



# Subtle Scientific Fallacies Undermine the Validity of Neuroendocrinological Research: Do Not Draw Premature Conclusions on the Role of Female Sex Hormones

Michael P. Hengartner\*

Department of Applied Psychology, Zurich University of Applied Sciences (ZHAW), Zurich, Switzerland

Major scientific flaws such as reporting and publication biases are well documented, even though acknowledgment of their importance appears to be lacking in various psychological and medical fields. Subtle and less obvious biases including selective reviews of the literature and empirically unsupported conclusions and recommendations have received even less attention. Using the literature on the association between transition to menopause, hormones and the onset of depression as a guiding example, I outline how such scientific fallacies undermine the validity of neuroendocrinological research. It is shown that in contrast to prominent claims, first, most prospective studies do not support the notion that the menopausal transition relates to increased risk for depression, second, that associations between hormone levels and depression are largely inconsistent and irreproducible, and, third, that the evidence for the efficacy of hormone therapy for the treatment of depression is very weak and at best inconclusive. I conclude that a direct and uniform association between female sex hormones and depression is clearly not supported by the literature and that more attention should be paid to the manifold scientific biases that undermine the validity of findings in psychological and medical research, with a specific focus on the behavioral neurosciences.

**Keywords:** bias, science, research, hormones, depression, menopause, methodology, statistics

## OPEN ACCESS

### Edited by:

Allan V. Kalueff,  
St. Petersburg State University,  
Russia

### Reviewed by:

Lesley J. Rogers,  
University of New England, Australia  
Matthew O. Parker,  
University of Portsmouth, UK

### \*Correspondence:

Michael P. Hengartner  
michaelpascal.hengartner@zhaw.ch

**Received:** 12 October 2016

**Accepted:** 05 January 2017

**Published:** 17 January 2017

### Citation:

Hengartner MP (2017) Subtle Scientific Fallacies Undermine the Validity of Neuroendocrinological Research: Do Not Draw Premature Conclusions on the Role of Female Sex Hormones. *Front. Behav. Neurosci.* 11:3. doi: 10.3389/fnbeh.2017.00003

## BACKGROUND

Based on extensive theoretical considerations, Ioannidis (2005a) drew the bold conclusion that, in biomedicine, “most published research findings are false”. Original studies subjecting psychological and medical research to close scrutiny have indeed provided ample evidence that many, and sometimes even a majority, of published findings are irreproducible, false positive, or severely inflated (Ioannidis, 2005b; Turner et al., 2008; Munafo et al., 2009; Bakker et al., 2012; Begley and Ellis, 2012; Open Science Collaboration, 2015; Müller et al., 2017); for reviews see (Nosek et al., 2012; Macleod et al., 2014; Naci and Ioannidis, 2015). More specifically, Open Science Collaboration (2015) tried to replicate 100 original studies published in 2008 in three leading psychology journals. While 97% of original studies had a significant result (i.e.,  $p < 0.05$ ), only 36% of the replication studies revealed a significant result, even though they all had high statistical power.

Moreover, in the original studies the average effect size was  $r = 0.403$  ( $SD = 0.188$ ), whereas in the replication studies it was only  $r = 0.197$  ( $SD = 0.257$ ), suggesting that the evidence for the original findings is rather weak (Open Science Collaboration, 2015). Very recently, Müller et al. (2017) meta-analyzed neuroimaging studies in persons with unipolar depression conducted between 1997 and 2015 and found not a single functional neuro-pattern that replicated consistently across studies, suggesting that the grand majority of published findings are irreproducible or false-positives. Finally, Contopoulos-Ioannidis et al. (2003) identified a total of 101 articles published between 1979 and 1983 in six leading biomedical journals, which clearly stated that the technology studied had novel therapeutic or preventive promises. However, most findings turned out to be false-positives or gross overestimations; by October 2002, only 27 of the promising technologies had resulted in at least one published randomized trial, five articles resulted in interventions with clinical licence, but only one finding lead to the development of an intervention that has been used extensively for the licensed indications (Contopoulos-Ioannidis et al., 2003).

