

University of Groningen

Evidence-b(i)ased psychiatry

de Vries, Ymkje Anna

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

de Vries, Y. A. (2018). *Evidence-b(i)ased psychiatry*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Chapter 6

Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene, life stress and depression

Ymkje Anna de Vries, Annelieke M. Roest, Minita Franzen,
Marcus R. Munafò, Jojanneke A. Bastiaansen

Psychological Medicine (2016), 46, 2971 - 2979

Abstract

Background: Caspi *et al.*'s 2003 report that 5-HTTLPR genotype moderates the influence of life stress on depression has been highly influential but remains contentious. We examined whether the evidence base for the 5-HTTLPR-stress interaction has been distorted by citation bias and a selective focus on positive findings.

Methods: Seventy-three primary studies were coded for study outcomes and focus on positive findings in the abstract. Citation rates were compared between studies with positive and negative results, both within this network of primary studies and in Web of Science. In addition, the impact of focus on citation rates was examined.

Results: Twenty-four (33%) studies were coded as positive, but these received 48% of within-network and 68% of Web of Science citations. The 38 (52%) negative studies received 42% and 23% of citations, respectively, while the 11 (15%) unclear studies received 10% and 9%. Of the negative studies, the 16 studies without a positive focus (42%) received 47% of within-network citations and 32% of Web of Science citations, while the 13 (34%) studies with a positive focus received 39% and 51% respectively, and the 9 (24%) studies with a partially positive focus received 14% and 17%.

Conclusions: Negative studies received fewer citations than positive studies. Furthermore, over half of the negative studies had a (partially) positive focus, and Web of Science citation rates were higher for these studies. Thus, discussion of the 5-HTTLPR-stress interaction is more positive than warranted. This study exemplifies how evidence-base-distorting mechanisms undermine the authenticity of research findings.

Introduction

Major depressive disorder (MDD) is a complex illness, caused by a combination of genetic and environmental risk factors [301]. One of the most robust risk factors for MDD is the experience of a stressor, such as a stressful life event or childhood abuse [302]. However, many who experience such a stressor do not develop depression. This individual variability has been suggested to be due, at least in part, to genetic variation [303].

In 2003, Caspi and colleagues reported that a polymorphism in the serotonin transporter gene (5-HTTLPR) moderates the relationship between life stress and depression: while carriers of at least one short (S) allele had a similar risk of depression as people homozygous for the long (L) allele in the absence of stress, S carriers were up to twice as likely to develop depression after stressful life events or childhood abuse [304].

This study has since been highly cited (>3,800 times, Web of Science, October 2015) and has become the seminal finding within the burgeoning field of gene-environment interactions (G×E). However, this finding also remains highly contentious. Even meta-analyses on this topic contradict each other, with some finding evidence of an effect [305, 306], while others do not [307, 308].

Many issues complicate the interpretation of G×E findings and replications, such as publication bias [309] and analytical flexibility [64, 310], which increases the chance of false-positives due to the multitude of analyses performed [311]. The likelihood of false-positives is further increased by low power and by the low prior probability of associations in candidate gene studies [309]. Although replication has been suggested as the solution to false-positive findings [312], many G×E studies are imprecise replications of the original finding, and a loose definition of replication may still permit propagation of false-positives [63].

Additionally, researchers may emphasize positive findings while downplaying negative findings. Within the randomized controlled trial literature, such reporting strategies, whether intentional or unintentional, that focus on positive (secondary) findings (in spite of non-significant results for the primary outcome) and that may distort the interpretation of results, are defined as ‘spin’ [20, 40]. A focus on positive findings has also been demonstrated in observational studies [300]. As a consequence, the published literature on a topic may appear more convincing than is justified by the strength of the evidence.

Selective citation may also affect the quality of the evidence base [44]. Statistically significant (positive) studies are cited more frequently than non-significant (negative) studies [42, 45, 313, 314], which may render non-supportive studies relatively invisible.

Citation bias and focus on positive findings can also work synergistically to hide negative results from view. A previous examination of citation patterns on a related topic, that of 5-HTTLPR and amygdala activation [315], showed that negative studies that had been spun were cited at a similar rate as positive studies, while negative studies that had not

been spun received almost no citations. The resulting invisibility of negative findings may create the impression that this effect has been proven beyond doubt, although meta-analyses have questioned its robustness [297, 298].

In the current study, we aimed to determine whether citation bias and selective focus on positive findings are also present in the literature on 5-HTTLPR, life stress, and depression. Achieving a better understanding of the etiology of depression is of vital importance to psychiatry, given the high burden of depression [316]. Distortion of the evidence base could mislead researchers and clinicians and thus pose a major obstacle to this goal.

