Functioning before and after a major depressive episode: pre-existing vulnerability or scar? A prospective three-wave population-based study

Author for correspondence:
E. H. Bos, E-mail: elske.bos@rug.nl

E. H. Bos[1], M. ten Have[2], S. van Dorsselaer[2], B. F. Jeronimus[1], R. de Graaf[2] and P. de Jonge[1]

[1]Department of Developmental Psychology, Behavioural and Social Sciences, University of Groningen, Groningen, The Netherlands and [2]Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands

Abstract

Background. The vulnerability hypothesis suggests that impairments after remission of depressive episodes reflect a pre-existing vulnerability, while the scar hypothesis proposes that depression leaves residual impairments that confer risk of subsequent episodes. We prospectively examined vulnerability and scar effects in mental and physical functioning in a representative Dutch population sample.

Methods. Three waves were used from the Netherlands Mental Health Survey and Incidence Study-2, a population-based study with a 6-years follow-up. Mental and physical functioning were assessed with the Medical Outcomes Study Short Form (SF-36). Major depressive disorder (MDD) was assessed with the Composite International Diagnostic Interview 3.0. Vulnerability effects were examined by comparing healthy controls ($n = 2826$) with individuals who developed a first-onset depressive episode during first follow-up but did not have a lifetime diagnosis of MDD at baseline ($n = 181$). Scarring effects were examined by comparing pre- and post-morbid functioning in individuals who developed a depressive episode after baseline that was remitted at the third wave ($n = 108$).

Results. Both mental ($B = -5.4$, S.E. $= 0.9$, $p < 0.001$) and physical functioning ($B = -8.2$, S.E. $= 1.1$, $p < 0.001$) at baseline were lower in individuals who developed a first depressive episode after baseline compared with healthy controls. This effect was most pronounced in people who developed a severe episode. No firm evidence of scarring in mental or physical functioning was found. In unadjusted analyses, physical functioning was still lowered post-morbidly ($B = -5.1$, S.E. $= 2.1$, $p = 0.014$), but this effect disappeared in adjusted analyses.

Conclusions. Functional impairments after remission of depression seem to reflect a pre-existing vulnerability rather than a scar.

Introduction

Major depressive disorder (MDD) is a prevalent disease (Kessler *et al.* 2005; De Graaf *et al.* 2012), characterized by a persistent depressed mood and/or marked loss of pleasure (American Psychiatric Association, 2013). MDD is a highly recurrent disorder and risk of recurrence is related to the number of previous episodes (Solomon *et al.* 2000; Kessing *et al.* 2004). This observation stimulated the idea that depressive episodes leave more or less permanent residual impairments or scars that confer risk of subsequent episodes; the 'scar hypothesis' (Rohde *et al.* 1990; Burcusa & Iacono, 2007). Indeed, many studies comparing remitted depressed patients with healthy controls have shown that the former differ in certain ways from the latter (Just *et al.* 2001; Bhagwagar & Cowen, 2008; Paykel, 2008). Such studies, however, cannot rule out the possibility that post-morbid impairments already existed before the onset of the episode, predisposing people to depression. Thus, individuals with a history of depression may already have had certain characteristics that made them vulnerable to recurrent depression beforehand. This is called the 'vulnerability hypothesis' (Just *et al.* 2001; Christensen & Kessing, 2006; Sowislo & Orth, 2013) or 'trait marker hypothesis' (Rohde *et al.* 1990).

To distinguish between pre-existing vulnerabilities and scars, prospective studies with pre- and post-morbid data of first-onset depressed subjects are required (i.e. subjects with a first-lifetime occurrence of an MDD episode). The few available prospective studies, most often done in population-based samples, have addressed scarring in psychosocial functioning (Burcusa & Iacono, 2007, review), cognitive functioning (Burcusa & Iacono, 2007; Allott *et al.* 2016, reviews), self-esteem (Sowislo & Orth, 2013, review), personality (Christensen & Kessing, 2006, review), and after stressful life events (Burcusa & Iacono, 2007, review).

CAMBRIDGE
UNIVERSITY PRESS

These reviews have shown that there is more evidence for the vulnerability hypothesis than for the scar hypothesis. There is slight evidence in favor of scarring in neuroticism (Kendler *et al.* 1993; Burcusa & Iacono, 2007) and some evidence that sensitization to stressful life events occurs (Monroe & Harkness, 2005; Burcusa & Iacono, 2007), but evidence for psychosocial or cognitive scars is virtually absent.

A limitation of the few longitudinal population studies done thus far is that in most studies the number of people who developed a depressive episode during the study follow-up was very small, leaving limited power to detect vulnerability or scarring effects. Further, it may be that scarring only occurs when depressive episodes are severe, long-lasting, or recurrent. Ormel *et al.* (2004*a*) indeed observed scarring effects in psychosocial disability only in individuals with severe recurrent episodes, although they could not replicate this regarding scars in neuroticism, depressive coping, and self-esteem (Ormel *et al.* 2004*b*), and did not control for possible confounders. Most longitudinal studies on scarring, however, did not separately investigate severe or recurrent episodes. Further, most studies focus on mental scars; there are hardly any studies that investigated physical functioning pre- and post-morbidly.