In psychology/psychiatry the rate of positive findings (i.e.,  $p < 0.05$ ) is currently around 92% (Fanelli, 2010). Because 92% of all published results cannot possibly be true-positives given the average sample size and statistical power in published research, obviously strong and systematic biases are taking effect (Bakker et al., 2012; Button et al., 2013). There are various reasons for this excess significance bias, including foremost, reporting and publication biases (Ferguson and Heene, 2012; Glasziou et al., 2014; Ioannidis et al., 2014b) as well as methodological biases such as statistical misconduct and questionable research practices (Simmons et al., 2011; John et al., 2012; Ioannidis et al., 2014a). Because many conflicting results and negative findings are not objectively reported or published, the scientific literature is systematically biased towards spectacular positive findings (Young et al., 2008; Nosek et al., 2012), which leads to inflated meta-analytic estimates of effect sizes (Cuijpers et al., 2010; Ioannidis, 2011; Bakker et al., 2012). As a result, effectiveness of pharmaceutical drugs and impact of neurobiological and psychological findings are systematically overestimated (Ferguson and Heene, 2012; Button et al., 2013; Turner, 2013).

However, there are also subtle and less obvious fallacies. Sometimes conflicting results such as negative findings are sufficiently published, but these are simply not adequately acknowledged and considered (Jannot et al., 2013; Chalmers et al., 2014). That is, many researchers selectively review the literature and come to rather arbitrary conclusions that are not unequivocally supported by the published evidence (Ioannidis et al., 2007; Tatsioni et al., 2007; Saraga and Stiefel, 2011). Using the literature on the association between transition to menopause, hormonal changes and the onset of depression as a guiding example, I will thoroughly outline that many prominent claims do not hold up to close scrutiny and are therefore grossly exaggerated or false. I will ponder my arguments by testing two prevalent claims, the first being that the transition to menopause increases the risk for depression and the second that changing

hormone levels directly and uniformly increase the risk for depression.

## FIRST CLAIM: THE TRANSITION TO MENOPAUSE INCREASES THE RISK FOR DEPRESSION

That claim has been put forward by various authors of original studies (e.g., Bromberger et al., 2011; Joffe et al., 2016) and narrative reviews (e.g., Freeman, 2010; Soares, 2010). For instance, Bromberger et al. (2011) stressed that “we have moved from the “belief” that women were particularly susceptible to depression after the menopausal transition to the current empirically supported conclusion that middle-aged women are at a greater risk for depression during the transition than before [references]” (p. 1879). In addition, Soares (2010) stated that “Unlike cross-sectional studies, most prospective studies [references] have systematically confirmed the menopausal transition as a period of heightened risk for development of depressive symptoms and/or depression” (p. 2). Both strengthen their statement by citing various prospective observational studies that ostensibly confirmed the association between depression and transition to menopause. However, that view does not withstand close scrutiny and has been challenged by various contradictory findings (Judd et al., 2012; Rössler et al., 2016). First of all, that claim misses that most prospective studies did actually NOT report an association (Vesco et al., 2007; Rössler et al., 2016), so here many researchers are producing a severe confirmation bias by selectively reviewing a minority of studies with positive results that are not representative of the broader literature (Ioannidis et al., 2007; Tatsioni et al., 2007; Chalmers et al., 2014). Of concern is also the substantial content overlap between depression and menopause symptoms such as sleep disturbance, fatigue and irritability (Judd et al., 2012; Davis et al., 2015). Part of the covariance between menopause and depression symptoms is therefore tautological, which artificially inflates the strength of association. This could at least in part explain why some studies relying of self-report inventories of depression find a significant association anyway. In addition, various prospective studies that did report a positive association appear to be systematically biased due to the inadequate statistical procedure of dichotomization of a continuous measure of depression. As demonstrated by Rössler et al. (2016), when a continuous measure of depression is artificially dichotomized into categories of present vs. absent based on arbitrary cut-offs, false-positive associations with menopause stages can emerge. That the dichotomization of continuous variables conveys severe bias is widely acknowledged in the methodological literature (Ragland, 1992; MacCallum et al., 2002; Royston et al., 2006). Unfortunately, that ill-advised practice is still very prevalent in psychological and medical research. Rössler et al. (2016) name further biases in highly cited reports of positive associations between menopause transition and the onset of depression. These comprise among others overfitting and multiple adjustments in multivariable regression models,

which can also produce inflated or false-positive associations (Babyak, 2004; Simmons et al., 2011).

