

# Hardy-Weinberg Equilibrium as Foundational

Alan E. Stark and Eugene Seneta\*

School of Mathematics and Statistics FO7, University of Sydney, NSW, 2006, Australia

**Abstract:** The Hardy-Weinberg Principle explains how random mating can produce and maintain a population in equilibrium, that is: with constant genotypic proportions. The Hardy-Weinberg formula is in constant use as a basis for developing population genetics theory. Here we give a complete description of a model which can sustain equilibrium but with a general mating system, thereby giving a much broader basis on which to develop population genetics. It was S. N. Bernstein who first showed how Mendel's first law could be justified simply on the basis of observations of populations in equilibrium. We show how the model can be applied to exploring the change in incidence of a genetic disorder.

**Keywords:** Hardy-Weinberg Equilibrium, Non-random Mating, Mendel's First Law, Population Genetics Theory, Tay-Sachs disease.

## INTRODUCTION

Penrose gives the conventional introduction to the Hardy-Weinberg Principle of population genetics [1]. He uses it in his superb exposition of the then current knowledge of the aetiology of mental defect. He derives the Hardy-Weinberg formula giving the frequencies of genotypes in a population in equilibrium:

$$\{q^2, 2pq, p^2\}. \quad (1)$$

Penrose uses (1), *inter alia*, to explain how, in the case of recessive inheritance, the great majority of defectives would be derived from normal parents.

Hartl and Jones [2], Edwards [3] and Mayo [4] give detailed accounts of the Hardy-Weinberg law.

The usual textbook derivation of the (1) takes as foundational Mendel's coefficients of heredity (for example, that the relative frequency of *aa* offspring in *Aa* x *Aa* parental crossings is  $\frac{1}{4}$ ), together with random mating between individuals in the parental generation. Mendel's experimental approach had relied on the collation of outcomes of single pollination events.

Around 1920 S. N. Bernstein (see Seneta [6]), using axiomatic reasoning, derived Mendel's first law of inheritance (see Bernstein, 1942 for the mathematical derivation): that is, the structure of the Mendelian coefficients of heredity [5]. By contrast with Mendel's experimental approach, Bernstein *postulated* a *population* with many members which was *maintained in equilibrium* by a system of random mating. From

these basic assumptions Bernstein *derived* Mendel's coefficients of heredity which we put in matrix form in the section on notation below.

Stark & Seneta [7] put Bernstein's result in a more general context, and gave a more transparent derivation, by modifying the mating system. The population is still assumed to be maintained in equilibrium, which thus becomes foundational, but is not necessarily in the form of (1). But a more elaborate set of conditions is needed to specify the equilibrium than is required in a model of random mating. However, there is a considerable gain in generality and realism beyond that provided by the Hardy-Weinberg model.

The first purpose of this note is to delineate visually the region of equilibrium which varies according to the gene frequency. It is three-dimensional, with simpler form when the gene frequency is less than or equal to  $\frac{1}{4}$ . Thus, although the basic nature of the system is simple, its full description of the region is less so. This is an extension of work of Stark [8].

We begin by introducing notation which enables a convenient description of the mating system without resort to cumbersome tables. This is followed by descriptions of the admissible regions of equilibrium. Then we point out the highly restrictive nature of the model of random mating.

Our final section is a discussion of how Hardy-Weinberg equilibrium, as described by (1), can be achieved under mating which is not necessarily random. Our illustration is in terms of the Tay-Sachs disease. We demonstrate by this example how the model of equilibrium specified by (7) below can provide a broader basis for developing many of the concepts treated by Penrose [1].

\*Address correspondence to this author at the School of Mathematics and Statistics FO7, University of Sydney, NSW, 2006, Australia;  
Tel: +61-2-93515806; Fax: +61-2-93514534;  
E-mail: eseneta@maths.usyd.edu.au

**NOTATION**

We deal only with a single autosomal locus with two alleles  $U$  and  $T$  with frequencies in the population  $q$  and  $p$  ( $q+p=1$ ). Throughout  $q$  remains constant because this is guaranteed by the nature of the selected mating system. A set of frequencies of genotypes  $\{UU, UT, TT\}$  can be represented in terms of  $q$  and a measure of departure from Hardy-Weinberg (HW) form  $F$  as, say,  $\{q^2 + Fpq, 2pq - 2Fpq, p^2 + Fpq\}$ . These will vary according to  $F$  and will be denoted generally by  $\{f_0, f_1, f_2\}$ , ( $f_0 + f_1 + f_2 = 1$ ), that is  $f_0 = q^2 + Fpq$ , etc.