Methods

Study selection

To establish the network of primary studies, we searched PubMed for the most recent meta-analysis on 5-HTTLPR, stress, and depression [305], which included 81 studies. For each study, we determined the outcome for the effect of interest (i.e., 5-HTTLPR x stress).

We included studies with continuous outcomes (e.g., score on a depression questionnaire), as well as studies with binary outcomes (e.g., depression diagnosis). We excluded studies in which the outcome was clearly a different construct than depression (e.g., cognitive dysfunction). Studies were included regardless of whether the 5-HTTLPR x stress interaction effect on depression was the primary outcome. No exclusion criteria were applied for stressors, which were very diverse.

Coding study outcomes

Coding was done in duplicate by two independent raters (YV and MF), and disagreements were resolved by discussion with AR and JB. Study outcome was coded as positive, negative, or unclear. We coded a study outcome as unclear if we could not determine whether the 5-HTTLPR x stress interaction was significant, for instance because only the p-value associated with a three-way interaction (e.g., 5-HTTLPR x stress x gender) was presented. Study outcome was coded as positive if the extracted p-value was <0.05 , provided that the interaction was in the expected direction (i.e., S allele associated with increased depression), and as negative otherwise.

P-values were extracted according to a hierarchical decision tree. We first determined whether the design of the study was “exposed-only”. In these studies, the entire sample was exposed to a stressor, such as a somatic illness. The effect of interest, in this case,

is not an interaction but the main effect of 5-HTTLPR. Hence, we extracted the p-value associated with the main effect for these studies. For all other studies, we determined whether a p-value was reported for a two-way interaction between 5-HTTLPR and stress, consistent with Caspi and colleagues (2003) [304].

If multiple relevant, independent outcomes or stressors were included in a study, we extracted all p-values. Following Sharpley and colleagues (2014) [305], we averaged these p-values to arrive at a conclusion. If multiple non-independent outcomes were given (e.g., a continuous symptom scale and a dichotomized version thereof), we only included the continuous outcome.

When studies provided p-values for both biallelic and triallelic genotyping, we erred towards coding a study as positive by selecting the smallest p-value, as it is unclear which genotyping approach should be preferred [317, 318, 319]. If both unadjusted and adjusted analyses were given, we also used the smallest p-value. We preferentially extracted the p-value of an overall test of interaction; however, if only post-hoc comparisons were available, we extracted the p-value associated with the SS vs LL homozygotes comparison.

Coding study abstracts

Two independent raters (YV and MF) coded the abstract of each study, and discrepancies were resolved by discussion with AR and JB. Abstracts were preferentially coded based upon their conclusions, but if these did not provide a clear statement, we used the results section of the abstract. In coding abstracts, we were interested in the way abstracts reported on how their findings reflected on the original result by Caspi and colleagues (2003) [304].

Abstracts were coded as positive if a claim was made that the results supported the existence and/or importance of the 5-HTTLPR x stress interaction. Abstracts were coded as partially supportive if a positive claim was made that was not directly related to the 5-HTTLPR x stress interaction (e.g., positive findings for a three-way interaction) or if the abstract mentioned findings for multiple outcomes or stressors and not all were positive. Abstracts that did not make a positive claim or that made an explicitly negative claim were coded as negative. If the abstract did not report on the effect of interest, the study was excluded (2 studies).

Citation outcomes

We examined citations both within the network of primary studies and outside of the network in the broader literature [315]. To examine within-network citations, we constructed a citation grid and marked for each study by which of the other included studies it was cited. Total citation counts for each study were calculated from the grid. To examine

out-of-network citations, we looked up the citation counts for each study on Web of Science (Core Collection, October 2015). To create non-overlapping outcomes, we pruned the within-network citations from the Web of Science citations. While within-network citations represent citations by other experts working within the 5-HTTLPR x stress field, Web of Science citations also include citations by researchers not directly involved in this area.

Analyses

For our citation analysis, we first compared the citations received by studies with positive, negative or unclear outcomes (irrespective of abstract coding). The sum of citations was calculated and the percentage of all citations received by studies with a given outcome was determined. In examining within-network citations, we excluded the most recent study, as it could not have been cited within the network. We also examined the study by Caspi and colleagues (2003) [304] separately, as we expected it to receive many citations.

To determine whether a (selective) focus on positive findings was present, we examined the number of negative studies with a negative abstract (studies without a positive focus), a partially supportive abstract (studies with a partially positive focus), or a positive abstract (studies with a positive focus). We then examined the impact of focus on citation rates by calculating the percentage of all citations to negative studies received by each type of negative study.

Within the network, we also examined whether positive studies, negative studies without positive focus, and negative studies with a (partially) positive focus showed different citation patterns, that is, whether positive studies were more likely to cite other positive studies and negative studies more likely to cite other negative studies.

