

# Graphical tools for assessing Hardy-Weinberg equilibrium for bi-allelic genetic markers

Jan Graffelman<sup>1,\*</sup>

1. Department of Statistics and Operations Research, Universitat Politècnica de Catalunya, Barcelona, Spain

\*Contact author: [jan.graffelman@upc.edu](mailto:jan.graffelman@upc.edu)

**Keywords:** Exact test, chi-square test, ternary plot, acceptance region, log-ratio transformation.

If there are no disturbing forces (migration, mutation, selection, etc.), then the genotype frequencies AA, AB, and BB of a bi-allelic genetic marker are expected to occur with frequencies  $p^2$ ,  $2pq$  and  $q^2$  respectively, where  $p$  is the allele frequency of A, and  $q = 1 - p$  is the allele frequency of B. This basic principle in genetics is known as Hardy-Weinberg equilibrium (HWE). In genetics, markers are statistically tested prior to their subsequent use in for instance, association studies. A significant deviation from HWE can be ascribed to many factors. Gross genotyping error can cause deviation from HWE and forms one of the reasons to test markers. The HapMap project (2007) excludes genetic markers that have a  $p$ -value below 0.001 in an exact test for HWE.

Several statistical tests are available to check markers for HWE: the  $\chi^2$  test, the exact test, the likelihood ratio test and Bayesian tests. Weir (1996, Chapter 4) gives an overview of several tests for HWE. Several of these tests are implemented in The *R* package **HardyWeinberg**. The package is currently being extended with routines for power calculation and Bayesian tests. Most testing for HWE is done in an entirely numerical manner. Graffelman and Morales (2008) showed that testing can be done graphically inside a ternary plot representation. The HWE law defines a parabola inside a ternary plot of the three genotype frequencies. An acceptance region for the different tests can be drawn around the HWE parabola. The graphical testing facilities are implemented in the function `HWTernaryPlot` of the *R* package **HardyWeinberg**. The obtained graphics are very informative because they display genotype frequencies, allele frequencies and the (statistical) equilibrium condition in a single graph.

The genotype counts can also be treated as three-way compositions that sum up to 1. Tools from the field of compositional data analysis yield alternative graphical representations of the HWE law. Several different log-ratio transformations (additive, centred or isometric) of the genotype counts can be used. Graphically, the HWE parabola in the ternary plot is converted into a straight line in log-ratio coordinates. Functions `HWalrPlot`, `HWclrPlot` and `HWilrPlot` of the **HardyWeinberg** package can be used to create the log-ratio plots for HWE. Zero genotype counts are a problem for the log-ratio approach, and some adjustment for zero counts is necessary.

The proposed graphics can be used to assess HWE for multiple samples that are all typed for one marker (e.g. cases and controls), but can also be used to screen many markers simultaneously. This way, a whole genomic region can be screened for an excess or lack of heterozygotes. Several genetic data sets will be used in the talk to illustrate the proposed graphs.

## References

Graffelman, J. and J. Morales-Camarena (2008). Graphical tests for Hardy-Weinberg equilibrium based on the ternary plot. *Human Heredity* 65(2), 77–84. DOI: 10.1159/000108939.

The International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–861.

Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, Massachusetts: Sinauer Associates.