Journal of Biometrics & Biostatistics

Zhang, J Biom Biostat 2016, 7:3
http://dx.doi.org/10.4172/2155-6180.1000297

Opinion Open Access

On The Use of P-Values in Genome Wide Disease Association Mapping

Yu Zhang*

Department of Statistics, The Pennsylvania State University, USA

***Corresponding author:** Yu Zhang, Department of Statistics, The Pennsylvania State University, USA, Tel: (814) 867-0780; E-mail: yzz2@psu.edu

Rec date: April 24, 2016; **Acc date:** April 25, 2016; **Pub date:** May 2, 2016

Copyright: © 2016 Zhang Y. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

In hypothesis testing, p-value is routinely used as a measure of statistical evidence against the null hypothesis, where a smaller p-value indicates stronger evidence substantiating the alternative hypothesis. P-value is the probability of type-I error made in a hypothesis testing, namely, the chance that one falsely reject the null hypothesis when the null holds true. In a disease genome wide association study (GWAS), p-value potentially tells us how likely a putative disease associated variant is due to random chance. For a long time p-values have been taken seriously by the GWAS community as a safeguard against false positives. Every disease-associated mutation reported in a GWAS must reach a stringent p-value cutoff (e.g., 10-8) in order to survive the multiple testing corrections. This is reasonable because after testing millions of variants in the genome, some random variants ought to yield small p-values purely by chance. Despite of p-value's theoretical justification, however, it has become increasingly evident that statistical p-values are not nearly as reliable as it was believed. It has not been uncommon for a GWAS to identify some very significant associations that later turn out to be false positives. The current routine is therefore to require every reported genetic variant for a disease to be replicated at least once, which is a much more reliable criterion against false positives.

Type-I Errors in GWAS

The problem often arises in the miscalculation p-values. Calculating p-values under the null hypothesis is not as simple as introduced in textbook, because the null hypothesis made in practice are often too simplistic. In a case control study, for instance, the samples are often assumed independent and sufficiently large, for which statistical theory allows us to calculate p-values analytically. In reality, such assumptions are almost never met due to human inheritance and sample availability, both of which may either inflate or deflate theoretical p-values. Such issues were recognized early on, and numerous statistical methods have been developed to account for sample relatedness, including genomic control, sample matching, model-based stratification detection, and more recently generalized mixed effect models. Numerical solutions for calculating p-values in finite samples have also been extensively developed, including various permutation methods, Monte Carlo Markov chain simulations, and methods adjusting for multiple comparisons. While these computational solutions have been useful, new methods are constantly needed to accommodate new study designs and new data characteristics generated by improved technologies in GWAS. Rare variant, for example, is currently thought by many as the key to disease missing heritability, and is being intensively studied. Because of their low frequency in the population, none of the existing asymptotic theories applies. In fact, how to best test rare variants, adjust for sample bias and evaluate their significance are open questions.

Type-II Errors in GWAS

Type-II error in hypothesis testing refers to falsely accepting the null hypothesis when the alternative is true, and its complement, power, is the more familiar term in GWAS. Most existing GWAS are underpowered due to limited sample size and small effects of disease variants. Optimizing power in GWAS is therefore an important and challenging problem. Without increasing sample size, there are several alternative approaches to improve. First, one may gain power by developing better test statistics or computational methods that more efficiently capture the true signals in a disease model. Most statistical methods belong to this category, such as tests for dominant, recessive, or haplotype effects, epistasis association mapping methods, burden tests and vibrational methods for testing cumulative effects in rare variants. Depending on the underlying disease model, some methods may outperform the others, but there will be no most powerful methods for all scenarios. Secondly, one may combine independent studies of the same or similar diseases together, followed by variant imputation and joint analysis. Assuming that multiple studies carry a similar set of disease variants, joint analysis of the combined results can effectively increase sample size and thus improve power. This strategy has been used to combine tens or hundreds of thousands of samples from different studies and successfully revealed many new disease variants that are otherwise undetectable in individual studies. Thirdly, one may leverage information from orthogonal data sets to help narrowing down the regions and sets of variants for disease risks. This approach can effectively reduce the search space and thus increase power. Examples include population origin information in admixed sample, haplotype structures, gene annotations, and various functional data sets (e.g., expression, epigenetic marks, chromatin accessibility, conservation, motif, etc.). This latter approach of using functional data in GWAS is relatively new, and it has quickly gained popularity, not only due to the availability of a plethora of functional data generated by high throughput sequencing, but also that it has the potential to unveil the functional roles of genetic variants in disease. With increasing evidence suggesting that there may be hundreds or thousands of genetic mutations affecting the risks of complex traits, most of which may have very small effects, one needs to think out of the box and develop novel integrative methods to combine all information to gain power in GWAS.

Type-III Errors in GWAS

Although this term is purely statistical, its meaning is not unfamiliar to the GWAS community. Type-III error refers to making a right decision by wrong reasons. A most typical example is that a significant association detected in GWAS is due to genotyping errors. That is, a variant violates the null hypothesis not because it is associated with the disease. In fact, genotyping issues are so common that every GWAS requires stringent quality control before inference is made. Another common scenario that causes type-III errors, which is much less

appreciated, is the effect of linkage disequilibrium (LD). Considering two variants in LD, where one is causal and the other is not, or both are not but tagging some untyped causal variants. When testing the two variants, both may turn out to be significant, although only one or perhaps none is causal. At first glance, this is a harmless known fact that tagging variants are not causal variants. What is not being realized is that LD effects can substantially increase false discoveries in the genome scale. Using the major histocompatibility complex (MHC) in human as an example, the region has mega-base pair long LD blocks that often yield many hundreds of significant variants in autoimmune diseases, most of which are due to LD effects. If one includes the MHC region with the rest of the genome and calculate an overall false discovery rate (FDR) at 0.05 levels, then 5% false positives (corresponding to ~50 false positives raised by MHC) will be tolerated. These 5% false positives are randomly distributed in the genome. So after grouping nearby significant variants, there will be just 1 true positive loci at MHC and 50 false positive loci elsewhere, leading to 98% FDR! This highlights the discrepancy between statistical and biological significance: the hundreds of significant variants in MHC are true positives in the statistical sense, but they are essentially all tagging a single causal variant in the biological sense, and thus should be counted as one.

Conclusion

While p-value tells us how much the data substantiate the alternative hypothesis, its usage has not been most appropriate in GWAS. On one hand, the null hypothesis setup in a GWAS is often overly simplistic that does not include all possible scenarios that may induce signals other than disease association. On the other hand, most disease association tests (and hence p-values) in GWAS are calculated using genetic data alone, which have not been accounting for the large amount of non-genetic information about the genome that may improve the power of association mapping. The importance of p-values in GWAS has therefore been decreasing, with each GWAS having to replicate its disease variants in addition to genome-wide significance. It is very likely that the use of p-values in GWAS will continue to decline. In fact, there are other statistics that also measure statistical evidence in the data, such as Bayes factor, but is better than p-values. Bayes factor can be more robust than p-values and are more flexible in terms of modeling a complex null or alternative hypothesis. This will be important for big data, because not only the true signals will be amplified in big data, but also non-random bias in the data will become detectable when there are enough samples, for which the null hypothesis will almost never be simple. Finally, one must remember that there is always a difference between statistical significance and biological significance. Additional important aspects of GWAS results, such as reproducibility and interpretability, should always be evaluated.