Twin Research and Human Genetics (2021), **24**, 160–162
doi:10.1017/thg.2021.26

CAMBRIDGE UNIVERSITY PRESS

Article

A Misconception About the Hardy–Weinberg Law

Alan E. Stark

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

Abstract

The Hardy–Weinberg law of population genetics is usually associated with the notion of random mating of parents. A numerical example for a triallelic autosomal locus shows that an uncountable set of mating combinations can maintain Hardy–Weinberg proportions. Therefore, one cannot infer random mating in a population from the observation of Hardy–Weinberg equilibrium. The mating system which ensures that the genotypic distribution of offspring is the same as that of the parents is specified.

Keywords: Hardy–Weinberg law; triallelic system; random mating; independent gametes

(Received 4 June 2021; accepted 28 June 2021; First Published online 22 July 2021)

The object of this paper is to show by a numerical example how the commonly held view of the Hardy–Weinberg principle is based on a misconception of its connection to random mating of parents and independent pairing of gametes to form zygotes. The conclusion that follows is that one cannot infer random mating in a population from the observation of Hardy–Weinberg equilibrium. However, this is done commonly — for example, Tallis (1966, p. 121), after giving a set of Hardy–Weinberg frequencies, writes: 'the population is assumed to be panmictic'.

Li (1988) introduced the term 'pseudo-random mating' when he specified a model showing that Hardy–Weinberg proportions could be so maintained at a locus for two alleles. That this is possible is implicit in a formula given by Stark (1980), and Stark (2006) shows that Hardy–Weinberg proportions can be reached in one generation from an arbitrary distribution in one round of nonrandom mating. In the light of the present paper, it can be conjectured confidently that it is possible to maintain the composition of a population with nonrandom mating for any number of alleles.

The ABO blood group series is referred to because it is an example of a locus with three allelomorphs. Crew (1947) is based on a course of genetics that the author gave to medical students for almost 30 years. When considering the difficulty of presenting such a course, he wrote (1947, viii): 'It is not to be expected that, to one whose conditioned ambition it is to treat the sick or injured human individual, genetics can exercise a strong appeal.' In presenting the ABO groups, Crew strives to overcome the resistance of medical students. The subdivision of antigen A into forms A_1 and A_2 noted by Penrose (1973, p. 25) is ignored here.

Crew points out the possible confusion that can arise as to whether A, B and O refer to the blood groups, the antigens or the genes. We use *A, B* and *O* for the genes and A, B, AB and O for the phenotypes. Crew gives a table of the rules to be observed for the transfusion of whole blood (1947, p. 66). The phenotypes and genotypes are grouped as follows: A (*AA* and *A O*), B (*BB* and *BO*), AB (*AB*) and O (*OO*). He gives the percentages of phenotypes: A — 42.2; B — 8.7; AB — 3.2; O — 45.8 (p. 67). Crew gives a table of the 10 different parental mating combinations and both the possible phenotypes of offspring and the phenotypes lacking among the offspring (p. 68).

Yamamoto (2004) gives a more comprehensive review of the ABO system, including the evolution of the system in humans and other species.

In the main, Crew does not touch on population genetics theory. In relation to 'total sex-linked inheritance', such as red–green blindness and hemophilia, he makes the following observation (p. 49): 'If the frequency of hemophilic males among the male population is p and the frequency of normal males is q, where $(p + q) = 1$, then with random mating, the frequency of hemophiliacs, carriers and normals among the female population is p^2:$2pq$: q^2.' The extension of this distribution to an autosomal locus with three alleles is the subject of this paper.

Cavalli-Sforza and Bodmer (1971, p. 53) state: 'The Hardy-Weinberg theorem (as shown by Weinberg, 1909) can be extended quite simply to cover multiple alleles. Thus, if we assume that random mating is equivalent to the random union of gametes, we may compute the frequencies of the various genotypes by the expansion of

$$(p_1 + p_2 + \cdots + p_m)^2$$

where the p_i's are the frequencies of the genes A_i.'

