

Supplementary Note 2: Winners' curse correction of post-hoc power analysis of replication studies

In order to investigate the numbers of replicated loci observed in our study we carried out a Winner's Curse correct power analysis. The power of a study is the probability of rejecting the null hypothesis given that the alternate hypothesis is true (*i.e.*, the true positive rate, or sensitivity). By linearity of expectation, the sum of the powers of each of the SNP associations is an estimate of the number of replications we should expect in a replication cohort, under the assumption that all of the associations are true. For example, in our discovery cohort we identified 1,262 associations at a $-\log_{10} p$ -value threshold of 7.5 (see Supplementary Table 5), and in the $N = 930$ replication cohort we replicated 455 of these associations at nominal significance. A power analysis can determine whether or not these 455 replications are more than we should expect, because it allows us to estimate (or 'predict') the number of replications we would see in a 'best case' scenario in which all the associations are true. This power analysis is based on [1].

For a study that employs a linear model with significance level α , if the true value of the effect size is β and if the true value of the standard error is σ , then the power of a study with a cohort of size N is:

$$p(\beta, \sigma, N, \alpha) = f\left(g\left(1 - \alpha/2; N - 2\right); N - 2, |\beta|/\sigma\right).$$

Equation 1

Here $f(x; d, c)$ is the distribution function of a non-central Student-t distribution with d degrees of freedom and non-centrality parameter c evaluated at x , and $g(x; d)$ is the quantile function of a Student-t distribution with d degrees of freedom evaluated at x [2].

To examine the power of a replication cohort, we may use Equation 1 and substitute the effect size found in the discovery cohort. However, this is a biased estimate of the true effect size. The bias arises from the fact that we are conditioning on the event that the association was found to be significant at level α in the discovery cohort. There is some fluctuation in effect size due to sampling noise, and so conditioning on significance will bias the sampling noise towards the tails of the distribution on the effect size. This is known as the Winner's Curse [3]. Methods have been provided for correcting this bias [1], and we describe those methods here as they are relevant, and we adapt them to our replication paradigm. In [1], the following approximation of the likelihood of the effect size is provided, conditioned on the observations of the study and conditioned on the truth of the alternate hypothesis:

$$L(\beta' | \hat{\beta}, \hat{\sigma}, N, H_1, \alpha) \approx \frac{\varphi(\beta'; \hat{\beta}, \hat{\sigma})}{p(\beta', \hat{\sigma}, N, \alpha)}.$$

Equation 2

Here $\varphi(x; \mu, \sigma)$ is the density function of a normal distribution with mean μ and standard deviation σ evaluated at x . This likelihood, as a function of $(\beta', \hat{\beta})$ is supported on the set:

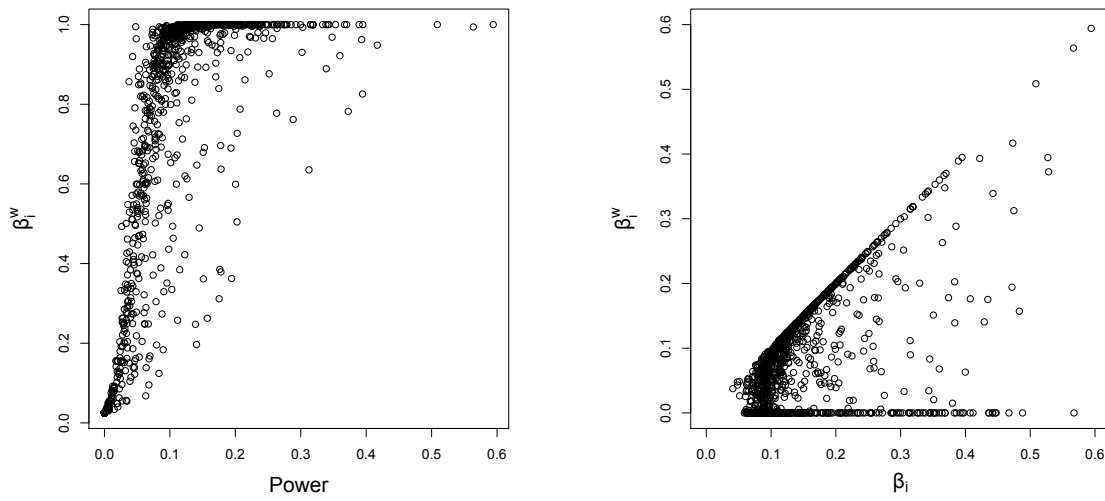
$$\left\{ (\beta', \hat{\beta}) : \text{sign}(\beta') = \text{sign}(\hat{\beta}), |\hat{\beta}|/\hat{\sigma} > g(1 - \alpha/2; N - 2) \right\}.$$

