

## Making Sense of Hardy-Weinberg Equilibrium

One of the more difficult topics for introductory biology students to understand and for teachers to teach is the Hardy-Weinberg equilibrium (H-W eq) principle. One reason for this difficulty is the students' mathematical background. More problematic than lack of manipulative skill<sup>1</sup> is the difficulty of understanding why the principle is true and understanding how the principle applies to specific populations or more importantly, the value of its application.



Figure 1 Wilhelm Weinberg (above) and G. H. Hardy (used with permission)

The H-W eq principle is, of course, the cornerstone of introductory population genetics and is therefore an important part of understanding evolution, as recognized in most science standards documents. For example, “adaptation by natural selection” is one of the “Core Ideas” in the Next Generation Science Standards (NGSS), focusing on how the distribution of traits in a population changes (NGSS, 2014). Likewise, the NGSS targets the ability to “use mathematical representations to support scientific conclusions and design solutions.”

Some NGSS objectives<sup>2</sup> address H-W eq more explicitly:

---

<sup>1 1</sup> The probability notions involved are at Grade 7 in the Common Core State Standards (Math) (CCSI, 2014). (The algebraic manipulations have always been in the Algebra I standards.)

<sup>2</sup> A recent review of the NGSS completed by the American Society of Human Genetics (M. Dougherty, personal communication; May 15, 2014) shows that items related to the H-W eq are included only at the middle school level and not at the high school level. (Likewise, Mendel's laws of segregation and independent assortment, inheritance of complex traits, etc. are not included at the high school level!)

“Students who demonstrate understanding can use mathematical representations to support explanations of how natural selection may lead to increases and decreases of specific traits in populations over time. (MS-LS4-6)

Given its importance and the well-recognized teaching and learning difficulty, in accord with these standards, in this article we explain basic concepts of the H-W eq, emphasizing distinctions that are sometimes ignored at the cost of coherent understanding.

I. What is Hardy-Weinberg equilibrium?

In general, a system is said to be in equilibrium if all competing influences are balanced. In the body, for example, we speak of homeostasis as the ability to maintain the internal equilibrium regardless of changes in the environment (*e.g.*, temperature). A basic precept of evolution is that under certain conditions the frequencies of genotypes (and therefore of alleles) do not change (they remain at “equilibrium”), but in the absence of these conditions the frequencies do change.

**Definition: A population is in Hardy-Weinberg equilibrium if the genotype frequencies are the same in each generation.**

Consider the simplest situation of a monogenic Mendelian trait: a pair of alleles, one dominant  $A$  and the other recessive  $a$ , within a population of  $n$  individuals. The frequency of the  $A$  allele is the number of  $A$  alleles divided by total number of alleles at this locus within the population (two times the number of individuals). For example, if  $n$  is 4000 and 2000 of the alleles of this locus in a population are  $a$ , the frequency of the  $A$  allele is  $3/4$  and that of the  $a$  allele (*i.e.*, the remaining non-  $A$  alleles) is therefore  $1/4$ . These would be the allele frequencies if there are 1000  $aa$  and 3000  $AA$  individuals (or 500  $aa$ , 1000  $Aa$ , and 2500  $AA$ ). But if these individuals randomly mate,<sup>3</sup> the next generation has the same allele frequency but a genotype frequency of  $1/16$   $aa$ ,  $3/8$   $Aa$ , and  $9/16$   $AA$ . If a finite population is at H-W eq, however, BOTH the genotype and the allele frequencies will be essentially the same in subsequent generations.

## 2. Requirements for Hardy-Weinberg Equilibrium

In Section 3, we will demonstrate that a population is in Hardy-Weinberg equilibrium if the following conditions hold (with respect to a particular gene):

1. There is no migration (“gene flow”) in or out of the population.
2. Natural selection is not occurring.
3. Mutation is not occurring.
4. Each *member* of the population is equally likely to breed.<sup>4</sup>
5. The population is infinitely large.

---

<sup>3</sup> This scenario occurs whenever two pure-bred homozygous populations are allowed to randomly interbreed.

<sup>4</sup> Often “random mating” is used to refer to both condition 4) and 6). “Random mating” means that the frequency of mating of an individual or of any pair of individuals does not depend on the genotype.