The above statement invites two comments: first, 'random union of gametes' should be pairing of gametes drawn independently from the gene pool, and second, Hardy–Weinberg frequencies can be maintained by nonrandom mating of parents, as is demonstrated in the next section.

Nonrandom Mating and Hardy–Weinberg Frequencies

Phenotypic identities are ignored so that the focus is on the three genes denoted *A, B, C* and genotypes *AA, BB, CC, AB, AC, BC,*

Author for correspondence: Alan Stark, Email: alans@exemail.com.au
Cite this article: Stark AE. (2021) A Misconception About the Hardy–Weinberg Law. *Twin Research and Human Genetics* **24**: 160–162, https://doi.org/10.1017/thg.2021.26

which are numbered, respectively, 1, 2, 3, 4, 5, 6. There are 36 possible mating combinations and the proportions are set out in a symmetric matrix with elements $c_{i,j}$, $i, j = 1, 2, \ldots, 6$.

The following identities give the conditions on the elements of the matrix which ensure that the offspring distribution is the same as the parental:

$$AA : c_{12} + c_{13} + c_{16} - \left(\frac{1}{4}c_{44} + \frac{1}{4}c_{55} + \frac{1}{2}c_{45}\right) = 0$$

$$BB : c_{12} + c_{23} + c_{25} - \left(\frac{1}{4}c_{44} + \frac{1}{4}c_{66} + \frac{1}{2}c_{46}\right) = 0$$

$$CC : c_{13} + c_{23} + c_{34} - \left(\frac{1}{4}c_{55} + \frac{1}{4}c_{66} + \frac{1}{2}c_{56}\right) = 0$$

$$AB : 2c_{12} + c_{16} + c_{25} + \frac{1}{2}c_{56} - \left(c_{34} + \frac{1}{2}c_{44} + \frac{1}{2}c_{45} + \frac{1}{2}c_{46}\right) = 0$$

$$AC : 2c_{13} + c_{16} + c_{34} + \frac{1}{2}c_{46} - \left(c_{25} + \frac{1}{2}c_{45} + \frac{1}{2}c_{55} + \frac{1}{2}c_{56}\right) = 0$$

$$BC : 2c_{23} + c_{25} + c_{34} + \frac{1}{2}c_{45} - \left(c_{16} + \frac{1}{2}c_{46} + \frac{1}{2}c_{56} + \frac{1}{2}c_{66}\right) = 0.$$

The above equations are satisfied if

$$c_{44} = 4c_{12}; c_{55} = 4c_{13}; c_{66} = 4c_{23}; c_{45} = 2c_{16}; c_{46} = 2c_{25}; c_{56} = 2c_{34}.$$

Table 1 is an example that illustrates how Hardy–Weinberg frequencies can be maintained with nonrandom mating. The gene frequencies are 2/9, 3/9 and 4/9, and the genotype frequencies 4/81, 9/81, 16/81, 12/81, 16/81 and 24/81. Each element in the table is to be divided by 6561 to convert it to a fraction. Table 2 gives the corresponding matrix for random mating of couples.

Table 1. Nonrandom mating proportions which produce the same Hardy–Weinberg frequencies in offspring as in parents (multiplied by 6561)

	AA	*BB*	*CC*	*AB*	*AC*	*BC*	Total
AA	9	75	144	0	0	96	324
BB	75	216	294	0	144	0	729
CC	144	294	666	192	0	0	1296
AB	0	0	192	300	192	288	972
AC	0	144	0	192	576	384	1296
BC	96	0	0	288	384	1176	1944

Table 2. Random mating frequencies (multiplied by 6561)

	AA	*BB*	*CC*	*AB*	*AC*	*BC*	Total
AA	16	36	64	48	64	96	324
BB	36	81	144	108	144	216	729
CC	64	144	256	192	256	384	1296
AB	48	108	192	144	192	288	972
AC	64	144	256	192	256	384	1296
BC	96	216	384	288	384	576	1944