We investigated vulnerability and scar effects in mental and physical functioning among adults with a major depressive disorder in a large sample representative of the general Dutch population. We used three waves from the Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2), a prospective epidemiological study with a 6-years follow-up. We specifically focus on the questions: (1) whether levels of premorbid functioning of individuals who develop a first-onset depressive episode differ from those of healthy controls (vulnerability effect); (2) whether residual impairments in functioning remain after remission of the depressive episode (scar effect); (3) whether vulnerability or scar effects are different for mental compared with physical functioning; and (4) whether vulnerability or scar effects are different for recurrent *v.* first-onset episodes, and for episodes of different severity. We hypothesized that vulnerability but not scar effects will be found in general, but that scarring will be detected in the subgroup with recurrent or severe depression. Further, we examined whether the results may be confounded by comorbid disorders or by prodromal or residual symptom severity.

Methods

Participants

We used data from the first three waves of NEMESIS-2, a prospective epidemiological cohort study in the Dutch adult general population (de Graaf *et al.* 2010). The inclusion criterion for this study was age between 18 and 64 at the time of the baseline interview. Insufficient fluency in Dutch was an exclusion criterion. Participants were randomly selected by means of a multistage, stratified sampling procedure, with one respondent being randomly sampled from randomly selected households sampled from randomly selected municipalities. In the baseline wave (T0), between November 2007 and July 2009, 6646 individuals were assessed by means of face-to-face interviews (response rate 65.1%). This sample was reasonably representative of the Dutch general population, though younger individuals were somewhat underrepresented (de Graaf *et al.* 2010). During the second wave (T1), 3 years after T0, 5303 participants were interviewed again (response rate 80.4%). In the third wave (T2), 3 years after T1, 4618 participants were re-interviewed (response rate 87.8%). NEMESIS-2 was approved by an independent Medical Ethics Committee and all participants provided written informed consent. For more details about the study design, see de Graaf *et al.* (2010).

Measures

The Medical Outcomes Study Short Form Health Survey (SF-36, Ware & Sherbourne, 1992) was used to assess mental and physical functioning at each wave. The SF-36 is a self-report questionnaire consisting of 36 items and eight subscales which assess the level of functioning during the previous 4 weeks (McHorney *et al.* 1994). The scale is composed of 36 items, rated on Likert-type scales. Item responses are converted to a scale ranging from 0 to 100, with higher scores indicating better functioning. The scores are averaged to create the subscale scores. The SF-36 has been widely used and its psychometric properties are good (McHorney *et al.* 1993, 1994; Aaronson *et al.* 1998). We combined four of the SF-36 subscales into a mental health component scale, namely mental health, role limitations due to emotional problems, social functioning, and vitality (Cronbach's $\alpha = 0.78$ at T0). The other four SF-36 subscales were combined into a physical health component scale, namely general health perceptions, physical functioning, role limitations due to physical health problems, and bodily pain (Cronbach's $\alpha = 0.79$ at T0). Earlier principal component analyses on the SF-36 have provided support for the validity of these two component scales (McHorney *et al.* 1993; Sanson-Fisher & Perkins, 1998). Scores on these scales range from 0 (poor) to 100 (good).

DSM-IV diagnoses were established with the Composite International Diagnostic Interview (CIDI) version 3.0, a fully structured lay-administered diagnostic interview (Kessler & Üstün, 2004). At all three waves, the disorders assessed were mood disorders (major depressive disorder, dysthymia, bipolar disorder), anxiety disorders (generalized anxiety disorder, panic disorder with or without agoraphobia, agoraphobia without panic disorder, specific phobia, and social phobia), and substance use disorders (alcohol/drug abuse and dependence). The CIDI 3.0 assesses these disorders with generally good validity in comparison with blinded clinical reappraisal interviews (Haro *et al.* 2006). At all waves, 12-month diagnoses were assessed. Lifetime diagnoses were assessed at T0. At T1 and T2, an adapted CIDI version was used with as timeframe the period between T0–T1 and T1–T2, respectively.

Healthy controls ($n = 2826$) were defined as participants who did not have any lifetime diagnosis of a mood, anxiety, or substance use disorder at T0, and who did not develop any within the T0–T1 interval (Table 1). MDD ($n = 295$) individuals were defined as participants who did not have a 12-month diagnosis of MDD at T0 and who developed an MDD within the T0–T1 interval. The distinction between a first-onset ($n = 181$) and recurrent MDD ($n = 114$) was determined according to the absence or presence of a CIDI lifetime diagnosis of MDD at T0. Severity of the MDD was assessed at T1 in those who had a CIDI 12-month diagnosis of MDD and was based on criteria used in previous studies (Demyttenaere *et al.* 2004; Ten Have *et al.* 2013). Disorders were classified as severe if participants reported severe impairments in at least two areas of role functioning on the Sheehan Disability Scales (SDS, Leon *et al.* 1997); as moderate in case of moderate role impairment in any domain of the SDS; and the remaining were classified as mild. The SDS is a self-report measure of disability in four role domains (home, work, social, close relationships)

Table 1. Baseline characteristics of participants who developed a 12-month major depressive disorder between T0 and T1 (first-onset and recurrent MDD) and healthy controls (HC)

