

Chapter 2

Terminology, Concepts, and Models in Genetic Epidemiology

M. Dawn Teare and Mauro F. Santibàñez Koref

Abstract

Genetic epidemiology brings together approaches and techniques developed in mathematical genetics and statistics, medical genetics, quantitative genetics, and epidemiology. In the 1980s, the focus was on the mapping and identification of genes where defects had large effects at the individual level. More recently, statistical and experimental advances have made possible to identify and characterise genes associated with small effects at the individual level. In this chapter, we provide a brief outline of the models, concepts, and terminology used in genetic epidemiology.

Key words: Population genetics, Mendelian segregation, Kinship, Identity by descent, Genetic components of variance

1. Introduction

Genetic epidemiology studies the influence of genes and environment on measures of health and disease susceptibility in populations. This discipline emerged relatively recently and brings together established methodologies arising from population genetics, quantitative genetics, medical genetics, and epidemiology. Much of the terminology currently used was conceived when little was known about the molecular mechanisms mediating inheritance (1). The term *gene* is now frequently used to refer to a functional segment of DNA, which is transcribed into RNA and may code for a protein. However, within the field of population genetics “gene” continues to be used in its original meaning, and refers to the basic unit of heredity. As with any speciality the terminology has become specialised, and this in itself can form a potential barrier to newcomers. The purpose of this chapter is to present the basic terminology and outline the basic models used in genetic epidemiology.

2. Mendelian Genetics and Modes of Inheritance

Gregor Mendel was the first to propose a discrete model to explain the inheritance of genetic factors and their impact upon an organism's *phenotype* (2). By phenotype we mean an individual's measurable characteristics or *traits*. When Mendel reported the results of his experiments on pea plants in 1865, he focused his attention on qualitative phenotypes of his plants, such as pea seed coat shape (wrinkled vs. round) or flower colour (white vs. violet). He postulated a mechanism of inheritance in which each organism carries two factors that determine together the organisms' phenotype and that an adult organism can only transmit one of the two factors to each of its offspring.

With respect to the phenotype of *wrinkled seed coat* or *round seed coat*, he labelled the two possible factors *r* and *w*. For this phenotype, he proposed that the factor *r* was dominant to *w* (or conversely *w* was recessive to *r*). This means that the recessive phenotype of "wrinkled" is seen only in peas with two copies of the recessive factor *w*. Conversely, those pea plants with one or two copies of the dominant factor *r* would express the dominant phenotype of round seeds. We would now refer to these factors *r* and *w* as *alleles* of the gene determining the variation in seed coat shape. The physical location of this gene in the pea genome is referred to as the *locus* (plural *loci*). There are examples of alleles that are *codominant*, where in individuals carrying two different alleles the phenotypes characteristic for both alleles are displayed. For example, the ABO human blood group system has three classes of alleles, A, B, and O. The allele O is recessive with respect to A or B, but A and B are codominant, this gives rise to a four phenotype system (namely, blood groups A, B, AB, and O).

Mendel's extensive experiments on peas led him to propose two laws: the law of segregation and the law of independent assortment. The law of segregation stipulates that when an organism produces gametes the two copies of the gene separate so that each gamete randomly receives one allele. The law of independent assortment states that alleles at different loci are inherited independently from alleles at other loci, or that alleles of different genes segregate independently during gamete formation. We now know that genes are physically linked to other genes due to their location on chromosomes. Mendel happened to study traits arising from unlinked loci. The genes determining his seven phenotypes are each located on a different chromosome. Though this was true for the traits Mendel studied, the law of independent assortment is generally true when loci are not genetically linked. We now use the term *Mendelian segregation* to describe the pattern of allele transmission from one generation to the next, meaning that the probability of a parent transmitting one specific allele to one specific offspring is 50%.

The pair of alleles found at a single locus in a diploid organism is referred to as the *genotype*. At a multi-allelic locus, such as ABO described above, there are six possible genotypes (AA, AB, AO, BB, BO, OO). When the two alleles are of the same type (e.g. AA) the individual is said to be *homozygous* at that locus. When both alleles are not the same, the individual is described as *heterozygous* at that locus. Loci are also defined using descriptors of their position in the genome (see Chapter 4).

