STUDY DESIGNS

The heritability of human disease: estimation, uses and abuses

Albert Tenesa[1,2] and Chris S. Haley[1,2]

Abstract | Relatives provide the basic material for the study of inheritance of human disease. However, the methodologies for the estimation of heritability and the interpretation of the results have been controversial. The debate arises from the plethora of methods used, the validity of the methodological assumptions and the inconsistent and sometimes erroneous genetic interpretations made. We will discuss how to estimate disease heritability, how to interpret it, how biases in heritability estimates arise and how heritability relates to other measures of familial disease aggregation.

Tetrachoric correlation
An estimate of the correlation between two bivariate normal variables obtained from the 2 × 2 contingency table containing the counts of two categorical variables (for example, disease and non-disease in two types of relative).

Fundamental to the study of the inheritance of human disease is the partition of the total phenotypic variation in a population into genetic and environmental components. The ratio of the genetic component to the total phenotypic variance is the proportion of genetically controlled variation and is termed the heritability[1]. Estimated heritabilities for human diseases are specific to the diseases, populations and particular circumstances from which they were derived, but they are often in the mid-range between the theoretical extremes of zero and one. The heritability helps to set bounds to the potential discriminative ability of disease classifiers based on genetic markers[2] and can be used to estimate familial recurrence risk of disease[3].

Human diseases are highly heterogeneous. Many diseases are dichotomous and are measured on a binary or 0/1 scale (that is, the disease is either absent (0) or present (1)). Some diseases, however, manifest variable levels of disease severity (for example, structuring and inflammatory ulcerative colitis), allowing classification into a larger number of phenotypic classes. Diseases can also be defined by imposing a threshold on a continuous trait to produce a discontinuity (for example, osteoporosis or hypertension). We focus here on diseases that are dichotomous or that are treated as such.

Unlike continuous traits[1], the phenotypic variance of a dichotomous disease depends on the trait mean (that is, the disease prevalence), and methods used for heritability estimation in quantitative traits do not directly apply. To overcome this problem, heritability is often estimated for the normally distributed liability that is postulated to underlie disease[4]. This is called the heritability of liability to disease (h_x^2) and should be distinguished from the heritability of disease in the observed scale ($h_{0/1}^2$)[5,6]. There are four main approaches to estimating the heritability of disease. The first approach is a general method based on the correlation of disease status in pairs of relatives of a particular type from a random sample of the population (the so-called tetrachoric correlation). The second method, the twin method, compares the resemblance among monozygotic (MZ) twins and dizygotic (DZ) twins[7,8]. The third approach, Falconer's method[9], measures the correlation in liability between relatives from the incidence of the disease in the general population and among relatives of affected individuals[9]. The fourth approach, which is based on generalized mixed linear models[10], is the most flexible. It uses Bayesian methods or maximum likelihood methods to estimate genetic (and other) variance components in families or pedigrees[11].

The various methods available, each with their own potential sources of bias, combined with differences in sampling strategies (for instance, random population data versus data from families of diseased individuals), small sample sizes or different genetic interpretation can lead to heterogeneous estimates of heritability that sometimes add more confusion than understanding. This is not a new issue as was highlighted in 1972 by Smith *et al.*[12], who estimated the heritability of diabetes using first-, second- and third-degree relatives and obtained estimates ranging from 0.19 to 1.02. In their words: "The range of heritability estimates from the different kinds of relatives raises doubts about both the data collected and about the methods of analysis used"[12].

This Review focuses on the validity of heritability in human disease genetics. How is it possible that despite its wide use (for example, ~16,000 titles in Web of Science), it remains such a controversial[13,14] and

[1]*Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK.*
[2]*MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK.*
Correspondence to A.T.
e-mail:
Albert.Tenesa@ed.ac.uk
doi:10.1038/nrg3377

Box 1 | **Basic concepts in heritability: causal components of resemblance**

The simplest genetic model of a complex trait can be outlined as

$P = G + E_T$

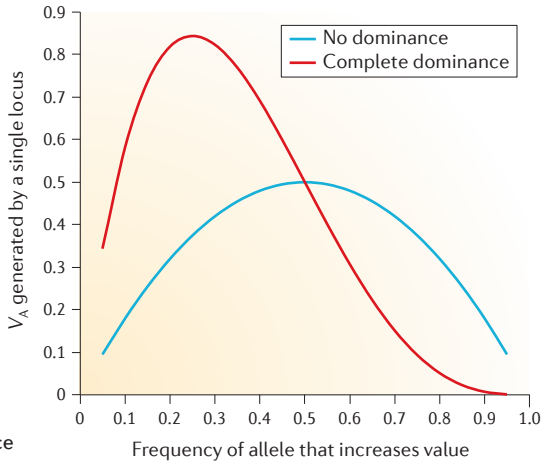
where P represents the phenotype, G is the genotype, and E_T is the combination of environmental factors that influence the expression of the phenotype. G and E_T are the causal components and are assumed to be independent. G can be partitioned further as follows:

$G = A + D + I$

where A is known as the additive genetic value, D is the dominance component, and I is the epistatic (interaction) component of the individual's genotype. Similarly, E_T can partitioned as

$E_T = C + E$

where C is the common environment shared by family members, and E is the remainder of the environmental variance that is not linked to whether the individuals are related or not. Common environment can be defined in different ways, depending on the background information available to the researcher. That is, although the unit of grouping is often assumed to be the family or the mother, the researcher could choose other groupings (such as village and school).



Thus, the total variance (V_P) can be partitioned into its contributing factors

$V_P = V_A + V_D + V_I + V_C + V_E$.

Often, human quantitative geneticists refer to ADE or ACE models when reporting heritability estimates. These would be models that partition V_P as $V_A + V_D + V_E$ and $V_A + V_C + V_E$, respectively.

The distinction between G and A gives two meanings to a genetically determined trait. The first meaning is that it is determined by the person's genotype (G), and the second is that it is transmitted from parents to children (A).

For most traits, the chief cause of resemblance among relatives is the additive genetic variance (V_A): that is, the variance of additive genetic values. The presence of additive variance, however, does not imply that the gene actions will be purely additive nor is it necessary to assume additive gene action to estimate the additive genetic variance. Indeed, additive genetic variance can arise from genes with varied levels of dominance and epistasis (see the figure, in which the difference among homozygous genotypes is fixed). Additive gene action is difficult to ascertain; only when complications from the environment can be ignored and the total genetic and additive variances are equal can it safely be assumed that the genes involved show no dominance or epistasis.

Mixed linear models
Statistical models used to analyse grouped data. Grouped data, such as repeated measurements, generate within-group correlations that need to be accounted for to make correct inferences. In the context of heritability estimation, they separate fixed effects (for example, gender or age) from random effects (individuals).

Bayesian methods
Bayesian methods of inference combine prior beliefs about a hypothesis with the information provided by the available data to modify those prior beliefs. The stronger the prior beliefs are, the more data will be required to modify them. Bayesian methods could help inference that is based on small sample sizes, where maximum likelihood methods may fail.

Maximum likelihood methods
Methods or techniques used for statistical inference. These methods are used for deriving functions of the sample (technically called estimators) that when applied to particular samples give estimates of the population parameters. The maximum likelihood estimates of the unknown parameters are the most likely parameters to have generated the observed data.

Bias
A population parameter (for example, a variance) is estimated from a random population sample using an estimator (for example, a formula). An estimator is unbiased if the mean of the estimates it produces over many samples, regardless of their size, is the population parameter.

Additive genetic values
Also called breeding values, these are defined as the sum of the average effects of the alleles an individual carries and, in the context of disease, as the average disease risk a person will confer to their children. Both definitions are equivalent only when there is no interaction between loci.

disputed parameter in human genetics? We discuss what fuels some irreconcilable views, how disease heritability is estimated, what produces bias and how it relates to other measures of familial disease aggregation. The 'missing heritability'[15] question, including the potential contribution of epigenetic inheritance[16], is extensively discussed elsewhere[17,18] and is only briefly mentioned here. We argue that despite the controversies that surround it, the heritability of disease remains, after more than 60 years[5], the most useful summary statistic for the genetic contribution to familial aggregation of complex diseases. Identifying the sources of familial resemblance and the practical exploitation of that knowledge should remain a focus of future research.

