Please cite this article in press as: Slatkin, Genotype-Specific Recurrence Risks as Indicators of the Genetic Architecture of Complex Diseases, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.06.011

REPORT

Genotype-Specific Recurrence Risks as Indicators of the Genetic Architecture of Complex Diseases

Montgomery Slatkin[1,*]

A statistic is introduced that relates discoveries made in genome-wide association (GWA) studies to patterns of disease risks among relatives. The genotype-specific recurrence risk (GSR) is the genotype-specific risk to relatives of known relationship to affected probands. The GSRs can be used for three purposes. (1) They can provide an independent test of whether an allele identified in a GWA study is associated with the disease. (2) They can provide a test of whether interactions among loci affecting the disease are multiplicative. (3) They can be used by genetic counselors to incorporate information from GWA studies for predicting the risk to relatives of known genotype. Under a multiplicative model of disease causation, the GSRs for a locus are the genotypic risks in probands for that locus multiplied by λ_R/λ_{jR}, where λ_R is Risch's recurrence risk ratio and λ_{jR} is the contribution to λ_R from the locus of interest. If there is saturation of risk with increasing numbers of causative alleles, then observed GSRs for individuals with high-risk genotypes will be lower than predicted by the multiplicative model.

Complex inherited diseases are affected by many genes. Although genome-wide association (GWA) studies have detected numerous single-neucleotide polymorphisms (SNPs) associated with elevated risk of inherited diseases, the ways in which causative alleles interact, the extent to which they are comparable in their effects on disease risk, and their utility for genetic counseling are not well understood. Here, I introduce a new set of statistics that quantify the extent of risk to relatives of affected individuals. The idea is closely related to affected-relative pair (ARP) methods of gene mapping.[1] In those methods, an allele that is shared between affected relatives significantly more often than expected from their relationship indicates disease association. As noted by Risch,[1,2] allele frequencies at each locus and the way loci interact affect the probability of allele sharing in affected-relative pairs and hence the power of ARP mapping.

In this paper, I reverse the information flow. If an allele has already been identified in a GWA study as being significantly associated with a disease, then the genotype-specific risks to relatives of cases in the GWA study can be estimated and those estimates can be compared with predictions made under a multiplicative model of interactions among loci. Testing groups of relatives of known relationship can both confirm the association found in the GWA study and indicate whether interactions among loci deviate significantly from the multiplicative model. In the following sections, I first define the genotype-specific recurrence risks (GSRs) and derive an expression for them under the multiplicative model of risk. Then, I present simulation results that demonstrate that GSRs predicted from the multiplicative model differ substantially from observed values if the multiplicative model is not valid. In particular, if interactions among loci result in saturation of risk (meaning that risk does not continue to increase with increasing numbers of causative alleles), predicted GSRs for high-risk genotypes will be larger than observed values. Finally, although no data are available to test the predictions made here, the feasibility of using the GSRs is illustrated by using allele frequencies and genotypic risks for a SNP associated with risk of age-related macular degeneration (AMD).

Genotype-Specific Recurrence Risk

The recurrence risk, denoted by K_R[3] where R indicates the relationship (sibling, half-sibling, etc.), of an inherited disease is the probability that a relative of an affected individual is also affected. Risch[3] introduced the risk ratio, $\lambda_R = K_R/K$, where K is the prevalence of the disease, and showed that under a multiplicative model of interactions among loci, λ_R is the product of effects attributable to each locus,

$$\lambda_R = \prod_{j=1}^{L} \lambda_{jR}. \tag{1}$$

The λ_{jR} are computed from the genotype frequencies and genotype-specific contributions to risk from each locus.

The GSRs are related to the risk ratios. Assume that an allele A at locus j is associated with increased disease risk. Let the average risk to individuals with 2, 1, or 0 copies of A be $w_{j,2}$, $w_{j,1}$, and $w_{j,0}$. It is shown in Appendix A that, for the multiplicative model, the disease risk in a relative with relationship R who carries $k = 2$, 1, or 0 copies of A is

$$\rho_{j,k}^{(R)} = w_{j,k}\left(\frac{\lambda_R}{\lambda_{jR}}\right). \tag{2}$$

The quantities on the right hand side of Equation (2) can all be estimated. The risk ratio, λ_R, is estimated from family studies, and λ_{jR} is estimated from results obtained in the GWA study.

Calculations analogous to those presented in Appendix A show that Equation (2) can be generalized to sets of

[1]Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA
*Correspondence: slatkin@berkeley.edu
DOI 10.1016/j.ajhg.2008.06.011.