	MDD			Overall test statistic	
	First-onset	Recurrent	HC	F/χ^2	p
	$n = 181$	$n = 114$	$n = 2826$		
Age, mean (S.D.)	41.2 (12.0)	43.4 (11.8)	45.5 (12.4)	11.5	<0.001
Sex, % female (n)	55.3 (100)	76.3 (87)	54.4 (1538)	21.2	<0.001
Education, % high (n)	69.1 (125)	72.8 (83)	70.2 (1983)	0.5	0.79
Comorbid anxiety disorder at T0, % (n)	18.2 (33)	26.3 (30)	0	639.1	<0.001
Any somatic disorder at T0, % (n)	36.9 (66)	56.5 (61)	32.0 (892)	29.3	<0.001

F, F-value from analysis of variance test; χ^2, Chi-square value from Chi-square test.
Post-hoc tests show that the difference in sex and somatic comorbidity is only significant for the recurrent subgroup versus HC comparison.

and is incorporated in the CIDI 3.0. The scale has shown good reliability and validity (Leon *et al.* 1997). It consists of four questions assessing, on a scale from 0 to 10, the extent to which a particular disorder interfered with activities in one of the four role domains during the month in the past year when the disorder was most severe. Note that this measure of severity is based on the impact the disorder has on a person's functioning rather than on symptom severity.

Covariates

Covariates used in the analyses consisted of demographic and clinical variables, and were chosen based on their presumed association with both predictor and outcome. Demographic variables were assessed at baseline and included age, sex, and education. Education was defined as either low (primary, basic vocational, and low secondary education) or high (high secondary education, higher professional education, and university). Clinical variables included the presence of a comorbid anxiety disorder and the presence of a somatic disorder. The presence of a comorbid anxiety disorder (yes/no) was defined as any 12-month anxiety disorder according to the CIDI. This was assessed at all three waves. The presence of somatic disorders was also assessed at all three waves and 'any somatic disorder' was considered present if one or more of 17 chronic somatic disorders from a standard checklist was reported and had been treated or monitored by a medical doctor in the previous 12 months. These disorders included the most common chronic somatic disorders such as hypertension, chronic backache, rheumatoid disease, asthma, and diabetes. Comparisons between self-reports of chronic somatic disorders and medical records show moderate to good concordance (Baker *et al.* 2004; NCHS, 2004). Subthreshold residual symptoms were assessed at T2 with the K10, a self-report screening questionnaire with strong psychometric properties assessing psychological distress in the depression-anxiety spectrum in the past 4 weeks with 10 items on a five-point scale (Kessler *et al.* 2002; Donker *et al.* 2010). The K10 has proven to be both reliable and valid (Kessler *et al.* 2002; Donker *et al.* 2010).

Statistical analysis

Vulnerability effect

To examine the presence of a vulnerability effect, we compared the level of functioning at baseline (T0) between the healthy

controls ($n = 2826$) and participants who did not have a lifetime diagnosis of MDD at T0 and who developed a first-onset MDD between T0 and T1 ($n = 181$). Differences in demographic characteristics between the healthy and MDD group were examined using Chi-square tests for categorical variables and one-way ANOVA for continuous variables. First, mental functioning and physical functioning at T0 (thus prior to the development of the depressive episode) were compared between the first-onset MDD and the healthy group using linear regression analyses. The analyses were subsequently adjusted for age, sex, and education. The analyses on mental functioning were additionally adjusted for 12-month comorbid anxiety disorders at baseline, and the analyses on physical functioning were additionally adjusted for 12-month somatic disorders at baseline. We also examined whether the recurrent MDD subgroup ($n = 114$) differed from the HC group, and whether the first-onset and the recurrent MDD subgroups differed from each other. We further examined whether vulnerability effects were different for those who developed a mild, *v.* a moderate, *v.* a severe first-onset depressive episode. Because severity could only be assessed in those who had a 12-month diagnosis at T1 ($n = 118$, of the total of 181 in the first-onset MDD group), we restricted the analyses on severity subtype to these participants. Multiple tests were conducted, with a maximum of nine tests for each outcome measure. To ensure that the cumulative type I error remained below 0.05, an effect was considered significant if the p value was below 0.0056. A power analysis showed that with this adjusted alpha level we still had a power of 0.99 to detect a difference with a moderate effect size ($d = 0.50$) between the first-onset MDD group and the HC group.

Scar effect

To assess the presence of a scar effect, we compared the level of functioning prior to the onset of the depressive episode with the level of functioning following remission of the depressive episode. We did so in participants without a 12-month MDD at T0, who had a 12-month MDD at T1, and who had no 12-month MDD at T2 (i.e. were remitted at T2). This sample consisted of 108 participants. We used the 12-month diagnosis at T1 (and not the 3 years interval diagnosis at T1) to be able to assess the severity of the disorder. To account for the repeated measurements in the data, linear mixed models were used with mental and physical functioning as dependent variables and dummy variables for time T1 and time T2 as independent variables (time T0

being the reference category). The restricted maximum likelihood was used as the estimation method and the Akaike Information Criterion was used for model selection. First, we examined a model with the dummy variables for time T1 and time T2 only. For mental functioning this resulted in a model with a random intercept, a random slope for time T2, and an independent covariance structure for the random effects and the residuals. For physical functioning, the same model without random slopes was optimal.