Defining heritability

For most human traits, relatives are more alike to each other than to random members of the population: that is, there is correlation between the phenotypes of relatives. This well-known phenomenon, the resemblance among relatives, underpins the study of the inheritance of complex traits. The degree of resemblance varies by trait, depending on the combined effects of common environmental factors (that is, those shared by family members) and inherited factors. Quantifying and partitioning the resemblance among relatives provides a foundation for disentangling the relative contribution of nature and nurture to the observed variation. The partition of human phenotypic variation is complicated because inference about the contributing factors, known as the causal components (BOX 1), is made through the observational components (that is, those that can be directly estimated from the data), and the observational components often include different proportions of the causal components that are difficult to separate[19].

The proportion of the phenotypic variance attributable to genetic differences is termed the heritability. However, more precisely, it is necessary to distinguish between two statistics: the narrow-sense heritability (h^2), which refers to differences among the additive genetic values, and the broad-sense heritability (H^2), which refers to genetic differences as differences between genotypic values. H^2 represents the amount of variation attributable to genetics if genetically identical individuals, such as clones, could be exposed to multiple environments (that is, the amount of variation explained by the clones over all environments). In human studies, the closest approximation is the study of MZ twins reared apart. The broad-sense heritability ($H^2 = V_G / V_P$, where V_G is the genotypic variance, and V_P is the phenotypic variance — see BOX 1

Table 1 | **Coefficients of correlations between relatives***

Type of relationship		r	s
MZ twins		1	1
First-degree	Full sibling	0.5	0.25
	Parent–child	0.5	0
Second-degree	Child–grandparent	0.25	0
Third-degree	Cousins	0.125	0

*$t = (rV_A + sV_D + V_C) / V_P$. Second-order terms or larger are omitted. *r*, the coefficient of relationship, is twice the probability that two gametes sampled at random from each parent will carry alleles identical by descent (that is, come from the same ancestor); *s*, the probability of two relatives having the same genotype identically by descent; *t*, the intra-class correlation; V_A, additive genetic variance; V_C, variance due to common environment; V_D, dominance variance; V_P, phenotypic variance.

for further explanations of the symbols) indicates the extent to which an individual's genotype determines its phenotype. This is the most useful concept for prediction of an individual's risk of disease from its own genotype: that is, for personalized medicine. The narrow-sense heritability ($h^2 = V_A / V_P$, where V_A is the additive genetic variance) reflects the degree to which the genes transmitted from the parents determine the phenotype of their children[20] and is most useful in predicting disease risk from parental family history. Obtaining 'pure' estimates of either h^2 or H^2 is limited by the possibility of separating the causal components given the study design (for example, relatives available, whether relatives share the same household and length of cohabitation). For instance, if only twins are available, it is not possible uniquely to estimate additive, dominance and common environmental variances as there are more variables to be estimated than there are sources of information.

Estimating heritability

Partitioning the phenotypic variance among groups of individuals forming defined family structures (for example, full siblings, twins or parents and children) allows the quantification of the level of phenotypic similarity among and within a particular family structure. The degree of resemblance is expressed as the intraclass correlation (*t*) for groupings of contemporary relatives, such as full siblings or cousins, and as the regression (*b*) of parents on offspring for parent–offspring pairs. Such a distinction is necessary because the family size affects the parent–offspring correlation but not the regression. Both the intraclass correlation and the parent–offspring regression provide estimates of the amount of variation that is common to the group as a proportion of the variation in the population[19].

Although the observational components of variance do not identify the causes of the among group variance, they provide, in combination with a 'causal' model (BOX 1), two important sources of information: first, a quantification of the importance of the different forces hypothesized in the model; and, second, a statistical framework for comparing competing models. It is likely that no model is perfect, but some models describe the data better than others. The heritability is estimated from the degree of resemblance among relatives as either *b* / *r* or *t* / *r*, where the coefficient of relationship (*r*) is the correlation that would be observed if all of the phenotypic variance observed were additive genetic. However, the estimates of *b* or *t* are composed of different proportions of the variance components, depending on the relationship between the pairs of relatives available for study (TABLE 1). This makes direct comparisons of studies using different types of relatives challenging. Generally, the larger *r* is, the more precise the estimate of heritability is, as the standard error of the *b* or *t* has to be divided by *r*. However, bias due to common environment and dominance (for example, in full siblings when estimating h^2) is often of more concern than precision is.

In the following sections, we present a succinct description and examples of the main methodology applied to estimate heritability. Detailed mathematical descriptions can be found elsewhere[11,20–22].

General method for population data. The methodology described above was originally developed for continuous traits and was extended to binary traits by assuming an underlying liability to disease (BOX 2). Usually, the liability to disease cannot be measured, and the correlation coefficient in liability between pairs of family members is estimated from the observed categorical data using tetrachoric correlations. The estimated correlations are then equated to their expectations (the 'causal' model) to estimate the degree of genetic and environmental variation. Assume that a binary trait for which population data (that is, a random sample of pairs of relatives) is available. Pairs of relatives have four possible disease status combinations. The disease status of the pairs can be expressed as (11), (10), (01) and (00), where the first and second values represent the disease status of relatives 1 and 2, respectively. The data for *N* pairs of relatives could be arranged in a 2 × 2 contingency, where N_{11} is the number of pairs in which both relatives have the disease, N_{00} is the number of pairs in which both relatives are healthy, and N_{10} and N_{01} are the numbers of relatives with discordant disease status. Then, *t* can be approximated by using the following formula[23]:

$$t = \frac{N_{11}N_{00} - N_{10}N_{01}}{N_{11}N_{00} + N_{10}N_{01}} \quad (1)$$

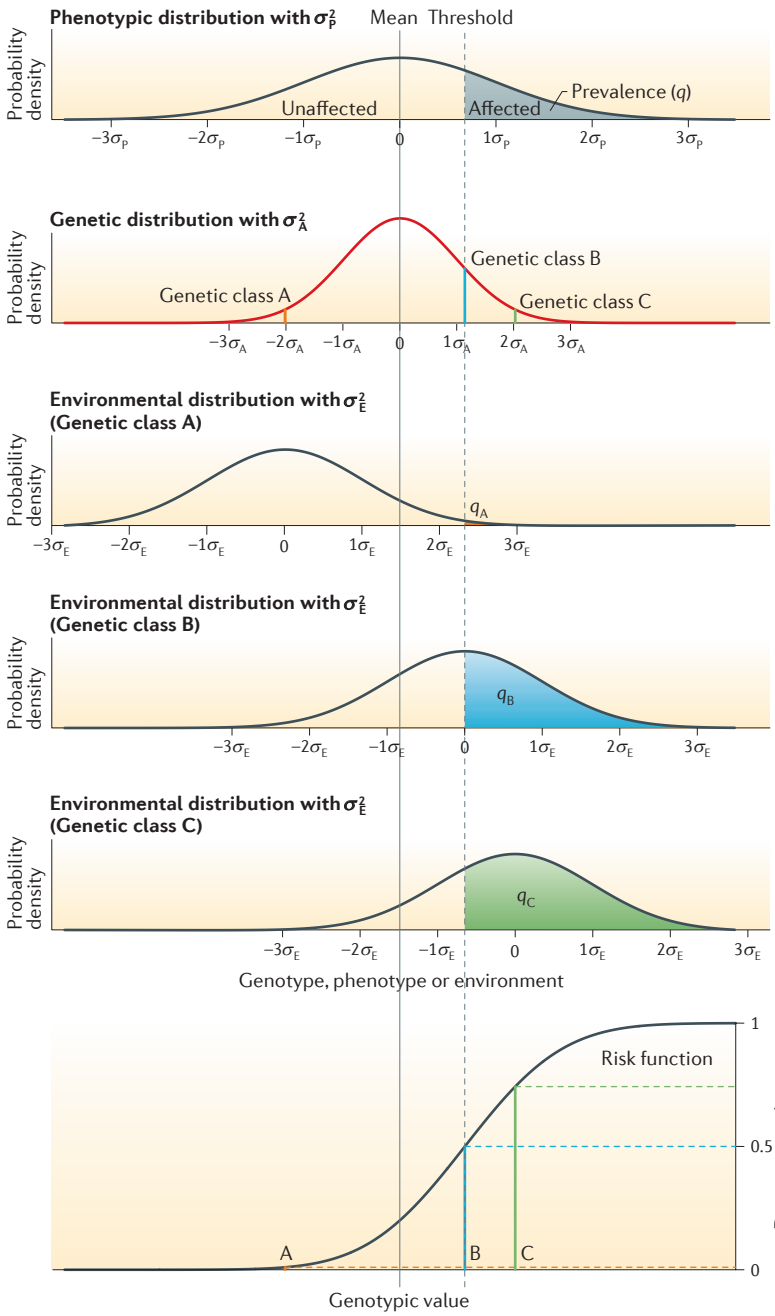
for which expectations for different types of relatives are shown in TABLE 1.

