Framing depression in a SN(a)Pshot

Woudstra, S.

2013

document version
Publisher's PDF, also known as Version of record

Chapter 4

Inferior Frontal and Cingulate Volume as Markers of Poor Outcome in Depression

Brain morphometry changes in MDD over 2 years

Saskia Woudstra
Dick J. Veltman
Nic J.A. van der Wee
Mark A. van Buchem
Esther M. Opmeer
André Aleman
Witte J.G. Hoogendijk
Marie-José van Tol

Under review

Abstract

Early prediction of recurrence of depressive episodes in major depressive disorder (MDD) is key for prevention and decreasing disease burden. We investigated 1) the predictive value of regional brain volume abnormalities (anterior and posterior cingulate, inferior frontal gyrus, and hippocampus) for recurrence or recovery over a 2-year interval; and 2) whether structural changes progressed with chronic course. Effects of depression severity and use of antidepressant medication were additionally investigated.

Structural MRI-data from the Netherlands Study for Depression and Anxiety (NESDA), acquired with a 2-year interval were included for longitudinal whole-brain voxel-based morphometry analyses. Predictive and longitudinal analyses were performed in non-remitters, (n=30), remitters (n=30), and healthy controls (n=45).

At baseline and follow-up, non-remitted MDD patients, but not remitted patients, showed smaller right inferior frontal, rostral anterior cingulate, and posterior cingulate volumes than healthy controls (all:$P_{FWE\text{-}corrected} < .05$). No progressive volumetric decline was observed in any region in the MDD patients. No effect of depression severity, anxiety comorbidity, age of onset, number or duration of previous depressive episode, or use of antidepressants was found.

Our findings indicate that smaller inferior frontal, rostral anterior and posterior cingulate volumes predict a non-remitted MDD outcome. Moreover, absence of progressive decline indicates that smaller volumes are not the result of prolonged depression, suggesting a premorbid vulnerability factor for developing recurrent depression. Volume abnormalities could potentially serve as biological markers for poor clinical outcome, indicating a possible step in MRI-based complementary MDD diagnosis.

Key words Major depressive disorder – Structural imaging – Longitudinal data – Recurrence – Remission – Inferior frontal gyrus.

Introduction

Major depressive disorder (MDD) is a common multifactorial psychiatric disorder severely impacting daily functioning, and one of the leading causes of years lost due to disability (Marcus *et al.*, 2012). Approximately 50% of MDD patients experience recurrence of MDD within 5 years (Burcusa & Iacono, 2007). Early prediction of recurrent depression is key for selecting patients for relapse preventing therapy in order to reduce disease burden and socioeconomic costs associated with MDD. To date, the pathogenetic mechanisms underlying recurrence are unclear and useful biomarkers are lacking.

So far, a limited number of studies have focused on morphometric abnormalities to predict disease course (Baldwin *et al.*, 2000; Hickie *et al.*, 1997; Li *et al.*, 2010a; Phillips *et al.*, 2012). In MDD outpatients, preserved grey matter (GM) volume of the anterior cingulate cortex (ACC) was shown to predict good clinical outcome (Frodl *et al.*, 2008b). Smaller hippocampal volume has often been found in MDD (Ballmaier *et al.*, 2008; Frodl *et al.*, 2006; Hickie *et al.*, 2005; Janssen *et al.*, 2007; Janssen *et al.*, 2004; Lloyd *et al.*, 2004; Lorenzetti *et al.*, 2009; MacMaster *et al.*, 2008; MacQueen *et al.*, 2003; Mervaala *et al.*, 2000; Neumeister *et al.*, 2005; O'Brien *et al.*, 2004; Savitz & Drevets, 2009; Sheline *et al.*, 1999; Sheline *et al.*, 1996), albeit not consistently (Eker & Gonul, 2010; van Tol *et al.*, 2010), and may relate to disease load (Frodl *et al.*, 2008c). It has been suggested that smaller volume of limbic (amygdala (Frodl *et al.*, 2008c), insula (Hatton *et al.*, 2012), hippocampus (Frodl *et al.*, 2008c; MacQueen *et al.*, 2003)) and prefrontal (e.g. ACC (Frodl *et al.*, 2008c) and dorsal medial prefrontal cortex [PFC] (Frodl *et al.*, 2008c)) regions may result from prolonged duration of the disorder (i.e., 'scarring' (McEwen, 2001; Sheline, 2000)), due to chronically elevated levels of stress hormones (Post *et al.*, 2012) and possibly abnormal BDNF levels (Gonul *et al.*, 2010). Such scarring could explain the heightened vulnerability for new depressive episodes, as these morphological abnormalities may underlie functional impairments related to emotional processing and mood regulation. However, empirical support for the scar hypothesis is ambiguous (Santesso *et al.*, 2008; Zeiss & Lewinsohn, 1988), and comparisons between studies investigating the effects of disease burden of depression on regional brain volumes are hampered by low power (Cho *et al.*, 2010; Frodl *et al.*, 2008a; Frodl *et al.*, 2004; Frodl *et al.*, 2008c), the inclusion of selected patient subgroups (Chen *et al.*, 2010; Nifosí *et al.*, 2010), clinical heterogeneity (Lagopoulos *et al.*, 2012), and the use of different data-analytical strategies. Longitudinal studies published so far were likewise low-powered (Chen *et al.*, 2007; Frodl *et al.*, 2008a; Frodl *et al.*, 2004; Frodl *et al.*, 2008c; Hou *et al.*, 2012), or were conducted in elderly (Hickie *et al.*, 1997; Hou *et al.*, 2012; O'Brien *et al.*, 2004; Soriano-Mas *et al.*, 2011). However, most studies to date were cross-sectional, and were therefore unable to infer on effects of 'allostatic' disease-load, or to disentangle state and trait effects of the disease.