We performed several sensitivity analyses. First, since older studies have had more opportunities to be cited, we re-examined citation rates based on measures taking into account publication year. For within-network citations, the percentage of subsequent studies citing a given study was calculated; for out-of-network citations, the yearly citation rate was calculated. Second, as the distribution of citations is right-skewed, we examined the median number of citations to each study type. Third, we recoded the outcome for studies with multiple relevant p-values based upon the smallest p-value. As it is often unclear what should be considered the primary outcome, we used average p-values in our main analysis; however, in some cases the smallest p-value may have been associated with the outcome considered most important by the authors, which is why we performed this sensitivity analysis.

Since the included studies form the total population of studies on the effect of interest, we used descriptive analyses rather than statistical tests [315], which are designed to generalize from a sample to a hypothetical larger population.

Results

Coding of studies and abstracts

We excluded ten of the 81 studies in Sharpley and colleagues (2014) [305]: eight studies were excluded because the outcome was not depression-related, no stressor was included, or the entire sample was depressed; one study was excluded because the abstract did not report on 5-HTTLPR; and one study was excluded because the abstract did not report on the depression outcome. Furthermore, we included two additional studies that had been excluded from the meta-analysis because the sample was a subset of those included in a later study.

Consequently, we included 73 studies, of which 24 studies were coded as positive, 38 studies as negative, and 11 studies as unclear in terms of outcome. Of the 11 unclear studies, four studies were coded as unclear because of the inclusion of three-way interactions in the model (e.g., with social support), while another study was coded as unclear because the 5-HTTLPR x stress interaction was only tested in males and females separately. Four studies were coded as unclear because the 5-HTTLPR x stress interaction was not tested (e.g., only the main effect of 5-HTTLPR in the different stress groups was tested). Finally, two studies were coded as unclear because we could not determine whether the (averaged) p-value was <0.05 , as one p-value was given as “non-significant” while another was <0.05 .

Inter-rater agreement was moderate ($\kappa=0.49$). Our agreement with Sharpley and colleagues (2014) [305] was good: within the subset of studies included in both Sharpley and colleagues (2014) and our own paper and that we coded as positive or negative (rather than unclear), the percentage of positive studies was 38% (23 out of 60) by both our coding and Sharpley’s coding; coding was identical for 54 out of 60 (90%) papers.

Of the 73 studies, we coded 40 abstracts as positive, 16 abstracts as negative, and 17 abstracts as partially supportive. Inter-rater agreement for abstract coding was good ($\kappa=0.71$). A full table of studies with characteristics and coding is given in Table 6.1 in the Appendix.

Citations by study outcome

Figure 6.1 shows the percentage of citations to positive, negative, and unclear studies (outer circle) compared to the percentage of studies of each type (inner circle).

The total number of citations was 488 within the network and 9160 on Web of Science. Positive studies, comprising 33% of all studies, received 236 (48%) within-network citations and 6187 (68%) Web of Science citations. Negative studies (52% of all studies) received 205 (42%) within-network citations and 2113 (23%) Web of Science citations,

while unclear studies (15% of all studies) received 47 (10%) within-network citations and 860 (9%) Web of Science citations. The study by Caspi and colleagues (2003) [304] received a large share of the citations to positive studies, particularly in Web of Science. However, even after exclusion of this study, positive studies still received 40% of within-network and 45% of Web of Science citations, as compared to 48% and 39%, respectively, for negative studies.

On average, negative studies received 5.5 (standard deviation (SD)=9.3) within-network citations, while unclear studies received 4.3 (SD=6.2) and positive studies received 9.8 (SD=14.6). Positive studies other than Caspi and colleagues (2003) [304] received 7.4 (SD=8.9) within-network citations on average. For Web of Science, negative studies received, on average, 55.6 (SD=72.3) citations, while unclear studies received 78.2 (SD=61.6) and positive studies received 257.8 (SD=765.5) citations. Positive studies other than Caspi and colleagues (2003) received 103.8 (SD=132.4) citations on average.

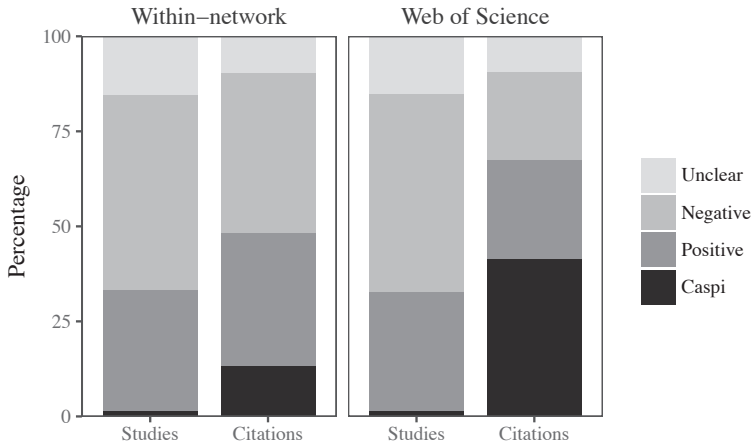


Figure 6.1: *Percentage of studies of each type (the study by Caspi and colleagues (2003), other positive studies, negative studies, and unclear studies) and the percentage of citations received by each type of study for within-network citations and Web of Science citations.*

