The role of BDNF in depression : will the neurothrophin hypothesis sparkle on, long after the glitter of the firework is gone?

Molendijk, M.L.

CHAPTER 5

Serum BDNF concentrations as peripheral manifestations of depression

Evidence from a systematic review and meta-analyses on 179 associations (N = 9,484)

Molendijk ML
Polak M
Bus BAA
Penninx BWJH
Spinhoven P
Elzinga BM

Published as: Serum BDNF concentrations as peripheral manifestations of depression: evidence from a systematic
review and meta-analyses on 179 associations (*N* = 9,484)
Molecular Psychiatry AOP, doi: 10.1038/mp.2013.105

SIGNIFICANCE: This systematic review and meta-analyses is noteworthy in that it confirms that alternations in serum BDNF concentrations are peripheral manifestations of depresion. Yet, the evidence for this is way slimmer as was initially thought. An important implication of the message that this paper brings is that serum BDNF concentrations probably are of little clinical use.

ABSTRACT

Meta-analyses, published in 2008/2010, have confirmed abnormally low serum BDNF concentrations in depressed patients and normalisation of this by antidepressant treatment. These findings are believed to reflect peripheral manifestations of the *neurotrophin hypothesis*, which states that depression is secondary to an altered expression of BDNF in the brain. Since the publication of these meta-analyses, the field has seen a huge increase in studies on these topics. This motivated us to update the evidence on the aforementioned associations and, in addition, to compile the data on serum BDNF concentrations in relation to the symptom severity of depression. Using a manifold of data as compared to earlier meta-analyses we find low serum BDNF concentrations in 2,384 antidepressant-free depressed patients relative to 2,982 healthy controls and to 1,249 antidepressant-treated depressed patients (Cohen's *d* = -0.71 and -0.56, *P*-values < .0000001). When publication bias is accounted for, these effect-sizes become substantially smaller (*d* = -0.47 and -0.34 respectively, *P*-values < .0001). We detect between-study heterogeneity in outcomes for which only year of publication and sample size are significant moderators, with more recent papers and larger samples sizes in general being associated with smaller between-group differences. Finally, the aggregated data negates consistent associations between serum BDNF concentrations and the symptom severity of depression. Our findings corroborate the claim that altered serum BDNF concentrations are peripheral manifestations of depression. However, here we highlight that the evidence for this claim is slimmer as was initially thought and amidst a lot of noise.

INTRODUCTION

The *neurotrophin hypothesis*, originally formulated in 1997 by Duman, Heninger and Nestler, characterizes major depressive disorder as being secondary to aberrant neurogenesis in brain regions that regulate emotion and memory. According to this hypothesis; aberrant neurogenesis is brought about by a (stress induced) lower expression of Brain-derived Neurotrophic Factor (BDNF). In addition, the neurotrophin hypothesis predicts that antidepressants are efficacious because they increase BDNF expression and herewith resolve aberrant neuronal plasticity (Duman and Monteggia, 2006; Park and Poo, 2013). A large pre-clinical literature, allowing for mechanistic insights, fits very well with these predictions. Taliaz and colleagues (2010) for instance, showed in rats that a reduction of BDNF in the dentate gyrus impairs neurogenesis and induces depressive-like behavior. Human post-mortem studies have indicated similar alternations in the brains of persons who were depressed at the time of dying (Thompson Ray et al., 2011). Further support for abnormalities in BDNF expression in depressed patients comes from clinical studies. Karege *et al.* (2002) as the first, found serum BDNF concentrations to be low in depressed patients as compared to healthy controls and lowest in persons with the highest levels of symptom severity. Shimuzu *et al.* (2003) were the first to show an increase in serum BDNF concentrations in the course of antidepressant treatment.

These findings generated a buzz of research activity and in 2008/2010 the clinical data were summarized in three meta-analyses (Sen *et al.,* 2008; Brunoni *et al.,* 2008; Bocchio-Chiavetto *et al.,* 2010). These meta-analyses, basically including the same 11 studies (N ~ 968) confirmed the finding of low serum BDNF concentrations in untreated depressed patients (effect size [Cohen's d] ~ -1) and normalization of this by antidepressant treatment (d ~ 1) whilst suggesting that these associations were not hampered by between-study heterogeneity or publication bias. Accordingly, the conclusion was: *BDNF may have potential use as biomarker for psychiatric disorders or as a predictor of antidepressant efficacy* (Sen *et al.,* 2008; page 527). Since then, the field has seen an abundance of new data on these topics. Important is that this new data entails striking variation in outcomes across studies (see for instance Basterzi *et al.,* 2009 or Elfing *et al.,* 2012). This, and the abundance of new data, motivated us to update the current state of knowledge by calculating pooled effect-size estimates on differences in serum BDNF concentrations among:

- Antidepressant-free depressed patients and healthy controls subjects
- Antidepressant-free- and antidepressant-treated depressed patients
- Antidepressant-treated depressed patients and healthy controls subjects

An additional aim was to compile the data on the putative relation between serum BDNF concentrations and the symptom severity of depression in:

- Antidepressant-free depressed patients
- Antidepressant-treated depressed patients
- Healthy control subjects

A final aim, made possible by a large amount of studies, was to learn on the potential influence that some relevant moderators might have on the outcomes of our interest.

Method

We adhered to the guidelines that are recommended by the preferred reporting items for systematic reviews and meta-analyses statement (Moher *et al.,* 2009). The literature search, decisions on inclusion, data extraction, and quality control were all performed independently by ≥ two of the authors.

