

## How to calculate the Hardy-Weinberg equilibrium for polymorphisms teaching? A facilitating spreadsheet for molecular biology remote class

*Como calcular o equilíbrio de Hardy-Weinberg para o ensino de polimorfismos genéticos? Uma planilha facilitadora para aula remota de biologia molecular*

Howard Lopes Ribeiro Junior <sup>1,2,3\*</sup>, João Vitor Caetano Goes <sup>1,3</sup>

<sup>1</sup> Post-Graduate Program of Pathology, Federal University of Ceara - CE.

<sup>2</sup> Post-Graduate Program in Translational Medicine, Federal University of Ceara - CE.

<sup>3</sup> Center for Research and Drug Development (NPDM), Federal University of Ceara - CE.

\*e-mail: [howard@ufc.br](mailto:howard@ufc.br)

### Abstract

This article describes a complete self-filling spreadsheet for Hardy-Weinberg Equilibrium (HWE) calculate to facilitate understanding of the importance of calculating HWE for experimental research on polymorphism analysis in a molecular biology remote class during the Coronavirus 2019 (COVID-19) pandemic period.

**Keywords:** Polymorphism; Teaching; Remote class.

### Resumo

Este artigo descreve uma planilha de autopreenchimento completa para o cálculo de Hardy-Weinberg Equilibrium (HWE) para facilitar a compreensão da importância do cálculo de HWE para pesquisa experimental sobre análise de polimorfismo em uma classe remota de biologia molecular durante o período pandêmico do Coronavirus 2019 (COVID-19).

**Palavras-chave:** Polimorfismo; Ensino; Aulas remotas.

## Record activity performed

<b>Title</b>	How to calculate the hardy-weinberg equilibrium for polymorphisms teaching? A facilitating spreadsheet for molecular biology remote class
<b>Target audience</b>	Higher education students in the field of biological sciences and health.
<b>Related disciplines</b>	Molecular Biology and Human Genetics.
<b>Educational objectives</b>	To demonstrate the use of simple spreadsheet for teaching and calculation of Hardy-Weinberg equilibrium for polymorphisms studies.
<b>Justification of use</b>	Due to the difficulties in conducting practical activities in the teaching of genetic polymorphism, especially in relation to concepts of equilibrium Hardy-Weinberg, a playful and efficient spreadsheet was proposed for teaching this content, either by classroom or by remote teaching and to experimental studies of population genetics research.
<b>Worked contents</b>	Hardy-Weinberg Equilibrium and Genetic polymorphism.
<b>Estimated duration</b>	40 minutes
<b>Materials used</b>	Microsoft Excel.
<b>Link</b>	<a href="https://docs.google.com/spreadsheets/d/1xgJvLly5JK7he5_w75UwwPqE1ZR8NrLCOmjUaWei4SA/edit?usp=sharing">https://docs.google.com/spreadsheets/d/1xgJvLly5JK7he5_w75UwwPqE1ZR8NrLCOmjUaWei4SA/edit?usp=sharing</a>

## **1 Introduction**

Since the announcement of the Coronavirus disease 2019 (COVID-19) pandemic by the World Health Organization in March 2020, all global organizations need to readjust to the social isolation necessary to minimize the spread of the COVID-19 virus, especially the face-to-face teaching [1]. Some studies have shown that polymorphisms in the human angiotensin I converting enzyme 2 (ACE-2) gene, the main target for viral binding to the membrane of lung cells could influence the COVID-19 susceptibility infection rate and the poor outcome of the COVID-19 disease [2,3]. It is known that an important topic to be addressed in molecular biology classes on studies of genetic polymorphisms is the presentation of concepts about the Hardy-Weinberg Equilibrium (HWE) and its assumptions [4,5].

Based on the studies by Hardy [4] and Weinberg [5], Smith and Baldwin [6] didactically summarize that HWE advocates that HWE assumes that a population is in genetic equilibrium if the genotype frequencies and allele frequencies are the same in each generation at birth. Among the seven assumptions of the HWE [7], it is necessary to expand the experimental analysis in such a way as to maintain a balance of the genotype frequencies in a population in which the size (read as “sample number”) of the population must be infinitely large. Therefore, analyzing whether the genotypes obtained in the molecular analyzes are in HWE or not is essential for understanding whether the evaluated sample size is sufficient to verify differences in genotype distribution. It is known that testing deviation from the HWE is generally performed by Pearson's chi-squared test using the observed genotype frequencies obtained from the data and the expected genotype frequencies obtained using the HWE.