Presence of positive focus in abstracts

Figure 6.2 depicts the presence of a positive focus within the set of studies. Of the 24 positive studies, 21 (88%) abstracts were positive and 3 (13%) abstracts were partially supportive. These partially supportive abstracts focused on gender differences (2 abstracts) or on a three-way interaction (1 abstract). Of the 11 unclear studies, 5 (45%) abstracts were partially supportive and 6 (55%) abstracts were positive. Of the 38 negative studies, 16 (42%) abstracts were negative, 9 (24%) abstracts were partially supportive, and 13 (34%) abstracts were positive. Thus, 22 out of 38 (58%) negative studies had a (partially) positive focus.

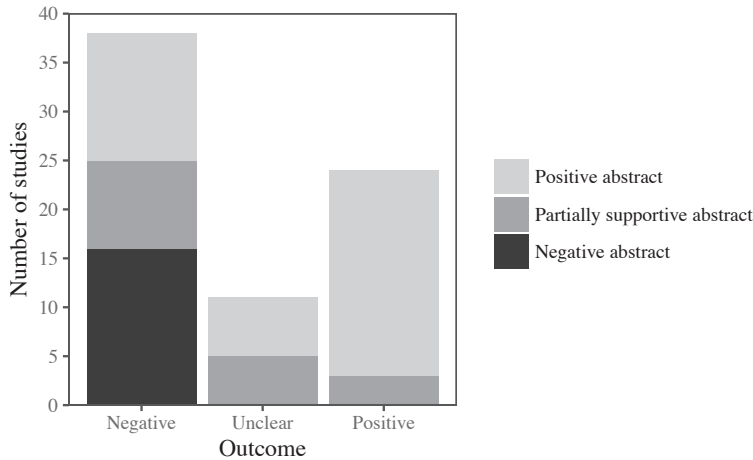


Figure 6.2: Abstract coding by study outcome. The categories on the x-axis represent the outcome of the study, while the different sections of the bars indicate the abstract coding.

Effect of focus on citation

Figure 6.3 shows the distribution of citations (outer circle) by presence of a positive focus (inner circle) in negative studies. Studies without a positive focus, which comprised 42% of all negative studies, received 97 (47%) out of 205 within-network citations and 679 (32%) out of 2113 Web of Science citations to negative studies. Studies with a partially positive focus (24% of all studies) received 28 (14%) within-network citations and 366 (17%) Web of Science citations, while studies with a positive focus (34% of all studies) received 80 (39%) within-network citations and 1068 (51%) Web of Science citations.

On average, a negative study without a positive focus received 6.1 (SD=9.5) citations within the network, while a study with a partially positive focus received 3.1 (SD=5.9) citations and a study with a positive focus received 6.7 (SD=11.3) citations. For Web of Science, a study without a positive focus received 42.4 (SD=44.8) citations on average, while a study with a partially positive focus received 40.7 (SD=40.5) citations and a study with a positive focus received 82.2 (SD=106.6) citations.

Citation patterns by study category

Within the network, both positive and negative studies showed preferential citation of positive studies. Although only 33% of all studies were positive, 55% of citations made by positive studies were to other positive studies, as were 45% of citations made by negative studies. Only negative studies without a positive focus (22% of all studies) additionally showed increased citation of other negative studies without a positive focus, allocating 30% of citations to these studies.

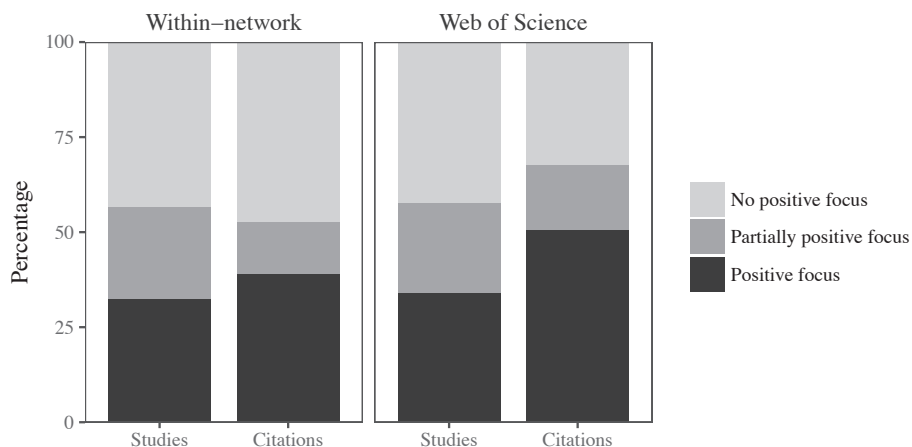


Figure 6.3: Percentage of studies of each type (positive focus, partially positive focus, and no positive focus) and the percentage of citations received by each type of study for within-network citations and Web of Science citations.

