



de Vries, Y. A., Roest, A. M., Franzen, M., Munafo, M., & Bastiaansen, J. A. (2016). Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene (5-HTTLPR), life stress and depression. *Psychological Medicine*, 46(14), 2971-2979. <https://doi.org/10.1017/S0033291716000805>

Peer reviewed version

Link to published version (if available):
[10.1017/S0033291716000805](https://doi.org/10.1017/S0033291716000805)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via Cambridge University Press at <http://dx.doi.org/10.1017/S0033291716000805>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/pure/user-guides/explore-bristol-research/ebr-terms/>

1 **Citation bias and selective focus on positive findings in the literature on 5-**
2 **HTTLPR, life stress, and depression**

3

4 Ymkje Anna de Vries¹, Annelieke M. Roest¹, Minita Franzen², Marcus R. Munafò^{3,4},
5 Jojanneke A. Bastiaansen¹

6

7 ¹ Interdisciplinary Center Psychopathology and Emotion regulation, Department of
8 Psychiatry, University of Groningen, University Medical Center Groningen, Groningen,
9 The Netherlands.

10 ² Department of Psychology, University of Groningen, Groningen, the Netherlands

11 ³ MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, UK

12 ⁴ UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology,
13 University of Bristol, Bristol, UK

14

15

16 Corresponding author:

17 Ymkje Anna de Vries, MSc

18 Department of Psychiatry

19 University Medical Center Groningen

20 Hanzeplein 1, 9713 GZ Groningen, The Netherlands

21 Email: y.a.de.vries@umcg.nl

22 Phone: +31 503615730

23

24 Word count: 4,500

25

1 Financial support:

2 This research received no specific grant from any funding agency, commercial or not-
3 for-profit sectors. Marcus R. Munafò is a member of the United Kingdom Centre for
4 Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence.
5 Funding from British Heart Foundation, Cancer Research UK, Economic and Social
6 Research Council, Medical Research Council, and the National Institute for Health
7 Research, under the auspices of the UK Clinical Research Collaboration, is gratefully
8 acknowledged.

9

10 Conflicts of interest:

11 None.

12

1 **Abstract**

2

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

7

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

12

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

20

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

1 **Introduction**

2 Major depressive disorder (MDD) is a complex illness, caused by a combination of
3 genetic and environmental risk factors (Sullivan et al. 2000). One of the most robust risk
4 factors for MDD is the experience of a stressor, such as a stressful life event or childhood
5 abuse (Hammen 2005). However, many who experience such a stressor do not develop
6 depression. This individual variability has been suggested to be due, at least in part, to
7 genetic variation (Caspi et al. 2010).

8 In 2003, Caspi and colleagues reported that a polymorphism in the serotonin
9 transporter gene (5-HTTLPR) moderates the relationship between life stress and
10 depression: while carriers of at least one short (S) allele had a similar risk of depression
11 as people homozygous for the long (L) allele in the absence of stress, S carriers were up
12 to twice as likely to develop depression after stressful life events or childhood abuse
13 (Caspi et al. 2003). This study has since been highly cited (>3,800 times, Web of Science,
14 October 2015) and has become the seminal finding within the burgeoning field of gene-
15 environment interactions (G×E). However, this finding also remains highly contentious.
16 Even meta-analyses on this topic contradict each other, with some finding evidence of an
17 effect (Sharpley et al. 2014; Karg et al. 2011), while others do not (Risch et al. 2009;
18 Munafò et al. 2009).

19 Many issues complicate the interpretation of G×E findings and replications, such
20 as publication bias (Duncan & Keller 2011) and analytical flexibility (Simmons et al.
21 2011; Zammit et al. 2010), which increases the chance of false-positives due to the
22 multitude of analyses performed (Heininga et al. 2015). The likelihood of false-positives
23 is further increased by low power and by the low prior probability of associations in
24 candidate gene studies (Duncan & Keller 2011). Although replication has been
25 suggested as the solution to false-positive findings (de Jonge et al. 2011), many G×E

1 studies are imprecise replications of the original finding, and a loose definition of
2 replication may still permit propagation of false-positives (Sullivan 2007).

3 Additionally, researchers may emphasize positive findings while downplaying
4 negative findings. Within the randomized controlled trial literature, such reporting
5 strategies, whether intentional or unintentional, that focus on positive (secondary)
6 findings (in spite of non-significant results for the primary outcome) and that may
7 distort the interpretation of results, are defined as ‘spin’ (Boutron et al. 2010; Roest et al.
8 2015). A focus on positive findings has also been demonstrated in observational studies
9 (Park et al. 2014). As a consequence, the published literature on a topic may appear
10 more convincing than is justified by the strength of the evidence.

11 Selective citation may also affect the quality of the evidence base (Greenberg
12 2009). Statistically significant (positive) studies are cited more frequently than non-
13 significant (negative) studies (Nieminen et al. 2007; Jannot et al. 2013; Kjaergard &
14 Gluud 2002; Etter & Stapleton 2009), which may render non-supportive studies
15 relatively invisible. Citation bias and focus on positive findings can also work
16 synergistically to hide negative results from view. A previous examination of citation
17 patterns on a related topic, that of 5-HTTLPR and amygdala activation (Bastiaansen et al.
18 2015), showed that negative studies that had been spun were cited at a similar rate as
19 positive studies, while negative studies that had not been spun received almost no
20 citations. The resulting invisibility of negative findings may create the impression that
21 this effect has been proven beyond doubt, although meta-analyses have questioned its
22 robustness (Bastiaansen et al. 2014; Murphy et al. 2013).