Imagine a hypothetical disease with a population prevalence of 8% and for which a random sample of 5,000 sibling pairs is available. All sibling pairs are screened and diagnosed with the same rigour. We assume, without a loss of generality, that $V_A = 2V_C = 0.16V_P$, where V_P, the liability variance, is 1, and V_C is the variance due to the common environment. As the liability model assumes that gene actions are additive on the liability scale, we assume the dominance variance $V_D = 0$ (REF. 24). Under this model, $h_x^2 = V_A / V_P = 0.16$. The sample counts are $N_{11} = 45$, $N_{01} = 340$, $N_{10} = 400$ and $N_{00} = 4{,}215$ (note that N_{01} and N_{10} are expected to be equal but will differ by chance).

Genotypic values
For a given genotype, these values are the expected phenotypes that arise from the combined expression of all of the genes contributing to the trait. In the context of disease, in the observed scale is the penetrance (that is, the probability of disease given the genotype).

Box 2 | **Underlying disease liability model and disease risk function**

The liability model assumes an underlying normal distribution of liability to disease. Individuals with a liability value above a threshold that is determined by the population prevalence of disease express the disease and those below do not. The liability is made up of additive genetic and environmental factors. In the example in the figure, the h^2 is 0.35 and the population prevalence (q) is 25%. The figure shows the equivalence of the model to a normally distributed genetic liability with a cumulative normal risk function. Three genetic classes (A–C) are represented by vertical lines on the genetic distribution. Their environmental distribution and associated risk ($q_{A–C}$) are shown below. The bottom panel of the figure represents the risk function for the disease. The risk is the probability of disease given the genotype (that is, the penetrance). The *y* axis of the top five figure panels is the probability density of the appropriate normal distribution and the *x* axis represents the phenotype, genotype or environmental values in standard deviations (σ) of the appropriate normal distribution. The bottom panel of the figure shows the disease prevalence on the *y* axis as a function of the genotypic value (or, equivalently, the liability value).



Equation 1 estimates $t = 0.16$, which would provide an estimate of $h_x^2 = t/r = 0.32$, and thus the heritability estimate is inflated by the common family variance.

This methodology allows fitting of a liability model to the data, thereby estimating the correlation and the two liability thresholds that define disease in the two types of relatives[22]. However, the method does not provide a test for whether the liability model is valid (that is, with a 2×2 contingency table, it is possible to estimate only three parameters).

Twin method. The general method outlined above ignores the contribution of dominance and common environmental factors and therefore inflates the estimates of h^2. A way of overcoming this is to sample pairs of MZ and DZ twins. By assuming that variation due to common environment and total variation are both equal in MZ and DZ twins, the inflation due to common environment can be removed.

Assume that twin pairs have been recruited, regardless of their disease status, into a twin registry (that is, a random sample of twin pairs from the population of twins has been screened for the disease) and that there are n_{11} diseased twin pairs, n_{10} discordant twin pairs and n_{00} healthy twin pairs. Then, t can be approximated by using the following equation[23]:

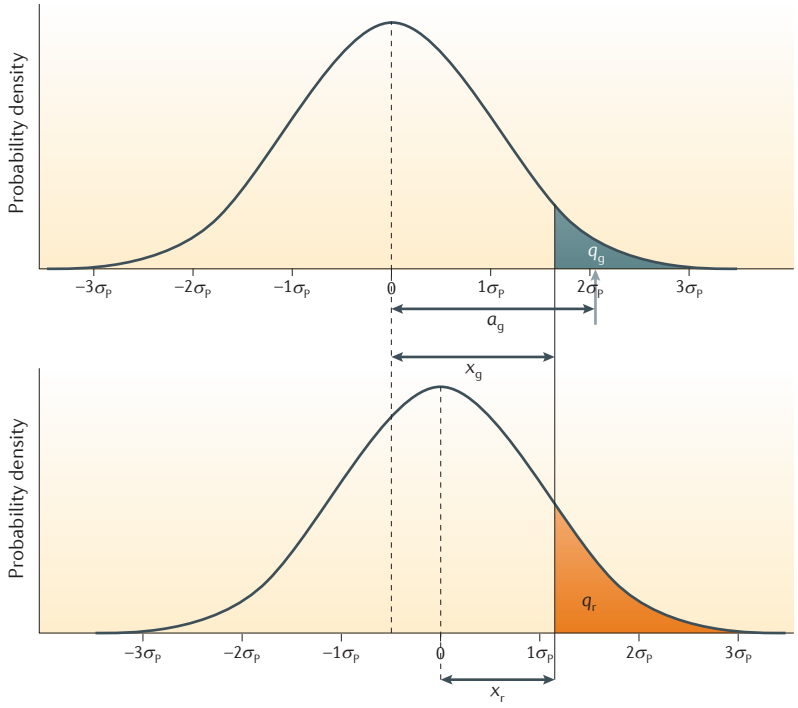
$$t = \frac{n_{11}n_{00} - \left(\frac{n_{10}}{2}\right)^2}{n_{11}n_{00} + \left(\frac{n_{10}}{2}\right)^2} \quad (2)$$

In the hypothetical disease in which $h_x^2 = 0.16$ (and in which $V_A = 2V_C = 0.16V_P$; $V_D = 0$) and the population prevalence is 8%, we could expect $n_{11} = 62$, $n_{10} = 791$ and $n_{00} = 4{,}147$ MZ pairs, and $n_{11} = 45$, $n_{10} = 740$ and $n_{00} = 4{,}215$ DZ pairs. Applying equation 2 and the coefficients of relationship for MZ and DZ twins (TABLE 1), we obtain $t_{MZ} = 0.24 = (V_A + V_D + V_C)/V_P$ and $t_{DZ} = 0.16 = (0.5V_A + 0.25V_D + V_C)/V_P$. Assuming that V_C is the same for MZ and DZ twins, h_x^2 can be estimated as $2(t_{MZ} - t_{DZ})$, giving $h_x^2 = 2(0.24 - 0.16) = 0.16$. An estimate of V_C could be obtained as $2t_{DZ} - t_{MZ} = 0.08$. These calculations assume that $V_D = 0$, which in our example is true; however, if this was not true, the estimates of h_x^2 would be inflated by 1.5 times the amount of V_D. Similarly, the estimates of V_C would be underestimated by $0.5V_D$.

Difficulties associated with this approach arise from the definition of twin concordance rates[25] when ascertainment is not complete and the assumptions that the total environmental variation and common environment variation are equal among MZ and DZ twins.

Falconer's method. Sometimes, patients are recruited to a study (that is, they are ascertained) when they are diagnosed in the clinic; thereafter, the researcher identifies the relatives of the index case and screens them for the disease. This strategy does not screen healthy relative pairs ($N_{00} = 0$), generating a censored sample that requires a different analytical approach. Falconer's[9] approach is based on selection theory. As with previous

Box 3 | **Falconer's method to estimate h^2 of liability to disease**



The V_P of liability is assumed to be one. As illustrated in the figure, the prevalence of the disease in the general population (q_g) determines the threshold of liability above which a person expresses the disease. Let Φ be the standard normal cumulative distribution function. Then, $q_g = 1 - \Phi^{-1}(x_g)$ and $x_g = \Phi^{-1}(1 - q_g)$, where Φ^{-1} is the inverse of the standard normal cumulative distribution, and x_g is the threshold of liability above which the disease is expressed. The mean liability of the index cases (a_g) is equal to z_g / q_g, where z_g is the height of the ordinate (that is, the density of the normal distribution) at the threshold (x_g). Similarly, q_r is the prevalence of disease among relatives of index cases and x_r is the normal deviate for q_r. The regression of a relative's liability on the liability of index cases is estimated as

$$b = \frac{x_g - x_r}{a_g} \quad (3)$$

The h_x^2 is estimated as b / r, where r is obtained from TABLE 1. If q_g is known with a high accuracy, then the standard error (SE) of b can be estimated as

$$SE(b) = \frac{1}{b a_g^2} \sqrt{\frac{1 - q_r}{N q_r}} \quad (4)$$

where N is the number of relative pairs.

methods, this method assumes that there is a normally distributed liability or that such a liability could be transformed to normality. Although the assumption is reasonable for a highly polyfactorial trait, this assumption is difficult to test because liability is usually not observed. Moreover, the method is sensitive to departures from normality, and even if a trait is polyfactorial — for example, birthweight — it is not always easy to find an appropriate transformation[26].