Sensitivity analyses

Analyses examining the percentage of subsequent studies citing a study (within-network), the yearly Web of Science citation rate, or the median number of citations (rather than the mean) yielded similar results as our main analyses (data not shown).

When we recoded studies based upon the smallest p-value rather than the average p-value, 10 negative studies and 2 unclear studies became positive. Of the smallest p-values from these 12 studies, 2 were between 0.04 and 0.05, 5 were between 0.01 and 0.05, 4 were less than 0.01, and 1 was only given as <0.05 . After recoding, 36 studies were positive, 28 studies were negative, and 9 studies were unclear. The prevalence of a (partially) positive focus in the remaining negative studies decreased from 58% to 43% (12 out of 28). Recoding did not markedly affect citation patterns (data not shown).

Discussion

Principal findings

We examined citation patterns within the literature on 5-HTTLPR, life stress, and depression. In line with previous research [45, 313], we found that positive studies received more citations than negative studies. This effect was present both within the network of primary studies and within the broader literature (as represented by Web of Science citations), but it was more pronounced within the broader literature. This more pronounced

difference appeared to be largely driven by the study of Caspi and colleagues (2003) [304], which was cited especially frequently, illustrating how such a premier finding may continue to exert considerable influence even as other studies accumulate. Excluding this study reduced, but did not eliminate, citation differences between positive and negative studies.

Furthermore, we found that a (partially) positive focus was present in the abstract of over half of the negative studies. Consequently, although the majority of studies (52%) were negative, these appeared to form a fairly small minority (22%), judging by the abstracts. A positive focus did not affect citation rates within the network, but it increased citation rates within the broader literature.

This suggests that authors of other primary studies are not affected by a positive focus in abstracts. However, upon examining within-network citations to negative studies, we found that studies without a positive focus were overwhelmingly cited as negative (95%), while studies with a positive focus were usually cited as positive (56%) or partially supportive (38%), and only rarely as negative (6%). Thus, the positive focus was still propagated through these citations. Studies with a partially positive focus were actually cited less frequently than studies without a positive focus, particularly within the network. This may be because these studies, which often focused on three-way interactions, appear less relevant to the authors of primary studies on the two-way interaction itself.

Our results resemble those found previously for the literature on 5-HTTLPR and amygdala activation [315], although citation bias toward positive studies and in particular positive abstracts was more pronounced in the amygdala activation literature. This difference may be due to the controversy surrounding gene-environment interactions: both opponents and proponents may be more likely to cite negative studies when there is controversy, the former to cast doubt upon the value of gene-environment research, the latter to point out potential flaws in these negative studies. However, when we examined early citations (prior to 2010) and late citations separately, there was little evidence that citation bias toward positive studies has changed since the publication of critical meta-analyses in 2009 [307, 308], although there did seem to be a decrease in citation bias toward negative studies with a positive focus.

In this study, we extended the concept of spin, which originated within the clinical trial literature, to observational studies. Given the differences between observational studies and clinical trials, we use the term “positive focus” instead of spin. Unlike clinical trials, which are usually narrowly focused on the efficacy of an intervention, observational studies tend to have more wide-ranging topics and often lack a clearly defined, a priori primary outcome.

In this study, we specifically examined whether abstracts suggested that the results supported the 5-HTTLPR, life stress, and depression hypothesis, although some studies had other (primary) hypotheses (e.g., three-way interactions). However, all studies were

clearly inspired by Caspi and colleagues (2003) [304] and have a bearing on the original finding. As discussed by Kapur and colleagues (2012) [67], novel findings in biological psychiatry often become surrounded by a penumbra of subsequent studies with a multiplicity of measures and significant findings that are, at best, “approximate replications”. A finding thus appears to be supported, even though it has not been decisively replicated (or refuted) and even though some supportive findings may have been accompanied by negative findings on a more precise replication of the original finding. We therefore deemed it important to specifically investigate how papers report on their findings with respect to the original finding by Caspi and colleagues (2003) [304].

Duncan and Keller (2011) have previously shown that negative replications of $G \times E$ findings were often published alongside positive findings [309]. This tendency, which is distinct from, although related to a focus on positive findings in the abstract, further illustrates that authors are inclined to present a positive message. The tendency for the hypothesis to expand, as reflected in the study of three- or even four-way interactions between 5-HTTLPR, life stress, and gender, other genes or environmental factors, may also be rooted, in part, in the search for positive findings.

