool did not include well-established risk factors and was not externally validated.

The goal of this study was to develop and externally validate a simple and easily applicable clinical prediction tool based on well-established risk factors that predicts risk of recurrence in individuals with a history of recurrent depression.

2. Material and methods

2.1. Data and participants

Data were used from two pragmatic randomised controlled trials, further referred to as the development and validation data. The studies were performed in accordance with the latest version of the Declaration of Helsinki, approved by an independent medical ethics committee (METIGG), and registered at trialregister. nl (identifiers NTR2503 and NTR 1907). After explaining the procedure and before randomisation, participants provided a written informed consent to participate in the trials. In both studies, participants aged between 18 and 65 were included that 1) had experienced at least two MDEs with the last occurring in the past two years; 2) were remitted according to DSM-IV criteria assessed with the Structured Clinical Interview for DSM-IV Axis-I Disorders (SCID-I) (Spitzer et al., 1992) and a score on the Hamilton Rating Scale for Depression (HRSD) (Hamilton, 1960) less than or equal to ten. Exclusion criteria were mania/hypomania, a psychotic or bipolar disorder (past or present), alcohol/drug abuse, a primary diagnosis of an anxiety disorder, and organic brain damage. In the development data (Bockting et al., 2011a), 288 individuals were eligible of whom 264 were randomised between mid-September 2010 and August 2013 to Mobile Cognitive Therapy (M-CT), an internet-based Preventive Cognitive Therapy (PCT) with minimal therapist support added to Treatment As Usual (TAU) or TAU alone. In the validation data (Bockting et al., 2011b), participants additionally used antidepressants (ADs) for at least 6 months and were restricted to attend psychotherapy for a maximum of twice a month. Between mid-July 2009 and the end of April 2015, 289 participants were assessed for eligibility and randomised to PCT and AD, AD alone, or PCT with tapering of AD. Participants were recruited via media, general practitioners, pharmacists, and secondary mental health care institutions. More information is provided elsewhere (Bockting et al., 2011a, 2011b).

2.2. Outcome

The primary outcome was time to recurrence up to 2 years assessed with the SCID-I by trained interviewers after 3, 12, and 24 months in the development study and after 3, 9, 15, and 24 months in the validation study. In case of a recurrence, the exact date of onset was retrospectively determined and if necessary the life-chart of the SCID-I was used to determine the exact date of onset based on specific triggering events. In both studies, interviewers were blinded by treatment condition.

2.3. Predictors

Based on the literature (Burcusa and Iacono, 2007; Carr et al., 2013; Hardeveld et al., 2010; Monroe, 2010; Nanni et al., 2012; Nelson et al., 2017; Roca et al., 2011), the following variables measured at baseline were included as candidate predictors: 1) number of previous MDEs (life-chart of the SCID-I) categorised into less than three, three or four, and five or more; 2) number of residual depressive symptoms (Inventory of Depressive Symptomatology - Self Report (IDS-SR) (Rush et al., 1996)), entered into the model as a continuous variable; 3) severity of the last MDE (SCID-I), categorised into mild or moderate versus severe; 4) a chronic somatic illness (NEMESIS somatic illness list, De Graaf et al. (2002)); 5) Childhood adverse events (Dutch version of the Life Events Questionnaire (LEQ) (Kraaij and De Wilde, 2001)), where we examined whether participants had lost a parent or had experienced sexual or physical abuse before the age of 16; and 6) axis-I comorbidity (SCID-I).

We explicitly modeled treatment with M-CT/PCT during the study as predictor, based on recommendations of Groenwold et al. (2016) to include treatment in prognostic modelling using data from randomised controlled trials. The main findings of the development study showed no statistically significant effect of M-CT added to TAU compared to TAU alone (Klein et al., 2018). The main results of the validation study showed that AD was not superior to tapering AD with PCT and that adding PCT to AD was effective in preventing recurrence compared to AD alone (Bockting et al., 2018). In the current study, we could not take into account differences in treatment arms between the development (two arms) and validation (three arms, including a tapering arm) study and therefore added a variable indicating whether participants received PCT (online/face to face). Using this classification, a model only including treatment showed a Nagelkerke R^2 of 0.008 in the development data and 0.007 in the validation data.

2.4. Statistical analyses

At baseline, in the development data 12.3% and in the validation data 5.7% of the cases had missing data. Multiple imputations by chained equations was used to impute missing data (number of imputations: 40), assuming data were missing at random. Conforming to White et al. (2011), the imputation model included all predictors of the analyses, the event indicator and cumulative baseline hazard function, and variables predicting whether data were missing.

2.5. Model building and internal validation

The rule of thumb of at least ten events (recurrences) per parameter was followed to obtain sufficient statistical power and prevent overfitting (Peduzzi et al., 1995). We selected seven variables of which one comprised three categories, resulting in a total number of eight parameters in the validation data where 104 recurrences were observed. A Cox proportional hazards model was used to quantify the association between predictors and recurrence or censoring, whichever came first. The proportional hazards assumption was examined using log-minus-log plots. To select variables most strongly and independently associated with time to recurrence, a backward selection procedure was used with an alpha of 0.05. In the model selection process, the pooled p-value was derived from the imputed datasets after pooling the total covariance matrix or D1 method (Enders, 2010) and regression coefficients, Hazard Ratios (HRs), and confidence intervals of the final model were determined. The cumulative absolute risk of recurrence within 2 years was calculated for each participant using the baseline survival function at 2 years follow-up and the individual regression coefficients. In the internal validation process, the backward selection procedure was incorporated. To report the performance of the model, the median Harrell's C-statistic, which is a generalisation of the C-statistic to survival analysis, and Nagelkerke R^2 over the imputed datasets were used.

