

Preface

The explosive growth in the field of molecular biology over the last two decades has started to make a great impact on clinical medicine. Genes have been cloned for diseases that were poorly understood only a decade ago. Additionally, investigators are increasingly aware that there are strong genetic components to complex disorders, such as osteoporosis, that are not classically thought of as genetic disorders. New insights into the pathogenesis of metabolic bone diseases have been obtained from investigations into the molecular biology of these diseases and new therapies will become available based on these new insights.

In *The Genetics of Osteoporosis and Metabolic Bone Disease*, I have assembled an internationally renowned group of experts to write the various chapters. Each of the authors is an expert in his/her field who is currently performing research on the content of their chapter and have made important contributions to the understanding of the clinical features and pathophysiology of metabolic bone disease and genetics.

The first part of *The Genetics of Osteoporosis and Metabolic Bone Disease* addresses issues related to genetic contributions to the development of osteoporosis and the many factors that must be considered when searching for genes that predispose to osteoporosis. The second section addresses recent advances in the clinical and molecular biological aspects of inherited metabolic bone disorders. The last section reviews the latest techniques for finding genes that predispose to metabolic bone diseases.

Michael J. Econs, MD

CHAPTER 2

How to Determine If, and by How Much, Genetic Variation Influences Osteoporosis

John L. Hopper

Genetic Variation: Within and Between Populations

Discussion of nature vs nurture, or “genes vs environment,” is often obscured by a failure to understand that what is being considered is the *variation* in genetic make-up of individuals, and how it relates to differences between them in the characteristic or trait of interest. Therefore, a clear distinction needs to be made between genetic differences *within* a population, and genetic differences *between* populations (e.g., between different races, or between blacks and whites). For example, genetic factors may explain much of the difference in a characteristic between two racial groups, but within any such group, the variation may be entirely due to nongenetic factors. Consequently, discussion about the roles of genes and environment in explaining variation in a trait must depend on, first, whether one is considering within or between population comparisons.

Historically, discussion has been in terms of the *relative* roles of genetic and environmental factors in explaining trait variation across a given population, as reflected by the often cited but little understood concept of heritability (see below). However, in order to understand and quantify properly the impact of genetic factors, the critical concept is not the proportion but the *absolute* size of genetic variation. This is because the amounts of variation caused by genetic and nongenetic factors are not universal constants; rather, they can vary both within and across populations according to age, sex, lifestyle, and a multitude of factors that cannot be controlled for, at least not in studies of humans. It is easier to understand that there can be different environmental factors, but genetic factors also can be expressed at different stages of life, or only when the individual is subject to particular environmental challenges. The effect of genes may also depend on exposures or lifestyle, i.e., there may be gene-environment interactions or covariations. Consequently, the size of the genetic variance may differ for different populations and for subgroups within a population.

Therefore, if the ratio of genetic to total variance (heritability) differs between populations or subgroups, one cannot conclude that it is due to differences in the genetic component alone. There could be differences in the nongenetic component, or in both the genetic and the nongenetic components. That is, a trait does not have a unique “heritability”; its heritability may depend on the ethnicity and environmental milieu of the population, and also may differ according to the characteristics used to describe the average of the population under consideration (see the next two sections).

To Study the Causes of Variation, the Mean Must First Be Specified

One cannot define variation, let alone study its causes, without specifying the mean, or “expected value,” of the trait in question. The mathematical definition of the variance of a trait is the sum, over all possible values the trait can take, of the squared deviation of each value from its mean, weighted by the probability of that value. That is, for a trait Y that can take at most a finite number of values, by definition:

$$\text{variance}(Y) = S (y_i - m)^2 p(y_i) \quad (1)$$

where y_i represents an observed value of Y , $p(y_i)$ represents its probability, summation is over all possible values of i , and $m = \sum y_i p(y_i)$ is the mean. (For a continuously distributed trait, the summation is replaced by integration, and the probabilities by a probability density function.)

The term $(y_i - m)$ is called the residual of Y about its mean, for individual i . The residuals, not the trait values, are the focus of attention in analyses that aim to determine the role of genetic and environmental factors in explaining the causes of variation in a trait.

Suppose a trait varies, on average, according to factors such as age, sex, lifestyle, environmental exposures, body characteristics, and so forth. The variance of the residuals will then differ according to which of these factors are taken into account in specifying the trait mean, m .

In practice, the mean of a trait can be modeled in terms of the effects of measured factors. These are called “fixed effects.” The remaining variation in residuals can then be modeled in terms of the effects of unmeasured factors (both genetic and environmental). These are called “random” effects. This statistical modeling process is a well-established application of linear regression and analysis of variance.

The random effects modeling of unmeasured factors must do more than take into account genetic factors as causes of similarities between blood relatives. It must also allow for environmental factors shared by relatives. The strength of these factors could depend on whether the relatives are currently living together—and if so, for how long and how intensely—and if not, how long since they cohabited, and how often they are in contact with one another.