Next, we adjusted the models for age, sex, and education, and additionally for comorbid anxiety disorders (mental functioning model) and somatic disorders (physical functioning model). The latter variables were time-varying. When a scar effect was present after remission, lower functioning scores would be found at time T2 compared with time T0. A state effect would be visible in worse functioning scores at time T1 compared with T0. Finally, we examined the presence of a scar effect in specific subgroups (first-onset *v.* recurrent episode; mild *v.* moderate *v.* severe episode), by adding the subgroup variable and the interaction between this variable and the time variables to the adjusted model (two separate models). In view of the smaller sample size in the scar analyses and the lower total number of tests, we adopted a somewhat less conservative *p* value adjustment here (Leon, 2014), using an α-level of 0.0167. A power analysis showed that with this adjusted alpha level we still had a power of 0.92 to detect a difference with a moderate effect size in functioning before *v.* after the depressive episode in the sample of 108 MDD patients.

Prodromal and residual symptoms

Vulnerability effects may be biased by prodromal symptoms, because premorbid assessments done just before the onset of the depressive episode may reflect prodromal symptomatology (Ormel *et al.* 2004*a*). Scar effects, as estimated in the pre-post comparison, may be biased by both prodromal and residual symptoms; if the post-morbid assessment takes place just after remission of the depressive episode, scarring effects may be over-estimated due to the presence of residual symptoms. If the premorbid assessment takes place just before the onset of the depressive episode, scarring effects may be obscured due to the presence of prodromal symptoms. To examine whether residual symptoms may have biased the results on scarring, we checked whether the type and severity subgroups differed with respect to their K10 scores at T2 using one-way ANOVA. Because the K10 was only assessed at T2, not at T0, we examined whether pro-dromal symptoms may have biased the results in an indirect way, by comparing MDD participants with a depression onset within 1 year after T0 *v.* those with a later onset in the T0–T1 period on mental and physical functioning.

Results

Baseline demographic and clinical characteristics of the 181 first-onset and 114 recurrent MDD participants are presented in Table 1 and compared with those of the 2826 healthy individuals. Individuals in both MDD subgroups were on average younger and more often suffered from a comorbid anxiety disorder at baseline than healthy individuals. Individuals in the recurrent subgroup were also more often female and more often suffered from a somatic disorder at baseline than healthy individuals.

Vulnerability effect

The regression models showed that mental functioning at T0, before the depressive episode developed, was significantly lower in the first-onset MDD group compared to the healthy group ($B = -7.8$, s.e. = 0.8, $p < 0.001$). Adjusting the analyses for age, sex, education, and anxiety comorbidity did not substantially change these results (see Table 2). Physical functioning was also significantly lower at baseline in the first-onset MDD group than in the healthy group in unadjusted ($B = -8.8$, s.e. = 1.2, $p < 0.001$) and adjusted analyses (Table 2). Thus, in participants who developed a first depressive episode after T0, both mental and physical functioning were already impaired prior to the onset of the depressive episode.

Table 2. Evaluating vulnerability effects; mental and physical functioning at T0 of participants who developed a 12-month major depressive disorder between T0 and T1 (first-onset and recurrent MDD) and healthy controls

	MDD			Adjusted regression Test statistic		
	First onset	Recurrent	Healthy controls			
	$n = 181$	$n = 114$	$n = 2826$			
	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	B	s.e.	p
Mental functioning	80.5 (14.5)	74.3 (18.9)	88.3 (9.9)			
				−5.4	0.9	<0.001
				−10.0	1.1	<0.001
				−4.6	1.3	<0.001
Physical functioning	78.6 (20.5)	72.6 (23.1)	87.4 (14.9)			
				−8.2	1.1	<0.001
				−10.9	1.4	<0.001
				−2.7	1.7	0.118

Regression analyses on mental functioning adjusted for age, sex, education, and any anxiety disorder; regression analyses on physical functioning adjusted for age, sex, education, and somatic disorders.

Mental functioning at T0 was significantly lower also in the recurrent MDD subgroup ($n = 114$) compared with the healthy group (Table 2). The same was true for physical functioning. Further, the recurrent subgroup showed lower mental functioning than the first-onset subgroup (Table 2). The difference in physical functioning between the recurrent and first-onset subgroups was not significant.

Next, we examined whether the severity of the depressive episode influenced the results. Of the 118 first-onset participants in which episode severity was assessed, 12 had a mild, 39 a moderate, and 67 a severe depressive episode. Mental functioning was significantly lower at T0 in the severe subgroup compared with the healthy group (Table 3). The mild and moderate subgroup did not differ significantly from the healthy group. The severe subgroup also had lower mental functioning at T0 than the moderate subgroup, while the difference between the severe and mild subgroup showed a trend ($p = 0.010$). The moderate subgroup did not differ significantly from the mild subgroup. Similar results were found for physical functioning, which was also significantly lower at T0 only in the severe subgroup compared with the healthy group (Table 3). The severe subgroup also had lower physical functioning than the moderate subgroup, while the difference between the severe and mild subgroup showed a trend ($p = 0.011$). The moderate subgroup did not differ significantly from the mild subgroup. Thus, those who developed the most severe depressive episode between T0 and T1 also had the worst level of mental and physical functioning prior to episode onset compared with those with less severe or no episodes.

To test whether prodromal symptoms may have biased the results, we examined whether first-onset participants whose depressive episode started within 1 year after T0 ($n = 87$) differed from participants with a later onset ($n = 94$) on baseline mental and physical functioning. This was not the case (adjusted analyses; mental functioning, $B = -0.3$, S.E. = 2.1, $p = 0.88$; physical functioning, $B = -2.3$, S.E. = 2.7, $p = 0.40$). Thus, the observed vulnerability effects in mental and physical functioning were not likely to be reflections of prodromal symptoms.