These figures indicate the discriminatory power and overall performance, respectively. After bootstrapping, the amount of overfitting and shrinkage was determined for all statistics and subtracted from the apparent performance statistic to correct for overfitting. The baseline survival was re-estimated after shrinking the coefficients. After shrinkage, the performance is more likely to reflect the performance when the model is applied to future studies (Harrell et al., 1996).

The results were used to create a score that clinicians can easily apply in clinical practice to evaluate the individual risk of recurrence. To this end, each coefficient was divided by the coefficient closest to zero and subsequently rounded to the nearest integer to obtain a number of points per unit of the predictor variable. For each participant, a total score was calculated as the sum of these numbers. The relationship between total score and risk of recurrence (1-survival probability) was presented graphically. Finally, the total score was subdivided into the categories 'high risk', 'medium risk', or 'low risk'. These categories were arbitrarily chosen based on recurrence rates in the literature, clinical sensibility, and statistical stability, i.e., that the sample size and recurrence rates in each category remained sufficient. In addition, we evaluated our tool as a binary prognostic test by dividing it using the category cutoffs into 1) high or medium risk combined versus low risk, and 2) high risk versus medium and low risk combined. For these cutoffs we calculated the sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). We could not use the observed numbers of recurrence since the true number of recurrences would be higher as a result of censoring. Therefore, we used the pooled predicted risks in each risk stratum and multiplied them with the total numbers in the risk strata to obtain estimated numbers of recurrence. These calculations were made for a hypothetical population of 1000 individuals with the same distribution across risk categories as in our development sample. Subsequently, the estimated numbers of recurrence were used to evaluate the tool as binary prognostic test. All probabilities in this study were pooled according to Marshall et al. (2009), using the complementary log-log transformation before pooling the results and back-transformation thereafter.

2.6. External validation

Based on the literature (Royston and Altman, 2013), we evaluated the external validation in several steps. Using the coefficients and cumulative baseline survival of the development data, the linear predictor was applied in the validation data and the median calibration slope and C-statistic (i.e., Harrell's C-statistic) of the imputed datasets was presented. A Calibration plot was constructed and Kaplan Meier curves according to risk groups were displayed and compared between both datasets, providing evidence of external validity in terms of discrimination and calibration. The calibration plot and Kaplan Meier curves were compared in all imputed datasets and one randomly chosen imputed dataset was used for illustration.

3. Results

In total, 264 participants were randomised in the development and 289 in the validation study. As 29 participants in the development and 80 in the validation study dropped-out immediately after randomisation, they were excluded from further analyses in the current study. In the development study, 104 out of 235 experienced a recurrence within 2 years compared to 116 out of 205 in the validation study. According to Kaplan-Meier estimates, the overall 2-year cumulative proportion of recurrence was 46.2% and 55.1% in the development and validation study respectively. Demographic and clinical characteristics are described in Table 1.

Table 1
Baseline demographic and clinical characteristics.

Characteristics	Development data (n = 235)	Validation data (n = 209)
Age, mean (SD)	46.8 (10.6)	48.3 (9.9)
Female gender, %	74.5 (175/235)	66.5 (139/209)
Country of birth (% The Netherlands)	90.6 (211/233)	96.2 (201/209)
Marital status, %		
Single	26.1 (61/234)	28.2 (59/209)
Married or cohabiting	65.4 (153/234)	61.7 (129/209)
Divorced or widowed	8.5 (20/234)	10.1 (21/209)
Education		
Primary or secondary education	14.0 (33/235)	20.6 (43/209)
Vocational education	24.7 (58/235)	26.3 (55/209)
Higher education	61.3 (144/235)	53.1 (111/209)
Employed, %	67.4 (157/233)	63.2 (132/209)
Treatment As Usual (TAU), %		
General practitioner	29.8 (70/235)	67.0 (140/209)
Mental health care	37.9 (89/235)	33.0 (69/209)
Treatment with antidepressants	55.0 (127/231)	100
Age of first MDE, mean (SD)	28.9 (12.2)	27.8 (11.9)
Months in remission, mean (SD)	8.5 (6.5)	8.0 (6.1)
Previous episodes MDD, %		
Two	22.1 (52/235)	16.3 (34/209)
Three or four	41.7 (98/235)	37.3 (78/209)
Five or more	36.2 (85/235)	46.4 (97/209)
Depressive symptoms (IDS-SR_{30}), mean (SD)	16.2 (9.6)	19.1 (11.2)
Severity last episode, %		
Mild or moderate	77.9 (183/235)	66.8 (139/208)
Severe	22.1 (52/235)	33.2 (69/208)
Baseline axis-I comorbidity		
Anxiety disorder	13.6 (32/235)	9.1 (19/209)
Dysthymia	3.0 (7/235)	1.9 (4/209)
Somatoform disorder	2.6 (6/235)	1.0 (2/209)
Other	1.3 (3/235)	1.0 (2/209)
More than one	2.1 (5/235)	4.2 (9/209)
Chronic somatic illness	34.4 (78/227)	21.1 (42/199)
Adverse events in childhood	28.4 (66/232)	33.3 (69/207)

3.1. Model building and internal validation

Table 2 displays the univariable and multivariable association with recurrence. Number of previous MDEs, residual depressive symptoms, and severity of the last MDE were retained in the multivariable model at the 0.05 alpha level. As described in the methods, treatment with PCT during the study was included in the multivariable model although it was not significant at the 0.05 alpha level. After correcting for overfitting, the Nagelkerke R^2 was 13% and the model had a C-statistic of 0.56, the latter indicating low discriminatory power. The calibration slope was 0.81 and the final model was adjusted for overfitting.