The population is maintained in discrete generations according to the mating scheme

$$\begin{bmatrix} UU \times UU & UU \times UT & UU \times TT \\ UT \times UU & UT \times UT & UT \times TT \\ TT \times UU & TT \times UT & TT \times TT \end{bmatrix} \quad (2)$$

with commensurate pairing frequencies given by the matrix

$$C = \begin{bmatrix} f_{00} & f_{01} & f_{02} \\ f_{10} & f_{11} & f_{12} \\ f_{20} & f_{21} & f_{22} \end{bmatrix} \quad (3)$$

$C$  is symmetric, that is  $f_{ij} = f_{ji}$ , with row and column sums  $\{f_0, f_1, f_2\}$ . This triplet of sums is the parental frequency distribution.

Below we use  $C$  in the extended (row vector) form

$$u' = \{f_{00}, f_{01}, f_{02}, f_{10}, f_{11}, f_{12}, f_{20}, f_{21}, f_{22}\} \quad (4)$$

To follow the progression of generations we need Mendel's coefficients of heredity given in matrix form by

$$M = \begin{bmatrix} 1 & 1/2 & 0 & 1/2 & 1/4 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1 & 1/2 & 1/2 & 1/2 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/4 & 1/2 & 0 & 1/2 & 1 \end{bmatrix} \quad (5)$$

Then the frequency distribution of juveniles is calculated from

$$j' = (Mu)'$$

which in detail is

$$j = \left\{ \begin{array}{l} f_{00} + \frac{f_{01} + f_{10}}{2} + \frac{f_{11}}{4}, \frac{f_{01}}{2} + f_{02} + \frac{f_{10} + f_{11} + f_{12}}{2} \\ + f_{20} + \frac{f_{21}}{2}, \frac{f_{11}}{4} + \frac{f_{12} + f_{21}}{2} + f_{22} \end{array} \right\} \quad (6)$$

The population is in equilibrium, that is: the distribution of juveniles is the same as that of adults, if and only if matrix  $C$  has, in addition to the properties given above, the special property

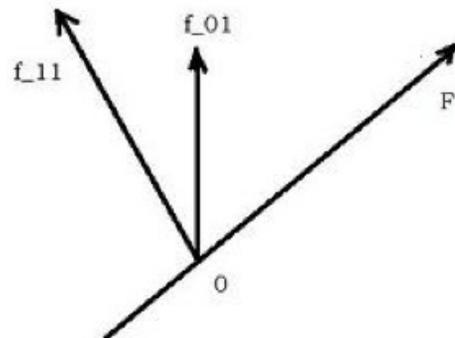
$$f_{11} = 4f_{02} = 4f_{20} \quad (7)$$

The notation used here is a modified version of that given in Stark & Seneta [7, 9].

Identity (7) allows for non-random mating (NRM) as well as random mating (RM).

**REGIONS OF ADMISSIBLE POINTS**

Mating matrix  $C$ , displayed in (3), is determined by combinations of  $q$ ,  $F$ ,  $f_{11}$ , and  $f_{01}$ . We first fix  $q$  then consider sets of  $F$ ,  $f_{11}$ , and  $f_{01}$  as points in 3-dimensional space by giving their coordinates with reference to the orthogonal axes shown in Figure 1. Without loss of generality we take  $0 < q \leq 1/2$  throughout.



**Figure 1:** Orthogonal axes used to specify coordinates  $F$ ,  $f_{11}$ , and  $f_{01}$ , for given  $q$ .

The shape of the region of admissible points is governed by the elements  $f_{ij}$  of  $C$  - that they sum to 1, etc. So, while the requirements are basically simple, they require detailed description. The regions are of three main types, depending on  $q$ : (i)  $1/4 < q < 1/2$ ; (ii)  $q \leq 1/4$ ; (iii)  $q = 1/2$ . We describe these separately. Points  $(F, f_{11}, f_{01})$  on and between the bounding planes are admissible, that is to say correspond to admissible  $C$ . The faces of the region are planar, the planes

defined by  $f_{11} = 0 (\Lambda_1)$  and  $f_{01} = 0 (\Lambda_2)$  being two main ones. The remaining planes are defined by the equations