An alternative hypothesis to be considered is that structural abnormalities represent a premorbid vulnerability factor, associated with genetic or familiar risk for depression, and thus a predictor rather than a consequence of chronic course. A direct prediction from this hypothesis is that prolonged duration of the disorder is not accompanied by progressive volume loss, but that regional brain volume decreases should differentiate those with an unfavourable course from those with a favourable course early during the disease, or even before onset of the disease. Previous work has provided some support for this suggestion

(Chen *et al.*, 2007; Cho *et al.*, 2010; Frodl *et al.*, 2008c; Lagopoulos *et al.*, 2012), but these results may have been confounded by use of antidepressant medication. Moreover, in these studies, manual delineation techniques were applied, limiting the analysis to a small number of regions. Regions often associated with MDD include frontal regions (anterior cingulate and orbitofrontal cortex) and limbic regions (hippocampus, putamen and caudate nucleus) (Arnone *et al.*, 2012; Koolschijn *et al.*, 2009). In a previous whole-brain study in a large sample of MDD patients and healthy controls (HC), we found smaller GM volumes of the rostral ACC, inferior frontal gyrus (IFG), and posterior cingulate cortex (PCC) in MDD patients relative to controls (van Tol *et al.*, 2010). However, the cross-sectional design of this study precluded assessment of prognostic implications of these results.

Therefore the aim of the present study is two-fold: first, following the 'vulnerability' hypothesis, we tested whether volumes of regions associated with MDD (i.e. ACC, PCC, IFG, and hippocampus) at baseline are predictive of MDD recurrence or remission at two year follow-up. Second, following the scar hypothesis, we tested whether structural changes seen in MDD are progressive with prolonged MDD course. For both research questions we controlled for the effects of illness severity and the use of selective serotonin reuptake inhibitors (SSRI) as potential confounders.

Methods

Participants

This study is part of the Netherlands Study of Depression and Anxiety (NESDA), a multisite longitudinal cohort study aimed to provide insight into the long-term course of depression and anxiety disorders in patients selected from primary care and mental health organizations (Penninx *et al.*, 2008). Participants in this study underwent extensive clinical assessment and a subgroup was additionally included for Magnetic Resonance Imaging (MRI) measurements on average within two months following clinical assessment in one of three imaging centres (Leiden University Medical Center (LUMC); Academic Medical Center (AMC) Amsterdam, or University Medical Center Groningen (UMCG). Participants in the present study underwent a neuroimaging protocol at two time points; at baseline (S1, described by van Tol and colleagues (2010), and after two years (S2). At baseline, 301 participants were included in the NESDA-neuroimaging study and 199 participants were scanned again at S2 (66.11%). 'Drop-outs' did not differ from 'included participants' regarding depression severity at baseline ($F(1,300)=1.15$; $P=0.29$). Reasons for dropping out included 'too busy', pregnancy, and MRI-contraindications including metal in body. To assess depressive symptom characteristics and severity scores and severity at the day of scanning, the Inventory of Depressive Symptomatology (IDS) (Rush *et al.*, 1986) and the Montgomery–Åsberg Depression Rating Scale (MADRS) (Montgomery & Asberg, 1979) were used. Anxiety severity was assessed using the Beck Anxiety Inventory (BAI) (Beck *et al.*, 1988). To establish current half-year diagnosis, the Composite International Diagnostic Interview (CIDI) (Robins LN, 1988) was administered by a trained interviewer on average within two months from the scanning sessions at both S1 and S2. Percentage of time with depressive symptoms between S1 and S2 was assessed using the life chart interview (Lyketsos *et al.*, 1994; Penninx *et al.*, 2011). Number and duration

of previous episodes was determined by the number of depressive episodes prior to S1 and average months of these episodes prior to S1, respectively (self-report).

For the present report, we included participants with high-quality MR-scans at both S1 and S2. Patients had to meet criteria for current MDD (six-month recency) at S1. HC were without lifetime DSM-IV diagnosis at S1 and S2. Patients with only a diagnosis of anxiety disorder (n=83) at S1 or S2 were excluded from current analysis. Eleven subjects were excluded because of inferior data quality. Exclusion criteria were: presence of MRI-contraindications, dependence or recent abuse of alcohol and/or drugs, hypertension, major internal and/or neurological disorders and use of psychotropic medication (other than stable SSRI-use or incidental use of benzodiazepines). Additional exclusion criteria for patients were: presence of DSM-IV axis-I disorder other than MDD or comorbid anxiety disorders, and for HC presence of any DSM-IV disorder. This resulted in a final sample of 105 participants, which were divided into three groups with the following criteria: non-remitted MDD patients (non-remMDD; n=30) with a MADRS score >10 at both time points; remitted MDD patients (remMDD; n=30) with a MADRS score >10 at S1 and <10 at S2; and HC (n=45) with a MADRS score <10 and a 'normal' BAI score (based on cut-off score of <10 [Beck *et al.*, 1988]), measured at S1 and S2 (Table 4.1). The study was approved by the Ethics Committees at the VU University Medical Center, AMC, LUMC, and UMCG and all participants provided written informed consent.

Magnetic Resonance Image data acquisition

Images were acquired using 3T MR-systems, situated at different locations (AMC, LUMC. UMCG) equipped with SENSE-8 channel head coils. In Amsterdam a SENSE-6 channel head coil was used at S1. Images at both time points were acquired parallel to the anterior-posterior commissure plane using the same anatomical imaging sagittal 3-dimensional gradient-echo T1-weighted sequence (TR=9ms, TE=3.5ms, matrix 256x256, voxel size 1x1x1mm, 170 slices).