Falconer's method is described in BOX 3 and can be applied to groups of relatives of the same structure (for example, parent–offspring pairs or twin pairs). Now assume that the researcher of our hypothetical disease (introduced in the general method) has followed a clinic-based approach to recruit siblings through sibling 2 (the proband). Then, $N_{11} = 45$, $N_{01} = 340$, $N_{10} = 0$ and $N_{00} = 0$. Clearly, the general method cannot be applied to these data. However, we know from a population registry that the prevalence of the disease (q_g) is 8%. Following Falconer's approach, we have that $x_g = \Phi^{-1}(1 - q_g) = \Phi^{-1}(1 - 0.08) = 1.4$, where Φ is the standard normal cumulative distribution. We can estimate that $q_r = 45 / (45 + 340) = 0.12$ and that $x_r = 1.17$. As $a_g = 3.27$, $b = (1.4 - 1.17) / 3.27 = 0.07$ and $h_x^2 = b / r = 0.14$. Note that had we decided to call sibling 1 the proband, we would have obtained a slightly different estimate (that is, $h_x^2 = 0.07$), albeit well within their expected precisions (standard error = 0.18 and 0.38, respectively).

A similar sampling strategy for our twin example would deplete the study of healthy twin pairs. In this case, twin concordance could be computed in several ways[25] but the simplest and most appropriate measure is the proband concordance rate (q_c); q_c replaces q_r in the calculations (BOX 3). Applied to our hypothetical disease, we would obtain $h_x^2(DZ) = 0.18$, $h_x^2(MZ) = 0.33$ or $h_x^2 = 0.15$ if we apply the twin method.

The main sources of bias in Falconer's method are common familiar environmental factors and ascertainment. The consequences of the selection are that the distribution of liability among prevalent cases and among their relatives is skewed and that the variance of liability among relatives is reduced. Falconer's equations underestimate the correlation among relatives by ~10%, hence in terms of absolute values the bias becomes important only when the correlation is high. MZ twins, however, are an oddity and may produce estimates of heritability that are too high when the correlation is high[27,28]. Another source of error is the assumption of a linear regression of one relative on another on the liability scale. This is violated if a major gene with dominance underlies the disease (for example, as in families in which the breast cancer gene *BRCA1* is segregating[29]). Although Falconer's method is simple and provides a reasonably good approximation, other approaches (described in REFS 30,31) provide more accurate approximations and should be favoured where possible.

Mixed linear models. The methods described so far work only for regular pedigrees in which the data are structured into defined families of the same size. For instance, a study with a mix of parent–offspring trios and pairs would apply the methodology to the two types of data structures, obtain two estimates of heritability and pool those into some average. Mixed linear models, however, offer a unified framework to handle complex pedigrees of varying size and structure. They also allow testing of the distributional assumptions and the fit of the models. Complex pedigrees are advantageous because data collection is simplified, information use is optimized and partitioning of the liability variance into its 'causal' components is improved. The correlation structure within pedigrees can be formulated for binary and normally distributed data by using different link functions[11,32–34]. However, in practice, the analysis of a binary trait as continuous gives parameter

Ascertainment
The process of identifying cases of disease in the population. Ascertainment and sampling are often used synonymously, especially when talking of ascertainment or sampling bias.

estimates that are similar (after appropriate transformation) to those obtained using the binary formulation[35,36] and as convergence is generally improved, especially for sophisticated genetic models, the data are frequently analysed as if the trait were continuous. These heritability estimates, which are obtained on the 0/1 scale (that is, they estimate $h_{0/1}^2$), need to be transformed to the liability scale (h_x^2).

Mixed linear models can be fitted using maximum likelihood methods, restricted maximum likelihood (REML) methods and Bayesian approaches[37,38]. Bayesian implementations of the liability model are described elsewhere[31,32]. Structural equation models (SEMs) are a class of mixed linear model[39,40]. SEMs describe the covariance structure among relatives as random effects and the mean structure as fixed effects[41]. SEMs are often used by human geneticists[42] and are fitted by maximum likelihood methods. Evolutionary biologists[43] and animal geneticists[44] often favour the use of mixed linear models fitted using REML methods. REML methods are favoured over maximum likelihood methods because such methods give biased estimates of the variance components when the fixed effects (for example, gender, age and deprivation score) are estimated. This can become important if the sample size is small and if there are many fixed effects. Maximum likelihood methods inflate the estimates of heritability by underestimating V_E. REML methods remove the bias by maximizing only the part of the likelihood that is not dependent on fixed effects (the residual likelihood). Another source of bias common to maximum likelihood and REML methods is that the methodology constrains the estimates of the variance components to be within the parameter space. Variances cannot be negative, and when estimates are negative, they are fixed to zero, giving upwardly biased estimates. The bias will generally be small except for small components and when estimation accuracy is lowest.

Fitting mixed linear models is computationally intensive, but their proven performance in experimental populations and their ability to exploit data composed of various relatives combined with statistical[44] and computing developments make them the preferred methods in practice.

Biased heritability

Biased heritability estimates can arise for various reasons. However, identifiability and model misspecification are often the main general issues. In this section, we describe the main reasons for confounding (that is, a lack of identifiability) and model misspecification.

Identifiability of the variance components mainly hinges on study design. It refers to the possibility of partitioning the phenotypic variation among all of the causal components. For instance, twin data cannot uniquely estimate V_A, V_D and V_C because there are only two estimates of correlation (t_{MZ} and t_{DZ}) with which it is possible to construct only two equations, as shown above. In this case, the investigator overcomes the problem by assuming either V_D or V_C to be zero and estimating the remaining variance components.

Complex pedigrees with varied degrees of relatedness and the inclusion of family members reared apart facilitate identifiability because the rate of decay of environmental and genetic sources of correlation are different as a function of the level of relatedness. In complex pedigrees, non-identifiability in mixed linear models generates numerical problems (that is, matrix singularities). Mixed linear models allow for the comparison of competing identifiable genetic models. The fit of nested models can be tested using a likelihood ratio test[39,45,46], and non-nested models can be compared using the Akaike information criterion[47].

Model misspecification is more difficult to notice than identifiability. In this case, the data may allow for the variance components to be partitioned correctly; however, the researcher fails to fit the 'correct' model. There are several reasons why this might happen. Lack of power can be one. For example, if there is dominance but not sufficient power to detect it, then it might be concluded that all genetic variance is additive genetic variance, thereby inflating h^2. Over-parameterization due to the fact that different models fit the data more or less equally well also causes model misspecification problems.

Common or shared environment. In human studies, the most important sources of confounding with additive genetic factors are shared environmental factors among relatives. The pattern of correlations is predicted to decay with the level of relatedness in both cases, albeit the rate of decay will be different for genetic and environmental factors. Mixed linear models may tend to attribute the observed patterns of correlations to additive genetic rather than environmental sources of correlation if the latter sources are not appropriately modelled. This leads to an inflation of h^2. Estimates of h^2 obtained from siblings, including twins, are likely to be the worst affected.

A class of common environment is maternal effects, which may be a direct effect of the mother's phenotype on the same phenotype of her children or the effect of the mother's phenotype on a different phenotype of her children. Maternal effects have been documented for traits such as birthweight[20] but are poorly studied for complex diseases. There is some indirect evidence that low birthweight may affect the development of diseases such as hypertension[48,49] later in life[50,51]. Whether this link is mediated through a different phenotype in the mother (for example, smoking) or the same phenotype (for example, hypertension) is unknown. Note that the phenotype of the mother is likely to have a genetic and environmental component, and therefore models can become quite cumbersome. Correlations of maternal first cousins and children of MZ mothers are best to estimate maternal genetic effects[20].

The twin method relies on the crucial assumption that common environmental factors in MZ and DZ twins are equal in magnitude[52,53]. The validity of such assumption is controversial, as even dichorionic and monochorionic MZ twins show differences in the magnitude of shared environmental effects for some traits[54–56]. Also, it is unclear the extent to which estimates of V_C from twin studies are useful for the general population.