In conclusion, firstly, most prospective studies failed to provide support for a positive association between the transition to menopause and the occurrence of depression (Vesco et al., 2007; Rössler et al., 2016). Unfortunately, when looking at citation patterns one will easily recognize that the studies with positive findings receive much more attention than studies with negative findings (Kjaergard and Gluud, 2002; Jannot et al., 2013; Glasziou et al., 2014). For example, the negative findings from Avis et al. (1994) and Kaufert et al. (1992) received 499 and 380 citations (as of October 2016), whereas the much more recent positive findings by Cohen et al. (2006) and Freeman et al. (2006) received already 562 and 538 citations. Second, some studies that did report a positive association applied problematic statistical procedures indicating that data were possibly “tortured until they confess” (Wagenmakers et al., 2012) to obtain statistically significant associations at  $p < 0.05$  (Simmons et al., 2011; John et al., 2012; Ioannidis et al., 2014a). That phenomenon is also well documented as p-hacking and referred to journals’ and researcher’s aversion towards negative findings (Young et al., 2008; Ferguson and Heene, 2012; Nosek et al., 2012).

## SECOND CLAIM: HORMONE LEVELS DIRECTLY INFLUENCE DEPRESSIVE SYMPTOMS

Due to a selective reference to positive associations between the transition to menopause and depression (see above), it was concluded that the hormonal changes during that period may cause depression. As a result, hormone replacement therapy has been proposed as a potent first-line treatment for depression in peri-menopausal women (Riecher-Rössler and de Geyter, 2007; Georgakis et al., 2016). As above, that claim is not sufficiently supported by the literature. First of all, the association between hormone levels and depression during the transition to menopause is unclear, that is, positive associations are merely anecdotal and have not been replicated thus far (see reviews by Freeman, 2010; Vivian-Taylor and Hickey, 2014). So, once rapidly increasing FSH relates to lower risk of depression (Freeman et al., 2004) and once to higher risk of depression (Ryan et al., 2009), then again it is testosterone (Bromberger et al., 2010) and according to another study it is the fluctuating estradiol (Freeman et al., 2006) that causes depression. Moreover, as we would expect from a true null association, it comes as no surprise that some studies did not find any association between hormone levels and depression at all (Woods et al., 2008; Bromberger et al., 2011). In accordance, results from randomized clinical trials testing the efficacy of hormone therapy for depressive symptoms in menopausal women are also inconsistent (Soares et al., 2001; Morrison et al., 2004; Joffe et al., 2011) and provide overall only weak evidence in support of hormone therapy for depressed peri-menopausal but not post-menopausal women (see systematic review by Rubinow et al., 2015). However, due to substantial reporting and publication biases in the evaluation of drug trials (Turner,

2013; Naci and Ioannidis, 2015), these suggestions must be taken with reservation. Also, in initially non-depressed women who make the transition to menopause, hormone therapy has no preventive value (Rubinow et al., 2015). Recently, Georgakis et al. (2016) showed in a comprehensive meta-analysis that in women who naturally enter menopause ( $n = 67434$ ), higher age at menopause was associated with a marginally smaller risk of depression (OR = 0.98 for a 2-year increment,  $p < 0.05$ ). They supposed that this association was due to estrogen exposure and concluded that estrogen-based therapies could be useful to prevent depression in women who naturally enter the menopause before population average, even though the systematic review by Rubinow et al. (2015) concluded that hormone therapy has no preventive value. In a commentary on Georgakis et al. (2016), Hengartner (2016) transformed that odds ratio effect size into the number needed to treat (NNT), which is a convenient measure to quantify the effectiveness of medical treatments (Cook and Sackett, 1995). He showed that the NNT was 500, suggesting that 500 women would need to undergo continuous estrogen substitution for 2 years in order to prevent depression in only one woman (Hengartner, 2016). Of course, such a treatment is ineffective and no option to prevent depression in women making the transition to menopause. Unfortunately, misleading and empirically unsupported conclusions are quite common in medical research (Ioannidis et al., 2007; Saraga and Stiefel, 2011). In addition, another very prevalent scientific fallacy in contemporary psychological and medical research is the confusion of statistical significance with practical significance (Kirk, 1996). When the sample size is large enough (say  $n > 10000$ ), even negligibly small deviations from the null that bear absolutely no practical implications will become statistically significant (Cohen, 1994; Hengartner, 2016).