Statistics

Demographic and clinical data were analysed using SPSS 20.0 (SPSS for Windows, 2011, IBM Corporation, Armonk, USA). Longitudinal imaging data were analysed using diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) (Ashburner, 2007) following unified voxel-based morphometry (VBM), as part of the VBM8 toolbox of the Statistical Parametric Mapping software suite (SPM8) implemented in Matlab 7.5.0 (MathWorks, Natick, MA, USA).

All images were manually reoriented to the anterior commissure. Next, the two images of each participant were realigned, and subsequently registered to their mean image for time-unbiased registration (VBM8-Toolbox Manual,2010). After unified segmentation of each image into GM, WM, and cerebro-spinal fluid, we used the DARTEL-approach for registration, normalization into DARTEL-space, modulation, and further normalization of the images into Montreal Neurological Institute (MNI) space. Smoothing of GM images using an 8-mm full-width at half-maximum Gaussian kernel was performed to increase signal to noise ratio and to comply with our previous preprocessing procedure (van Tol *et al.*, 2010).

To optimize voxel residual smoothness estimation, achieve maximal sensitivity, and exclude false positives in non–grey matter tissue, voxelwise comparisons were masked using a comparison-specific explicit optimal threshold grey matter mask created using the Masking toolbox (Ridgway *et al.*, 2009).

Based on our previous findings and on previous reports on volumetric differences in MDD, we set the following *a priori* regions of interest (ROIs) using the Automated Anatomical Labelling (AAL) atlas (Maldjian *et al.*, 2003), implemented in the Wake Forest University (WFU) Pick-Atlas toolbox: bilateral anterior cingulate cortex (ACC), bilateral posterior cingulate cortex (PCC), bilateral inferior frontal gyrus (IFG; pars opercularis, pars triangularis, and pars orbitalis), and hippocampus. These labels were used for small volume correction after initial whole-brain exploration at an initial threshold of $P<.005$. To account for the number of *a priori* regions of interest, we corrected the critical corrected *p*-value for the number of regions (n=4: ACC, PCC, IFG, and hippocampus) and took the interdependency of the volumes of these regions into account to optimally balance between Type-I and Type-II error using the Simple Interactive Statistical Analysis Bonferroni tool (http://www.quantitativeskills.com/sisa /calculations/bonfer.htm). Using a Bonferroni-type correction that treats the variables as independent (proper Bonferroni: *alpha*/number of tests), would lead to a too stringent correction and thus increase the false negative rate, as the dependent variables (e.g. mean volume of the four ROIs) are not obtained in independent subgroups but within the same participant. Volumes in the four *a priori* regions of interest over all sessions and subjects showed a mean correlation of r=0.68, leading to a critical *p*-value of .032 to hold $P<.05$ FWE correction for the spatial extent of our regions of interest. For non-ROIs, a voxel level threshold of $P<.05$ whole-brain FWE-corrected was set.

For our main comparison, we performed a 2x3 factorial analysis with group as between-subject factor and time as within-subject factor. To test for effects of depression severity on regional brain volumes, a regression analysis was performed, using MADRS score as independent variable, with the patients groups pooled. In addition, to exclude effects of SSRI-use, analyses were repeated after omitting SSRI-users. Scan location, and head coil (by means of 3 dummy variables to model any differences between the 6- and 8-channel head coil) at S1 and S2 as well as total GM+WM volume at S1 and S2 were entered as covariates in all analysis.

Results

Sample descriptives

Table 4.1 lists sample characteristics. Groups were matched for gender, age, handedness, and scan location. Years of education at baseline differed significantly across the three groups ($F(2,102)=8.59$, $P<.001$) and *post-hoc* comparisons indicated that HC had more years of education than remMDD ($P=0.002$) and non-remMDD ($P=0.003$). For MADRS scores, a trend was observed between patient groups, showing trend-wise lower MADRS scores at S1 in remMDD relative to non-remMDD ($F(1,58)=3.28$; $P=0.075$). No significant differences between patient groups were however seen for IDS or BAI scores at S1. A main effect of group of depressive and anxiety scores was found at S2 (all: $F(2,102)>35$, $P<.001$): *post-hoc* analyses showed higher scores of both non-remMDD and remMDD at follow-up compared to controls

on all questionnaires (all: $F(1,73>35$, $P<.001$). Between patient group differences were also found at S2 on all questionnaires (MADRS and IDS: $F(1,58)>14$, $P<.001$; BAI: $F(1,58)=5.15$, $P=0.027$). Longitudinal within-group comparisons showed significant lower scores over time on MADRS and IDS (non-remMDD and remMDD). BAI score changed over time in remMDD only (MADRS, non-remMDD: $t(29)=2.07$, $P=0.048$; remMDD: $t(29)=10.3$, $P<.001$; HC: $t(44)=0.31$, $P=0.06$; IDS, non-remMDD: $t(29)=3.06$, $P=0.005$; remMDD: $t(29)=7.06$, $P<.001$; HC: $t(44)=0.28$, $P=0.78$; BAI, non-remMDD: $t(29)=1.28$, $P=0.21$; remMDD: $t(29)=5.65$, $P<.001$; HC: $t(44)=0.39$, $P=0.69$).

