What Makes the Hardy-Weinberg Equation So Useful?*

Laura Martin

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 2.0†

The Hardy-Weinberg Equation

- p2 (**AA**) + 2pq (**Aa**) + q2 (**aa**) = 1

is incredibly useful because it describes mathematically the genetic product of a population in which all individuals are equally likely to survive and to produce surviving offspring. Specifically, it calculates the genotype frequencies that will be observed in a population that is not evolving. This information functions as a null hypothesis or standard against which we can judge if a population is evolving.

1 Summarizing the Concepts on which Hardy-Weinberg is Based

To understand how Hardy-Weinberg enables this, let's review what we have learned:

- When agents of evolution favor the survival and/or reproduction of some individuals (and their alleles) over others, the population's allele frequencies change over time and the population is said to evolve.
- When all individuals are equally likely to survive and to produce offspring that survive, allele frequencies do not change from one generation to the next because alleles end up in fertilization events, and thus the offspring generation, in proportion to their relative commonness (frequency) in the parental generation.
- When alleles end up in fertilization events (and the offspring generation) in proportion to their relative commonness in the parental generation, **genotype** frequencies of the offspring generation can be determined from the **allele** frequencies of the parental generation using simple rules for calculating the probability of an event composed to two independent events.
- Specifically, these rules say that, if we let p equal the frequency of the **A** allele and q the frequency of the **a** allele in the parental generation, and thus p equals the probability of picking an **A** allele and q an **a** allele from the parental generation, the:
- frequency of the **AA** genotype in the offspring generation is equal to the probability of picking two **A** alleles from the parent generation or p x p = p2
- frequency of the **aa** genotype in the offspring generation is equal to the probability of picking two **a** alleles from the parent generation or q x q = q2
- frequency of the **Aa** genotype in the offspring generation is equal to two times the probability of picking an **A** and **a** allele from the parent generation or 2 x (p x q) = 2pq
- This relationship, which is often written as p2 (**AA**) + 2pq (**Aa**) + q2 (**aa**) = 1, is known as the Hardy-Weinberg equation. It tells us what genotype frequencies (p2, 2pq, q2) we will see in a population that is not evolving based on parental allele frequencies, p and q.

*Version 1.1: Oct 16, 2007 6:03 pm -0500
†http://creativecommons.org/licenses/by/2.0/

Finally, please recall that the Hardy-Weinberg equation only applies to a locus for which there are two alleles in a population. And that, although some individuals in the population will die or fail to reproduce, population allele frequencies will be maintained because these losses will remove alleles in proportion to their commonness in the population.

2 So what is the power of these concepts?

Notice that these principles apply to the production of a **single** generation. Thus, if an offspring generation is the product of a parental generation that was not subject to an agent of evolution, then the offspring generation must have the following two properties.

1. The allele frequencies must be the same as those of the parental generation.
2. Genotypes must occur with the frequencies predicted by Hardy-Weinberg: **AA** = p2, **aa** = q2 and **Aa**=2pq

Because of this, we can use information from a **single** offspring generation to test the hypothesis that individuals in the parental generation were all equally likely to survive and to produce surviving offspring and, therefore, that the population is not evolving with respect to this locus.

3 How does this work?

If parents are all equally likely to survive and to produce surviving offspring, then the allele frequencies of the offspring generation can be used as proxy for the allele frequencies of the parental generation. They should be identical if the no evolution condition is met.

Now we can use these hypothetical parental generation allele frequencies, that are based on the assumption of no evolution, in the Hardy-Weinberg equation to calculate the genotype frequencies we would see in the offspring generation if the population were not evolving with respect to these alleles.

Finally, we can draw a conclusion about whether a population is subject to an agent of evolution by comparing the genotype frequencies we calculated using Hardy-Weinberg to those we actually observe in the offspring generation.

To test your understanding of these relationships, answer the questions below.

Exercise 1 ***(Solution on p. 5.)***

What would you conclude about a) the actual similarity of the parental and offspring generation allele frequencies and b) whether the population was evolving if

1. The observed offspring genotype frequencies equaled those predicted by Hardy-Weinberg?
2. The observed offspring genotype frequencies did not equal those predicted by Hardy-Weinberg?

Now, let's apply this logic to an actual problem.

Exercise 2 ***(Solution on p. 5.)***

In the mid-1990's, researchers discovered that despite repeated exposure to HIV-1, a strain of the Human immunodeficiency virus (HIV), some individuals remained uninfected (Samson *et al.*, 1996).

Subsequent investigation revealed the existence of an allele that confers immunity to HIV-1 infection in homozygotes. This allele, known as delta-*ccr-5* or CCR5-delta-32, is a mutant version of the cell-surface receptor protein CCR-5. It inhibits HIV infection because it codes for a form of the CCR-5 receptor to which HIV-1 viruses are unable to bind and, thus, to enter white blood cells and thereby establish an infection (Samson *et al.*, 1996).

For reasons unrelated to its effects on susceptibility to HIV-1 infection, this allele is found most commonly in caucasian Europeans and is absent or virtually absent from African, Asian, Middle Eastern and American Indian populations (Galvani and Slatkin, 2003). Table 1 contains original

data from Samson *et al.* (1996) documenting the genotypes of 704 caucasian Europeans. Use these data to answer the questions that follow.