6. (Full random mating) Each *pair* from the population is equally likely to breed (Females of a species often prefer males with certain traits. Examples originally identified by Darwin include peacock feather displays, antlers in deer, and the manes of lions. ).<sup>5</sup>

**Observation 1. As long as a population satisfies biological conditions 1-5 the allele frequencies ( $p$  and  $q$ ) are the same in each generation.**

Why is this so? Conditions 1 and 2 guarantee that there is no change in the allele frequencies between the birth and maturity of the next generation; there are no unaccounted forces that would change the allele frequencies (*i.e.*, one phenotype is not more fit than the other). Conditions 3 and 4 guarantee that at birth the pool of alleles in the next generation is the same as in the current generation; mating just reshuffles the alleles; the allele frequencies remain the same.

The population needs to be infinite to guarantee that the frequencies remain **exactly**  $p$  and  $q$ . The probabilities  $p$  and  $q$  represent the averages over many trials and so it will only be approximate in a particular trial on a finite population. (A change in **allele frequencies** can be caused by “genetic drift” or a “bottleneck.”) Of course, no

---

<sup>5</sup> This condition implies that the number of males and females is equal.

population is truly infinite; therefore Condition 5 can never be strictly met. If a population is large enough, however, it is considered to be “effectively infinite.”<sup>6</sup>

Likewise, the other assumptions are rarely if ever true of a given population (e.g., the mutation rate is rarely zero). H-W eq is largely a theoretical state, like a frictionless plane, an absolute vacuum, or travel at the speed of light. As with those concepts in physics, it nevertheless plays a fundamental conceptual role in biology and is a valuable tool for understanding evolution. The H-W principle has many applications in the modern practice of evolutionary biology where its value often lies in identifying when H-W eq does NOT exist and then determining which factor (or combination of factors) most likely explains the observed change in allele/phenotype frequencies over time, *i.e.*, what are the drivers of evolution in this population. Also, for evolution-neutral mutations the population is often close enough to equilibrium to provide a tool for comparing their frequencies against the frequencies of linked genes of interest to determine how close the latter are to H-W eq (Chen, 2010).

### 3. History and derivation of the Hardy Weinberg Principle



Figure 2 R. C. Punnett  
(used with permission)

Building on the work of other biologists and mathematicians, in 1908 Wilhelm Weinberg (1862-1937), a German obstetrician-gynecologist, and G. H. Hardy (1877-1947), a leading mathematician of his day, independently demonstrated the conditions required for genotype equilibrium. In a famous lecture earlier that same year, R.C. Punnett had combined

we mean within the limits of our ability to count; sampling with replacement would yield the same results as sampling without replacement.



Figure 3 Udny Yule (used with permission)

Mendelian genetics with natural selection (Edwards, 2008).

After the talk Udny Yule, one of the founders of modern statistics, asked whether a dominant/recessive allele pair would not eventually achieve a 3:1 ratio (Yule, 1908). (He was apparently assuming an initial frequency of 1/2 for each allele.)

In 1902 Yule had in fact shown that genotype frequencies would remain constant under random mating in the special case of a simple Mendelian trait with only two alleles of equal frequency ( $p = q = 1/2$ ), although he failed to recognize that this fact holds for all initial allele frequencies (Edwards, 2008). Punnett's (1908) response though not entirely apt, was a suggestion that a dominant allele should eventually drive the recessive out (which is not the case). Punnett later asked his friend Hardy about this question prompting the analysis we now describe.

The first of two contributions of Hardy and Weinberg was to remove the restriction that  $p = q = 1/2$ . Let  $A$  and  $a$  represent the two possible alleles of a simple Mendelian trait; let  $p$  and  $q$  represent the frequencies of  $A$  and  $a$ , respectively in the parent generation. Hardy and Weinberg argued that if every pair of individuals is equally likely to mate (Condition 6), then the frequencies of the three possible genotypes at birth can be determined by thinking of a Punnett square but labeling the rows and columns with allele frequencies instead of alleles:

	$p$	$q$
$p$	$p^2$	$pq$
$q$	$pq$	$q^2$

This Punnett square demonstrates the **Crucial H-W Insight**: Under fully random mating, the frequency of AA homozygotes in the next generation is  $p^2$ , of heterozygotes is  $2pq$ , and of aa homozygotes is  $q^2$ .