The final model was used to calculate a total risk score. For example, an individual that has experienced five or more MDEs, has an IDS-SR score of 30, had a severe last MDE, and is not treated with PCT has a total risk score of 69 (26 + 30 + 13 + 0). The corresponding 2-year recurrence risk can be read from Fig. 1, which presents the relationship between total risk score and 2-year risk of recurrence. Thus, the individual's risk is approximately 80%.

For each individual, a total risk score was calculated and categorised into low (< 35), moderate (35–50), and high (≥50) risk of depressive recurrence within 2 years. Table 3 shows that the tool is able to stratify individuals in predicted risk classes of 37, 55%, and 72%. The corresponding observed cumulative recurrence risks were similar in each class. Table 4 displays test characteristics of the tool when it would be used as binary prognostic test using the predicted probabilities and absolute numbers in the predicted risk categories. The sensitivity of the prediction tool was moderate in individuals scoring 35 or more and low in individuals scoring 50 or more, whereas the specificity in both risk categories was moderate to high. For example, when individuals score

Table 2
Univariable and multivariable association of predictors with depressive recurrence within 2 years.

	Univariable analyses		Multivariable analyses		
	Coefficient	HR (95% CI)	Coefficient[a]	HR (95% CI)	Risk score[b]
Childhood adverse events	0.48	1.61 (1.08–2.41)*			
Chronic somatic illness	0.17	1.19 (0.79–1.79)			
Axis-I comorbidity	0.14	1.15 (0.74–1.81)			
Number of depressive episodes					
2[c]					0
3 or 4	0.53	1.70 (0.93–3.11)	0.46	1.56 (0.86–2.90)	13
≥ 5	0.97	2.64 (1.46–4.80)**	0.91	2.48 (1.35–4.53)	26
IDS-SR score (per point)	0.04	1.04 (1.02–1.06)***	0.04	1.04 (1.01–2.47)	1
Severity last episode (severe)	0.43	1.53 (0.99–2.36)	0.46	1.58 (1.01–2.47)	13
Treatment PCT	−0.28	0.76 (0.52–1.12)	−0.16	0.85 (0.57–1.26)	−5

Baseline survival at 2 years follow-up = 0.55. *p < 0.05, **p < 0.01, ***p < 0.001.
[a] Corrected for overoptimism after bootstrapping. Shrinkage factor: 0.81.
[b] Risk score for depressive recurrence within 2 years follow-up. Each coefficient was divided by the coefficient closest to zero.
[c] Reference category.

35 or more, 52% who actually experienced a recurrence and 69% who did not experience a recurrence was correctly classified. The NPV and PPV were acceptable in both risk categories. For example, in individuals scoring 50 or more, 72% actually experienced a recurrence and in individuals that tested negative, 57% actually did not experience a recurrence.

3.2. External validation

The median C-statistic was 0.59, indicating a low discriminative power. The median calibration slope of 0.56 was considerably smaller than unity, suggesting overfitting. The calibration plot (Fig. 2) indicates poor calibration, with an underestimation of low risk individuals and an overestimation of high risk individuals. Fig. 3 shows Kaplan Meier curves in both datasets according to risk category. In both studies, high risk scores were associated with a lower cumulative probability of surviving (no recurrence) whereas moderate and low risk scores were associated with a higher cumulative probability of surviving.

Table 3
Risk of recurrence within 2 years according to score categories in the development dataset.

Score	*n* (%)	Recurrence	Censored[a]	Predicted probability (%)	Observed probability (1-Kaplan Meier estimate) (%)
< 35	140 (59.6)	47	13	0.37	0.36
35–50	71 (30.2)	40	3	0.55	0.59
≥50	24 (10.2)	17	0	0.72	0.71

[a] Individuals that were lost to follow-up without experiencing a recurrence.

4. Discussion

4.1. Principal findings

We developed a prediction tool based on well-established risk factors and transformed it into a practically applicable score to estimate the absolute risk of recurrence within 2 years for an individual. Besides

