Author/s:
Wray, NR;Goddard, ME

Title:
Multi-locus models of genetic risk of disease

Date:
2010-02-02

Citation:
Wray, N. R. & Goddard, M. E. (2010). Multi-locus models of genetic risk of disease. Genome Medicine, 2 (2), https://doi.org/10.1186/gm131.

Persistent Link:
https://hdl.handle.net/11343/264414

Genome Medicine

RESEARCH Open Access

Multi-locus models of genetic risk of disease

Naomi R Wray*[1] and Michael E Goddard[2]

Abstract

Background: Evidence for genetic contribution to complex diseases is described by recurrence risks to relatives of diseased individuals. Genome-wide association studies allow a description of the genetics of the same diseases in terms of risk loci, their effects and allele frequencies. To reconcile the two descriptions requires a model of how risks from individual loci combine to determine an individual's overall risk.

Methods: We derive predictions of risk to relatives from risks at individual loci under a number of models and compare them with published data on disease risk.

Results: The model in which risks are multiplicative on the risk scale implies equality between the recurrence risk to monozygotic twins and the square of the recurrence risk to sibs, a relationship often not observed, especially for low prevalence diseases. We show that this theoretical equality is achieved by allowing impossible probabilities of disease. Other models, in which probabilities of disease are constrained to a maximum of one, generate results more consistent with empirical estimates for a range of diseases.

Conclusions: The unconstrained multiplicative model, often used in theoretical studies because of its mathematical tractability, is not a realistic model. We find three models, the constrained multiplicative, Odds (or Logit) and Probit (or liability threshold) models, all fit the data on risk to relatives. Currently, in practice it would be difficult to differentiate between these models, but this may become possible if genetic variants that explain the majority of the genetic variance are identified.

Background

Complex genetic diseases are defined as those influenced by multiple genes and by environmental effects. In the past, individual genetic variants contributing to the risk of disease were usually not known, so the contribution of genes to disease was recognised through increased risk of disease in relatives of affected probands. Modeling allowed the genetic component of disease to be expressed as variance components and heritabilities. However, with the advent of genome-wide association studies (GWAS), individual genetic risk factors, or at least markers linked to them, are identifiable. This provides a description of the genetics in quite different terms to the traditional use of variance components. The new description is based on the frequency of individual risk alleles and their effect sizes expressed either as the relative risk or the odds ratio.

A clear picture is emerging as more and more results from GWAS are published about the effect sizes of individual loci that contribute to disease. For instance, allelic odds ratios at markers are typically estimated to be <1.5 and risk alleles can be the minor or major frequency allele. At present, there is little evidence of departure from a multiplicative model (on the observed disease risk scale) of disease [1], within and across loci, but this is based on combining only a limited number of markers and explaining only a small proportion of the genetic variance.

To reconcile the traditional description in terms of risk to relatives with the description based on individual risk loci, we need a model of how the risk loci combine to determine the total genetic risk for an individual person. Simple models are unlikely to be a true representation of complex diseases, but they allow us to explore the boundaries of possible genetic architectures that remain consistent with observed data. Several models are commonly used. Unfortunately the terms used to describe these models are confusing. For example, the terms 'additive' and 'multiplicative' can both be used to describe

*Correspondence: naomi.wray@qimr.edu.au
[1]Genetic Epidemiology and, Queensland Institute of Medical Research, Herston Road, Brisbane, Queensland 4006, Australia
Full list of author information is available at the end of the article

Table 1. Recurrence risk (λ_R) to relatives (of type *R*) for several common complex genetic diseases ordered by prevalence (*K*)

Disease	Reference	*K*	λ_{MZ} [a]	λ_{Sib} [b]	λ_{OP}	H^2_{01} [c] = (λ_{MZ} – 1) / (1 – *K*)	(λ_{Sib} – 1) [d] / (λ_{OP} – 1)	(λ_{MZ} – 1) [e] / (λ_{Sib} – 1)	λ_{MZ} [f] / λ^2_{Sib}	h^2_L [g]
Major depression (population cohort)	[27]	0.24	2	1.3		0.32		3.3	1.2	0.34
Age related macular degeneration	[28,29]	0.12	4.7	2.1		0.50		3.4	1.1	0.64
Myocardial infarction	[30]	0.056	4.6	3.2		0.21		1.6	0.4	0.72
Breast cancer	[31]	0.036	4.1	2.2	1.9	0.12	1.3	2.6	0.8	0.37
Type II diabetes	[32]	0.028	10.4	3.5		0.27		3.8	0.8	0.58
Asthma	[33]	0.019	6.6	3.4		0.11		2.3	0.6	0.49
Rheumatoid arthritis	[34]	0.01	12.2	3.6		0.11		4.3	0.9	0.42
Bipolar disorder	[5]	0.01	60	7	7	0.60	1.0	10	1.2	0.70
Schizophrenia	[3]	0.0085	52.1	8.6	10	0.44	0.8	6.7	0.7	0.76
Type I diabetes	[35]	0.005	79	14		0.39		6.0	0.4	0.85
Multiple sclerosis	[36]	0.001	190	20		0.19	~1	9.9	0.5	0.68
Crohn's disease	[37]	0.001	600	64		0.60		10	0.1	1.00
Ankylosis spondylitis	[6]	0.001	630	82	79	0.63	1.0	7.8	0.1	1.00
Systemic lupus erythematosus	[38]	0.001		29	27		1.1			0.80
	[39,40]	0.0003	774	65		0.24		12	0.2	0.84

[a]The maximum prevalence for K_{MZ} is 1, so $\lambda_{MZ} = K_{MZ}/K$ is constrained to be ≤1/*K*. λ_{MZ} was calculated from probandwise concordance rates K_{MZ} and prevalence rates if λ_{MZ} was not directly reported. [b]Estimated from either sibling, dizygotic twin or first degree relative risks. [c]Broad sense heritability on the risk scale (Equation 1). [d]This ratio is expected to be 1 in the absence of dominance effects on the risk scale. [e]This ratio is expected to be 2 under an additive model on the risk scale. [f]This ratio is expected to be 1 under the unconstrained Risch model. [g]Calculated from the estimates of *K* and λ_{Sib} [41,42], constrained to a maximum of 1.

the same fundamental model because a multiplicative model on the observed disease risk scale (the 'risk scale') is equivalent to an additive model on the logarithm of the risk scale. Moreover, the multiplicative model can imply multiplicativity of allelic relative risks [2,3], or of odds ratios [4], or that risk alleles are needed at all loci in order to develop disease [5].

In this paper we show how the parameters for the individual risk loci (effect, allele frequency and number of loci) plus a model for combining the effects of individual loci determine the traditional parameters such as risk to relatives. The purpose of the paper is to compare the predictions made by different models and to determine which model(s) best fit the observed data. Before explaining the different models of genetic risk we first describe the genetic population parameters of recurrence risk to relatives.

Recurrence risk to relatives

The genetic epidemiology of complex genetic diseases can be described in terms of the observable parameters of disease prevalence and relative risk to relatives of diseased probands (Table 1). Risks of disease in relatives provide an upper limit to the genetic component because common environmental factors may also increase risk to relatives. However, for the purposes of this paper we will assume risk to relatives is due to their genetic similarity. The recurrence risk for relatives of type *R* (λ_R) is calculated as the ratio of the prevalence in the population of relatives of type *R* (K_R) to the overall population prevalence (*K*), $\lambda_R = K_R/K$. As the maximum value for K_R is 1 and the prevalence in monozygotic (MZ) twins of probands, K_{MZ}, will be the highest of all relative types, there is a constraint that $\lambda_{MZ} \leq 1/K$, so that higher values of λ_{MZ} (and all λ_R) are often observed for diseases of lower prevalence (Table 1). Despite being observable, the parameters *K* and λ_R are subject to considerable sampling variance. For Table 1, we have tried, where possible, to take estimates from reviews or large studies, but large study samples simply do not exist for low prevalence disorders - for example, the λ_{MZ} for ankylosis spondylitis [6] is based on only 27 MZ twin probands. Nonetheless, we can use these examples as a guide to assessing realistic scenarios for disease.