23 In the current study, we aimed to determine whether citation bias and selective
24 focus on positive findings are also present in the literature on 5-HTTLPR, life stress, and
25 depression. Achieving a better understanding of the etiology of depression is of vital

1 importance to psychiatry, given the high burden of depression (Whiteford et al. 2013).
2 Distortion of the evidence base could mislead researchers and clinicians and thus pose a
3 major obstacle to this goal.

4

5 **Methods**

6 *Study selection*

7 To establish the network of primary studies, we searched PubMed for the most recent
8 meta-analysis on 5-HTTLPR, stress, and depression (Sharpley et al. 2014), which
9 included 81 studies. For each study, we determined the outcome for the effect of interest
10 (i.e., 5-HTTLPR x stress). We included studies with continuous outcomes (e.g., score on a
11 depression questionnaire), as well as studies with binary outcomes (depression
12 diagnosis). We excluded studies in which the outcome was clearly a different construct
13 than depression (e.g., cognitive dysfunction). Studies were included regardless of
14 whether the 5-HTTLPR x stress interaction effect on depression was the primary
15 outcome. No exclusion criteria were applied for stressors, which were very diverse.

16

17 *Coding study outcomes*

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

1 P-values were extracted according to a hierarchical decision tree. We first
2 determined whether the design of the study was “exposed-only”. In these studies, the
3 entire sample was exposed to a stressor, such as a somatic illness. The effect of interest,
4 in this case, is not an interaction but the main effect of 5-HTTLPR. Hence, we extracted
5 the p-value associated with the main effect for these studies. For all other studies, we
6 determined whether a p-value was reported for a two-way interaction between 5-
7 HTTLPR and stress, consistent with Caspi and colleagues (2003).

8 If multiple relevant, independent outcomes or stressors were included in a study,
9 we extracted all p-values. Following Sharpley and colleagues (2014), we averaged these
10 p-values to arrive at a conclusion. If multiple non-independent outcomes were given
11 (e.g., a continuous symptom scale and a dichotomized version thereof), we only included
12 the continuous outcome. When studies provided p-values for both biallelic and triallelic
13 genotyping, we erred towards coding a study as positive by selecting the smallest p-
14 value, as it is unclear which genotyping approach should be preferred (Martin et al.
15 2007; Hu et al. 2005; Wendland et al. 2006). If both unadjusted and adjusted analyses
16 were given, we also used the smallest p-value. We preferentially extracted the p-value of
17 an overall test of interaction; however, if only post-hoc comparisons were available, we
18 extracted the p-value associated with the SS vs LL homozygotes comparison.

19

20 *Coding study abstracts*

21 Two independent raters (YAdV and MF) coded the abstract of each study, and
22 discrepancies were resolved by discussion with AR and JAB. Abstracts were
23 preferentially coded based upon their conclusions, but if these did not provide a clear
24 statement, we used the results section of the abstract. In coding abstracts, we were
25 interested in the way abstracts reported on how their findings reflected on the original

1 result by Caspi and colleagues (2003). Abstracts were coded as positive if a claim was
2 made that the results supported the existence and/or importance of the 5-HTTLPR x
3 stress interaction. Abstracts were coded as partially supportive if a positive claim was
4 made that was not directly related to the 5-HTTLPR x stress interaction (e.g., positive
5 findings for a three-way interaction) or if the abstract mentioned findings for multiple
6 outcomes or stressors and not all were positive. Abstracts that did not make a positive
7 claim or that made an explicitly negative claim were coded as negative. If the abstract
8 did not report on the effect of interest, the study was excluded (2 studies).

9

10 *Citation outcomes*

11 We examined citations both within the network of primary studies and outside of the
12 network in the broader literature (Bastiaansen et al. 2015). To examine within-network
13 citations, we constructed a citation grid and marked for each study by which of the other
14 included studies it was cited. Total citation counts for each study were calculated from
15 the grid. To examine out-of-network citations, we looked up the citation counts for each
16 study on Web of Science (Core Collection, October 2015). To create non-overlapping
17 outcomes, we pruned the within-network citations from the Web of Science citations.
18 While within-network citations represent citations by other experts working within the
19 5-HTTLPR x stress field, Web of Science citations also include citations by researchers
20 not directly involved in this area.

21

22 *Analyses*

23 For our citation analysis, we first compared the citations received by studies with
24 positive, negative or unclear outcomes (irrespective of abstract coding). The sum of
25 citations was calculated and the percentage of all citations received by studies with a

1 given outcome was determined. In examining within-network citations, we excluded the
2 most recent study, as it could not have been cited within the network. We also examined
3 the study by Caspi and colleagues (2003) separately, as we expected it to receive many
4 citations.

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

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

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

1 Since the included studies form the total population of studies on the effect of
2 interest, we used descriptive analyses rather than statistical tests (Bastiaansen et al.
3 2015), which are designed to generalize from a sample to a hypothetical larger
4 population.