Table 1: The number of individuals homozygous for either the CCR-5 or ccr-5 allele or heterozygous for these two alleles in a sample of 704 caucasian Europeans.

Genotype	**Number of Individuals**
CCR-5/CCR-5	582
CCR-5/*ccr-5*	114
ccr-5/*ccr-5*	8
Total	704

Table 1

Is this population of Europeans evolving with respect to this allele? How do you know? To answer this question, work your way through the steps below.

1. Review the question. Using information from earlier in this module or in related modules, complete the chart below on a piece of paper.

What do I need to know to answer this question?	How do I get this information?

Table 2

Exercise 3 ***(Solution on p. 5.)***

Samson *et al.* (1996) determined that CCR-5 allele occurred with a frequency of 0.91 and the *ccr-5* allele with a frequency 0.09. Do you agree?

Exercise 4 ***(Solution on p. 5.)***

Samson *et al.* (1996) also concluded that '*The genotype frequencies observed in this population were not significantly different from the expected Hardy-Weinberg distribution...*' (Samson *et al.*, 1996, p.724.) Do you agree with their conclusion? Why or why not? Please explain using data to support your conclusion.

Exercise 5 ***(Solution on p. 6.)***

Finally, we are ready to answer our original question. Is this population of caucasian Europeans evolving with respect to this locus? Yes or no? Please explain.

Exercise 6 ***(Solution on p. 6.)***

Are you confident that this result is probably true of the caucasian European population at large? Why or why not? Please explain.

If you are not confident, please describe one activity you could undertake to improve your confidence.

Exercise 7 ***(Solution on p. 6.)***

Imagine Europeans were not careful to practice safe sex, use clean needles, and maintain a virus-free blood supply causing exposure to the virus and infection rates to increase dramatically. Predict what would happen, if anything, to the frequency of these two alleles over time? Why? Please explain.

Exercise 8 ***(Solution on p. 6.)***

It turns out that European caucasian heterozygotes for the CCR-5 locus have lower infection rates and, when infected, progress more slowly to AIDS, the syndrome that ultimately kills HIV infected individuals, than CCR-5/CCR-5 homozygotes (Galvani and Slatkin, 2003).

Under a scenario in which exposure to HIV-1 and actual infection rates soar in Europe, what do you think would happen, if anything, to the frequency with which these CCR-5/CCR-5, CCR-5/*ccr-5* and *ccr-5*/*ccr-5* genotypes occur in this population? Why? Please explain.

Works Cited

- Galvani, A.P. and M. Slaktin. 2003. Evaluating plagues and small pox as historical selective pressures for the *CRR5 delta-32* HIV-resistance allele. *Proceedings of the National Academy of Sciences.* 100:15276-15279.
- Samson, M., F. Libert, B.J. Doranz, *et al.*. 1996. Resistance to HIV-1 in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature.* 382:722-725.
- Schliekelman, P., C. Garner and M. Slatkin. 2001. Natural selection and resistance to HIV. *Nature.* 411:545

Solutions to Exercises in this Module

Solution to Exercise (p. 2)

1. The parental generations allele frequencies must have been identical to those of the offspring, therefore the population is not evolving with respect to these alleles.
2. The offspring generation allele frequencies, which you used as a proxy for the parental allele frequencies, cannot be the same as those of the actual parental generation and therefore the population is evolving with respect to these alleles. This is true because actual offspring genotype frequencies did not equal those predicted by Hardy-Weinberg. The follow-up question would then be: what agent of evolution is responsible for this departure from genetic equilibrium?

Solution to Exercise (p. 2)

To answer this question, you need to know

- the frequency with which the CCR-5 and *ccr-5* alleles occur in this population
- the genotype frequencies you would see if this population was not evolving with respect to these alleles
- the actual genotype frequencies observed in this population

To get this information, you need to

- calculate allele frequencies from the genotype frequencies provided in Table 1.
- use these frequencies as a proxy for the allele frequencies of the generation that produced this population and calculate, via the Hardy-Weinberg equation, the genotype frequencies you would see if this population is not evolving with respect to these alleles.
- calculate the actual genotype frequencies from the data provided in Table 1.
- compare the expected versus actual genotype frequencies and draw a conclusion about whether this population is evolving with respect to these alleles.

Complete these steps and move on to the following problems to confirm your answers.

Solution to Exercise (p. 3)

To answer this question, we need to determine the total number of alleles in the population and what fraction of that total is due to the CCR-5 and *ccr-5* alleles respectively.

Since each of the 704 individuals in this sample has two alleles for this locus, one per chromosome, we are working with a total of 1408 alleles.

Of these 704 individuals, 582 have two copies of the CCR-5 allele and another 114 have one copy for a total of 1278 (2 x 582 + 114) CCR-5 alleles. Therefore, 1278 of 1408 alleles are CCR-5 alleles, a value equal to a frequency of 0.91 or 91% (1278/1408 x 100).