Search Strategy

We searched the PUBMED, Embase, and PsychInfo through April 1st 2013 to identify eligible human studies on serum BDNF concentrations in healthy controls, depressed patients or in both. These digital searches were supplemented by backward searches in which the references to the seminal papers of interest were screened (Karege *et al.,* 2002; Shimuzu *et al.,* 2003) and by examining the reference sections of the retrieved papers.

We included peer-reviewed human studies that reported data on serum BDNF concentrations in healthy controls, and antidepressant-free and treated depressed patients. Inclusion was independent of clinical- and the methodological characteristics of the sample or study. Non-empirical studies were excluded, as were studies that were not written in English, Dutch, German or Spanish. Papers that reported on overlapping samples were excluded except for the one that reported on the largest number of subjects.

Data Extraction

We extracted, as primary outcomes, mean serum BDNF concentrations and Standard Deviation (*SD*) as a function of diagnostic status and antidepressant use and/or indices on the relation between BDNF concentrations and the symptom severity of depression (*e.g.,* Pearson's *r*). When BDNF concentrations were assessed at multiple time points we extracted the data recorded at baseline and at the longest follow-up period.

We also extracted data on mean age, gender distribution, depression severity, antidepressant use (subdivided by SSRIs, TCAs, and SNRIs), duration of antidepressant use, and the number of subjects in the study. Where records did not provide sufficient information, corresponding authors were contacted and the required data was requested. In those cases where non-significant results were reported in a paper (*e.g., P* > .05) and authors did not reply to our request, we assigned an estimated effect-size of zero.

Quality Assessment

We used the Newcastle-Ottawa Scale (NOS; Wells *et al.,* 2013) to assess the quality of the included studies. Overall quality score was defined as the frequency of criteria that were met by the particular study. We excluded NOS items 4 and 7 because these are meaningless in the context of the current paper. Mean-quality score of the included studies was 3.18 (Standard Deviation [*SD*] = 0.14). The agreement between the independent raters was excellent (Cohen's Kappa = 0.89, Standard Error [*SE*] = 0.03).

Statistical analysis

All calculations were performed using comprehensive meta-analyses 2.0 (Borenstein *et al.,* 2009). Random effects models were applied to calculate pooled Cohen's *d* (Cohen 1988) on between-group differences in serum BDNF concentrations. Pooled correlation coefficients were calculated on the relation between serum BDNF concentrations and the symptom severity of depression. All outcomes were weighted using inverse variance methods (Mosteller and Golditz, 1996). Statistical significance of the pooled effect-sizes was assessed using a Confidence Interval (*CI*) of 95%. The I^2 measure was used to quantity the amount of between-study heterogeneity and considered to be high when I^2 > 50% (Higgins and Thompson, 2002). Statistical significance of heterogeneity was assessed using the *Q*-statistic (Borenstein *et al.,* 2009).

Through meta-regression analyses the possible moderating effects of between-study differences on outcomes was evaluated. We considered the number of subjects included in the study, year of publication, mean age, symptom severity of depression of the patient sample, gender distribution, and the NOS score as potential moderators for all outcomes of interest. The severity rating scales that were used differed

between studies. These instruments use different values to quantify severity (*e.g.,* Hamilton 1960 or Rush *et al.,* 1996) that do not necessarily equate to each other. Therefore, we used the validated severity categories: none, mild, moderate, severe, and very severe that can be derived from the continuous scores on each of these instruments as potential moderating variable. The moderation analysis on the difference in serum BDNF concentrations between healthy controls and antidepressant-treated depressed patients in addition included variables coding for the class of antidepressant and the duration of treatment. For the meta-analysis on antidepressant-free and treated depressed patients, the set of moderators was extended with a variable coding for change in depression severity over treatment defined as the percentage of improvement on the depression rating scale that was used.

Publication bias was assessed by inspection of funnel-plots and the Egger test (Egger et al., 1997). The trim-and-fill procedure, a validated manner to estimate an effect-size after bias has been taken into account (Duval and Tweedie, 2000; Peters *et al.,* 2007), was performed in case of publication bias. Power and sample size calculations were performed using G*Power (Faul *et al.,* 2009). Stability of our results was evaluated by sensitivity analyses in which each study was excluded from analyses at a time.

Results

Our initial search generated 730 papers of which 55 fulfilled the inclusion criteria for at least one of our meta-analyses. From these papers we could extract 124 between-group effect-size estimates and 55 correlation coefficients. For details on the search strategy we refer to the flow chart (**Figure 1 ↓**). **Table 1 ↓** lists in which meta-analysis the papers were included and provides demographic and clinical characteristics of the included studies.