If genetic factors cause variation in a trait, then genetically related individuals will be more similar. Just as the variance of trait residuals depends on which factors are used to model the trait mean, the covariance between two traits also depends on their trait means. For example, for discretely distributed traits Y_1 and Y_2 , by definition:

$$\text{Covariance}(Y_1, Y_2) = \sum \sum (y_1 - m_1)(y_2 - m_2)p(y_1, y_2) \quad (2)$$

(Y_1 and Y_2 could be two realizations of the same trait, such as when they represent the trait values of the first- and second-born twins of a pair.) Here y_1 and y_2 represent the observed values of Y_1 and Y_2 , respectively, $p(y_1, y_2)$ represents the probability of the joint occurrence, $m_1 = \sum y_1 p(y_1)$ is the mean of Y_1 and $m_2 = \sum y_2 p(y_2)$ is the mean of Y_2 , and the double summation is over all values of 1 and 2. The correlation is a measure of how similar traits are. It can take any value from -1 to 1 , and is defined as

$$\text{Correlation}(Y_1, Y_2) = \text{covariance}(Y_1, Y_2) / [\text{variance}(Y_1) \text{variance}(Y_2)]^{1/2} \quad (3)$$

As discussed in a later section, the process of modeling unmeasured genes as random effects works by matching, for pairs of individuals, the similarity of their trait residuals against their genetic relatedness. The amount of variance in the residuals that can be explained by the covariance between relatives appearing to fit the pattern expected under a genetic model is called the genetic variance. If genetic factors are measured, and are modeled as fixed effects, the total variance will be reduced. This should result in a reduction in the genetic variance, provided the model of genetic and environmental causes of variation across the population, and of covariation within families, is a close approximation to reality.

An Example

When considering height, what factors determine variations in height from individual to individual; i.e., what are the “causes of variation”? The age and sex of an individual are of primary importance. Other critical factors might be nutrition—especially during early childhood—developmental diseases, and variables related to socioeconomic status and lifestyle, and these effects could be confounded with one another. Some of these factors, such as age and sex, and possibly even the occurrence of any developmental diseases, can be measured. Some can be assessed by a surrogate measure; e.g., socioeconomic status is often inferred from measures of income, occupation, and/or residential location. Others, such as childhood nutrition, can be very difficult to determine, yet may be quite similar in genetically related individuals such as siblings, especially twins, and even more especially, within monozygotic (MZ) twin pairs.

There are also genetic factors to consider. There may be (if not now, perhaps sometime in the near future) known genes for which different variants are predictors of height, in both the statistical and biological meaning of the word. There are

likely to be, however, a large number of genetic loci involved in the expression of height. How are we to find out if such loci exist, given that they are as yet unidentified and therefore not measured?

First, we have to take into account the nongenetic, or environmental, risk factors. Those that are measured can be incorporated as fixed effects when modeling the mean. The residual height of an individual, after adjusting for age, sex, and any other relevant measured factors of that individual, becomes the focus of attention. It is the correlation or covariation between related individuals in these residuals that forms the basis of analyses that aim to determine the role of genetic and environmental factors in explaining the causes of variation in height. Note that if height was not adjusted for age, the residual variance would be larger, and relatives of the same or similar age (such as twins and siblings) would appear to be more strongly correlated. This could have a substantial effect on the outcome of genetic modeling.

Defining Osteoporosis: Fractures vs Risk Factors

Distinction needs to be made between a fracture (i.e., an event which is represented by a binary trait and can take only two possible values), and a risk factor for that event. The latter could include variables such as bone mineral density (BMD), bone geometry, propensity to fall, and so forth; *see* Chapter 1. These risk factors may also be binary variables (e.g., sex), but are more often scalar variables distributed along a continuum. Moreover, distinction must be made between fractures at different sites, and there is the possibility that different risk factors may be operating at different sites.

There are numerous risk factors for fractures, and these may be interrelated (i.e., correlated within a population). Depending on what factors are taken into account in specifying the mean of these risk factors, some genetic factors may cause variation, and some of the genes involved in causing variation in one risk factor may also cause variation in other risk factors.

The extent to which genes causing variation in a risk factor for fracture explain the incidence of fractures in the population will depend on: (1) The strength of association between the risk factor and fracture risk, and (2) The proportion of the population at different levels of genetic risk (allele frequencies). Therefore, although variation in a risk factor may be strongly genetically determined, it may have little consequence for the disease in question in terms of explaining cases, and why it runs in families. This latter issue will be quantified in a later section.

Making Inference about Possible Genetic Causes of Variation: Biometric Modeling of Twin and Family Data

In trying to infer a role for genes in causing variation, distinction must be made between whether genes are measured or not measured. As discussed earlier,

the effect of measured genetic variation can be assessed by modeling the mean, while the effect of unmeasured genes can be assessed by modeling variation about the mean.

For the latter case, inference is made by developing models based on assumptions about the action of genes, and then testing the extent to which the data are compatible with the different models. This process has been referred to as biometric modeling. Note that biometric modeling cannot prove that genetic factors are causing variation, it can only be used to demonstrate that the data are consistent with one or more genetic causes of variation. It has very limited ability to discern how many genetic loci are involved (1). Furthermore, the utility of this process depends critically on the extent to which the design, sample size, and methods of analysis allow the effects of nongenetic causes of variation to be discriminated from those of genetic causes.

