


Article

The Impossible Dream — Panmixia

Alan E. Stark 

School of Mathematics and Statistics F07, The University of Sydney, Sydney, New South Wales, Australia

Abstract

This study starts with a simple model by which Hardy-Weinberg proportions are attained in a single generation while maintaining gene frequencies. The question of differentiating between random and non-random mating is explored by simulation. Sample mating proportions are generated using the model as base. The difficulty of differentiating between random and non-random mating is illustrated.

Keywords: Hardy-Weinberg law; Quasi-random mating; Simulated coupling parameter; Estimation standard error

(Received 8 November 2023; revise received 1 February 2024; accepted 5 February 2024; First Published online 27 February 2024)

Despite the fact that in 1988 respected geneticist C. C. Li demonstrated that Hardy-Weinberg proportions can be maintained by non-random mating, and that the subject was worth further investigation, the genetics community maintains the pretence of ‘random mating’. This paper illustrates how a sample generated by non-random mating frequently can be perceived as coming from a population practising random mating.

The title of this article has been borrowed from the song in the musical *Man of La Mancha*. It expresses the notion of an unattainable ideal. Perfect random mating and Hardy-Weinberg proportions are unattainable in practice. As an abstraction it is not obvious how to define random mating for a finite population. This does not exclude Hardy’s law from practical application. Weinberg (1908) used the formulae to study the propensity to produce twins in a segment of the German population. The textbook presentation of the Hardy-Weinberg principle should be rewritten to include pseudo- and quasi-random mating.

Smith (1969, p. 200) expresses the traditional view of the Hardy-Weinberg law as follows: ‘Now married couples do not come together on a basis of their MN blood-groups, whatever other reasons they may have. So we may well expect that, as far as blood-groups are concerned, mating will be at random (“panmixia”)’. Following this assumption, Smith produces the frequency of couple MN, MM chosen at random from the population as $P_{MN} \times P_{MM}$. Smith then finds the frequencies of a child drawn at random from the population to be the Hardy-Weinberg proportions.

Stark (2023) presents a concise genetics model that produces Hardy-Weinberg frequencies for an autosomal locus with two alleles, in one generation, starting from arbitrary frequencies. This model provides some justification for the fact that approximate Hardy-Weinberg proportions are often observed. In this model mating proportions are controlled by a single parameter h which allows flexibility. Setting $h = 0$ defines ‘random mating’. The object

of this paper is to illustrate, by simulation, that it is difficult to distinguish random from non-random mating in samples of mating proportions generated by setting a non-zero value of h .

It appears that the details of Li’s paper are not well known. For this reason a brief outline is given in the next section. This is followed by the model of quasi-random mating, the simulation and a section that attempts to create an impression of how the genetics community views the Hardy-Weinberg law.

Pseudo-Random Mating

The Wikipedia version of the Hardy-Weinberg principle has obscured the basic simplicity of G. H. Hardy’s (1908) original law. Also, it has diverted attention from the observation by C. C. Li (1988) that Hardy-Weinberg proportions can be maintained by non-random mating which Li called ‘pseudo-random mating’. In Li’s notation random mating is defined by $f_{ij} = f_i f_j$ and non-random mating by $f_{ij} = f_i f_j + d_{ij}$, where d_{ij} is the deviation from random-mating frequency. Li (1988) says: ‘The properties of panmixia are well known’ (p. 731).

C. C. Li (1988) defined pseudo-random mating by means of the model shown in Table 1, with minimal explanation, except by pointing out that the mating proportions given can be shown to produce offspring in the same proportions as the parents. He says that Table 1 has the ‘pseudo-random mating property’ (p. 732). Table 1 does not have equal reciprocal mating frequencies, unlike Hardy’s model. Later, Li reaches the symmetrical model characterized by the condition $c_{33} = 4c_{12}$, as dealt with in the next section. This condition reproduces any genotypic distribution.

The mating proportions given in Table 1 follow Li’s prescription given in the first paragraph of the introduction above. The quantity x can be varied providing it conforms to the constraints of the table. Li considers other aspects of pseudo-random mating that do not change the main property of Table 1, which is that the offspring have the same genotypic proportions as the parents. The terms involving x cancel each other as can be

Corresponding author: Alan E. Stark; Email: alans@exemail.com.au

Cite this article: Stark A. E. (2024) The Impossible Dream — Panmixia. *Twin Research and Human Genetics* 27: 174–178. <https://doi.org/10.1017/thg.2024.6>



Table 1. Algebraic mating proportions reproducing Hardy-Weinberg offspring devised by C. C. Li (1988). The order of genotypes has been rearranged.