The risk to different classes of relatives (that is, λ_R) depends on the magnitude of genetic variance components. The total genetic variance is traditionally decomposed into additive variance, dominance variance and various types of epistatic variance. The relationship between relative risks and variance components on risk scale was derived by James [7], who showed that the probability of disease in relatives of type *R* can be expressed as:

$$K_R = K + \mathrm{cov}(X,R)/K$$

with cov(*X*,*R*) the genetic covariance between the proband, *X* , and a relative, *R*. For individuals *X* and *R* we

define r to be the relationship between them, r = 2 × Probability of identity by descent (IBD) of random alleles (that is, twice the ancestry or kinship coefficient) and u is the probability of both alleles being IBD at a locus, so that

$$\mathrm{cov}(X, R) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} r^k u^l V_{A(k)D(l)}$$

where $V_{A(k)D(l)}$ denotes the genetic variance component with k A and l D terms [3,5,8,9]. So for R = MZ twin, r = 1, u =1, then:

$$\mathrm{Cov}(X, MZ) = V_{A_{01}} + V_{D_{01}} + V_{AA_{01}} + V_{AD_{01}} + V_{DD_{01}} + V_{AAD_{01}} + V_{AAA_{01}} + \ldots = V_{G_{01}}$$

We use the '01' subscript to emphasize the observed zero-one (not diseased-diseased) risk scale of measurement. Therefore, an estimate of the broad sense heritability on the risk scale (H^2_{01}) is:

$$H^2_{01} = \frac{V_{G_{01}}}{V_{P_{01}}} = \frac{(\lambda_{MZ} - 1)K^2}{K(1-K)} = \frac{(\lambda_{MZ} - 1)K}{(1-K)} \qquad \text{(Equation 1)}$$

since the phenotypic variance on the risk scale is $V_{P_{01}} = K(1 - K)$. For the diseases listed in Table 1, H^2_{01} ranges from 0.11 to 0.63, but the heritability on this scale is not a normally reported statistic because of its dependence on disease prevalence. When the relatives are sibs, R = Sib, r = ½, u = ¼, then:

$$\mathrm{Cov}(X, Sib) = \frac{V_{A_{01}}}{2} + \frac{V_{D_{01}}}{4} + \frac{V_{AA_{01}}}{4} + \frac{V_{AD_{01}}}{8} + \frac{V_{DD_{01}}}{16} + \frac{V_{AAA_{01}}}{8} + \frac{V_{AAD_{01}}}{16} + \ldots$$

When the relatives are parents or offspring, R = OP, r =1/2, u = 0, then:

$$\mathrm{Cov}(X, OP) = \frac{V_{A_{01}}}{2} + \frac{V_{AA_{01}}}{4} + \frac{V_{AAA_{01}}}{8} + \ldots$$

Therefore, $\lambda_{Sib} \geq \lambda_{OP}$ since the former includes dominance terms; the magnitude of the ratio:

$$\frac{(\lambda_{Sib} - 1)}{(\lambda_{OP} - 1)} = \frac{Cov(X,Sib)}{Cov(X,OP)}$$

reflects the relative importance of dominance effects.

Often $\frac{(\lambda_{Sib} - 1)}{(\lambda_{OP} - 1)} \approx 1$ (Table 1) and so dominance effects are considered to be negligible. This approximate equality also implies that common environmental effects between sibs is not different to that between parent and offspring, and, for many diseases, assuming common environmental effects are negligible seems plausible. Similarly, the ratio:

$$\frac{(\lambda_{MZ} - 1)}{(\lambda_{Sib} - 1)} = \frac{Cov(X,MZ)}{Cov(X,Sib)}$$

is expected to be 2 under a model that contains only additive genetic variance; if individual risk loci combined additively on the risk scale, then only additive variance would be observed. This ratio is often greater than 2 (Table 1), implying that epistatic genetic variance on the risk scale is not negligible.

Methods

Genetic model

We define K, as before, as the disease prevalence and g_x as the genetic risk (or probability) of disease of an individual given their multilocus genotype of x risk alleles out of a possible $2n$, where n is the number of loci that contribute to the genetic variance of the disease; by definition E(g) = K. For simplicity, we will assume that all risk alleles have equal frequency, p, and equal relative risks, τ, compared to the non-risk (wild type allele). We discuss the implications of these assumptions later. We assume that all loci are independent and that each locus is biallelic and is in Hardy-Weinberg equilibrium so that the frequency of wild type, carrier and homozygous risk genotypes in the population are $(1 - p)^2$, $2p(1 - p)$ and p^2 and x is distributed Binomial ($2n$,p), which approximates a normal distribution for n > ~5. We also assume random mating, no inbreeding and equal fertility of diseased and non-diseased individuals.

We consider three widely used genetic models of risk that are additive on some underlying scale. We assume that risk alleles act additively on the underlying scale both within a locus and between loci so that the critical contributor to genetic risk of disease is the number of risk alleles in an individual's multilocus genotype. We do not consider models that are additive on the risk scale as these were rejected by Risch [3] and confirmed in preliminary simulations as being unable to generate the patterns of recurrence risks to relatives observed for complex genetic diseases. After describing the disease risk models, we use numerical analysis and simulation to compare them. We compare the models to determine if they make the same predictions about observable recurrence risks and to investigate which model best fits the observed estimates.

Risch risk model

Additive on the log (risk) = log(g) scale: $\log(g_x) = \log(f_n) + x \log(\tau)$

Multiplicative on the risk (g) scale: $g_x = f_n \tau^x$

Under this model the relative risk of the risk allele compared to the other (wild-type) allele is τ, the homozygous risk genotype at each risk locus is τ^2 and the risks of the individual loci are multiplicative on the risk scale

$g_x = f_n \tau^x$, where f_n is the probability of disease in a person with only wild-type alleles at all n contributing loci and f_n can be expressed explicitly as $f_n = K/(1 + p(\tau - 1))^{2n}$ [10]. This model of disease risk was introduced by Risch [3,11] and is the model that we [10] and others [2,12,13] have used in the prediction of genetic risk to disease from multiple loci. The multiplicative Risch model is attractive because of its mathematical properties, but an undesirable feature (often not apparent in the mathematical expressions) is that there is no constraint placed on g_x, so that under some combinations of model parameters the probability of disease can have impossible values greater than 1 (that is, $g_x > 1$ for some x). This occurs when $x \geq -\ln(f_n)/\ln(\tau)$ (after solving $f_n \tau^x = 1$). We define the constrained Risch (CRisch) model to be the same as the Risch model except that g_x is truncated to 1 [13]. In this case, if K is considered known, f_n must be derived by numerically solving $K = E(g)$ for f_n assuming that n, p and τ are known.

Odds of risk model

Additive on the logit of risk scale: $\text{logit}(\text{risk}) = \log(g_x/(1 - g_x)) = \log(c_n K/(1 - K)) + x\log(\gamma)$

Multiplicative on the odds of risk scale: $\text{Odds} = g_x/(1 - g_x) = \gamma^x c_n K/(1 - K) = \gamma^x C_n$

and so $g_x = \gamma^x C_n/(1 - \gamma^x C_n)$

Under this model, $g_x/(1 - g_x)$ is the odds of disease given the multilocus genotype and $C_n = c_n K/(1 - K)$ is the odds of disease for an individual with all wild-type alleles at the n contributing loci, following Janssens *et al.* [4] and Lu and Elston [2]. The odds of disease without any information on multilocus genotype is $K/(1 - K)$. Under this model the relative odds of risk of carriers and the homozygous risk genotypes are γ and γ^2, where γ is the odds of the risk and where the γ are multiplicative on the odds of disease risk scale across loci. There is no explicit solution for $K = E(g_x)$ so that an explicit expression for c_n cannot be derived. For given input parameters c_n is derived by solving $K = E(g_x)$ numerically. Janssens *et al.* [4] used the approximation of $c_n = c_1$, but in preliminary studies we recognized that this approximation meant that the equality of $E(g_x)$ with the input (and key benchmark) parameter K was lost.

