European Journal of Human Genetics (2011) 19, 180–185
© 2011 Macmillan Publishers Limited All rights reserved 1018-4813/11
www.nature.com/ejhg

ARTICLE

Discriminative accuracy of genomic profiling comparing multiplicative and additive risk models

Ramal Moonesinghe*[,1], Muin J Khoury[2], Tiebin Liu[2] and A Cecile JW Janssens[3,4]

Genetic prediction of common diseases is based on testing multiple genetic variants with weak effect sizes. Standard logistic regression and Cox Proportional Hazard models that assess the combined effect of multiple variants on disease risk assume multiplicative joint effects of the variants, but this assumption may not be correct. The risk model chosen may affect the predictive accuracy of genomic profiling. We investigated the discriminative accuracy of genomic profiling by comparing additive and multiplicative risk models. We examined genomic profiles of 40 variants with genotype frequencies varying from 0.1 to 0.4 and relative risks varying from 1.1 to 1.5 in separate scenarios assuming a disease risk of 10%. The discriminative accuracy was evaluated by the area under the receiver operating characteristic curve. Predicted risks were more extreme at the lower and higher risks for the multiplicative risk model compared with the additive model. The discriminative accuracy was consistently higher for multiplicative risk models than for additive risk models. The differences in discriminative accuracy were negligible when the effect sizes were small (<1.2), but were substantial when risk genotypes were common or when they had stronger effects. Unraveling the exact mode of biological interaction is important when effect sizes of genetic variants are moderate at the least, to prevent the incorrect estimation of risks.
European Journal of Human Genetics (2011) **19,** 180–185; doi:10.1038/ejhg.2010.165; published online 17 November 2010

Keywords: discriminative accuracy; genomic profiles; sensitivity; specificity; additive models

INTRODUCTION

Common diseases such as type 2 diabetes, cardiovascular disease, asthma and osteoporosis are caused by a complex interplay of many genetic and non-genetic factors, each of which conveys only a minor increase in the risk of disease. Genetic associations typically have odds ratios, with effect sizes ranging from 1.1 to 1.5, with each single polymorphism explaining only a minor fraction of the variation in a phenotype. Because the predictive accuracy of testing for a single genetic variant is limited, genetic prediction of disease will be based on testing for multiple genetic variants simultaneously (*genomic profiling*).

Studies on the combined effect of multiple genetic variants generally assume that joint effects result from multiplying the risk of single variants rather than from adding them.[1–4] Further, empirical studies implicitly assume multiplicative risk models, as the joint effects of multiple genes are typically examined using logistic regression or Cox proportional hazard analyses.[5–7] The assumption that susceptibility genes interact in a multiplicative manner may not be correct.[8] On the basis of the ratio of disease risks in monozygotic and dizygotic twins of cancer cases, an additive model provided the best fit for most common cancers, including breast cancer.[9] Conversely, a very high incidence was reported for monozygotic twins of women with breast cancer, which would be more consistent with a multiplicative model for joint genetic effects.[10]

Because estimated disease risks of genomic profiles may differ depending on whether a multiplicative or additive risk model is assumed, the clinical validity of the combined testing may change accordingly. The aim of this study was to investigate predicted risks and the discriminative accuracy of predictive testing using multiple genetic variants, by comparing multiplicative and additive risk models for the estimation of the joint genetic effects. The two risk models were compared using hypothetical scenarios with equal genotype frequencies and risks and using realistic scenarios of predictive testing for multiple genetic variants in prostate cancer.

METHODS

Distribution of predicted risks and discriminative accuracy were obtained from formulae. The formulae are presented for the hypothetical scenario in which we considered genomic profiles based on K genetic variants with equal genotype frequencies (m) and equal relative risks (R). In the hypothetical scenario, for ease of interpretation, we consider genetic variants that have dominant or recessive effects yielding two possible results for each gene: a risk genotype and a non-risk or referent genotype. Under the assumption of equal frequencies and relative risks, genomic profiles can be expressed as risk genotype scores, values of which range from zero to K. The formulae to calculate disease risk when the risk score equals c for both multiplicative models and additive models are given by equation (A4) in Appendix A and by equation (B1) in Appendix B, respectively.

The discriminative accuracy, quantified as the area under the receiver operating characteristic curve (AUC), is determined by the distribution of disease risks in those who will develop the disease and those who will not. The AUC indicates the discriminative accuracy of a continuous test. The AUC ranges from 0.5 (total lack of discrimination) to 1.0 (perfect discrimination)

[1]Office of Minority Health and Health Disparities, Centers for Disease Control and Prevention, Atlanta, GA, USA; [2]Office of Public Health Genomics, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA; [3]Department of Public health, Erasmus MC University Medical Center, Rotterdam, The Netherlands; [4]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands
*Correspondence: Dr R Moonesinghe, Office of Minority Health and Health Disparities, 4770 Buford Highway, Mailstop E-67, Atlanta, GA 30341, USA.
Tel: +1 404 498-2342; Fax: +1 770 488-8336; E-mail: rmoonesinghe@cdc.gov
Received 24 February 2010; revised 26 August 2010; accepted 14 September 2010; published online 17 November 2010

and is independent of the prevalence of disease. The AUC can be mainly considered as the probability that the test correctly identifies the affected subject from a pair in whom one is affected and one is unaffected. An AUC of 0.95 means that 95% of the pairs are correctly classified, whereas a test with an AUC of 0.50 is non-discriminative – as accurate as tossing a fair coin.