Now that we have calculated the frequency of the CCR-5 allele, there are two ways to confirm that the *ccr-5* allele occurs with a frequency equal to 0.09. The first is to use the relationship p + q = 1. If we set p equal to 0.91, the frequency of the CCR-5 allele, then q, the frequency of the *ccr-5*, must equal 0.09 because q = 1 - p.

We can also confirm this calculation using the genotype data from Table 1. If 114 individuals have one copy of the *ccr-5* allele and another 8 have two copies, there are 130 copies (2 x 8 +114) of the *ccr-5* allele in this population of 1408 alleles for a frequency of 130/1408 or 0.09, equivalent to 9% of the allele population.

Solution to Exercise (p. 3)

To answer this, we must calculate the genotype frequencies we would see if this population is not evolving with respect to these alleles, the actual genotype frequencies observed in this population and finally, compare these two frequencies and draw a conclusion.

If the population is not evolving with respect to these alleles, the population of parents that produced these 704 offspring and the population of offspring themselves must have the same allele frequencies. Therefore, we can use the allele frequencies we calculated above and the Hardy-Weinberg equation to determine the genotype frequencies we would see in this offspring population under the assumption of no evolution.

According to the Hardy-Weinberg equation, the frequency of the

- CCR-5/CCR-5 genotype is equal to the square of the frequency with which this allele occurs or p2 = 0.91 x 0.91 = 0.83.
- *ccr-5*/*ccr-5* genotype is equal to the square of the frequency with which this allele occurs or q2 = 0.09 x 0.09 = 0.008.
- CCR-5/*ccr-5* genotype is equal to the two times the product of frequency with which of these two alleles occur or 2pq = 2(0.91 x 0.09) = 0.16.

Thus, if this population is not evolving, 83%, 0.8% and 16% of these 704 offspring should exhibit the CCR-5/CCR-5, *ccr-5*/*ccr-5*, and CCR-5/*ccr-5* genotypes respectively. We can confirm our mathematics and that we haven't missed any potential offspring genotypes by checking that these three frequencies sum to 1 as they do (0.83 + 0.16 = 0.008 = 1.0).

We now need to calculate the actual frequency with which these genotypes occur in this population of 704 offspring (and therefore 704 genotypes). From Table 1:

- 582 of 704 genotypes are CCR-5/CCR-5 for a frequency of 582/704=0.83.
- 8 of 704 genotypes are *ccr-5*/*ccr-5* for a frequency of 8/704=0.011
- 114 of 704 genotypes are CCR-5/*ccr-5* for a frequency of 114/704=0.16

Thus, 83%, 1.1% and 16% of this population of 704 offspring exhibit the CCR-5/CCR-5, *ccr-5*/*ccr-5*, and CCR-5/*ccr-5* genotypes respectively. We can confirm our mathematics by checking that these three frequencies sum to 1 as they do (0.83 + 0.16 + 0.011 = 1.0).

We are now prepared to compare the genotype frequencies we actually see in this population of 704 individuals to those we would see if it is not evolving with respect to these alleles. Reviewing the summarized data below, we can see that both sets of genotype frequencies are nearly identical confirming Samson *et al.'s* (1996) conclusion.

- 83% of the genotypes are CRR-5/CRR-5 in both data sets
- 16% of the genotypes are CRR-5/*crr-5* in both data sets.
- 0.08 and 1.1% , or ∼0.1%, of the genotypes are *crr-5*/*crr-5* in the observed and expected data sets respectively.

Solution to Exercise (p. 3)
The nearly identical expected and observed genotype frequencies calculated in problem 4 suggest that this population of caucasian Europeans is not evolving with respect to this locus.

Solution to Exercise (p. 3)
The best answers to this question will discuss the lack of information describing the source of the 704 individuals genotyped for this study. Are they representative of caucasian Europeans at large or is this sample biased in some way? If you are not confident, your confidence might be improved by gathering background information on these individuals to assess their representativeness or by looking for or conducting a complimentary study to see if its results support these reported here.

Solution to Exercise (p. 3)
One would expect the frequency of the CCR-5 allele to decline and the frequency of the *ccr-5* allele to increase as individuals carrying the *ccr-5* allele resist infection and therefore leave behind more offspring relative to those who, without it, contract and succumb to the disease and leave behind fewer offspring as a result.

Significantly, something like this appears to be happening in Africa, where different variations in the CCR-5 receptor are conferring increased resistance to HIV infection and, in those infected, increased time to development of AIDS, the condition that ultimately kills HIV infected individuals. Individuals with these alleles are expected to produce 15 to 30% more children than individuals that do not carry these alleles (Schliekelman, Garner and Slatkin, 2001).

Solution to Exercise (p. 4)
Because individuals carrying at least one copy of the *ccr-5* allele would be less likely to both contract HIV-1 and, if infected, to progress more slowly to AIDS, these individuals would leave behind more offspring

than CCR-5/CCR-5 homozygotes. Thus, one would expect the frequency of CCR-5/CCR-5 homozygotes to decline and the frequency of *ccr-5*/*ccr-5* homozygotes and heterozygotes to increase over time at least initially.