Probit of risk model or liability threshold model

Additive on an underlying liability scale: $u_x = (x - 2np)a$

Probit on the risk scale: $g_x = \Phi\left(\frac{u_x - t}{\sqrt{(1 - h_L^2)}}\right)$

Under this model we define a to be the effect of a risk allele on the underlying liability scale and u_x is the genetic value on the underlying scale of an individual with x risk alleles, distributed about a mean of zero (since the mean number of risk alleles is $2np$). Φ is the cumulative normal distribution function and t is a constant. The liability threshold model [14-16] assumes that liability to disease is normally distributed and that the presence of the disease arises if the liability exceeds a threshold, with the threshold positioned so that the proportion of the population that exceeds the threshold is equal to the population prevalence, K. The threshold, t, is derived from the inverse probability of the normal distribution, $t = \Phi^{-1}(1 - K)$, $\Phi(t) = 1 - K$; for example, if $K = 0.05$, $t = 1.645$. The model is parameterized in terms of variance components and heritability (h_L^2) on the underlying liability scale and can be scaled so that the phenotypic variance is 1. An individual's liability to disease is the sum of a genetic component (purely additive on this scale) distributed $N(0,h_L^2)$ and an environmental component distributed $N(0,1\text{-}h_L^2)$. The number (that is, n) and frequency (that is, p) of risk alleles determine the value of a:

$$a = \sqrt{\frac{h_L^2}{2np(1 - p)}}$$

Although this model is often referred to as the liability threshold model, we will use the name 'Probit model' so that all three models are named on the risk scale.

Relationship between relative risk (τ) and odds ratio (γ)

Under the Risch model, considering a single locus, the risk of the heterozygote is τ and the homozygote relative to the wild-type homozygote is τ^2. Under this model the heterozygous odds ratio is:

$$OR_{het} = \tau(1 - f_1)/(1 - \tau f_1)$$

Similarly, the homozygous odds ratio:

$$OR_{hom} = \tau^2(1 - f_1)/(1 - \tau^2 f_1)$$

Therefore, $OR_{hom} > OR_{het}^2$. In contrast, under the Odds model $OR_{het} = \gamma$, $OR_{hom} = \gamma^2$ and $OR_{hom}/OR_{het}^2 = 1$. For example, $K = 0.1$, $p = 0.1$, $\tau = 2$ under the Risch model, we can see that $OR_{het} = 2.49$ and $OR_{hom}/OR_{het}^2 = 1.13$, which shows the Risch and Odds models to be quite different. However, under parameters more relevant to human disease, for example, $K = 0.01$, $p = 0.1$, $\lambda = 1.05$, then $OR_{het} = 1.0506$ and $OR_{hom}/OR_{het}^2 = 1.00003$. Hence, odds risks and relative risks are often used interchangeably because, at the single locus level, they are equivalent for practical purposes. However, under a multi-locus model, the differences between the models compound. Establishing a mathematical relationship between the multi-locus models is not tractable. So we have investigated this relationship by simulation.

Comparison of models

One of the problems with comparing the models is to find a fair benchmark. We chose two parameters that are

directly measurable in real populations for benchmarking models: disease prevalence and the effect size of a single risk allele. To achieve this benchmarking, four input parameters were needed for the Probit model from which all other variables are derived: disease prevalence, number of risk loci, frequency of risk allele and heritability on the liability scale (that is, K, n, p and h_L^2). To benchmark our comparisons, we set τ, the effect size of a single risk allele, to be equal to g_{2np+1}/g_{2np} with g_{2np+1} and g_{2np} calculated from the Probit model. We use τ together with K, n and p as the input parameters for the Risch, CRisch and Odds models. Models are compared for the shape of the risk function, g_x and on the broad sense heritability on the risk scale:

$$H_{01}^2 = \frac{1}{K(1-K)}[E(g^2) - E(g))^2] \qquad \text{(Equation 2)}$$

where $E(g^2) = \sum_{x=0}^{2n} g_x^2 q_x$, and q_x is the probability of an individual carrying x risk alleles.

To compare models we have used results from GWAS to inform us of realistic values of τ. We use K = 0.1, 0.01, 0.001, to be representative of common, complex genetic diseases and we use K = 0.5 to benchmark comparison at the most extreme prevalence rate and maximum phenotypic variance ($K/(1 - K)$) on the risk scale. Since the number of loci underlying complex diseases is an unknown, we use n =100, 1,000, 10,000 since it is now considered unlikely that less than 100 loci will influence risk to common complex genetic diseases. We examined a range of n, p and h_L^2, but have limited the results reported to situations that generate $\tau < 2$. Although a few loci with $\tau > 2$ have been identified (for example, for the late age of onset disorder, age related macular degeneration [17]), GWAS results suggest that the average τ will be less than this [18]. From simulation of 10^6 families over three generations, we calculate λ_{MZ}, λ_{Sib}, λ_{OP} and the recurrence risk of disease in grandchildren of affected grandparents, λ_{OG}. From these we calculate H_{01}^2 (using equation 1) and $H_{01}^2 \approx 4(\lambda_{OG} - 1)K/(1 - K)$, which is an estimate of narrow sense heritability that is less contaminated by non-additive variance than the estimate $2(\lambda_{OP} - 1)K/(1 - K)$. More detailed descriptions of the simulations are provided in Additional file 1.

Results

Risch versus constrained Risch model

In the unconstrained Risch model we found that the occurrence of the impossible probabilities of disease ($g_x > 1$) had a significant impact on the results for some realistic combinations of parameters. For example, when n = 1,000, K = 0.1, p = 0.1, τ = 1.1, the mean number of risk alleles per person is 200 and $g_x > 1$ when $x > 232$, which occurs with frequency 0.009. Despite the low frequency of occurrence, these extreme risks contribute disproportionately to the genetic variance and heritability. In this example, the heritability (calculated using equation 2) is 0.51, but falls to only 0.17 when these impossible risks are truncated to 1.

Combined effect of *n*, *p* and τ

Results for a representative combination of parameters (n = 100, 1,000, 10,000, K = 0.1, 0.01, 0.001, p = 0.1, 0.3 and h_L^2 = 0.5, 0.7; Additional file 2) show that although the broad sense heritability on the observed (that is, H_{01}^2; Equation 2) scale differs markedly between the Probit, CRisch and Odds models, there is little dependence on n, p and τ provided h_L^2 is held constant. This is because, for a given h_L^2, the parameters n and p control the variance contributed by each locus, so that when n is small, the effect size of each locus τ is necessarily high. These results imply that the key parameter in determining heritability on the risk scale is the total genetic variance rather than the variance at each locus. Consequently, the results are presented in terms of h_L^2 (see 'Comparison of models' section above) because this allows translation to multiple combinations of n, p and τ.

Shape of risk function and heritabilities on the risk scale

In Figure 1 we illustrate risk functions for combinations of parameters relevant to human complex genetic diseases. The x-axis is the number of risk alleles harbored by individuals in a population; theoretically, this can be between 0 and $2n$, but in practice the number of risk alleles takes on the range $2np \pm 4\sqrt{2np(1 - p)}$, that is, 4 standard deviations about the mean. The number of risk alleles has an approximate normal distribution since the binomial distribution with large n tends to normality. In Figure 1, the black dotted line represents the proportion of individuals with x or more risk alleles. The 'S'-shaped curves are the risks or probability of disease given the number of risk loci, rising from g_x = 0 to g_x = 1. The positioning of this rise along the x-axis reflects the disease prevalence (that is, K) showing that, for low prevalence diseases, a greater number of risk alleles relative to the population mean is required for disease. The steepness reflects the broad sense heritabilities on the risk scale (that is, H_{01}^2) so that a steeper rise reflects a higher correlation between genotype and phenotype. Of these examples, only when h_L^2 = 0.2 and K = 0.001 (Figure 1b) was there no need to constrain the Risch risk model as g_x never reaches 1 even for the maximum values of x found in the population.

