The significance filter, the winner's curse and the need to shrink

Zwet, E.W. van; Cator, E.A.

WILEY

The significance filter, the winner's curse and the need to shrink

Erik W. van Zwet[1] | **Eric A. Cator**[2]

[1]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[2]Faculty of Science, Radboud University, Nijmegen, The Netherlands

Correspondence
Erik W. van Zwet, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands.
Email: E.W.van_Zwet@lumc.nl

The "significance filter" refers to focusing exclusively on statistically significant results. Since frequentist properties such as unbiasedness and coverage are valid only before the data have been observed, there are no guarantees if we condition on significance. In fact, the significance filter leads to overestimation of the magnitude of the parameter, which has been called the "winner's curse." It can also lead to undercoverage of the confidence interval. Moreover, these problems become more severe if the power is low. These issues clearly deserve our attention. They have been studied mostly through empirical observation and simulation, while there are relatively few mathematical results. Here we study them both from the frequentist and the Bayesian perspective. We prove that the relative bias of the magnitude is a decreasing function of the power and that the usual confidence interval undercovers when the power is less than 50%. We conclude that it is important to apply the appropriate amount of shrinkage to counter the winner's curse.

KEYWORDS
bias, exaggeration ratio, shrinkage, type M error

[Correction added on 11 June 2021 after first online publication: 'SE' was changed to 'se' and subheadings in 'APPENDIX' have been corrected in this version.]

1 | INTRODUCTION

The ongoing debate about the role of statistical significance in scientific research Roze-boom (1960), Meehl (1978) has recently intensified Wasserstein and Lazar (2016), Benjamin et al. (2018), Wasserstein, Schirm, and Lazar (2019), McShane, Gal, Gelman, Robert, and Tack-ett (2019), Amrhein and Greenland (2018), and Ioannidis (2019). Looking back to the beginning, we find that Ronald Fisher wrote in 1926 Fisher (1992):

> "Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and ignore entirely all results which fail to reach this level."

In other words, Fisher considered the familiar 5% level to be quite liberal and recommended that results that fail to reach even that level can be safely ignored. Now, more than 90 years later, Fisher's advice to apply the "significance filter" is widely followed. Recently, Barnett and Wren (2019) collected over 968,000 confidence intervals extracted from abstracts and over 350,000 intervals extracted from the full-text of papers published in Medline (PubMed) from 1976 to 2019. We converted these to z-values and their distribution is shown in Figure 1. The under-representation of z-values between -2 and 2 is striking.

As time and resources are always limited, it certainly makes sense to focus on significant results to avoid chasing noise. However, there is a problematic side-effect; considering only results that have reached statistical significance leads to overestimation Ioannidis (2008). This is sometimes called the "winner's curse." See Hedges (1984), Hedges and Vevea (2005), Lane and Dunlap (1978), Sterling (1959), Sterling, Rosenbaum, and Weinkam (1995) and in particular the monograph Rothstein, Sutton, and Borenstein (2005) for the many issues related to publication bias. Moreover, it has been demonstrated informally, that is, by simulation, that the winner's curse is especially severe when the power is low (Gelman & Carlin, 2014; Ioannidis, 2008). Here, we provide the first formal proof of this important fact.

As it turns out, low power is very common in the biomedical sciences (Button et al., 2013; Dumas-Mallet, Button, Boraud, Gonon, & Munafò, 2017). In particular, so-called pilot studies often have extremely low power. When such a study yields a significant result, the effect is likely grossly overestimated. Unfortunately, effect estimates from significant pilot studies are often used to inform the sample size calculation of a larger trial (Gelman & Carlin, 2014; Leon, Davis, & Kraemer, 2011). Low power also occurs when some correction is used to adjust for multiple comparisons. Such corrections are especially severe in genomics research, and the resulting
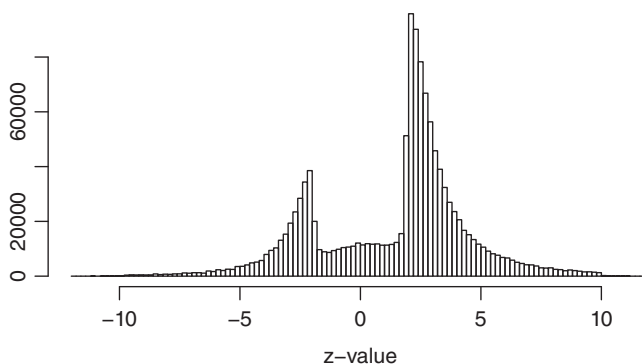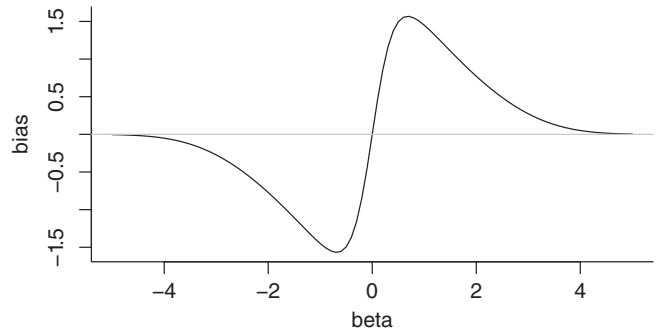

FIGURE 1 The distribution of more than one million z-values from Medline (1976–2019)

FIGURE 2 The bias as a function of β when se = 1, conditional on significance at the 5% level (i.e., $|b| > 1.96$)

overestimation of effects is well known (Göring, Terwilliger, & Blangero, 2001). The recent suggestion to lower the significance level to improve reproducibility (Benjamin et al., 2018) also reduces power, and therefore may backfire by aggravating the winner's curse (McShane et al., 2019).