Time x group interaction analysis between patients showed a significant difference for MADRS score ($F(1,58)=44.3$, $P<.001$), and IDS score ($F(1,58)=5.8$, $P=0.019$) but not for BAI score ($F(1,58)=0.95$, $P=0.33$). Furthermore, SSRI-use between the two patient groups did not significantly differ (S1: $\chi^2(1)=0$; $P=1$; S2: $\chi^2(1)= 2.1$; $P=0.25$).

Non-remitted patients showed a higher percentage of time with depressive symptoms between S1 and S2 than remitted patients ($F(1,58)=6.59$, $P=0.013$). Patients who were remitters at S2 showed a trend-wise larger number of previous (i.e. prior to S1) depressive episodes ($F(1,57)=3.79$; $P=0.057$) but with similar duration ($F(1,55)=2.5$; $P=0.12$) and occurrence of comorbid anxiety than non-remitters ($^2(2)=4.06$; $P=0.13$). No difference was found between patient groups for age of onset ($F(1,58)=0.08$, $P=0.78$).

VBM results

Group effects

We found smaller volumes in non-remMDD relative to HC at S1 in the right IFG (MNI [57,9,7]; $Z=3.36$; P_{FWE-}corrected$=0.009$), right rostral ACC (MNI [3,44,16]; $Z=3.09$; P_{FWE-}corrected$=0.019$), and PCC (MNI[-3,-33,34]; $Z=2.69$; P_{FWE-}corrected$=0.024$) and these smaller volumes were not found in remMDD (Figure 4.1b and 4.2). At a subthreshold level, we found smaller volume in non-remMDD relative to HC at S1 in the left rostral ACC (MNI [-5,48,9]; $Z=2.82$; P_{FWE-}corrected$=0.04$). No hippocampal differences between remitted and non-remitted patients were observed at S1. Groups differed in total GM+WM volume at S1, but not between patient groups ($F(2,104)=4.4$; $P=0.015$; Table 4.1).

Smaller volumes in non-remMDD relative to HC were also observed at S2 (right IFG: MNI [57,9,7]; $Z=3.03$; P_{FWE-}corrected$=0.024$; right ACC: MNI [3,44,19]; $Z=2.94$; P_{FWE-}corrected$=0.029$; PCC: MNI [-3,-33,34]; $Z=2.86$; P_{FWE-}corrected$=0.015$; figure 4.1a and 4.2) and these smaller volumes were not found in remMDD (Figure 4.1b and 4.2). At a subthreshold level, we found smaller volume in non-remMDD relative to HC at S2 in

Table 4.1: Clinical and demographic characteristics

	Group; mean (SD)							
	Patients				Across groups		Patients only	
Characteristics	Patients, n=60	non-remMDD, n=30	remMDD, n=30	HC, n=45	F / c^2	*P*	F / c^2	*P*
Age at baseline (years)	38.4 (10.6)	39.70 (11.10)	37.17 (10.18)	40.40 (8.90)	.99	0.374	.85	.36
Education at baseline (years)	11.9 (3.1)	11.97 (3.49)	11.80 (2.82)	14.36 (2.83)	8.59	**<0.001**	.04	.84
Gender (female/male)	42/18	21/9	21/9	30/15	*0.13* (c^2)	0.936	0 (c^2)	1
Scan location (A/L/G)	19/28/13	9/14/7	10/14/6	23/15/7	*1.95* (c^2)	0.162	.13 (c^2)	.93
GM+WM volume at baseline (cc)	1469.74 (56.75)	1469.43 (60.89)	1470.05 (53.34)	1441.34 (33.89)	4.39	**0.015**	.002	.97
GM+WM volume at follow-up	1447.61 (62.68)	1448.51 (65.37)	1446.73 (60.98)	1422.53 (36.95)	2.84	0.063	.012	.91
SSRI use at baseline (no/yes)	42/18	21/9	21/9	45/0	N/A	N/A	0 (c^2)	1
SSRI use at follow-up (no/yes)	43/17	24/6	19/11	45/0	N/A	N/A	2.05 (c^2)	0.25
MADRS at baseline	20.2 (8.1)	22.07 (9.21)	18.33 (6.53)	0.91 (1.63)	127.98	**<0.001**	3.28	0.08
MADRS at follow-up	11.45 (8.4)	18.07 (6.72)	4.83 (2.59)	0.45 (1.04)	188.77	**<0.001**	101.30	**<0.001**
IDS at baseline	32.7 (9.9)	33.13 (10.57)	32.27 (9.43)	4.96 (3.41)	159.48	**<0.001**	.11	0.74
IDS at follow-up	22.8 (11.2)	27.73 (10.73)	17.90 (9.44)	4.82 (3.75)	76.27	**<0.001**	14.21	**<0.001**
BAI at baseline	16.8 (8.6)	16.23 (8.81)	17.33 (8.43)	1.98 (3.65)	58.89	**<0.001**	0.24	0.62
BAI at follow-up	11.8 (8.5)	14.27 (9.82)	9.43 (6.29)	1.80 (2.34)	35.93	**<0.001**	5.15	**0.027**
Comorbidity at S1	34	20	14	N/A	N/A	N/A	4.06	.13
# of previous depressive episodes	6.3 (9.4)	3.97 (5.65)	8.6 (11.56)	N/A	N/A	N/A	3.79	0.06
duration of previous depressive episodes(months)	21.4 (16.5)	24.76 (16.78)	17.89 (15.63)	N/A	N/A	N/A	2.55	0.12
% time with depressive symptoms between S1 and S2	0.40 (0.30)	0.49 (0.29)	0.31 (0.28)	N/A	N/A	N/A	6.59	**.013**
age of onset MDD (years)	24.4 (10.7)	24.83 (10.82)	24.03 (10.79)	N/A	N/A	N/A	0.08	0.78