Scar effect

The characteristics of the 108 MDD participants who had developed a 12-month diagnosis of depression at T1 and were remitted at T2 are described in Table 4. Mean mental functioning in this group was lower at T1 than at T0, prior to the development of the depressive episode, and returned to premorbid levels following remission at T2 (see Table 5). The linear mixed models showed a state effect but no scar effect with regard to mental functioning, both in unadjusted (time T1, $B = -15.1$, S.E. 2.4, $p < 0.001$; time T2, $B = -2.3$, S.E. 1.8, $p = 0.18$) and adjusted analyses (Table 5). The random slope for time T2 was significant (S.D. = 17.3, 95% CI 13.7–21.7), which implies that there was large heterogeneity in the scarring effect. Physical functioning was higher at T0 than at T1 and T2. In the unadjusted model for physical functioning, both the state effect (time T1, $B = -5.7$, S.E. = 2.1, $p = 0.006$) and the scar effect (time T2, $B = -5.1$, S.E. = 2.1, $p = 0.014$) showed significance. In the adjusted analyses, neither the state nor the scar effect was significant (Table 5).

Table 3. Evaluating vulnerability effects in different severity groups; mental and physical functioning at T0 of participants who developed a first-onset 12-month major depressive disorder between T0 and T1 (mild, moderate, and severe MDD) and healthy controls

	First-onset MDD			Healthy controls	Adjusted regression Test statistic		
	Mild	Moderate	Severe				
	$n = 12$	$n = 39$	$n = 67$	$n = 2826$			
	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	Mean (S.D.)	B	S.E.	p
Mental functioning	88.1 (7.6)	85.1 (9.8)	76.4 (16.4)	88.3 (9.9)			
					−0.7	2.9	0.800
					−0.9	1.6	0.580
					−8.9	1.3	<0.001
					−0.2	3.3	0.957
					−8.1	3.2	0.010
					−8.0	2.0	<0.001
Physical functioning	87.5 (8.6)	84.1 (16.4)	72.7 (23.3)	87.4 (14.9)			
					−2.4	4.0	0.553
					−3.2	2.3	0.158
					−13.4	1.7	<0.001
					−0.82	4.6	0.858
					−11.0	4.3	0.011
					−10.2	2.8	<0.001

Regression analyses on mental functioning adjusted for age, sex, education, and any anxiety disorder; regression analyses on physical functioning adjusted for age, sex, education, and somatic disorders.
Severity refers to role functioning impairment due to the depressive episode, as assessed by the Sheehan Disability Scales.

Table 4. Characteristics of participants who developed a 12-month major depressive disorder between T0 and T1 and were remitted at T2, as used to examine a scar effect

	$n = 108$
Age, mean (s.d.)	42.3 (12.2)
Sex, % female (n)	65.7 (71)
Education, % high (n)	71.3 (77)
Comorbid anxiety disorder at T1, % (n)	29.6 (32)
Any somatic disorder at T1, % (n)	52.8 (57)

Subsequently, we examined whether any scar effects were different for specific subgroups, according to the type of disorder (first-onset, $n = 70$ v. recurrent, $n = 38$) and disorder severity (mild, $n = 10$ v. moderate, $n = 41$ v. severe, $n = 57$). In both the mental functioning and the physical functioning model, neither of the interaction terms were significant, indicating that scar effects were not more pronounced in individuals with recurrent of more severe episodes (results not shown).

Next, we examined whether scarring effects were different in participants with an onset within one year after T0 ($n = 35$) and participants with a later onset ($n = 73$). Also in these models, no significant interaction effects were observed. Finally, we examined whether residual symptoms levels (K10 scores) were different for those with a first-onset v. a recurrent episode, which was not the case [$F_{(1106)} = 0.37$, $p = 0.54$]. Neither were K10 scores different for subgroups with different disorder severity [$F_{(2105)} = 0.43$, $p = 0.52$]. So, the results were not likely influenced by prodromal or residual symptoms.
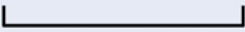
Discussion

The present three-wave population-based study was aimed to distinguish between pre-existing vulnerability and scar effects in mental and physical functioning after the experience of an episode of major depressive disorder. We did so by comparing individuals who developed a first-onset depressive episode during follow-up with healthy controls (vulnerability effect), and with themselves after remission of the depressive episode (scar effect). The results provide more support for the vulnerability than the scar hypothesis: both mental and physical functioning were already impaired prior to the onset of the depressive episode, and no firm evidence of scarring after remission was found. Subgroup analyses showed that the vulnerability effects mainly applied to persons with a severe depressive episode ('severe' in terms of role functioning impairment).