Male\female	AA	BB	AB	Total
AA	$q^4 + x$	$p^2q^2 + 2x$	$2pq^3 - 3x$	q^2
BB	$p^2q^2 + 3x$	$p^4 + 4x$	$2p^3q - 7x$	p^2
AB	$2pq^3 - 4x$	$2p^3q - 6x$	$4p^2q^2 + 10x$	$2pq$
Total	q^2	p^2	$2pq$	1

Table 2. Symbolic mating proportions reproducing offspring

Male\female	AA	BB	AB	Total
AA	c_{11}	c_{12}	c_{13}	G_1
BB	c_{21}	c_{22}	c_{23}	G_2
AB	c_{31}	c_{32}	c_{33}	G_3
Total	G_1	G_2	G_3	1

shown by applying Mendel’s rule to the mating proportions. The other terms follow Hardy’s rule.

Other sets of coefficients for x can be found as explored by Li. The details of the logic involved may be the reason why the genetics community has not embraced Li’s findings. The other aspects of pseudo-random mating considered by Li are parental correlation, gametic independence and sib-sib correlation. Following Wright (1921), Li gives values 0, 1 and 2 to genotypes. Li makes an ambiguous comment: ‘Since there are an infinite number of possible pseudo-random mating patterns, we cannot be sure that none of them exists in nature’ (p. 733). Li concludes: ‘When we observe reasonable agreement with random mating expectations from a sample, we really could not be sure what the mating pattern might be in the population’ (p. 737).

Quasi-Random Mating

Consider an autosomal locus with two alleles A and B and genotypes AA, BB and AB numbered 1, 2 and 3. Mating pairs are formed in the current generation to produce offspring in the next. The proportions of the mating pairs are given symbolically in Table 2. The elements c_{ij} are non-negative and symmetrical in value ($c_{ij} = c_{ji}$) and sum to 1.

Suppose that the genotypic proportions are

$$G_1 = q^2 + Fpq; G_2 = p^2 + Fpq; G_3 = 2pq(1 - F),$$

F measures departure from Hardy-Weinberg form and the gene frequencies are $q = (2G_1 + G_3)/2; p = (2G_2 + G_3)/2 = 1 - q$.

The mating frequencies are

$$c_{ij} = G_iG_j(1 + he_i e_j / v) \tag{1}$$

where $e_1 = p(F - 1)/(q + Fp); e_2 = q(F - 1)/(p + Fq); e_3 = 1;$ and $v = pq(1 - F^2)/((q + Fp)(p + Fq))$. For fixed q , v increases monotonically from zero to 1 as F decreases from 1 to zero.

The gene frequencies are not changed through the action of (1). Subject to constraints h can be chosen over a wide range allowing uncountable possibilities for varying the mating regime but still

producing, in one generation, offspring distributed according to the Hardy-Weinberg formulae:

$$H_1 = q^2; H_2 = p^2; H_3 = 2pq. \tag{2}$$

This can be verified by calculating the offspring frequencies by applying Mendel’s rule to (1):

$$AA \ c_{11} + c_{13} + c_{33}/4; BB \ c_{22} + c_{23} + c_{33}/4; AB \ 2c_{12} + c_{13} + c_{23} + c_{33}/2.$$

A numerical example illustrating model (1) is given in Stark (2023).

The Simulation

The object of this note is to explore the question of whether a standard statistical test is likely to be able to identify a departure from random mating in a hypothetical collection of genetic data.

The performance of an estimator of h assuming that couples are formed following model (1) is studied by simulation. This is done by drawing random samples of couples and using the observed numbers of pairs of genotypes to calculate q and F from each sample and, using these values, to make a maximum likelihood estimate of h . It is assumed that the only information available is the number of pairs of each type and so estimates of q and F must first be made from these counts.