Please cite this article in press as: Slatkin, Genotype-Specific Recurrence Risks as Indicators of the Genetic Architecture of Complex Diseases, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.06.011

two or more unlinked loci. For example, if two loci, j and j', are identified as causative in a GWA study, then the genotype-specific risk in relatives can be shown to be

$$\rho^{(R)}_{jj',kk'} = w_{j,k}w_{j',k'}\left(\frac{\lambda_R}{\lambda_{jR}\lambda_{j'R}}\right) \quad (3)$$

under the multiplicative model.

Simulation Study

In order to determine how well the GSRs are predicted when the risk model is not multiplicative, I carried out a simulation study. Each simulation has several steps. First, the model of risk and its parameters are chosen. In some cases, this requires the random generation of parameters for each locus according to rules specified below. Second, allele frequencies at each locus are determined in such a way that the prevalence (K) of the disease is close to a specified value. Third, with the risk model and set of allele frequencies determined, observable statistics are estimated by randomly generating genotypes of pairs of individuals, computing the statistics for each pair, and then taking the average of a large number of replicate pairs. In a single simulation, neither the parameters of the risk model nor the set of allele frequencies change.

The model of risk assumes L diallelic loci, with one allele (denoted by +) associated with increased risk. The frequency of + at each locus is p_j and the loci are assumed to be at Hardy-Weinberg and linkage equilibrium. The genotype of each locus is denoted by k_j, which takes values 2, 1, and 0 if the locus has 2, 1, or 0 + alleles. The genotype of an individual is denoted by a L-vector $\mathbf{k} = \{k_1, \ldots, k_L\}$. I considered three models of the dependence of risk f on $\mathbf{k}$: the unconstrained multiplicative, the constrained multiplicative, and the threshold models.

In the unconstrained multiplicative model, the contribution of genotype k at locus j, $u_{j,k}$, to overall risk is $u_{j,2} = b^{1/L}(1+r_j)$, $u_{j,1} = b^{1/L}(1+h_jr_j)$, and $u_{j,0} = b^{1/L}$. The overall risk is the product across loci:

$$f(\mathbf{k}) = \prod_{j=1}^{L} u_{j,k}. \quad (4)$$

The parameter b is the background risk and the r_j are the maximum effects of each locus. In the simulations, either the r_j are random variables drawn from an exponential distribution with a specified mean, or they are all set to a specified value. The h_j are the dominance parameters, which are either drawn randomly and independently of the r_j from a uniform distribution on (0,1) or set to a specified value.

For some parameter values, f determined by Equation (4) exceeds 1 for some genotypes. In the unconstrained multiplicative model, values of $f > 1$ are allowed. The constrained multiplicative model is the same as the unconstrained model except that, if the value in Equation (4) is > 1, f is set to 1.

The threshold model comes from quantitative genetics. The model assumes that disease risk depends on an underlying liability, z, which is the sum of a genetic component, ξ, and an environmental component, e. The genetic component is the sum of contributions of each locus: $\xi(\mathbf{k}) = \sum_{j=1}^{L} v_j(k_j)$, where $v_j(2) = s_j$, $v_j(1) = h_js_j$, and $v_j(0) = 0$. The s_j are either set to the same specified value or are generated from an exponential distribution with a specified mean. The h_j are either set to the same specified value or generated from a uniform distribution on (0, 1). The environmental component, e, is assumed to be a normally distributed random variable that is independent of ξ and has mean 0 and variance σ_e^2. The model assumes there is a threshold value of liability, T: the risk is b if $z < T$ and 1 otherwise. With these definitions,

$$f(\mathbf{k}) = b + \frac{1}{2}(1-b)\mathrm{erfc}\left[(T - \xi(\mathbf{k}))/\left(\sigma_e\sqrt{2}\right)\right], \quad (5)$$

where erfc is the complementary error function, $\mathrm{erfc}(z) = (2/\sqrt{\pi})\int_z^\infty \exp(-t^2)dt$. If $\sigma_e < 1/4$, f is equivalent to a step function, considered by Lindsey.[4] If $\sigma_e < 1/4$ and $0 < T < 1$, the threshold model is equivalent to the heterogeneous model analyzed by Risch.[3] For larger values of σ_e, f is a sigmoid function of ξ centered at $\xi = T$.

For each risk model, allele frequencies are generated randomly and independently of r_j and h_j with a specified coefficient of variation, CV. This is done by first generating a set of frequencies from a beta distribution with mean 1/2 and coefficient of variation CV. Then, each frequency is multiplied by a factor that is adjusted until the prevalence K is between 0.009 and 0.011. The coefficient of variation is preserved when each frequency is multiplied by the same factor.