There is a consensus that negative results are difficult to publish, which is supported by the finding that the sample size of purely negative $G \times E$ studies was six times greater than that of positive studies [309]. Although cohort studies have not found a greater journal acceptance rate for positive papers compared to negative papers [320], these studies often examined high-impact general medical journals, and authors may not submit negative studies that they judge to have little chance of acceptance to such journals. The perception that negative studies are unpublishable, as well as the conviction that the effect is real, may lead researchers to use motivated reasoning to justify presenting their findings in a positive light (without necessarily any conscious intentions of doing so) [321].

Strengths and limitations

One of the strengths of our study is our examination of positive focus in abstracts and its influence on citation patterns, as the decision to cite a study and the manner of citation may be based on the abstract only. An additional strength is that we examined citations within the network of primary studies as well as in the broader literature, since authors of other primary studies are likely to have different citation motives than authors writing on a broader or different topic. We also corrected for differences in opportunity to be cited by looking at yearly rates and the percentage of studies citing a given study, which yielded similar results.

Finally, we performed a sensitivity analysis based upon the smallest p-values, when studies had multiple relevant stressors or outcomes. Using only the smallest p-value accounts for studies in which the analysis considered most important by the authors is statistically significant, whereas other analyses are not. This lenient approach does not account for

multiple testing, although many p-values were not highly significant (only 4 out of 12 were smaller than 0.01). While this approach increased the proportion of positive studies, 43% of the remaining negative studies still had a (partially) positive focus in the abstract, and citation patterns were comparable, showing that the overall pattern remains the same even as some individual studies shift categories.

A limitation of our study is that the inter-rater agreement for coding study outcomes was only moderate. Although some disagreements were easily resolved, others reflect the opacity of some of the studies we included, which often included a multitude of stressors, outcomes, analyses, and p-values. Unfortunately, the G×E field is characterized by a proliferation of approaches, hampering easy interpretability and comparability. Pre-specification of a primary outcome and analytical approach, such as proposed in the protocol of a collaborative meta-analysis [322], may help curb this proliferation and yield clear results.

A second limitation is that we did not incorporate meta-analyses, although citations are probably diverted from primary studies to meta-analyses once these are published. However, both the negative and positive meta-analyses in this field [306, 307, 308] have been highly cited, suggesting that inclusion of meta-analyses would not undo the preferential citation of positive studies. Finally, we did not assess study quality. Arguably, high-quality studies should receive more citations, and it is possible, although not very likely [309], that positive studies were of higher quality than negative studies.

Conclusions

Although we have examined a specific, highly prominent finding, selective focus on positive findings and citation bias are unlikely to be isolated problems, limited to this particular example. On the contrary, like other biases, they are probably widespread in many scientific disciplines. Our research therefore illustrates evidence-base-distorting mechanisms that may be at work in other areas as well.

Consequently, our findings have broad implications. The frequent presence of positive conclusions in the abstracts of negative studies suggests that readers should endeavor to read the full study and personally assess its results whenever possible. Furthermore, researchers are well-advised to perform an independent search to obtain all relevant studies, as combing through reference lists may yield a disproportionate number of positive studies. Researchers should also be encouraged to cite all relevant studies, and peer reviewers may play a part in ensuring that relevant negative studies are cited and that abstracts provide an accurate and complete representation of the results.

Our study is not a meta-analysis and is not intended to provide a definitive answer to the question of whether 5-HTTLPR moderates the association between life stress and the development of depression. Instead, we examined whether there is a tendency

within this literature to preferentially cite some studies over others. We have shown that positive studies receive a disproportionate amount of attention and that negative studies are frequently presented as positive, which distorts the apparent evidence base. In the $G \times E$ field, where individual studies often include a variety of analyses and p-values, it is difficult for any reader to tell the forest from the trees. The presence of a selective focus on positive findings and citation bias further compounds this difficulty by hiding published negative results from view and rendering the “forest” greener than it truly is.

Appendix

Table 6.1: *Supplemental table of studies*

Study	Stressor	Depression measure	P-value	Outcome	Abstract
Aguilera (2009)	Child sexual abuse	SCL-90-R subscale	<0.0001	P	P
Antypa (2010)	Child emotional abuse	MDQ	0.59	N	N
Araya (2009)	SLEs	SDQ	0.23	N	N
Aslund (2009)	Childhood maltreatment	DSRS (symptoms) DSRS (diagnosis)	0.016 0.015	P	PS
Beaver (2012)	Perceived stress	CES-D	0.001	P	PS
Benjet (2010)	Relational peer victimization	CDI	0.03	P	P
Brown (2013)	Childhood maltreatment Life events	SCAN (prospective) SCAN (retrospective) SCAN	0.3916 0.0017 0.9745	N	N
Brummett (2008)	Caregiving stress Father's low education	CES-D ODS	n/a n/a	U	PS
Bull (2009) [E]	Interferon- α treatment	BDI/ZSDS	0.03	P	P
Carli (2011)	Childhood abuse	HDRS	0.0026 (O)	N	N
Caspi (2003)	SLEs Childhood maltreatment	Depression symptoms Depression diagnosis Informant report Depression diagnosis	0.02 0.056 <0.01 0.05	P	P
Cervilla (2006)	Threatening life events	CIDI	0.04	P	P
Chipman (2007)	SLEs Childhood adversity Family stress Persistent family adversity	GoDS SMFQ	0.584 0.613 0.903, 0.812 0.215, 0.007 (O)	N	N
Chorbov	Traumatic	C-SSAGA	<0.0001 (O)	N	N