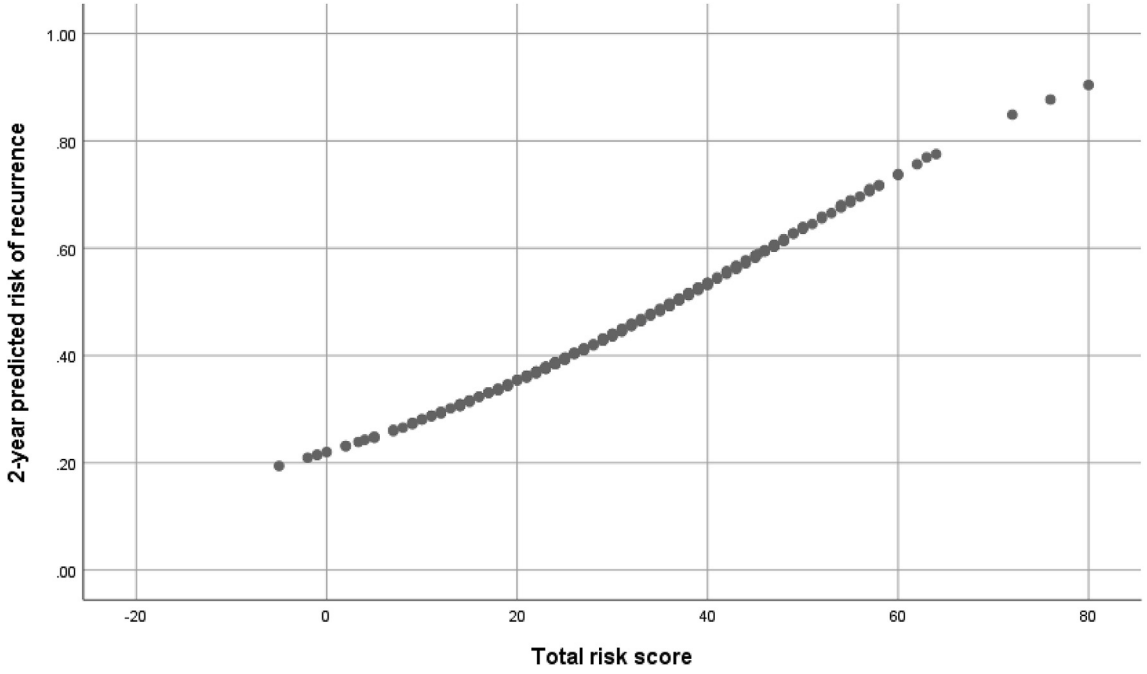

Fig. 1. Relationship between total risk score and predicted risk of recurrence within 2 years based on individuals in the development data.

Table 4
Binary prognostic test characteristics based on predicted probabilities.

Cut-off score	*n* (%)	Sensitivity	Specificity	Positive predictive value	Negative predictive value
≥35	404 (40.4)	52% (239/460)	69% (375/540)	59% (239/404)	63% (375/596)
≥50	102 (10.2)	16% (73/460)	95% (511/540)	72% (73/102)	57% (511/898)

This table is based on a hypothetical sample of 1000 individuals with the same distribution across risk categories as in the development data. The number of recurrences are based on the predicted probabilities from the development data. Applying the same distribution across risk categories in the development data, 596 individuals had a total risk score of 35 or lower, 302 had a total risk score between 35 and 50, and 102 had a total risk score of 50 or more. In this Table, ≥ 35 refers to the summation of the risk categories 35–50 and ≥ 50.

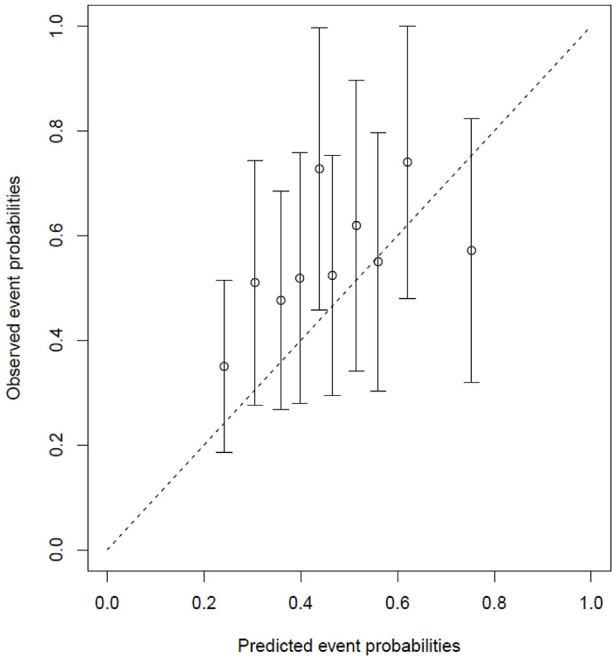

Fig. 2. Calibration plot of the prediction rule in the validation data.

treatment with PCT, our final tool included number of previous MDEs, severity of the last MDE, and residual depressive symptoms. Although the Kaplan Meier curves showed that the tool could differentiate between risk classes and the predictive values of the tool used as binary prognostic test varied from poor to excellent, the calibration and discriminative power of the model was poor and the figures based on the development data should be interpreted with caution due to overfitting.

4.2. Comparison with other studies

The risk factors in our final model are consistent risk factors of depressive recurrence according to literature reviews (Burcusa and Iacono, 2007; Hardeveld et al., 2010; Monroe, 2010). A possible explanation why axis-I comorbidity did not predict recurrence is the low rate of comorbidity in this sample. Our exclusion criteria may have influenced these rates and the presence of specific types of comorbidities associated with recurrence. Literature reviews do show mixed results on psychiatric comorbidity as risk factor of depressive recurrence (Burcusa and Iacono, 2007; Hardeveld et al., 2010). More recent studies are also mixed, with some demonstrating the importance of psychiatric comorbidity on recurrence (e.g., Hoertel et al., 2017; Kennedy et al., 2018) and suicide (e.g., Hoertel et al., 2015), and others demonstrating no association with recurrence (e.g., Hardeveld et al., 2013; Kuehner and Huffziger, 2013). Differences between studies might be caused by differences in study population, sample size, and operationalisation. Operationalisation may also be an explanation for not finding a unique effect of childhood adverse events in the multivariable model. For example, our tool did not include childhood neglect and recent studies suggest that of the childhood trauma categories, emotional neglect is an important predictor of depressive recurrence (e.g., Hovens et al., 2015; Paterniti et al., 2017). It is also suggested that clinical characteristics mediate the association between childhood trauma and the occurrence (Hovens et al., 2015) and course (Hovens et al., 2012) of depression, which corroborates with our finding that childhood adverse events lost statistical significance when added to the multivariable model. Although clinical guidelines presume that chronic somatic illness is a risk factor for depressive recurrence (American Psychiatric Association, 2010; National Institute for Health and Clinical Excellence, 2009, 2010), the systematic review of Kok et al. (2013) only identified four studies and found no association. A recent study found that pain and not specific types of chronic somatic illness predicted recurrence of MDD and that subthreshold depression mediated this association (Gerrits et al., 2014). Literature is mixed on the role of sociodemographic characteristics in recurrence. Based on literature reviews that concluded sociodemographic characteristics are not consistently