While the winner's curse is a relatively well-known phenomenon, mathematical results are largely lacking. In the first part of this paper, we study the winner's curse from the frequentist point of view. We prove that the relative bias in the magnitude is a decreasing function of the power. We also examine the effect of the significance filter on the coverage of confidence intervals and find it results in undercoverage when the power is less than 50%.

In the second part of the paper, we study the significance filter from the Bayesian perspective. We conclude that it is necessary to apply shrinkage. We end the paper with a short discussion.

2 | THE FREQUENTIST PERSPECTIVE

Suppose that b is a normally distributed, unbiased estimator of β with se > 0. We have in mind that β is some regression coefficient such as a difference of means, a slope, a log odds ratio or log hazard ratio, and we shall sometimes refer to β as the "effect".

Of course, b is no longer unbiased if we condition on $|b|$ exceeding some threshold, for example $|b| > 1.96$ se. This "significance filter" happens when journals preferentially accept results that are statistically significant but *also* when authors or readers choose to focus on such promising results as per Fisher's advice. In Figure 2, we show the conditional bias $\mathbb{E}(b - \beta|\text{se}, \beta, |b| > 1.96 \text{ se})$ as a function of β (for se = 1).

We will now proceed to study the relative conditional bias of $|b|$ as an estimator of $|\beta|$, which is also known as the exaggeration ratio or type M error (Gelman & Carlin, 2014).

2.1 | Bias of the magnitude

By Jensen's inequality, $|b|$ is positively biased for $|\beta|$. Indeed, given β, $|b|$ has the folded normal distribution with mean

$$\mathbb{E}(|b||\text{se}, \beta) = |\beta| + \sqrt{\frac{2}{\pi}} \text{se } e^{-\beta^2/2\text{se}^2} - 2|\beta|\Phi\left(-\frac{|\beta|}{\text{se}}\right). \tag{1}$$

Proposition 1. *The bias* $\mathbb{E}(|b||\text{se}, \beta) - |\beta|$ *is positive for all* se *and* β. *Moreover, it is decreasing in* $|\beta|$ *and increasing in* se.
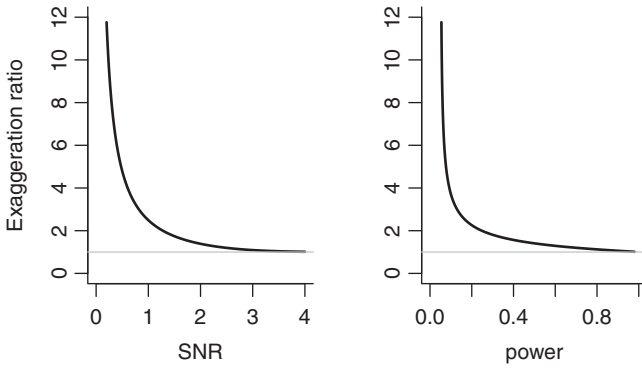
FIGURE 3 The exaggeration factor as a function of the signal-to-noise ratio and the power, when conditioning on significance at the 5% level ($c = 1.96$)

The proposition asserts that in low powered studies (small effects and large standard errors), the magnitude of the effect tends to be overestimated. For fixed se, the bias $\mathbb{E}(|b||\text{se}, \beta) - |\beta|$ is maximal at $\beta = 0$ where it is equal to $\sqrt{2/\pi}$ se ≈ 0.8 se.

Importantly, the bias in the magnitude becomes even larger if we condition on $|b|$ exceeding some a threshold. We have the following extension of Proposition 1.

Theorem 1. *The conditional bias* $\mathbb{E}(|b||\text{se}, \beta, |b|/\text{se} > c) - |\beta|$ *is positive for all* se *and* β*. Moreover, it is decreasing in* $|\beta|$ *and increasing in* se *and* c*.*

We define the *relative* conditional bias as

$$\frac{\mathbb{E}(|b||\text{se}, \beta, |b|/\text{se} > c) - |\beta|}{|\beta|},$$

and the exaggeration ratio or type M error (Gelman & Carlin 2014) as $\mathbb{E}(|b||\text{se}, \beta, |b| > c)/|\beta|$.

Corollary 1. *The relative conditional bias is positive and the exaggeration factor is greater than 1. Both depend on* β *and* se *only through the signal-to-noise ratio (SNR)* $|\beta|/$se*. Both quantities are decreasing in the SNR and increasing in* c*.*

We illustrate this result in Figure 3. Now the power for two-sided testing of $H_0 : \beta = 0$ at level 5% is

$$P(|b| > 1.96 \text{ se}|\beta, \text{se}) = \Phi(\text{SNR} - 1.96) + 1 - \Phi(\text{SNR} + 1.96),$$

which is a strictly increasing function of the SNR. Hence, the relative conditional bias and the exaggeration factor are decreasing functions of the power, as was already noted on the basis of simulation in Ioannidis (2008) and Gelman and Carlin (2014).