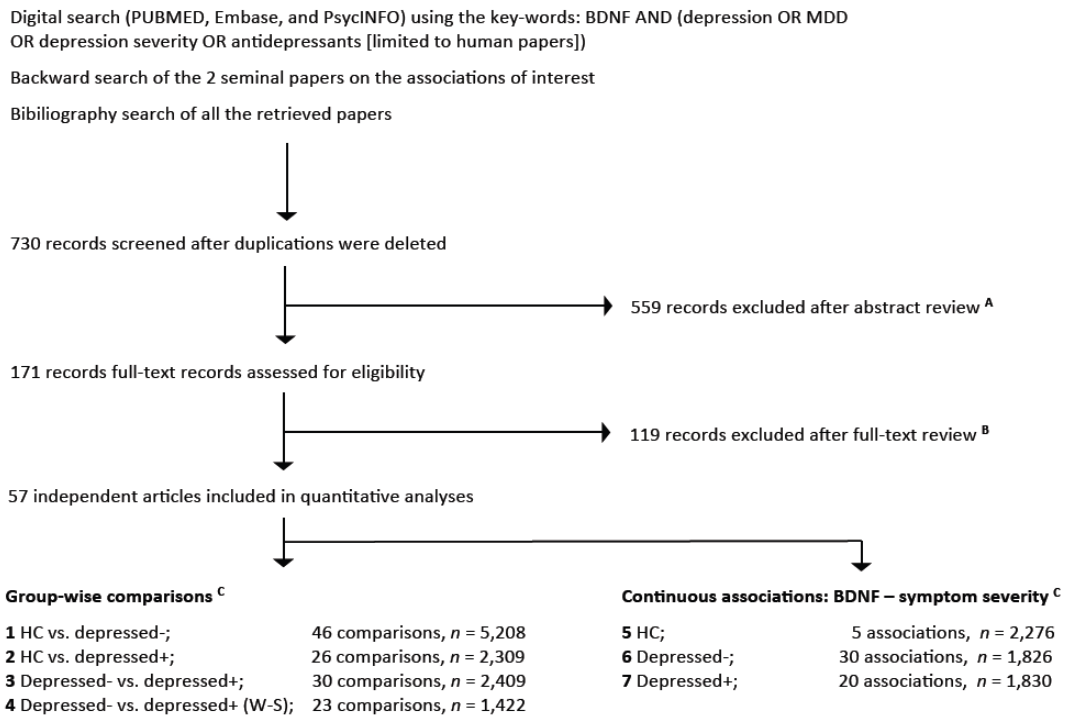

Figure 1. Flow-chart of the search strategy and results

Abbreviations: BDNF; Brain-Derived Neurotrophic Factor; HC; Healthy Controls, MDD; Major Depressive Disorder.

[A] 192 records reported on the BDNF gene, 193 records were reviews, perspectives, comments or hypotheses, 36 records reported on animal data, 14 records were post-mortem studies, 12 records were in vitro studies, and 111 records did not rapport on BDNF.

[B] 2 records reported overlapping data, 3 records reported on the BDNF gene, 64 records reported on plasma BDNF concentrations, 3 records were reviews, 43 records did not reported on serum BDNF concentrations in illnesses other than depression and did not indicate that depression related assessments were performed.

[C] Most of the articles provided input for > 1 meta-analytical effect-size. The numbersof comparisons/associations therefore do not add up to 57.

Table 1. Summary of study characteristics of included studies (studies are sorted by year and month of publication)

Author, year	meta-analysis [A]	design [B]	*N*	% female	mean age	patient status	*n* [C]	severity
Karege *et al.,* 2002	(1)(5)	B-S	60	50	37	HC	30	MADRS
						Depressed +	30	
Shimuzu *et al.,* 2003	(1)(2)(3)(5)(6)	both	83	43	43	HC	50	HAMD
						Depressed -	16	
						Depressed +	17	
Gervasoni *et al.,* 2005	(1)(2)(3)(4)	both	52	54	40	HC	26	MADRS
						Depressed -	26	
						Depressed +	26	
Gonul *et al.,* 2005	(1)(2)(3)(4)(5)	both	46	71	36	HC	18	HAMD
						Depressed -	28	
						Depressed +	28	
Karege *et al.,* 2005	(1)(4)	B-S	78	56	34	HC	35	MADRS
						Depressed -	43	
Aydemir *et al.,* 2005	(1)(2)(3)	both	20	80	36	HC	10	HAMD
						Depressed -	10	
						Depressed +	10	
Zanardini *et al.,* 2006	(6)	W-S	16	69	56	Depressed +	16	HAMD
Lommatzsch *et al.,* 2006	(1)(5)	B-S	80	100	28	HC	62	EPDS
						Depressed -	18	
Ayedemir *et al.,* 2006	(1)(2)(3)	both	40	100	35	HC	20	HAMD
						Depressed -	20	
						Depressed +	20	
Bocchi-Chiavetto *et al.,* 2006	(6)	W-S	12	70	53	Depressed +	12	MADRS
Lang *et al.,* 2006	(4)	B-S	24	*NK*	46	Depressed -	8	MADRS
						Depressed +	16	
Aydemir *et al.,* 2007	(1)	B-S	50	74	33	HC	26	HAMD
						Depressed-	24	
Yoshimura *et al.,* 2007	(1)(2)(3)(4)	both	72	65	46	HC	30	HAMD
						Depressed -	42	
						Depressed +	42	
Ziegenhorn *et al.,* 2007	(1)(5)	B-S	465	48	85	HC	259	HAMD
						Depressed-	91	
Hellweg *et al.,* 2007	(3)	W-S	40	71	51	Depressed -	40	HAMD
						Depressed +	40	
Okamoto *et al.,* 2008	(6)	B-S	18	50	61	Depressed +	18	HAMD
Stanek *et al.,* 2008	(4)	B-S	34	56	73	HC	34	PRIME
Huang *et al.,* 2008	(1)(2)(3)	both	218	72	33	HC	107	HAMD
						Depressed -	111	
						Depressed +	79	
Piccini *et al.,* 2008	(1)(2)(3)	both	30	83	42	HC	15	HAMD
						Depressed -	15	
						Depressed +	15	
Matrisciano *et al.,* 2009	(1)(2)(3)	both	41	51	37	HC	20	HDRS
						Depressed -	21	
						Depressed +	21	
Basterzi *et al.,* 2009	(1)(2)(3)	both	58	67	33	HC	15	HAMD
						Depressed -	43	
						Depressed +	43	
Gorgulu *et al.,* 2009	(1)(2)(3)	both	72	69	36	HC	31	HAMD
						Depressed -	41	
						Depressed +	22	
Grønli *et al.,* 2009	(6)	B-S	15	60	70	Depressed +	15	HAMD
Umene-Nakano *et al.,* 2009	(1)(5)	B-S	40	25	44	HC	20	HAMD
						Depressed -	20	
Fernandes *et al.,* 2009	(2)(6)	B-S	40	60	42	HC	30	HAMD
						Depressed +	10	
Lee *et al.,* 2009	(1)	B-S	132	61	74	HC	98	GDS
						Depressed -	34	
Ozan *et al.,* 2010	(1)	B-S	122	70	34	HC	56	HAMD
						Depressed -	66	
Diniz *et al.,* 2010	(1)(4)	B-S	71	83	70	HC	42	HAMD
						Depressed -	29	
Eker *et al.,* 2010	(1)(4)	B-S	47	75	31	HC	22	HAMD
						Depressed -	25	
Bocchi-Chiavetto *et al.,* 2010	(1)(4)	B-S	84	81	43	HC	59	MADRS
						Depressed -	25	

