J Med Microbiol Infect Dis, 2018, 6 (1): 35-36
DOI: 10.29252/JoMMID.6.1.35

Letter to the Editor

Hardy Weinberg Equilibrium Testing and Interpretation Focus on Infection

Sana Eybpoosh[1,2*]

[1]Department of Epidemiology and Biostatistics, Research Centre for Emerging and Reemerging Infectious Diseases, Pasteur Institute of Iran, Tehran, Iran; [2]HIV/STI Surveillance Research Center, and WHO Collaborating Center for HIV Surveillance, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

Received Aug 01, 2018; Accepted Aug 08, 2018

To the editor: Recent advances in genetic association (GA) studies have reported novel associations between common genetic markers and human susceptibility to many major infectious diseases. In parallel, an increasing number of rare genetic variants have been identified that are associated with susceptibility to infection. Together, these achievements have highlighted the genetic architecture of infectious disease susceptibility and illuminated immune and cellular pathways that are associated with the infection process. However, for a genetic association study to yield valid results, particular considerations on study design and conduct are required to be taken. Application of GA studies in the field of infectious diseases is at its early stage, calling the need for further elaborations on design considerations of these studies. Perhaps one of the most commonly referred approach in GA studies is to test for "Hardy-Weinberg equilibrium" (HWE, described below) principle. However, this principle is frequently disregarded or miss-interpreted in GA studies of infectious diseases. So, in this piece, I would further elaborate on this concept, its applications, and its testing and interpretation.

The HWE principle deals with the interdependence of allele and genotype frequencies in a population. Allele frequency is the proportion of a variant of a gene (allele) in a given locus among all alleles in the population. Genotype frequency is the number of individuals with a particular genotype in a population divided by the total number of individuals in that population. Both allele and genotype frequencies are proportions in nature, which mean that they can hold a range of valuse between 0 and 1, and the sum of allele (as well as genotype) frequencies will equal one. Now, assume a large yet close population with no emigrations, immigrations, or population bottlenecks, where individuals randomly mate and there are no evolutionary forces on individuals' genome. Under this scenario, alleles will remain at a constant frequency over generation. Since the frequency of alleles are stable, genotype frequencies can be predicted from allele frequencies using a simple function. For example, if *a* and *b* are the allele frequencies of a bi-allelic trait, then the expected frequency of genotypes *aa*, *ab*, and *bb* would be $(a)^2$, 2 $(a \times b)$, and $(b)^2$, respectively. These genotype frequencies will remain constant over generations. This phenomenon is termed Hardy-Weinberg equilibrium [1-2]. HWE principle can also be applied to loci with more than two alleles. In this case, the expected genotype frequencies are derived by multinomial expansion of all k alleles: $(a+b+c+. . .+k)^2$.

Testing for HWE is now a common practice in population genetics and genetic association studies, where conforming to HWE expectations is usually desired. Although most of the assumptions of the HWE principle are not expected to hold for most human populations, there is a common expectation from HWE principle to hold for most populations of healthy individuals. Deviations from HWE principle at particular markers are considered as a suggestion of population sub-structure, genotyping error or, in samples of diseased individuals, an association with the disease [1, 3].

Despite its use in everyday practice and its practical importance, empirical data suggest that HWE reporting may be suboptimal in both genetic [4] and non-genetics [5] journals. For example, re-analysis of the data from 92 studies (150 associations) published in high-profile genetic journals showed significant deviation from HWE in the disease-free controls of 13 studies (20 associations), but less than 20% of them were admitted in the published articles. Evidence from this study also suggested that most studies conforming to HWE simply were mostly underpowered to detect HWE deviation [4]. There are also cases where the HWE test result is misinterpreted. For example, statistically significant results of the HWE test have been interpreted as the existence of HWE, while this shows a violation of the HWE assumption [6].

***Correspondence:** Sana Eybpoosh
Department of Epidemiology and Biostatistics, Pasteur Institute of Iran, No. 69, Pasteur Ave, Tehran, Iran, 1316943551.
Email: s_eybpoosh@pasteur.ac.ir, sana.eybpoosh@gmail.com
Tel: +98 (21) 64112121 **Fax:** +98 (21) 66465132

This empirical evidence suggests that, even in high profile genetics journals, testing and reporting for HWE is often neglected and deviations are rarely admitted in the published reports. Hence, I have provided a brief guide on the applications of HWE, the corresponding statistical test for that, and proper interpretation of its results, in order to prevent similar issues in the future.

