A Century of Hardy–Weinberg Equilibrium

Oliver Mayo

CSIRO Livestock Industries, Adelaide, Australia

Hardy–Weinberg equilibrium (HWE) is the state of the genotypic frequency of two alleles of one autosomal gene locus after one discrete generation of random mating in an indefinitely large population: if the alleles are *A* and *a* with frequencies *p* and *q* (=1-*p*), then the equilibrium gene frequencies are simply *p* and *q* and the equilibrium genotypic frequencies for *AA*, *Aa* and *aa* are p^2, $2pq$ and q^2. It was independently identified in 1908 by G. H. Hardy and W. Weinberg after earlier attempts by W. E. Castle and K. Pearson. Weinberg, well known for pioneering studies of twins, made many important contributions to genetics, especially human genetics. Existence of this equilibrium provides a reference point against which the effects of selection, linkage, mutation, inbreeding and chance can be detected and estimated. Its discovery marked the initiation of population genetics.

Hardy–Weinberg equilibrium (HWE) is the state of the genotypic frequency of two alleles of one gene locus after one generation of random mating in an indefinitely large population with discrete generations, in the absence of mutation and selection: if the alleles are *A* and *a* with frequencies *p* and *q* (= 1-*p*), then the equilibrium gene frequencies are just *p* and *q* and the equilibrium genotypic frequencies for *AA*, *Aa* and *aa* are p^2, $2pq$ and q^2. Thus, there is equilibrium at both the allelic and the genotypic level. Not excessively fancifully, one could compare this Hardy–Weinberg rule with Newton's first law of motion: a physical body will remain at rest, or continue to move at a constant velocity, unless an external force acts upon it. If such stability is the rule, then it provides the basis for the detection and estimation of the effects on the population of 'the thousand natural shocks the flesh is heir to', including natural and artificial selection, mutation, assortative mating, migration, inbreeding and random sampling (through finite population size).

Each word in the topic concept deserves explication: Hardy was a notable pure mathematician, Weinberg was a pioneering human geneticist and doctor to the poor who made a special contribution to twin studies, and the concept of equilibrium is simple and attractive to those in permanent disequilibrium, like human beings. The concept made immediately clear that human populations were essentially stable in genetical terms: a system, like the genome of a population, is at equilibrium in time when no net change occurs or is expected to occur from its state at that time. Furthermore, equilibria can be stable, meaning that when a small displacement occurs, the system is expected to return to the equilibrium.

Soon after the rediscovery of Mendel's remarkable work in 1900, interest arose in the properties of Mendelian genes in populations; this was the dawn of population genetics. Castle (1903) and Pearson (1903a, 1903b) were among the first to investigate these. As Edwards (in press) has pointed out, Castle did not derive a generalization equivalent to Pearson's, and will be considered no further in this paper.

Mendel (1865) had hypothesized that inheritance of a trait was particulate, that the units of inheritance did not change from generation to generation, that they were contributed equally by an organism's two parents, and that each parent contributed its unit at random from the two it contained. He had shown that a cross between two pure-breeding lines, termed *A* and *a* in regard to some trait, gave a first generation resembling one of the two parents (*A*) identically and that crosses among members of this first generation gave a ratio of 3:1 of the two parental types, the more frequent type being the same as this first generation (*A*). He had described the first generation type as dominant, the other as recessive.

Mendel produced a model of the following kind: the genetic make-up of the two parental lines was *AA* and *aa* respectively, and their offspring were *Aa*. Crossing two *Aa* gave, by the binomial expansion (½*A* + ½*a*)(½*A* + ½*a*), ¼*AA*, ¼*Aa*, ¼*aA*, ¼*aa*. If *Aa* = *aA* (since the units of inheritance are unchanged), and if *Aa* resembles *AA* exactly (the phenomenon of dominance, deduced from the disappearance of *a* in the cross of the two lines), then the proportions of the phenotypes A:a will be ¾:¼. Mendel made further confirmatory crosses, for example showing that two thirds of the *A* types were *Aa* and one third were *AA*.

Mendel did not consider directly what would happen in a population of an organism, but this was

essential if evolutionary phenomena were to be explained or even studied at the level of the hypothesized 'essential character', as Mendel called his fundamental particles of inheritance. (The name 'gene' was introduced in about 1905 by Wilhelm Johannsen, as an abbreviation of Darwin's and De Vries's 'pangen'.)

Pearson immediately saw that dominance/recessiveness was not essential to the dynamics of the model, but was rather an additional assumption of Mendel's. He generalized the model by removing this assumption, and also began the analysis of multiple independent genes.