5

6 **Results**

7 *Coding of studies and abstracts*

8 We excluded ten of the 81 studies in Sharpley and colleagues (2014): eight studies were
9 excluded because the outcome was not depression-related, no stressor was included, or
10 the entire sample was depressed, one study was excluded because the abstract did not
11 report on 5-HTTLPR, and one study was excluded because the abstract did not report on
12 the depression outcome. Furthermore, we included two additional studies that had been
13 excluded from the meta-analysis because the sample was a subset of those included in a
14 later study (see flow chart in Supplementary Material). Consequently, we included 73
15 studies, of which 24 studies were coded as positive, 38 studies as negative, and 11
16 studies as unclear in terms of outcome. Of the 11 unclear studies, four studies were
17 coded as unclear because of the inclusion of three-way interactions in the model (e.g.,
18 with social support), while another study was coded as unclear because the 5-HTTLPR x
19 stress interaction was only tested in males and females separately. Four studies were
20 coded as unclear because the 5-HTTLPR x stress interaction was not tested (e.g., only
21 the main effect of 5-HTTLPR in the different stress groups was tested). Finally, two
22 studies were coded as unclear because we could not determine whether the (averaged)
23 p-value was <0.05, as one p-value was given as “non-significant” while another was
24 <0.05. Inter-rater agreement was moderate ($\kappa=0.49$). Our agreement with Sharpley
25 and colleagues (2014) was good: within the subset of studies included in both Sharpley

1 and colleagues (2014) and our own paper and that we coded as positive or negative
2 (rather than unclear), the percentage of positive studies was 38% (23 out of 60) by both
3 our coding and Sharpley's coding; coding was identical for 54 out of 60 (90%) papers.

4 Of the 73 studies, we coded 40 abstracts as positive, 16 abstracts as negative, and
5 17 abstracts as partially supportive. Inter-rater agreement for abstract coding was good
6 ($\kappa=0.71$). A full table of studies with characteristics and coding is given in the
7 Supplementary Material.

8

9 *Citations by study outcome*

10 Figure 1 shows the percentage of citations to positive, negative, and unclear studies
11 (outer circle) compared to the percentage of studies of each type (inner circle).

12 The total number of citations was 488 within the network and 9160 on Web of
13 Science. Positive studies, comprising 33% of all studies, received 236 (48%) within-
14 network citations and 6187 (68%) Web of Science citations. Negative studies (52% of all
15 studies) received 205 (42%) within-network citations and 2113 (23%) Web of Science
16 citations, while unclear studies (15% of all studies) received 47 (10%) within-network
17 citations and 860 (9%) Web of Science citations. The study by Caspi and colleagues
18 (2003) received a large share of the citations to positive studies, particularly in Web of
19 Science. However, even after exclusion of this study, positive studies still received 40%
20 of within-network and 45% of Web of Science citations, as compared to 48% and 39%,
21 respectively, for negative studies.

22 On average, negative studies received 5.5 (standard deviation (SD)=9.3) within-
23 network citations, while unclear studies received 4.3 (SD=6.2) and positive studies
24 received 9.8 (SD=14.6). Positive studies other than Caspi and colleagues (2003) received
25 7.4 (SD=8.9) within-network citations on average. For Web of Science, negative studies

1 received, on average, 55.6 (SD=72.3) citations, while unclear studies received 78.2
2 (SD=61.6) and positive studies received 257.8 (SD=765.5) citations. Positive studies
3 other than Caspi and colleagues (2003) received 103.8 (SD=132.4) citations on average.

4

5 *[Insert Figure 1 here]*

6

7 *Presence of positive focus in abstracts*

8 Figure 2 depicts the presence of a positive focus within the set of studies. Of the 24
9 positive studies, 21 (88%) abstracts were positive and 3 (13%) abstracts were partially
10 supportive. These partially supportive abstracts focused on gender differences (2
11 abstracts) or on a three-way interaction (1 abstract). Of the 11 unclear studies, 5 (45%)
12 abstracts were partially supportive and 6 (55%) abstracts were positive. Of the 38
13 negative studies, 16 (42%) abstracts were negative, 9 (24%) abstracts were partially
14 supportive, and 13 (34%) abstracts were positive (see Supplemental Table 1 for a list of
15 these 13 studies, with the relevant sentence(s) from the abstract). Thus, 22 out of 38
16 (58%) negative studies had a (partially) positive focus.

17

18 *[Insert Figure 2 here]*

19

20 *Effect of focus on citation*

21 Figure 3 shows the distribution of citations (outer circle) by presence of a positive focus
22 (inner circle) in negative studies. Studies without a positive focus, which comprised 42%
23 of all negative studies, received 97 (47%) out of 205 within-network citations and 679
24 (32%) out of 2113 Web of Science citations to negative studies. Studies with a partially
25 positive focus (24% of all studies) received 28 (14%) within-network citations and 366

1 (17%) Web of Science citations, while studies with a positive focus (34% of all studies)
2 received 80 (39%) within-network citations and 1068 (51%) Web of Science citations.