The relationship between H_{01}^2 and τ or h_L^2 is illustrated in Figure 2 and depends on both disease prevalence and model. Apparently small differences in the risk functions can have a big impact on the H_{01}^2. For the Probit model

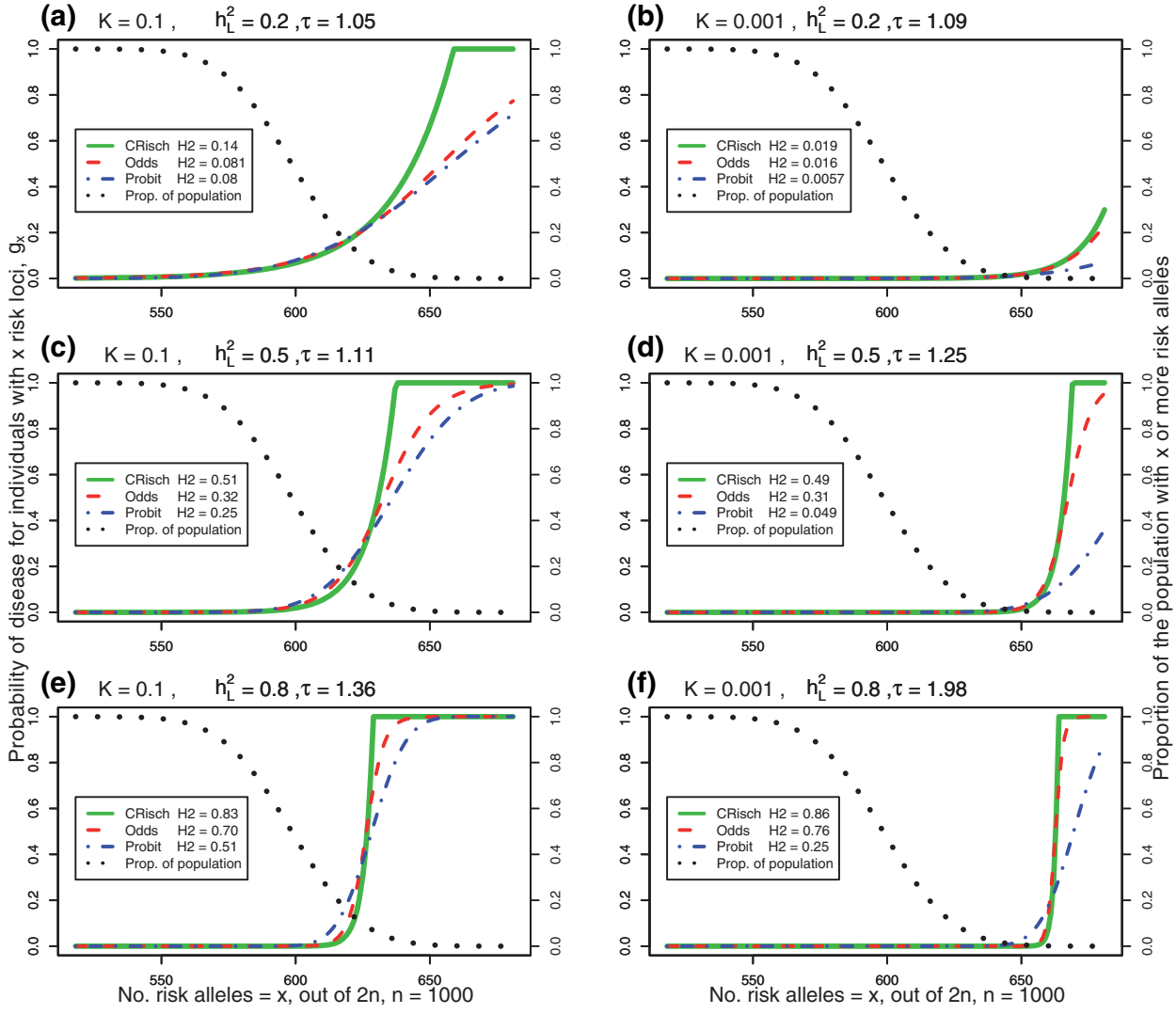

Figure 1. Risk functions for the CRisch, Odds and Probit models using parameters relevant to human complex genetic diseases. (a-f) Risk or probability (g_x) of disease for an individual with x out of $2n$ risk alleles where the number of risk loci, n = 1,000 and the frequency of each risk allele, p = 0.3. The black dotted lines represent the proportion of individuals in the population who have x or more risk alleles. The parameters n, p, heritability on the underlying liability scale, h^2_L, and disease prevalence, K, determine the relative risk of a single locus, τ. The legend lists the resulting broad sense heritability on the risk scale, H^2_{01} (H2 in the legend). The shape of the risk functions is achieved with other combinations of n and p for the same K and h^2_L.

H^2_{01} is a function of K, whereas for the CRisch and Odds models the dependence on K is of much less importance. This reflects the choice of benchmarking between the models. In the Probit model, the ratio g_{x+1}/g_X decreases as x (number of risk alleles) increases, whereas in the CRisch model this ratio is constant until the limit on probability of disease is reached. Therefore, the probability of disease rises more steeply with number of risk alleles for the CRisch model than the Probit model and this is more pronounced for rarer diseases when the difference between g_{x+1}/g_X at the average x and a high x is greater for the Probit model; the Odds model is intermediate.

Figure 3 presents the estimates of $\lambda_{MZ}/\lambda^2_{Sib}$ across the full range of h^2_L and for different prevalences. Risch [3] predicted this relationship to be 1 under a multiplicative model. However, this relationship only holds when K = 0.5, or as $h^2_L \to 0$ but becomes <<1 as K decreases and $h^2_L \to 1$, a consequence of the need to constrain the probability of disease for an individual (g_x) to a maximum value of 1. Values of λ_{MZ} and λ_{Sib} and the ratio $\lambda_{MZ}/\lambda^2_{Sib}$ are presented for a range of scenarios (Table 2) to allow comparison with diseases listed in Table 1.

The relationship between h^2_{01} and H^2_{01} is almost the same for all models (Figure 4), confirming the similarity

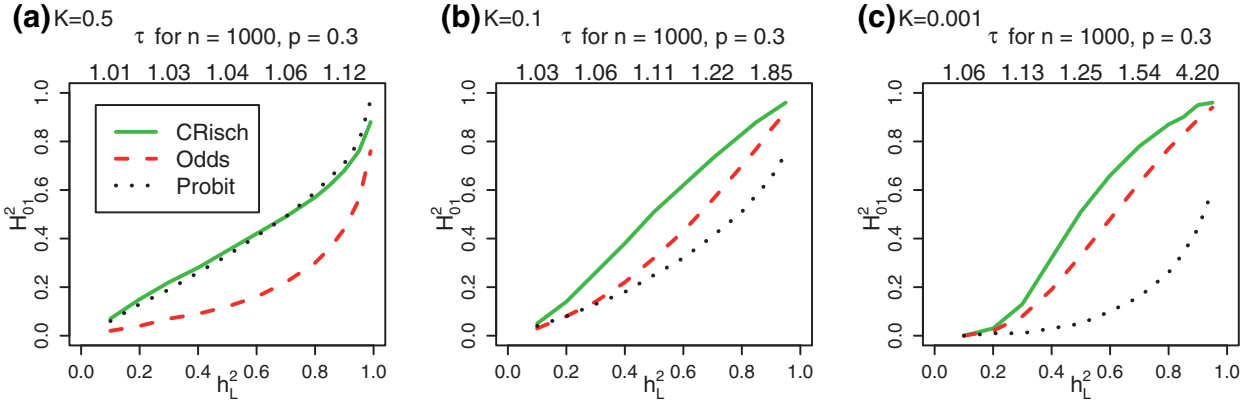

Figure 2. Relationship between H^2_{01} for the CRisch, Odds and Probit models and h^2_L, heritability on the underlying liability scale. (a-c) For each h^2_L, τ is estimated from the Probit model simulation and used as an input for the other models, so that all three models are benchmarked by K and τ. The shape of the relationship is not dependent on the choice of n and p; the τ when h^2_L = 0.1, 0.3, 0.5, 0.7 and 0.9 are listed above each graph when n = 1,000 and p = 0.3. From simulations of a single population of 10^6 individuals.