Estimating Gene Frequencies

Race and Sanger (1975) comment on the value of estimating gene frequencies and cite Bernstein (1930) in relation to ABO; for example: 'knowing the gene frequencies we can calculate the expected frequency of children of different groups, from any type of mating.' (p. 12). Looking back over the 50 years preceding the 1975 edition of their work they write:

> We sometimes wonder whether since 1911, or say 1925 to take in Bernstein, the only contributions of the first magnitude to the system [ABO] are to be found in the biochemical work on the ABH substances and in the work on the 'Bombay' phenomenon; and in the recognition of the *cis* phenomenon and perhaps, on a more practical level, the finding of specific agglutinins in extracts of seeds and snails. (Race & Sanger, 1975, p. 15)

Presumably, Race and Sanger were not considering studies of associations between ABO phenotypes and diseases. Mueller and Young (1995, pp. 188–189) summarize the associations between types and duodenal and gastric ulcers, and investigations were done before 1975. Many other such studies have been done.

Hartl and Clark (1989, pp. 40–42) describe a method for calculating gene frequencies from a set of ABO phenotypic frequencies. This uses the assumption that the population is in Hardy–Weinberg equilibrium so that the genotypic proportions are

$$\{p_1^2, p_2^2, p_3^2, 2p_1p_2, 2p_1p_3, 2p_2p_3\}.$$

These authors use the following sample counts from Mourant et al. (1976):

O — 702; A — 862; B — 365; AB — 131 and give the estimated gene frequencies : *A*—˜ 0.2813; *B* — 0.1291; *O* — 0.5895.

Kempthorne (1957, pp. 172–177) gives a different method, also iterative, for estimating gene frequencies. He uses the following phenotypic counts taken from Taylor and Prior (1938):

O — 202; A — 179; B — 35; AB — 6, and gives estimated gene frequencies: *A* — 0.25156; *B* — 0.05001; *O* — 0.69843. He also gives standard errors of the estimates.

Both methods of estimating gene frequencies use the theoretical proportions of the Hardy–Weinberg distribution to split the A and B phenotypic counts into two parts. It is used also to calculate starting frequencies for the iterative methods. The method of Hartl and Clark uses gene counts to make repeated revisions of gene frequency estimates. The important point is the crucial role played by the Hardy–Weinberg assumption.

Discussion

The goal of Zhu and colleagues (2020) was to explore the association between the ABO system and human longevity. They sampled 2201 centenarians (570 males, 1631 females) and a regionally matched control group of 2330 middle-aged individuals (793 males, 1537 females). They found no significant difference in ABO phenotypic frequencies between the two groups, so concluded that there was no effect of ABO on longevity. The A and B gene frequencies were each about 21%.

By contrast, the study of Groot et al. (2020) found the ABO blood group system to be associated with several parameters of healthy aging and disease development. The analysis was based on data of 406,755 unrelated individuals from the UK Biobank cohort. They summarized the result as follows:

https://doi.org/10.1017/thg.2021.26 Published online by Cambridge University Press

> In this large community-based population, we determined ABO blood group phenotypes based on inherited allelic combinations and observed numerous associations between the ABO blood group system with healthy aging and the development of a multitude of diseases. The ABO blood groups were primarily associated with cardiovascular outcomes. The present study observed that individuals with blood group A and B were at higher risk of developing thromboembolic diseases, but lower risk of hypertension, when compared with O-group individuals. Individuals with blood group A were at higher risk of developing hyperlipidemia, atherosclerosis, and heart failure compared with blood group O, whereas individuals with blood group B were at higher risk of myocardial infarction compared with individuals with blood group O. The observed differences suggest blood group-specific approaches for the maintenance of human health and the prevention and treatment of a multitude of diseases. (Groot et al., 2020, p. 834)

The gene frequencies in the sample are A — 0.2718; B — 0.0698; O — 0.6583.