Historically, biometric modeling has been focused on fitting more and more elaborate genetic models, trying to interpret variations from simple genetic descriptions in terms of sophisticated modes of action of the genes (e.g., sex-limited expression, or nonadditive effects such as dominance and epistasis). This process is unconvincing, however, to the skeptic.

It is important to realize that familial aggregation does not necessarily imply genetic causation. In theory, familial aggregation in a trait can always be fully explained by an environmental model, simply making the postulated effect of sharing the environment match the observed correlations. For example, monozygotic (MZ) twin pairs might be more similar for a trait than dizygotic (DZ) twin pairs simply because:

1. They live or lived more similar lifestyles, especially during their formative years;
2. They meet with each other more often; and/or
3. They are being or were treated more alike when children.

Similar arguments might explain why first-degree relatives are more similar for a trait than second-degree relatives, and so on. The extent to which this issue is important varies from trait to trait. For example, a large difference in sibling correlation for blood lead levels was observed between adolescent pairs living together ($r = 0.5$) and adult pairs living apart for 20 or more years ($r = 0.1$) (2). On the other hand, for body mass index ($BMI = \text{weight}/[\text{height}]^2$) we observed a small decrease in the correlation, about 2,000 DZ pairs, from 0.6 for those who were cohabiting to 0.5 for those living apart (unpublished data). Moreover, that decrease occurred over an age range of less than 5 yr, suggesting that the effects of shared environment on BMI variation dissipate rapidly. For bone density, a similar rapid dissipation of the effects of shared adolescent environment appears to occur (3).

Unfortunately, little attention has been placed on trying to disprove or falsify genetic hypotheses. On the other hand, classic biometric models make simplistic

assumptions about the roles of nongenetic factors. Accordingly, only major and specific types of environmental effects can be detected by the statistical approaches typically used, and even then large samples are needed for there to be reasonable statistical power to detect such effects. Designs that allow for contrasts between the effects of shared genes and those of shared environments, measures of environmental exposure, and large data sets ascertained by unbiased sampling are needed if the modeling process is to have credibility in teasing apart the roles and genes from those of the environment. In practice, this is not easy to do and tends to have been the exception, rather than the rule (4).

Genes Measured

Suppose a genetic marker, such as a polymorphism at a candidate gene hypothesized on biological grounds to be involved in the trait of interest, can be measured, and takes the values m_i , for $i=1, \dots, n$, say. To test that hypothesis, a simple test would be to select individuals of genotype m_j and compare their trait values against those of genotype m_k , for all pairs of j and k not equal to each other. If these individuals are unrelated, careful consideration needs to be given to how the subjects were sampled, and to the possibility that there could be other factors (such as race and ethnicity) associated with genotype at this marker, and with the trait itself. These are called “potential confounders.” If all such potential confounders are known, and measured on all individuals, a statistical adjustment can be made for their effect(s) on the mean of the trait of interest using, e.g., linear regression techniques. In practice, one never knows all the confounders, and it is impossible to adjust for unmeasured or unknown confounders. In particular, care should be taken if the study sample contains different racial or ethnic groups. It is well known that these genetic association studies can give misleading conclusions due to such confounding, or population-stratification as it is referred to in the genetics literature.

Another approach is to select related individuals who differ in genotype. Dizygotic twin pairs of the same sex are perhaps the most useful design. Twins within a same-sex pair are perfectly matched for age and sex, typically among the most important determinants of the mean of a human characteristic. They are also matched, at least to some extent, for a range of nongenetic factors related to their shared environment during gestation and upbringing. Whereas the latter matching may reduce within-pair trait differences, with a consequent loss of statistical power to detect effects associated with the genotyping, it will not bias results if matching is taken into account in the analysis.

An appropriate method of statistical analysis would be to consider within-pair trait differences as a function of within-pair differences in genotype. If there are n different genotypes possible, there will be $n(n-1)/2$ combinations of differences between genotypes, and it is likely there may be some combinations with few or no pairs. Analyses that suggest differences between some combinations, but not others, are difficult to interpret.

Because each genotype is a combination of two alleles, one from the mother and the other from the father, it is useful to consider the within-pair difference in the number of shared alleles: it can take the values 0, 1, or 2. The first group of pairs, who are concordant for genotype, is uninformative for association studies, and can be excluded from the analysis. The within-pair trait differences of the second and third groups can each be compared with zero, using e.g., a paired Student's t-test, or with each other. Again, the analyses may be difficult to interpret if, e.g., pairs differing by one allele are different, but those differing by two are no different, or differ in the opposite direction.