$$\Lambda_3 \quad 4pqF - f_{11} - 4f_{01} + 4q^2 = 0 \tag{8}$$

$$\Lambda_4 \quad 2pqF + f_{11} + f_{01} - 2pq = 0 \tag{9}$$

$$\Lambda_5 \quad 12pqF + 3f_{11} + 4f_{01} + 4p(p - 2q) = 0 \tag{10}$$

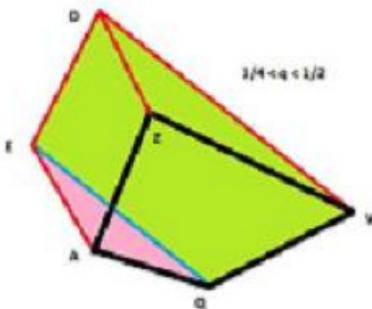
**(i) Region  $1/4 < q < 1/2$**

Figure 2 is a schematic of this region. The coordinates of the vertices are given in Table 1. Other points of reference, not shown on Figure 2, are: O  $(-q/p, 0, 0)$ ; B  $((p - 2q)/(3p), 0, 0)$ ; N  $((2p - q)/(3p), 0, 0)$ . The distance between O and V is  $1/p$ , that between O and B is  $1/(3p)$ , that between O and N is  $2/(3p)$ , and therefore that between B and N is  $1/(3p)$ . The bounding planes, identified by their vertices, are as follows:

$\Lambda_1$ : AQVZ;  $\Lambda_2$ : QVDE;  $\Lambda_3$ : AZDE;  $\Lambda_4$ : VDZ;  $\Lambda_5$ : AQE

**Table 1: The Coordinates of the Vertices of the Admissible Regions as Functions of  $q$**

| Vertex | $F$                        | $f_{11}$      | $f_{01}$  |
|--------|----------------------------|---------------|-----------|
| A      | $-(p - q)^2 / 4pq$         | 0             | $q - 1/4$ |
| V      | 1                          | 0             | 0         |
| D      | $(p - 2q) / (3p)$          | $4q/3$        | 0         |
| E      | $-(1 - 4q + 6q^2) / (6pq)$ | $2(4q - 1)/3$ | 0         |
| Z      | $(2p - q) / (3p)$          | 0             | $2q/3$    |
| Q      | $(3q - 1) / (3q)$          | 0             | 0         |



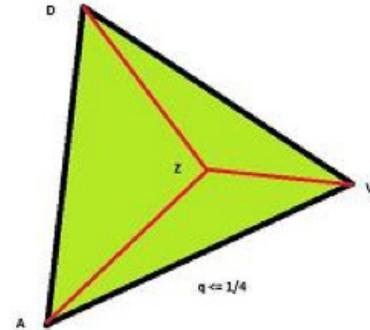
**Figure 2:** Schematic illustration of the bounding region of admissible sets of  $F$ ,  $f_{11}$  and  $f_{01}$  for  $1/4 < q < 1/2$ .

In Figure 2, the plane shaded in lime-green is  $\Lambda_2$ , that is when  $f_{01} = 0$ . To assist with orientation, the line segment QV lies along the  $F$  axis, so that the corner letters of  $\Lambda_2$  lie in the  $F \sim f_{11}$  plane.

**(ii) Region  $q \leq 1/4$**

This region is shown in Figure 3. It does not contain  $\Lambda_5$ . The planes are:

$\Lambda_1$ : AVZ;  $\Lambda_2$ : AVD;  $\Lambda_3$ : AZD;  $\Lambda_4$ : VDZ



**Figure 3:** Schematic illustration of the bounding region of admissible sets of  $F$ ,  $f_{11}$  and  $f_{01}$  for  $q \leq 1/4$ .

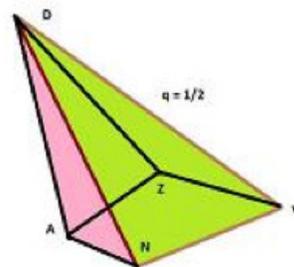
In Figure 3, the plane shaded in lime-green is  $\Lambda_2$ , that is when  $f_{01} = 0$ . To assist with orientation, the line segment AV lies along the  $F$  axis.

**(iii) Region  $q = 1/2$**

This region is shown in Figure 4. The planes are:

$\Lambda_1$ : ANVZ;  $\Lambda_2$ : NVD;  $\Lambda_3$ : AZD;  $\Lambda_4$ : VDZ;  $\Lambda_5$ : AND

When  $q = 1/2$ , points D and E coalesce, as do Q and N. When  $q = 1/4$ , points A, E and O coalesce.