Equation 3

And so for the i -th association in our study, if we observe an effect size of $\hat{\beta}_i$ and a standard error of $\hat{\sigma}_i$ in the discovery cohort, and if the size of the discovery cohort is N^d , and if the significance level of the discovery is α , then in [1] the winners' curse corrected effect size is β_i^w and is given as follows:

$$\beta_i^w = \underset{\beta'}{\text{argmin}} L(\beta' | \hat{\beta}_i, \hat{\sigma}_i, N^d, H_1, \alpha).$$

Equation 4



Supplementary Note 3; Figure 1: Winners' curse corrected effect sizes. Left) The y-axis shows the winners' curse corrected effect sizes for the 1,262 associations reported in Supplementary Table 5 and the x-axis shows the power to replicate the association in the $N = 3,456$ replication cohort. Right) The action of the winners' curse correction is shown; the x-axis shows the effect sizes before correction, and the y-axis shows the effect sizes after correction. Zeros in the corrected effect sizes indicate loss of significance in the discovery cohort.

As in [2], We compute these winners' curse corrected effect sizes β_i^w using Nelder—Mead optimization [4]. The action of this correction is shown above in Supplementary Note 3; Figure 1. Using these winners' curse corrected effect sizes, we then compute an unbiased estimate of the expected number of associations that replicate at nominal significance in the replication cohort:

$$E[\#\text{replications}|H_1] = \sum_i p(\beta_i^w, \hat{\sigma}_i^r, N^r, 0.05).$$

Equation 5

Here $\hat{\sigma}_i^r$ is the observed standard error in the replication cohort, and N^r is the size of the replication cohort. Thus, Equation 5 provides a post-hoc winners' curse corrected power analysis to examine the expected number of replicated clusters.

<i>Condition</i>	<i>Predicted replications</i>	<i>Actual replications</i>
N = 930	486	455
N = 3,456	740	844
N = 3,456, singleton clusters	72	68
N = 3,456, clusters	161	148

Supplementary Note 3; Table 1: Predicted replications from post-hoc winners' curse corrected power analysis compared to actual number of replications in various replication conditions. The number of replications we actually find are less than what would be predicted under the alternate hypothesis for all conditions except N = 3,456.

Using Equation 5, we calculated that the expected number of replications under the alternate hypothesis among the 1,262 associations listed in Supplementary Table 5, and found that for the $N = 930$ replication cohort, we expected 486 replications (we found 455 in the experiment). And, for the $N = 3,456$ replication cohort, we expected 740 replications (and we found 844 in the experiment).

While the number of replicated associations we found for $N = 3,456$ is in excess of the expectation, we note that linkage disequilibrium among SNPs and correlation among phenotypes act to increase the variance of the estimate of this expectation. Associations that are all in the same "cluster" tend to all replicate or fail to replicate more or less together. To understand this variance, we restricted Equation 5 to sum only over the 314 "singleton" clusters — clusters that contained only a single association, and therefore exhibit less correlation. We calculate that in the $N = 3,456$ replication cohort, the expected number of replicated singleton clusters was 72 (we found 68 in the experiment).

We also used Equation 5 to calculate the expected number of replicated clusters in the $N = 3,456$ replication cohort. Assuming independence among the associations, the expected number of replicated clusters is as follows (a 'noisy or'):

$$E[\#\text{replications}] = \sum_j 1 - \prod_{i \in C_j} (1 - p(\beta_i^w, \hat{\sigma}_i^r, N^r, 0.05)).$$

Equation 6

Here C_j is the set of indices of the associations assigned to the j -th cluster. We calculated that the expected number of replicated clusters was 161 (and we found 148 in the experiment). So, even though the number of replications we find in the raw associations in the $N = 3,456$ replication cohort is larger than what is predicted by post-hoc winners' curse corrected power analysis; when we aggregate over clusters, or examine singleton clusters only, or examine the $N = 930$ replication cohort, we replicate fewer associations than what is predicted by post-hoc winners' curse corrected power analysis. These results are tabulated in Supplementary Note 3; Table 1. This analysis therefore suggests that our results are not outside of the realm of what could be expected through a post-hoc power analysis of the replication cohorts.

References

1. R. Xiao and M. Boehnke. *Quantifying and correcting for the winner's curse in quantitative-trait association studies*. Genetic Epidemiology. **35**(5), 133-138 (2011)
2. T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical Learning*. Springer. (2001)
3. R. H. Thaler. *Anomalies: The Winner's Curse*. Journal of Economic Perspectives. **2**(1), 191-202 (1988)
4. J. A. Nelder and R. Mead. *A simplex method for function minimization*. Computer Journal. **7**(4), 308-313 (1965)