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

9

10 *[Insert Figure 3 here]*

11

12 *Citation patterns by study category*

13 Within the network, both positive and negative studies showed preferential citation of
14 positive studies. Although only 33% of all studies were positive, 55% of citations made
15 by positive studies were to other positive studies, as were 45% of citations made by
16 negative studies. Only negative studies without a positive focus (22% of all studies)
17 additionally showed increased citation of other negative studies without a positive
18 focus, allocating 30% of citations to these studies (Supplemental Table 2).

19

20 *Sensitivity analyses*

21 Analyses examining the percentage of subsequent studies citing a study (within-
22 network), the yearly Web of Science citation rate, or the median number of citations
23 (rather than the mean) yielded similar results as our main analyses (Supplemental
24 Tables 3 and 4).

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

9

10 **Discussion**

11 We examined citation patterns within the literature on 5-HTTLPR, life stress, and
12 depression. In line with previous research (e.g., Nieminen et al. 2007; Jannot et al. 2013),
13 we found that positive studies received more citations than negative studies. This effect
14 was present both within the network of primary studies and within the broader
15 literature (as represented by Web of Science citations), but it was more pronounced
16 within the broader literature. This more pronounced difference appeared to be largely
17 driven by the study of Caspi and colleagues (2003), which was cited especially
18 frequently, illustrating how such a premier finding may continue to exert considerable
19 influence even as other studies accumulate. Excluding this study reduced, but did not
20 eliminate, citation differences between positive and negative studies.

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

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

10 Our results resemble those found previously for the literature on 5-HTTLPR and
11 amygdala activation (Bastiaansen et al. 2015), although citation bias toward positive
12 studies and in particular positive abstracts was more pronounced in the amygdala
13 activation literature. This difference may be due to the controversy surrounding gene-
14 environment interactions: both opponents and proponents may be more likely to cite
15 negative studies when there is controversy, the former to cast doubt upon the value of
16 gene-environment research, the latter to point out potential flaws in these negative
17 studies. However, when we examined early citations (prior to 2010) and late citations
18 separately, there was little evidence that citation bias toward positive studies has
19 changed since the publication of critical meta-analyses in 2009 (Risch et al. 2009;
20 Munafò et al. 2009), although there did seem to be a decrease in citation bias toward
21 negative studies with a positive focus.

22 In this study, we extended the concept of spin, which originated within the
23 clinical trial literature, to observational studies. Given the differences between
24 observational studies and clinical trials, we use the term “positive focus” instead of spin.
25 Unlike clinical trials, which are usually narrowly focused on the efficacy of an

1 intervention, observational studies tend to have more wide-ranging topics and often
2 lack a clearly defined, *a priori* primary outcome. In this study, we specifically examined
3 whether abstracts suggested that the results supported the 5-HTTLPR, life stress, and
4 depression hypothesis, although some studies had other (primary) hypotheses (e.g.,
5 three-way interactions). However, all studies were clearly inspired by Caspi and
6 colleagues (2003) and have a bearing on the original finding. As discussed by Kapur and
7 colleagues (2012), novel findings in biological psychiatry often become surrounded by a
8 penumbra of subsequent studies with a multiplicity of measures and significant findings
9 that are, at best, “approximate replications”. A finding thus appears to be supported,
10 even though it has not been decisively replicated (or refuted) and even though some
11 supportive findings may have been accompanied by negative findings on a more precise
12 replication of the original finding. We therefore deemed it important to specifically
13 investigate how papers report on their findings with respect to the original finding by
14 Caspi and colleagues (2003).

15 Duncan and Keller (2011) have previously shown that negative replications of
16 G×E findings were often published alongside positive findings. This tendency, which is
17 distinct from, although related to a focus on positive findings in the abstract, further
18 illustrates that authors are inclined to present a positive message. The tendency for the
19 hypothesis to expand, as reflected in the study of three- or even four-way interactions
20 between 5-HTTLPR, life stress, and gender, other genes or environmental factors, may
21 also be rooted, in part, in the search for positive findings. There is a consensus that
22 negative results are difficult to publish, which is supported by the finding that the
23 sample size of purely negative G×E studies was six times greater than that of positive
24 studies (Duncan & Keller 2011). Although cohort studies have not found a greater
25 journal acceptance rate for positive papers compared to negative papers (Song et al.

1 2009), these studies often examined high-impact general medical journals, and authors
2 may not submit negative studies that they judge to have little chance of acceptance to
3 such journals. The perception that negative studies are unpublishable, as well as the
4 conviction that the effect is real, may lead researchers to use motivated reasoning to
5 justify presenting their findings in a positive light (without necessarily any conscious
6 intentions of doing so) (Nosek et al. 2012).

7 One of the strengths of our study is our examination of positive focus in abstracts
8 and its influence on citation patterns, as the decision to cite a study and the manner of
9 citation may be based on the abstract only. An additional strength is that we examined
10 citations within the network of primary studies as well as in the broader literature, since
11 authors of other primary studies are likely to have different citation motives than
12 authors writing on a broader or different topic. We also corrected for differences in
13 opportunity to be cited by looking at yearly rates and the percentage of studies citing a
14 given study, which yielded similar results. Finally, we performed a sensitivity analysis
15 based upon the smallest p-values, when studies had multiple relevant stressors or
16 outcomes. Using the smallest p-value only accounts for studies in which the analysis
17 considered most important by the authors is statistically significant, whereas other
18 analyses are not. This lenient approach does not account for multiple testing, although
19 many p-values were not highly significant (only 4 out of 12 were smaller than 0.01).
20 While this approach increased the proportion of positive studies, 43% of the remaining
21 negative studies still had a (partially) positive focus in the abstract, and citation patterns
22 were comparable, showing that the overall pattern remains the same even as some
23 individual studies shift categories.