It might make biological sense to presume that the within-pair trait difference increases linearly as the difference in the number of a certain disease allele (or subgroup of alleles) increases. The trait difference can then be modeled as a linear function of the difference in number of disease alleles, in which case it is presumed that pairs differing in both alleles are twice as different (and in the same direction!) as those differing by only one allele. Again, linear regression techniques can be used for the modeling, making sure that the line of best fit is constrained to pass through the origin. Care must be taken, however, if there are only a few pairs differing by two disease alleles, as they will have a strong influence on the fitted line. Robust regression methods designed to be insensitive to influential points should also be used. Remember that for these *association* studies, the dependent variable is the actual within-pair trait difference, not the absolute or squared difference. There is an implied order within members of a pair, which could be based on an exposure or covariate of interest (5), or may be arbitrary. The independent variable is the ordered difference in the number of disease alleles, and can take the values -2 , -1 , 0 , 1 or 2 . The fitted line should be constrained to pass through the origin; see Fig. 1. Note that pairs with the same number of disease alleles have no influence on this fitted line. They will, however, contribute information on the variation about the fitted line, and hence may improve statistical inference.

If multiple regression techniques are used, statistical adjustment for within-pair differences in other measured factors likely to explain trait variation — either genetic or environmental — can be made while concurrently estimating the effect, if any, of the measured genetic markers on the mean; i.e., estimating a genetic association (6). The formula is:

$$\Delta Y = Y_1 - Y_2 = (m_1 - m_2) + (x_{11} - x_{12}) + \dots = \Delta m + \Delta x_1 + \Delta x_2 + \dots \quad (4)$$

where the x_{ij} are the observed values for trait i in twin j , $\Delta x_i = x_{i1} - x_{i2}$, and Δx_1 could be the difference in number of disease alleles at the genetic locus of interest.

Another way the effect of a measured genetic marker on a trait can be assessed is in terms of possible genetic linkage. This can be done by testing if related individuals who share 2 alleles are more similar for a trait (in the sense defined below) than those who share 1 allele, and if the latter are more similar than those who share 0 alleles. This is called identity-by-state (6). The parents each

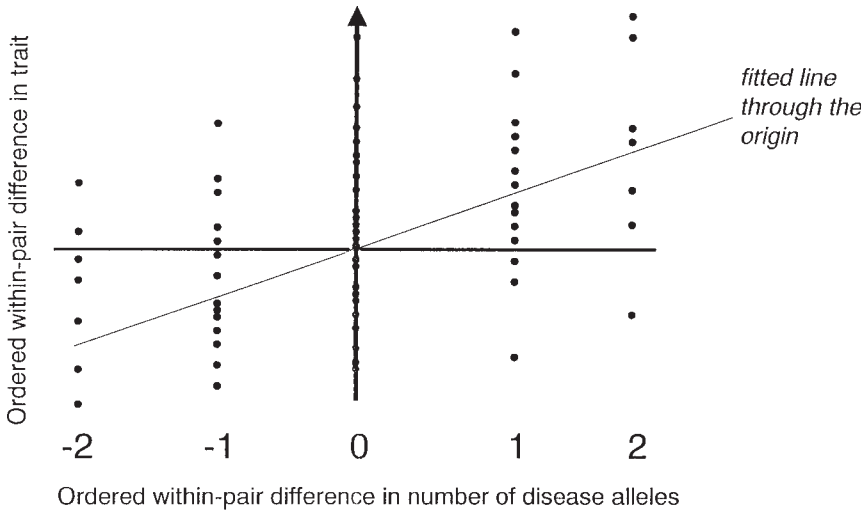


Fig. 1. Best-fitting straight line through the origin, for data from ordered twin pairs plotting their within-pair difference in the trait vs their within-pair difference in number of disease alleles.

have two alleles, although they may not be distinct, and each offspring inherits a copy of one allele from the father and one from the mother. Therefore two offspring can share exactly the same two alleles, just one of the alleles, or neither. This is referred to as identity-by-descent. To conduct an identity-by-descent analysis, parental genotypes must be known or inferred probabilistically. Analysis of whether similarity in such measures of identity at a locus predicts similarity in a trait can be incorporated using the method of analysis described in the next section (7). For further discussion on aspects of genetic linkage *see* Chapter 20.

Genes Not Measured

In 1918, the Royal Society in Edinburgh published a seminal paper in the history of genetics and statistics (8), written by the then 28-year-old Ronald Aylmer Fisher. (The paper had been rejected by the Royal Society in London 2 yr earlier.) Fisher derived the expected pattern of correlations one would observe in a trait if its variance (about its mean) was due to genetic variation at one or more loci, under the assumption that genetic status was transmitted from generation to generation under the rules of Mendelian inheritance. The paper itself is long and not easy, but the basic ideas have been summarized and interpreted many times (9–13).

In brief, suppose the effect of alleles at each loci are additive and independent of one another, in that the trait mean for an individual who has two copies of an allele differs from that for an individual who has no copies of that allele by

twice as much as for an individual who has one copy of the allele. Fisher showed that the correlation will be 0.5 between first-degree relatives, 0.25 between second-degree relatives, 0.125 between third-degree relatives, and so on. MZ twin pairs would be identical under this model which assumes additive genetic factors, A, only. There are, however, numerous prenatal factors that could cause MZ pairs to be less than identical for a given trait (1).