continued

Table 6.1: *Supplemental table of studies*

Study	Stressor	Depression measure	P-value	Outcome	Abstract
(2007)	life events				
Cicchetti (2007)	Childhood maltreatment	DISC YSR	NS NS	N	P
Cicchetti (2011)	Childhood maltreatment	CDI & TRF	0.276	N	PS
Comasco (2011a)	Season of delivery	EPDS	n/a	U	PS
Comasco (2011b)	SLEs	EPDS	n/a	U	P
Conway (2010)	Chronic family stress	BDI SCID	NS	N	N
Coventry (2010)	SLEs	SSAGA DSSI	NS NS or <0.05 (O)	N	N
Dick (2007)	SLEs	SSAGA	n/a	U	P
Eley (2004)	Environmental risk	SMFQ	0.09	N	P
Fergusson (2011)	Childhood adversity SLEs Adult adversity	CIDI	NS or <0.05 (O) NS NS	N	N
Gibb (2009)	Maternal criticism	CDI	n/a	U	PS
Gillespie (2005)	Personal SLEs Network SLEs Personal SLEs Network SLEs	SSAGA SCL-90 + DSSI SSAGA SCL-90 + DSSI	0.43 0.66 0.74 0.15	N	N
Goldman (2010)	Lifetime trauma Major life events	CES-D	0.025 0.294	N	PS
Grabe (2005)	Unemployment Chronic disease	Von Zerksen scale	n/a n/a	U	PS
Grabe (2012)	Child abuse Adult trauma	BDI	0.2218, 0.1680 0.4030, 0.3858	N	PS
Grassi (2010)	SLEs	HAD-D	0.18	N	N
Hammen (2010)	Acute life events	BDI	NS	U	PS

continued

Table 6.1: *Supplemental table of studies*

Study	Stressor	Depression measure	P-value	Outcome	Abstract
	Chronic family stress				
Hankin (2011)	Negative life events	CDI	>0.05	N	P
Jacobs (2006)	Life events	SCL-90 subscale	0.04	P	P
Jenness (2011)	Episodic stressors	CDI	0.02	N	PS
	Chronic family stress		0.88		
Kaufman (2004)	Childhood maltreatment	MFQ	0.01	P	P
Kaufman (2006)	Childhood maltreatment	MFQ	<0.03	P	PS
Kendler (2005)	SLEs	Depression diagnosis	0.04	P	P
Kilpatrick (2007)	Hurricane exposure	SCID-IV	n/a	U	P
Kim (2007)	SLEs	GMS (B3)	n/a	U	P
Kim (2009)	Somatic disorders	GMS (B3)	0.048	P	P
Kohen (2008) [E]	Stroke	GeDS	0.045	P	P
Kraus (2007) [E]	Interferon- α treatment	HAD-D	0.413	N	N
Kumsta (2010)	Institutional deprivation	Rutter scales, SDQ, CAPA	0.14	N	P
Laucht (2009)	SLEs Family adversity	SCID BDI	p<0.05 (O) NS	N	N
Lazary (2008)	Threatening life events	ZSDS	0.0049	P	P
Lenze (2005) [E]	Hip fracture	PRIME-MD HDRS	0.026 p<0.001	P	P
Lotrich (2007) [E]	Interferon- α treatment	BDI SCID	NS p<0.05	U	P
Mitchell	Low education	CIDI-SF	NS	N	P