In each simulation, K is estimated when the p_j are determined. The next step is to estimate the genotype-specific risks in probands ($w_{j,k}$) by randomly generating replicate individuals with specified genotypes at each locus ($k_j = 2$, 1, 0) and averaging over replicates. To estimate recurrence risks and GSRs, genotypes of pairs of relatives at each locus with various relationships are drawn independently from the joint probabilities of genotypes of relatives in an outbred population. Independently drawing genotypes for each locus is equivalent to assuming the loci are unlinked. The simulated genotype-specific recurrence risks ($\rho^{(R)}_{j,k}$) are found by conditioning on specific genotypes at each locus in the relatives and taking averages over the remaining loci. The predicted GSRs, denoted by $\widehat{\rho}^{(R)}_{j,k}$, are computed from Equation (2). The results in all the figures were based on averages over 10^6 replicate pairs.

The figures show results only for full siblings ($R = S$), but the program produces results for first (parent-offspring), second (half-siblings), and third (first cousins) degree relatives also. The patterns for these other classes of relatives are the same as for full siblings, but the effects are weaker in second and third degree relatives because the risk ratios are smaller.

Please cite this article in press as: Slatkin, Genotype-Specific Recurrence Risks as Indicators of the Genetic Architecture of Complex Diseases, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.06.011

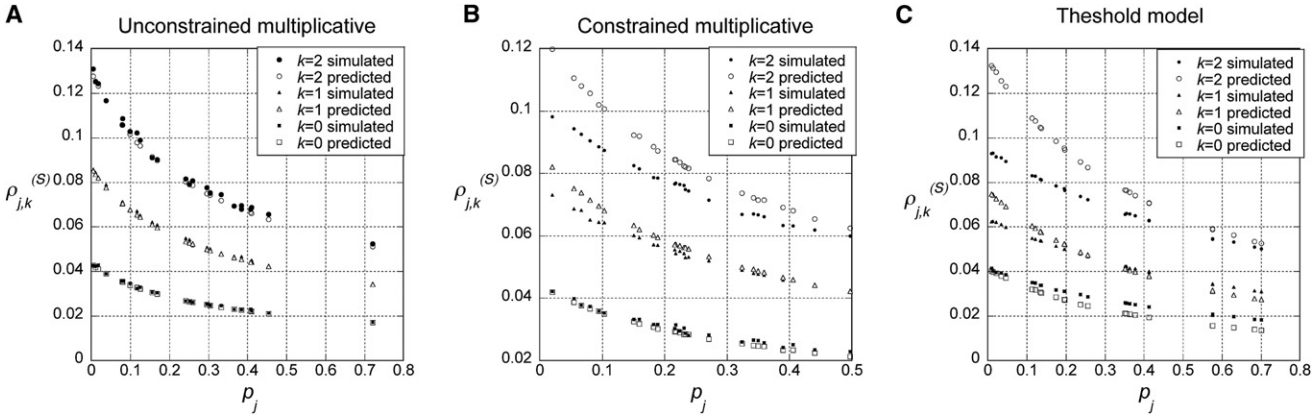

Figure 1. Predicted and Simulated Values of GSRs
Comparison of predicted and simulated values of the genotype-specific risk ratios (GSRs) in full siblings, $\rho_{j,k}^{(S)}$, in the three models of disease risk defined in the text. In all cases, $L = 25$, $b = 10^{-6}$, $h = 0.5$, CV = 0.75, and all results are based on 10^6 replicates of the simulation program described in the text.
(A) Unconstrained multiplicative model with $r_j = 2$ for all j. For the set of parameter values used, $K = 0.0104$ and $\lambda_S = 4.65$.
(B) Constrained multiplicative model with $r_j = 2$ for all j ($K = 0.00972$, $\lambda_S = 4.56$).
(C) Threshold model with $s_j = 1$ for all j, $T = 12$, and $\sigma_e = 1.5$ ($K = 0.00924$ and $\lambda_S = 4.43$).

Simulation Results

In the first set of results, parameter values are the same at all loci. In all three models, $h_j = 1/2$. In both multiplicative models, $r_j = 2$ and in the threshold model $s_j = 1$. With these restrictions, the models are exchangeable, meaning that the risk depends only on the numbers of loci homozygous and heterozygous for the + allele, and not on the genotypes at individual loci. In all cases, $L = 25$, CV = 0.75, and frequencies were adjusted so that $0.009 < K < 0.011$. These parameter values were chosen so that the average effect of each locus and the recurrence risk to full siblings ($\lambda_S \approx 5$) are comparable to what was found by Maller et al.[5] in their study of AMD (discussed below). Results for other parameter values are similar.

In Figure 1A, the predicted and observed values of $\rho_{j,k}^{(S)}$ are shown for the unconstrained multiplicative model. As expected, the observed (filled symbols) and expected (open symbols) values are essentially the same. Similar results were obtained with $h = 0$ and $h = 1$.