Mendel's model allows us to introduce the concept of a *penetrance function*. The penetrance function is used in both discrete and quantitative genetics, in the discrete setting it is the probability of having the trait or phenotype state of interest conditional on the genotype. For example, for Mendel's seed shape experiment we discussed earlier, there are three possible genotypes labelled *ww*, *rw*, and *rr*. If we define *round* as the normal or common state and *wrinkled* as the phenotype of interest, then the penetrance function is defined by the three conditional probabilities:

1. $\text{Prob}(\text{wrinkled}/rr) = 0$
2. $\text{Prob}(\text{wrinkled}/rw) = 0$
3. $\text{Prob}(\text{wrinkled}/ww) = 1$

Here, the *mode of inheritance* for the phenotype *wrinkled* is recessive.

3. Population Genetics

Sexual reproduction is a mechanism by which the genetic units are transmitted from one generation to the next. Mendel's model came from his experiments on peas. The diploid system that he discovered determines the distribution or patterns of phenotypes in a population. To illustrate these patterns, let us consider a locus that has two types of alleles which we designate as alleles of type A and of type B. The population relative frequency of the alleles of type A (termed *gene frequency* or *allele frequency*) is denoted as p . As there are only two types of alleles at this locus, and on a single chromosome the allelic state must be A or B, the frequency of alleles of type B is $(1 - p)$. It is important to remember that the population gene frequency refers to the population of chromosomes and not diploid organisms, whereas the term genotype refers to the type of the pair of alleles found at the locus in a single (diploid) organism. In large populations, when alleles are inherited independently, the expected frequencies of the genotypes in each generation are a simple function of the allele frequency, and they do not vary from one generation to the next. When such a state exists, the locus is said to be in Hardy-Weinberg Equilibrium

(HWE). Under these circumstances, the expected *genotype frequencies* can be derived from a binomial distribution, where the probability of success is p and the number of trials is two. The three genotypes AA, AB, and BB should be seen in the frequencies p^2 , $2p(1-p)$, and $(1-p)^2$.

The relation above (which can be extended to multi-allelic systems) holds under *random mating*, when alleles are inherited independently, and in the absence of *selection* or *mutation*. Random mating or *panmixia* means that sexual partners randomly select their partner, i.e. without reference to their genotypic state, their degree of relatedness or physical proximity. When sexual partners do exhibit some preference or selection, this is termed *non-random mating*. Mutation is the mechanism through which new alleles arise or one allele may change from one type to another.

HWE also implies that the gene frequency remains constant from one generation to the next. For this to be the case, all genotypes must be equivalent with respect to *viability* or *fitness* of the organism. Viability can be thought of as an individual's probability of survival or the fraction of the population surviving to reaching maturity. If organisms of one genotype have an advantage over another genotype (e.g. a better chance of survival to reproductive age), then that genotype group will be over represented in the parents of the next generation. The presence of HWE is often used to confirm that a locus is *neutral* (no variation in fitness associated with genotype variation), and hence may be useful as a genetic marker.

3.1. Effects of Population Size

The previous section requires a large population for these general properties to hold. However, all large populations must have gone through a small population phase at some time in their history. When the population is small, HWE may not hold even in the presence of random mating and equal viability. This is due to the random sampling of the gametes, and results in the population gene frequency varying from one generation to the next (*random genetic drift*). As the population size increases, the effect of genetic drift reduces and once the population becomes sufficiently large the gene frequency becomes effectively stabilised.

When a new population is established from a small number of founder individuals, the *founder effect* means the descendent population's genetic variation is limited by the genetic diversity of the *founder population*. When isolated populations remain at small numbers over several generations (through disaster or migration to new territory), this is described as a *bottleneck*, and a significant amount of genetic variation can be lost from the gene pool. The founder effect can be responsible for the different genetic profile in neutral markers seen between populations (3), and also is responsible for some of the different rates of genetic disease seen when comparing isolated populations (4).