2.2 | Coverage

The significance filter also has consequences for the coverage of confidence intervals. We start by recalling their definition. Suppose a random variable X is distributed according to some distribution f_θ. A $(1 - \alpha) \times 100\%$ confidence set $S(X)$ is a random subset of the parameter space such that

$$P(\theta \in S(X)|\theta) = 1 - \alpha,$$

for all θ (Lehmann & Romano, 2006). A negatively biased semi-relevant (or recognizable) set R is a subset of the sample space such that

$$P(\theta \in S(X)|\theta, X \in R) < 1 - \alpha,$$

for all θ . It is quite problematic if such a set R exists, for is it still reasonable to report $S(X)$ with $(1 - \alpha) \times 100\%$ confidence, after the event $X \in R$ has been observed?

Semi-relevant sets have been constructed in various situations, most notably in case of the standard one-sample t-interval (Lehmann & Romano, 2006). Lehmann and Romano (2006) called the existence of certain relevant sets "an embarrassment to confidence theory". Now suppose b is normally distributed with mean β and known standard deviation se. If we define

$$S(b) = \{\beta : |b - \beta|/\text{se} < z_{1-\alpha/2}\}$$

where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution, then we have the following confidence statement

$$P(\beta \in S(b)|\beta, \text{se}) = P(|b - \beta|/\text{se} < z_{1-\alpha/2}|\beta, \text{se}) = 1 - \alpha,$$

for all $0 < \alpha < 1$, β and se > 0. Lehmann and Romano (2006) show that in this particular setting, there do not exist any negatively biased semi-relevant sets. This is certainly reassuring. However, if $c > 0$, then the conditional coverage

$$P(|b - \beta|/\text{se} < z_{1-\alpha/2}|\beta, \text{se}, |b|/\text{se} > c),$$

depends on β and se. This dependence is not simple. For instance, it is *not* monotone in β. We do have the following theorem.

Theorem 2. *Suppose b is normally distributed with mean β and standard deviation se. If the SNR $|\beta|/$se is less than $z = z_{1-\alpha/2}$ then*

$$P(|b - \beta|/\text{se} < z|\beta, \text{se}, |b|/\text{se} > z) < P(|b - \beta|/\text{se} < z|\beta, \text{se}) = 1 - \alpha. \tag{2}$$

Note that if the SNR is equal to $z_{1-\alpha/2}$, then the power for testing $H_0 : \beta = 0$ at level α is slightly more than 50%. So the theorem implies that if we have a significant result while the power is 50% or less, then the confidence interval will not reach its nominal coverage.

The result is quite sharp. By inspecting the proof, we can see that if the SNR is slightly larger than $z_{1-\alpha/2}$ then the conditional coverage *exceeds* the nominal (unconditional) coverage.

3 | THE BAYESIAN PERSPECTIVE

Bayesian inference is valid conditionally on the data, and so the significance filter should not pose any difficulties. On the other hand, Bayesian estimators are naturally biased. In this section we compare the performance of the unbiased estimator b and the Bayes estimator.

Let us assume that β has a normal prior distribution with mean 0 and known standard deviation $\tau > 0$. This is a rather simple model, but it serves to highlight the main issues.

Moreover, the normal prior is the standard choice in mixed models (in particular, random effects meta-analysis). The conditional distribution of β given b is normal with mean $b^* = \tau^2 b/(\text{se}^2 + \tau^2)$ and variance $v = \text{se}^2 \tau^2/(\text{se}^2 + \tau^2)$. We will write $s = \sqrt{v}$. Note that b^* is the Bayes estimator (under squared error loss) of β. Clearly, $|b^*| < |b|$ and for that reason b^* is called a shrinkage estimator.

We can evaluate b and b^* as estimators of β conditionally on the parameter and averaged over the distribution of the data, which is the frequentist point of view. Alternatively, we can condition on the data and average over the distribution of the parameter, which is the Bayesian point of view. We have the following nicely symmetric situation, where we consider se and τ to be fixed and known.

$$\mathbb{E}(b - \beta|\beta) = 0 \quad \text{and} \quad \mathbb{E}(b^* - \beta|\beta) = -\frac{\text{se}^2}{\text{se}^2 + \tau^2}\beta, \tag{3}$$

$$\mathbb{E}(b - \beta|b) = \frac{\text{se}^2}{\text{se}^2 + \tau^2}b \quad \text{and} \quad \mathbb{E}(b^* - \beta|b) = 0, \tag{4}$$

So, from the frequentist point of view, b is unbiased for β and b^* is biased. However, from the Bayesian point of view, it is the other way around!

3.1 | Bias of the magnitude

Now, if we are interested in the magnitude of β, then we could take the posterior mean of $|\beta|$ as an estimator. However, it is still relevant to evaluate the performance of $|b^*|$ as an estimator of $|\beta|$ from the Bayesian point of view. Conditionally on s and b^*, β has the normal distribution with mean b^* and standard deviation s and hence $|\beta|$ has the folded normal distribution. Similarly to Proposition 1, we have the following.

Proposition 2. *The difference $\mathbb{E}(|\beta||s, b^*) - |b^*|$ is positive. It is decreasing in $|b^*|$ and increasing in s. Moreover, the difference vanishes as $|b^*|$ tends to infinity.*

So, conditionally on the data, $|b^*|$ underestimates $|\beta|$ on average, but the difference disappears if we focus on large or significant effects. So now the significance filter actually *reduces* the bias in the magnitude! In other words, shrinkage lifts the winner's curse.