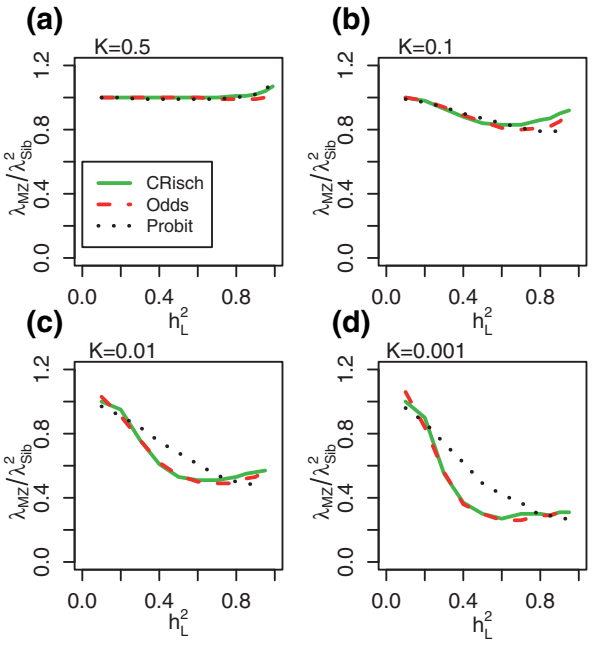

Figure 3. Relationship between $\lambda_{MZ}/\lambda^2_{Sib}$ and h^2_L for the CRisch, Odds and Probit models. (a-d) Relationship for different disease prevalences (K).

of the models on the risk scale. The maximum value of h^2_{01} is 0.64, which occurs as $H^2_{01} \to 1$ when K = 0.5 as derived by Robertson (Appendix of Dempster and Lerner [14]). As K decreases or h^2_L increases the proportion of H^2_{01} that is additive declines so that, for diseases of prevalence ≤ 0.01 almost all of the heritability on the risk scale is explained by epistatic variance (as shown by the steep increase in the risk function [14]).

Distinguishing between models based on risk to relatives

Although we assume that each risk locus has the same individual effect size, the models differ in the way that the effect sizes combine. In the CRisch model each additional risk allele multiplies probability of disease by the same amount until the number of risk alleles harbored reaches the limit of disease being certain, g_x = 1. In contrast, the Odds and Probit models have 'built-in' constraints so that $g_x \leq 1$, which means that each additional risk allele contributes proportionally less to the probability of disease. This effect can be seen in Figure 1 where the risk function is steepest for the CRisch model and least steep for the Probit model with the Odds model usually in between the other two. The steeper the risk function the higher the broad sense heritability H^2_{01}, so this is usually highest for the CRisch model and least for the Probit model. This effect of the risk function on heritability on the risk scale also applies to the narrow sense heritability, h^2_{01}, so the relationship between the two remains constant (Figure 4). The similarity of the models on the risk scale is not perfect as shown by differences in $\lambda_{MZ}/\lambda^2_{Sib}$ in Figure 3. However, if this ratio is graphed against a function of observable parameters, such as H^2_{01} instead of h^2_L, the differences between models are small (Additional file 3) and could not be demonstrated in practice given the sampling errors of the parameters. Thus, the three models could not be distinguished using only traditional data, that is, recurrence risk of relatives.

Distinguishing between models based on relative risks of individual loci, τ

If we identify one or more loci affecting a disease, we can directly observe the risk in people carrying different numbers of risk alleles and compare this with the model

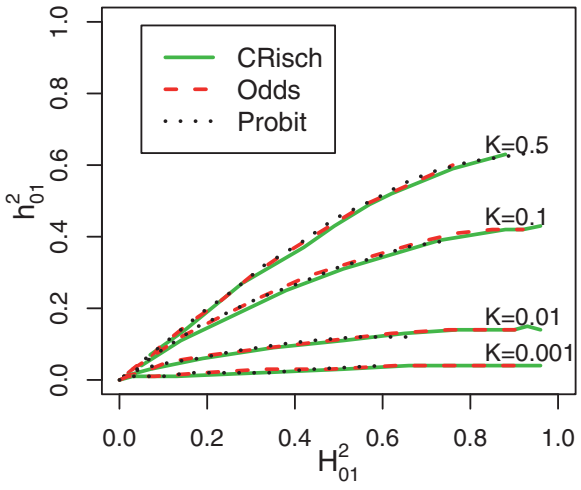

Figure 4. Relationship between narrow sense (additive) h^2_{01} and broad sense heritability H^2_{01} on the risk scale for different disease prevalences (*K*). From simulations of a single population of 10^6 individuals, with h^2_{01} calculated as $4(\lambda_{OG} - 1)K/(1 - K)$ where λ_{OG} is the recurrence risk of disease in grandchildren of affected grandparents and H^2_{01} calculated from Equation 2.

predictions. The numerical example in the 'Relationship between τ and γ' section shows that, for a single locus, the models do make different predictions when τ values are large but not when they are small, as is expected to be the usual case. However, even for small τ values the models differ when all risk loci are included. To obtain the same heritability on the risk scale, the models required different effect sizes (τ) of associated variants (Figure 2). Similarly, by comparing Tables 1 and 2, we can see that combinations of observed λ_{MZ} and λ_{Sib} correspond to a much lower τ, which translates to a lower heritability on the liability scale under the CRisch or Odds model compared to the Probit model. For example, for a disease with prevalence $K = 0.01$, $\lambda_{MZ} = 52$, $\lambda_{Sib} = 10$ (parameters representative of schizophrenia), the τ for $n = 1{,}000$ loci each with risk allele frequency $p = 0.3$ were 1.19, 1.26 and 1.41 for the CRisch, Odds and Probit models, respectively. However, only if it is possible to identify the majority of the risk variants will it be possible to differentiate between the models in practice.

Another way to look at this difference between the models is that, for a given value of λ_{MZ} (or λ_{Sib}) and τ and p, a higher value of n is required for the Probit model than for the CRisch model. This means that a given risk locus with observed τ and p explains a smaller proportion of the risk to relatives under a Probit model than under a CRisch model. Or equally, it means that the CRisch models generate higher risks to relatives in our benchmarked comparisons - for example, when $K = 0.01$, $n = 1{,}000$, $p = 0.3$, $\tau = 1.2$ and $h^2_L = 0.5$, λ_{MZ} for the CRisch, Odds and Probit models were 52, 35 and 13, respectively; the λ_{Sib} for the same models were 10, 8 and 4, respectively. If risk loci are identified that account for a significant proportion of the sibling risk, then it may be possible to test which model better fits observed data, but this will require a large number of families to be genotyped for the risk loci.

Discussion

With the advent of GWAS we are gaining a clearer understanding of the genetic architecture of common complex diseases. Empirical evidence suggests an architecture of many genetic loci with many variants of small effect. Interest in genomic profiling, the use of a genome-wide markers to predict genetic disease risk, is growing (for example, [19,20]), as is the establishment of companies offering profiling services. The prediction of disease risk from many risk loci or markers requires a model that combines the effects of these loci and the choice of this model is the topic of this paper.

Total variance of risk loci is the driving force

We chose two parameters that are directly measurable in real populations for benchmarking models: disease prevalence (that is, K) and the effect size of a single risk allele (that is, τ). We recognized that many combinations of the number of loci (that is, n) allele frequency (that is, p) and τ were consistent with the same heritability on the underlying scale in the Probit model (that is, h^2_L) and that the predictions of all the models were insensitive to the exact combination of n, p and τ provided h^2_L was held constant. Therefore, we have compared the models while holding constant K and h^2_L. In Figures 1 and 2 we present results for $n = 1{,}000$ and $p = 0.3$, to provide some comparison to empirical estimates of τ. Since the distribution of genetic risk of disease in a population is driven by total genetic variance rather than the variance contributed by each locus, it is unlikely that relaxing the restriction of equal allele frequencies and effect sizes will impact the results; this is consistent with the results of other studies [4,10,21].