3.2. Linkage Disequilibrium

When considering more than one genetic locus, alleles tend to be co-transmitted to the same gamete when they are located physically close on the same chromosome. Genetic linkage between loci is generally a consequence of their being located on the same chromosome. When the segregation of alleles is followed through the generations in a pedigree, the allelic states tend to be co-inherited as the loci are “linked” by both occurring on the same chromosome, forming a *haplotype* (see Chapters 4 and 5).

The term *linkage disequilibrium* (LD) is used in a slightly different context. Rather than relating to the probability of an exchange of information at meiosis, LD is observed at the population level. LD is a general term which exists when allelic association is seen between two loci. Sometimes, this is referred to as non-random association of alleles. LD can arise through several mechanisms: by chance in small populations, by new mutation and or selection, or by intermixture of previously isolated populations (5).

Such association arises when alleles at distinct loci are found together in gametic phase (the alleles originate from the same gamete) at frequencies different to those expected based on the allele frequencies alone. The existence of LD does not necessarily imply that the loci are “linked”, i.e. are in close proximity on a chromosome, however, when two loci are in close physical proximity, LD implies that the population frequency of the two locus haplotypes are not as expected based on the allele frequencies. For example, consider two genetic loci with alleles labelled A and B at locus 1 and C and D at locus 2. At the population level, these alleles occur at the following frequencies A: 30%, B: 70%; C: 40%, D: 60%. While these loci may be “linked” and hence the probability of recombination between them at meiosis may be less than 0.5, this “linkage” is not seen at the population level. After many generations, you would expect the alleles to be randomly associated with haplotypes occurring at frequencies dictated by the product of their population allele frequencies AC: 12%, AD: 18%; BC: 28%, BD: 42%. If the population haplotype frequencies differ from these expected numbers, the loci are said to be in LD. The extent of LD is quantified by the disequilibrium parameter D (6). LD is discussed further in Chapter 6.

3.3. Kinship, Gene Identity by Descent and Inbreeding

Mendel’s laws imply certain patterns of allele sharing between pairs of relatives. For example, consider the four alleles found in two siblings at one specific locus, we would expect the siblings to share alleles inherited from their shared or common ancestors, the probability that they would share 0, 1, or 2 alleles inherited in common is 0.25, 0.5, and 0.25, respectively. If we consider half siblings, they are equally likely to share exactly 0 or 1 allele through their common parent, but they cannot share both alleles. In this example, we are considering the probability that they share an allele inherited from a common ancestor, these alleles are said

Table 1
Kinship coefficients and IBD sharing probabilities
for relative pairs

Relative pair	ψ	(k_2, k_1, k_0)
Full siblings	1/4	(1/4, 1/2, 1/4)
Half siblings	1/8	(0, 1/2, 1/2)
Monozygous twins	1/2	(1, 0, 0)
Parent offspring	1/4	(0, 1, 0)
Cousins	1/16	(0, 1/4, 3/4)

to be *identical by descent* (IBD) (each allele is a descended copy from a common ancestor). Two alleles may be identical but have not been inherited from a recent common ancestor. In this case, the alleles are said to be *identical by state*. The relationship between a pair of individuals, labelled X_1 and X_2 , can be summarised by the *coefficient of kinship* $\psi(X_1, X_2)$ (7). This coefficient is defined as the probability that an allele randomly sampled from X_1 and an allele randomly sampled from X_2 at the same locus are identical by descent. The *coefficient of inbreeding* $\alpha(X)$ for a single individual X , is defined as the probability that the pair of alleles that constitute the genotype of individual X at an arbitrary locus are IBD. The inbreeding coefficient for individual X is equal to the kinship coefficient for the parents of X . The *gene identity states* comprise the possible IBD sharing patterns for a pair of individuals. In the absence of inbreeding, pairs can share 2, 1, or 0 alleles IBD as argued above. The expected IBD sharing probabilities for each of these states are reported as a vector $\mathbf{k} = (k_2, k_1, k_0)$. Pairs of individuals with the same kinship coefficient do not necessarily have the same \mathbf{k} vectors. Table 1 lists some kinship coefficients and IBD sharing probabilities. It is interesting to note, though obvious from Mendelian segregation, that although parents and offspring have the same expected kinship as full siblings, the parent offspring pairs always share exactly one allele IBD.