So far, we have conditioned either on the parameter or the data, and averaged over the other. However, in practice we do not keep the parameter fixed and repeat the experiment many times. We also do not keep the data fixed and vary the parameter. So, it is also relevant to consider the performance of b and b^* on average over the distribution of *both* the parameter *and* the data. If the distribution of the parameter represents some field of research, then this averaging will provide insight into how our statistical procedures perform when used repeatedly in that field.

Under our simple model, the marginal distribution of b is normal with mean zero and variance $\text{se}^2 + \tau^2$ and the marginal distribution of b^* is normal with mean zero and variance $\tau^4/(\text{se}^2 + \tau^2)$. So, trivially, $\mathbb{E}(b) = \mathbb{E}(b^*) = \mathbb{E}(\beta) = 0$. Moreover, it is easy to see that the variance of b^* is less than the variance of b. Marginally, $|\beta|$, $|b|$ and $|b^*|$ have half-normal distributions with means

$$\mathbb{E}|b^*| = \frac{\tau^2}{\sqrt{\text{se}^2 + \tau^2}}\sqrt{\frac{2}{\pi}}, \quad \mathbb{E}|\beta| = \tau\sqrt{\frac{2}{\pi}}, \quad \mathbb{E}|b| = \sqrt{\text{se}^2 + \tau^2}\sqrt{\frac{2}{\pi}}. \tag{5}$$

It is easy to see that

$$\mathbb{E}\,|b^*| < \mathbb{E}|\beta| < \mathbb{E}|b|. \tag{6}$$

The negative bias of $|b^*|$ may be preferable to the positive bias of $|b|$ in situations where it is important not to overstate the magnitude of an effect. This will depend on the application and indeed different stakeholders for a given application may have different opinions. It is interesting to note that the factor by which $|b|$ *over*estimates $|\beta|$ is the same as the factor by which $|b^*|$ *under*estimates it. That is,

$$\frac{\mathbb{E}|b|}{\mathbb{E}|\beta|} = \frac{\mathbb{E}|\beta|}{\mathbb{E}|b^*|} = \frac{\sqrt{\text{se}^2 + \tau^2}}{\tau}. \tag{7}$$

Moreover, the following proposition says that the bias of $|b^*|$ is smaller (on average) than the bias of $|b|$.

Proposition 3. *Suppose β has a normal prior distribution with mean 0 and standard deviation $\tau > 0$. Suppose that conditionally on β, b is normally distributed with mean β and standard deviation se > 0. Let $b^* = \mathbb{E}(\beta|b)$, then*

$$\mathbb{E}(|b| - |\beta|) > \mathbb{E}(|\beta| - |b^*|). \tag{8}$$

Most importantly, however, while the bias of $|b|$ increases as we condition on $|b|$ exceeding some threshold, the bias of $|b^*|$ vanishes!

Theorem 3. *As c goes to infinity, $\mathbb{E}(|b^*| - |\beta|\,||b| > c)$ vanishes.*

This result formalizes a blog post by Andrew Gelman (2016).

3.2 | Coverage

We now return to the coverage issue we discussed in Section 2.2. It might seem that Theorem 2 is not much of a problem in practice because conditional on a significant result, the power is unlikely to be small. But such an argument would depend on the (prior) distribution of the signal-to-noise ratio $|\beta|/\text{se}$. We have the following result.

Theorem 4. *Suppose β and se are distributed such that the SNR $|\beta|/\text{se}$ has a decreasing density and $|\beta|/\text{se}$ and se are independent. Also suppose that conditionally on β and se, b is normally distributed with mean β and standard deviation se. For every $0 < \alpha < 1$*

$$P(|b - \beta| < z_{1-\alpha/2}\text{se}\,||b|/\text{se} > c) < 1 - \alpha. \tag{9}$$

This result suggests that across research fields where the SNR has a decreasing density, confidence interval undercover on average. But how realistic is it to assume such a decreasing density? Clearly, it would imply a decreasing density of the absolute z-value, and this is certainly not the case in Figure 1. However we believe that this is due to selective reporting.

We have made an effort to collect an unselected sample of z-values as follows. It is a fairly common practice in the life sciences to build multivariate regression models by "univariable screening." First, the researchers run a number of univariable regressions for all predictors that
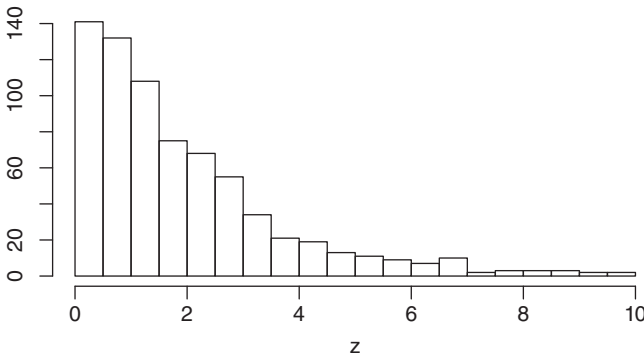
FIGURE 4 The distribution of 732 absolute z-values from Medline. Here we made an effort to avoid the significance filter

they believe could have an important effect. Next, those predictors with a p-value below some threshold are selected for the multivariate model. While this approach is statistically unsound, we believe that the univariable regressions should be largely unaffected by selection on significance, simply because that selection is still to be done. For further details, we refer to van Zwet (2019). We do note that in that article, we discarded p-values below .001, but these are included here.