Now we can deduce the **Hardy-Weinberg Equilibrium Principle**: Consider a population satisfying biological conditions 1-6. If in a certain generation the allele frequencies are  $p$  and  $q$  and the genotype frequencies are  $p^2$ ,  $2pq$ ,  $q^2$ , then BOTH the genotype and allele frequencies remain the same for as many generations as conditions 1-6 continue to hold.

Here is why this principle holds. When a population satisfies conditions 1-5, Observation 1 ensures that allele frequencies will remain unchanged in every succeeding generation satisfying those conditions. Applying condition 6 and the crucial H-W insight, in each generation, therefore,<sup>7</sup> the genotype frequencies are  $p^2$ ,  $2pq$ , and  $q^2$ .

This is the genius of the H-W principle. After one generation of fully random mating, BOTH the genotype and allele frequencies are fixed until one of Conditions 1-6 is violated.

---

<sup>7</sup> Note that conditions 1-6 could hold without the genotype frequencies being  $p^2$ ,  $2pq$ , and  $q^2$ . For example, a breeder could begin a new population of  $p$  per cent AA individuals and  $q$  per cent aa individuals and allows fully random mating. In the first generation the AA frequency will be  $p$ , but ever after it will be  $p^2$ . In that case, the population enters Hardy Weinberg equilibrium in the second generation.

In any population where all three genotypes can be identified, (incomplete/co-dominance [e.g. red, pink and white flowers] in 1908 or alleles identified by DNA analysis in 2008), regardless of whether the population is at H-W eq or not, Mendelian genetics allows us to determine the allele frequencies from the genotype frequencies. Namely, in ANY population the frequency of the dominant allele of a Mendelian pair is the sum of two times the number of dominant homozygotes ( $AA$ ) + the number of heterozygotes ( $Aa$ ) divided by the total number of these alleles.

To summarize,

a) allele frequencies can always be computed from the genotype frequencies in the same generation (if all genotypes can be identified), but not vice versa.

If the population is in H-W equilibrium,

b) genotype frequencies in the current or the next generation can be computed from the current allele frequencies.

We have not yet mentioned the “H-W equations.” In fact, Hardy and Weinberg NEVER mentioned them! These two equations are widely used in biology teaching, but all too often they are used as a mathematics exercise that does not promote understanding.

**Equation 1:**  $p + q = 1$

is true for any monogenic Mendelian trait because there are only two outcomes.

Squaring equation 1 yields:



**Equation 2:**  $p^2 + 2pq + q^2 = 1$

So equation 2 simply follows mathematically from Equation 1. There is no assumption about random mating and no other biological assumption in the step from Equation 1 to Equation 2.<sup>8</sup>

Interestingly, using only these formulae, we can determine whether H-W eq exists in a SINGLE generation of a population by determining whether the genotype distribution matches that predicted from the allele distribution, but this requires that **both** the allele frequencies and the genotype frequencies are known. Hardy (1908) provided an ingenious way to determine whether H-W eq exists in a single generation given only the genotype frequencies. A short account of Hardy's proof in modern language accessible to advanced students along with several other proofs of the HW-principle appears at (Baldwin, 2014).

## 5. Five Example Problems

The following textbook problems are built on the assumption that, if a population is in H-W eq (which is often a dubious assumption), then it is possible to calculate the allele frequencies from the frequency of the homozygous recessives (which can be found by observation). That is, under the assumption of H-W eq, if  $b$  percent of the population is homozygous, then  $q^2 = b$ , so  $q = \sqrt{b}$ . And because  $p = 1 - q$ , then  $p = 1 - \sqrt{b}$ .

---

<sup>8</sup> There is a noteworthy potential misunderstanding in the two standard H-W eq equations. The terms in Equation 1 ( $p$  and  $q$ ) represent ALLELE frequencies, but the terms in Equation 2 ( $p^2$ ,  $2pq$ , and  $q^2$ ) represent GENOTYPE frequencies. Biologists and biology teachers implicitly understand this biological sleight of hand; making it explicit for students can reduce confusion.

Problem 1. 'The data below demonstrate the frequency of tasters and non-tasters in an isolated population at Hardy-Weinberg equilibrium. The allele for non-tasters is recessive. How many of the tasters in the population are heterozygous for tasting?'

tasters	non-tasters
8235	4325

Solution provided:<sup>9</sup> "An acceptable answer would be any number in the range of 6030-6156, depending on how the students rounded the variables in the Hardy-Weinberg equation."