4. Quantitative Genetics

The terms phenotype and trait are often used interchangeably, however, trait is commonly used in the quantitative context, and phenotype in the qualitative context. A state of health such as diagnosis of diabetes (phenotype is “affected with diabetes”) is often the result of consideration of a single quantitative trait, such as blood glucose levels; if the level of the trait is above a specified

threshold, the individual is classed as affected. Most clinical conditions fall into the discrete or qualitative phenotype, though the diagnosis may reflect the presence of an extreme value for the underlying quantitative trait, such as the relationship between body mass index (BMI) and obesity.

The terminology introduced so far has focussed on the influence of genes on discrete or qualitative phenotypes. In the late nineteenth century, Francis Galton first used the term “regression” when describing the correlation he observed between traits measured in parents and offspring such as height (8). Galton’s work laid the foundations for later researchers who made inferences about genetic models or trait inheritance by applying statistical methods to observations on pairs of relatives.

4.1. Components of Variance Models

While quantitative and Mendelian genetics use the same principles regarding the inheritance of genes, in the former the penetrance function (the relationship between genotype and phenotype) links a discrete with a continuous variable. A normally distributed quantitative trait can be summarily described by its mean and variance. Quantitative genetics models assume that genetic variation contributes to phenotype variation. Hence, the quantitative phenotype observed in an individual, *the phenotypic value*, can be thought of as made up of several components, one of which may be due to genes and another due to environment. This allows us to decompose the trait value seen in an individual into a linear expression.

$$\mathcal{Y}_{g,c} = \mu + G_g + E_c$$

where $\mathcal{Y}_{g,c}$ represents the phenotype value observed in a person with genotype g and environment c . The genetic (G_g) and environmental (E_c) contributions are generally represented as deviations from a population mean μ (3). By breaking down the phenotype into these components and using Mendelian segregation to derive the expected IBD sharing between pairs of relatives, we can model the “correlation” between pairs of relatives. The model can also be used to predict trait values but is more usually used to assess the evidence for a genetic component. In the simplest form, we might assume that the genetic contribution to the trait is due to a single locus with two alleles. A locus with two alleles, A and A' , has three associated genotypes AA , AA' , and $A'A'$. Each of these genotypes makes a specific contribution to the trait value. However, the contribution is rescaled so that the origin is at the value mid way between the two homozygote (AA and $A'A'$) values. If the alleles act in a simple *additive* fashion, the heterozygote value is exactly the mid-point between the two homozygote values. If there is an *interaction* between alleles at the same locus (*dominance*), then the value associated with the heterozygote, d , will deviate from this

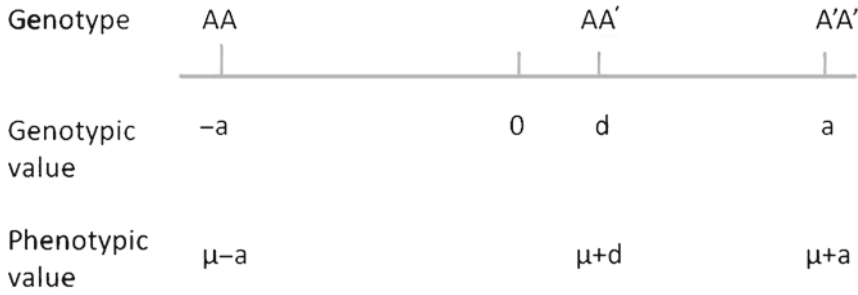


Fig. 1. Illustrating the relationship between the genotypes, the genotypic values and the phenotypic values. In this linear model, the impact of the genotype on the quantitative trait is described in terms of three parameters, μ , a , and d .

mid-point. According to our new scale illustrated on Fig. 1, the homozygote value ranges from $-a$ to $+a$. If $d=0$, we say there is no dominance, the alleles are codominant, or act additively. If $d=-a$, then A' is recessive to A , if $d=a$, then A' is dominant to A . If d is greater than $+a$ or less than $-a$, then we have *overdominance*. The degree of dominance is sometimes reported as d/a . This model can be extended to allow for multilocus genotypes, where each locus contributes additive and dominance effects.