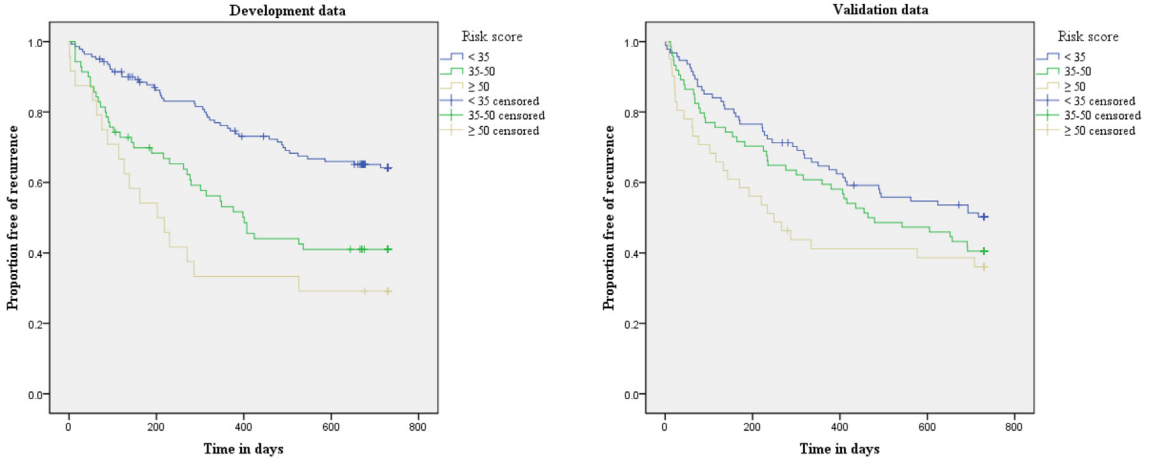

Fig. 3. Kaplan Meier curves for external validation. The Kaplan Meier curves were constructed in one imputed dataset and were highly comparable in the other imputed datasets.

associated with recurrence (Burcusa and Iacono, 2007; Hardeveld et al., 2010), we did not include them in our model. However, as some studies suggest they might play a role in recurrence (e.g., Hardeveld et al., 2013; Hoertel et al., 2017), we post-hoc examined whether specific sociodemographic characteristics independently added to our model, but this was not the case. More studies are needed to understand under which circumstances specific predictors are associated with recurrence. It must be noted that the model building strategy in the current study was focused on optimal predictive value of a set of variables and not on establishing unconfounded relationships and therefore the predictors should not be interpreted as causal factors for recurrence. Further, differences between datasets need to be addressed. The main findings of the development study showed no statistically significant effects of M-CT added to TAU compared to TAU alone (Klein et al., 2018). The main findings of the validation study showed that AD was not superior to tapering AD with PCT and that adding PCT to AD was effective in preventing relapse/recurrence (Bockting et al., 2018). In the current study, all three treatment conditions of the validation data were collapsed into M-CT/PCT or no M-CT/PCT, which explains the low Nagelkerke R^2.

The C-statistic of our model was lower compared to the other two studies that developed a multivariable algorithm for predicting recurrence (Van Loo et al., 2015; Wang et al., 2014). Wang et al. (2014) used longitudinal data of participants with current or lifetime MDD and found C-statistics of 0.75 and 0.72 in their development and validation data. Van Loo et al. (2015) used longitudinal data of female twins that had experienced an MDE in the last year and found C-statistics of 0.79 and 0.61 in their development and validation data. However, these studies were population-based, which limits generalisability to clinical populations. Furthermore, they did not report predicted risks according to individual characteristics summarised in a practical score, which is crucial for clinical use. Their inclusion of multiple variables is in line with the upcoming view that MDD is highly heterogeneous (Fried, 2017; Fried and Nesse, 2015). Moreover, it is in line with the large population-based study of (Hoertel et al., 2017) that built a comprehensive model predicting persistence and recurrence of MDD within 3 years in individuals with an MDE at baseline and found that combined effects of multiple factors determined the risk. However, using a tool with so many factors is not feasible for clinical practice. Thus far, only the study of Judd et al. (2016) developed a simple clinical prediction tool to differentiate between depressive relapse and non-relapse. No C-statistic was mentioned in this study, but overall comparable results were found regarding performance of specific risk categories. For example, a low sensitivity and high specificity was found in participants with a high (72.7%) risk of depressive relapse. However, their prediction tool was not externally validated. Our tool is easy applicable, has a profound empirical foundation in predictors selected, and was externally validated in an independent sample.

4.3. Clinical implications

Several factors need to be considered to determine whether a prediction tool can be used in clinical practice, including model performance and practical applicability, but also current risk stratification practice and treatment options (Moons et al., 2009; Royston and Altman, 2013).