Although we show that the unconstrained Risch model is not a practical model, its mathematical tractability can still provide valuable insight into our understanding of the factors influencing genetic risk. We show (Additional file 4) that the scaled contribution to the genetic variance on the risk scale by each risk allele (v) is a function of p and τ, $v = p(1 - p)(\tau - 1)^2/[1 + p(\tau - 1)]^2$ and the total genetic variance on this scale is proportional to nv. For small values of τ (that is, $\tau \to 1$), $nv \approx np(1 - p)(\tau - 1)^2$, which can be used to derive the proportion of genetic variance explained by one locus.

Rejection of simple additive and simple multiplicative models on the risk scale

Risch [3], using schizophrenia as an example, was the first to show that recurrence risk to relatives in complex

Table 2. Relative risks to relatives of affected individuals calculated within the stochastic simulation for Probit, CRisch and Odds models

		Probit			CRisch			Odds		
K	h_L^2	λ_{MZ}	λ_{Sib}	$\frac{\lambda_{MZ}}{\lambda_{Sib}^2}$	λ_{MZ}	λ_{Sib}	$\frac{\lambda_{MZ}}{\lambda_{Sib}^2}$	λ_{MZ}	λ_{Sib}	$\frac{\lambda_{MZ}}{\lambda_{Sib}^2}$
0.1	0.1	1.3	1.2	0.99	1.4	1.2	1.00	1.3	1.1	1.00
0.1	0.5	3.2	1.9	0.87	5.6	2.6	0.84	3.9	2.1	0.85
0.1	0.7	4.7	2.4	0.81	7.6	3.0	0.83	6.0	2.8	0.80
0.1	0.95	7.8	3.1	0.82	9.7	3.2	0.92	9.3	3.2	0.90
0.01	0.1	1.9	1.4	0.97	2.4	1.5	1.00	1.7	1.3	1.03
0.01	0.5	13.0	4.4	0.68	51.7	9.9	0.53	34.8	8.1	0.54
0.01	0.7	26.6	7.0	0.54	76.8	12.3	0.51	62.3	11.3	0.49
0.01	0.95	67.3	11.7	0.49	97.0	13.0	0.57	94.6	12.9	0.57
0.001	0.1	2.8	1.7	0.96	4.0	2.0	1.00	1.2	1.1	1.06
0.001	0.5	54.8	10.5	0.49	516.5	41.6	0.30	342.5	34.0	0.30
0.001	0.7	157.8	20.6	0.37	796.8	51.4	0.30	638.5	49.5	0.26
0.001	0.95	599.8	47.5	0.27	989.9	57.6	0.30	968.6	55.9	0.31

h_L^2 is an input parameter for the Probit model. For each h_L^2 τ is estimated from the Probit model simulation and used as input to the CRisch and Odds model simulations. h_L^2 is used as the benchmark as τ is dependent on n, p and K.

diseases is better explained by a multiplicative than an additive model of gene action on the risk scale because $(\lambda_{MZ} - 1)/(\lambda_{sib} - 1) > 2$ as shown in Table 1. In preliminary simulations (not reported) we confirmed that additivity on the risk scale of all risk loci simply could not produce the steep rise in probability of disease (Figure 1) necessary to achieve the disease prevalences and recurrence risks to relatives typical of complex diseases. In contrast, Slatkin [13], under his thesis of exchangeable models, demonstrated that an additive model on the risk scale could explain complex disease. However, to achieve the steep rise in disease risk, he imposed stringent constraints, so that the additive effect of risk alleles only occurred in the (very narrow) range of the number of risk alleles associated with the steep rise in probability of disease. Outside this range probability of disease was either zero or 1. In this way, the shape of the risk function is similar to the models that are multiplicative on the risk scale.

Other theoretical studies have used the Risch model [2,13], the CRisch model [13], the Odds model [4] and the Probit model [22]. Although there is a generally accepted dogma that these models are similar, in trying to compare studies it is important to know if any differences are a function of the choice of risk model. In a previous study [10] we made derivations under the Risch model and for the parameter combinations considered the probability of disease being greater than 1 was rare. However, in this study, where we have considered the full range of parameters, we have recognized that under the unconstrained Risch model, individuals for whom probability of disease is greater than 1 ($g_x > 1$) make a huge contribution to the genetic variances.

Risch [3] investigating schizophrenia and Brown *et al*. [6] studying ankylosing spondilitis recognized that the observed ratio $\lambda_{MZ}/\lambda_{Sib}^2$ was less than one, whereas this ratio is expected to be 1 under the Risch model [3]. The sampling variance on estimates of recurrence rates is high and so the greater consistency with multiplicative rather than additive models (risk scale) was their main conclusion. However, by looking at a range of complex diseases (Table 1) there is consistent evidence that $\lambda_{MZ}/\lambda_{Sib}^2$ is less than 1, particularly for low prevalence diseases. These observed ratios are consistent with our simulation results, which show that under the CRisch, Odds and Probit models, the ratio $\lambda_{MZ}/\lambda_{Sib}^2 \to 1$ only as $K \to 0.5$ and $h_L^2 \to 0$, but under parameters typical of common complex genetic diseases $\lambda_{MZ}/\lambda_{Sib}^2 << 1$, particularly as $K \to 0$ and $h_L^2 \to 1$. The mathematical tractability of the Risch model has often made it the method of choice in theoretical studies and the equality $\lambda_{MZ}/\lambda_{Sib}^2 = 1$ has been used to underpin predictions (for example, see the Supplement of Clayton [23]); in the mathematical expressions the impact of not constraining the probability of disease to be less than 1 is not obvious, but it is because of this important constraint that equality $\lambda_{MZ}/\lambda_{Sib}^2$ is often much less than 1.

Therefore, we conclude that the unconstrained Risch model is simply not realistic, particularly for parameters typical of human complex disease (K < 0.1 and $h_L^2 > 0.5$),

so here we have made comparisons on the more realistic constrained (CRisch) model.

Differences between the models unlikely to be detectable in practice

Since we reject the additive and Risch models, we concentrate on the comparison of the CRisch, Odds and Probit models. We chose to compare models with two fixed benchmarks, disease prevalence and effect size of an individual risk allele, taken at the average number of risk alleles (that is, τ). Under this benchmarking, the probability of disease associated with carrying the minimum number of alleles in the population differs between models, but in all models this will be very close to zero given the number or risk loci now expected to contribute to complex genetic disease. Although we assume that each risk locus has the same individual effect size, the models differ in the way that the effect sizes combine. For example, a given risk locus with observed τ and p explains a smaller proportion of the risk to relatives under a Probit model than under a CRisch model. However, we conclude that for all operational purposes, in the foreseeable future, it is unlikely that we will be able to distinguish between the models either on the basis of recurrence risks to relatives or on the basis of estimates of effect sizes of risk loci. Slatkin [13] also compared the CRisch and Probit models and benchmarked on a range of parameters. Our results are complementary to, and consistent with, his, although direct comparison is prevented by his models distinguishing between heterozygotes and homozygotes at each locus, so that the multiplicativity of risk alleles was only between loci and not within loci. Inability to distinguish between multi-locus risk models on the basis of recurrence risks is perhaps not surprising given that Smith [24] was unable to distinguish between more extreme models on this basis. Ability to distinguish between the models is only possible in the very tail of the risk curve and would only be achievable if genomic profiles could be constructed using measured variants that accounted for the totality of the genetic variance. If this were possible, sets of individuals could be identified with high predicted risk and the proportion succumbing to disease could be measured and compared to the proportion expected under different models. Such hypothetical scenarios at present seem unattainable.