In conclusion, a critical examination of the literature does not support the view that sex hormone levels uniformly relate to depression. A systematic review of the literature indicates that most published findings are anecdotal, irreproducible and inconsistent, especially in the behavioral neurosciences (Ioannidis, 2011; Rosmalen and Oldehinkel, 2011; Ioannidis et al., 2014b; Sundström Poromaa and Gingnell, 2014; Müller et al., 2017). Nevertheless, many researchers tend to selectively review the literature (Tatsioni et al., 2007; Jannot et al., 2013) and to uncritically draw conclusions that are not sufficiently supported by the literature (Ioannidis et al., 2007; Saraga and Stiefel, 2011; Chalmers et al., 2014). Premature treatment recommendations can expose thousands of patients to treatments that are largely ineffective or that may cause more harm than good (Naci and Ioannidis, 2015). In this respect it is also necessary to evaluate the net benefit of a given therapy. Too often researchers equate statistical significance with clinical significance (Hengartner, 2016), ignoring that statistical significance does not allow to make inferences on the effectiveness of a given treatment (Cohen, 1994; Kirk, 1996).

## CONCLUDING REMARKS

By now it is increasingly documented that some psychological and medical science is systematically flawed (Ioannidis, 2005a;

Ferguson and Heene, 2012; Pashler and Harris, 2012; Macleod et al., 2014). However, both researchers and journals have an aversion towards null results, because they neither boost a researcher's career nor a journal's reputation (Young et al., 2008; Nosek et al., 2012). Major scientific fallacies such as reporting and publication biases cause a systematic overestimation of reported effect sizes to the point that many associations in the scientific literature are eventually marginally small or false-positives (Bakker et al., 2012; Ferguson and Heene, 2012; Ioannidis et al., 2014b; Naci and Ioannidis, 2015). However, reporting and publication biases are only the tip of the iceberg. Using the association between transition to menopause, hormones and depression as a guiding example, I demonstrated that, first, the scientific literature is often selectively reviewed and synthesized (Tatsioni et al., 2007) and second, that unsupported conclusions and treatment recommendations are readily made (Saraga and Stiefel, 2011). Specifically, in contrast to prominent claims that an increased risk for depression during the transition to menopause is clearly confirmed by the literature (Soares, 2010; Bromberger et al., 2011), a critical examination of the literature reveals that most prospective studies did not show a meaningful association (Vesco et al., 2007; Rössler et al., 2016). As the transition to menopause is a critical life event accompanied by major psychosocial changes, the most promising explanation is that these psychosocial risk factors may predispose vulnerable women to depression, but not the hormonal changes *per se* (Kaufert et al., 1992; Nelson, 2008; Rössler et al., 2016). In line with that notion it has been shown that an association between hormonal changes during the menopausal transition and depression is largely inconsistent (compare: Freeman et al., 2004; Woods et al., 2008; Ryan et al., 2009; Bromberger et al., 2010) and that the efficacy of hormone therapy to treat depression is at best inconclusive (Nelson, 2008; Rubinow et al., 2015). Therefore, recommending estrogen as a useful (preventive) treatment for depression in peri-menopausal women appears misguided and premature.