On his model, the first two generations described above would be

$$(AA') \times (aa') = (A+A')(a+a') =$$
$$(Aa) + (Aa') + (A'a) + (A'a'),$$

representing the parents, the gametes and the offspring in turn. If the gametes identified by the prime ´ are actually identical, that is $A = A'$ and so on, then this second generation is identical, Aa. This can then be extended to multiple independent genes. Pearson (1903b) wrote:

> If these hybrids now breed at random and are equally fertile among themselves, segregation takes place. If the process of random mating with equal fertility be continued generation by generation, what further changes, if any, take place, and what are the laws of inheritance within such a population? (p. 506)

His conclusion was 'that when the members of this segregating population cross at random the population accurately reproduces itself, and supposing no artificial, natural or reproductive selection takes place, a stable population or 'race' is created, which is permanent and shows a permanent proportional frequency for each sub-class of the population'. From this important conclusion, Pearson went on to calculate parent-offspring correlations, r_{OP}, and other attributes of the quantitative inheritance which he was developing. Unfortunately, having pointed out that he did not need Mendel's hypothesis of dominance, he calculated $r_{OP} = \frac{1}{3}$ and noted that this was not in agreement with empirical observations which lay round 0.5. For this and other reasons he abandoned particulate inheritance of the Mendelian kind. Had he assumed that the heterozygote was intermediate, he would have obtained $r_{OP} = \frac{1}{2}$. After the publication of Hardy's note, Pearson (1909a, 1909b) obtained correct results, without referring to his earlier errors as such. It is perhaps unsurprising but nevertheless noteworthy that the teutonophile Pearson, aware of Weinberg's fine work on the familial incidence of cancer (Weinberg & Gaspar, 1904) at least through attendance at and participation in a major meeting on genetics of human disease in 1908 (see Church, 1908), did not cite Weinberg (1908).

The Initial Work of Hardy and Weinberg

Consider a population in which a diallelic gene with two alleles, A_1 and A_2, are segregating. Suppose that the frequencies of the three possible genotypes are as shown:

A_1A_1	A_1A_2	A_2A_2
P	Q	R

Then the frequencies of the alleles A_1 and A_2 are $P + \frac{1}{2}Q$ and $\frac{1}{2}Q + R$ respectively. Call these p and q respectively. Call this population the parental generation.

Some very simple algebra shows that the frequencies of the three genotypes in the offspring produced by random mating among this parental generation will be

A_1A_1	A_1A_2	A_2A_2
$(P + \frac{1}{2}Q)^2$	$2(P + \frac{1}{2}Q)(\frac{1}{2}Q + R)$	$(\frac{1}{2}Q + R)^2$
$= p^2$	$= 2pq$	$= q^2$

It is also simple to show that these genotypic frequencies, which are also those chosen by pairwise sampling of gametes at random in the population, will be the same in the next generation. (At this equilibrium, $Q^2 = 4PR$.) The process of pairwise sampling is simple binomial sampling with replacement, justifiable because the populations of gametes can be regarded as indefinitely large. For a careful and complete mathematical account of HWE, see Edwards (2000).

The English pure mathematician G. H. Hardy (1908), notable for contributions to number theory and analysis, simply showed that the relationship given above would hold; he had been asked what would happen to gene and genotype frequencies in a population mating at random, and gave the answer. He participated no further in population genetics. Diaconis (2002) has speculated that Hardy had 'a true antipathy to the subject' of probability, which could explain his failure to contribute further, but it could equally well be explained by his love of pure mathematics and total lack of interest in applications. Hardy's place in mathematical history is secure; that in genetical history minor but significant. For an accessible portrait, see Snow (1967) and Hardy's own memoir (1940), and for detailed comments on Hardy's (1908) paper, see Edwards (in press).

Weinberg (1908), who was a human geneticist of the first rank, though widely regarded in his own country as an *Armenarzt* (a doctor employed by a local authority to treat the indigent, an honorable calling, perhaps, but hardly a sign of success in his career), did much more work on the topic.

Weinberg's Contribution

Wilhelm Weinberg was born in Stuttgart in 1862, was educated in Stuttgart, Tübingen and Munich, worked as 'poor doctor', public health adviser and private practitioner in Stuttgart, and died after some years of poor health in Tübingen in 1937, though he remained scientifically productive until his death.

As well as demonstrating how HWE must arise in the diallelic case, Weinberg (1908, 1909a, 1909b) also considered multiple alleles and multiple independent genes. In this last case, he showed that the approach to multilocus HWE would be asymptotic, not the result of one generation of panmixia.

Apart from HWE, his major contributions were to quantitative inheritance (correlations between relatives), twin studies, segregation analysis and the merits (or otherwise) of eugenics. Crow (1999) gives a thoughtful appraisal of many aspects of his work.

Hill (1984) and Crow (1999) give an account of Weinberg's (1909a, 1909b, 1910) work on quantitative genetics, which in some ways anticipated developments by Fisher and Wright. See the next section for further discussion.

'Weinberg's studies on the frequency of twins and higher multiple births are the best studies ever published on this subject' (Bulmer, 2003). Bulmer (1970) had earlier shown how good these studies are.