The Kaplan Meier curves in the development and validation data showed that our tool was able to differentiate between risk classes and when using the tool as binary prognostic test the values varied from poor to excellent. However, the figures based on the development data must be interpreted with caution due to overfitting and the overall model performance, calibration, and discriminatory power was poor. Nevertheless, even a model with modest discrimination might be better than no model at all regarding clinical decision making (Moons et al., 2009; Royston and Altman, 2013). This notion may apply to the prediction of depressive recurrence as it is currently based on clinical judgment, which can be susceptible to bias (Croskerry et al., 2013). Albeit, validated questionnaires are available to detect the presence of acute MDD, no tools are available for predicting future risk of recurrence in remitted individuals. Therefore, in the absence of existing models, one might argue that a clinician could include our tool to facilitate clinical decision making by using it as a prognostic yardstick for the time being. However, we recommend more research to develop more accurate tools that can be recommended for implementation into clinical practice. Eventually, it is important to integrate a high performance tool into the clinical decision making process. By doing so, it may aid in determining the intensity of monitoring on potential depressive recurrence and in selecting relapse prevention strategies (e.g., self-help PCT (Biesheuvel-Leliefeld et al., 2017) or cognitive psychoeducation (Stangier et al., 2013) for the lower risk group, and face-to-face PCT, Mindfulness-Based Cognitive Therapy (MBCT), or the combination of PCT with maintenance AD for the medium to high risk group (Bockting et al., 2015, 2018). Altogether, more studies are needed to improve performance, validate risk categories, and test outcomes of implementing such a tool using impact studies.

4.4. Strengths and limitations

Strengths of the current study include the prospective longitudinal study design, the use of a structured clinical interview, and the external validation. Only a small number of prospective studies externally validate their tool and report clinical impact or usefulness, especially related to specific risk categories (Steyerberg et al., 2013). Several limitations have to be acknowledged. First, although we believe that the most consistent predictors were included, according to the literature there might be other easily assessable clinical factors contributing to the prediction, such as DSM personality disorders, history of suicide attempt, family history of MDD, and age at onset of MDD (i.e., Burcusa and Iacono, 2007; Hardeveld et al., 2010; Hoertel et al., 2017). Of these variables, only age at onset was assessed in the current study but not included because of its strong correlation with previous number of MDEs. We examined post-hoc whether this variable independently added to the model but this was not the case. Second, participants aged between 18 and 65 were included that had experienced at least two MDEs and were predominantly highly educated females, which might limit the generalisability. Third, in the literature no formal description of risk classes is available and therefore the risk classes in the current study were arbitrarily constructed, taking into account the literature, clinical sensibility, and statistical stability (i.e., sufficiently large numbers). Fourth, calibration in the validation data was poor. Although model development and external validation should focus on moderate rather than perfect calibration (Van Calster et al., 2016), our model should be further enhanced.

Acknowledgment

We are grateful to all participants in the two studies. In addition, we thank all recruitment sites for their efforts. We also thank the therapists for supporting M-CT and conducting the PCT. Finally, we are grateful to all PhD-students (dr. G.D. Kok, dr. G.D. Van Rijsbergen, H.J. Elgersma, MSc), master students, honour students, and volunteers for their help in the data-collection and coordination process.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jpsychires.2018.06.006.

References

American Psychiatric Association, 2010. Practice Guideline for the Treatment of Patients with Major Depressive Disorder, third ed. . http://psychiatryonline.org/guidelines.