We have collected 732 absolute z-values from 51 recent articles from Medline. We show the distribution in Figure 4 which suggest a decreasing distribution of the absolute z-values, which implies that the distribution of the SNR is decreasing as well.

4 | DISCUSSION

In this paper we have considered the generic situation where we have an unbiased, normally distributed estimator b of a parameter β, with known standard error se. Frequentist properties, such as the unbiasedness of b and the coverage of the confidence interval, are only meaningful before the data have been observed. Once the data are in, they become meaningless since b is just some fixed number and the confidence interval either covers β or it does not. Nothing more can be said without specifying a (prior) distribution for β.

However, suppose we condition not on (b, se) but only on the event $|b| > 1.96$ se. That is, we condition on statistical significance at the 5% level. Now b is still random and we can talk about bias and coverage. Conditionally on significance, b is biased away from zero. This tendency to overestimate the magnitude of significant effects is sometimes called the "winner's curse." It is especially severe when the SNR $|\beta|/$se is low. Also, if the SNR is low, then conditionally on significance the confidence interval will undercover. By providing mathematical proofs of these facts, we hope to contribute to the awareness of these very serious problems.

The goal of hypothesis testing is to try to avoid chasing noise, which is perfectly reasonable. However, the consequence of focusing on significant results is that all the nice frequentist properties no longer hold. Many proposals have been made to address this issue. From a frequentist point of view, one could condition throughout on statistical significance. See, for example, Ghosh, Zou, and Wright (2008) and references therein. Alternatively, one can take a Bayesian approach, such as proposed by Xu, Craiu, and Sun (2011) and ourselves (van Zwet, 2019). Of course, the Bayesian approach relies on correct specification of the prior.

Shrinkage is often viewed as a method to achieve a lower mean squared error by reducing the variance at the expense of increasing the bias. Our most important point is that it is

necessary to apply shrinkage to *reduce* the bias that results from focusing on interesting results. We explore how this may be achieved in van Zwet and Gelman (2020) and van Zwet, Schwab, and Senn (2020).

ORCID

Erik W. van Zwet https://orcid.org/0000-0001-5537-3179

REFERENCES

Amrhein, V., & Greenland, S. (2018). Remove, rather than redefine, statistical significance. *Nature Human Behaviour*, *2*(1), 4.

Barnett, A. G., & Wren, J. D. (2019). Examination of CIs in health and medical journals from 1976 to 2019: An observational study. *BMJ Open*, *9*(11), 1–4.

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., … Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour*, *2*(1), 6.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365.

Dumas-Mallet, E., Button, K. S., Boraud, T., Gonon, F., & Munafò, M. R. (2017). Low statistical power in biomedical science: A review of three human research domains. *Royal Society open science*, *4*(2), 160254.

Fisher, R. A. (1992). *The arrangement of field experiments*. In *Breakthroughs in statistics* (pp. 82–91). New York, NY: Springer.

Gelman, A. (2016). *Bayesian inference completely solves the multiple comparisons problem. Statistical modeling, causal inference, and social science*. Retrieved from https://statmodeling.stat.columbia.edu/

Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651.

Ghosh, A., Zou, F., & Wright, F. A. (2008). Estimating odds ratios in genome scans: An approximate conditional likelihood approach. *The American Journal of Human Genetics*, *82*(5), 1064–1074.

Göring, H. H., Terwilliger, J. D., & Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *The American Journal of Human Genetics*, *69*(6), 1357–1369.

Hedges, L. V. (1984). Estimation of effect size under nonrandom sampling: The effects of censoring studies yielding statistically insignificant mean differences. *Journal of Educational Statistics*, *9*(1), 61–85.

Hedges, L. V., & Vevea, J. L. (2005). *Selection method approaches*. In *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 145–174). Hoboken, New Jersey: John Wiley & Sons, Ltd.

Ioannidis, J. P. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648.

Ioannidis, J. P. (2019). The importance of predefined rules and prespecified statistical analyses: Do not abandon significance. *JAMA*, *321*(21), 2067–2068.

Lane, D. M., & Dunlap, W. P. (1978). Estimating effect size: Bias resulting from the significance criterion in editorial decisions. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 107–112.

Lehmann, E. L., & Romano, J. P. (2006). *Testing statistical hypotheses*. Berlin, Germany: Springer Science & Business Media.

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research*, *45*(5), 626–629.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, *73*(Suppl 1), 235–245.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, *46*, 806–834.

Rosenbaum, S. (1961). Moments of a truncated bivariate normal distribution. *Journal of the Royal Statistical Society Series B (Methodological)*, *23*, 405–408.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. In *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 1–7). Hoboken, New Jersey: John Wiley & Sons, Ltd.

Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*(5), 416.

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance or vice versa. *Journal of the American Statistical Association*, *54*(285), 30–34.

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*(1), 108–112.

van Zwet, E. W. (2019). A default prior for regression coefficients. *Statistical Methods in Medical Research*, *28*(12), 3799–3807.

van Zwet, E. W., & Gelman, A. (2020). *A proposal for informative default priors scaled by the standard error of estimates*. Retrieved from http://arxiv.org/abs/2011.15037/

van Zwet, E. W., Schwab, S., & Senn, S. J. (2020). *The statistical properties of RCTs and a proposal for shrinkage*. Retrieved from http://arxiv.org/abs/2011.15004

Wasserstein, R. L., & Lazar, N. A. (2016). *The ASA's statement on p-values: Context, process, and purpose*. New York, NY: Taylor & Francis.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). *Moving to a world beyond p < 0.05*. New York, NY: Taylor & Francis.