24 A limitation of our study is that the inter-rater agreement for coding study
25 outcomes was only moderate. Although some disagreements were easily resolved,

1 others reflect the opacity of some of the studies we included, which often included a
2 multitude of stressors, outcomes, analyses, and p-values. Unfortunately, the G×E field is
3 characterized by a proliferation of approaches, hampering easy interpretability and
4 comparability. Pre-specification of a primary outcome and analytical approach, such as
5 proposed in the protocol of a collaborative meta-analysis (Culverhouse et al. 2013), may
6 help curb this proliferation and yield clear results.

7 A second limitation is that we did not incorporate meta-analyses, although
8 citations are probably diverted from primary studies to meta-analyses once these are
9 published. However, both the negative and positive meta-analyses in this field (Munafò
10 et al. 2009; Risch et al. 2009; Karg et al. 2011) have been highly cited, suggesting that
11 inclusion of meta-analyses would not undo the preferential citation of positive studies.
12 Finally, we did not assess study quality. Arguably, high-quality studies should receive
13 more citations, and it is possible, although not very likely (Duncan & Keller 2011), that
14 positive studies were of higher quality than negative studies.

15 Although we have examined a specific, highly prominent finding, selective focus
16 on positive findings and citation bias are unlikely to be isolated problems, limited to this
17 particular example. On the contrary, like other biases, they are probably widespread in
18 many scientific disciplines. Our research therefore illustrates evidence-base-distorting
19 mechanisms that may be at work in other areas as well. Consequently, our findings have
20 broad implications. The frequent presence of positive conclusions in the abstracts of
21 negative studies suggests that readers should endeavor to read the full study and
22 personally assess its results whenever possible. Furthermore, researchers are well-
23 advised to perform an independent search to obtain all relevant studies, as combing
24 through reference lists may yield a disproportionate number of positive studies.

25 Researchers should also be encouraged to cite all relevant studies, and peer reviewers

1 may play a part in ensuring that relevant negative studies are cited and that abstracts
2 provide an accurate and complete representation of the results.

3 Our study is not a meta-analysis and is not intended to provide a definitive
4 answer to the question of whether 5-HTTLPR moderates the association between life
5 stress and the development of depression. Instead, we examined whether there is a
6 tendency within this literature to preferentially cite some studies over others. We have
7 shown that positive studies receive a disproportionate amount of attention and that
8 negative studies are frequently presented as positive, which distorts the apparent
9 evidence base. In the G×E field, where individual studies often include a variety of
10 analyses and p-values, it is difficult for any reader to tell the forest from the trees. The
11 presence of a selective focus on positive findings and citation bias further compounds
12 this difficulty by hiding published negative results from view and rendering the “forest”
13 greener than it truly is.

14

1 **References**

2

3 **Bastiaansen JA, Servaas MN, Marsman JBC, Ormel J, Nolte IM, Riese H & Aleman A**
4 (2014). Filling the gap: relationship between the serotonin-transporter-linked
5 polymorphic region and amygdala activation. *Psychological Science*, 25(11), 2058–
6 2066.

7 **Bastiaansen JA, de Vries YA & Munafò MR** (2015). Citation distortions in the literature
8 on the serotonin-transporter-linked polymorphic region and amygdala activation.
9 *Biological Psychiatry*, 78(8), E35–36.

10 **Boutron I, Dutton S, Ravaud P & Altman DG** (2010). Reporting and interpretation of
11 randomized controlled trials with statistically nonsignificant results for primary
12 outcomes. *JAMA*, 303(20), 2058–64.

13 **Caspi A, Hariri AR, Holmes A, Uher R & Moffitt TE** (2010). Genetic sensitivity to the
14 environment: the case of the serotonin transporter gene and its implications for
15 studying complex diseases and traits. *American Journal of Psychiatry*, 167, 509–527.

16 **Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington HL, McClay J, Mill J,**
17 **Martin J, Braithwaite AW & Poulton R** (2003). Influence of life stress on
18 depression: moderation by a polymorphism in the 5-HTT gene. *Science*, 301(5631),
19 386–9.

20 **Culverhouse RC, Bowes L, Breslau N, Nurnberger JI, Burmeister M, Fergusson DM,**
21 **Munafò MR, Saccone NL & Bierut LJ** (2013). Protocol for a collaborative meta-
22 analysis of 5-HTTLPR, stress, and depression. *BMC Psychiatry*, 13, 304.

23 **Duncan L & Keller M** (2011). A critical review of the first 10 years of candidate gene-

1 by-environment interaction research in psychiatry. *American Journal of Psychiatry*,
2 168, 1041–1049.

3 **Etter J-F & Stapleton J** (2009). Citations to trials of nicotine replacement therapy were
4 biased toward positive results and high-impact-factor journals. *Journal of Clinical*
5 *Epidemiology*, 62, 831–837.