Comment on Problem 1. This is a standard H-W eq problem. The frequency of non-tasters (homozygous recessive individuals— $q^2$ ) is 4325 / 12560. Assuming that the population is in Hardy-Weinberg equilibrium, the frequency of the homozygous recessive allele ( $q$ ) is computed as the square root of the frequency of homozygous individuals ( $q^2$ ):  $q = \sqrt{(4325/12560)}$ ; therefore  $q = 0.58$  and  $p = 1 - q = 0.42$ . Then the frequency of the heterozygotes ( $2pq$ ) = 2 (.58) (.42) = 0.49. This yields the number of heterozygotes as  $.49 \times 12560 = 6119$ .

Problem 2 (Trout, 2012). 'The ability to taste PTC is due to a single dominant allele  $T$ . You sampled 215 individuals and determined that 150 could detect the bitter taste of PTC and 65 could not. Calculate the following frequencies. a. The frequency of the recessive allele. b. The frequency of the dominant allele. c. The frequency of the heterozygous individuals.'

Solution 2 (from the teacher's guide):

---

<sup>9</sup> This is a sample problem for the revised AP biology exam, 2010 presented by Jim Pellegrino at the Chicago Symposium on Teaching and Learning.

a. The frequency of the recessive allele.

$$q^2 = 65 \div 215$$

$$q^2 = 0.30$$

$$q = 0.55$$

b. The frequency of the dominant allele.

$$p = 1 - q \quad p = 1 - 0.55 \quad p = 0.45$$

c. The frequency of the heterozygous individuals.

$$\chi = 2pq \quad \chi = 2(0.45 \times 0.55) \quad \chi = 0.495$$

Comments on Problems 1 and 2. Both problems focus on making calculations that students can do without understanding what H-W eq is. In Problem 1 H-W eq is explicitly assumed so that the problem is technically correct. But it doesn't say what chemical was being tasted (presumably PTC), so it doesn't ask students whether HW-eq conditions COULD be met for this trait. And students can only solve Problem 2 by assuming H-W eq, which is not justified.

The ability to taste PTC (phenylthiocarbamide), a bitter substance that cannot be tasted by some individuals, is frequently used in H-W eq problems, likely because it is assumed to be selected neither for nor against because PTC does not occur in nature. Thus, the student is expected to deduce (or more likely assume) that the H-W conditions apply. Teachers and textbooks, however, rarely make this reasoning explicit, leaving students with the misperception that understanding PTC tasting is just a game or a puzzle which likely seems unimportant to them because it doesn't relate to their daily lives.

In fact, recent research has shown that the ability to taste PTC is strongly correlated with the ability to taste other bitter substances that DO occur naturally, many

of which are toxins. In fact, humans have about 30 genes that code for bitter taste receptors, allowing people to taste a wide variety of bitter substances (Genetic Science Learning Center, 2014). Thus, it seems likely that the ability to taste bitter substances (such as PTC) is positively selected for.

Problem 3 (K-State Parasitology Laboratory, 2000). "Sickle-cell anemia is an interesting genetic disease. Normal homozygous individuals (SS) have normal blood cells that are easily infected with the malarial parasite. Thus, many of these individuals become very ill from the parasite and many die. Individuals homozygous for the sickle-cell trait (ss) have red blood cells that readily collapse when deoxygenated. Although malaria cannot grow in these red blood cells, individuals often die because of the genetic defect.

However, individuals with the heterozygous condition (Ss) have some sickling of red blood cells, but generally not enough to cause mortality. In addition, malaria cannot survive well within these "partially defective" red blood cells. Thus, heterozygotes tend to survive better than either of the homozygous conditions. If 9% of an African population is born with a severe form of sickle-cell anemia (ss), what percentage of the population will be more resistant to malaria because they are heterozygous (Ss) for the sickle-cell gene?"

Solution to Problem 3 (from teacher's guide). "9% = 0.09 = ss =  $q^2$ . To find  $q$ , simply take the square root of 0.09 to get 0.3. Since  $p = 1 - 0.3$ , then  $p$  must equal 0.7.  $2pq = 2(0.7 \times 0.3) = 0.42$  or 42% of the population are heterozygotes (carriers)."