Index case
Also called proband; Falconer used the term propositi to refer to the probands. This is the patient within a family who is first recruited to the study. Because other relatives are actively recruited as a consequence of the index case recruitment, the families are not a representative sample of the general population.

Proband concordance rate
(q_c). Defined as the proportion of twins with the disease of interest among twins who are independently ascertained. That is, each twin pair is counted once for each twin independently brought to the study. It can be computed as $q_{twin} = 2n_{11}/(2n_{11} + n_{10})$.

Nested models
Two statistical models are nested if both models contain the same terms and one model has at least one additional term. The model with the larger number of terms is the full model, and the other is the reduced model. For instance, model $P = A + C + E$ is nested within $P = A + D + C + E$. Models $P = A + C + E$ and $P = A + D + E$ are non-nested.

Likelihood ratio test
(LRT). Used to compare how well a model (full model) and a subset of that model (reduced model) fit the data. It is calculated as $LRT = -2ln(L_{Reduced}/L_{Full})$ and distributed as χ_r^2 where r is the difference of parameters fitted in the two models.

Akaike information criterion
An approach used to compare non-nested models. The Akaike information criterion penalizes complicated models by adding two times the number of fitted parameters to twice the negative value of the maximum likelihood. The model with the smallest Akaike information criterion is chosen as the most parsimonious.

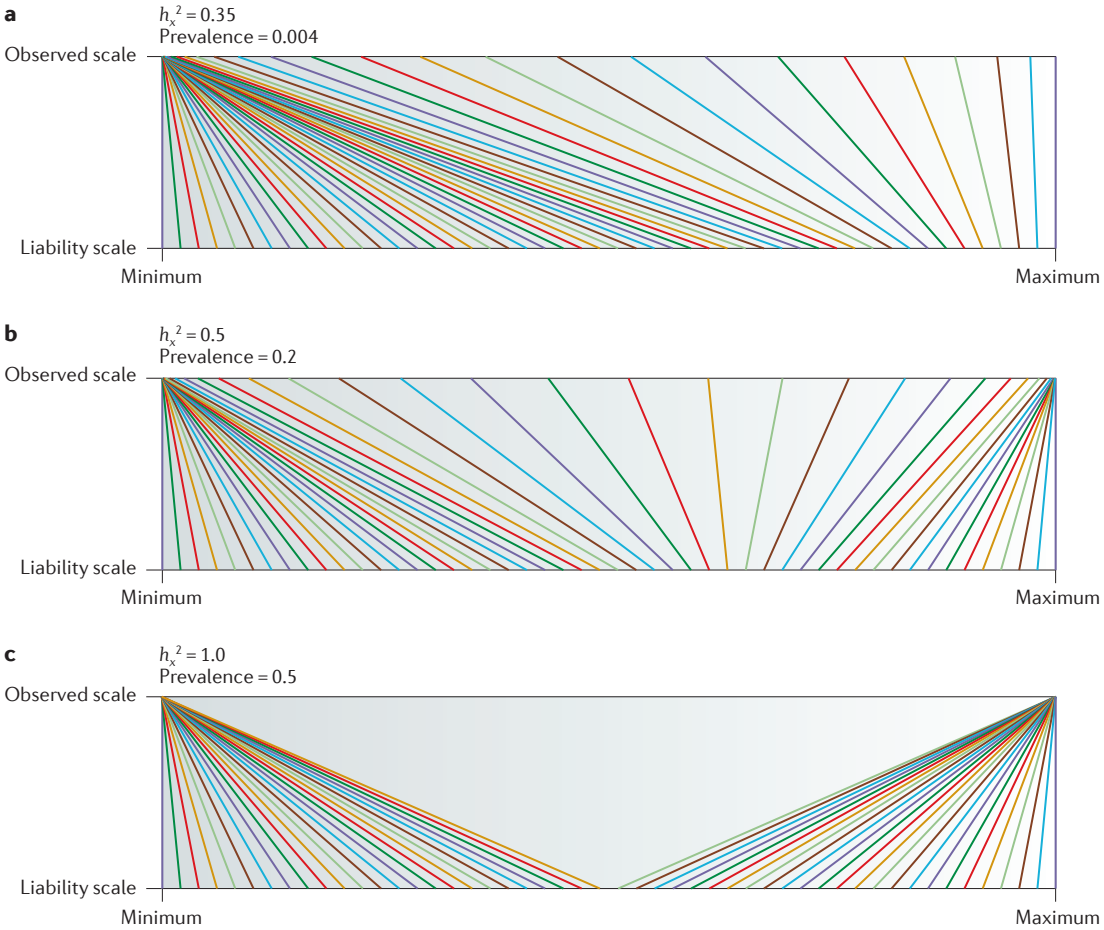


Figure 1 | **Comparison of the genotypic values of individuals on the observed and liability scale.** Different liability-scale heritabilities (h_x^2) and prevalences are shown to demonstrate how the change of scale generates epistatic variance. The slanting lines in each example connect the genotypic values on the liability scale with the genotypic values on the observed scale. On the liability scale, genetic differences are assumed to be additive (that is, all genetic variance is additive). The variation observed on the top of the three examples shows that the gene substitution effects are no longer additive in this scale, thereby generating epistatic variance. The observed scale represents the penetrance and ranges from 0 to 1. The liability scale represents a deviate of the assumed normally distributed liability to disease and ranges from −4 to 4.

Genotype-by-environment interactions. Partitioning the observed or unobserved liability variation (V_P) into components attributable to different causes (for simplicity, $V_G + V_E$) has two potential complications that arise from the assumption of independence between genotype (G) and environment (E). The first complication is that G and E might be correlated. In this case, $V_P = V_G + V_E + 2*COV_{GE}$, where COV_{GE} is the covariance between G and E. When the COV_{GE} is not modelled, it usually becomes incorporated into estimates of V_G, thereby biasing the estimates of H^2 (and h^2). The direction of the bias will depend on the sign of the COV_{GE}. For instance, a parental genotype with an inclination for a high-caloric diet could lead to a high-caloric food environment on which their children will grow up, leading to a positive COV_{GE}. However, if as a consequence of the high-caloric diet one of the parents has suffered a major health scare (for example, a heart attack) that has lead to a change in habits, the opposite might be true.

The second complication arises from genotype-by-environment (G×E) interactions. This means that genotypes vary in their environmental sensitivity. In the presence of G×E interactions, $V_P = V_G + V_E + V_{G\times E}$. Misspecification of this model will provide biased estimates of H^2. The direction of bias will depend on whether the environmental component is familial[57] and whether the trait is multifactorial. If the environment is familial and the trait is multifactorial, then there will be inflated estimates of V_G (REFS 58–60). Moreover, if the environment is not familial (that is, if it is not genetically determined), then environmental sensitivity will increase V_E, and H^2 will be underestimated.

Change of scale. Estimates of heritability obtained on the 0/1 observed scale ($h_{0/1}^2$) need to be adjusted to the liability scale (h_x^2) because the heritability in the observed scale depends on the disease prevalence. The transformation allows for the comparison of heritabilities of

Identity-by-descent
Two alleles are identical-by-descent if they are a copy of the same allele carried in an ancestral individual.

Realized identity-by-descent
Actual, as opposed to expected, identity-by-descent sharing between pairs of individuals as estimated from their genotypes. It accounts for the deviations from the expected identity-by-descent values that arise from the random segregation of alleles.

diseases with varying prevalence or heritability estimates for a disease when the prevalence varies across populations, at the expense of accepting the validity of the liability model.

Theoretically, the heritability estimated on the observed scale is $z_g^2/q_g(1-q_g)$ times as large as the heritability estimated directly on the underlying normally distributed liability scale[24], where z_g is the ordinate of the normal distribution at the threshold point that determines the population prevalence of the disease (q_g). However, in practice, the performance of this adjustment depends on the study design, the heritability and the prevalence, as well as on their interactions[61]. Inflated estimates are expected for most study designs when the disease prevalence drops below 25%.

What is the source of the inflation? FIGURE 1 shows the genotypic values of individuals on the liability scale and the corresponding genotypic values of the observed scale. On the liability scale, the difference between adjacent genotypes remains the same across the whole scale; however, on the observed scale, adjacent genotypes have very different genotypic values in different parts of the risk scale. The differences are a function of the prevalence and the heritability. In general, little difference is observed between the two scales when the heritability on the liability scale is low and the prevalence is close to 50%, but substantial disparities arise when the heritability is high and the prevalence is close to either 0 or 100%. These disparities generate substantial amounts of epistatic variance on the observed scale (FIG. 2). The estimates of correlations obtained from different sorts of relatives capture some of this epistatic variance, which is then transformed to the liability scale as additive, leading to an inflation of h_x^2.