Weinberg (1901) systematically developed his differential method for determining the frequencies of monozygotic and dizygotic twinning. In Bulmer's notation, suppose

L = number of like-sexed twin maternities in a total sample of N maternities

U = number of unlike-sexed twin maternities in the sample.

Then the monozygotic twinning rate is given by
$$m = (L-U)/N$$
and the dizygotic rate by
$$d = 2U/N.$$
Bulmer shows that
$$Var(m) = (m + d)/N$$
and
$$Var(d) = 2d/N$$
approximately.

In an example given by Bulmer, in 791,584 maternities in Wales in 1960, $L = 5,894$ and $U = 3,192$. Then $m = 0.0034 \pm 0.0001$ and $d = 0.0081 \pm 0.0001$. As presented, a constant sex-ratio of unity is assumed, but the bias engendered thereby is very small, as Weinberg understood. Indeed, in 1934, he published a paper on this topic, including a method for estimating the precision of the estimates rather more complex than Bulmer's (see also James 2007).

Weinberg (1909a, 1909b) investigated the inheritance of twinning, showing that a propensity to produce dizygotic twins is inherited, though he could not investigate this further by statistical methods. He also concluded that there was no inherited propensity to produce monozygotic twins. These results have largely been borne out by subsequent work (Bulmer, 1970; Fisher, 1928; Hoekstra et al., 2007).

In human genetics, breeding experiments not being possible (even ignoring ethical issues), methods have had to be developed to detect Mendelian inheritance, determine its form, and to investigate linkage and interaction of Mendelian factors, by analysis of observed (ascertained) families. In these circumstances, random sampling and inference therefrom are not always possible, and bias and other statistical problems have to be carefully avoided. To take a very simple example, suppose that a deleterious trait is suspected to be a simple recessive. The segregation ratio in the offspring of two normal carriers is then expected to be Mendel's normal dominant:affected recessive ratio of 3:1. However, human families are generally small, and for sibships of size 1, 2, 3… the probability of all dominants is ½, ¼, ⅛ … Thus, if the trait is ascertained through affected children, the observed ratio will be lower than 3:1. As noted by Bailey (1961), the problem is to fit a binomial distribution with its initial term missing, and Weinberg was the first to analyze this case in human genetics. Weinberg's solution (the proband method) is the maximum likelihood solution if ascertainment has been both random and complete (Fisher, 1934). That this condition is not often met was recognized by Weinberg (1912a, 1912b, 1927), who consequently preferred what he called the sib method, whereby the segregation ratio in sibs of propositi is measured, weighting each sibship by the number of affected individuals. He also developed a method for the case where ascertainment was incomplete. In developing these methods, Weinberg was drawing on his experience of working with poor families, both in public employment and privately, recognizing how genetic studies were often not the mainspring for data collection, and how, consequently, data might be incomplete and biased. However, he also showed remarkable statistical insight and expertise.

Demmler (2003) has pointed out that Weinberg was one of the earliest German medical scientists to understand and apply Mendel's laws. Thus, he was always having to battle against others, such as the psychiatric geneticist Wilhelm Strohmeyer (1874-1936) who had accepted Darwinism, including Darwin's fallacious blending inheritance (Demmler, 2003, p. 74–76). Strohmeyer, writing to his wife from a conference in 1912, reported that he had met his 'antagonist' Weinberg, who 'talked a great deal, but was clever and industrious' (Demmler, 2003, p. 75). Strohmeyer was also rather in favor of eugenics, more so as the years passed, whereas Weinberg was against it; he 'noted that for "race-hygienic" reasons tight boundaries would have to be drawn round any intervention. An improvement in "national efficiency" without reduction of population size Weinberg saw as the more favorable way; 'a sober statistical view leading thus to the result that eugenics does not require abortion and sterilization particularly strongly and indeed prominent personalities would not support such an approach' (Weinberg, 1918, as cited in Demmler, 2003, p. 81). Weinberg was right about the science, but not about the way that 'prominent personalities' would fail to support 'sterilization of the unfit.' For further detail, see Früh (1996, 1999).

Subsequent Work on HWE

The importance of HWE as a basis for investigation of the effects on gene frequencies of selection, mutation, inbreeding and other factors was immediately recognized, though not all workers in the next decade acknowledged the earlier workers.