Table 1 *continues on the next page*

Table 1 *continued*

Author, year	meta-analysis [1]	design [2]	*N*	% female	mean age	patient status	*n*	severity
Hu	(1)	B-S	84	73	43	HC	28	HAMD
						Depressed a -	28	
						Depressed b -	28	
Zhou *et al.,* 2011	(1)	B-S	123	*NK*	*NK*	HCa	30	HAMD
						HCb	58	
						Depressed -	35	
Su *et al.,* 2011	(1)	B-S	52	0	23	HC	21	*NK*
						Depressed -	31	
Rojas *et al.,* 2011	(3)	B-S	34	71	42	Depressed -	34	HAMD
						Depressed +	34	
Yoshimura *et al.,* 2011	(3)(4)	W-S	132	60	51	Depressed –	132	HAMD
						Depressed +	132	
Wolkowitz *et al.,* 2011	(1)(2)(3)	B-S	57	36	39	HC	28	HAMD
						Depressed -	29	
						Depressed +	25	
Kobayakawa *et al.,* 2011	(1)	B-S	162	30	65	HC	81	HADS
						Depressed -	81	
Terraciano *et al.,* 2011	(5)	B-S	2,099	62	51	HC	1,661	CES-D
						Depressed -	438	
Molendijk *et al.,* 2011	(1)(2)(3)(4)(5)(6)	B-S	1,344	65	42	HC	382	IDS
						Depressed -	541	
						Depressed +	421	
Toups *et al.,* 2011	(6)	B-S	70	80	47	Depressed +	70	HAMD
Satomura *et al.,* 2011	(2)(4)(5)	B-S	272	63	53	HC	163	HAMD
						Depressed +	109	
Sasaki *et al.,* 2011	(1)(2)(3)(5)(6)	B-S	52	56	13	HC	22	CDRS-R
						Depressed -	19	
						Depressed +	11	
Sozeri-Varma *et al.,* 2011	(1)(4)	B-S	70	73	37	HC	40	HAMD
						Depressed -	30	
Bus *et al.,* 2012	(4)	B-S	1,230	50	61	HC	1,230	BDI
Gedge *et al.,* 2012	(5)	W-S	29	69	45	Depressed +	29	HAMD
Gazal *et al.,* 2012	(1)	B-S	72	100	25	HC	36	BDI
						Depressed -	36	
Birkenhäger et al., 2012	(6)	W-S	42	43	47	Depressed -	42	HAMD
Deuschle *et al.,* 2012	(1)(2)(3)(4)	W-S	70	72	52	HC	14	HAMD
						Depressed -	56	
						Depressed +	56	
Harvey *et al.,* 2012	(1)(5)	W-S	200	49	44	HC	89	PHQ-9
						Depressed -	111	
Oral *et al.,* 2012	(1)(5)	B-S	79	68	27	HC	40	BDI
						Depressed -	39	
Karlovic *et al.,* 2012	(1)	B-S	264	50	46	HC	142	HAMD
						Depressed -	122	
Jeon *et al.,* 2012	(1)(2)(3)(4)	W-S	155	71	44	HC	50	HAMD
						Depressed -	105	
						Depressed +	105	
Yoshida *et al.,* 2012	(2)(5)	B-S	147	56	38	HC	78	SIGH-D
						Depressed +	69	
Elfving *et al.,* 2012	(1)(2)	B-S	406	81	46	HC	289	ICD-10
						Depressed -	117	
						Depressed +	45	
Papakostas *et al.,* 2013	(1)	B-S	79	52	36	HC	43	HAMD
						Depressed -	36	

Abbreviations: HC, Healthy controls; Depressed -, antidepressant free; Depressed +, antidepressant treated; *NK*, Not Known.