Denoting the random number of pairs of genotypes i and j by n_{ij} , the likelihood of the sample is proportional to

$$\prod_{i,j} (G_iG_j(1 + he_i e_j / v))^{n_{ij}}. \tag{3}$$

Heuristic estimates of q and F are formed:

using the notation given above the estimates from each sample are

$$q = (4n_{11} + 2n_{12} + 3n_{13} + 2n_{21} + n_{23} + 3n_{31} + n_{32} + 2n_{33})/(4N),$$

and

$$F = 1 - (n_{13} + n_{23} + n_{31} + n_{32} + 2n_{33})/(4Nq(1 - q)),$$

where N is the number of couples in the sample. Using these estimates of q and F , taking the logarithm of expression (3), for each sample the maximum likelihood estimate of h is obtained. This is done by taking a fine grid of potential values of h and choosing the one of greatest likelihood.

Figure 1 illustrates a typical outcome, in this case by generating couples taking $h = 1/200$. The mean of the data displayed in the histogram is 0.005442 and the standard error is 0.0201. In this case 2500 couples were generated in each sample and 1000 estimates of h were made. Although this is of lesser interest the starting value of $q = 1/4$ and of $F = 1/3$. The important point to note is that there is no clear distinction between $h = 0$ and values near this.

As Figure 1 illustrates, there is no reason why a test of goodness of fit might not contradict the Hardy-Weinberg expectation but may be underpinned by non-random mating.

Figure 2 gives an indication of the dependence of the estimate of h on the size of the sample of pairs of mating couples. The number of couples is plotted on the horizontal axis and the standard error of estimate on the vertical.

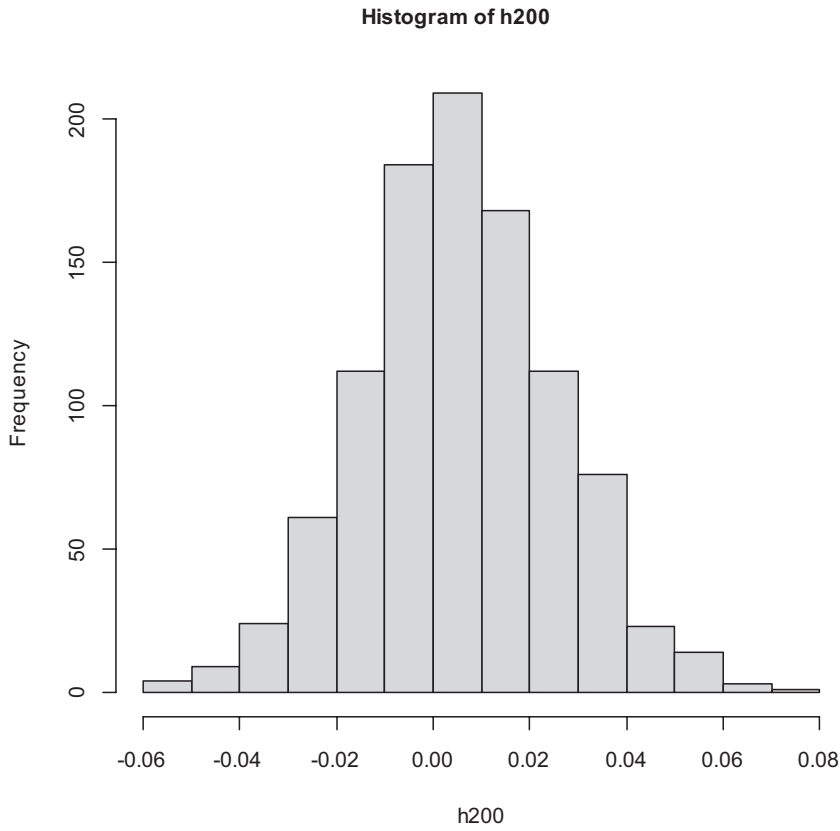


Figure 1. Histogram of estimates of h based on samples generated from a population value $h = 1/200$.