Allowing then for nongenetic factors, E, which are assumed to be independent between relatives in the same family—such as random environmental effects and random measurement error—the correlation between MZ twins within a pair will be reduced depending on the relative amount of variation explained by E. That is, if the variance of A is σ_a^2 and that of E is σ_e^2 , the MZ correlation will be $\sigma_a^2/(\sigma_a^2+\sigma_e^2) = \rho_{MZ}$. In this case, the correlation between first-, second-, and third-degree relatives will become $1/2 \rho_{MZ}$, $1/2 \rho_{MZ}$, and $1/8 \rho_{MZ}$, and so on.

The model can be extended to allow for nongenetic factors shared by, or common to, members of the same family (7,14). These cohabitation effects, C, that result from sharing or having shared a “common” environment, will result in a perturbation of the pattern of ratios above that should occur when genetic factors only are causing trait correlations. For example, if first-degree but not second-degree relatives share a common environment effect with variance σ_c^2 , the ratio of first- to second-degree trait correlations will be greater than 2:1.

Fisher allowed for nonadditive effects at a genetic locus (i.e., for the effect of having two alleles to differ from twice the effect of having one allele), and in doing so introduced the concept of dominance variance, σ_d^2 . If there are such nonadditive effects within one or more loci, the MZ pair and sibling correlations will be differentially increased relative to correlations between other pairs of relatives. Fisher also derived the expected correlations under other forms of genetic nonadditivity, such as interactions between the effects of alleles at different loci. There is strong confounding, however, between additive and nonadditive components of variance, and dominance components can mimic common environment effects (7).

Whereas there is a theoretical basis on which genetic models can be based, cohabitation or common environment effects can take virtually any structure. Some examples of plausible models have been derived and fitted (2,7,13). In practice, however, it is very difficult to clearly distinguish the effects of shared genes from those of shared environment. Designs which have the potential to overcome this problem include those using twins reared apart, adoptees and their biological and adopted families, and migrant and nonmigrant families.

The classic twin model purports to allow estimation of the effects of shared genes and shared environment, but does so only by making the very strong assumption that the effects on the trait in question of sharing environment are the same for MZ pairs as they are for DZ pairs. This assumption is difficult to test using twin data alone, unless detailed information is collected on the amount and extent of cohabitation that has or is occurring within the pairs (15). Also, the modeling process is biased toward concluding that genetic factors alone are

causing twin correlations (4). Large sample sizes are needed to detect even modest proportions of variance attributable to σ_c^2 , the variance attributable to shared, or common environments during or as a result of cohabitation (16). Nevertheless, application to adolescent and young adult female twins has suggested plausible cohabitation effects on bone mineral density, at the spine and the hip, that abate rapidly as the twins begin to live apart (3).

Fisher's theory is best expressed in terms of variance components. For example, suppose a model allows for additive genetic effects, and for a zygosity-independent effect of common environment as in the classic twin model. For a given trait mean, the (residual) variance can be expressed as the sum of variance components $\sigma_a^2 + \sigma_c^2 + \sigma_e^2 = \sigma^2$. The covariance between a pair of relatives will be: $\sigma_a^2 + \sigma_c^2$ for MZ pairs; $1/2 \sigma_a^2 + \sigma_c^2$ for DZ pairs; $1/4 \sigma_a^2$ for parent-offspring pairs, and for non-twin sibling pairs; $1/4 \sigma_a^2$ for second-degree pairs, etc.

Note that the effects of measured genes can be incorporated either in modeling the mean, as fixed effects, or as random effects by including additional component(s) of variance. For example, the covariance between a pair of individuals could be modeled as σ_M^2 if they share two alleles (IBD), or $1/2 \sigma_M^2$ if they share one allele (IBD) (7). That is, detection of "quantitative trait loci" by variance components modeling can be carried out within the usual framework for fitting genetic and environmental models.

If the variance, σ^2 , is the same for all individuals, the difference between the MZ and DZ pair correlations, ρ_{MZ} and ρ_{DZ} , respectively, will be $1/2 \sigma_a^2 / \sigma^2$. This has led historically to the simplistic formula: $2(\rho_{MZ} - \rho_{DZ}) = \sigma_a^2 / \sigma^2$. The latter term is referred to as the narrow-sense heritability, the proportion of variance attributed to additive genetic factors. This formula is problematic, if only because values greater than unity can occur (i.e., it could lead to the implausible conclusion that more than 100% of the variance is attributable to heritable factors).

The statistical fitting of genetic and environmental models has changed considerably over the last few decades due principally to improvements in computational resources. The underpinning of the statistical approaches, however, is in likelihood theory derived by the young Fisher way back in 1912. They almost always rely on the assumption that families have been sampled at random. This is, of course, not usually true, but small sampling biases may not have a major effect on results. Sampling through related traits, however, can have a major impact. Methods for adjusting for some forms of nonrandom ascertainment exist.

Either the raw data, or summary measures in terms of correlations between relative-pairs, are fitted to variance components or path analytic models. Statistical packages, such as FISHER (17) and M_x (18), are available for analysis of continuous traits, and the latter handles categorical traits under some stringent assumptions; see the next section. Newer robust statistical methods have also been developed and compared (19). Standard errors, confidence intervals, correlations between estimates based on large-sample theory are available. Some

tests of lack of fit which generally have little power, yet are often mistakenly used to infer a “good fit”, are also available with these packages.