We can extend this notation to multiple loci, say we have three loci with alleles A and A' , B and B' , and C and C' adding a suffix to indicate the source of the genotype effects. The genotypic value associated with the *compound genotype* AA' , BB , and $C'C'$ would be $d_A - a_B + a_C$. This assumes no interaction between alleles at different loci. Interaction between genotypes at distinct loci is termed *epistasis*. When studying the correlation between pairs of relatives in a pedigree, it is important to remember that it is the allele that is transmitted from one generation to the next and not the genotype values directly. For this reason, “breeding values” are sometimes used when referring to the genotypic values of parents (3). The breeding value is the additive genotypic value, as the dominance effect arises only in the individual who receives the interacting alleles; it is not transmitted directly (though covariance due to dominance effects can be seen in some relative pairs).

This model describing the relationship between phenotype and genetic factors gives rise to a *variance components* framework. The total (or population) trait variance is made up of variance components attributable to the genetic component and the environmental component.

$$\sigma_p^2 = \sigma_A^2 + \sigma_D^2 + \sigma_I^2 + \sigma_c^2 + \sigma_e^2$$

The terms above for each component of population variance (σ_p^2) are defined as the *additive variance* (σ_A^2), *dominance variance*

(σ_D^2), *common environment variance* (σ_c^2) and random non attributed variance (σ_e^2). *Epistatic variance* (σ_I^2) or the variance attributable to interaction between loci, is included in the expression above for completeness but is very difficult to characterise in practice. The expression above assumes that there is no interaction between the environment and genotype. The *genetic variance* is the sum of all the genetic components ($\sigma_A^2 + \sigma_D^2 + \sigma_I^2$). While we have stated the model in terms of these components, these components cannot be identified by sampling from a population, unless relative pairs are studied. The expected sharing of alleles between pairs of relatives enables inferences to be made on the components of genetic variance. We can write down the expected covariance for pairs of relatives (see Table 2).

While Mendelian segregation dictates how the alleles are shared among relatives, the degree of shared environment is more open to discussion. In Table 2, you can see that only full sibs are assumed to share a common sibling environment. This type of shared environment is commonly assumed, but other models can be proposed depending upon the characteristic of interest (9).

We have presented a framework where a trait value is made up of contributions from many sources. The genetic component may arise from additive effects of alleles, an interaction between alleles at the same locus and interaction between genotypes at different loci. Similarly, the influence of the environment can be dissected in more detail. Particularly, if we want to know how much of the correlation in relatives is due to shared environment.

Table 2
Expected co-variances between relative pairs

Relative pair	Expected covariance
Full siblings	$\frac{1}{2}\sigma_A^2 + \frac{1}{4}\sigma_D^2 + \sigma_c^2$
Half siblings	$\frac{1}{4}\sigma_A^2$
Monozygous twins	$\sigma_A^2 + \sigma_D^2 + \sigma_c^2$
Parent offspring	$\frac{1}{2}\sigma_A^2$
Cousins	$\frac{1}{8}\sigma_A^2$

4.2. Heritability

The magnitude of the genetic contribution is frequently summarised as *heritability*. Heritability is defined as the proportion of the trait variance that is attributable to genetic variation. It is therefore the ratio of the genetic variance compared to the total variance. *Heritability in the broad sense* (H^2) includes additive, dominant, and epistatic effects. Heritability in the narrow sense (h^2) restricts attention to additive effects only. Given these definitions heritability must always lie between 0 and 1. Values close to 1 suggest a strong genetic component with most trait variation due to genetic variation. Conversely, values close to zero suggest that genetic variation only weakly contributes to trait variation. Caution needs to be used when comparing heritability estimates from different populations as the heritability is defined relative to the population phenotypic variation and hence is population specific.