The above studies show that there is considerable interest in exploring the relation between the ABO system and disease and the management of disease. They may appear to be rather a blunt approach to understanding compared with, for example, the HapMap method described by Collins (2010, pp. 64–68).

Collins (2010, p. 28) makes a general statement about common diseases such as diabetes, heart disease and cancer, referring to them as *polygenic*, with the 'power of each individual genetic risk factor is generally quite low'. For heart disease, this might appear to be not convincing to Groot et al. (2020).

Using the HapMap approach, Levinsson et al. (2014) sought to find which NOS variants were most strongly associated with cardiovascular pathology. They studied 560 CHD cases and 2791 controls using 58 SNPs. They found the strongest additive protective effect (*OR* 0.59) was related to rs3782218 of NOS1 (the T-allele). Could this be connected in some way with ABO?

Acknowledgments. The author would like to thank a reviewer for suggesting ways to improve the paper.

References

Bernstein, F. (1930). Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, *56*, 233–273.

Cavalli-Sforza, L. L. & Bodmer, W. F. (1971). *The genetics of human populations*. W. H. Freeman and Company.

Collins, F. S. (2010). *The language of life*. Profile Books.

Crew, F. A. E. (1947). *Genetics in relation to clinical medicine*. Oliver and Boyd.

Groot, H. E., Villegas S. L. E., Said, M. A., Lipsic, E., Karper, J. C., & van der Harst, P. (2020). Genetically determined ABO blood group and its associations with health and disease. *Arteriosclerosis Thrombosis and Vascular Biology*, *40*, 830–838.

Hartl, D. L., & Clark, A. G. (1989). *Principles of population genetics* (2nd ed.). Sinauer Associates.

Kempthorne, O. (1957). *An introduction to genetic statistics*. John Wiley.

Levinsson, A. Olin, A.-C., Björck, L., Rosengren, A., & Nyberg, F. (2014). Nitric oxide synthase (NOS) single nucleotide polymorphisms are associated with coronary heart disease and hypertension in the INTERGENE study. *Nitric Oxide*, *39*, 1–7.

Li, C. C. (1988). Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics*, *119*, 731–737.

Mourant, A. E., Kopec, A. C., & Domaniewska-Sobczak, K. (1976). *The distribution of the human blood groups and other polymorphisms* (2nd ed.). Oxford University Press.

Mueller, R. F., & Young, I. D. (1995/6). *Emery's elements of medical genetics* (9th ed.). Churchill Livingstone.

Penrose, L. S. (1973). *Outline of human genetics* (3rd ed.). Heinemann Educational Books.

Race, R. R., & Sanger, R. (1975). *Blood groups in man* (6th ed.). Blackwell Scientific Publications.

Stark, A. E. (1980). Inbreeding systems: classification by a canonical form. *Journal of Mathematical Biology*, *10*, 305.

Stark, A. E. (2006). A clarification of the Hardy-Weinberg law. *Genetics*, *14*, 1695–1697.

Tallis, G. M. (1966). Equilibria under selection for k alleles. *Biometrics*, *22*, 121–127.

Taylor, G. L., & Prior, A. M. (1938). Blood groups in England. *Annals of Eugenics (London)*, *8*, 343–355.

Weinberg, W. (1909). Über Vererbungsgesetze beim Menschen. *Zeitschrift für induktive Abstammungs- und Vererbungslehre*, *1*, 377–392, 440–460; 2, 276–330.

Yamamoto, F. (2004). Review: ABO blood group system ¾ ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases and ABO genes. *Immunohematology*, *20*, 3–22.

Zhu, Y., Liang, Y., Khan, A. H., Dong, M., Wan, Y., Sun, Z., . . . Tian, X.-L. (2020). Allelic distribution of *ABO* gene in Chinese centenarians. *Aging Medicine*, *3*, 195–204.

https://doi.org/10.1017/thg.2021.26 Published online by Cambridge University Press