continued

Table 6.1: *Supplemental table of studies*

Study	Stressor	Depression measure	P-value	Outcome	Abstract
(2011)					
Mossner (2001) [E]	Parkinson's disease	HDRS	0.02	P	P
Nakatani (2005) [E]	Acute myocardial infarction	ZSDS	0.01	P	P
Otte (2007) [E]	Coronary disease	DIS	0.04	P	P
Petersen (2012)	SLEs	YSR + CBCL	0.035	P	P
Phillips-Bute (2008) [E]	Coronary artery bypass surgery	CES-D	0.70, 0.02	N	N
Power (2010)	Recent life events	MINI CES-D	0.08, 0.70 0.47, 0.82	N	N
Quinn (2012)	Early life stress	MINI	NS	N	PS
Ramasubbu (2006) [E]	Stroke	SCID	0.025	P	P
Ressler (2010)	Childhood trauma	BDI SCID	NS 0.016	N	P
Ritchie (2009)	Excessive problem sharing by parents Poverty Other childhood trauma	MINI + CES-D + antidepressant use (composite)	0.027 (O) 0.025 (O) NS	N	PS
Scheid (2007)	Life stressors	CES-D	>0.24 ($\times 6$), 0.04	N	P
Scheid (2011)	Life stressors	CES-D	NS	N	PS
Sen (2010) [E]	Medical internship	PHQ-9	0.002	P	P
Sjöberg (2006)	Residence type Parental separation Traumatic family conflict Composite	DSRS	0.106 0.106 0.01 0.004	N	PS
Stefanis	Military	SCL-90-R	0.11	N	P

continued

Table 6.1: *Supplemental table of studies*

Study	Stressor	Depression measure	P-value	Outcome	Abstract
(2011) [E]	conscription				
Sugden (2010)	Bullying	CBCL + TRF	0.013, 0.508	N	P
Surtees (2006)	Adverse life events	HLEQ	NS	N	N
Taylor (2006)	Early life stress Negative life events	BDI	<0.008 <0.024	P	P
Tsuboi (2011)	Perceived stress	CES-D	<0.05	P	P
Uher (2011)	Childhood maltreatment	DIS	0.003, 0.2312	N	P
Wichers (2008)	Childhood trauma	SCL-90-R subscale SCID	0.4 0.4	N	PS
Wilhelm (2006)	Adverse life events	DIS & CIDI	0.036, 0.435	N	P
Wilhelm (2012) [E]	Diabetes	Clinical interview PHQ-9 K10	0.81 NS 0.047	N	P
Zalsman (2006)	SLEs Child abuse	HDRS BDI HDRS BDI	0.04 0.51 0.20 0.19	N	P
Zhang (2009a)	Negative life events	DIGS	0.005 - 0.006 (O)	N	P
Zhang (2009b) [E]	Parkinson's disease	CES-D	0.804	N	N

[E] indicates a study with an exposed-only sample.

List of acronyms for stressors and depression measures: BDI: Beck Depression Inventory; CAPA: Child and Adolescent Psychiatric Assessment; CBCL: Child Behavior Checklist; CDI: Children's Depression Inventory; CES-D: Center for Epidemiological Studies Depression Scale; CIDI: Composite International Diagnostic Interview; CIDI-SF: Composite International Diagnostic Interview – Short Form; C-SSAGA: Child Semi-Structured Assessment for the Genetics of Alcoholism; DIGS: Diagnostic Interview for Genetic Studies; DIS: Diagnostic Interview Schedule; DISC: Diagnostic Interview Schedule for Children; DSRs: Depression Self-Rating Scale; DSSI: Delusions-Symptoms-States Inventory; EPDS: Edinburgh Postnatal Depression Scale; GeDS: Geriatric Depression Scale; GoDS: Goldberg Depression Scale; GMS: Geriatric

Mental State schedule; HAD-D: Hospital Anxiety and Depression scale – depression subscale; HDRS: Hamilton Depression Rating Scale; HLEQ: Health and Life Experiences Questionnaire; K10: Kessler Psychological Distress Scale; MDQ: Major Depression Questionnaire; MFQ: Mood and Feelings Questionnaire; MINI: Mini International Neuropsychiatric Interview; ODS: Obvious Depression Scale; PHQ-9: Patient Health Questionnaire; PRIME-MD: Primary Care Evaluation of Mental Disorders; SCAN: Schedules for Clinical Assessment in Neuropsychiatry; SCID: Structured Clinical Interview for DSM Disorders; SCL-90-R: Symptom Checklist 90 Revised; SDQ: Strengths and Difficulties Questionnaire; SLEs: stressful life events; SMFQ: Short Mood and Feelings Questionnaire; SSAGA: Semi-Structured Assessment for the Genetics of Alcoholism; TRF: Teacher Report Form; YSR: Youth Self-Report; ZSDS: Zung Self-rating Depression Scale.

P-values: *P-values apply to the specific combination of stressor and depression measure listed. “NS” stands for non-significant, in cases where an exact p-value was unavailable. n/a indicates those studies for which the p-value for a two-way interaction could not be found and which were subsequently coded as unclear. Multiple p-values may be given for a single stressor-depression measure combination in case of, for example, multiple time points. For significant p-values, (O) indicates a significant interaction in the opposite direction from that expected. For non-significant p-values, the direction of the effect (expected or opposite) is not indicated.*

Outcome and abstract coding acronyms: *P: Positive; N: Negative; U: Unclear; PS: Partially supportive.*