Figure 1B shows that the simulated and predicted values of $\rho_{j,k}^{(S)}$ are no longer equal for the constrained multiplicative model. For the high-risk genotype, $\widehat{\rho}_{j,2}^{(S)} > \rho_{j,2}^{(S)}$. The threshold model, with parameters chosen to have roughly the same λ_S, yields results similar to the constrained multiplicative model. For both models, relatives with the high-risk genotypes have a lower risk than is predicted by the multiplicative model. The reason is that risk saturates with the number of loci carrying causative alleles. If an individual has a high-risk genotype at one locus, there is less potential for high-risk genotypes at other loci to increase risk in those models than there is in the unconstrained multiplicative model. Figure 2

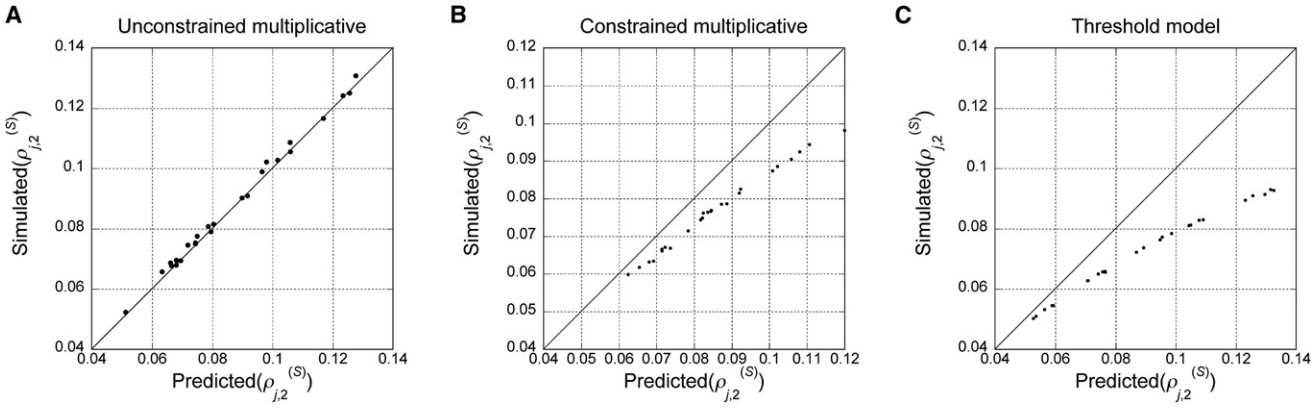

Figure 2. Comparison of Predicted and Simulated GSRs
Graphical comparison of simulated and predicted values of the GSRs for the high-risk genotype ($k = 2$) in full siblings, $\rho_{j,2}^{(S)}$, in the three models of disease risk defined in the text. The data are from Figure 1 for $k = 2$. The 45° lines are drawn to facilitate comparison.

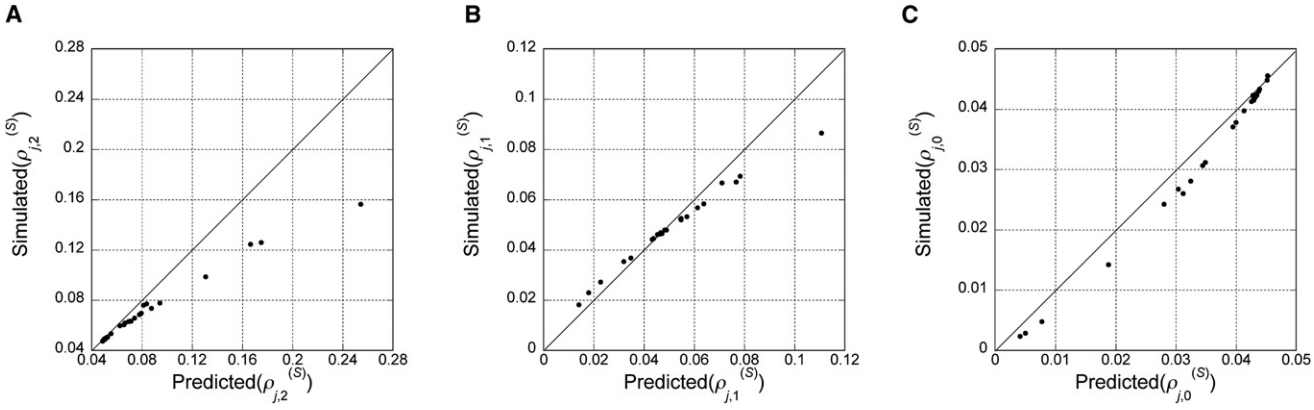

Figure 3. Effects of Varying Allelic Contributions to Risk
Comparison of predicted, $\widehat{\rho}_{j,k}^{(S)}$, and simulated, $\rho_{j,k}^{(S)}$, values of the GSRs in full siblings with genotypes +/+ ($k = 2$) (A), +/− ($k = 1$) (B), and −/− ($k = 0$) (C) in the threshold models in which the r_j vary among loci but $h = 0.5$ for every locus. The parameters are the same as in the threshold model in Figures 1 and 2, except that the s_j are drawn from an exponential distribution with mean 1 ($K = 0.01097$ and $\lambda_S = 4.13$).

shows the systematic difference for the high-risk genotypes in another way that makes the patterns easier to visualize.