I strongly believe that hormones are crucial for affect and behavior. However, their effect on psychopathology is certainly not direct and uniform, otherwise research would have shown it. It is therefore supposed that hormones modulate complex psychobiological mechanisms such as the stress response system, and that these, further modulated through

even more psychobiological factors, impact on psychopathology (Zahn-Waxler et al., 2008; Naninck et al., 2011; Handa and Weiser, 2014). However, research into complex neurobiological systems is just at its beginning. As long as the neurosciences are so fundamentally hampered by manifold methodological biases, including notoriously underpowered samples, opaque experimental designs, almost unrestricted flexibility in data analysis and flawed statistical methods (e.g., Kriegeskorte et al., 2009; Carp, 2012; Button et al., 2013; Eklund et al., 2016), increased consistency and reproducibility of neurobiological research findings will not be readily achieved. A particularly striking example was provided by the senior authors of the Tracking Adolescents' Individual Lives Survey (TRAILS; Rosmalen and Oldehinkel, 2011). This large multisite project was designed to focus on the effects of cortisol on psychopathology. Rosmalen and Oldehinkel (2011) self-critically admit, that when they soon realized that the initial cross-sectional analyses yielded mainly null results or negligibly weak associations, the principal investigators started to test manipulations and different definitions of predictor and outcome variables; they arbitrarily included varying sets of potential confounders, and they used alternative statistical procedures in order to obtain statistically significant associations. The result of these inappropriate data manipulations (see Simmons et al., 2011) is that they achieved to publish spectacular positive findings in the leading journals. However, ultimately these confusing findings were highly inconsistent and most likely irreproducible false-positives that did not advance our knowledge in this field (Rosmalen and Oldehinkel, 2011). Looking at other hot topics in the behavioral neurosciences (e.g., Sundström Poromaa and Gingnell, 2014), one will easily recognize that this issue is the rule rather than the exception. It will therefore presumably take some time until we know how sex hormones and complex neurobiological systems influence psychopathology. In the meantime we are best advised to abstain from oversimplified and premature conclusions and instead to pay more attention to scientific and methodological biases.

## AUTHOR CONTRIBUTIONS

MPH wrote the entire manuscript.

## REFERENCES

- Avis, N. E., Brambilla, D., McKinlay, S. M., and Vass, K. (1994). A longitudinal analysis of the association between menopause and depression. Results from the Massachusetts women's health study. *Ann. Epidemiol.* 4, 214–220. doi: 10.1016/1047-2797(94)90099-x
- Babak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom. Med.* 66, 411–421. doi: 10.1097/01.psy.0000127692.23278.a9
- Bakker, M., van Dijk, A., and Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspect. Psychol. Sci.* 7, 543–554. doi: 10.1177/1745691612459060
- Begley, C. G., and Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. *Nature* 483, 531–533. doi: 10.1038/483531a
- Bromberger, J. T., Kravitz, H. M., Chang, Y.-F., Cyranowski, J. M., Brown, C., and Matthews, K. A. (2011). Major depression during and after the menopausal transition: Study of Women's Health Across the Nation (SWAN). *Psychol. Med.* 41, 1879–1888. doi: 10.1017/S003329171100016X
- Bromberger, J. T., Schott, L. L., Kravitz, H. M., Sowers, M., Avis, N. E., Gold, E. B., et al. (2010). Longitudinal change in reproductive hormones and depressive symptoms across the menopausal transition: results from the Study of Women's Health Across the Nation (SWAN). *Arch. Gen. Psychiatry* 67, 598–607. doi: 10.1001/archgenpsychiatry.2010.55
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475
- Carp, J. (2012). The secret lives of experiments: methods reporting in the fMRI literature. *Neuroimage* 63, 289–300. doi: 10.1016/j.neuroimage.2012.07.004



- Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gulmezoglu, A. M., et al. (2014). How to increase value and reduce waste when research priorities are set. *Lancet* 383, 156–165. doi: 10.1016/S0140-6736(13)62229-1
- Cohen, J. (1994). The earth is round ( $p < 0.05$ ). *Am. Psychol.* 49, 997–1003. doi: 10.1037/0003-066X.49.12.997
- Cohen, L. S., Soares, C. N., Vitonis, A. F., Otto, M. W., and Harlow, B. L. (2006). Risk for new onset of depression during the menopausal transition: the Harvard study of moods and cycles. *Arch. Gen. Psychiatry* 63, 385–390. doi: 10.1001/archpsyc.63.4.385
- Contopoulos-Ioannidis, D. G., Ntzani, E., and Ioannidis, J. P. (2003). Translation of highly promising basic science research into clinical applications. *Am. J. Med.* 114, 477–484. doi: 10.1016/S0002-9343(03)00013-5
- Cook, R. J., and Sackett, D. L. (1995). The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 310, 452–454. doi: 10.1136/bmj.310.6977.452
- Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., and Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *Br. J. Psychiatry* 196, 173–178. doi: 10.1192/bjp.bp.109.066001
- Davis, S. R., Lambrinou, I., Lumsden, M., Mishra, G. D., Pal, L., Rees, M., et al. (2015). Menopause. *Nat. Rev. Dis. Primers* 1:15004. doi: 10.1038/nrdp.2015.4
- Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U S A* 113, 7900–7905. doi: 10.1073/pnas.1602413113
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS One* 5:e10068. doi: 10.1371/journal.pone.0010068
- Ferguson, C. J., and Heene, M. (2012). A vast graveyard of undead theories: publication bias and psychological science’s aversion to the null. *Perspect. Psychol. Sci.* 7, 555–561. doi: 10.1177/1745691612459059
- Freeman, E. W. (2010). Associations of depression with the transition to menopause. *Menopause* 17, 823–827. doi: 10.1097/gme.0b013e3181db9f8b
- Freeman, E. W., Sammel, M. D., Lin, H., and Nelson, D. B. (2006). Associations of hormones and menopausal status with depressed mood in women with no history of depression. *Arch. Gen. Psychiatry* 63, 375–382. doi: 10.1001/archpsyc.63.4.375
- Freeman, E. W., Sammel, M. D., Liu, L., Gracia, C. R., Nelson, D. B., and Hollander, L. (2004). Hormones and menopausal status as predictors of depression in women in transition to menopause. *Arch. Gen. Psychiatry* 61, 62–70. doi: 10.1001/archpsyc.61.1.62
- Georgakis, M. K., Thomopoulos, T. P., Diamantaras, A. A., Kalogirou, E. I., Skalkidou, A., Daskalopoulou, S. S., et al. (2016). Association of age at menopause and duration of reproductive period with depression after menopause: a systematic review and meta-analysis. *JAMA Psychiatry* 73, 139–149. doi: 10.1001/jamapsychiatry.2015.2653
- Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., et al. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 383, 267–276. doi: 10.1016/S0140-6736(13)62228-x
- Handa, R. J., and Weiser, M. J. (2014). Gonadal steroid hormones and the hypothalamo-pituitary-adrenal axis. *Front. Neuroendocrinol.* 35, 197–220. doi: 10.1016/j.yfrne.2013.11.001
- Hengartner, M. P. (2016). Estrogen-based therapies and depression in women who naturally enter menopause before population average. *JAMA Psychiatry* 73:874. doi: 10.1001/jamapsychiatry.2016.0709
- Ioannidis, J. P. A. (2005a). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124
- Ioannidis, J. P. A. (2005b). Contradicted and initially stronger effects in highly cited clinical research. *JAMA* 294, 218–228. doi: 10.1001/jama.294.2.218
- Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Arch. Gen. Psychiatry* 68, 773–780. doi: 10.1001/archgenpsychiatry.2011.28
- Ioannidis, J. P. A., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., et al. (2014a). Increasing value and reducing waste in research design, conduct and analysis. *Lancet* 383, 166–175. doi: 10.1016/S0140-6736(13)62227-8
- Ioannidis, J. P. A., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., and David, S. P. (2014b). Publication and other reporting biases in cognitive sciences: detection, prevalence and prevention. *Trends Cogn. Sci.* 18, 235–241. doi: 10.1016/j.tics.2014.02.010