6 **Greenberg SA** (2009). How citation distortions create unfounded authority: analysis of
7 a citation network. *BMJ*, 339, b2680.

8 **Hammen C** (2005). Stress and depression. *Annual Review of Clinical Psychology*, 1, 293–
9 319.

10 **Heininga VE, Oldehinkel AJ, Veenstra R & Nederhof E** (2015). I just ran a thousand
11 analyses: benefits of multiple testing in understanding equivocal evidence on gene-
12 environment interactions. *PLoS ONE*, 10(5), e0125383.

13 **Hu X, Oroszi G, Chun J, Smith TL, Goldman D & Schuckit MA** (2005). An expanded
14 evaluation of the relationship of four alleles to the level of response to alcohol and
15 the alcoholism risk. *Alcoholism: Clinical & Experimental Research*, 29(1), 8–16.

16 **Jannot A-S, Agoritsas T, Gayet-Ageron A & Perneger T V** (2013). Citation bias
17 favoring statistically significant studies was present in medical research. *Journal of*
18 *Clinical Epidemiology*, 66, 296–301.

19 **de Jonge P, Conradi HJ, Thombs BD, Rosmalen JGM, Burger H & Ormel J** (2011).
20 Prevention of false positive findings in observational studies: registration will not
21 work but replication might. *Journal of Epidemiology and Community Health*, 65(2),
22 95–96.

23 **Karg K, Burmeister M, Shedden K & Sen S** (2011). The serotonin transporter

1 promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited:
2 evidence of genetic moderation. *Archives of General Psychiatry*, 68(5), 444–54.

3 **Kjaergard LL & Gluud C** (2002). Citation bias of hepato-biliary randomized clinical
4 trials. *Journal of Clinical Epidemiology*, 55, 407–410.

5 **Martin J, Cleak J, Willis-Owen SAG, Flint J & Shifman S** (2007). Mapping regulatory
6 variants for the serotonin transporter gene based on allelic expression imbalance.
7 *Molecular Psychiatry*, 12(5), 421–422.

8 **Munafò MR, Durrant C, Lewis G & Flint J** (2009). Gene X environment interactions at
9 the serotonin transporter locus. *Biological psychiatry*, 65(3), 211–9.

10 **Murphy SE, Norbury R, Godlewska BR, Cowen PJ, Mannie ZM, Harmer CJ & Munafò**
11 **MR** (2013). The effect of the serotonin transporter polymorphism (5-HTTLPR) on
12 amygdala function: a meta-analysis. *Molecular Psychiatry*, 18(4), 512–20.

13 **Nieminen P, Rucker G, Miettunen J, Carpenter J & Schumacher M** (2007).
14 Statistically significant papers in psychiatry were cited more often than others.
15 *Journal of Clinical Epidemiology*, 60, 939–946.

16 **Nosek BA, Spies JR & Motyl M** (2012). Scientific Utopia: II. Restructuring incentives and
17 practices to promote truth over publishability. *Perspectives on Psychological Science*,
18 7(6), 615–631.

19 **Park I-U, Peacey MW & Munafò MR** (2014). Modelling the effects of subjective and
20 objective decision making in scientific peer review. *Nature*, 506(7486), 93–6.

21 **Risch N, Herrell R, Lehner T, Liang K, Eaves L, Hoh J, Griem A, Kovacs M, Ott J &**
22 **Merikangas KR** (2009). Interaction between the serotonin transporter gene (5-
23 HTTLPR), stressful life events, and risk of depression. *JAMA*, 301(23), 2462–2471.

- 1 **Roest AM, de Jonge P, Williams CD, de Vries YA, Schoevers RA & Turner EH (2015).**
2 Reporting bias in clinical trials investigating the efficacy of second-generation
3 antidepressants in the treatment of anxiety disorders: a report of 2 meta-analyses.
4 *JAMA Psychiatry*, 72(5), 500 – 510.
- 5 **Sharpley CF, Palanisamy SKA, Glyde NS, Dillingham PW & Agnew LL (2014).** An
6 update on the interaction between the serotonin transporter promoter variant (5-
7 HTTLPR), stress and depression, plus an exploration of non-confirming findings.
8 *Behavioural Brain Research*, 273, 89–105.
- 9 **Simmons JP, Nelson LD & Simonsohn U (2011).** False-positive psychology:
10 undisclosed flexibility in data collection and analysis allows presenting anything as
11 significant. *Psychological Science*, 22(11), 1359–66.
- 12 **Song F, Parekh-Bhurke S, Hooper L, Loke YK, Ryder JJ, Sutton AJ, Hing CB & Harvey**
13 **I (2009).** Extent of publication bias in different categories of research cohorts: a
14 meta-analysis of empirical studies. *BMC Medical Research Methodology*, 9, 79.
- 15 **Sullivan PF (2007).** Spurious genetic associations. *Biological psychiatry*, 61(10), 1121–6.
- 16 **Sullivan PF, Neale MC & Kendler KS (2000).** Genetic epidemiology of major
17 depression: review and meta-analysis. *American Journal of Psychiatry*, 157, 1552–
18 1562.
- 19 **Wendland JR, Martin BJ, Kruse MR, Lesch K-P & Murphy DL (2006).** Simultaneous
20 genotyping of four functional loci of human SLC6A4, with a reappraisal of 5-
21 HTTLPR and rs25531. *Molecular Psychiatry*, 11(3), 224–6.
- 22 **Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson**
23 **FJ, Norman RE, Flaxman AD, Johns N, Burstein R, Murray CJL & Vos T (2013).**

1 Global burden of disease attributable to mental and substance use disorders:
2 findings from the Global Burden of Disease Study 2010. *Lancet*, 382(9904), 1575–
3 1586.