Each individual carries a unique portfolio of risk loci

From Figure 1 it becomes clear that when there are many risk loci contributing to disease each of small effect, that all individuals in the population necessarily carry a large number of risk alleles. For example, when 1,000 loci with risk alleles of frequency 0.1 underlie a complex disease, all individuals in the population carry at least 150 risk alleles, an average individual carries 200 risk alleles and, when disease prevalence is low and heritability is high, most of those with disease carry 230 to 250 risk alleles. Since, in this example, there is a total of 2,000 risk alleles, each individual will carry their own unique portfolio, which could underlie the phenotypic heterogeneity typical of many complex diseases.

Large amounts of epistasis on the risk scale despite additivity on underlying scales

Our results show that additivity of individual genetic variants on some underlying scale can convert to, sometimes considerable, non-additive genetic variance on the risk scale, particularly when the disease prevalence is low. These results are not new and were presented by Dempster and Lerner [14], but are sometimes overlooked. Human diseases usually have prevalences of less than 0.1, in which case the majority of the genetic variance on the risk scale is epistatic. These results imply that the models underpinning GWAS already account for one type of gene-gene interaction, if each τ could be estimated without error. Likewise, our usual models also imply genotype-environment interaction on the risk scale because the effect of an environmental factor is greater in people with higher genetic risk. Our definition of epistasis is one of statistical interaction; the extent to which statistical interaction relates to biological or functional interaction has been much debated (see [25] for a review) and will not become clear until more of the genetic variance can be explained by identified genomic variants.

True versus estimated τ

We set out to benchmark models on the basis of two observable parameters, disease prevalence (that is, K) and the effect size of a single risk allele (that is, τ). In building the models we have assumed that the true τ is known and have defined it as the effect of a single risk locus in the background of the average number of risk loci. However, the estimates of τ made from experimental data may be quite different to these true values. If the genotypes at all risk loci were known and a complete model was fitted to the data, then the correct estimate of τ would be obtained (within experimental sampling error). In practice, however, usually only the effect of a single risk locus is included in the statistical model and under these circumstances we will estimate the effect of an extra risk allele averaged across all background genotypes rather than the effect at the mean background genotype. The effect of this may be dependent on the true way in which loci combine to influence risk of disease, which, of course, is unknown. Under the CRisch model of Figure 1a, all individuals with >650 risk alleles get the disease, so above 650 risk alleles there is no effect of an

extra risk allele. Conversely, below 650 risk alleles each extra risk allele increases the probability of disease by τ. The experimental estimate will be a weighted average of these two estimates (zero and τ). In practice, therefore, variants detected with small relative risk may reflect greater biological importance than might otherwise be inferred. Under the Probit model the τ calculated at the average number or risk loci is:

$$\Phi\left(\frac{a-t}{\sqrt{(1-h_L^2)}}\right) \Big/ \Phi\left(\frac{-t}{\sqrt{(1-h_L^2)}}\right)$$

whereas the τ estimated when a single risk locus is in the statistical model is $\Phi(a - t)/\Phi(-t)$ because then all other risk loci are part of the residual variance in liability and so the residual variance approaches the phenotypic variance, which is 1.0.

Comparison of the models in practice is difficult and distinguishing between them may be impossible, especially if the true n is large and the true τ is small. Since we have demonstrated that the models are difficult to differentiate, the use of the Probit model, which has mathematical tractability and a known relationship between the estimates of τ in different genetic backgrounds, is likely to be the model of choice. The estimated variance on the liability scale explained by a locus with estimated effect size $\hat{\tau}$ is $2p(1-p)(\hat{\tau}-1)^2/i^2$ [26], so that the estimated effect on the liability scale is $\hat{a}(\hat{\tau}-1)/i$, where i is the mean liability of the diseased group, $i = z/K$, where z is the height of the normal curve at the threshold t.

Limitations

The true genetic architecture (in terms of number, frequency and effect size of risk variants and the way in which they combine) is unknown and may be quite different for the different diseases listed in Table 1. For simplicity, we have described disease in terms of affected/unaffected, ignoring time-dependent onset, and we have ignored phenotypic heterogeneity (which may reflect genetic heterogeneity) in the definition of disease status and other real-life complications. In principle, our approach could reflect any definition of disease if the genetic epidemiology and genetic risk variants can be defined - for example, early and late onset disease may be considered as different diseases - but despite this any simple model is likely to be a poor representation of disease. None of the models we have considered are likely to be the true model, but since they can all generate recurrence risks consistent with complex genetic diseases (given the right combination of parameters), they can give useful insight until empirical data provide evidence for them to be rejected. These simple models provide some boundaries, demonstrating some properties that must be upheld by the true genetic architecture in order to be consistent with observed data.

Conclusions

In this paper we set out to compare different models that combine the effects of multiple risk loci into an overall genetic risk. We conclude that a model that is additive or multiplicative on the risk scale across all loci is incompatible with the observed recurrence risks to relatives. The constrained multiplicative (CRisch), Odds and Probit models are all compatible with the observed data and, in fact, it is difficult to distinguish between them when the relative risk at an individual locus is small. Importantly, we show that the unconstrained multiplicative (Risch) model, often used in theoretical studies because of its mathematical tractability, is not a realistic model as impossible probabilities of disease are implied. Specifically, the multiplicative Risch model generates a relationship of $\lambda_{MZ}/\lambda_{Sib}^2 = 1$, but we have demonstrated that this not possible under many disease scenarios and occurs in the theoretical derivation because probabilities of disease are not constrained and can exceed 1. We have demonstrated that under more realistic models in which probabilities of disease are constrained to 1, the ratio $\lambda_{MZ}/\lambda_{Sib}^2$ is often much less than 1, a result that is consistent with empirical estimates from a range of diseases. Finally, we conclude that it will only be possible to distinguish between the CRisch, Odds and Probit models in practice if genetic risk profiles are able to reconstruct the majority of the known genetic variance; this is unlikely for the foreseeable future.

Additional file 1. A detailed description of simulations.

Additional file 2. A table showing broad sense heritabilities on the disease risk scale. A table showing broad sense heritabilities on the disease risk scale, H_{01}^2 (Equation 2), for different combinations of disease prevalence, K, number of risk loci, n, risk allele frequency, p, heritability on the liability scale, H_L^2, and risk of a single risk allele compared to the non-risk allele, τ.

Additional file 3. A figure showing the relationship between $\lambda_{MZ}/\lambda_{Sib}^2$ and $H_{01}^2 = [(\lambda_{MZ} - 1)K]/[(1 - K)]$ for the CRisch, Odds and Probit models and different disease prevalences (K).

Additional file 4. A PDF document providing variance components on the risk scale using the unconstrained risk model. Uses the mathematical tractability of the unconstrained Risch model to examine the contribution of each risk allele to genetic variance on the risk scale.

Abbreviations

CRisch, constrained Risch; GWAS, genome-wide association study; MZ, monozygotic. γ, odds of disease for risk allele compared to wild-type allele; λ_{MZ}, recurrence risk of disease in monozygotic twins of diseased individuals; λ_{OG}, recurrence risk of disease in grandoffspring of diseased grandparents; λ_{OP}, recurrence risk of disease in offspring of diseased parents; λ_R, recurrence

risk of disease in relatives of diseased individuals for relatives of type R; λ_{Sib}, recurrence risk of disease in sibs of diseased individuals; τ, the risk (probability) of disease of a risk allele relative to the other (wild-type) allele for a single locus (for the unconstrained Risch model $\tau = g_{x-1}/g_x$ for all $x = 0, 2n - 1$); a, additive effect size of each risk allele on the liability scale in Normal standard deviation units; f_n, probability of disease in a person with wild-type alleles only at all n contributing loci; g_x, the genetic risk (or probability) of disease of an individual given their multilocus genotype of x risk alleles; h^2_{01}, narrow sense (that is, additive genetic) heritability on the risk scale; h^2_L, heritability on the liability scale, on this scale all genetic variance is additive; H^2_{01}, broad sense (that is, total genetic) heritability on the risk scale - on this scale the phenotype, disease, is either not diseased (0) or diseased (1); K, disease prevalence in a population; K_R, disease prevalence in relatives of diseased individuals for relatives of type R; n, the number of loci that contribute to the genetic variance of the disease; p, frequency of risk allele; t, threshold truncating proportion K in the right-hand tail of the normal distribution; x, number of risk alleles harbored by an individual, between 0 and $2n$.