[A] This column indicates in which meta-analysis the study that is indicated in the corresponding row is included: (1) HCs vs. depressed -; (2) HCs vs. depressed +; (3) Depressed - patients vs. MDD+; (4-6) regard meta-analyses on continuous associations between serum BDNF concentrations and depression severity scores: (4) in HC's; (5) in depressed -; (6) in depressed +.

[B] This column, design, indicates whether Within-Subjects data (W-S), a Between-Subjects data (B-S), or a combination of these types of data (both) is used by the study that is indicated in the corresponding row.

[C] Note that the numbers in the column *n* do not add to the numbers as they are given in the column *N*. This is because the numbers in column *n*, in some instances, are counted double (*e.g.,* before and after antidepressant treatment in longitudinal designs).

Meta-analyses

Random-effects meta-analyses showed that antidepressant-free depressed patients had lower BDNF concentrations as compared to healthy controls (*d* = -0.71, 95% *CI* = -0.89 — -0.53*, P* < .0000001; 46 comparisons, *n* = 5,203; see **Figure 2 ↓**) and to those of antidepressant-treated depressed patients (*d* = -0.56, 95% *CI* = -0.77 — -0.35*, P* < .00001, 28 comparisons, *n* = 4,204). Repeating this latter analysis using only studies that reported pre- and post-treatment BDNF concentrations gave a somewhat higher effect-

size estimate (*d* = -0.74, 95% *CI* = -1.04 — -0.45*, P* < .0000001, 23 comparisons, within-subjects data on 711 patients pre- and post-treatment). Differences in BDNF concentrations among healthy controls and antidepressant-treated depressed patients were not observed (*d* = 0.07 *P* = .52; 24 comparisons, *n* = 3,720). Forest plots (except **Figure 2 ↓**) are provided as supplementary materials (**Figure S1**–**S3**) in **Appendix III** of this thesis.

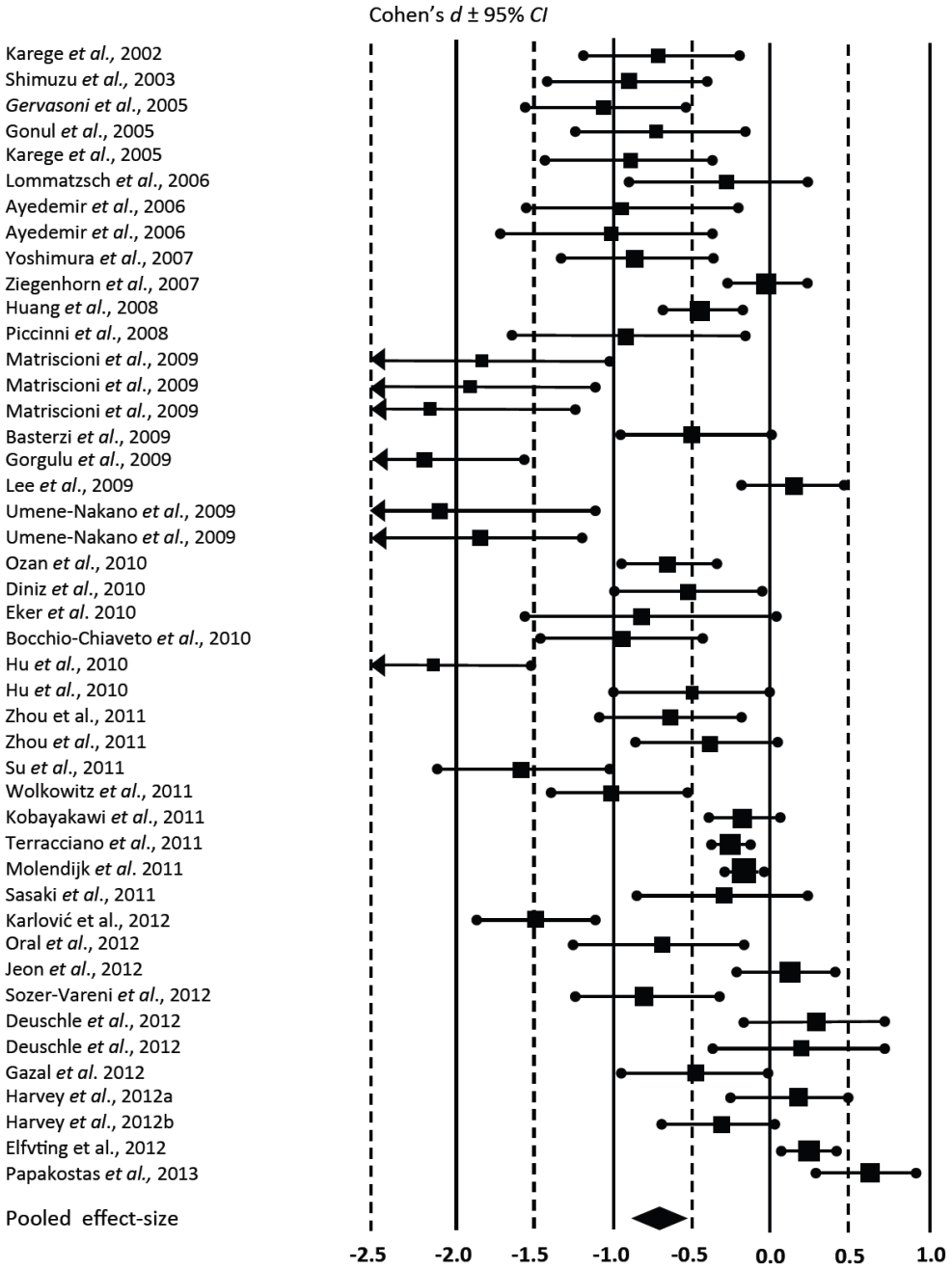

Figure 2. Forrest plots for random effect meta-analyss on differences in serum BDNF concentrations between healthy control subjects and antidepressant-free depressed patients. The sizes of the squares are proportional to sample size.