Scan location A =Amsterdam, L=Leiden, G=Groningen; GM =Gray Matter volume; WM=White Matter volume; SSRI = selective serotonin reuptake inhibitors; MADRS = Montgomery–Åsberg Depression Rating Scale; IDS= Inventory of Depressive Symptomatology; BAI= Beck Anxiety Inventory; non-remMDD = non-remitted MDD patients; remMDD=remitted MDD patients; HC=healthy controls.

the left rostral ACC (MNI [-6,47,10]; Z=2.74; P_{FWE-}corrected=0.049). At S2, smaller hippocampal volume was only observed at a very low threshold ($P<.05$ uncorrected): compared to healthy controls, non-remitted patients, but not remitted patients, showed lower hippocampal volume at S2. To test for a possible effect of scan centre x time, we performed an ANOVA *post-hoc* with difference in volume between S1 and S2 as dependent factor and scan location at S1 as independent factor for each ROI and found no effect of centre x time on regional brain volume. No effects of course were observed in other regions. Groups did not differ in total GM+WM volume at S2 (Table 4.1).

Figure 4.1: Panel A: Overview of smaller IFG and ACC volumes in non-remitted patients relative to healthy controls. Red circles indicate smaller IFG volume in non-remitted MDD patients relative to healthy controls, blue circles indicate smaller ACC volume in non-remitted MDD patients relative to healthy controls (IFG: MNI [57,9,7]; Z=3.36; PFWE-corrected=0.009; right ACC: MNI [3,44,16]; Z=3.09; PFWE-corrected=0.019).
Panel B: Overview of contrast estimate plots and 90% confidence intervals of smaller IFG and ACC volumes in non-remitted patients relative to healthy controls. NR: non-remitted MDD patients; R: remitted MDD patients; HC: healthy controls.

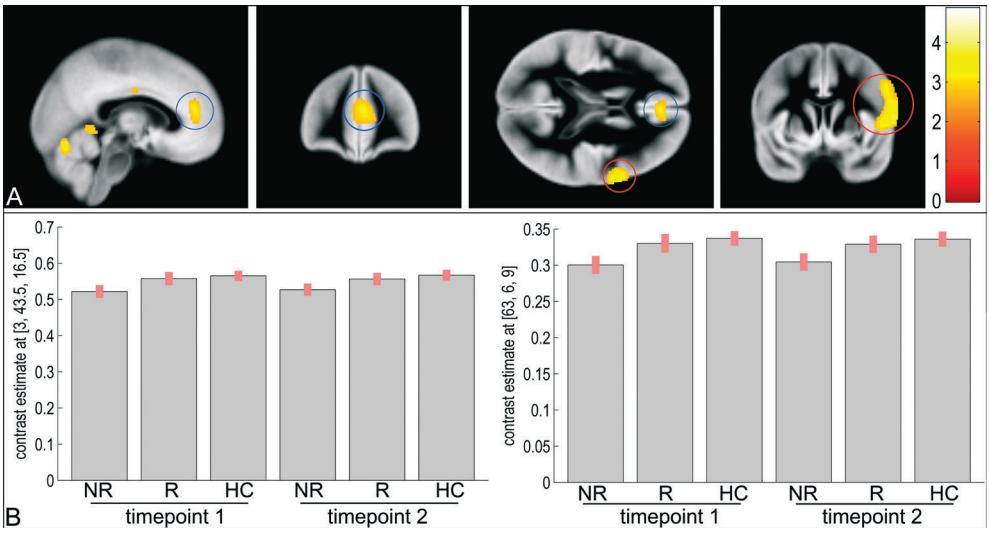
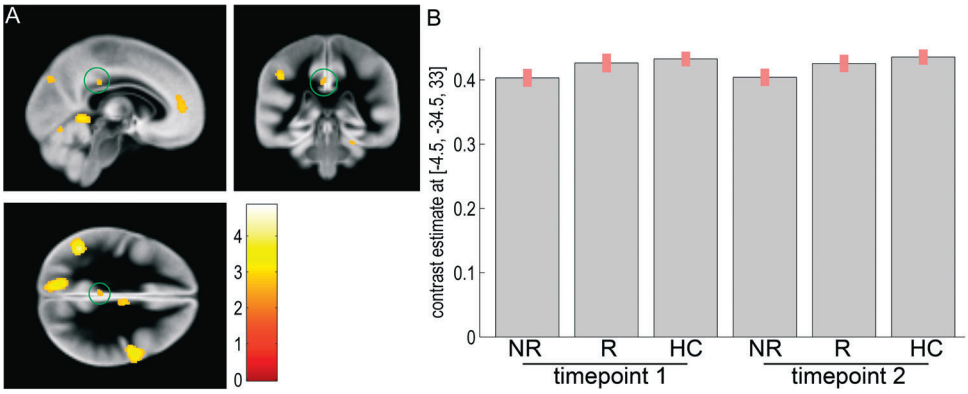


Figure 4.2: Panel A: Overview of smaller PCC volume in non-remitted patients relative to healthy controls. Green circles points at MNI [-3,-33,34].
Panel B: Overview of contrast estimate plots and 90% confidence intervals for smaller PCC volume in non-remitted patients relative to healthy controls. NR: non-remitted MDD patients; R: remitted MDD patients; HC: healthy controls.



Longitudinal effects

Effect of time

Within group comparison revealed smaller total GM+WM in non-remMDD ($t(29)=3.83$, $P<.001$), remMDD ($t(29)=7.12$, $P<.001$), and HC ($t(44)=7.37$, $P<.001$). Whole-brain analysis showed a main effect of time on GM volume in our ROIs (Table S4.1 in supplemental information).