The results are congruent with those of previous studies done on this topic, the majority of which showed vulnerability but not scar effects on diverse outcomes (for reviews, see Christensen & Kessing, 2006; Burcusa & Iacono, 2007; Sowislo & Orth, 2013; Allott et al. 2016). Notably, studies that did find effects congruent with the scar hypothesis most often had a design in which individuals with remitted depression were compared with healthy controls, a design in which vulnerability and scar effects are confounded (Just et al. 2001). Possibly, much of the impairments observed in remitted patients can be accounted for by pre-existing characteristics that predispose these people to depression (vulnerability) or are due to common causes. Similarly, the often-made assertion that risk of recurrence increases with every further episode (Kessing et al. 2004), which has often been considered as support for the scar hypothesis, may actually be explained by pre-existing between-subjects differences: people with more previous episodes than other people may have a higher risk of recurrence just because they were vulnerable to recurrent episodes right from the start. Also in our study persons with recurrent MDD showed lower mental functioning before the onset of the episode than healthy persons, and their functioning was also lower than persons with first-onset depression. This may thus reflect a pre-existing differential vulnerability. Alternatively, the latter difference could be a reflection of scarring that occurred earlier in life. However, our subsequent analyses did not show evidence of scarring.

Although we did not find firm evidence for scarring in our study (only a scarring effect in physical functioning in unadjusted analyses), there was substantial heterogeneity in the degree to which mental functioning was still lowered after remission of the depressive episode. This heterogeneity could not be explained by type, severity, or timing of the depressive episode. There are not many studies who examined such subgroups. Ormel et al. (2004a) did so, and found some scarring in psychosocial functioning in individuals with severe recurrent episodes, in unadjusted analyses. In our unadjusted analyses, we found significant scarring in physical functioning, but this effect disappeared

Table 5. Evaluating scar effects; mental and physical functioning of participants who developed a 12-month major depressive disorder between T0 and T1 and were remitted at T2

	T0 $n = 108$ Mean (s.d.)	T1 $n = 108$ Mean (s.d.)	T2 $n = 108$ Mean (s.d.)	Adjusted regression Test statistic		
				B	s.e.	p
Mental functioning	79.6 (15.9)	63.8 (25.0)	77.3 (16.4)			
	⌐_____⌐			−14.0	2.4	<0.001
		⌐_____⌐		−2.1	1.7	0.19
Physical functioning	77.9 (20.2)	72.2 (23.6)	72.8 (23.6)			
	⌐_____⌐			−2.8	2.0	0.17
	⌐_____⌐			−2.3	2.0	0.26

Regression analyses on mental functioning adjusted for age, sex, education, and any anxiety disorder; regression analyses on physical functioning adjusted for age, sex, education, and any somatic disorder.

after adjustment, which suggests confounding by the presence of somatic disorders. An alternative explanation for this finding might be that the presence of MDD increases the likelihood of developing a somatic disorder. The high percentage of people with recurrent MDD having comorbid somatic disorders (Tables 1 and 4) is remarkable in this respect. If this alternative explanation would be true, we might have over-adjusted by using somatic disorders as a confounder, since this somatic disorder may be considered a scar. These reflections point to the more general problem that it is always hard to decide which variables should be used as potential confounders. In a recent study on scarring after remission of an anxiety disorder (Schopman *et al.* 2017), scarring was found in mental but not physical functioning in the subgroup with recurrent anxiety, also after adjustment. Power issues are also prevailing here, because subgroups were often very small, also in our study. Further, it might be that any scarring already occurred before T0 in those with previous episodes, limiting further scarring. Thus, although it may still be too early to discard the idea that scarring occurs altogether – maybe scarring does occur, but in a subtle way (Wichers *et al.* 2010), in subgroups we did not investigate, or in outcome measures we did not investigate (Burcusa & Iacono, 2007; Bhagwagar & Cowen, 2008) – our study suggests there is no robust scarring effect.

In contrast to the scarring effects, the vulnerability effects observed in the present study were robust and substantial. Effect sizes, calculated by dividing the adjusted estimated mean difference between the groups by the pooled standard deviation (Cohen, 1988, p.44), were 0.43 for mental functioning and 0.46 for physical functioning (first-onset MDD group), and especially large in the severe subgroup (0.66 for mental functioning; 0.69 for physical functioning). These effect sizes are slightly larger than those found for the average prospective association between neuroticism and depression, one of the most notorious risk factors for depression (Jeronimus *et al.* 2016). Functional limitations following remission from depression thus presumably reflect pre-existing vulnerabilities rather than scars. This may have clinical implications, as people with impaired functioning in either mental, social, or physical areas are at increased risk of (recurrence of) depression (Lewinsohn *et al.* 1998; Oldehinkel *et al.* 2001; Solomon *et al.* 2004; Conradi *et al.* 2008), and these impairments are not merely reducible to subthreshold symptoms (Spijker *et al.* 2007). Treatment and prevention strategies may be thus improved by targeting patients' level of functioning. Solely focusing on symptom reduction may be 'trying to dry out a flooded room without turning off the taps', if the underlying vulnerability is not addressed.

A vulnerability perspective of depression may also have implications for how we conceive this disorder. Depression is, even though operationalized as episodic, seldom an isolated event, but rather an expression of an underlying vulnerability. Many persons suffering from a current depressive episode have had other mental or somatic disorders in their past, which may be indicative of this vulnerability. Previous work on lifetime comorbidity of mental disorders has shown that disorders tend to cluster in large domains of psychopathology, such as internalizing and externalizing disorders (Kessler *et al.* 2011*a*, *b*). Depression has a median age of onset around 30 (Bromet *et al.* 2011), while some of the other disorders in the internalizing domain, such as specific phobia (age 8) or social phobia (age 14), have a considerably earlier onset (Lijster *et al.* 2017; Wardenaar *et al.* 2017). Perhaps such disorders with an early age of onset can be seen as markers of an underlying internalizing vulnerability that

eventually leads to disorders such as depression and generalized anxiety disorder. Viewing psychopathology from a lifespan perspective opens the way of preventing depression in such targeted subpopulations. It should also be noted that although disorders are typically defined as discrete, the underlying vulnerability may well be continuous (Ten Have *et al.* 2016).