It must not be forgotten, however, that no matter how sophisticated the modeling, how good the fit or how elegant the method of statistical analysis, a fitted model is just that. Especially when trying to make inference about unmeasured genes and unmeasured environmental factors, there may be many quite dissimilar models that provide equally good—or bad—descriptions of the data, and it is not possible to discriminate between them.

A significant genetic component of variance in the most parsimonious model does not prove that genes exist (20). For example, although under the classic twin model MZ pairs will be more correlated than DZ pairs, finding that MZ pairs are more correlated does not prove that the classic twin model is correct! In the parlance of mathematical logic, that “statement A implies statement B” is true does not mean that “statement B implies statement A” is also true. If genetic factors exist (statement A), MZ pairs will be more correlated than DZ pairs (statement B). The observation of statement B, however, does not definitively prove the veracity of statement A.

Evidence for Genetic Variation Influencing Fractures

A fracture is an event that either happens or doesn't. Inference about the possible roles of genetic and environmental factors in explaining variation in binary traits (i.e., traits that divide individuals into two groups, such as affected versus unaffected) can be made by a similar approach to that outlined previously.

Association in a binary trait can be measured and modeled a number of ways, such as by a correlation coefficient, odds ratio, or tetrachoric correlation. The latter was first used in this context by Karl Pearson, around the start of the 20th century. It supposes that there is an underlying, normally distributed, but unmeasured “latent” trait. In medical applications, it is referred to as a “liability.” For a given 2×2 table of association in a pair of individuals for the binary trait, a unique bivariate normal distribution of the latent liability trait can be derived, with appropriate cut-off points — one horizontal and one vertical — so that the proportion of the distribution in each of the four regions equals the observed numbers in the 2×2 table. The tetrachoric correlation is then defined as the correlation in the bivariate normal distribution that is needed to achieve this fit.

For whatever measure of association is used, its strength can be compared across different categories of relatives, and between cohabiting and noncohabiting pairs. The same issues discussed previously, to do with confounding and the ability of the design and sample size to discriminate between different models, apply. Typically, very large samples in the order of hundreds if not thousands of pairs are needed to have reasonable power.

The tetrachoric correlation modeling has become popular in some disciplines, perhaps because it allows calculation of heritability. There are many problems,

however, with this approach. It presumes there is an underlying, but unmeasurable liability, that both predicts the binary trait status and is correlated in relatives. Inference is made about the proportion of variance in liability statistically attributed to genetic or nongenetic factors, along the same arguments as used for continuous traits in 5(ii) above. However, the strength of the quantitative estimates rely implicitly on the assumption that the liability distribution in families is multivariate normal, and it is often impossible to test this assumption, at least not with any substantial statistical power.

Furthermore, the heritability of liability is a difficult concept to understand intuitively. It is usually, and mistakenly, referred to as “heritability of the trait.” When the trait is a disease state (affected vs unaffected), the slip is often made of interpreting heritability as the proportion of disease due to genetic factors, when in fact it is the proportion of the variance of liability that is due to genetic factors (21). Unfortunately, judging by the frequency with which this false implication is made even in professional circles, this point is not well-appreciated.

Comparison of twin disease concordance between MZ pairs and DZ pairs is often used to infer a genetic aetiology for the disease in question, but again this inference is predicated by the strong shared environment assumption of the classic twin method.

There has been considerable confusion about the concepts of pairwise, casewise, and probandwise concordance. Pairwise concordance is the probability that a given twin is affected, given that at least one member of the pair is affected. The pairwise estimator is the number of pairs in which both twins are affected, divided by this number plus the number of discordant pairs. Casewise concordance is the probability that one twin is affected, given that the other is also. The casewise estimator is twice the number of pairs in which both are affected, divided by the twice the number of pairs in which both are affected plus the number of pairs in which twins are disease discordant. The decision of which concordance to work with depends on the question(s) being asked. For example, if as is the case in genetic counseling when one wants to predict a twin’s disease status from knowledge of the other pair’s status, the casewise concordance is indicated. On the other hand, if one is interested in predicting the pair disease status when all one knows is that at least one twin is affected, the pairwise concordance is indicated.

If twin pairs are sampled nonrandomly, e.g., because at least one of the pairs is known to be affected, proper adjustment must be made (22). In this case, it is possible that a pair could be sampled because it is known that both members are affected; this is called a doubly-ascertained pair. It is also possible that a pair will be sampled because one particular member is known to be affected, but once the pair is sampled it is found out that the other twin is also affected; this is called a singly-ascertained pair. “Incomplete ascertainment” is said to occur if the probability of singly ascertained pairs is greater than zero. Otherwise, ascertainment is said to be “complete,” although this expression can be confusing, because it does not necessarily mean that all pairs are sampled!

Although reference is often made in the twin literature to the “probandwise concordance rate,” “probandwise concordance” actually refers to an estimator,

not a concordance, and it certainly is not a rate. The probandwise estimator is defined as the ratio of a numerator to a denominator. The numerator is twice the number of pairs in which both twins are affected and doubly ascertained, plus the number of singly-ascertained pairs both affected. The denominator is equal to the numerator plus the number of disease discordant pairs.

