GWAS Explorer (GWAX) Formulae

Jeremy Yang

May 2019

1 GWAX Gene-Trait Association Scoring

1.1 RCRAS = Relative Citation Ratio (RCR) Aggregated Score

The purpose is to evaluate the evidence for a gene-trait association, by aggregating multiple studies and their corresponding publications. The iCite[1] RCR is itself a statistic designed to evaluate the evolving empirical impact of a publication (in contrast to the non-empirical impact factor). Hence by aggregating RCRs we seek an corresponding measure of scientific community impact.

$$RCRAS_{gt} = \sum_{study} \left(\frac{1}{grc} \sum_{pmid} \frac{log_2(RCR) + 1}{spp} \right) \tag{1}$$

Where

$$
\begin{aligned}
grc &= \quad \text{(gene-reported count)} \\
pmid &= \quad \text{(PubMed ID)} \\
spp &= \quad \text{(study-per-pmid count)}
\end{aligned}
$$

RCR $median = 2.0$ and $90\%ile = 8.5$. The $log_{10}()$ function is used with the belief that each additional publication adds less evidence. Division by spp effects a partial count for papers associated with multiple studies. Similarly division by grc reflects a partial count since papers and studies may associate to few or many findings and reported genes.

From GWAS Catalog[2] and iCite PubMed statistics.

As with loss functions, the absolute value is less important than the gradient, which solely determines ranking. For example, ten supporting publications may not be twice the evidence than five, but is certainly more.

---

[1] https://icite.od.nih.gov/
[2] https://www.ebi.ac.uk/gwas/