Suppose that a sample is obtained from a population, and the numbers and frequencies observed are:

A_1A_1	A_1A_2	A_2A_2	Total
a	b	c	n
P	Q	R	1

Then an obvious test to determine whether HWE holds is Pearson's χ^2 whereby

$$(a-n(P + \tfrac{1}{2}Q)^2)^2/(n(P + \tfrac{1}{2}Q)^2) +$$
$$(b-2n(P + \tfrac{1}{2}Q)(R + \tfrac{1}{2}Q))^2/n(P + \tfrac{1}{2}Q)(R + \tfrac{1}{2}Q)) +$$
$$(c-n(R + \tfrac{1}{2}Q)^2)^2/(n(R + \tfrac{1}{2}Q)^2)$$

will be distributed as χ^2 with one degree of freedom if HWE holds. This test was available to the early workers, though there was uncertainty about degrees of freedom. (In this case, there are three classes, yielding two degrees of freedom, but one is associated with estimation of gene frequency in order to estimate genotypic expectations). It is immediately clear that, on the assumption of HWE, the observed frequency of recessives c/n is an estimator for q^2, and this relationship was used early, for example in estimating gene frequencies for the ABO blood group system (see Kempthorne, 1957). Later developments will be discussed in the next section.

Norton, in an appendix to Punnett (1915), appears to have been the first to have considered the effects of selection on a gene in a panmictic, indefinitely large population. Since Punnett had led Hardy to consider the problem, and Norton was a student of Hardy's (see Edwards, in press), Norton must have been familiar with Hardy (1908), but he did not cite the paper.

Fisher (1918) used HWE without comment; we do not know whether he knew of Hardy's or Weinberg's papers, though he was acquainted with Pearson's earlier work, cited above, and knew and valued Hardy highly (Fisher, 1958). HWE was the basis for Fisher's derivation of correlations between related individuals under Mendelian inheritance. In this notable paper, Fisher also presented one of the first analyses of departures from panmixia. His development of the concept of balanced polymorphism also required HWE (see Fisher, 1922; Lewontin, 1958; Mayo, 2007).

It is perhaps worth noting that Fisher wrote to Weinberg on August 29, 1930, in the course of a cordial correspondence mainly about ascertainment: 'I am sure you will always be honoured abroad, and I hope also in your own country for your pioneer work upon the Mendelian or other interpretation of human data' (Fisher, 1930, August 29). Even here, Fisher does not notice the contribution honoured as HWE.

Haldane (1924a) called HWE the Hardy–Pearson rule, and used it as the basis for all his important early work on selection in populations. He derived the recurrence relationship for the approach to equilibrium of gene and genotypic frequencies of an X-linked diallelic gene; both the gene frequencies in the two sexes and the genotypic frequencies in females approach equilibrium asymptotically (see Bennett & Oertel, 1965, for a definitive analysis). In a series of papers, Haldane (1924a, 1924b, 1926a, 1926b, 1927

etc.) also showed how selection and mutation could be taken into account, introducing the concept of mutation-selection balance. To take a very simple example, consider selection against a deleterious recessive:

	AA	Aa	aa
Frequency	p^2	$2pq$	q^2
Fitness	1	1	$1-s$

Here, the frequency of A in the progeny is

$$p' = (p^2 + pq)/(p^2 + 2pq + q^2(1-s)) = p/(1-q^2s).$$

Then the change in p through selection is $q^2/(1-q^2s)$. If gene frequency is not to change over time, this increase must be balanced by mutation at the rate μ from A to a, that is $\mu p = q^2/(1-q^2s)$. If s is not small relative to μ, then $q = (\mu/s)^{1/2}$ approximately.

On a similar argument, in the X-linked case, $q = 3\mu/s$. If the mutation rate is in males and in females and these are unequal, $q = (\mu + 2\nu)/s$ to the same level of approximation. Using this argument, Haldane (1935, 1947) was the first to estimate human mutation rates (3.2×10^{-5} for haemophilia) with μ > ν, possibly by a factor of 10. Recent analyses show this to be of the correct order (e.g., Ellergren, 2002). Weinberg (1912b) was one of the first to recognize that mutation might maintain deleterious traits in this way, and also that mutation might be more frequent in males than in females, at a time when mutation was an underdeveloped and misunderstood concept.

Chetverikov (1926) referred to HWE as Hardy's law, and (possibly through misunderstanding of Pearson's (1903a, 1903b) work) a slightly different rule as Pearson's law. Despite his familiarity with the German literature, Chetverikov did not cite Weinberg. Chetverikov was one of the first experimental scientists to understand the implications of population genetics, especially the role of finite population size, and to try to investigate natural populations from this standpoint.

Fisher, Haldane, Wahlund (1928), Wright (1931) and others showed that departures from random mating such as inbreeding, whether systematic (as in plants, or in animal and plant breeding) or chance (as in the effect of finite population size), could induce departures from expected Hardy–Weinberg proportions. Perhaps it should be mentioned that such departures from random mating must influence genotypic frequencies for every gene, which is not the case for selection and mutation.

Wahlund showed that HWE proportions would not be found in a population composed of isolated subpopulations, even if each subpopulation were itself in HWE. The frequencies would be

A_1A_1	A_1A_2	A_2A_2
$p^2 + V_p$	$2(pq - V_p)$	$q^2 + V_p$

where p is the mean frequency of A_1 in the whole population and V_p the variance in gene frequency among the subpopulations.