Limitations of the present study include the self-report nature of our measures and the fact that they were assessed retrospectively, which might have induced some mood-related reporting bias. However, functioning was assessed prior to onset and 1 year after remission of the depression, presumably reducing this bias to acceptable levels. Further, we examined whether the timing of the onset (as a proxy for prodromal symptom severity) and residual symptom severity influenced the results, which was not the case. Another limitation may be that we used the Sheehan disability scale as a measure of depression severity, which may potentially overlap with our outcome measures. However, the SDS assesses disability *due to* the presence of the disorder, whereas the SF-36 assesses mental and physical functioning in general. A further limitation is that we did not assess prodromal symptom severity directly, but instead used the timing of the onset, because the K10 was not assessed at T0.

To conclude, this study suggests that limitations in mental and physical functioning following remission of depression reflect a pre-existing vulnerability rather than a scar. Depressive disorders might be conceived as life-stage-dependent manifestations of an underlying internalizing vulnerability dimension. A lifespan perspective of psychopathology opens the way of prevention strategies targeting this vulnerability early in life.

References

Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R et al. (1998) Translation, validation, and norming of the Dutch language version of the SF-36 health survey in community and chronic disease populations. *Journal of Clinical Epidemiology* **51**(11), 1055–1068.

Allott K, Fisher CA, Amminger GP, Goodall J and Hetrick S (2016) Characterizing neurocognitive impairment in young people with major depression: state, trait, or scar? *Brain and Behavior* **6**(10), e00527.

American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders*, 5th edn. Washington DC: American Psychiatric Association.

Baker M, Stabile M and Deri C (2004) What do self-reported, objective, measures of health measure? *Journal of Human Resources* **39**(4), 1067–1093.

Bhagwagar Z and Cowen PJ (2008) 'It's not over when it's over': persistent neurobiological abnormalities in recovered depressed patients. *Psychological Medicine* **38**(3), 307–313.

Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G *et al.* (2011) Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine* **9**(1), 90.

Burcusa SL and Iacono WG (2007) Risk for recurrence in depression. *Clinical Psychology Review* **27**(8), 959–985.

Christensen MV and Kessing LV (2006) Do personality traits predict first onset in depressive and bipolar disorder? *Nordic Journal of Psychiatry* **60**(2), 79–88.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Conradi HJ, de Jonge P and Ormel J (2008) Prediction of the three-year course of recurrent depression in primary care patients: different risk factors for different outcomes. *Journal of Affective Disorders* **105**(1–3), 267–271.

de Graaf R, ten Have M and van Dorsselaer S (2010) The Netherlands Mental Health Survey and Incidence Study-2 (NEMESIS-2): design and methods. *International Journal of Methods in Psychiatric Research* **19**(3), 125–141.

de Graaf R, ten Have M, van Gool C and van Dorsselaer S (2012) Prevalence of mental disorders and trends from 1996 to 2009. Results from the Netherlands Mental Health Survey and Incidence Study-2. *Social Psychiatry and Psychiatric Epidemiology* **47**(2), 203–213.

Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP *et al.* (2004) Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association* **291**(21), 2581–2590.

Donker T, Comijs H, Cuijpers P, Terluin B, Nolen W, Zitman F *et al.* (2010) The validity of the Dutch K10 and extended K10 screening scales for depressive and anxiety disorders. *Psychiatry Research* **176**(1), 45–50.

Haro JM, Arbabzadeh-Bouchez S, Brugha TS, de Girolamo G, Guyer ME, Jin R *et al.* (2006) Concordance of the composite international diagnostic interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *International Journal of Methods in Psychiatric Research* **15**(4), 167–180.

Jeronimus BF, Kotov R, Riese H and Ormel J (2016) Neuroticism's prospective association with mental disorders halves after adjustment for baseline symptoms and psychiatric history, but the adjusted association hardly decays with time: a meta-analysis on 59 longitudinal/prospective studies with 443 313 participants. *Psychological Medicine* **46**(14), 2883–2906.

Just N, Abramson LY and Alloy LB (2001) Remitted depression studies as tests of the cognitive vulnerability hypotheses of depression onset: a critique and conceptual analysis. *Clinical Psychological Review* **21**(1), 63–83.

Kendler KS, Neale MC, Kessler RC, Heath AC and Eaves LJ (1993) A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry* **50**(11), 853–862.

Kessing LV, Hansen MG, Andersen PK and Angst J (2004) The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders – a life-long perspective. *Acta Psychiatrica Scandinavica* **109**(5), 339–344.

Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SL *et al.* (2002) Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine* **32**(6), 959–976.

Kessler RC, Berglund PA, Demler O, Jin R and Walters EE (2005) Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication (NCS-R). *Archives of General Psychiatry* **62**(6), 593–602.

Kessler RC, Cox BJ, Green JG, Ormel J, Mclaughlin KA, Merikangas KR *et al.* (2011a) The effects of latent variables in the development of comorbidity among common mental disorders. *Depression and Anxiety* **28**(1), 29–39.

Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ *et al.* (2011b) Development of lifetime comorbidity in the World Health Organization World Mental Health Surveys. *Archives of General Psychiatry* **68**(1), 90–100.

Kessler RC and Üstün TB (2004) The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods Psychiatry Research* **13**(2), 93–121.

Leon AC (2004) Multiplicity-adjusted sample size requirements: a strategy to maintain statistical power with Bonferroni adjustments. *Journal of Clinical Psychiatry* **65**(11), 1511–1514.

Leon AC, Olfson M, Portera L, Farber L and Sheehan DV (1997) Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine* **27**(2), 93–105.

Lewinsohn PM, Rohde P and Seeley JR (1998) Major depressive disorder in older adolescents: prevalence, risk factors, and clinical implications. *Clinical Psychology Review* **18**(7), 765–794.

Lijster JM, Dierckx B, Utens EM, Verhulst FC, Zieldorff C, Dieleman GC *et al.* (2017) The age of onset of anxiety disorders. *Canadian Journal of Psychiatry* **62**(4), 237–246.

McHorney CA, Ware JE, Rachel Lu JF and Sherbourne CD (1994) The MOS 36-item Short-Form Health Survey (SF-36). III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Medical Care* **32**(1), 40–66.

McHorney CA, Ware JE and Raczek AE (1993) The MOS 36-item Short-Form Health Survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care* **31**, 247–263.

Monroe SM and Harkness KL (2005) Life stress, the "kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective. *Psychological Review* **112**(2), 417–445.

NCHS (2004) *Evaluation of National Health Interview Survey Diagnostic Reporting. Series 2: Data Evaluation and Methods Research.* (No. 120). Hyattsville, MD: Department of Health and Human Services: National Center for Health Statistics; Vital and Health Statistics.

Oldehinkel AJ, Bouhuys AL, Brilman EI and Ormel J (2001) Functional disability and neuroticism as predictors of late-life depression. *The American Journal of Geriatric Psychiatry* **9**(3), 241–248.

Ormel J, Oldehinkel AJ, Nolen WA and Vollebergh W (2004a) Psychosocial disability before, during, and after a major depressive episode – A 3-wave population-based study of state, scar, and trait effects. *Archives of General Psychiatry* **61**(4), 387–392.

Ormel J, Oldehinkel AJ and Vollebergh W (2004b) Vulnerability before, during, and after a major depressive episode: a 3-wave population-based study. *Archives of General Psychiatry* **61**(10), 990–996.

Paykel E (2008) Partial remission, residual symptoms, and relapse in depression. *Dialogues in Clinical Neuroscience* **10**(4), 431–437.

Rohde P, Lewinsohn PM and Seeley JR (1990) Are people changed by the experience of having an episode of depression? A further test of the scar hypothesis. *Journal of Abnormal Psychology* **99**(3), 264–271.

Sanson-Fisher RW and Perkins JJ (1998) Adaptation and validation of the SF-36 Health Survey for use in Australia. *Journal of Clinical Epidemiology* **51**(11), 961–967.

Schopman SME, ten Have M, van Dorsselaer S, de Graaf R and Batelaan NM (2017) Limited functioning after remission of an anxiety disorder: trait- and/or scar effect? Results of a longitudinal general population study. *Journal of Clinical Psychiatry.* pii: 16m11256. doi: 10.4088/JCP.16m11256. [Epub ahead of print].

Solomon DA, Keller MB, Leon AC, Mueller TI, Lavori PW, Shea MT *et al.* (2000) Multiple recurrences of major depressive disorder. *American Journal of Psychiatry* **157**(2), 229–233.

Solomon DA, Leon AC, Endicott J, Mueller TI, Coryell W, Shea MT *et al.* (2004) Psychosocial impairment and recurrence of major depression. *Comprehensive Psychiatry* **45**(6), 423–430.

Sowislo JF and Orth U (2013) Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychological Bulletin* **139**(1), 213–240.

Spijker J, de Graaf R, Oldehinkel AJ, Nolen WA and Ormel J (2007) Are the vulnerability effects of personality and psychosocial functioning on depression accounted for by subthreshold symptoms? *Depression and Anxiety* **24**(7), 472–478.

Ten Have M, Lamers F, Wardenaar K, Beekman A, de Jonge P, van Dorsselaer S *et al.* (2016) The identification of symptom-based subtypes of depression: a nationally representative cohort study. *Journal of Affective Disorders* **190**, 395–406.

Ten Have M, Nuyen J, Beekman A and de Graaf R (2013) Common mental disorder severity and its association with treatment contact and treatment intensity for mental health problems. *Psychological Medicine* **43**(10), 2203–2213.

Wardenaar K, Lim C, Al-Hamzawi A, Alonso J, Andrade L, Benjet C *et al.* (2017) The cross-national epidemiology of specific phobia in the World Mental Health Surveys. *Psychological Medicine* **47**(10), 1744–1760.

Ware JE and Sherbourne CD (1992) The MOS 36-item Short-Form Health Survey (SF-36). I. conceptual framework and item selection. *Medical Care* **30**(6), 473–483.

Wichers M, Geschwind N, van Os J and Peeters F (2010) Scars in depression: is a conceptual shift necessary to solve the puzzle? *Psychological Medicine* **40**(3), 359–365.