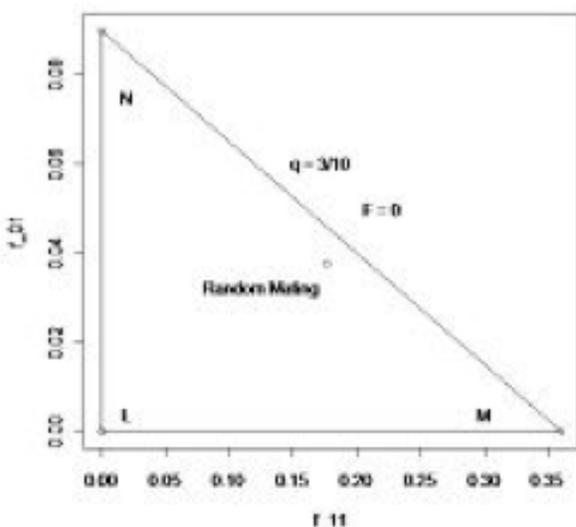
**Figure 4:** Schematic illustration of the bounding region of admissible sets of  $F$ ,  $f_{11}$  and  $f_{01}$  for  $q = 1/2$ .

In Figure 4, the plane shaded in lime-green is  $\Lambda_2$ , that is when  $f_{01} = 0$ . To assist with orientation, the line segment NV lies along the  $F$  axis.

**HARDY-WEINBERG PROPORTIONS AND RANDOM MATING**

The fact that Hardy-Weinberg proportions can be maintained by non-random as well as random mating is

supported by Figure 5 which displays the section of the region depicted by Figure 2 corresponding to  $F=0$ , when  $q=3/10$ . Provided the mating system follows the restrictions applying to matrix  $C$  outlined above, any point within and on the edges of the triangle gives an admissible system. This shows the highly restrictive nature of random mating which corresponds to a single point within the triangle. The  $(f_{11}, f_{01})$  coordinates of the points in the diagram are: L:- (0,0); M:- (0.36,0); N:- (0,0.09); random mating:- (0.1764,0.0378).



**Figure 5:** The region of admissible combinations of values of  $f_{11}$  and  $f_{01}$  when  $q=3/10$  and  $F=0$ , that is for a population in Hardy-Weinberg form. The letters L, M, and N identify the vertices. The point corresponding to random mating is shown within the region. All points within and on the edges are admissible.

## APPLICATION

The trinity of a population *in equilibrium* under *random mating* with *Hardy-Weinberg proportions* is a standard of monographs in human genetics. Penrose [1, page 103] writes:

The concept of gene frequency is of fundamental importance in the genetics of wild populations, which human populations resemble much more closely than selected breeds of laboratory animals. The idea is essential to the mathematical study of evolution because many of the processes of natural selection can be expressed in terms of progressive increase or decrease in gene frequencies. In the shorter-term problems of human populations, the concept is also indispensable. The elementary theoretical

results were discovered independently by Hardy, Pearson and Weinberg.

As can be seen from the preceding sections, the Hardy-Weinberg Principle is less elementary than commonly supposed. Figure 5 shows that random mating is associated with a single point in the wider set of matings consistent with Hardy-Weinberg proportions. Furthermore a more general equilibrium is sustainable *even without natural selection*.

Penrose has a section entitled "Amaurotic Idiocy" [1]. He includes, in particular, infantile amaurotic idiocy, which he says is known as Tay-Sachs disease. This disorder is the focus of our application. Penrose notes the then recently-established connection with hexosaminidase A established by O'Brien [10]. Penrose included Tay-Sachs disease in the category of cerebromacular degeneration. [1, pages 163, 382]. He notes: "Slome's (1933) analysis of published sibships clearly indicates that it is due to a single autosomal gene in spite of a slight excess of female cases." [11] Since then extensive research has mapped the relevant HEXA gene to chromosome 15.

O'Brien writes: "The widespread absence of hexosaminidase A in the tissues, the normal or increased activity of other lysosomal hydrolases, the normal or increased activity of hexosaminidase A in other gangliosidoses, and the partial reduction of this enzyme in the serum of individuals heterozygous for the Tay-Sachs gene, suggest that the absence of hexosaminidase A is the fundamental enzymic defect in Tay-Sachs disease." [10]

The finding of O'Brien led to a test for Tay-Sachs heterozygotes. Lewis, an Ashkenazim, gives an account of her own experience [12, pp. 126-130]. For the purposes of this paper our population is the people of Ashkenazi Jewish background, that is Jews whose ancestors came from central or eastern Europe. Westman [13] gives additional information about the Ashkenazi community. Lewis states that, before genetic testing, about 1 in 3,600 Ashkenazi newborn were diagnosed with Tay-Sachs disease. Using Hardy-Weinberg proportions we take the gene frequency in the population to be  $q = \sqrt{1/3600} = 1/60$ . Lewis notes that, as a consequence of screening, the incidence of Tay-Sachs in the Ashkenazi population has been substantially reduced. Just for the purpose of illustration, we suppose the incidence now to be  $1/36000$ , although it is lower than that.