The estimates and large sample standard errors for pairwise and casewise concordance have been derived, and the distinctions above between concordances and estimators clarified, using a likelihood theory approach (22–23). For large samples and under random or complete ascertainment, the casewise estimator is unbiased for the casewise concordance, and the pairwise estimator is unbiased for the pairwise concordance. Under incomplete ascertainment, the casewise estimator is biased for the casewise concordance, and the pairwise estimator is biased for the pairwise concordance. The probandwise estimator, however, is unbiased for the casewise concordance.

Interpreting Genetic Variation in a Risk Factor in Terms of How Much Familial Aggregation in the Disease Is Explained

Finally the question arises: if genetic factors explain a proportion of variation in a risk factor, how much of the familial aggregation for the disease is explained by this source of genetic variation? To address this, suppose that X is a risk factor measured on a continuous scale. Suppose the conditional probability of being affected, $p(x)$, given a value of the risk factor, x , can be represented by the linear logistic model

$$p(x) = \Pr(D = 1 \mid X = x) = \exp\{\alpha + \beta x\} / [1 + \exp\{\alpha + \beta x\}] \quad (5)$$

where $D = 1$ signifies the disease is present, and the parameters α and β describe the dependence of the probability of disease on the observed value of X . Assume also, without loss of generality, that X has a standard normal distribution with mean 0 and variance 1, and that for a pair of related individuals, (X_1, X_2) has a bivariate normal distribution with correlation parameter ρ .

On a population-basis, the correlated risks within pairs of relatives translates into clustering of disease. A measure of disease association between relatives is the odds ratio, OR, the ratio of the odds of being affected when a relative is affected to that when the relative is unaffected. The value of OR depends on the correlation, ρ , and the strength of the risk factor on probability of disease. The latter is conveniently represented by the inter-quartile risk ratio, RR, which is the risk for individuals at the upper-quartile level of X divided by the risk for individuals at the lower-quartile level. Hopper and Carlin (*see ref. 24*) give tables for these relationships.

Bone density at the hip is a risk factor for hip fractures that is itself correlated between first degree relatives. The correlation, ρ , is about 0.4, while the inter-

quartile risk ratio, RR, is about 2.5. From Table 1 of Hopper and Carlin (24) this translates into an odds ratio, OR, of about 1.1. Given that the increased risk for a daughter consequent upon her mother having had hip fracture is roughly twofold, it is seen that whatever causes bone density at the hip to be correlated in first-degree relatives explains about 10% of familial aggregation for hip fractures. Simple twin models suggest that all of the familial aggregation for hip bone density in adult women is attributable to genetic factors.

Therefore, even if the heritability of hip bone density is 80%, the genes causing variation in hip bone density are not responsible for most of familial aggregation in hip fractures. Similar arguments apply to lumbar spine bone density and spinal fractures. Furthermore, genes involved in causing variation in other risk factors, such as hip axis length, may explain just as much familial aggregation in hip fractures (25). Nevertheless, identifying genes that influence variation in bone density could have important implications for prevention and treatment, for example by providing molecular targets for altering bone density. Finding these genes may provide insight into molecular pathways that are important in regulating osteoclast or osteoblast activity, for example, and this knowledge could be used to manipulate the pathway so as to increase osteoblast activity without increasing bone resorption.

Summary

“Familial aggregation” is the tendency for a trait to be more similar, or positively correlated, in family members. This applies both to the occurrence of disease in an individual—often expressed as a binary trait representing the two states, being or not being affected—and to indices of morbidity and risk factors which are measured on continuums, often referred to as continuous traits. Depending on the strength of association between risk factors and disease, and on the strength of familial aggregation in the risk factors, this can result in familial aggregation in the disease itself. Therefore, familial aggregation in risk factors could in part explain why family history is a risk factor for osteoporosis. Knowing how genetic and environmental factors explain familial associations in risk factors will help understand the causes of osteoporosis.

A theory under which correlations between relatives in a continuous trait can be explained in terms of Mendelian inheritance at one or more genetic loci was published by R. A. Fisher in 1918. Application of this theory has since provided much insight into the design and analysis of studies to resolve the relative contributions of, and interactions between, genetic and environmental factors, and has identified the following statistical and design issues:

1. Prior to genetic and environmental modeling the data should be carefully explored, and relationships between trait mean and covariates examined. To be able to understand properly the genetic determinants of a trait, it is important to first account for the effects of major nongenetic determinants.