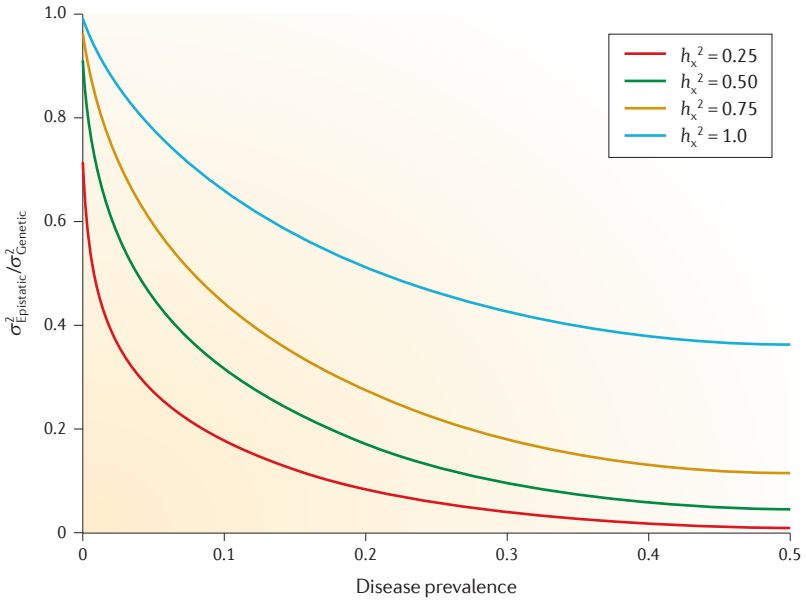


Figure 2 | **Proportion of total genetic variance, which is additive on the underlying scale, that becomes epistatic genetic variance on the observed scale.** The proportion is shown for four different levels of heritability. h_x^2, heritability of liability to disease.

Disease diagnosis and ascertainment. Often, the diagnosis of the probands prompts the recruitment of their relatives. In this case, the sample is not a representative random sample of the population, and the estimates of heritability obtained are biased. Also, because not all instances of a disease in the population may be diagnosed, a population-based sample may not diagnose all of the available patients. This may lead to different rates of ascertainment and diagnosis among probands and other relatives. For instance, the rate of diagnosis among relatives of probands might be higher than the rate of diagnosis in the population because their families have prompted the interest of the investigator. This would inflate the estimates of heritability[62].

An additional problem caused by incomplete ascertainment and diagnosis is that the level of ascertainment and diagnosis affects the correlation in liability among relatives. The liability model considers this to be a source of environmental variance, thereby reducing the correlation among relatives. Let us assume a disease with $q_g = 0.05$ and $q_r = 0.1$. Using Falconer's method, we estimate $t = 0.18$; however, if the diagnosis rate was halved, then $q_g = 0.025$, $q_r = 0.05$ and $t = 0.13$. Heritability is substantially underestimated when the disease prevalence is high and rates of diagnosis and ascertainment are low[63]. Moreover, different estimates of heritability are expected from different studies if the data have been gathered in different ways.

Improvements in modelling and data. There are several possibilities that would allow better separation of the causal components. At the genetic level, complex pedigrees allow a detailed description of the correlation structure of the data (that is, the relationships among pedigree members) by exploiting the expected identity-by-descent sharing across the genome. The availability of genotyping arrays allows estimation of the realized identity-by-descent among individuals[64] and relaxation of the assumption that individuals with unknown relationships are unrelated. Indeed, the availability of large numbers of genetic markers allows the estimation of relationships among individuals of unknown pedigree and even among individuals considered to be unrelated, allowing inferences about heritability[65–67]. A discussion of the differences between family-based estimates (that is, from related samples) and population-based estimates (that is, from unrelated samples) of heritability are outside the scope of this paper and are comprehensively treated elsewhere[68].

Further gain, perhaps substantial, could be obtained by substituting simple models of environmental correlation by more realistic ones. Modelling the environmental correlation as a function of, for instance, age, gender, time living together and time living apart would improve discrimination of genetic and environmental factors. Additional improvements could be achieved if proxies of environmental exposures could be identified that allowed the modelling of the environmental correlation among relatives. An attractive, although speculative, proxy could be epigenetic modifications on the genome. For instance, tobacco smoking is known to

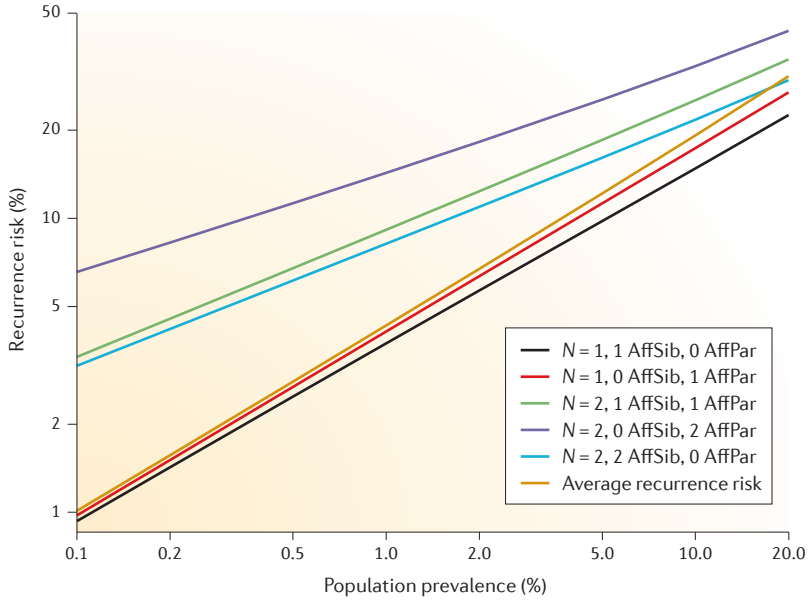


Figure 3 | **Sibling recurrence risk for different patterns of family history as a function of the population prevalence.** A narrow-sense heritability (h^2) of liability to disease equal to 0.5 is assumed. It is also assumed that the disease status of the two parents and *N* of their children is known. *N* children were phenotyped in each case, as indicated. Note that, by having an estimate of h^2, it is possible to calculate recurrence risks to match a person's family history and to update those estimates as information on relatives accumulates. AffPar, affected parent; AffSib, affected sibling.

cause differential methylation of *F2RL3* (REF. 69). Other epigenetic modifications could be triggered by other risk factors, and the combination of epigenetic modifications could be used to model environmental correlation.

Using and abusing heritability

Controversies surrounding heritability estimates. Several controversies have surrounded the use of heritability estimates and the biases that can be associated with them. These controversies are misguided and are unrelated in their nature.

Variation in heritability across studies has led to suggestions that estimates are either wrong or of little use. The variation could arise from any of the points previously highlighted, from other points that we have not discussed[70–72] or simply because the heritability varies from population to population. Estimates of heritability quantify how much of the variation in disease liability in a population can be attributed to genetic variation. This justifies further research into the genetic causes of variation in this population and opens the possibility of exploiting that genetic variation to our benefit (for example, genomic prediction). Despite its caveats (such as difficulty of estimation and potential biases), heritability is the single most useful measure of familial aggregation of disease available for two reasons. First, it has the potential to simultaneously capture information from multiple relatives (for example, the sibling relative risk is useful only for siblings). Second, it attempts to separate environmental and genetic sources of familial correlation. Other epidemiological measures, such as the population attributable risk, are useful in the context of exposures but are difficult to interpret in the context of genetics.

Another source of controversy and misunderstanding arises from the change in disease prevalence with time (for example, the increase of obesity prevalence in recent years[73]) and the mistaken interpretation that this provides evidence that genes play little part in trait variation, as allele frequencies have not changed. If the environment changes, then the heritability may change because the threshold in liability will change. There is, however, little evidence that this has happened for obesity over the past 40 years[74–76].

Another common misconception is that a low concordance rate among MZ twins implies a low contribution of genetic factors to variation in disease liability. Under a liability model, this is not correct because the concordance rate depends on the disease prevalence. For example, the MZ concordance for prostate cancer[77] (q_{MZ} = 0.21) and multiple sclerosis[78] (q_{MZ} = 0.25) are similar, but their estimated heritabilities are ~0.42 and 0.72, respectively, because prostate cancer is ~40-fold more prevalent than multiple sclerosis.