Comment on Problem 3. The problem solution above, which assumes H-W eq and that natural selection is not occurring with regard to this gene, contradicts the statement of

the problem, which notes selective pressures for one and against another of two blood cell phenotypes.

Problem 4. A more sophisticated version of this problem (Trout, 2012) states that sickle-cell disease affects approximately 9% of the African population and then asks the students to use the Hardy-Weinberg equations to calculate the predicted genotype frequencies. The students are then asked, 'Based on this analysis, is the African population in Hardy-Weinberg equilibrium? Justify your answer.'

Solution to Problem 4: (as provided in the teacher's guide): "No. Because the members of the population that contract sickle cell because they are homozygous recessive will likely die before reproducing, the frequency of alleles in the population is not stable. There is natural selection taking place."

Comments on Problem 4: Although this problem instructs students to use the H-W equations, again the known effects of natural selection at this locus mean that H-W eq is impossible. The problem therefore asks for what is in fact a meaningless calculation. Then it asks students to answer a question that demonstrates that the computation was meaningless but does not ask them to recognize that it was meaningless! The H-W eq-based frequencies are irrelevant. In fact, the decision of whether the population is in H-W eq is NOT "based on this analysis." (Note: The response also states, "contract sickle cell". As an inborn, genomic error, one does not "contract" this disease.)

A second issue that arises in problems about sickle cell anemia is that two opposing selective pressures are at work—a positive selection for heterozygosity and a negative selection against affected homozygotes. (Such a situation can produce balanced polymorphism equilibrium (of alleles) but not H-W equilibrium because the

genotype frequency will change.) This makes sickle cell anemia a poor choice for the context of most introductory-level H-W problems.

A third objection to this entire analysis of sickle-cell anemia is that Hardy-Weinberg analysis requires addressing the idea of a single “generation.” This is a very difficult concept to apply to human populations without careful data collection because people living at any one time can represent three or four generations.

A more conceptual shortcoming of all these problems is that there is no readily apparent value to the calculation. This is the “So what? Who cares?” question. How might such a calculation be used to answer a research question or be applied to a case that is at least interesting to the students? When students see the utility of such calculations or find the case interesting, they are more likely to engage in this learning.

Here is a homework problem that addresses some of these issues:

Problem 5. People who are homozygous for (have two copies of) a certain 32-bp deletion mutation in a gene known as CCR5 are known to be largely resistant to HIV infection. (CCR5 is the main co-receptor molecule that allows the virus to attach to certain white blood cells and enter them, establishing an infection: Jones, Maguire, and Davenport, 2011.) In a study of 1318 random Caucasians of childbearing age in the US 1102 individuals were found to be homozygotes free of this deletion(Glass et al, 2006),. Assume that the US Caucasian population is at H-W equilibrium at this locus.

Susan is a Caucasian American woman at increased risk of HIV infection because she has multiple sex partners.

1. What is the probability that Susan has little reason to worry about HIV infection, *i.e.*, that she is homozygous for the deletion?
2. What is the predicted frequency of US Caucasians of this age who are “carriers” of the protective deletion? How many heterozygotes would be expected in this sample?
3. Before HIV appeared, would you have expected the population to have been at H-W equilibrium at this locus? Why or why not? State your assumptions. In the absence of effective HIV treatments, what would you expect to happen to the allele frequencies over time? How would you expect the allele frequencies to change over time once effective HIV treatment was in use? Why? How would your answer change if the HIV treatment was only effective for people past child-bearing age?

Answer:

1. Let  $p$  be the frequency of the allele without the deletion.  $p^2 = 1102/1318 = .836$

Since we assumed HW-eq,  $p = \sqrt{0.836} = 0.914$ . Since there are only two alleles,  $q$  the frequency of the allele with the deletion, satisfies

$$q = 1 - p = 1 - 0.914 = 0.086$$

$q^2 = 0.007$ , so Susan has a 0.7 % chance of being at lower risk.