4 **Zammit S, Owen MJ & Lewis G** (2010). Misconceptions about gene-environment
5 interactions in psychiatry. *Evidence-Based Mental Health*, 13(3), 65–68.

6

1 **Figure captions**

2

3 Figure 1: Percentage of citations received by positive, negative, and unclear studies.

4 The inner ring indicates the percentage of studies of each type. The outer ring indicates

5 the percentage of total citations received by studies with positive, negative or unclear

6 outcomes.

7

8 Figure 2: Abstract coding by study outcome.

9 The categories on the x-axis represent the outcome of the study, while the different

10 sections of the bars indicate the abstract coding.

11

12 Figure 3: Percentage of citations received by negative studies without a positive focus,

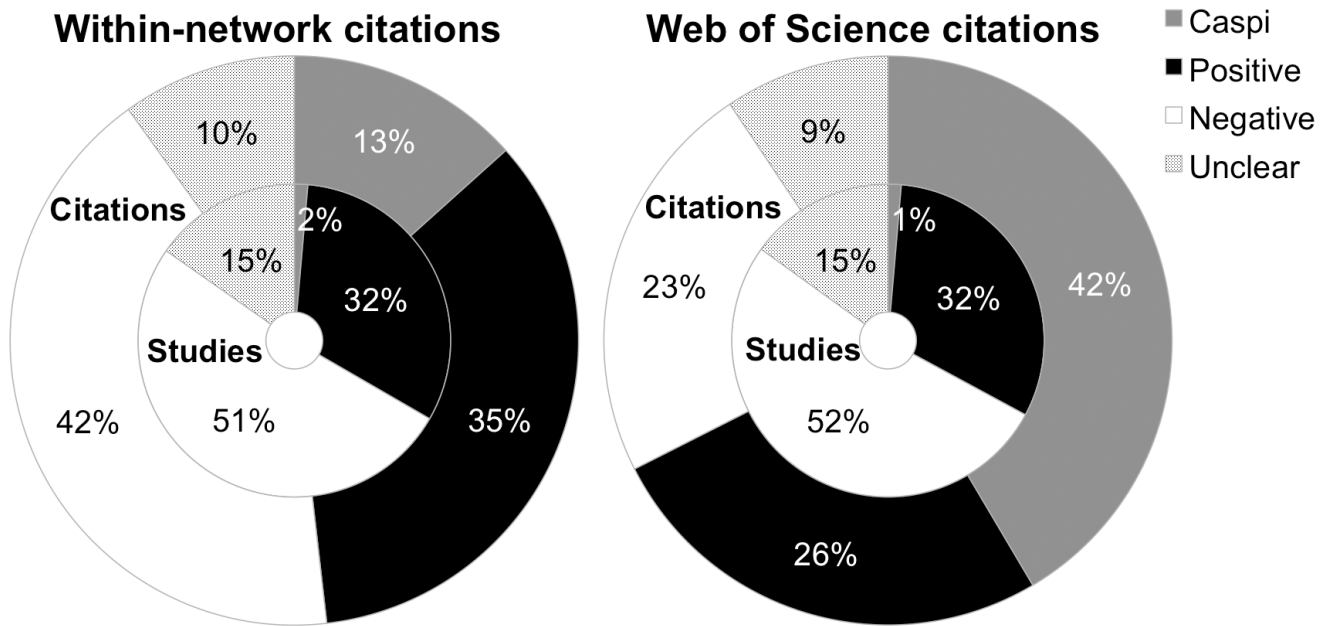
13 with a partially positive focus, and with a positive focus.

14 The inner ring indicates the percentage of studies of each type. The outer ring indicates

15 the percentage of total citations received by studies of each type.