Finally, the last source of controversy arises from the validity of the liability model. This model was proposed to study diseases for which liability was determined by large numbers of small additive genetic and environmental factors. Current evidence suggests that the model may be reasonable for complex diseases[79,80]. Hence, controversies about the utility of the model and heritability estimates based on it would be best not exemplified by rare monogenic traits, such as phenylketonuria[81]. The model was not envisaged for such diseases and may not always give a reasonable description of them[82]. Proposing alternative models[83,84] and ways of testing them seems to be a more constructive way forward.

Uses of heritability estimates. In a population, H^2 and h^2 set the upper limits of the utility of genetic variation for individual prediction of disease risk[2,85] and prediction based on family history[3,86], respectively (FIG. 3). However, more insight will be gained by identifying the causes underlying the estimated genetic contribution. The technology is available, and genome-wide association studies will soon be followed by sequence-wide association studies. However, to capitalize fully on the genomic technology, we would need large familial cohorts on top of large prospective studies, such as the UK Biobank[87], and we would also need to develop better ways of measuring environmental sources of familial covariance. These resources could, as a first aim, be used to estimate heritability, but they would be far more useful to identify the causes of familial resemblance (including private as well as common disease variants), which should, without doubt, be the overarching aim.

From a public health perspective, environmental interventions should be pursued that can modify the liability threshold to benefit the population. But would some sort of genetic intervention be useful? Could preventive medicine offer something different from genetic

Population attributable risk
Also called the population attributable fraction. For a given disease, risk factor and population, the population attributable risk for the population incidence rate is the fraction by which the incidence rate of the disease in the population would be reduced if the risk factor was eliminated.

Sequence-wide association studies
The extension of array-based genome-wide association studies to whole-genome sequence-based association studies.

classes A and C in BOX 2? It is unlikely that public health interventions could sufficiently shift the threshold to make genetic variation irrelevant. This happened, for instance, for some infectious diseases after population-wide vaccination but may be more complicated for age-related complex diseases. It is equally unlikely that by focusing only on understanding the genetics of disease, any meaningful advice could be offered to class C members. What class C members would probably ask is 'what can we do to move our liability score to the left side of the threshold?' We would not know that answer without a clear understanding of the environmental sources of variation, including those that are familial. In summary, the two interventions are not incompatible, and ways of understanding and implementing them should be pursued.

1. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era—concepts and misconceptions. *Nature Rev. Genet.* **9**, 255–266 (2008).
This Review paper is a clear and concise introduction to the concept of heritability.
2. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6**, e1000864 (2010).
3. Smith, C. Recurrence risks for multifactorial inheritance. *Am. J. Hum. Genet.* **23**, 578–588 (1971).
This paper shows the correspondence of Falconer's abrupt liability model with a normal genetic distribution of liability with a cumulative normal risk function.
4. Gianola, D. Theory and analysis of threshold characters. *J. Animal Sci.* **54**, 1079–1096 (1982).
5. Lush, J. L., Lamoreux, W. F. & Hazel, L. N. The heritability of resistance to death in the fowl. *Poultry Sci.* **27**, 375–388 (1948).
The authors of this paper suggest, for the first time, to transform the heritability estimates obtained on the 0/1 scale ($h_{0/1}^2$) to estimates on the liability scale (h_x^2).
6. Robertson, A. & Lerner, I. M. The heritability of all-or-none traits — viability of poultry. *Genetics* **34**, 395–411 (1949).
7. Boomsma, D., Busjahn, A. & Peltonen, L. Classical twin studies and beyond. *Nature Rev. Genet.* **3**, 872–882 (2002).
8. Edwards, J. H. Familial predisposition in man. *Br. Med. Bull.* **25**, 58–64 (1969).
9. Falconer, D. S. Inheritance of liability to certain diseases estimated from incidence among relatives. *Ann. Hum. Genet.* **29**, 51–76 (1965).
This article describes how to transform prevalence information among relatives into an estimate of correlation and hence heritability.
10. Eisenhart, C. The assumptions underlying the analysis of variance. *Biometrics* **3**, 1–21 (1947).
11. Hopper, J. L. Variance components for statistical genetics: applications in medical research to characteristics related to human diseases and health. *Statist. Methods Med. Res.* **2**, 199–223 (1993).
12. Smith, C., Falconer, D. S. & Duncan, L. J. P. Statistical and genetic study of diabetes: II. Heritability of liability. *Ann. Hum. Genet.* **35**, 281–299 (1972).
13. Rose, S. P. R. Commentary: heritability estimates — long past their sell-by date. *Int. J. Epidemiol.* **35**, 525–527 (2006).
14. Lewontin, R. C. Annotation: the analysis of variance and the analysis of causes. *Am. J. Hum. Genet.* **26**, 400–411 (1974).
15. Maher, B. Personal genomes: the case of the missing heritability. *Nature* **456**, 18–21 (2008).
16. Slatkin, M. Epigenetic inheritance and the missing heritability problem. *Genetics* **182**, 845–850 (2009).
17. Eichler, E. E. *et al.* Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Rev. Genet.* **11**, 446–450 (2010).
18. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
19. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Longman, 1996).
This is a classic introductory book on quantitative genetics for anyone new to the field.
20. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).
This is a comprehensive and more advanced book on quantitative genetics than Reference 19. Volume 2 is available at the author's Web page.
21. Elston, R. C. & Rao, D. C. Statistical modeling and analysis in human genetics. *Annu. Rev. Biophys. Bioengineer.* **7**, 253–286 (1978).
22. Sham, P. *Statistics in Human Genetics* (Arnold, 1998).
This is an excellent reference book. Concepts of statistical genetics are clearly explained in the context of human genetics.
23. Pearson, K. & Lee, A. Mathematical contributions to the theory of evolution VII — on the application of certain formulae in the theory of correlation to the inheritance of characters not capable of quantitative measurement. *Proc. R. Soc.* **66**, 324–327 (1900).
24. Dempster, E. R. & Lerner, I. M. Heritability of threshold characters. *Genetics* **35**, 212–236 (1950).
This paper and its appendix, by A. Robertson, show the relationship between the heritability estimates obtained in observed and liability scales.
25. Allen, G., Harvald, B. & Shields, J. Measures of twin concordance. *Acta Genet. Stat. Med.* **17**, 475–481 (1967).
26. Trimble, B. K. *An Empirical Simulation of Quasi-Continuous Inheritance Using Human Birthweight Data.* Thesis, McGill Univ. (1971).
27. Falconer, D. S. Inheritance of liability to diseases with variable age of onset with particular reference to diabetes mellitus. *Ann. Hum. Genet.* **31**, 1–20 (1967).
28. Smith, C. Heritability of liability and concordance in monozygous twins. *Ann. Hum. Genet.* **34**, 85–91 (1970).
29. Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
30. Reich, T., Morris, C. A. & James, J. W. Use of multiple thresholds in determining mode of transmission of semi-continuous traits. *Ann. Hum. Genet.* **36**, 163–168 (1972).
31. Thompson, R. Maximum likelihood approach to estimate of liability. *Ann. Hum. Genet.* **36**, 221–231 (1972).
32. Gilmour, A. R., Anderson, R. D. & Rae, A. L. The analysis of binomial data by a generalized linear mixed model. *Biometrika* **72**, 593–599 (1985).
This paper describes a method that fits mixed linear models to binomial data and that allows estimation of the variances directly on the liability scale.
33. Nelder, J. A. & Wedderburn, R. W. M. Generalized linear models. *J. R. Stat. Soc. A* **135**, 370–384 (1972).
34. Harville, D. A. & Mee, R. W. A. Mixed-model procedure for analyzing ordered categorical-data. *Biometrics* **40**, 393–408 (1984).
35. Visscher, P. M., Haley, C. S., Heath, S. C., Muir, W. J. & Blackwood, D. H. Detecting QTLs for uni- and bipolar disorder using a variance component method. *Psychiatr. Genet.* **9**, 75–84 (1999).
36. Visscher, P. M., Haley, C. S. & Knott, S. A. Mapping QTLs for binary traits in backcross and F-2 populations. *Genet. Res.* **68**, 55–63 (1996).
37. Burton, P. R. *et al.* Genetic variance components analysis for binary phenotypes using generalized linear mixed models (GLMMs) and Gibbs sampling. *Genet. Epidemiol.* **17**, 118–140 (1999).
38. Sorensen, D. & Gianola, D. *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics* (Springer, 2004).
39. Visscher, P. M. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Res. Hum. Genet.* **9**, 490–495 (2006).
40. Martin, N., Boomsma, D. & Machin, G. A twin-pronged attack on complex traits. *Nature Genet.* **17**, 387–392 (1997).
41. Posthuma, D. & Boomsma, D. I. Mx scripts library: structural equation modeling scripts for twin and family data. *Behav. Genet.* **35**, 499–505 (2005).
42. Neale, M. C. & Cardon, L. R. *Methodology for Genetic Studies of Twins and Families* (Kluwer Academic Publishers, 1992).
43. Kruuk, L. E. Estimating genetic parameters in natural populations using the "animal model". *Phil. Trans. R. Soc. Lond. B* **359**, 873–890 (2004).
44. Patterson, H. D. & Thompson, R. Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–554 (1971).
This is a classic and advanced paper that describes the method of restricted maximum likelihood estimation.
45. Stram, D. O. & Lee, J. W. Variance-components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–1177 (1994).
46. Self, S. G. & Liang, K. Y. Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605–610 (1987).
47. Akaike, H. A new look at statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
48. Law, C. M. & Shiell, A. W. Is blood pressure inversely related to birth weight? The strength of evidence from a systematic review of the literature. *J. Hypertension* **14**, 935–941 (1996).
49. Godfrey, K. M. & Barker, D. J. Fetal programming and adult health. *Publ. Health Nutr.* **4**, 611–624 (2001).
50. Keita, S. O., Payne, P., Pascalev, A. K. & Roya, C. Abstract 37 in 'Abstracts of the 32nd Annual Meeting of the Human Biology Association Philadelphia, Pennsylvania March 28–29, 2007'. *Am. J. Hum. Biol.* **19**, 261–262 (2007).
51. Wells, J. C. K. The thrifty phenotype as an adaptive maternal effect. *Biol. Rev.* **82**, 143–172 (2007).
52. Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. A test of the equal-environment assumption in twin studies of psychiatric illness. *Behav. Genet.* **23**, 21–27 (1993).
53. Hopper, J. Genes for osteoarthritis: interpreting twin data — commentary. *Br. Med. J.* **312**, 943–944 (1996).
54. Christian, J. C. *et al.* Variance of plasma free and esterified cholesterol in adult twins. *Am. J. Hum. Genet.* **28**, 174–178 (1976).
55. Reed, T., Uchida, I. A., Norton, J. A. Jr & Christian, J. C. Comparisons of dermatoglyphic patterns in monochorionic and dichorionic monozygotic twins. *Am. J. Hum. Genet.* **30**, 383–391 (1978).
56. Kaminsky, Z. A. *et al.* DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genet.* **41**, 240–245 (2009).
57. Scarr, S. & Mccartney, K. How people make their own environments — a theory of genotype-environment effects. *Child Dev.* **54**, 424–435 (1983).
58. Cavalli-Sforza, L. L. & Feldman, M. W. Cultural versus biological inheritance: phenotypic transmission from parents to children. (A theory of effect of parental phenotypes on children's phenotypes). *Am. J. Hum. Genet.* **25**, 618–637 (1973).
59. Lathrope, G. M., Lalouel, J. M. & Jacquard, A. Path analysis of family resemblance and gene-environment interaction. *Biometrics* **40**, 611–625 (1984).
60. Rao, D. C. & Morton, N. E. Path analysis of family resemblance in presence of gene-environment interaction. *Am. J. Hum. Genet.* **26**, 767–772 (1974).
61. Vanvleck, L. D. Estimation of heritability of threshold characters. *J. Dairy Sci.* **55**, 218–225 (1972).
62. Hrubec, Z. Effect of diagnostic ascertainment in twins on assessment of genetic factor in disease etiology. *Am. J. Hum. Genet.* **25**, 15–28 (1973).
63. Smith, C. Concordance in twins — methods and interpretation. *Am. J. Hum. Genet.* **26**, 454–466 (1974).
64. Visscher, P. M. *et al.* Genome partitioning of genetic variation for height from 11,214 sibling pairs. *Am. J. Hum. Genet.* **81**, 1104–1110 (2007).