If loci differ in their effects on risk (i.e., r_j or s_j vary), the results are essentially the same, as illustrated in Figure 3 for the threshold model. The GSRs for the high-risk genotypes differ substantially from the predictions of the multiplicative model. Similar results are obtained when the h_j also vary among loci, as shown in Figure 4.

Example

To illustrate how the theory presented here can be applied, I use parameter values for a locus associated with elevated risk of AMD (MIM 603075), the leading cause of blindness among elderly people in developed countries. Maller et al.[5] presented evidence that five SNPs are associated with higher risk of AMD and that together they accounted for 33%–67% of the recurrence risk to full siblings. Maller et al. could not reject a multiplicative model of interaction among these SNPs, even though the relatively large effects on risk gave their test of deviations from the multiplicative model considerable power. Effect sizes are larger than in more recent GWA studies but the potential use of the data is the same.

For one SNP, rs10490924 (MIM 61131) on chromosome 10, the frequency of T among cases was 0.455 and among controls 0.194 (Table 2 in Maller et al.). From the assumption of multiplicative interactions within a locus, these observations imply $w_2/w_0 = 12.03$ and $w_1/w_0 = 3.45$. The absolute risks are age dependent. Vingerling et al.[6] estimated the prevalence of AMD in the Rotterdam study to be 1.7%, with risk increasing from 0.1% in individuals 55–64 to 3.7% for individuals older than 85. If w_0 is set to 0.017, $w_2 = 0.204$ and $w_1 = 0.059$.

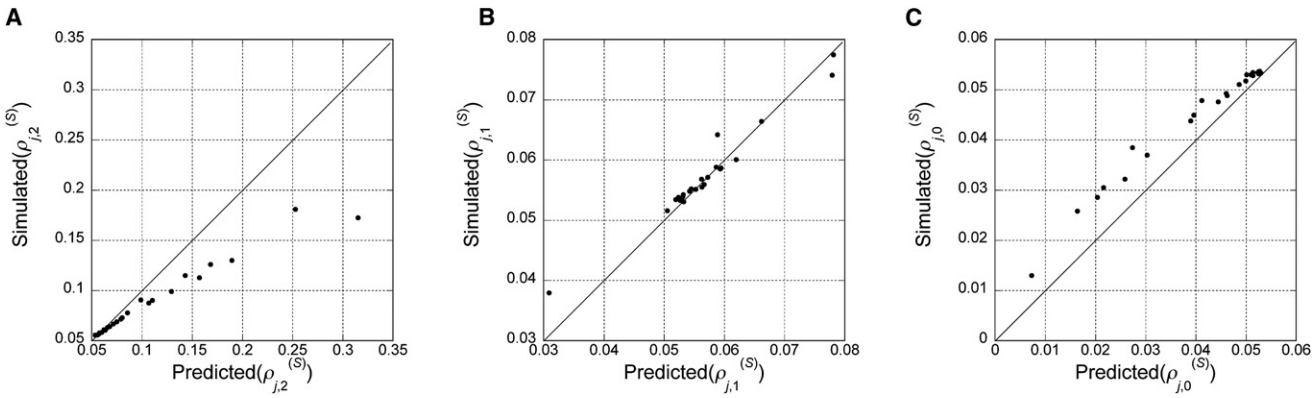

Figure 4. Effects of Varying Allelic Contributions and Heritabilities
Comparison of predicted, $\widehat{\rho}_{j,k}^{(S)}$, and simulated, $\rho_{j,k}^{(S)}$, values of the GSRs in full siblings with genotypes +/+ (A), +/− (B), and −/− (C) in the threshold models in which the r_j and h_j vary among loci. The parameters are the same as in the threshold model in Figures 1 and 2, except that $\sigma_e = 2$, the s_j are drawn from an exponential distribution with mean 1, and the h_j are drawn from a uniform distribution on (0,1). For the set of parameter values used, $K = 0.0109$ and $\lambda_S = 4.89$.