2. Again assuming H-W eq,  $2pq = (.914) (.086) = 0.157$  .

$$0.157 \times 1318 = 207$$

3. Before HIV appeared, the population was likely at H-W equilibrium at this locus.

This assumes that the deletion was selection neutral. With the appearance of HIV and in the absence of effective treatment, we would expect that the frequency of people without the homozygous deletion ( $p^2$  and  $2pq$ ) should decrease, shifting the population out of H-W equilibrium. The advent of an effective treatment should move the allele frequency distribution back toward equilibrium. A treatment that only has an effect after child-bearing age would have effects on the allele distribution similar to no treatment because, for a mutation to have a natural selection effect, it must affect reproductive success.

### 3. Summary and Implications

Many textbook and internet H-W eq problems have substantial shortcomings. They may:

1. Fail to focus on understanding,
2. Be unclear about which specific gene is involved in the problem (e.g., “tasters”),
3. Be unclear about the characteristics of the population being studied (especially size),
4. Assume that the H-W eq exists in the population but not say so explicitly,
5. Assume students have certain biological knowledge about the gene involved in the problem,
6. Make assumptions that are contradicted in fact or are likely impossible,
7. Ask for judgments about populations that are constituted by members of multiple generations,
8. Ask for calculations that are meaningless in the given context, or



9. Ask for solutions that have no apparent value/are not related to genuine research questions (fail the “So what” test).

**Put simply, a population is in Hardy-Weinberg equilibrium if the genotype frequencies are the same in each generation.** This equilibrium requires a set of conditions that ensure that there are no unaccounted forces that would change the allele frequencies. For such a population, then, the genotype frequencies in the current or the next generation can be computed from the current allele frequencies. Focusing on this level of understanding for students as well as avoiding confusing misstatements and flawed problems is the primary key to effective teaching about the H-W equilibrium.

Our next paper on the H-W eq will provide more practical suggestions for how to implement these ideas in introductory biology instruction.

## REFERENCES

- Baldwin, John T. *The Hardy Criterion* (20 June, 2014) <  
<http://homepages.math.uic.edu/~jbaldwin/pub/Hardycriterion.pdf>> (accessed August 30, 2014).
- CCSI (2014). “Common Core Science Standards: Math”  
<<http://www.corestandards.org/Math/>> (accessed August 30, 2014).
- Chen, John J. “The Hardy-Weinberg Principle and Its Applications in Modern Population Genetics.” *Frontiers of Biology* 5(2010): 348-53.

Edwards, A. W. F. "G. H. Hardy, (1908) and Hardy–Weinberg Equilibrium." *Genetics* 179(2008): 1143–50.

Genetic Science Learning Center, "*PTC: Genes and Bitter Taste*," Learn.Genetics, 22 June 2014, <<http://learn.genetics.utah.edu/content/inheritance/ptc/>> (accessed August 30, 2014).

Glass, W. G., McDermott, D. H., Lim, J. K., Lekhong, S., Frank, W. A., Pape, J., Cheshier, R. C. and Murphy, P. M. (2006). CCR5 deficiency increases risk of symptomatic West Nile virus infection. *Journal of Experimental Medicine*, 203(1):35-40.

Hardy, G. W. "*Mendelian proportions in a mixed population*." *Science*, 28(1908):49–50, Holsinger, Kent. *Lecture notes in Population Genetics*, 2012.

<<http://darwin.eeb.uconn.edu/eeb348/lecture-notes/book.pdf>> (accessed August 30, 2014).

Jones, K. L., Maguire, J. J., and A. P. Davenport (2011). Chemokine receptor CCR5: from AIDS to atherosclerosis. *British Journal of Pharmacology*, 162(7): 1476-5381.

K-State Parasitology Laboratory, 21 August, 2000. Population Genetics and the Hardy-Weinberg Law <<http://www.k-state.edu/parasitology/biology198/hardwein.html>> (accessed 30 August 2014).

NGSS Lead States (2013). "Next Generation Science Standards: For States, By States" Achieve, Inc. <<http://www.nextgenscience.org/next-generation-science-standards>> (accessed August 30, 2014).

Punnett, R. C. "Mendelism in Relation to Disease", *Proceedings of the Royal Society of Medicine* 1(1908): 135-61

Trout, L., ed., *POGIL Activities for AP Biology* (Batavia, IL: Flinn Scientific, 2012).

Yule, G. U., "Contribution," *Proceedings of the Royal Society of Medicine* 1(1908): 165.