- Ioannidis, J. P. A., Palyzos, N. P., and Trikalinos, T. A. (2007). Selective discussion and transparency in microarray research findings for cancer outcomes. *Eur. J. Cancer* 43, 1999–2010. doi: 10.1016/j.ejca.2007.05.019
- Jannot, A. S., Agoritsas, T., Gayet-Ageron, A., and Perneger, T. V. (2013). Citation bias favoring statistically significant studies was present in medical research. *J. Clin. Epidemiol.* 66, 296–301. doi: 10.1016/j.jclinepi.2012.09.015
- Joffe, H., Crawford, S. L., Freeman, M. P., White, D. P., Bianchi, M. T., Kim, S., et al. (2016). Independent contributions of nocturnal hot flashes and sleep disturbance to depression in estrogen-deprived women. *J. Clin. Endocrinol. Metab.* 101, 3847–3855. doi: 10.1210/jc.2016-2348
- Joffe, H., Petrillo, L. F., Koukopoulos, A., Viguera, A. C., Hirschberg, A., Nonacs, R., et al. (2011). Increased estradiol and improved sleep, but not hot flashes, predict enhanced mood during the menopausal transition. *J. Clin. Endocrinol. Metab.* 96, E1044–E1054. doi: 10.1210/jc.2010-2503
- John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23, 524–532. doi: 10.1177/0956797611430953
- Judd, F. K., Hickey, M., and Bryant, C. (2012). Depression and midlife: are we overpathologising the menopause? *J. Affect. Disord.* 136, 199–211. doi: 10.1016/j.jad.2010.12.010
- Kaufert, P. A., Gilbert, P., and Tate, R. (1992). The manitoba project: a re-examination of the link between menopause and depression. *Maturitas* 14, 143–155. doi: 10.1016/0378-5122(92)90006-p
- Kirk, R. E. (1996). Practical significance: a concept whose time has come. *Educ. Psychol. Meas.* 56, 746–759. doi: 10.1177/0013164496056005002
- Kjaergard, L. L., and Gluud, C. (2002). Citation bias of hepato-biliary randomized clinical trials. *J. Clin. Epidemiol.* 55, 407–410. doi: 10.1016/S0895-4356(01)00513-3
- Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., and Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* 12, 535–540. doi: 10.1038/nn.2303
- MacCallum, R. C., Zhang, S. B., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7, 19–40. doi: 10.1037/1082-989x.7.1.19
- Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P., et al. (2014). Biomedical research: increasing value, reducing waste. *Lancet* 383, 101–104. doi: 10.1016/S0140-6736(13)62329-6
- Morrison, M. F., Kallan, M. J., Ten Have, T., Katz, I., Tweedy, K., and Battistini, M. (2004). Lack of efficacy of estradiol for depression in postmenopausal women: a randomized, controlled trial. *Biol. Psychiatry* 55, 406–412. doi: 10.1016/j.biopsych.2003.08.011
- Müller, V. I., Cieslik, E. C., Serbanescu, I., Laird, A. R., Fox, P. T., and Eickhoff, S. B. (2017). Altered brain activity in unipolar depression revisited. Meta-analyses of neuroimaging studies. *JAMA Psychiatry* 74, 47–55. doi: 10.1001/jamapsychiatry.2016.2783
- Munafò, M. R., Durrant, C., Lewis, G., and Flint, J. (2009). Gene X environment interactions at the serotonin transporter locus. *Biol. Psychiatry* 65, 211–219. doi: 10.1016/j.biopsych.2008.06.009
- Naci, H., and Ioannidis, J. P. (2015). How good is “evidence” from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? *Annu. Rev. Pharmacol. Toxicol.* 55, 169–189. doi: 10.1146/annurev-pharmtox-010814-124614
- Naninck, E. F., Lucassen, P. J., and Bakker, J. (2011). Sex differences in adolescent depression: do sex hormones determine vulnerability? *J. Neuroendocrinol.* 23, 383–392. doi: 10.1111/j.1365-2826.2011.02125.x
- Nelson, H. D. (2008). Menopause. *Lancet* 371, 760–770. doi: 10.1016/S0140-6736(08)60346-3
- Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific utopia: II. restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* 7, 615–631. doi: 10.1177/1745691612459058
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- Pashler, H., and Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspect. Psychol. Sci.* 7, 531–536. doi: 10.1177/1745691612463401