Inbreeding at the rate F will give rise to a similar increase in the frequencies of the homozygotes and decrease in that of heterozygotes. In the case of a finite

population of size N, the chance of identity by descent of two randomly chosen alleles is $1/2N$, and this is the increase in inbreeding each generation. See Wright (1922, 1931) for the original work. It should be noted once more that HWE arises from two phenomena: binomial sampling of gametes and panmixia. Hence, if subpopulations or lines are isolated so that inbreeding and differentiation arise but both phenomena continue to apply within a line or subpopulation, there can still be HWE within it.

Most methods for the analysis of gene and genotype frequencies in finite populations require HWE as a starting point, from which gene trajectories in time etc. can be pursued. Even when departures from HWE are caused by breeding systems, random mating generates, because of the independent binomial sampling of alleles in the two parents, genotypic regularities in the progeny that allow assessment of the effects of the breeding system on the genome. For example, gametophytically determined and other self-incompatibility systems which prevent selfing have been thoroughly analyzed at both the infinite and finite population level (Wright 1939, Leach & Mayo, 2005).

HWE Today

Li (1988), followed and elaborated by Stark (2006a, 2006b), showed that panmixia is not the only breeding structure that can yield HW proportions, so that panmixia is a sufficient but not a necessary condition for HWE. However, no natural population is known to manifest the other possible breeding structures so that it appears unlikely that they need to be considered in data collection and analysis. HWE continues to be an important starting point for any population analysis. This will indeed be true even when what is being analyzed is something that must initially disrupt the regularity of the meiotic processes that provide the basis for HWE (e.g. the investigation of the fate of newly arising duplications; see Force et al., 1999; Hittinger & Carroll, 2007).

Testing for departure from HWE began, as noted above, with simple χ^2 analysis. Problems inherent in such χ^2 analyses, especially the dependence on sample size, and the low power of the χ^2 HWE test (investigated thoroughly by Lewontin & Cockerham, 1959), meant that Haldane (1954) and others sought 'exact' tests based on the expectation that, under HWE, $Q^2 - 4PR = 0$. These tests have been extensively developed and used; see Mantel & Li (1974) and Rousset & Raymond (1995). Rousset & Raymond consider the issue of the alternative hypothesis: is one concerned with selection or a disturbance to panmixia? Guo & Thompson (1992) and Rousset & Raymond consider 'exact' tests for the case of multiple alleles; microsatellite markers are widely used in population studies, and multiple alleles are the norm with these markers. Many coding genes also exhibit multiple alleles and are significantly associated with disease. For example, consider the data on human *APOE* in

Table 1

Distribution of *APOE* Genotype by Ethnic Classification (adapted from Table 2 of Kimmel et al. 2008)

APOE genotype	African–American	Caucasian	Total
2–2	3	0	3
2–3	15	16	31
2–4	7	3	10
3–3	51	73	124
3–4	30	28	58
4–4	5	1	6
Total	111	121	232

Table 1. By Fisher's 'exact' test, the two ethnic classifications differ significantly in genotype frequencies. Both groups fit HWE by various tests. Given that *APOE* is known to be strongly associated with certain disease states (see Song et al., 2004; Kimmel et al., 2008, the source of these data), is this agreement surprising or not?

Today, following the development of automated methods of DNA sequencing and the consequent numerous genomic analyses, we know that single nucleotide polymorphisms (SNPs) are prodigiously numerous in the genome. In the human genome, for example, there may be ten million. Since they arise from DNA misreplication at a rate of, perhaps, 10^{-8} per base per generation (Nachman & Crowell, 2000), the overwhelming majority of these will be diallelic. Suppose that there are ten million SNPs in humans. Then a mutation yielding a third allele is expected every ten generations. In cattle, more than 98% of SNPs are diallelic (Wade & Adelson, personal communication). Neutral variants are fixed at a rate approximately given by the mutation rate (Kimura, 1968, 1983).

Considered individually, each diallelic SNP may be expected to be in HWE. Obviously, inbreeding, selection, mutation etc. are certain to be present in every population studied, but one has no expectation a priori that they will influence any particular SNP. Since, to the contrary, each pair of linked SNPs is unlikely to be in linkage disequilibrium, testing individual SNPs for departure from HWE has been widely regarded as a useful first step in genome scans for regions associated with important traits (multifactorial disorders in humans, production traits in plants and animals). Since disturbances to panmixia are expected to affect all sites in the genome, rare departures from HWE found for a few SNPs in the early stages of a genome-wide scan are likely to arise by chance or possibly from mistyping.

However, Zou & Donner (2006) dispute the validity and hence utility of such testing in case-control studies. (See their paper for many references supporting such preliminary screening).