4.3. Twin Studies

In human genetics, twin studies are commonly used to establish and identify the strength of a genetic component. This study design uses the variance component framework and is frequently used to estimate heritability. Monozygous (MZ) twins are genetically identical, whereas dizygous (DZ) twins can be thought of as age-matched siblings. So the classic twin study design (contrasting covariances between MZ and DZ twins) offers a means to estimate the components of genetic variance. If the MZ and DZ correlation are similar, then this would be evidence that any genetic component is weak. However, if the MZ correlation is greater than the DZ correlation, this is evidence for a genetic component. As can be seen from the table the co-variance is a function of three parameters, but there are only two equations to link the observations and the model. Hence, only two of the three components of variance can be estimated and investigators can only report if the evidence for a shared environmental component is stronger than the evidence for a dominance component (10). If the basic twin design can be extended to include observations on other relatives, such as additional siblings or parents, more specific components of variance can be modelled and potentially estimated (see Chapter 11).

One extension to the twin study is to compare co-variances between twins reared together and those reared apart (*adoption studies*). This design allows the estimation of both the dominance and shared environment component. A common criticism of the twin study design is the validity of the key assumption that the shared environment is equivalent for MZ and DZ twins. This may not be valid for the analysis of behavioural traits as MZ twins, and DZ twins can have socially very different experiences (10).

4.4. Major Genes and Polygenes

When a single gene has a strong influence on a trait, i.e. a large a , then this gene is called a *major gene* and the allele-specific

effects of the gene can be identified, through the model outlined above. If, however, many genes are involved, it becomes difficult to isolate the allele-specific effects, and it is more common to then assume that the several (unlinked) genes involved all have a small but equivalent effect. As we allow more loci to contribute to the variation, the individual allelic effect must reduce. If we then further assume that all the alleles at these unlinked loci have equivalent and only additive effects, then the distribution of the compound genotypic values will approach the normal distribution. This leads to the *polygenic model*, the joint effect of an infinitely large number of loci results in *polygenic values* distributed about a mean of zero and variance σ_{PG}^2 .

The *mixed model* (11) allows for both a major gene and a polygenic effect, assuming no interaction between these two components. Hence, the variation in a phenotype can be attributable to a major gene effect, a polygenic effect and environmental effects. It is important to note here that the source of the environmental sharing is not directly measured but is often assumed due to familial factors.

5. Familial Aggregation, Segregation Analysis, and Qualitative Traits

The framework described in Section 4 relates to the variation in quantitative phenotypes, which lend themselves naturally to a variance component model. However, the same approach can be used to make inferences about binary traits with one extension to the framework. Instead of assuming that the model predicts phenotype, we allow the model to predict an underlying latent variable *liability*. The link between the model and our binary phenotype is established by defining a *threshold*. If an individual's liability value exceeds a threshold, the individual becomes affected with the disease. This extension enables the calculation or estimation of risk of disease or penetrance function. In some variable age at onset models a log-normal distribution of risk is assumed rather than the liability threshold (12).

Approaches to identify the genetic component for binary phenotypes frequently take a different form than for quantitative traits. If a major gene is suspected, the genotype-specific penetrance estimates will be reported along with an estimate of disease allele frequency. These models can be fitted without the need for a variance component framework. Often families have been selected due to the presence of at least one relative with the disease or phenotype of interest. This individual is

designated as the *proband*. Proband-based sampling is common when a disease is rare, and it is expensive to study and record information on families who have no cases of disease occurring. The manner in which such families are identified and the members of the family studied is called the *ascertainment scheme*. This biased sampling scheme, resulting in an oversampling of affected individuals, needs to be taken account of in any subsequent analysis so that statistical inferences are not biased. Further constraints can be applied in segregation analysis to ensure that the model predicts incidence or prevalence rates consistent with population data.

Simple Mendelian traits or simple genetic models imply that one genotype determines one phenotype, such as the dominant and recessive examples above for Mendel's peas. A deviation from this simple one to one correspondence is termed "complex". A disorder is called a *single gene disorder* when it only arises when mutations occur in a specific gene. However, if the probability of being affected with the disease conditional on the risk genotype is less than 1, then the term *incomplete penetrance* is used. Cystic Fibrosis (CF) is an example of a single recessive gene disorder with variable severity of phenotype and showing extensive *allelic heterogeneity*. Over 1,000 distinct mutations (alleles) in the CFTR gene have been described, and the clinical phenotype varies from severe when detected soon after birth, to mild and clinically undetectable until well into adulthood. The term *locus heterogeneity* is used when several genes can each independently give rise to the same phenotype. In qualitative phenotype analysis, the term *sporadic case* or *phenocopy* is used to indicate an affected individual whose phenotype has arisen due to an environmental cause and not the genetic predisposition. When the model permits phenocopies and incomplete penetrance, both phenotypes (e.g. affected and unaffected) are possible for all genotypes, hence all penetrance probabilities are greater than 0.