- Ragland, D. R. (1992). Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. *Epidemiology* 3, 434–440. doi: 10.1097/00001648-199209000-00009
- Riecher-Rössler, A., and de Geyter, C. (2007). The forthcoming role of treatment with oestrogens in mental health. *Swiss Med. Wkly.* 137, 565–572. doi: 2007/41/smw-11925
- Rosmalen, J. G., and Oldehinkel, A. J. (2011). The role of group dynamics in scientific inconsistencies: a case study of a research consortium. *PLoS Med.* 8:e1001143. doi: 10.1371/journal.pmed.1001143
- Rössler, W., Ajdacic-Gross, V., Riecher-Rössler, A., Angst, J., and Hengartner, M. P. (2016). Does menopausal transition really influence mental health? Findings from the prospective long-term Zurich study. *World Psychiatry* 15, 146–154. doi: 10.1002/wps.20319
- Royston, P., Altman, D. G., and Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat. Med.* 25, 127–141. doi: 10.1002/sim.2331
- Rubinow, D. R., Johnson, S. L., Schmidt, P. J., Girdler, S., and Gaynes, B. (2015). Efficacy of estradiol in perimenopausal depression: so much promise and so few answers. *Depress. Anxiety* 32, 539–549. doi: 10.1002/da.22391
- Ryan, J., Burger, H. G., Szoek, C., Lehert, P., Ancelin, M. L., Henderson, V. W., et al. (2009). A prospective study of the association between endogenous hormones and depressive symptoms in postmenopausal women. *Menopause* 16, 509–517. doi: 10.1097/gme.0b013e31818d635f
- Saraga, M., and Stiefel, F. (2011). Psychiatry and the scientific fallacy. *Acta Psychiatr. Scand.* 124, 70–72. doi: 10.1111/j.1600-0447.2011.01708.x
- Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632
- Soares, C. N. (2010). Can depression be a menopause-associated risk? *BMC Med.* 8:79. doi: 10.1186/1741-7015-8-79
- Soares, C. N., Almeida, O. P., Joffe, H., and Cohen, L. S. (2001). Efficacy of estradiol for the treatment of depressive disorders in perimenopausal women: a double-blind, randomized, placebo-controlled trial. *Arch. Gen. Psychiatry* 58, 529–534. doi: 10.1001/archpsyc.58.6.529
- Sundström Poromaa, I., and Gingnell, M. (2014). Menstrual cycle influence on cognitive function and emotion processing—from a reproductive perspective. *Front. Neurosci.* 8:380. doi: 10.3389/fnins.2014.00380
- Tatsioni, A., Bonitsis, N. G., and Ioannidis, J. P. (2007). Persistence of contradicted claims in the literature. *JAMA* 298, 2517–2526. doi: 10.1001/jama.298.21.2517
- Turner, E. H. (2013). Publication bias, with a focus on psychiatry: causes and solutions. *CNS Drugs* 27, 457–468. doi: 10.1007/s40263-013-0067-9
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., and Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* 358, 252–260. doi: 10.1056/NEJMsa065779
- Vesco, K. K., Haney, E. M., Humphrey, L., Fu, R., and Nelson, H. D. (2007). Influence of menopause on mood: a systematic review of cohort studies. *Climacteric* 10, 448–465. doi: 10.1080/13697130701611267
- Vivian-Taylor, J., and Hickey, M. (2014). Menopause and depression: is there a link? *Maturitas* 79, 142–146. doi: 10.1016/j.maturitas.2014.05.014
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., and Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspect. Psychol. Sci.* 7, 632–638. doi: 10.1177/1745691612463078
- Woods, N. F., Smith-DiJulio, K., Percival, D. B., Tao, E. Y., Mariella, A., and Mitchell, S. (2008). Depressed mood during the menopausal transition and early postmenopause: observations from the Seattle Midlife Women's Health Study. *Menopause* 15, 223–232. doi: 10.1097/gme.0b013e3181450fc2
- Young, N. S., Ioannidis, J. P., and Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Med.* 5:e201. doi: 10.1371/journal.pmed.0050201
- Zahn-Waxler, C., Shirtcliff, E. A., and Marceau, K. (2008). Disorders of childhood and adolescence: gender and psychopathology. *Annu. Rev. Clin. Psychol.* 4, 275–303. doi: 10.1146/annurev.clinpsy.3.022806.091358

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hengartner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution and reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.