A meta-analysis aggregating 30 associations (*n* = 1,807) on the relation between BDNF concentrations and the symptom severity of depression in antidepressant-free depressed patients yielded a statistically significant, negative correlation (*r* = -0.19; 95% *CI* = -0.28 — -0.10*, P* < .00001). There was no evidence for a relation between serum BDNF concentrations and depression severity in antidepressant-treated depressed patients (*r* = -0.02; *P* = .36, 20 associations, *n* = 1,820) or in healthy controls (*r* = -0.02; *P* = .41, 5

associations, *n* = 2,276). Forest plots are provided as supplement (**Figure S4–S6**) in **Appendix III** of this thesis.

Between–study heterogeneity and moderation analyses

A large amount of between-study heterogeneity in outcomes was identified in all meta-analyses that yielded significant outcomes (55% < I^2 < 87%, for I^2-, *Q*-, and *P*-values we refer to **Table 2 ↓**).

Table 2. Statistics on between study heterogeneity and publication bias for the meta-analysis indicated in the row

	No. of associations	No. of subjects			Heterogeneity			Publication bias	
		HC	Depressed-	Depressed+	I^2	*Q*	*P*	Egger's *t*	*P*
Group-wise comparisons									
HC vs depressed -	41	2,911	2,292	*NA*	86.1%	287.6	< .001	4.2	< .001
HC vs depressed+	24	2,591	*NA*	1,129	84.6%	150.2	< .001	1.4	.16
Depressed- vs depressed +	27	*NA*	2,955	1,249	84.4%	165.1	< .001	2.5	< .05
Depressed- vs depressed + W-S [1]	23	*NA*	711	711	83.9%	136.8	< .001	2.6	< .05
Continuous associations									
HC	5	2.276	*NA*	*NA*	14.8%	4.7	.32	1.0	.15
Depressed -	29	*NA*	1,807	*NA*	67.9%	87.2	< .001	2.5	< .05
Depressed+	19	*NA*	*NA*	1,820	18.3%	48.9	.36	0.6	.53

Abbreviations: HC, Healthy controls; depressed-, antidepressant free; depressed+, antidepressant treated; *NA*, Not Applicable; W-S, Within-Subjects data

[1] Here, only associations were included that were derived using a within-subjects designs (*i.e.,* treatment studies)

In a series of meta-regression analyses, we aimed to identify sources of heterogeneity in outcomes. We observed that differences in serum BDNF concentrations among antidepressant-free depressed patients and healthy control subjects could partly be explained by sample size ($r = -0.33$, $R^2 = 0.11$, $P = 0.03$) and by year of publication ($r = -0.30$, $R^2 = 0.09$, $P = 0.04$), with larger samples and more recently reported papers in general reporting smaller between-group differences. In the meta-analysis on changes in serum BDNF concentration over the course of antidepressant treatment, we found that a larger decrease in symptom alleviation was accompanied by a larger increase in BDNF concentrations ($r = -0.48$, $R^2 = 0.22$, $P = 0.01$). Other moderators, including NOS score, were not observed (see **Table 3 ↓** for all coefficients). Moderation analyses were not performed when between-study heterogeneity was not detected.

Publication bias and power

Visual inspection of the funnel plots suggested that there was evidence for publication bias in all meta-analyses that yielded a significant outcome. Egger's tests confirmed this (*t*-values in the range 2.5 – 4.2, *P*-values all < .05, see **Table 2 ↑** for exact values).

Trim-and-fill estimations were used to assess the impact of publication bias. The meta-analysis on differences in BDNF concentrations among healthy controls and untreated depressed patients suggested that 9 studies had to be imputed to result in a symetric funnel plot. Imputation led to a smaller, yet significant, effect-size (*d* = -0.47, 95% *CI* = -0.64 — -0.27, *P* < .000001). The pattern of publication bias was similar in the meta-analyses comparing group differences among antidepressant-free and treated subjects, where 5 (all data) and 4 studies (within-subjects data) needed to be imputed to yield a symetric funnel plot. Also here, imputation led to smaller effect-size estimates (*d* = -0.54 and -0.34 respectively, *P*-values < .001). Likewise, for the meta-analyses on the continous association between serum BDNF concentrations and the symptom severity of depression in untreated depressed persons, the trim-and-fill estimations suggested that 5 studies had to be imputed to result in a symetric funnel plot pattern. Herewith, the effect-size estimate (*r* = -0.07) was no longer statistically signinicant. Funnel plots are provided in **Appendix III** (**Figures S7–S10**).