In general, if a departure from HWE is shown to be of interest, it should indicate, through a deficiency of either or both homozygotes, or of heterozygotes, etc., the type of explanation that might be sought. Since an individual SNP is unlikely to be influential in selection (recognizing that exceptions, like the malaria-related polymorphisms, exist), patterns involving closely linked SNPs must then be examined. Weir *et al.* (2004) provide an appropriate statistical approach to follow in these circumstances.

While technical mistakes are more likely a priori than selection as a cause of disturbed segregation, nonconforming loci must be followed up. As Edwards (2007) wrote of one major continuing study:

> The HapMap set of data restricts analysis to loci conforming to the simple genetic background imposed by rejecting genotypes inconsistent with Mendel's first law, and consistent with what Stern termed the Hardy–Weinberg law (Hardy, 1908; Stern, 1943; Weinberg, 1908). The rejects — the golden dross for the recognition of recessive lethals — are not discussed in detail although they account for over 10% of loci even though rejection was based on very high levels of significance ($P < .001$). (p. 390)

As outlined in the previous section, HWE is the fundamental starting point for all population–genetical investigation, whether the goal is detection or estimation of the effects of all the forces that disrupt HWE. While Mendel conceived the independent binomial sampling of gametes from parents and hence could be regarded as the first to have considered a population–genetical example (the effects of crossing of pure lines) (Edwards, in press), the generalization to arbitrary gene frequencies to give HWE was the true foundation of population genetics.

Acknowledgments

References

Bailey, N. T. J. (1961). *Introduction to the Mathematical Theory of Genetic Linkage*. New York: Oxford University Press.

Bennett, J. H., & Oertel, C. R. (1965). The approach to random association of genotypes with random mating. *Journal of Theoretical Biology, 9*, 67–76.

Bulmer, M. G. (1970). *The biology of twinning in man*. Oxford: Clarendon Press.

Bulmer, M. G. (2003). *Francis Galton, pioneer of heredity and biometry*. Baltimore, MD: Johns Hopkins University Press.

Castle, W. E. (1903). The laws of Galton and Mendel and some laws governing race improvement by selection. *Proceedings of the American Academy of Arts and Science, 35*, 233–242.

Chetverikov, S. S. (1926). On certain aspects of the evolutionary process from the standpoint of modern genetics. *Zhurnal Eksperimental'noi Biologii, A2*, 3–54. (Originally published in Russian; this translation by M. Barker originally published in 1961 in *American Philosophical Society Proceedings, 105*, 167–195 and reprinted in D. L. Jameson (Ed.) (1977), *Benchmark papers in genetics, vol. 8 evolutionary genetics*. Stroudsburg, PA: Dowden, Hutchinson & Ross, pp. 234–262.)

Church, W. B. (1908). The influence of heredity on disease, with special reference to tuberculosis, cancer and diseases of the nervous system (with discussion). *Journal of the Royal Society of Medicine, 2*, 8–142.

Crow, J. F. (1999). Hardy, Weinberg and language impediments. *Genetics, 152*, 821–825.

Demmler, A. (2003). Wilhelm Strohmeyer (1874–1936) Ein Wegbereiter der Kinder- und Jugendpsychiatrie. MD dissertation, Friedrich-Schiller-Universität Jena.

Diaconis, P. (2002). G. H. Hardy and probability??? *Bulletin of the London Mathematical Society, 34*, 385–402.

Edwards, A. W. F. (2000). *Foundations of Mathematical Genetics* (2nd ed.). Cambridge: Cambridge University Press.

Edwards, A. W. F. (in press). G. H. Hardy (1908) and Hardy–Weinberg Equilibrium. *Genetics*.

Edwards, J. H. (2007). Genome scans and the 'old genetics'. In O. Mayo and C. R. Leach (Eds.), *Fifty years of human genetics a Festschrift and liber amicorum to celebrate the life and work of George Robert Fraser* (pp. 385–401). Adelaide, Australia: Wakefield Press.

Ellergren, H. (2002). Human mutation — blame (mostly) me. *Nature Genetics, 31,* 9–10.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399–433.

Fisher, R. A. (1922). On the dominance ratio. *Proceedings of the Royal Society of Edinburgh, 42*, 321–341.

Fisher, R. A. (1928). Triplet children in Great Britain and Ireland. *Proceedings of the Royal Society of London, Series B, 102*, 286–311.

Fisher, R. A. (1930, August 29). Archived personal communication. Available from http://digital.library. adelaide.edu.au/coll/special/fisher/corres/weinberg/ Weinberg300829a.html

Fisher, R. A. (1934). The effect of method of ascertainment upon the estimation of frequencies. *Annals of Eugenics, 6,* 13–25.

Fisher, R. A. (1958). The nature of probability. *Centennial Review, 2,* 261–274.

Force, A., Lynch, M. T., Pickett, F. B., Amores, A., Yan, Y., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics, 151,* 1531–1545.