Under random mating, the former mating matrix, (3), is

$$12960000^{-1} \begin{bmatrix} 1 & 118 & 3481 \\ 118 & 13924 & 410758 \\ 3481 & 410758 & 12117361 \end{bmatrix}. \quad (11)$$

Because of the programme to reduce the incidence of Tay-Sachs in the population, the frequency of marriage between carriers has been reduced. Our model, incorporating incidence  $1/36000$ , is

$$36000^{-1} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 4 & 1194 \\ 1 & 1194 & 33606 \end{bmatrix}. \quad (12)$$

Note that the gene frequency is  $1/60$  and the population is in Hardy-Weinberg form in both (11) and (12). Neither model is fully valid in that children with Tay-Sachs die long before reproductive age. Consequently there is natural selection against the gene on this account but its reduction is slow because of the mode of inheritance. Mating matrix (12) could be replaced by many others which satisfy equation (7).

Sebro *et al.* show the importance of taking account of the stratification of populations in epidemiological research [14].

## REFERENCES

- [1] Penrose LS. The Biology of Mental Defect. 4<sup>th</sup> ed. London: Sidgwick & Jackson 1972.
- [2] Hartl DL, Jones EW. Essential Genetics: A Genomics Perspective. 4<sup>th</sup> ed. Sudbury, Massachusetts: Jones and Bartlett Publishers 2006.
- [3] Edwards AWF, Hardy GH. (1908) and Hardy-Weinberg equilibrium. Genetics 2008; 179: 1143-50. <http://dx.doi.org/10.1534/genetics.104.92940>
- [4] Mayo O. A century of Hardy-Weinberg equilibrium. Twin Res Hum Genet 2008; 11: 249-46. <http://dx.doi.org/10.1375/twin.11.3.249>
- [5] Bernstein SN. Solution of a mathematical problem connected with the theory of heredity. Ann Math Stat 1942; 12: 53-61. <http://dx.doi.org/10.1214/aoms/1177731642>
- [6] Seneta E. Sergei Natanovich Bernstein. In: Heyde CC, Seneta E, editors. Statisticians of the Centuries. New York: Springer 2001; pp. 339-342. [http://dx.doi.org/10.1007/978-1-4613-0179-0\\_73](http://dx.doi.org/10.1007/978-1-4613-0179-0_73)
- [7] Stark A, Seneta E, On SN. Bernstein's derivation of Mendel's law and 'rediscovery' of the Hardy-Weinberg distribution. Genet Mol Biol 2012; 35: 388-94. <http://dx.doi.org/10.1590/S1415-47572012005000025>
- [8] Stark AE. On extending the Hardy-Weinberg law. Genet Mol Biol 2007; 30: 664-66. <http://dx.doi.org/10.1590/S1415-47572007000400026>
- [9] Stark AE, Seneta E. A reality check on Hardy-Weinberg. Twin Res Hum Genet 2013; 16: 782-89. <http://dx.doi.org/10.1017/thg.2013.40>
- [10] O'Brien JS. Five gangliosidoses. Lancet 1969; 2: 805. [http://dx.doi.org/10.1016/S0140-6736\(69\)90524-8](http://dx.doi.org/10.1016/S0140-6736(69)90524-8)
- [11] Slome D. The genetic basis of amaurotic family idiocy. J Genet 1933; 27: 363-72, plus Appendix.
- [12] Lewis R. Human Genetics: the Basics. London and New York: Routledge 2011.
- [13] Westman JA. Medical Genetics for the Modern Clinician. Philadelphia: Lippincott Williams & Wilkins 2006.
- [14] Sebro R, Hoffmann TJ, Lange C, Rogus JJ, Risch NJ. Testing for non-random mating: evidence for ancestry-related assortative mating in the Framingham Heart Study. Genet Epidem 2010; 34: 674-79. <http://dx.doi.org/10.1002/gepi.20528>