Table 3. Associations (Pearson's correlation coefficients for continuous- and Spearman's Rho correlation coefficients for categorical variables) between study characteristics and study effect size (by meta-analysis)

Group differences	HC vs. depressed- 41 effect-sizes *n* = 5,203	HC vs. depressed+ 24 effect-sizes *n* = 3,720	Depressed- vs. depressed+ 27 effect-sizes *n* = 4,204	Depressed- vs. depressed+ W-S 23 effect-sizes *n* = 1,422
Gender (percentage female)	0.16	0.11	0.06	0.08
Age (mean, years)	0.13	-0.11	0.08	0.11
Depression severity (cat.)	-0.17	-0.10	-0.21	-0.07
Percentage SSRI	*NA*	0.29	-0.35 #	-0.34
Percentage TCA	*NA*	-0.21	0.13	0.11
Percentage SNRI	*NA*	-0.10	0.17	0.17
Percentage NaSSA	*NA*	-0.14	0.14	0.15
Duration of treatment (weeks)	*NA*	-0.34	0.04	0.04
Clinical response on treatment	*NA*	*NA*	*NA*	-0.48 *
Sample size (*n*)	0.33 *	-0.15	0.25	0.21
Year of publication	0.30 *	-0.16	0.18	0.18
Study quality (criteria met)	0.04	0.06	0.35 #	0.34

Abbreviations: HC, Healthy controls; depressed-, antidepressant free; depressed+, antidepressant treated; *NK*, Not Known; W-S, Within-Subjects data only (*i.e.,* associations were that were derived using a within-subjects design.
[1] Given that there was no evidence for between-study heterogeneity, moderation analysis was not performed in these sub-groups.
* Statistically significant at $P < .05$ # Trend-like finding at $P < .10$

We calculated the numbers of subjects that are needed to detect differences with a power of 0.80 at an α-level of .05 (one-sided). Hereto we used the pooled effect-size estimates that were corrected for publication bias. These calculations suggested that 57 subjects in each group would be neccesary to reliably detect differences in serum BDNF concentrations between healthy controls and antidepressant free depressed subjects. For differences in serum BDNF concentrations among antidepressant-free and treated persons, this number would be 108. Based on this, the majority of the included samples was not sufficiently powered (observed median sample size = 36). Sample-size calculations were not performed for continuous associations between serum BDNF concentrations and the symptom severity of depression since these were not statistically significant.

Sensitivity analyses indicated that none of the study findings was unduly driven by the effect of a particular study.

Discussion

Here we confirm, based on a manifold of data as compared to previous meta-analyses (Sen *et al.,* 2008; Brunoni *et al.,* 2008; Bocchio-Chiavetto *et al.,* 2010) that serum BDNF concentrations are low in untreated depressed patients and normalized by antidepressant treatment. The moderate to large effect-sizes that we rapport on these differences (random-effects meta-analyses, *d* = -0.71 and -0.56 respectively) are similar to the ones that were reported in the seminal studies (Karege *et al.,* 2002; Shimuzu *et al.,* 2003) and in previous meta-analyses. These findings are not new. The novelty of our work, instead is that our analyses highlight a large amount of unexplained between-study heterogeneity in outcomes and publication bias that together may call for a critical interpretation of the claim that altered serum BDNF concentrations are related to, and a clinical useful marker for, the illness depression.

We find a large amount of between-study heterogeneity in outcomes and none of the theoretically relevant variables that we tested (*e.g.,* the symptom severity of depression or gender distribution of the sample) was associated with this. Understanding the sources of the observed heterogeneity is essential and obviously, it may have come from between-sample characteristics that were not tested in our study, such as alcohol consumption and smoking (Bus *et al.,* 2011), sleep problems (Giese *et al.,* 2013), seasonality (Molendijk *et al.,* 2012), or exposure to trauma (Elzinga *et al.,* 2011). Given that depression is a

heterogeneous illness (Rush 2007), heterogeneity in outcomes may also have come from diversity in clinical characteristics of patient samples. The severity of depression, however, did not explain it. Unfortunately, we did not have the opportunity to test many of the other clinical characteristics because most of the included studies did not report on these variables.

We did find an artificial base for the heterogeneity in outcomes. First, a large part of the studies included in our meta-analysis was underpowered. Given that a low level of power increases the false versus true positive ratio (Sterne and Smith, 2001), some overly positive findings may have been among the studies that we included, causing heterogeneity in outcomes. Second, we found that sample size and year of publication were significant predictors of between-study heterogeneity, with larger samples and more recently published findings being associated with smaller between-group differences. This indicates publication bias; a particular threat to the validity of a meta-analysis (Dickersin 1990). We indeed found evidence for publication bias in funnel-plots (Egger *et al.,* 1997) and we applied validated trim-and-fill procedures to provide effect-size estimates that account for this (Peters *et al.,* 2007). These yielded attenuated effect-size estimates that were about half as large as those reported in previous meta-analysis (Sen *et al.,* 2008; Bocchio-Chiavetto *et al.,* 2010) and of moderate magnitude at best (*d* = -0.47 through -0.34). The often discussed association between serum BDNF concentrations and the symptom severity of depression (*e.g.,* Karege et al., 2002), for which we initially found some evidence, even lost its statistical significance after correcting for publication bias and thus likely does not exist. Given that the relevance of a diagnostic biomarker (*i.e.,* a variable that is able to distinguish between diagnostic groups; Kapur *et al.,* 2012) depends on the magnitude of an effect-size (and not on statistical significance *per se*; Kapur *et al.,* 2012), we conclude that serum BDNF concentrations are likely to be of little clinical use (as has been suggested in two earlier excellent reviews Groves 2007; Gass and Hellweg, 2010). Complicating this even more is that low serum BDNF concentrations have been reported in persons diagnosed with schizophrenia (Green *et al.,* 2011), bipolar disorder (Fernandes *et al.,* 2011), eating disorders (Montleone *et al.,* 2005), and anxiety (Molendijk *et al.,* 2012) indicating that serum BDNF concentrations are not specific enough to differentiate among diagnoses. Multiple-assay methods may serve a role as biomarker better, as recently has been shown (Papakostas *et al.,* 2013).