Time x group interaction

Using a repeated Measures ANOVA for total GM+WM volume we found a time x group interaction ($F(1,102)=3.69$, $P=0.028$). *Post-hoc* analyses showed decreased volume in HC relative to non-remMDD and remMDD, but Bonferonni corrected results were below threshold (non-remMDD vs. HC: mean difference 27.03 (SE: 11.7), $P=0.071$; remMDD vs. HC: mean difference 26.45 [SE:11.7], $P=0.08$). Whole-brain analysis showed no time x group effect on GM volume in our ROIs. Whole-brain analysis did not reveal any time x group effects in other regions.

Effects of depression severity and SSRI-use

No effect of depressive state (MADRS-score) was observed within patient groups on IFG (S1: $r^2=-0.06$, $P=0.31$; S2: $r^2=-0.07$, $P=0.28$), ACC (S1: $r^2=0.07$, $P=0.30$; S2: $r^2=-0.13$, $P=0.16$), PCC (S1: $r^2=-0.06$, $P=0.32$; S2: $r^2=-0.09$, $P=0.26$) or hippocampus (S1: $r^2=0.08$, $P=0.28$; S2: $r^2=-0.04$, $P=0.38$). To exclude possible confounding effects of SSRI treatment, we repeated our analyses with 79 SSRI-free participants. Groups were similar to the overall groups with respect to demographic and clinical characteristics. Percentage of depressive symptoms between S1 and S2 between non-remitted and remitted patients was no longer significant

($F(1,32)=1.2$, $P=0.29$). The direction of the neuroimaging effects were similar, although observed sub-threshold in the ACC ($P_{FWE}>0.032$) (Table S4.2).

Discussion

In this study we examined (1) the predictive value of regional brain volume abnormalities for two-year naturalistic course of MDD and (2) whether structural abnormalities observed in MDD progress with prolonged disease duration. We studied baseline and 2-year follow-up structural MRI data from the NESDA-study using a DARTEL-VBM approach optimized for longitudinal imaging, while controlling for effects of illness severity and SSRI use. Compared to controls, smaller volumes of the right IFG, right rostral ACC, and PCC at baseline predicted a non-remitting course. At follow-up, these differences in right IFG and right rostral ACC were also observed and no progressive decline was observed, contradicting the scar hypothesis of chronic depression. These smaller volumes were absent in remitted patients at baseline and follow-up. Next, the smaller left PCC volume in non-remMDD relative to HC at baseline did not further progress at follow-up. No hippocampal abnormalities were observed at the set threshold. Although some meta-analyses found a relation with depression severity (Arnone *et al.*, 2012) or duration (Bora *et al.*, 2012) and ACC volume, other findings are in line with our results and found no effects of depression severity or duration of depressive episodes on volume (Koolschijn *et al.*, 2009). We found that patients who were remitters at follow-up went through trend-wise more previous depression episodes (but shorter) than chronic patients, controlling for this effect did not affect our results. Also, results were unaffected by comorbidity of anxiety disorders, age of MDD onset, and SSRI-use.

To our knowledge, this is the first whole-brain longitudinal structural MRI study in MDD to show predictive value of IFG, rostral ACC, and PCC for non-remitting MDD course in a large, longitudinal cohort study. Previously, we showed smaller GM volumes in the right IFG, rACC, and PCC to be a generic characteristic of MDD, with or without comorbid anxiety disorders (van Tol *et al.*, 2010). We now demonstrate that these smaller volumes are selectively observed in patients with an unfavourable course of the disease and thus relate to a non-remitting course of the disease. Moreover, we observed that unremitted course was not associated with volumetric decline over a two-year time period in these regions. Our findings contradict the scar hypothesis, and are also in line with previous longitudinal studies showing no progressive reduction (Frodl *et al.*, 2008a; Soriano-Mas *et al.*, 2011), although results have been inconclusive (Frodl *et al.*, 2008c). However, this conclusion needs to be interpreted with care, as scarring may be seen over a longer time period. or occur up to a certain level and then halt. Prospective studies are needed to fully corroborate this conclusion.

The (right) IFG has been implicated in inhibitory processes (Hampshire *et al.*, 2010), relevant for successful cognitive control and in cognitive processes related to negative affect (Lieberman *et al.*, 2004; Wang *et al.*, 2008). Of note, maladaptive recruitment of the IFG has been associated with failures in regulation of negative mood, a core deficiency in MDD. Smaller right IFG volume may thus represent a neuroanatomical basis for these abnormalities. Smaller ACC volume have been consistently associated with MDD (Arnone *et al.*, 2012; Bora *et al.*, 2012; Koolschijn *et al.*, 2009; van Tol *et al.*, 2010), and may underlie altered regulation of cognitive and emotional processing, with specific functions for

rostral, and anterior midcingulate subdivisions (Bush *et al.*, 2000). The rostral ACC has been implicated in regulation of negative mood (Bush *et al.*, 2000; Polli *et al.*, 2005).). Our results, therefore, support the hypothesis that in non-remMDD, smaller rostral ACC volume may reflect abnormalities in regulation of negative mood, thereby increasing the tendency to engage in negative mood states. Our results further emphasize the importance of the rostral ACC for the vulnerability and occurrence of MDD.