Xu, L., Craiu, R. V., & Sun, L. (2011). Bayesian methods to overcome the winner's curse in genetic studies. *The Annals of Applied Statistics*, *5*, 201–231.

APPENDIX

Proposition 1. *The bias* $\mathbb{E}(|b||\text{se}, \beta) - |\beta|$ *is positive for all* se *and* β*. Moreover, it is decreasing in* $|\beta|$ *and increasing in* se*.*

Proof. This is a special case of Theorem 1. ∎

Theorem 1. *The conditional bias* $\mathbb{E}(|b||\text{se}, \beta, |b|/\text{se} > c) - |\beta|$ *is positive for all* se *and* β*. Moreover, it is decreasing in* $|\beta|$ *and increasing in* se *and c.*

Proof. Let Z be a standard normal random variable and define

$$g(\theta, c) = \mathbb{E}\left(|\theta + Z| - \theta \mid |\theta + Z| \geq c\right). \tag{A1}$$

Since

$$\mathbb{E}(|b||\text{se}, \beta, |b|/\text{se} > c) - |\beta| = \text{seg}(|\beta|/\text{se}, c) \tag{A2}$$

it is clear that it is enough to prove that $g(\theta, c)$ is decreasing in $\theta > 0$ and increasing in $c > 0$.

Suppose $c_1 < c_2$. For any random variable X we have that

$$\mathbb{E}(X|X \geq c_2) \geq \mathbb{E}(X),$$

since $\mathbb{E}(X|X < c_2) \leq c_2 \leq \mathbb{E}(X|X \geq c_2)$ and $\mathbb{E}(X)$ is a convex combination of the two conditional expectations. Now we can replace X by $\tilde{X} = \mathbb{E}(X|X \geq c_1)$, and we conclude that g is increasing in $c > 0$.

To prove that $g(\theta, c)$ is decreasing in $\theta > 0$, note that the density of $|Z + \theta|$ is given by

$$f(y) = \begin{cases} \phi(y - \theta) + \phi(y + \theta), & y \geq 0 \\ 0, & y < 0. \end{cases}$$

Here ϕ is the standard normal density. Using that $z\phi(z) = -\phi'(z)$,

$$\begin{aligned} g(\theta, c) &= \frac{\int_c^\infty (y - \theta)(\phi(y - \theta) + \phi(y + \theta))\, dy}{P(|Z + \theta| \geq c)} \\ &= \frac{\int_c^\infty -\phi'(y - \theta) - \phi'(y + \theta) - 2\theta\phi(y + \theta)\, dy}{P(|Z + \theta| \geq c)} \\ &= \frac{\phi(c - \theta) + \phi(c + \theta) - 2\theta(1 - \Phi(c + \theta))}{2 - \Phi(c - \theta) - \Phi(c + \theta)}. \end{aligned} \tag{A3}$$

We split $g(\theta, c)$ into the numerator and the denominator:

$$N = \phi(c - \theta) + \phi(c + \theta) - 2\theta\left(1 - \Phi(c + \theta)\right),$$

and

$$D = 2 - \Phi(c - \theta) - \Phi(c + \theta).$$

Now it is enough to check that

$$D \cdot \frac{\partial N}{\partial \theta} \leq N \cdot \frac{\partial D}{\partial \theta}.$$

So

$$\begin{aligned} \frac{\partial N}{\partial \theta} &= -\phi'(c - \theta) + \phi'(c + \theta) - 2(1 - \Phi(c + \theta)) + 2\theta\phi(c + \theta) \\ &= (c - \theta)(\phi(c - \theta) - \phi(c + \theta)) - 2(1 - \Phi(c + \theta)), \end{aligned}$$

and

$$\frac{\partial D}{\partial \theta} = \phi(c - \theta) - \phi(c + \theta).$$

Introduce

$$z_- = c - \theta \quad \text{and} \quad z_+ = c + \theta.$$

Then

$$D \cdot \frac{\partial N}{\partial \theta} \leq N \cdot \frac{\partial D}{\partial \theta} \qquad \Leftrightarrow$$

$$D \cdot z_- \left(\phi\left(z_-\right) - \phi\left(z_+\right) \right) - 2D \cdot \left(1 - \Phi\left(z_+\right)\right) \leq N \cdot \left(\phi\left(z_-\right) - \phi\left(z_+\right)\right) \quad \Leftrightarrow$$

$$\left(\phi\left(z_-\right) - \phi\left(z_+\right)\right)\left(D \cdot z_- - N\right) \leq 2D \cdot \left(1 - \Phi\left(z_+\right)\right). \tag{A4}$$

The right-hand side of (A4) is clearly positive, and it is not hard to see that the first factor of the left-hand side is also positive: for $|z_-| \leq z_+$ we have

$$\phi\left(z_-\right) - \phi\left(z_+\right) \geq 0.$$

Therefore, we can show that (A4) is true, if we can show that

$$D \cdot z_- - N \leq 0. \tag{A5}$$

We can see that

$$D \cdot z_- - N = z_- \left(1 - \Phi(z_-)\right) + z_- \left(1 - \Phi(z_+)\right) - \phi(z_-) - \phi(z_+) + 2\theta(1 - \Phi(z_+))$$

$$= z_- \left(1 - \Phi(z_-)\right) - \phi(z_-) + z_+ \left(1 - \Phi(z_+)\right) - \phi(z_+). \tag{A6}$$