Although limited in scope with regard to clinical utility, our findings do not dismiss the possibility that abnormalities in BDNF expression reflect the pathophysiological processes that may underlie depressive illnesses (Duman *et al.,* 1997; Duman and Monteggia, 2006). Even more, the associations that we report on, also when adjusting for publication bias, stand out as being strong when compared to other biological abnormalities in depression, for instance blood markers for immune dysregulation (*e.g.,* CRP and IL-6 [*d* = 0.15 and 0.25 respectively]) or HPA-axis activity (*e.g.,* adrenocorticotropin hormone [*d* = 0.28] for a review on these abnormalities see Penninx *et al.,* 2013).

A difficulty that remains however is that we studied peripheral BDNF concentrations. There are indications that BDNF concentrations measured in serum reflect BDNF activity in the brain (*e.g.,* Dawood *et al.,* 2007; Klein *et al.,* 2010). However, it has never been proven that peripheral BDNF concentrations directly reflect or influence the pathophysiology of depression. A complication is that other tissues than the brain, including immune-, liver-, smooth muscle-, and vascular endothelial cells serve as sources of BDNF (Cassiman *et al.,* 2001; Karege *et al.,* 2002b). The lower peripheral BDNF concentrations in depression and up-regulation of this in the course of antidepressant treatment therefore may be an epiphenomenon resulting from an altered BDNF expression (or metabolism) by these peripheral organs. Therfore, the alternations that we rapport on do not neccesarly indicate that similar alternations occur at a central level and conclusions should not be overbearing.

Strengths and limitations

The work presented herein has as obvious strength that it is based on a large amount of data (total *N* = 9,484), yielding in general accurate effect-size estimates (Ioannidis 2005). Another strength is that through sensitivity- and moderation analyses we addressed the potential influence of single studies and sources of heterogeneity. Notwithstanding this, our work carries limitations that need to be reflected upon.

Some limitations regard the methods that we used. First, we relied on funnel-plot assymetry and trim-and-fill estimations to assess publication bias. These methods are limited in that one never knows whether asymmetry in a funnel-plot is due to publication bias or to unmeasured differences between studies (Munafo and Flint, 2004) and whether the most extreme effect-sizes are the ones that are left unpublished (Peters *et al.,* 2007). Second, in at least some regards the methods that we used were limited with regard to their ability to detect associations. The meta-regression analyses, for instance, may have been underpowered. Besides, *P*-values were not adjusted for multiple comparisons. Also important is that there may have been noise in our assessment of individual study quality. The NOS scale that we used to this end, although recommended by the Cochrane Collaboration (www.cochrane.org) is not rigorously validated and therefore our quality assessments may have been unreliable (Sanderson *et al.,* 2007). Together, this may have limited our ability to detect true associations (*i.e.,* false negatives) or may have led to the detection of associations that in reality do not exist (*i.e.,* false positives). Finally, our findings are limited in scope in that they cannot be directly generalized to other BDNF parameters such as plasma or whole blood BDNF concentrations since there is no one–to–one relationship among these measures (see for instance Terracciano *et al.,* 2010).

Future work

There are several issues that deserve future research attention. First, our finding of a greater increase in serum BDNF concentrations in the course of antidepressant treatment is associated with a larger decrease in depression symptom severity may fuel work into the temporal dynamics between BDNF expression and treatment efficacy. It would be interesting if future studies could address early changes in the course of (non-)pharmacological treatment, a notion for which some evidence exists (Lang *et al.,* 2006; Machado-Vieira *et al.,* 2009; aan het Rot *et al.,* 2012). Besides, the prediction of how successful a given treatment will be, based on changes in serum BDNF concentrations (*i.e., a treatment biomarker*) is clinically relevant (see for instance Schmidt *et al.,* 2011). In our meta-analysis we did not have the possibility to address this because most of the included studies reported on pre- and post BDNF concentrations only. Another venue for future investigations regards the distinction between the pro- and the mature BDNF variant. The ELISA kits that currently are in use to quantify BDNF are not sensitive enough to make this distinction. Given the proposed opposing effects of these two BDNF variants (proBDNF is believed to induce apoptosis; Park and Poo, 2013) it would be interesting to study pro/mature BDNF ratios and whether these differ among diagnostic groups. The tools hereto were only recently developed and validated (Yoshida *et al.,* 2012).

With regard to future work on peripheral BDNF concentrations we finally wish to note that analyses would gain credibility if they were controlled for relevant confounding factors and performed using data (preferably within-subject) on a sufficiently large sample (*N* ~ 150, according to our power-analyses).

Concluding remarks

Our meta-analyses (aggregating 179 effect-size estimates; *N* = 9,484) initially yielded support for the claim that alternations in serum BDNF concentrations are peripheral manifestations of depression. This is not new. The important contribution of our work however is that we clearly show that between-study heterogeneity, underpowered designs, and publication bias are at play that together give rise to inflated

effect-size estimates. Together this suggest that the evidence base for the claim that altered serum BDNF concentrations are peripheral manifestations of depression is slimmer as was initially thought and amidst a lot of noise.