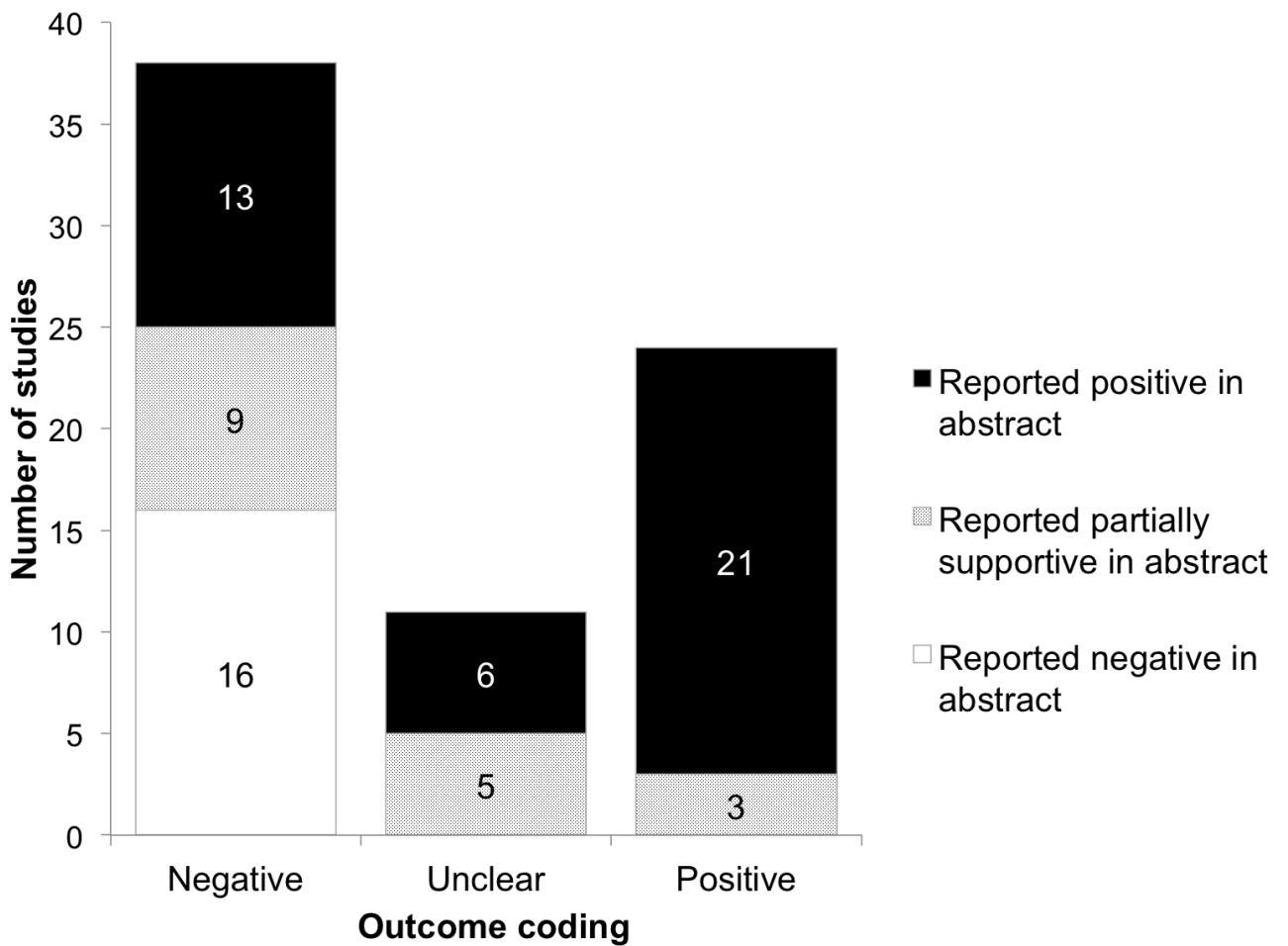
16

1 Figure 1



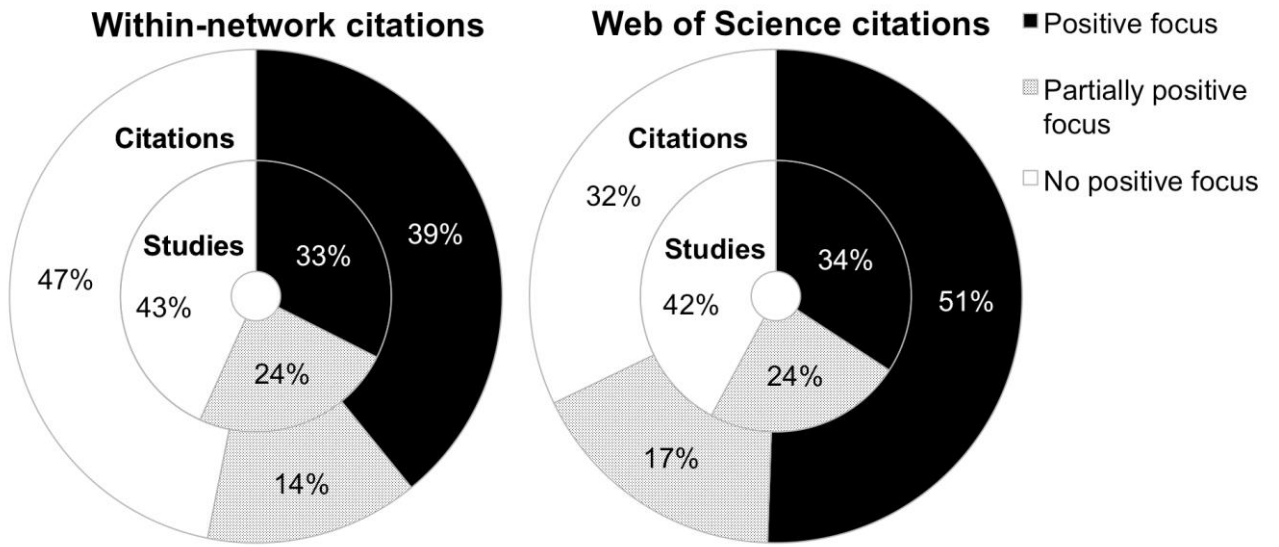
2

3 Figure 2



4

1 Figure 3



2