Additionally, we found smaller PCC volume, a region that is involved in rumination (Whitfield-Gabrieli & Ford, 2012) and sustaining self-processing during rest (Cavanna & Trimble, 2006; Liu *et al.*, 2012). The PCC is a core structure in the default mode network (DMN), a network involving medial and lateral frontal and parietal regions associated with abnormal intrinsic connectivity in depression (Greicius *et al.*, 2007; Raichle *et al.*, 2001; Sheline *et al.*, 2010; Whitfield-Gabrieli & Ford, 2012). Our results indicate that smaller PCC volume may underlie an increased propensity to engage in ruminative behavior and the vulnerability for chronic outcome of depression. Overall, our results suggest that smaller volumes of regions that together comprise a circuit important for self-referential and emotional processing is predictive of an unfavorable clinical course. Furthermore, they indicate that the vulnerability for non-remitting course is mediated by the functions sustained by these brain regions, including affective regulation, inhibitory control, and self-referential processing. As no progressive decline was observed in any of our ROIs, our findings may be reflective of neurodevelopmental abnormalities that predispose the development of depression, which is in line with findings from other mood disorders (Giuliani *et al.*, 2011; McCormick *et al.*, 2008).

The present study has a number of strengths. In this longitudinal neuroimaging study, we included a large number (n=60) of outpatients, and were able to follow them for two years and compared them to a large number of healthy controls (n=45). Having structural neuroimaging data available from two time-points allowed us to compare both between-group effects and within-group effects. Data was analysed using state of the art image processing methods: longitudinal data with DARTEL following VBM. To avoid bias of the images to the first image, the two images of each participant were registered to their mean image for time-unbiased registration. A mask was applied on the voxelwise comparisons to optimize voxel residual smoothness estimation, achieve maximal sensitivity, and exclude false positives in non–grey matter tissue.

In addition, we recruited a large and representative outpatients sample through the NESDA study, which made it possible to investigate the natural course effects of depression. A number of limitations should also be noted. Owing to the design of the present study, and to limit the amount of tests, we only included ROIs which previously were found to show volumetric differences between MDD and HC at baseline (described by van Tol *et al.* (2010). Although we included the hippocampus as ROI, our results did not provide clear evidence for detrimental effects of chronic course on hippocampal volume, as suggested by den Heijer (den Heijer *et al.*, 2011). Furthermore, since the epidemiological NESDA cohort was recruited through general practitioners and outpatient clinics, we may not have been fully able to capture the most severe MDD patients. Our non-remitted patients showed only moderate mean severity scores, and therefore our outpatient sample may not be representative of the severe subgroup of chronic depressed patients. Moreover, the difference between non-remitted and remitted patients regarding the percentage of depressive symptoms between

S1 and S2 was no longer visible after omitting the SSRI users. This is likely an effect of the non-remitted patient group: The SSRI-free non-remitted patient group showed less depressive symptoms between S1 and S2 than the total non-remitted patient group. The percentage of depressive symptoms between S1 and S2 in the remitted patient group was similar with the SSRI-free remitted patient group. This indicates that SSRI-free non-remitted patients are less severe depressed than the total non-remitted patient group. Decreased ACC volume was found subthreshold in the analysis with 79 SSRI-free participants, which likely reflects an effect of power. Omitting SSRI-users from our analyses did not affect our results in the IFG and PCC, suggesting that the predictive value of smaller IFG and PCC volume is not influenced by SSRI-use. This is supported by animal studies indicating that neurogenesis promoting effects of SSRIs may only be achieved in youth, but not in adulthood or old age (Couillard-Despres *et al.*, 2009; Yucel *et al.*, 2009). However, due to the naturalistic design of the study, effects of SSRI administration were not fully tested and it is therefore difficult to corroborate the SSRI effects in this study. Future studies should therefore specifically investigates treatment response.

In conclusion, our results suggest a predictive value of volume of the inferior frontal, anterior cingulate, and posterior cingulate cortex for non-remitting and unfavourable MDD outcome, independent of comorbidity, illness severity at baseline, SSRI use, age of MDD onset, and previous duration of the disorder. These areas have been related to self-referential and cognitive processing of negative mood. Additionally, disturbances of functions sustained by these regions, such as affective regulation and inhibitory control, may mark a vulnerability factor for prolonged depression (Hampshire *et al.*, 2010; Lieberman *et al.*, 2004; Wang *et al.*, 2008). Furthermore, smaller PCC together with rostral ACC volume may contribute to an altered default mode network activity and ruminative activity associated with depression, which may forecast an unfavourable course. Our results do not support the scar hypothesis, but instead indicate that smaller volumes may be present before or at the onset of the disorder. Future research should investigate whether smaller volumes may also predict a poor clinical outcome in high-risk groups, such as subthreshold MDD, and groups with a high familial load. Taken together, volumes of these areas could potentially serve as biological markers for poor clinical outcome, which may be a step in MRI-based complementary diagnosis of MDD. Future studies using a longitudinal design that specifically investigates treatment response should elucidate the contribution of smaller brain volume to treatment response in MDD to improve personalized treatment programs.