Author details
[1]Genetic Epidemiology and, Queensland Institute of Medical Research, Herston Road, Brisbane, Queensland 4006, Australia
[2]Faculty of Land and Food Resources, University of Melbourne, Royal Parade, 3010 and Department of Primary Industries, Research Avenue, 3086, Melbourne, Victoria, Australia

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
NRW and MEG together devised the study, interpreted the results and wrote the manuscript. NRW conducted all simulations. MEG derived the 'Variance components on the risk scale using the unconstrained risk model' in the Additional files. Both authors read and approved the final manuscript.

Acknowledgements
This work was supported by the Australian National Health and Medical Research Council (grant 496688) and by the Australian Research Council (grant DP0770096). We would like to thank Bill Hill and Peter Visscher for their helpful comments on earlier versions of this manuscript.

Submitted: 13 September 2009 Revised: 22 January 2010
Accepted: 2 February 2010 Published: 2 February 2010

References

1. WTCCC: **Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.** *Nature* 2007, **447:**661-678.
2. Lu Q, Elston RC: **Using the optimal receiver operating characteristic curve to design a predictive genetic test, exemplified with type 2 diabetes.** *Am J Hum Genet* 2008, **82:**641-651.
3. Risch N: **Linkage strategies for genetically complex traits. I. Multilocus models.** *Am J Hum Genet* 1990, **46:**222-228.
4. Janssens AC, Aulchenko YS, Elefante S, Borsboom GJ, Steyerberg EW, van Duijn CM: **Predictive testing for complex diseases using multiple genes: fact or fiction?** *Genet Med* 2006, **8:**395-400.
5. Craddock N, Khodel V, Van Eerdewegh P, Reich T: **Mathematical limits of multilocus models: the genetic transmission of bipolar disorder.** *Am J Hum Genet* 1995, **57:**690-702.
6. Brown MA, Laval SH, Brophy S, Calin A: **Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis.** *Ann Rheum Dis* 2000, **59:**883-886.
7. James JW: **Frequency in relatives for an all-or-none trait.** *Ann Hum Genet* 1971, **35:**47-49.
8. Kempthorne O: **The correlation between relatives in a random mating population.** *Proc R Soc Lond B Biol Sci* 1954, **143:**102-113.
9. Lynch M, Walsh B: *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates; 1998.
10. Wray NR, Goddard ME, Visscher PM: **Prediction of individual genetic risk to disease from genome-wide association studies.** *Genome Res* 2007, **17:**1520-1528.
11. Risch N, Merikangas K: **The future of genetic studies of complex human diseases.** *Science* 1996, **273:**1516-1517.
12. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA: **Polygenic susceptibility to breast cancer and implications for prevention.** *Nat Genet* 2002, **31:**33-36.
13. Slatkin M: **Exchangeable models of complex inherited diseases.** *Genetics* 2008, **179:**2253-2261.
14. Dempster ER, Lerner IM: **Heritability of threshold characters.** *Genetics* 1950, **35:**212-236.
15. Crittenden LB: **An interpretation of familial aggregation based on multiple genetic and environmental factors.** *Ann N Y Acad Sci* 1961, **91:**769-780.
16. Falconer DS: **The inheritance of liability to certain diseases, estimated from the incidence among relatives.** *Ann Hum Genet* 1965, **29:**51-71.
17. Swaroop A, Branham KEH, Chen W, Abecasis G: **Genetic susceptibility to age-related macular degeneration: a paradigm for dissecting complex disease traits.** *Hum Mol Genet* 2007, **16:**R174-R182.
18. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.** *Proc Natl Acad Sci U S A* 2009, **106:**9362-9367.
19. Jakobsdottir J, Gorin MB, Conley YP, Ferrell RE, Weeks DE: **Interpretation of genetic association studies: markers with replicated highly significant odds ratios may be poor classifiers.** *PLoS Genet* 2009, **5:**e1000337.
20. Evans DM, Visscher PM, Wray NR: **Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk.** *Hum Mol Genet* 2009, **24:**24.
21. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P: **Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.** *Nature* 2009, **460:**748-752.
22. Daetwyler HD, Villanueva B, Woolliams JA: **Accuracy of predicting the genetic risk of disease using a genome-wide approach.** *PLoS ONE* 2008, **3:**e3395.
23. Clayton DG: **Prediction and interaction in complex disease genetics: experience in type 1 diabetes.** *PLoS Genet* 2009, **5:**e1000540.
24. Smith C: **Recurrence risks for multifactorial inheritance.** *Am J Hum Genet* 1971, **23:**578-588.
25. Cordell HJ: **Detecting gene-gene interactions that underlie human diseases.** *Nat Rev Genet* 2009, **10:**392-404.
26. Falconer D, Mackay T: *Introduction to Quantitative Genetics*. Fourth edn. England: Longman; 1996.
27. Mosing MA, Gordon SD, Medland SE, Statham DJ, Nelson EC, Heath AC, Martin NG, Wray NR: **Genetic and environmental influences on the comorbidity between depression, panic disorder, agoraphobia and social phobia: A twin study.** *Depression Anxiety* 2009, **26:**1004-1011.
28. Seddon JM, Cote J, Page WF, Aggen SH, Neale MC: **The US twin study of age-related macular degeneration - Relative roles of genetic and einivironmental influences.** *Arch Ophthalmol* 2005, **123:**321-327.
29. Scholl HPN, Fleckenstein M, Issa PC, Keilhauer C, Holz FG, Weber BHF: **An update on the genetics of age-related macular degeneration.** *Mol Vision* 2007, **13:**196-205.
30. Zdravkovic S, Wienke A, Pedersen NL, de Faire U: **Genetic susceptibility of myocardial infarction.** *Twin Res Hum Genet* 2007, **10:**848-852.
31. Risch N: **The genetic epidemiology of cancer: interpreting family and twin studies and their implications for molecular genetic approaches.** *Cancer Epidemiol Biomarkers Prev* 2001, **10:**733-741.
32. Das SK, Elbein SC: **The genetic basis of type 2 diabetes.** *Cellscience* 2006, **2:**100-131.
33. Nieminen MM, Kaprio J, Koskenvuo M: **A population-based study of bronchial asthma in adult twin pairs.** *Chest* 1991, **100:**70-75.
34. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, Silman AJ: **Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins.** *Arthritis Rheum* 2000, **43:**30-37.
35. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J: **Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs - A nationwide follow-up study.** *Diabetes* 2003, **52:**1052-1055.
36. Sadovnick AD, Dyment D, Ebers GC: **Genetic epidemiology of multiple sclerosis.** *Epidemiol Rev* 1997, **19:**99-106.
37. Halfvarson J, Bodin L, Tysk C, Lindberg E, Jarnerot G: **Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics.** *Gastroenterology* 2003, **124:**1767-1773.
38. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, Pons-Estel BA, Gladel: **Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in**

1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum* 2005, **52:**1138-1147.
39. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD: **Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.** *Nat Genet* 2008, **40:**204-210.
40. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Royburman P, Walker A, Mack TM: **A revised estimate of twin concordance in systemic lupus-erythematosus.** *Arthritis Rheum* 1992, **35:**311-318.
41. Reich T, James JW, Morris CA: **The use of multiple thresholds in determining the mode of transmission of semi-continuous traits.** *Ann Hum Genet* 1972, **36:**163-184.
42. Yang J, Visscher PM, Wray NR: **Sporadic cases are the norm for common disease.** *Eur J Hum Genet* 2009, 2009, Oct 14 [Epub ahead of print].

doi:10.1186/gm131
Cite this article as: Wray NR, Goddard ME: **Multi-locus models of genetic risk of disease.** *Genome Medicine* 2010, **2:**10.