Please cite this article in press as: Slatkin, Genotype-Specific Recurrence Risks as Indicators of the Genetic Architecture of Complex Diseases, The American Journal of Human Genetics (2008), doi:10.1016/j.ajhg.2008.06.011

Maller et al.[5] estimated the contribution of rs10490924 to overall recurrence risk in full siblings (corresponding to λ_{jS} above) to be 1.45 and reported that estimated values of λ_S were in the range 3–6. With the average $\lambda_S = 4.5$, the above theory predicts that $\rho_2^{(S)} = 0.204 \times 4.5/1.45 = 0.633$ and $\rho_1^{(S)} = 0.059 \times 4.5/1.45 = 0.183$. These are the estimated risks to a proband's full sibling carrying 2 and 1 copies of T at rs10490924 under the multiplicative model. The confidence intervals of these estimates have to reflect the uncertainty in both λ_S and λ_{jS}. If we assume the uncertainty in λ_S dominates and use the range 3–6 as a confidence interval, then $0.422 < \rho_2^{(S)} < 0.844$ and $0.039 < \rho_1^{(S)} < 0.079$ under the assumption of the multiplicative model.

As described in the Summary, there are three ways these predictions can be used. The first is to provide confirmation of the correlation between T at rs10490924 and AMD. To illustrate, assume a group of full siblings of the cases in the Maller et al. study can be tested for AMD and that n of that group has genotype TT. If T were in fact not associated with AMD, then the risk conditional on the genotype at rs10490924 is just the risk expected in full siblings, $\rho_2^{(S)} = 0.017 \times 6 = 0.102$ where the upper bound of λ_S is used to be conservative. If n = 100, then observing 19 or more affected TT full siblings would reject the null at the 1% level, by a one-tailed binomial test.

A second use of these predictions is to test the hypothesis of multiplicative interactions. If we use the lower bound of the predicted value of $\rho_2^{(S)}$, 0.422, and again assume n = 100, then observing 30 or fewer TT affected full siblings would reject the multiplicative model at the 1% level by a one-tailed binomial test.

A third use of these predictions is for genetic counseling. For example, an individual who is homozygous for T at rs10490924 and whose full sibling has AMD can be told that the (non-age-adjusted) chance of getting AMD is between 42.2% and 84.4%, provided that further study confirms that rs10490924 is associated with AMD in the population to which the individual belongs and that a model of multiplicative interactions among loci is valid.

Discussion and Conclusions

This report shows that standard population genetics theory can be used to predict genotype-specific risks in relatives of affected probands, thereby providing (1) another way to test whether SNPs identified in a GWA study are causative, (2) another way to test whether the multiplicative model of gene interactions across loci applies, and (3) information needed to counsel relatives of affected individuals. The main result is that the genotype-specific risks to relatives of affected individuals are increased by a factor that may be nearly as large as λ_R.

It is becoming accepted that results from GWA studies offer little to genetic counselors. For example, the Wellcome Trust Consortium[7] concluded, "These estimates demonstrate the limited potential of the variants thus far identified (singly or in combination) to provide clinically useful prediction of disease." That is true for individuals of known genotype at one or more SNPs identified as conferring higher risk. The increase in risk is too small to be of much predictive value. For relatives of cases, however, the increase in risk may be substantially larger. For example, if $\lambda_S = 5$ and the high-risk genotype of an identified SNP has a 20% higher risk, then a sibling with the high-risk genotype has a risk almost 6 times higher than the background risk if the multiplicative model is valid.

The simulations verified that the predictions made by the analytic theory for the unconstrained multiplicative model were correct and also showed that if risk does not continue to increase with increasing numbers of causative alleles, predicted GSRs for the high-risk genotypes are too large. The unconstrained multiplicative model is widely used in theoretical studies both because of its mathematical simplicity and because it embodies the parsimonious idea that each locus contributes to risk independently. Furthermore, several studies, including that of Maller et al.,[5] have found no significant deviation from a multiplicative model of interactions among loci. However, as has been shown elsewhere,[8] it is difficult to parameterize the multiplicative model in such a way that the average risk (K) is low (≈ 0.01), the risk ratio to full siblings λ_S is relatively high (≈ 5), and the risk is ≤ 1 for all genotypes. The reason is that low K and high λ_S together require a large variance in risk among genotypes found in significant frequency in the population.[9] The only way to have a high variance in risk and still ensure that $f \leq 1$ is to carefully adjust the parameters so that $f = 1$ for an individual homozygous for high-risk alleles at every causative locus.[8] It is likely, however, that the number of potentially causative loci exceeds the number needed to create significant risk. In that case, individuals with even more causative alleles will have $f > 1$ under the unconstrained multiplicative model.

Constraining the multiplicative model so that $f \leq 1$ results in little change in either K or λ_S because they are determined by the genotypes in highest frequency in a population. Genotypes that result in $f > 1$ will have such low frequencies that setting their risks to 1 does not change K or λ_S by much. As we have seen, however, that constraint does affect the GSRs. The reason is that, under the unconstrained multiplicative model, relatives of affected individuals are predicted to have unfeasibly high risks with significant probability. If instead the risk model implies that there is saturation of risk with increasing numbers of causative alleles, as is the case for both the constrained multiplicative model and the threshold model, then relatives with high-risk genotypes have risks lower than predicted by the unconstrained multiplicative model. Consequently, detecting a lower-than-predicted genotype-specific recurrence risk indicates saturation of the risk as a function of the number of causative alleles.