Früh, D. (1996). Wilhelm Weinberg (1862–1937), Armenarzt und Populationsgenetiker — Anmerkungen zu Leben und Werk. [Wilhelm Weinberg (1862–1937), doctor to the poor and population geneticist — notes on his life and work]. *Biologisches Zentralblatt, 115,* 112–119.

Früh, D. (1999). Die Genealogie als Hilfswissenschaft der Humangenetik. [Genealogy as supporting science for human genetics]. *Jahrbuch für Geschichte und Theorie der Biologie, 6,* 141–162.

Guo, S. W. & Thompson, E. A. (1992). Performing the exact test of Hardy–Weinberg proportions for multiple alleles. *Biometrics, 48,* 361–372.

Haldane, J. B. S. (1924a). Part I. A mathematical theory of natural and artificial selection. *Transactions of the Cambridge Philosophical Society, 23,* 19–41.

Haldane, J. B. S. (1924b). Part II. The influence of partial self-fertilization, inbreeding, assortative mating and selective fertilization on the composition of Mendelian populations and on natural selection. *Proceedings of the Cambridge Philosophical Society, 1,* 158–163.

Haldane, J. B. S. (1926a). Part III. *Proceedings of the Cambridge Philosophical Society, 23,* 363–372.

Haldane, J. B. S. (1926b). Part IV. *Proceedings of the Cambridge Philosophical Society, 23,* 607–615.

Haldane, J. B. S. (1927). Part V. Selection and mutation. *Proceedings of the Cambridge Philosophical Society, 23,* 838–844.

Haldane, J. B. S. (1935). The rate of spontaneous mutation of a human gene. *Journal of Genetics, 31,* 317–326.

Haldane, J. B. S. (1947). The mutation rate for haemophilia, and its segregation ratios in males and females. *Annals of Eugenics, 13,* 262–277.

Haldane, J. B. S. (1954). An exact test for randomness of mating. *Journal of Genetics, 52,* 631–635.

Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science, 28,* 49–50.

Hardy G. H. (1940). *A mathematician's apology* (reprint edition, 1992). Cambridge: Cambridge University Press

Hill, W. G. (Ed.). (1984). *Quantitative genetics, part I.* New York: Van Nostrand Reinhold.

Hittinger, C. T. & Carroll, S. B. (2007). Gene duplication and the adaptive evolution of a classic genetic switch. *Nature, 449,* 677–682.

Hoekstra, C., Zhao, Z. Z., Lambalk, C. B., Willemsen, G., Martin, N. G., Boomsma, D. I. & Montgomery, G. W. (2007). Dizygotic twinning [Electronic Version]. *Human Reproduction Update Advance Access, 14,* 37–47.

James, W. H. (2007). The validity of Weinberg's differential rule. *Twin Research and Human Genetics, 10,* 771–2.

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

Kimmel, S. E., Christie, J., Kealey, C., Chen, Z., Price, M., Thorn, C. F., Brensinger, C. M., Newcomb, C. W. & Whitehead, A. S. (2008). Apolipoprotein E genotype and warfarin dosing among Caucasians and African Americans. *The Pharmacogenomics Journal, 8,* 53–60.

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217,* 624–626.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

Leach, C. R. & Mayo, O. (2005). *Outbreeding mechanisms in flowering plants: An evolutionary perspective from Darwin onwards.* Stuttgart: Gebrüder Borntraeger Verlagsbuchhandlung.

Lewontin, R. C. (1958). A general method for investigating the equilibrium of gene frequency in a population. *Genetics, 43,* 419–434.

Lewontin, R. C., & Cockerham, C. C. (1959). The goodness-of-fit test for detecting natural selection in random mating populations. *Evolution, 13,* 561–564.

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

Mantel, N. & Li, C. C. (1974). Estimation and testing of a measure of non-random mating. *Annals of Human Genetics, 37,* 445–454.

Mayo, O. (2007). The rise and fall of the common disease-common variant (CD-CV) hypothesis: How the sickle cell disease paradigm led us all astray (or did it?). *Twin Research and Human Genetics, 10,* 793–804.

Mendel, J. G. (1865). Versuche über Pflanzenhybriden. [Experiments in plant hybridisation]. *Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr1865,* 3–47.

Nachman, M. W. & Crowell, S. L. (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics, 156,* 297–304.

Pearson, K. (1903a). Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Philosophical Transactions of the Royal Society of London, Series A, 200,* 1–66.

Pearson, K. (1903b). Mathematical contributions to the theory of evolution. XII. On a generalised theory of alternative inheritance, with special reference to

Mendel's laws. *Proceedings of the Royal Society of London, 71*, 505–509.

Pearson, K. (1909a). The theory of ancestral contributions in heredity. *Proceedings of the Royal Society of London, 81*, 219–224.