Supplemental material

Supplementary table S4.1: Longitudinal effects

Region of interest		Side	MNI-coordinate			Z	P_{FWE}-corrected
			X	Y	Z		
IFG	nonremMDD	R	35	20	34	4.10	0.001
		R	35	14	39	3.65	0.004
		R	36	17	25	3.36	0.011
	remMDD	R	35	20	34	5.17	<.001
		R	35	14	39	5.13	<.001
		R	36	12	22	3.74	0.003
		R	51	6	12	4.16	0.001
	HC	R	36	20	34	4.50	<.001
ACC	non-remMDD	R	9	48	3	3.30	0.013
	remMDD	L	-8	47	7	5.19	<.001
		L	-11	44	7	5.03	<.001
		R	9	48	6	4.91	<.001
		L	-6	47	16	4.46	<.001
	HC	R	9	45	9	3.11	0.023
		L	-5	48	12	3.02	0.029
PCC	non-remMDD	L	-12	-42	7	3.49	0.012
	remMDD	L	-18	-43	7	4.92	<.001
		L	-2	-40	7	4.76	<.001
		R	9	-46	16	4.07	0.002
	HC	L	-6	-42	6	3.59	0.009
Hippocampus	non-remMDD	L	-14	-40	-2	3.75	0.014
	remMDD	L	-21	-42	7	5.55	<.001
		R	24	-42	7	4.98	<.001
	HC	L	-8	-40	4	3.53	0.028

Within group effects of longitudinal volumetric changes: Longitudinal changes within groups per region of interest. MNI: Montreal Neurological Institute; Z: Z-value; FWE: family-wise error corrected for multiple comparison; IFG: inferior frontal gyrus; ACC: anterior cingulate cortex; PCC: posterior cingulate cortex; non-remMDD: non-remitted MDD patients; remMDD: remitted MDD patients; HC: healthy controls; L:left; R: right. To account for the number of a priori regions of interest, we corrected the critical corrected p-value for the number of regions (n=4) and took the interdependency of the volumes of these regions into account, resulting in a critical corrected p-value of P< .032.

Supplementary table S4.2a: Behavioural and clinical characteristics SSRI free patients

	Group; mean (SD)							
	Patients				Across groups		Patients only	
Characteristics	Patients, n=34	non-remMDD, n=17	remMDD, n=17	HC, n=45	F / c^2	*P*	F / c^2	*P*
Age at baseline (years)	38.2 (10.7)	40.3 (11.3)	36.2 (9.9)	40.4 (8.9)	1.3	0.29	1.3	.27
Education at baseline (years)	11.7 (3.4)	11.9 (4.2)	11.6 (2.7)	14.4 (2.8)	6.8	**0.002**	.06	.81
Gender (female/male)	23/11	13/4	10/7	30/15	1.2 (c^2)	0.55	1.2 (c^2)	.46
Scan location (A/L/G)	11/14/9	5/6/6	6/8/3	23/15/7	4.7 (c^2)	0.32	1.3 (c^2)	.50
GM+WM volume at baseline (cc)	1463.1 (49.4)	1458.2 (49.0)	1468.0 (50.8)	1441.34 (33.89)	2.9	0.06	.33	.57
GM+WM volume at follow-up (cc)	1438.8 (57.3)	1436.4 (59.7)	1441.2 (56.5)	1422.53 (36.95)		0.31	.06	.81
MADRS at baseline	19.3 (7.2)	21.3 (7.6)	17.4 (6.3)	0.91 (1.63)	104.7	**<.001**	2.7	.11
MADRS at follow-up	11.0 (7.5)	17.0 (5.7)	5.0 (2.4)	0.45 (1.04)	193.8	**<.001**	63.6	**<.001**
IDS at baseline	31.7 (10.2)	30.4 (10.5)	33.1 (10.1)	4.96 (3.41)	136.4	**<.001**	.59	.45
IDS at follow-up	23.3 (11.9)	28.9 (11.2)	17.7 (10.0)	4.82 (3.75)	69.1	**<.001**	9.5	**.004**
BAI at baseline	15.2 (8.2)	13.3 (5.8)	17.2 (9.9)	1.98 (3.65)	49.8	**<.001**	1.9	.17
BAI at follow-up	11.2 (6.7)	13.0 (6.4)	9.4 (6.7)	1.80 (2.34)	42.4	**<.001**	**2.5**	.12
Comorbidity at S1	18	9	9	N/A	N/A	N/A	1.0	.61
# of previous depressive episodes	4.7 (6.9)	4.2 (6.0)	5.2 (7.9)	N/A	N/A	N/A	.18	.67
duration of previous depressive episodes(months)	19.3 (15.6)	19.8 (16.2)	18.6 (15.6)	N/A	N/A	N/A	.05	.83
% time with depressive symptoms between S1 and S2	.37 (.28)	.42 (.29)	.32 (.26)	N/A	N/A	N/A	1.2	.29
age of onset MDD (years)	24.7 (10.7)	23.6 (9.9)	25.8 (11.6)	N/A	N/A	N/A	.36	.55

SD: standard deviation; MDD: major depressive disorder; HC: healthy controls; Rem: remitted patients; NR: non-remitted patients; n: number; y:years; m:months; scan site A/L/G: A: Amsterdam; L: Leiden; G: Groningen; MADRS score= total scale score all items ; IDS score = total scale score; all items; BAI score = total scale score all items.

Supplementary table S4.2b: Neuroimaging results SSRI-free analyses

Region of interest	Side	MNI coordinates			Z	P_{FWE}-corrected
		x	y	z		
IFG	R	57	9	9	3.11	**.02**
ACC	R	3	44	16	2.85	.04
PCC	L	-8	-37	33	2.9	**.013**
Hippocampus		N/A	N/A	N/A	N/A	>.05

MNI: Montreal Neurological Institute; Z: Z-value; FWE: family-wise error corrected for multiple comparison; IFG: inferior frontal gyrus; ACC: anterior cingulate cortex; PCC: posterior cingulate cortex; non-remMDD: non-remitted MDD patients; remMDD: remitted MDD patients; HC: healthy controls; L: left; R: right.
To account for the number of a priori regions of interest, we corrected the critical corrected p-value for the number of regions (n=4) and took the interdependency of the volumes of these regions into account, resulting in a critical corrected p-value of P< .032.