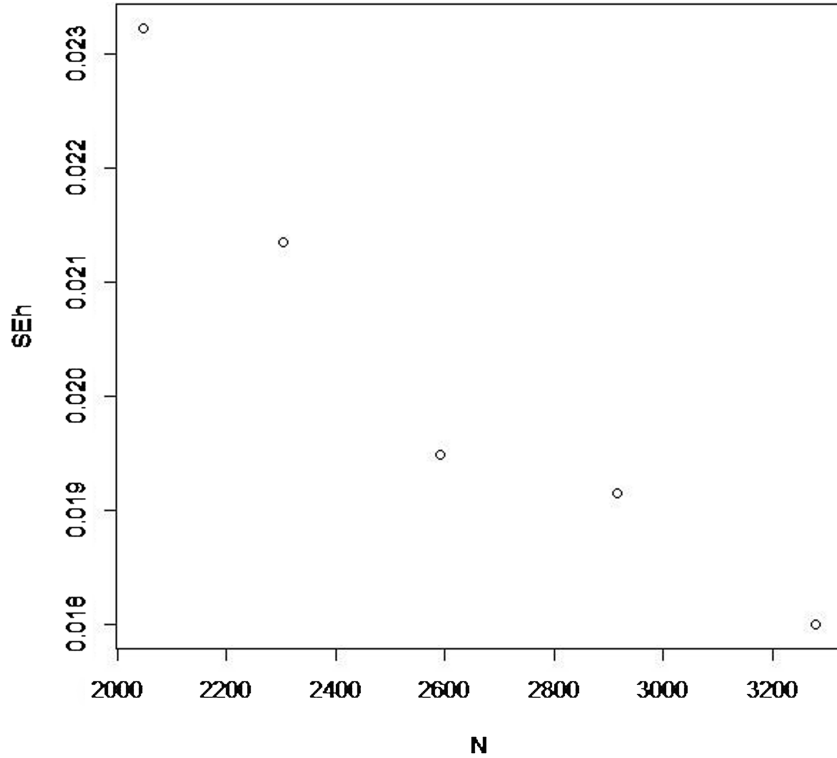


Figure 2. Simulated standard errors of estimates of h as a function of the size of samples of the number of couples N .

Grassroots

Wright (1959, p. 107) indicates the place of the Hardy-Weinberg principle as follows:

The theoretical genetics of populations may be considered to have begun with [Karl] Pearson’s demonstration that the 1:2:1 Mendelian ratio tends to maintain itself indefinitely in a large random-breeding population derived from F_2 of a cross. A few years later Hardy and Weinberg independently

pointed out that any array of gene frequencies at a locus ($\sum q_i A_i$, where A_i is a particular allele at locus A_i and q_i is its proportional frequency) tends to remain unchanged in a large self-contained population in the absence of disturbing factors such as mutation and selection, and thus that the frequency of the zygotes resulting from random mating becomes stable immediately after attainment of equality of gene frequencies in the sexes in the array $(\sum q_i A_i)^2$ for one locus. If mating is not at random, the zygotic array, in the absence of other disturbing factors, is given by $(1 - F)[\sum q_i A_i]^2 + F(q_i A_i A_i)$, where F is the inbreeding coefficient, defined as the correlation between uniting gametes with respect to additive effects.

This statement, which is typical of the genetics literature, conflates the essence of Hardy's original deterministic abstraction with qualifications about real populations that would make them consistent with Hardy's model. This leaves no room to accommodate Li's pseudo-random mating or the quasi-random mating model.

Dobzhansky (1951) introduces the Hardy-Weinberg law by combining two strains of a sexual and cross-fertilizing organism that differ in a single gene AA and aa , that they interbreed at random, and that they are introduced in the proportions q of AA and $(1 - q)$ of aa . He asks the question, what will be the resulting Mendelizing population in the next and following generations. He says that the solution is the Hardy-Weinberg, or the binomial square, law. He adds: 'If there is some breeding preference, such as a tendency towards inbreeding or self-fertilization, the relative frequencies of the homozygotes (AA and aa) and the heterozygotes (Aa) will be modified, but the gene frequencies, q and $(1 - q)$, will still remain constant' (p. 53).

Dobzhansky's explanation is consistent with the standard textbook version of the Hardy-Weinberg law. Malécot (1969) outlines population genetics theory in probabilistic terms, which permits him to switch between infinite and finite populations. 'Panmixia' is listed separately in the index and appears repeatedly in the book. One of the cases dealt with is described as follows: 'The parents mate at random; the probability of finding a mate is the same for all individuals; and fecundity is the same for all couples. This is <<random mating>>, panmixia.'

Malécot (1969) has a section dealing with the increase of coancestry in a finite population and gives the formulae for genotypic frequencies. He adds the footnote: 'The above formulas are based on the assumption that there is only random inbreeding (consistent with panmixia)' (p. 32).