Pearson, K. (1909b). On the ancestral gametic correlations of a Mendelian population mating at random. *Proceedings of the Royal Society of London, 81*, 225–229.

Punnett, R. C. (1915). *Mimicry in butterflies*. Cambridge: Cambridge University Press.

Rousset, F. and Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics, 140*, 1413–1419.

Snow, C. P. (1967). *Variety of men*. London: Macmillan.

Song, Y., Stampfer, M. J. & Liu, S. (2004). Meta-analysis: Apolipoprotein E genotypes and risk for coronary heart disease. *Annals of Internal Medicine, 141*, 137–147.

Stark, A. E. (2006a). A clarification of the Hardy–Weinberg law. *Genetics, 174*, 1695–1697.

Stark, A. E. (2006b). Stages in the evolution of the Hardy–Weinberg law. *Genetics and Molecular Biology, 4*, 589–594.

Stern, C. (1943). The Hardy–Weinberg law. *Science, 97*, 137–138.

Wahlund, S. (1928). Zusammensetzung von population und korrelationserscheinung vom standpunkt der vererbungslehre aus betrachtet. [Population composition and correlation structure considered from the viewpoint of genetics]. *Hereditas, 11*, 65–106.

Weinberg, W. (1901). Beiträge zur physiologie und pathologie der mehrlingsgeburten beim menschen. [Contributions to the physiology and pathology of human multiple births]. *Pflügers Archiv ges. Physiologie, 88*, 346–430.

Weinberg, W. (1908). Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg, Stuttgart* 64: 369–382. [On the demonstration of inheritance in humans]. Translation by R. A. Jameson printed in D. L. Jameson (Ed.), (1977). *Benchmark papers in genetics, Volume 8: Evolutionary genetics* (pp. 115–125). Stroudsburg, PA: Dowden, Hutchinson & Ross.

Weinberg, W. (1909a). Über Vererbungsgesetze beim Menschen I. [On laws of heredity in humans]. *Zeitschrift für induktive Abstammungen- und Vererbungslehre, 1*, 377–392, 440–460.

Weinberg, W. (1909b). Über Vererbungsgesetze beim Menschen II. [On laws of heredity in humans]. *Zeitschrift für induktive Abstammungen- und Vererbungslehre, 2*, 276–330.

Weinberg, W. (1910). Statistik und Vererbung in der Psychiatrie. [Statistics and genetics in psychiatry]. *Klinik für psychische und nervöse Krankheiten, 5*, 34–43.

Weinberg, W. (1912a). Methode und Fehlerquellen der Untersuchung auf Mendelschen Zahlen beim Menschen. [Methods and sources of error in the investigation of Mendelian ratios in humans]. *Archiv für Rassen- und Gesundheits Biologie, 9*, 165–174.

Weinberg, W. (1912b). Zur Vererbung der Anlage der Blutkrankheit mit methodologischen Ergänzungen meiner Geschwistermethode. [On the inheritance of predisposition to blood disease with methodological additions to my sib method]. *Archiv für Rassen- und Gesundheits Biologie*, 9, 694–709.

Weinberg, W. (1918). Künstliche Fehlgeburt und künstliche Unfruchtbarkeit vom Standpunkt der Statistik. [Artificial abortion and artificial infertility from the point of view of statistics]. In S. Placzek (Ed.), *Künstliche Fehlgeburt und künstliche Unfruchtbarkeit: ihre Indikationen, Technik und Rechtslage* (pp. 437–456). Leipzig: Thieme Verlag.

Weinberg, W. (1927). Grundlagen der Probandenmethode. [Foundations of the proband method]. *Zeitschrift für induktive Abstammungen- und Vererbungslehre, 48*, 179–228.

Weinberg, W. (1934). Differenzmethode und Geburtenfolge bei Zwillinge (nebst einem Anhang über dem mittleren Fehler der Geburtenfolgen-nummer). [Difference method and birth order in twins (together with an appendix on the mean error of the birth rank)]. *Genetica, 16*, 383–388.

Weinberg, W. & Gaspar, K. (1904). Die bösartigen Neubildungen in Stuttgart von 1878 bis 1902. [Malignant tumours in Stuttgart from 1878 to 1902]. *Zeitschrift für Krebsforschung, 2*, 195–260.

Weir, B. S., Hill, W. G. & Cardon, L. R. (2004). Allelic association patterns for a dense SNP map. *Genetic Epidemiology, 27*, 442–450.

Wright, S. (1922). Coefficients of inbreeding and relationship. *American Naturalist, 56*, 330–338.

Wright, S. (1931). Evolution in Mendelian populations. *Genetics, 16*, 97–159.

Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics, 24*, 538–552.

Zou, G. Y. & Donner, A. (2006). The merits of testing Hardy–Weinberg equilibrium in the analysis of unmatched case-control data. *Annals of Human Genetics, 70*, 923–933.