The theory presented in the previous sections assumes that the allele identified in a GWA study is causative and not simply in linkage disequilibrium with a causative allele. Linkage to a causative allele would make no difference provided that the two loci are closely enough linked that

no recombination between them is likely during the meioses separating the relatives. For GWAs based on 500,000 or more SNPs, implying an average map distance on the order of 0.01 cM between adjacent SNPs, that condition is almost certainly satisfied.

The GSR statistics for full siblings, $\rho_k^{(S)}$, are similar to Rybicki and Elston's[10] conditional recurrence risk ratio, λ_S^*, which is the recurrence risk ratio in relatives of probands having a specific "at risk" genotype. The difference is that $\rho_k^{(S)}$ is estimated from siblings of all affected probands whereas λ_S^* is estimated from siblings of only those probands with a specific genotype. Appendix B derives a formula for λ_S^* for the unconstrained multiplicative model. Although the two sets of statistics are closely related, λ_S^* cannot as easily be expressed in terms of observable quantities.

The results presented here add support the point made by Clerget-Darpoux and Elston[11] and others that family studies can provide important additional information in this era of GWAs. Unrelated individuals are essential for successful GWAs, but related individuals provide information that is not available even from large samples of unrelated individuals. Determining the genotype-specific risks to relatives of cases in a case-control study will minimize the effects of population heterogeneity and other factors, including the tendency for estimated effects of alleles identified in GWA studies to be biased upwards,[12] that contribute to difficulties in replicating associations with disease risk when independent populations are used.

Appendix A

The genotype-specific recurrence risks (GSRs) defined in the text can be computed from the multiplicative model via methods similar to those of Risch.[3] Assume that locus j of L loci has been identified as having a causative allele, +. Let $g_{X,j}$ and $g_{Y,j}$ be the genotypes at locus j of the proband (denoted by X) and the relative with relationship R (denoted by Y): $g_{X,j}$, $g_{Y,j} = 2$, 1, or 0, if the individual carries 2, 1, or 0 copies of + at locus j.

Let X and Y be the affected status of the two relatives. The GSR for locus j is defined to be

$$\rho_{j,k}^{(R)} = \Pr(Y = 1 \mid X = 1, g_{Y,j} = k) \tag{A1}$$

for $k = 2, 1, 0$.

Under the multiplicative model, $X = \prod_{j=1}^{L} x_j$ and $Y = \prod_{j=1}^{L} y_j$, where x_j and y_j are the contributions of locus j to the overall risk:

$$\Pr(x_j = 1 \mid g_{X,j} = k) = \Pr(y_j = 1 \mid g_{Y,j} = k) = u_{j,k}.$$

For unconstrained multiplicative model, we recall from Risch[3]

$$K = \Pr(X = 1) = \prod_{j=1}^{L} \Pr(x_j = 1) = \prod_{j=1}^{L} K_j, \tag{A2}$$

where $K_j = p_j^2 u_{j,2} + 2p_j(1 - p_j)u_{j,1} + (1 - p_j)^2 u_{j,0}$, and p_j is the frequency of + at locus j, and

$$\lambda_R = \frac{\Pr(Y = 1 \mid X = 1)}{K} = \prod_{j=1}^{L} \frac{\Pr(y_j = 1 \mid x_j = 1)}{K_j} = \prod_{j=1}^{L} \lambda_{jR}. \tag{A3}$$

Furthermore, the average risk given the genotype at locus j is

$$w_{j,k} = \Pr(Y = 1 \mid g_{Y,j} = k) = u_{j,k} \prod_{j' \neq j} K_{j'}. \tag{A4}$$

where the product is over all j' ($1 \leq j' \leq L$) except $j' = j$.

From these equations it follows that

$$\begin{aligned} \Pr(Y = 1 \mid X = 1, g_{Y,j} = k) &= u_{j,k} \prod_{j' \neq j} \Pr(y_{j'} = 1 \mid x_{j'} = 1) \\ &= u_{j,k} \prod_{j' \neq j} K_{j'} \prod_{j' \neq j} \frac{\Pr(y_{j'} = 1 \mid x_{j'} = 1)}{K_{j'}} \\ &= w_{j,k} \left(\frac{\lambda_R}{\lambda_{j,R}} \right). \end{aligned} \tag{A5}$$

This result can be generalized to any number of loci identified as causative.