Rather like the study of correlations between pairs of relatives, observed familial aggregation of binary phenotypes is often reported as the *familial relative risk* or *familial recurrence ratios* (FRR) (13). These are simply defined as the risk of disease in relatives of a case compared to the risk in the general population. The FRR can be reported for all relatives within kinship groups, such as first degree relatives, or by the specific form of the relationship, for example, sibling. Though a genetic model gives rise to predictable patterns of FRR, the FRR merely summarises the pattern of risk and does not necessarily imply a genetic cause to a correlation in risk. The FRR are often referred to as the "lambda" risks (Greek letter λ), with a subscript indicating which relative of the case is considered. Commonly considered relative types are the sibling (λ_s), parent (λ_p), and offspring (λ_o).

6. Prospects for Phenotype Studies

The use of segregation analysis and the variance components approaches rely only on measuring the phenotypes in relatives. Large extended pedigrees or observations on many different types of relative pair enables the exploration of more complex models than those outlined above. However, when only phenotype data is available, these models lack the power to distinguish between common genes with low penetrances and the polygenic components. Advances in both molecular genetics and statistical computing are now making it feasible to identify and characterise locus-specific effects, by incorporating measured genotypes into the analysis. It is the identification and characterisation of the environmental components that present the next major challenge to the field.

References

1. Gingeras, T.R. (2007) Origin of phenotypes: genes and transcripts. *Genome Res.* 17, pp 682–690.
2. Mendel, G. (2008) Experiments in Plant Hybridisation, in *Ending the Mendel-Fisher Controversy*. (Franklin, A., Edwards, A.W.F., Fairbanks, D.J., Hartl, D.L., Seidenfeld, T. eds.), University of Pittsburgh Press, Pittsburgh, PA. pp 79–116.
3. Falconer, D.S., Mackay, T.F.C. (eds.) (1996) *Introduction to Quantitative Genetics*, Fourth Edition. Longman, Essex, UK.
4. Conrad, D.F., Pritchard, J.K. (2007) Population genetics and disease, in *Genes and Common Diseases*. (Wright, A., Hastie, N. eds.), Cambridge University Press, Cambridge. pp 44–58.
5. Cordell, H.J., Clayton, D.G. (2005) Genetic association studies. *Lancet* 366, pp 1121–1131.
6. Weir, B.S. (ed.) (1996) *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
7. Cannings, C., Thompson, E.A. (eds.) (1981) *Genealogical and Genetic Structure*. Cambridge University Press, Cambridge.
8. Galton, F. (1886) Regression towards mediocrity in hereditary stature. *J Anthr Inst.* 15, pp 246–63.
9. Lynch, M., Walsh, B. (eds.) (1998) *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Sunderland, MA.
10. Sham, P. (ed.) (1998) *Statistics in Human Genetics*. Wiley, New York.
11. Morton, N.E., MacLean, C.J. (1974) Analysis of family resemblance. III. Complex segregation analysis of quantitative traits. *Am J Hum Genet.* 26, pp 489–503.
12. Pharoah, P.D.P., Antoniou, A.C., Easton, D.F., Ponder, B.A.J. (2008) Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med.* 358, pp 2796–2803.
13. Risch, N. (1990) Linkage strategies for genetically complex traits. I. Multilocus models. *Am J Hum Genet.* 46, pp 222–228.



<http://www.springer.com/978-1-60327-415-9>

Genetic Epidemiology

Teare, M.D. (Ed.)

2011, XII, 242 p., Hardcover

ISBN: 978-1-60327-415-9

A product of Humana Press