About this finding, one important question should be raised: Could teachers improve molecular biology classes, especially about polymorphism studies, to be accessible and understandable when taught remotely? This report will seek to present an online tool that demonstrates the step-by-step process of how to calculate the HWE to be applied to a group of students in the Postgraduate Course in Pathology at a public university in Northeast Brazil. In this report, we describe a complete self-completing spreadsheet for HWE calculations that is based on the Microsoft (MS) Excel platform to facilitate understanding of the importance of calculating HWE for experimental research on polymorphism analysis (Figure 1). This spreadsheet will be available in the cloud (Google drive®) for students to use in real time in class.

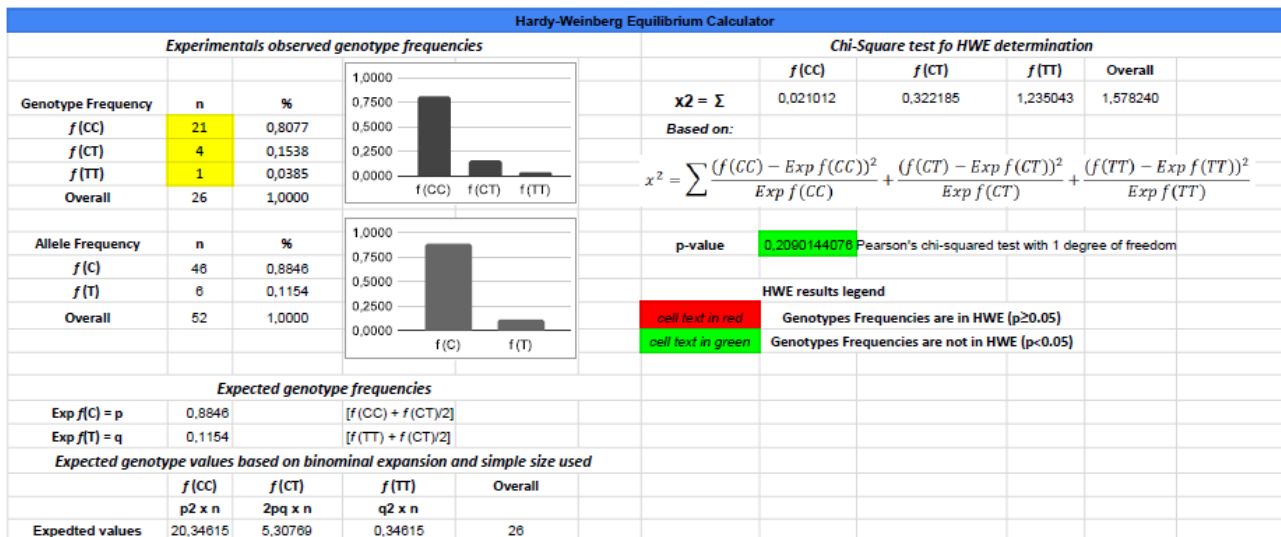


Figure 1. Schematic representation of the spreadsheet layout to calculate the HWE

## 2 Description of the Hardy-Weinberg Equilibrium calculator spreadsheet

According to the developed spreadsheet, in our virtual class, we didactically will present distinct concepts about HWE according to the steps below:

1. Initially, the researcher/student must use only the values related to the observed genotypic frequencies, stratified by each genotype, based on polymorphisms experiments (for example, real-time polymerase chain reaction [qPCR assays] or PCR-restriction length fragment polymorphism [RFLP]) in the yellow cells. The other analyses and a bar chart are calculated automatically (such as the allele frequency);

2. Subsequently, the spreadsheet will calculate the expected genotypic frequencies based on Punnet Square for HWE [7] in which the proportion of each genotype is equal to the product of the row and column allele frequencies from the current generation according to equations 1 and 2:

(1)

$$f(CC) = f(C) \cdot f(C) = p \cdot p = p^2$$

$$f(CT) = f(C) \cdot f(T) = p \cdot q + p \cdot q = 2pq$$

$$f(TT) = f(T) \cdot f(T) = q \cdot q = q^2$$

(2)

$$p = Exp f(C) = f(CC) + f(CT)/2$$

$$q = Exp f(T) = f(TT) + f(CT)/2$$

3. Based on the binominal expansion and selected sample size, the spreadsheet will calculate the expected genotypic values according to equation 3:

(3)

$$q^2 \times n + 2pq \times n + p^2 \times n$$

\* n means sample number.