65. Ritland, K. Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **50**, 1062–1073 (1996). **Ritland was the first to propose the use genetic markers to estimate relationships and to use these estimates to estimate heritability.**
66. Ritland, K. & Ritland, C. Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. *Evolution* **50**, 1074–1082 (1996).
67. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nature Genet.* **42**, 565–569 (2010). **This paper shows that SNP arrays can be used to estimate distant relationships among individuals considered to be unrelated and these used for REML estimation.**
68. Visscher, P. M., Yang, J. & Goddard, M. E. A commentary on 'Common SNPs explain a large proportion of the heritability for human height' by Yang *et al.* (2010). *Twin Res. Hum. Genet.* **13**, 517–524 (2010).
69. Breitling, L. P., Yang, R. X., Korn, B., Burwinkel, B. & Brenner, H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am. J. Hum. Genet.* **88**, 450–457 (2011).
70. Bishop, S. C. & Woolliams, J. A. On the genetic interpretation of disease data. *PLoS ONE* **5**, e8940 (2010).
71. Price, B. Primary biases in twin studies — a review of prenatal and natal difference-producing factors in monozygotic pairs. *Am. J. Hum. Genet.* **2**, 293–352 (1950).
72. Bundey, S. Uses and limitations of twin studies. *J. Neurol.* **238**, 360–364 (1991).
73. Veerman, J. L. On the futility of screening for genes that make you fat. *PLoS Med.* **8**, e1001114 (2011).
74. Maes, H. H. M., Neale, M. C. & Eaves, L. J. Genetic and environmental factors in relative body weight and human adiposity. *Behav. Genet.* **27**, 325–351 (1997).
75. Wardle, J., Carnell, S., Haworth, C. M. A. & Plomin, R. Evidence for a strong genetic influence on childhood adiposity despite the force of the obesogenic environment. *Am. J. Clin. Nutr.* **87**, 398–404 (2008).
76. Musani, S. K., Erickson, S. & Allison, D. B. Obesity — still highly heritable after all these years. *Am. J. Clin. Nutr.* **87**, 275–276 (2008).
77. Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *New Engl. J. Med.* **343**, 78–85 (2000).
78. Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D. & Ebers, G. C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl Acad. Sci. USA* **100**, 12877–12882 (2003).
79. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
80. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genet.* **44**, 247–250 (2012).
81. Pearce, N. Epidemiology in a changing world: variation, causation and ubiquitous risk factors. *Int. J. Epidemiol.* **40**, 503–512 (2011).
82. Kidd, K. K. & Cavalli-Sforza, L. L. An analysis of the genetics of schizophrenia. *Biodemography Soc. Biol.* **20**, 254–265 (1973).
83. Slatkin, M. Exchangeable models of complex inherited diseases. *Genetics* **179**, 2253–2261 (2008).
84. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl Acad. Sci. USA* **109**, 1193–1198 (2012).
85. Rowe, S. J. & Tenesa, A. Human complex trait genetics: lifting the lid of the genomics toolbox — from pathways to prediction. *Curr. Genom.* **13**, 213–224 (2012).
86. Curnow, R. N. The multifactorial model for the inheritance of liability to disease and its implications for relatives at risk. *Biometrics* **28**, 931–946 (1972).
87. Peakman, T. C. & Elliott, P. The UK Biobank sample handling and storage validation studies. *Int. J. Epidemiol.* **37**, 2–6 (2008).

Acknowledgements
This work was supported by Cancer Research UK (C12229/A13154) and the UK Biotechnology and Biological Sciences Research Council (BB/K000195/1). We acknowledge the financial support provided by the MRC–HGU Core Fund and the Roslin Institute through its Strategic Programme Grant. We thank W. G. Hill for helpful comments on an earlier version of the manuscript.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Albert Tenesa's homepage: http://www.roslin.ed.ac.uk/albert-tenesa
Chris S. Haley's homepage: http://www.hgu.mrc.ac.uk/people/c.haley.html
***Nature Reviews Genetics* Series on Study designs:** http://www.nature.com/nrg/series/studydesigns/index.html
ALL LINKS ARE ACTIVE IN THE ONLINE PDF