We now use the fact that for all $z \in \mathbb{R}$,

$$z(1 - \Phi(z)) - \phi(z) \leq 0,$$

which follows from the fact that the derivative of this function (i.e., $1 - \Phi(z)$) is positive, and the limit for $z \to \infty$ equals 0. So (A6) is indeed negative, which proves (A4), and therefore the fact that $g(\theta, c)$ is decreasing in $\theta \geq 0$. ∎

Corollary 1. *The relative conditional bias is positive and the exaggeration factor is greater than 1. Both depend on depend on β and* se *only through the SNR $|\beta|/$se. Both are decreasing in $|\beta|/$se and increasing in c.*

Proof. Recall the definition of the function $g(\theta, c)$ from (A1). The relative bias is equal to $\text{seg}(|\beta|/\text{se}, c)/|\beta|$ and the exaggeration factor is $\text{seg}(|\beta|/\text{se}, c)/|\beta| + 1$. We refer to the proof of Theorem 1 where we show that $g(\theta, c)$ is decreasing in $\theta \geq 0$ and increasing in $c \geq 0$. This also establishes the present claim. ∎

Proposition 2. *The difference $\mathbb{E}(|\beta||s, b^*) - |b^*|$ is positive. It is decreasing in $|b^*|$ and increasing in s. Moreover, the difference vanishes as $|b^*|$ tends to infinity.*

Proof. Comparing to Proposition 1, we see that this is also a special case of Theorem 1. ∎

Theorem 2. *Suppose b is normally distributed with mean β and standard deviation* se. *If $|\beta|/$se $\leq z = z_{1-\alpha/2}$ then*

$$P(|b - \beta|/\text{se} < z|\beta, \text{se}, |b|/\text{se} > z) < P(|b - \beta|/\text{se} < z|\beta, \text{se}) = 1 - \alpha. \tag{A7}$$

Proof. There is no loss of generality if we assume $\beta > 0$ and se $= 1$. In this proof, we will drop conditioning on β and se from our notation. In fact, without loss of generality we will prove the corresponding statement for $X \sim N(\mu, 1)$. Also, it is more convenient to work with the

complementary event $|X - \mu| > z$. Since

$$P(|X - \mu| > z||X| > z) = \frac{P(|X| > z||X - \mu| > z)P(|X - \mu| > z)}{P(|X| > z)},$$

is suffices to prove that

$$P(|X| > z|||X - \mu| > z) - P(|X| > z) > 0,$$

for all $0 < \mu \le z$. Now,

$$
\begin{aligned}
& P(|X| > z|||X - \mu| > z) - P(|X| > z) \\
& = P(|X| > z|X - \mu > z)/2 + P(|X| > z|X - \mu < -z)/2 - P(|X| > z) \\
& = \frac{1}{2} + P(X > z|X - \mu < -z)/2 + P(X < -z|X - \mu < -z)/2 - P(X > z) - P(X < -z) \\
& = \frac{1}{2} + \frac{\Phi(-z - \mu)}{2\Phi(-z)} - 1 + \Phi(z - \mu) - \Phi(-z - \mu).
\end{aligned}
$$

Taking the derivative with respect to μ, it is easy to see that this expression is decreasing in $\mu > 0$. Moreover, if we take $\mu = z$, then we get $\Phi(-2z)/2\Phi(-z) - \Phi(-2z)$, which is positive because $\Phi(-z) < 1/2$. ∎

Proposition 3. *Suppose β has a normal prior distribution with mean 0 and standard deviation $\tau > 0$. Suppose that conditionally on β, b is a normally distributed with mean β and standard deviation* se > 0. *Let $b^* = \mathbb{E}(\beta|b)$, then*

$$\mathbb{E}(|b| - |\beta|) > \mathbb{E}(|\beta| - |b^*|). \tag{A8}$$

Proof. We have to show that

$$\sqrt{\text{se}^2 + \tau^2} - \tau \ge \tau - \frac{\tau^2}{\sqrt{\text{se}^2 + \tau^2}}.$$

Multiplying by $\sqrt{\text{se}^2 + \tau^2}$ and rearranging we obtain

$$\text{se}^2 + \tau^2 - 2\tau\sqrt{\text{se}^2 + \tau^2} + \tau^2 \ge 0.$$

The left-hand side of this equality is equal to $(\sqrt{\text{se}^2 + \tau^2} - \tau)^2$ which is clearly positive unless se is zero. ∎

Theorem 3. *As c goes to infinity, $\mathbb{E}(|b^*| - |\beta|||b| > c)$ vanishes.*

Proof. Since the marginal distribution of b is symmetric around zero, we have for positive c

$$\mathbb{E}(|b^*|||b| > c) = \frac{\tau^2}{\text{se}^2 + \tau^2}\mathbb{E}(|b|||b| > c) = \frac{\tau^2}{\text{se}^2 + \tau^2}\mathbb{E}(b|b > c). \tag{A9}$$