With respect to variations in populations, Philip (1938, p. 198) writes: 'It is possible ... to decide genetically in some cases whether a population is inbred, or whether there is random mating.' Philip's method is to test against Pearson's and Hardy's law expressed as

$$u^2 AA : 2uAa : 1aa,$$

u being the 'ratio of the dominant gene A to the recessive gene a '.

Philip used the beetle *Dermestes vulpinus*, recording two characters, wing colour and pigment density. The alleles of the first character were dark (dominant) and light (recessive). The beetles were collected either as larvae, pupae, or imagines. An elaborate procedure was required to test for random mating. The analysis was based on 237 beetles. A backcross was used to separate the dark animals into homozygotes and heterozygotes. The almost perfect agreement between observed and expected frequencies was

obtained by using the result of the backcross to estimate gene frequencies. Philip comments: 'This fits well with the assumption of random mating if the relative distribution of the genes in the population is 18.7% for L and 81.3% for l ' (p. 201). Malécot (1969, p. 14) cites Philip (1938) for finding an example (the beetle above) of a natural population exhibiting Hardy's law.

Haldane (1938) cites Philip (1938) as obtaining indirect evidence concerning the mating system. Haldane's comment, which is appropriate to Philip's methodology is: '... we may find that the frequencies of the three genotypes are so close to those which would be expected from random mating that the divergences may be ascribed to chance' (p. 213).

There is no single clear definition of 'random mating'. It is applied to abstract settings for finite and infinite populations and real populations. Often it is assumed that no definition is needed since everybody understands what it means. In setting out his Fundamental Theorem of Natural Selection Fisher (1930, p. 35) writes: 'Since the theorem is exact for idealized populations ... It will be sufficient ... to consider the special case of a population mating and reproducing at random.' Fisher does not say what this means but a definition is implicit when he gives the variance of a chance fluctuation in the gene frequency as that obtainable from the binomial distribution.

Uspensky (1937, pp. 9–11) chose Mendelian genetics to illustrate the application of probability to a branch of science. He used what he refers to as the classical definition of probability. In essence, if an event occurs in r out of n equally likely ways, its probability of occurrence is r/n . In Mendel's case, if gametes are joined from two heterozygotes, the four possible outcomes each have probability $1/4$. Uspensky emphasizes the precision of this theory and contrasts it with the lack of an equally precise model to predict the sex ratio in humans. Uspensky does not consider the Hardy-Weinberg law.

Acknowledgment. I thank the reviewer for critical comments and constructive suggestions.

Financial support. None.

Competing interests. None.

Ethical standards. No experimental subjects were involved.

Publication ethics. Sources acknowledged.

References

- Dobzhansky, T. (1951). *Genetics and the origin of species* (3rd ed., rev.). Columbia University Press.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon Press.
- Haldane, J. B. S. (1938). Indirect evidence for the mating system in natural populations. *Journal of Genetics*, 23, 213–220.
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28, 49–50. <https://doi.org/10.1126/science.28.706.49>
- Li, C. C. (1988). Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics*, 119, 731–737. <https://doi.org/10.1093/genetics/119.3.731>
- Malécot, G. (1969). *The mathematics of heredity*. W. H. Freeman and Company.
- Philip, U. (1938). Mating systems in wild populations of *Dermestes vulpinus* and *Mus musculus*. *Journal of Genetics*, 36, 197–211. <https://doi.org/10.1007/BF02982414>
- Smith, C. A. B. (1969). *Biomathematics: The principles of mathematics, for students of biological and general science* (vol. 2, 4th ed.). Charles Griffin & Company.

- Stark, A. E.** (2023). C. C. Li and quasi-random mating. *Twin Research and Human Genetics*, 26, 277–279. <https://doi.org/10.1017/thg.2023.37>
- Uspensky, J. V.** (1937). *Introduction to mathematical probability*. McGraw-Hill Book Company.
- Weinberg, W.** (1908). Über den Nachweis der Vererbung beim Menschen. *Jahresh Ver Vaterl Naturkd*, 64, 368–382.
- Wright, S.** (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, 6, 111–123. <https://doi.org/10.1093/genetics/6.2.111>
- Wright, S.** (1959). Physiological genetics, ecology of populations, and natural selection. *Perspectives in Biology and Medicine*, 3, 107–151. <https://doi.org/10.1353/pbm.1959.0044>