Appendix B

The GSR statistics defined in the text are closely related to the conditional relative risks defined by Rybicki and Elston.[10] The conditional risk for genotype k at locus j, $\lambda_{k,R}^*$, is defined to be

$$\lambda_{k,R}^* = \Pr(Y = 1 \mid X = 1, g_{X,j} = k) \tag{A6}$$

where, as above, X denotes the affected status of the proband and Y denotes the affected status of the relative. Note that in Equation (A6), the conditioning is on $g_{X,j}$ whereas in Equation (A1) it is on $g_{Y,j}$.

With the assumption of independence across loci,

$$\begin{aligned} \Pr(Y = 1 \mid X = 1, g_{X,j} = k) &= \Pr(y_j = 1 \mid x_j = 1, g_{X,j} = k) \\ &\quad \times \prod_{j' \neq j} \Pr(y_{j'} = 1 \mid x_{j'} = 1) \\ &= \Pr(y_j = 1 \mid x_j = 1, g_{X,j} = k) \prod_{j' \neq j} K_{j'} \prod_{j' \neq j} \lambda_{j'R}. \end{aligned} \tag{A7}$$

To evaluate the remaining conditional probability, we use the fact that, given the genotype at locus j in the proband ($g_{X,j}$), the contribution of locus j to the risk in the relative (y_j) depends only on $g_{X,j}$ and the relationship, and not on the contribution of locus j to the risk in the proband (x_j):

$$\begin{aligned} \Pr(y_j = 1 \mid x_j = 1, g_{X,j} = k) &= \sum_{k'=0}^{2} \Pr(y_j = 1 \mid g_{Y,j} = k') \\ &\quad \times \Pr(g_{Y,j} = k' \mid g_{X,j} = k) \\ &= \sum_{k'=0}^{2} u_{j,k'} \Pr(g_{Y,j} = k' \mid g_{X,j} = k). \end{aligned} \tag{A8}$$

where the conditional probability is obtained from the Hardy-Weinberg frequencies and the probabilities of identity by descent in pairs of relatives. Substituting in Equation (A7) yields

$$\Pr(Y = 1 \mid X = 1, g_{X,j} = k) = \left(\frac{\lambda_R}{\lambda_{jR}}\right) \sum_{k'=0}^{2} w_{j,k'} \times \Pr(g_{Y,j} = k' \mid g_{X,j} = k).$$

Acknowledgments

I thank N. Freimer and G.J. Thomson for helpful discussions of this topic and M.J. Daly for clarification about the Maller et al. paper. This research was supported in part by a grant from the US NIH, R01-GM40282.

Received: April 30, 2008
Revised: June 10, 2008
Accepted: June 13, 2008
Published online: July 3, 2008

Web Resources

The URLs for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/
Slatkin Laboratory, http://ib.berkeley.edu/labs/slatkin/

References

1. Risch, N. (1990). Linkage strategies for genetically complex traits: II. The power of affected relative pairs. Am. J. Hum. Genet. *46*, 229–241.
2. Risch, N. (1990). Linkage strategies for genetically complex traits: III. The effect of marker polymorphism on analysis of affected relative pairs. Am. J. Hum. Genet. *46*, 242–253.
3. Risch, N. (1990). Linkage strategies for genetically complex traits: I. Multilocus models. Am. J. Hum. Genet. *46*, 222–228.
4. Lindsey, J.W. (2005). Familial recurrence rates and genetic models of multiple sclerosis. Am. J. Med. Genet. A. *135A*, 53–58.
5. Maller, J., George, S., Purcell, S., Fagerness, J., Altshuler, D., Daly, M.J., and Seddon, J.M. (2006). Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. Nat. Genet. *38*, 1055–1059.
6. Vingerling, J.R., Dielemans, I., Hofman, A., Grobbee, D.E., Hijmering, M., Kramer, C.F.L., and Dejong, P. (1995). The prevalence of age-related maculopathy in the Rotterdam study. Ophthalmology *102*, 205–210.
7. Wellcome Trust Case Control Consortium (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661–678.
8. Schliekelman, P., and Slatkin, M. (2002). Multiplex relative risk and estimation of the number of loci underlying an inherited disease. Am. J. Hum. Genet. *71*, 1369–1385.
9. Slatkin, M. (2008). Exchangeable models of complex inherited diseases. Genetics, in press.
10. Rybicki, B.A., and Elston, R.C. (2000). The relationship between the sibling recurrence-risk ratio and genotype relative risk. Am. J. Hum. Genet. *66*, 593–604.
11. Clerget-Darpoux, F., and Elston, R.C. (2007). Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. Hum. Hered. *64*, 91–96.
12. Goring, H.H.H., Terwilliger, J.D., and Blangero, J. (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. Am. J. Hum. Genet. *69*, 1357–1369.