HWE applications. The validity of a GA study depends considerably on the use of a proper control (non-infected) group [7]. As stated above, healthy (non-infected) human populations from outbred populations are expected to hold for HWE principle [8, 9]. So, the assessment of this assumption in the control (healthy) group of a GA study has been a common practice for long times [10]. The same applies to conditions where all subjects have a common infectious disease, for example, studies evaluating different treatments, whenever the disease risk per se is not influenced by the evaluated polymorphism [4]. Deviations from HWE can be indicative of genotyping error. This is mostly the case in genome-wide association studies where millions of single nucleotide polymorphisms (SNP) are genotyped, hence the chance of genotyping error is high. In these studies, HWE is usually checked and SNPs whose distribution deviates from HWE among controls, are discarded from the dataset. So, in the 'data analysis' phase, assessment of HWE in such datasets would be meaningless, as all SNPs violating from HWE has previously been excluded from the data set [11].

HWE test. HWE is not merely a theoretical law; deviations can signal essential problems, errors, or peculiarities in the analyzed data sets [12, 13]. The critical inferences from a GA study, for example, may be compromised if HWE is violated. Violation from HWE is tested using Pearson's $\chi 2$ goodness-of-fit test, or an asymptotically equivalent variant such as the log-likelihood-ratio test. The tests evaluate the degree of difference between observed genotype and allele frequencies with the frequencies that are expected if HWE assumption holds. If observed vs. expected frequencies are different to a large extent, then the test would turn statistically significant and suggest violations from HWE assumption [3, 14].

As stated previously, HWE test provides insights into a population's actual genetic structure. HWE test has five basic assumptions: 1) the population size is infinitely large; 2) between-populations gene flow, from migration or transfer of gametes, does not exist; 3) there are negligible amount of mutations and migrations; 4) individuals are mating randomly; and 5) there are no natural selections operating on the population [1, 15-17]. So, significant test results suggest that one or more of the abovementioned assumptions are being violated.

Overall, significant departures from HWE should invoke thinking for the underlying reason. Regardless of the underlying reason, departures from HWE may suggest that allele-infectious disease associations are biased. For these reasons, HWE testing should be routinely and appropriately used and adequately interpreted in the setting of genetic association studies of infectious diseases.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest associated with this manuscript.

REFERENCES

1. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet. 2005; 76 (6): 967-86.

2. Hardy GH. Mendelian proportions in a mixed population. Science. 1908; 28 (706): 49-50.

3. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet. 2004; 12 (5): 395-9.

4. Salanti G1, Amountza G, Ntzani EE, Ioannidis JP. Hardy-Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power. Eur J Hum Genet. 2005; 13 (7): 840-8.

5. Bardoczy Z, Gyorffy B, Kocsis I, Vasarhelyi B. Re-calculated Hardy–Weinberg values in papers published in Atherosclerosis between 1995 and 2003. Atherosclerosis. 2004; 173 (1): 141-3.

6. Alubadi AEM, Salih AM, Alkhamesi MBN, Ali NJ. Gene frequencies of ABO and rhesus blood groups in Sabians (Mandaeans), Iraq. Baghdad Sci J. 2014; 11 (2): 1035-42.

7. Eybpoosh S, Haghdoost AA, Mostafavi E, Bahrampour A, Azadmanesh K, Zolala F. Molecular epidemiology of infectious diseases. Electron Phys. 2017; 9 (8): 5149-58.

8. Khoury MJ, Little J, Burke W. Human genome epidemiology: a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004.

9. Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York: Oxford University Press, 1993.

10. Mehrbod P, Eybpoosh S, Fotouhi F, Shokouhi Targhi H, Mazaheri V, Farahmand B. Association of IFITM3 rs12252 polymorphisms, BMI, diabetes, and hypercholesterolemia with mild flu in an Iranian population. Virol J. 2017; 14 (1): 218.

11. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nat Rev Genet. 2008; 9 (5): 356-69.

12. Khoury MJ, Little J, Burke W: Human genome epidemiology. a scientific foundation for using genetic information to improve health and prevent disease. New York: Oxford University Press, 2004.

13. Sham P. Statistics in human genetics. London: Arnold Publishers, 2001.

14. Ward R, Carroll RJ. Testing Hardy-Weinberg equilibrium with a simple root-mean-square statistic. Biostatistics. 2014; 15 (1): 74-86.

15. Relethford JH. Human Population Genetics. 1 edition. New Jersey: Wiley-Blackwell; 2012.

16. Kimura A. Departure from the Hardy-Weinberg equilibrium. Gene. 2014; 537 (2): 357.

17. Lachance J. Hardy–Weinberg Equilibrium and Random Mating. Encycloped Evol Biol. 2016: 208-11.