4. Based on the observed and expected frequency values, the spreadsheet will calculate the chi-square values ( $x^2$ ) for all evaluated genotypes according to equation 4:

(4)

$$x^2 = (f(CC) - \text{Exp } f(CC))^2 / \text{Exp } f(CC)$$

$$x^2 = (f(CT) - \text{Exp } f(CT))^2 / \text{Exp } f(CT)$$

$$x^2 = (f(TT) - \text{Exp } f(TT))^2 / \text{Exp } f(TT)$$

5. Pearson's chi-square value will be calculated from the sum of all squares of calculated genotypes according to equation 5:

(5)

$$x^2 = \sum (f(CC) - \text{Exp } f(CC))^2 / \text{Exp } f(CC) + (f(CT) - \text{Exp } f(CT))^2 / \text{Exp } f(CT) + (f(TT) - \text{Exp } f(TT))^2 / \text{Exp } f(TT)$$

6. Finally, based on Pearson's chi-square value, a 5% significance level for one degree of freedom will be calculated based on Excel equation (6):

(6)

$$=CHIDIST(x; \text{degree of freedom})$$

\* x indicates Pearson's chi-square value. Degree of freedom equal 1

7. The result of the HWE analysis will be presented in color. If the cell is colored green, it means that the genotype frequencies are in HWE since the p value was  $\geq 0.05$ . This result will validate the sample number used in the survey. On the other hand, if the cell is colored red, it means that the genotype frequencies are not in HWE, since the p value was  $< 0.05$ . Therefore, the sample number of the research has not been validated; thus, it would be necessary to increase the number of cases evaluated. We reinforce that the spreadsheet will be of free access and unlimited access for use by teachers and/or researchers.

### 3 Perspectives to remote class

For this remote class, polymorphism analysis data presented in the study by Ribeiro-Jr and collaborators will be used [9]. To facilitate the virtual meet and prevent internet problems, students will be asked to download the worksheet onto their computers and to read this paper and the article with genetic data prior to class.

Khoury and colleagues describes that HWE is not a simple a theoretical law because its deviations can signal essential problems or errors in the analyzed data sets [10]. Thus, in summary, we have followed the current demands of remote education by addressing virtual practices and skills involving the concepts of genetic polymorphisms and HWE assumptions, also linked to biostatistics analyses. We know that the development of this teaching tool comes against the fact that it is essential and necessary that higher education teachers improve their classes with the use of didactic, accessible, and low-cost virtual tools. Regardless of the teaching method, we know that practical knowledge is indispensable for the researcher's educational and training process in biochemistry and molecular biology [11,12].

### References

- [1] Bogoch II, Watts A, Thomas-Bachli A, et al. Potential for global spread of a novel coronavirus from China. *J Travel Med.* 2020; 27(2):taaa011.
- [2] Devaux CA, Rolain JM, Raoult D. ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. *J Microbiol Immunol Infect.* 2020;53(3):425-435.
- [3] Delanghe JR, Speeckaert MM, De Buyzere ML. ACE Ins/Del genetic polymorphism and epidemiological findings in COVID-19. *Clin Chem Lab Med.* 2020;58(7):1129-1130.
- [4] Hardy GH. Mendelian proportions in a mixed population. *Science.* 1908; 28: 49-50.
- [5] Weinberg W. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg.* 1908; 64: 368–382.
- [6] Smith UM, Baldwim JT. Making Sense of Hardy-Weinberg Equilibrium. *The American Biology Teacher* (2015) 77 (8): 577–582.
- [7] Hartl DL, Clarke AG (2007). *Principles of population genetics.* Sunderland, MA: Sinauer. 2007.
- [8] Pierce, B. A. *Genetics: A conceptual approach.* New York: W.H. Freeman.
- [9] Ribeiro HL Jr, de Oliveira RT, Maia AR, et al. Polymorphisms of DNA repair genes are related to the pathogenesis of myelodysplastic syndrome. *Hematol Oncol.* 2015;33(4):220-228.
- [10] 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.

*Section name: Title (Arial, 10, italic, left, single paragraph)*

[11] Anwar YAS, Senam S, Laksono EW. The use of orientation / decision / do / discuss / reflect (OD3R) method to increase critical thinking skill and practical skill in biochemistry learning. *Biochem Mol Biol Educ.* 2018;46:107-113.

[12] Silverstein TP. CoViD-19 epidemic follows the "kinetics" of enzymes with cooperative substrate binding [published online ahead of print, 2020 Jun 30]. *Biochem Mol Biol Educ.*