Biesheuvel-Leliefeld, K.E.M., Dijkstra-Kersten, S.M.A., Van Schaik, D.J.F., Van Marwijk, H.W.J., Smit, F., Van der Horst, H.E., Bockting, C.L.H., 2017. Effectiveness of supported self-help in recurrent depression: a randomized controlled trial in primary care. Psychother. Psychosom. 86, 220–230.
Bockting, C.L.H., Hollon, S.D., Jarrett, R.B., Kuyken, W., Dobson, K., 2015. A lifetime approach to major depressive disorder: The contributions of psychological interventions in preventing relapse and recurrence. Clin. Psychol. Rev. 41, 16–26.
Bockting, C.L.H., Elgersma, H.J., Van Rijsbergen, G.D., De Jonge, P., Ormel, J., Buskens, E., Stant, A.D., De Jong, P.J., Peeters, F.P.M.L., Huibers, M.J.H., Arntz, A., Muris, P., Nolen, W.A., Schene, A.H., Hollon, S.D., 2011b. Disrupting the rhythm of depression: design and protocol of a randomized controlled trial on preventing relapse using brief cognitive therapy with or without antidepressants. BMC Psychiatr. 11, 1–22.
Bockting, C.L.H., Klein, N.S., Elgersma, H.J., Van Rijsbergen, G.D., Slofstra, C., Ormel, J., Buskens, E., Dekker, J., De Jong, P.J., Nolen, W.A., Schene, A.H., Hollon, S.D., Burger, H., 2018. Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicenter, randomised controlled trial. The Lancet Psychiatry 5, 401–410.
Bockting, C.L.H., Kok, G.D., Van der Kamp, L., Smit, F., Van Valen, E., Schoevers, R., Van Marwijk, H., Cuijpers, P., Riper, H., Dekker, J., Beck, A.T., 2011a. Disrupting the rhythm of depression using mobile cognitive therapy for recurrent depression: randomized controlled trial design and protocol. BMC Psychiatr. 11, 1–9.
Burcusa, S.L., Iacono, W.G., 2007. Risk for recurrence in depression. Clin. Psychol. Rev. 27, 959–985.
Carr, C.P., Martins, C.M., Stingel, A.M., Lemgruber, V.B., Juruena, M.F., 2013. The role of early life stress in adult psychiatric disorders: a systematic review according to childhood trauma subtypes. J. Nerv. Ment. Dis. 201, 1007–1020.
Croskerry, P., Singhal, G., Mamede, S., 2013. Cognitive debiasing 1: Origins of bias and theory of debiasing. BMJ Qual. Saf. 22, 58–65.
D'Agostino, R.B., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B., 2008. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation 117, 743–753.
De Graaf, R., Bijl, R.V., Ravelli, A., Smit, F., Vollebergh, W.A., 2002. Predictors of first incidence of DSM-III-R psychiatric disorders in the general population: findings from The Netherlands Mental Health Survey and Incidence Study. Acta Psychiatr. Scand. 106, 303–313.
Enders, C.K., 2010. Applied Missing Data Analysis. In: Little, T.D. (Ed.), Methodology in the Social Sciences. The Guilford Press, New York, NY.
Fried, E., 2017. Moving forward: how depression heterogeneity hinders progress in treatment and research. Expert Rev. Neurother. 17, 423–425.
Fried, E.I., Nesse, R.M., 2015. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J. Affect. Disord. 172, 96–102.
Fusar-Poli, P., Schultze-Lutter, F., 2016. Predicting the onset of psychosis in patients at clinical high risk: practical guide to probabilistic prognostic reasoning. Evid. Base Ment. Health 19, 10–15.
Geersing, G.J., Zuithoff, N.P., Kearon, C., Anderson, D.R., Ten Cate-Hoek, A.J., Elf, J.L., Bates, S.M., Hoes, A.W., Kraaijenhagen, R.A., Oudega, R., Schutgens, R.E., Stevens, S.M., Woller, S.C., Wells, P.S., Moons, K.G.M., 2014. Exclusion of deep vein thrombosis using the Wells rule in clinically important subgroups: individual patient data meta-analysis. BMJ 348, 1–13.
Gerrits, M.M.J.G., Van Oppen, P., Leone, S.S., Van Marwijk, H.W.J., Van der Horst, H.E., Penninx, B.W., 2014. Pain, not chronic disease, is associated with the recurrence of depressive and anxiety disorders. BMC Psychiatr. 14, 1–9.
Groenwold, R.H., Moons, K.G., Pajouheshnia, R., Altman, D.G., Collins, G.S., Debray, T.P., Reitsma, J.B., Riley, R.D., Peelen, L.M., 2016. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. J. Clin. Epidemiol. 78, 90–100.
Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56–62.
Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W.A., Beekman, A.T., 2010. Prevalence and predictors of recurrence of major depressive disorder in the adult population. Acta Psychiatr. Scand. 122, 184–191.
Hardeveld, F., Spijker, J., De Graaf, R., Nolen, W.A., Beekman, A.T., 2013. Recurrence of major depressive disorder and its predictors in the general population: results from The Netherlands Mental Health Survey and Incidence Study (NEMESIS). Psychol. Med. 43, 39–48.
Harrell, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med. 15, 361–387.
Hoertel, N., Blanco, C., Oquendo, M.A., Wall, M.M., Olfson, M., Falissard, B., Franco, S., Peyre, H., Lemogne, C., Limosin, F., 2017. A comprehensive model of predictors of persistence and recurrence in adults with major depression: results from a national 3-year prospective study. J. Psychiatr. Res. 95, 19–27.
Hoertel, N., Franco, S., Wall, M.M., Oquendo, M.A., Kerridge, B.T., Limosin, F., Blanco, C., 2015. Mental disorders and risk of suicide attempt: a national prospective study. Mol. Psychiatr. 20, 718–726.
Hovens, J.G., Giltay, E.J., Spinhoven, P., Van Hemert, A.M., Penninx, B.W., 2015. Impact of childhood life events and childhood trauma on the onset and recurrence of depressive and anxiety disorders. J. Clin. Psychiatr. 76, 931–938.
Hovens, J.G., Giltay, E.J., Wiersma, J.E., Spinhoven, P., Penninx, B.W., Zitman, F.G., 2012. Impact of childhood life events and trauma on the course of depressive and anxiety disorders. Acta Psychiatr. Scand. 126, 198–207.
Judd, L.L., Schettler, P.J., Rush, A.J., 2016. A brief clinical tool to estimate individual patients' risk of depressive relapse following remission: proof of concept. Am. J. Psychiatr. 173, 1140–1146.
Kennedy, J.C., Dunlop, B.W., Craighead, L.W., Nemeroff, C.B., Mayberg, H.S., Craighead, W.E., 2018. Follow-up of monotherapy remitters in the PReDICT study: maintenance treatment outcomes and clinical predictors of recurrence. J. Consult. Clin. Psychol. 86, 189–199.