2. Descriptive measures of familial aggregation should be explored, and tendencies for the associations between individuals to vary according to the genetic or cohabitational relationship between individuals should be noted. Empirical evidence has shown that genetic and environmental factors can produce a variety of patterns among the trait correlations between pairs of individuals.
3. As mentioned previously, theory shows that if genetic factors determine variations in a continuous trait, certain patterns will be evident among the correlations between relatives. It is therefore possible, in a rigorous statistical manner, to test if observations from independent groups of related individuals are consistent with a proposed genetic model. Note, however, that this does not prove that genetic factors are a cause of variation, let alone the only cause of covariation within a family.
4. For there to be statistical power to discriminate between causes of familial aggregation, first the confounding between genetic and environmental factors needs to be addressed in the design of a study. Second, the sample sizes must be large enough for discrimination and precise estimation of different effects. Third, the families need to be sampled in an unbiased manner, or else a correct statistical adjustment for their ascertainment must be made.
5. Analytical methods should be able to incorporate measurements from covariates, which may include measured genetic markers, in addition to modeling unmeasured genetic and environmental sources of variation.
6. Fitting a model is not an end in itself. Biological and statistical assumptions underlying models should be addressed before and during the modeling process. Selection of a “best” model from among a range of alternatives, even if shown not to provide a bad fit to the data, does not prove that the components of that model are true causes of variation.
7. The genes involved in causing variation in bone density only partially explain the familial aggregation in fractures.

References

1. Martin, N., Boomsma, D., and Machin, G. (1997) A twin-pronged attack on complex traits. *Nat. Genet.* **17**, 387–392.
2. Hopper, J. L. and Mathews, J. D. (1983) Extensions to multivariate normal models for pedigree analysis. II. Modeling the effect of shared environment in the analysis of blood lead levels. *Am. J. Epidemiol.* **117**, 344–355.
3. Hopper, J. L., Green, R. M., Nowson, C. A., Young, D., Sherwin, A. J., Kaymakci, B., Larkins, R. G., and Wark, J. D. (1998) Genetic, common environment, and individual specific components of variance for bone mineral density in 10- to 26-year-old females, a twin study. *Am. J. Epidemiol.* **147**, 17–29.
4. Hopper, J. L. (1999) On why “common environment effects” are so uncommon in the literature, in *Advances in Twin Sub-Pair Analysis* (Spector, T., Sneider, H., and MacGregor, A. J., eds.), Greenwich Medical Media, London, pp. 15–165.
5. Hopper, J. L. and Seeman, E. (1994) The bone density of twins discordant for tobacco use. *N. Engl. J. Med.* **330**, 37–92.

6. Lange, K. (1997) *Mathematical and Statistical Methods for Genetic Analysis*. Springer, New York.
7. Hopper, J. L. and Mathews, J. D. (1982) Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* **46**, 373–383.
8. Fisher, R. A. (1918) Correlations between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. (Edinburgh)* **52**, 399–433.
9. Bulmer, M. G. (1980) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, UK.
10. Falconer, D. S. (1989) *Introduction to Quantitative Genetics*. Wiley, New York.
11. Ewens, W. J. (1979) *Mathematical Population Genetics*. Springer-Verlag, Berlin, pp. 5–9.
12. Neale, M. C. and Cardon, L. R. (1992) *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
13. Hopper, J. L. (1993) Variance components for statistical genetics, applications in medical research to characteristics related to human diseases and health. *Statist. Methods Med. Res.* **2**, 199–223.
14. Hopper, J. L. and Mathews, J. D. (1994) A multivariate normal model for pedigree and longitudinal data and the software FISHER. *Aust. J. Stat.* **36**, 153–176.
15. Clifford, C. A., Hopper, J. L., Fulker, D. W., and Murray, R. M. (1984) A genetic and environmental analysis of a twin family study of alcohol use, anxiety and depression. *Genet. Epidemiol.* **1**, 63–79.
16. Christian, J. C., Norton, J. A., Sorbel, J., and Williams, C. J. (1995) Comparison of analysis of variance and maximum likelihood based path analysis of twin data, partitioning genetic and environmental sources of covariance. *Genet. Epidemiol.* **12**, 27–35.
17. Lange, K., Boehnke, M., and Weeks, D. (1987) Programs for pedigree analysis. Department of Biomathematics, University of California, Los Angeles, CA.
18. Neale, M. C. (1997) *M: Statistical Modeling*. 4th ed. Box 980126 MCV, Richmond, VA.
19. Huggins, R. M., Loesch, D. Z., and Hoang, N. H. (1998) A comparison of methods of fitting models to twin data. *Aust. N. Z. J. Stat.* **40**, 129–140.
20. Hopper, J. L. (1996) Commentary. Genes for osteoarthritis, interpreting twin studies. *Br. Med. J.* **312**, 943–944.
21. Hopper, J. L. (1998) Heritability, in *Encyclopedia of Biostatistics*, vol. 3, Wiley, London, pp. 1905–1906.
22. Hopper, J. L. (1998) Twin concordance, in *Encyclopedia of Biostatistics*, vol. 6, Wiley, London, pp. 4626–4629.
23. Witte, J., Carlin, J. B., and Hopper, J. L. (1999) A likelihood-based approach to estimating twin concordance for dichotomous traits. *Genet. Epidemiol.* **16**, 290–304.
24. Hopper, J. L. and Carlin, J. C. (1994) Familial aggregation of a disease consequent upon correlation between relatives in a risk factor measured on a continuous scale. *Am. J. Epidemiol.* **136**, 1138–47.
25. Flicker, L., Faulkner, K. G., Hopper, J. L., et al. (1996) Determinants of hip axis length in women aged 10–89 years, a twin study. *Bone* **18**, 41–45.