Conditionally on $b > c$, b has the truncated normal distribution. Hence

$$\frac{\tau^2}{\text{se}^2 + \tau^2}\mathbb{E}(b|b > c) = \frac{\tau^2}{\text{se}^2 + \tau^2}\sqrt{\text{se}^2 + \tau^2}\frac{\varphi\left(c/\sqrt{\text{se}^2 + \tau^2}\right)}{1 - \Phi\left(c/\sqrt{\text{se}^2 + \tau^2}\right)}$$

$$= \frac{\tau^2\varphi\left(c/\sqrt{\text{se}^2 + \tau^2}\right)}{\sqrt{\text{se}^2 + \tau^2}(1 - \Phi(c/\sqrt{\text{se}^2 + \tau^2}))} \tag{A10}$$

Turning to $|\beta|$, we have by symmetry,

$$\mathbb{E}(|\beta|||b| > c) = \mathbb{E}(|\beta||b > c). \tag{A11}$$

Moreover,

$$\mathbb{E}(|\beta||b > c) = \mathbb{E}(\beta|b > c) + \mathbb{E}(|\beta| - \beta|b > c). \tag{A12}$$

By a result due to Rosenbaum (1961) concerning the mean of a truncated bivariate normal distribution, we have

$$\mathbb{E}(\beta|b > c) = \frac{\tau^2\varphi\left(c/\sqrt{\text{se}^2 + \tau^2}\right)}{\sqrt{\text{se}^2 + \tau^2}(1 - \Phi(c/\sqrt{\text{se}^2 + \tau^2}))}. \tag{A13}$$

Since this expression is equal to (A10), we only need to show that

$$\lim_{c\to\infty}\mathbb{E}(|\beta| - \beta|b > c) = 0.$$

Since $P(\beta < 0|b > c) \to 0$ as $c \to \infty$, this is clearly true. ∎

To prove Theorem 4, we use the following Lemma.

Lemma 1. *Let X be a random variable and g an increasing function that is not constant on the support of X. Then for every x such that $P(X > x) > 0$*

$$\mathbb{E}(g(X)|X < x) < \mathbb{E}(g(X)).$$

Proof. $\mathbb{E}(g(X))$ is a convex combination of $\mathbb{E}(g(X)|X < x)$ and $\mathbb{E}(g(X)|X \geq x)$. Since g is increasing

$$\mathbb{E}(g(X)|X < x) \leq \mathbb{E}(g(X)|X \geq x).$$

If g is not constant on the support of X, then the inequality is strict and the claim follows. ∎

To prove Theorem 4, we first prove the following Proposition.

Proposition 4. *Suppose μ is distributed such that $|\mu|$ has a decreasing density f. Also suppose Z is independent of μ and has a distribution which is symmetric around zero and supported on the whole real line. Let $X = Z + \mu$. For every positive c and z*

$$P(|X - \mu| < z||X| > c) < P(|X - \mu| < z). \tag{A14}$$

Proof. Since $Z = X - \mu$, we can rewrite the claim as

$$P(|Z| < z||Z + \mu| > c) < P(|Z| < z).$$

Because,

$$P(|Z| < z||Z + \mu| > c) = \frac{P(|Z + \mu| > c||Z| < z)P(|Z| < z)}{P(|Z + \mu| > c)},$$

the claim is equivalent to

$$P(|Z + \mu| > c||Z| < z) < P(|Z + \mu| > c), \tag{A15}$$

for all positive z and c. If we define

$$g(z) = P(|Z + \mu| > c||Z| = z),$$

then

$$P(|Z + \mu| > c||Z| < z) = \mathbb{E}(g(|Z|)||Z| < z).$$

Now we can use Lemma 1 to prove our claim by showing that g is increasing and not constant. Since the distribution of Z is symmetric around zero, its sign and magnitude are independent. Therefore,

$$g(z) = P(|Z + |\mu|| > c||Z| = z). \tag{A16}$$

Using the independence of Z and μ, we have

$$g(z) = P(|z + |\mu|| > c)/2 + P(|-z + |\mu|| > c)/2$$

$$= P(z + |\mu| > c)/2 + P(z + |\mu| < -c)/2 + P(-z + |\mu| > c)/2 + P(-z + |\mu| < -c)/2$$

$$= \begin{cases} P(|\mu| > c - z)/2 + P(|\mu| > c + z)/2, & \text{if } z \le c \\ \frac{1}{2} + P(|\mu| > c + z)/2 + P(|\mu| < -c + z)/2, & \text{if } z > c. \end{cases}$$

Taking the derivative, we have

$$g'(z) = \begin{cases} f(c - z) - f(c + z), & \text{if } z \le c \\ -f(c + z) + f(-c + z), & \text{if } z > c \end{cases}$$

$$= f(|z - c|) - f(z + c) \ge 0.$$

f is not constant since it is a (proper) density. It follows that g' cannot be identically zero and hence g is not constant either. ∎

Theorem 4. *Suppose β and se are distributed such that $|\beta|/$se has a decreasing density and $|\beta|/$se and se are independent. Also suppose that conditionally on β and se, b is normally distributed with mean β and standard deviation se. For every $0 < \alpha < 1$*

$$P(|b - \beta| < z_{1-\alpha/2}\text{se}||b|/\text{se} > c) < 1 - \alpha. \tag{A17}$$

Proof. For every se > 0, it follows from Proposition 1 that

$$P(|X - \mu| < z_{1-\alpha/2}\text{se}|\text{se}, |X|/\text{se} > c) < P(|X - \mu| < z_{1-\alpha/2}\text{se}|\text{se}) = 1 - \alpha.$$

Averaging over se, the claim follows. ∎