King, M., Bottomley, C., Bellón-Saameño, J., Torres-Gonzalez, F., Svab, I., Rotar, D., Xavier, M., Nazareth, I., 2013. Predicting onset of major depression in general practice attendees in Europe: extending the application of the predictD risk algorithm from 12 to 24 months. Psychol. Med. 43, 1929–1939.
Klein, N.S., Kok, G.D., Burger, H., Van Valen, E., Riper, H., Cuijpers, P., Dekker, J., Smit, F., Van der Heiden, C., Bockting, C.L.H., 2018. No sustainable effects of an internet-based relapse prevention program over 24 months in recurrent depression: primary outcomes of a randomized controlled trial. Psychother. Psychosom. 87, 55–57.
Kok, G.D., Bockting, C.L.H., Burger, H., Hannig, W., Pijnenborg, G.H.M., Cuijpers, P., Hollon, S.D., 2013. Double trouble: does co-morbid chronic somatic illness increase risk for recurrence in depression? A systematic review. PLoS One 8, 1–7.
Kraaij, V., De Wilde, E.J., 2001. Negative life events and depressive symptoms in the elderly: a life span perspective. Aging Ment. Health 5, 84–91.
Kuehner, C., Huffziger, S., 2013. Factors predicting the long-term illness course in a cohort of depressed inpatients. Eur. Arch. Psychiatr. Clin. Neurosci. 263, 413–423.
Marshall, A., Altman, D.G., Holder, R.L., Royston, P., 2009. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med. Res. Meth. 9, 1–8.
Monroe, S.M., 2010. Recurrence in major depression: assessing risk indicators in the context of risk estimates. In: Richards, C.S., Perri, M.G. (Eds.), Relapse Prevention for Depression. American Psychological Association, Washington D.C, pp. 27–49.
Moons, K.G.M., Altman, D.G., Vergouwe, Y., Royston, P., 2009. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 338, 1487–1490.
Nanni, V., Uher, R., Danese, A., 2012. Childhood maltreatment predicts unfavorable course of illness and treatment outcome in depression: a meta-analysis. Am. J. Psychiatr. 169, 141–151.
National Institute for Health and Clinical Excellence, 2010a. Depression in Adults with a Chronic Physical Health Problem. Treatment and Management. National Collaborating Centre for Mental Health, London.
National Institute for Health and Clinical Excellence, 2009. Depression in adults: recognition and management. Clinical Guideline 90. https://www.nice.org.uk/guidance/cg90.
Nelson, J., Klumparendt, A., Doebler, P., Ehring, T., 2017. Childhood maltreatment and characteristics of adult depression: meta-analysis. Br. J. Psychiatry 210, 96–104.
Paterniti, S., Sterner, I., Caldwell, C., Bisserbe, J.-C., 2017. Childhood neglect predicts the course of major depression in a tertiary care sample: a follow-up study. BMC Psychiatr. 17, 1–13.
Peduzzi, P., Concato, J., Feinstein, A.R., Holford, T.R., 1995. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J. Clin. Epidemiol. 48, 1503–1510.
Roca, M., Armengol, S., García-García, M., Rodriguez-Bayón, A., Ballesta, I., Serrano, M.J., Comas, A., Gili, M., 2011. Clinical differences between first and recurrent episodes in depressive patients. Compr. Psychiatr. 52, 26–32.
Royston, P., Altman, D.G., 2013. External validation of a Cox prognostic model: principles and methods. BMC Med. Res. Meth. 13, 13–33.
Rush, A.J., Gullion, C.M., Basco, M.R., Jarrett, R.B., Trivedi, M.H., 1996. The inventory of depressive Symptomatology (IDS): psychometric properties. Psychol. Med. 26, 477–486.
Solomon, D.A., Keller, M.B., Leon, A.C., Mueller, T.I., Lavori, P.W., Shea, M.T., Coryell, W., Warshaw, M., Turvey, C., Maser, J.D., Endicott, J., 2000. Multiple recurrences of major depressive disorder. Am. J. Psychiatr. 157, 229–233.
Spijker, J., De Graaf, R., Ormel, J., Nolen, W.A., Grobbee, D.E., Burger, H., 2006. The persistence of depression score. Acta Psychiatr. Scand. 114, 411–416.
Spitzer, R.L., Williams, J.B., Gibbon, M., First, M.B., 1992. The structured clinical interview for DSM-III—R (SCID). I. History, rationale, and description. Arch. Gen. Psychiatr. 49, 624–629.
Stangier, U., Hilling, C., Heidenreich, T., Risch, A.K., Barocka, A., Schlösser, R., Kronfeld, K., Ruckes, C., Berger, H., Röschke, J., Weck, F., Volk, S., Hambrecht, M., Serfling, R., Erkwoh, R., Stirn, A., Sobanski, T., Hautzinger, M., 2013. Maintenance cognitive-behavioral therapy and manualized psychoeducation in the treatment of recurrent depression: a multicenter prospective randomized controlled trial. Am. J. Psychiatr. 170, 624–632.
Steyerberg, E.W., Moons, K.G.M., Van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H., Altman, D.G., PROGRESS Group, 2013. Prognosis research strategy (PROGRESS) 3: prognostic model research. PLoS Med. 10, 1–29.
Ten Haaf, K., Jeon, J., Tammemägi, M.C., Han, S.S., Kong, C.Y., Plevritis, S.K., Feuer, E.J., De Koning, H.J., Steyerberg, E.W., Meza, R., 2017. Risk prediction models for selection of lung cancer screening candidates: a retrospective validation study. PLoS Med. 14, 1–36.
Tran, T., Luo, W., Phung, D., Harvey, R., Berk, M., Kennedy, R.L., Venkatesh, S., 2014. Risk stratification using data from electronic medical records better predicts suicide risks than clinician assessments. BMC Psychiatr. 14, 1–20.
Van Calster, B., Nieboer, D., Vergouwe, Y., De Cock, B., Pencina, M.J., Steyerberg, E.W., 2016. A calibration hierarchy for risk models was defined: from utopia to empirical data. J. Clin. Epidemiol. 74, 167–176.
Van Loo, H.M., Aggen, S.H., Gardner, C.O., Kendler, K.S., 2015. Multiple risk factors predict recurrence of major depressive disorder in women. J. Affect. Disord. 180, 52–61.
Wang, J.L., Patten, S., Sareen, J., Bolton, J., Schmitz, N., Macqueen, G., 2014. Development and validation of a prediction algorithm for use by health professionals in prediction of recurrence of major depression. Depress. Anxiety 31, 451–457.
White, I.R., Royston, P., Wood, A.M., 2011. Multiple imputation using chained equations: issues and guidance for practice. Stat. Med. 30, 377–399.
Whiteford, H.A., Degenhardt, L., Rehm, J., Baxter, A.J., Ferrari, A.J., Erskine, H.E., Charlson, F.J., Norman, R.E., Flaxman, A.D., Johns, N., Burstein, R., Murray, C.J.L., Vos, T., 2013. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet 382, 1575–1586.