The AUC can be estimated non-parametrically from the empirical ROC curve. In this study, we derive equations for sensitivity and specificity for both additive and multiplicative models and construct empirical ROC curves. An empirical ROC curve is constructed by connecting all combinations of sensitivity and specificity obtained at all possible cutoff levels of the genotype score. AUC is calculated using the Trapezoidal rule. ROC curves and its characteristics have been described in many papers.[11,12]

The sensitivity and specificity of genotype scores are calculated from the sensitivity and specificity of each single genetic variant. The sensitivity of a single genetic variant is the percentage of carriers of the risk genotype among individuals who will develop the disease, and specificity is the percentage of non-carriers among those who will not develop the disease. The sensitivity, β_S, and specificity, α_S, of a single gene test are given by the following equation:[13]

$$\beta_S = \frac{\mathrm{Rm}}{(\mathrm{Rm}+(1-m))} \text{ and}$$

$$\alpha_S = 1 - \frac{m - p\beta_s}{1-p}$$

where R is the relative risk of the risk genotype, p is the risk of the disease in the population and m is the risk genotype frequency.

For a multiplicative model, the sensitivity (β_{Mc}) at each cutoff value of the genotype score is a function of the sensitivities of the single genetic variants:

$$\beta_{Mc} = \sum_{i=c}^{K} \binom{K}{i} \beta_S^i (1-\beta_S)^{K-i}$$

where β_S is the sensitivity of each single genetic variant, K is the total number of variants that are considered in the genotype score, and c is the cutoff value of the genotype score. The derivation is presented in Appendix A When β_{Mc} is known, the specificity at each cutoff value (α_{Mc}) can be obtained from Table 1:

$$\alpha_{Mc} = 1 - \frac{\sum_{i=c}^{K} \binom{K}{i} m^i (1-m)^{(K-i)} - p\beta_{Mc}}{1-p}$$

For the additive model, sensitivity β_{Ac} and specificity α_{Ac} are given by the following equation:

$$\beta_{Ac} = \frac{\sum_{i=c}^{K} \binom{K}{i} (i(R-1)+1) m^i (1-m)^{K-i}}{1+Km(R-1)} \text{ , and}$$

$$\alpha_{Ac} = 1 - \frac{\sum_{i=c}^{K} \binom{K}{i} m^i (1-m)^{(K-i)} - p\beta_{Ac}}{1-p} \text{, respectively.}$$

The derivations for these formulae are given in Appendix B.

We considered genotype scores that were based on 40 genes, and investigated risk distributions and AUC for different combinations of relative risks and genotype frequencies. We assumed that the disease risk in the population was 10%.

RESULTS

Figure 1 shows the disease risks for genotype scores of up to 20 for the two models when relative risks are 1.1, 1.2 and 1.5, and genotype frequency is 0.1 and 0.3. We considered genotype scores of up to 20 because anyone in the population having a genotype score of more than 20 is almost zero for these genotype frequencies. The disease risk for the additive model increases linearly with increasing risk genotypes. When genotype frequency is 10%, the disease risk for the additive model is higher than the disease risk for the multiplicative model for up to a genotype score of 5, and when genotype frequency is 30%, the disease risk for the additive model is higher than the disease risk for the multiplicative model for up to a genotype score of 13 for relative risk 1.2 and up to a genotype score of 14 for relative risk 1.5. When the genotype score is 4 in the genomic profile, with relative risk 1.2 and genotype frequency 10%, the increase in disease risk for the additive model is only 0.006; when relative risk is increased to 1.5, the increase is 0.028; when genotype frequency is 30% and relative risk 1.2, the increase is 0.033; and when genotype frequency is 30% and relative risk 1.5, the increase is 0.041. Theoretically, multiplicative models give higher disease risk compared with additive models when there are a large number of risk genotypes.

Table 1 The relationship of sensitivity and specificity of genomic profiling to genotype and disease frequencies (multiplicative models)

	Will develop disease	*Will not develop disease*	*Total*
Genotype score $>c$	$p\beta_{Mc}$	$(1-\alpha_{Mc})(1-p)$	$\sum_{i=c}^{K} \binom{K}{i} m^i (1-m)^{(K-i)}$
Genotype score $\leq c$	$(1-\beta_{Mc})p$	$\alpha_{Mc}(1-p)$	$1 - \sum_{i=c}^{K} \binom{K}{i} m^i (1-m)^{(K-i)}$
Total	P	$1-p$	1

The vertical lines in Figure 1 denote the distribution of risk genotypes in the population. When the genotype frequency is 0.1, only 8 out of 10 000 in the population can have more than 10 risk genotypes, whereas when the genotype frequency is 0.3, more than 75% of the population can have more than 10 risk genotypes. For more than 79% of the population, the risk under the additive model is higher than the risk under the multiplicative model. However, when relative risk is around 1.1, the difference in risk between the two models is almost negligible. For the population for which the multiplicative model gives higher risk than the additive model, the difference in risk is very large. This situation is to be expected, as both models assume identical prevalence of disease in the population.

We also considered 40 risk genotypes with genotype frequency 0.1 and relative risk 1.2, and another risk genotype, A, with the same genotype frequency and relative risk of 3.5 in a genomic profile. The risks for the two models for the populations with and without risk genotype A are shown in Figure 2. For the sub-population that has risk genotype A, almost 77% shows a higher risk under the multiplicative model compared with the additive model. However, the proportion of the sub-population that has the risk genotype is only 10% of the overall population.

Figure 3 shows the ROC curves for the different combinations of relative risks and genotype frequencies studied. The AUC for the multiplicative model (AUC_M) and the AUC for the additive model (AUC_A) increase with increasing values of relative risks. When relative risk is 1.1 and genotype frequency is 10%, the AUC_M and AUC_A are 0.56 and 0.53, respectively. When the relative risk is 1.5, AUC_M

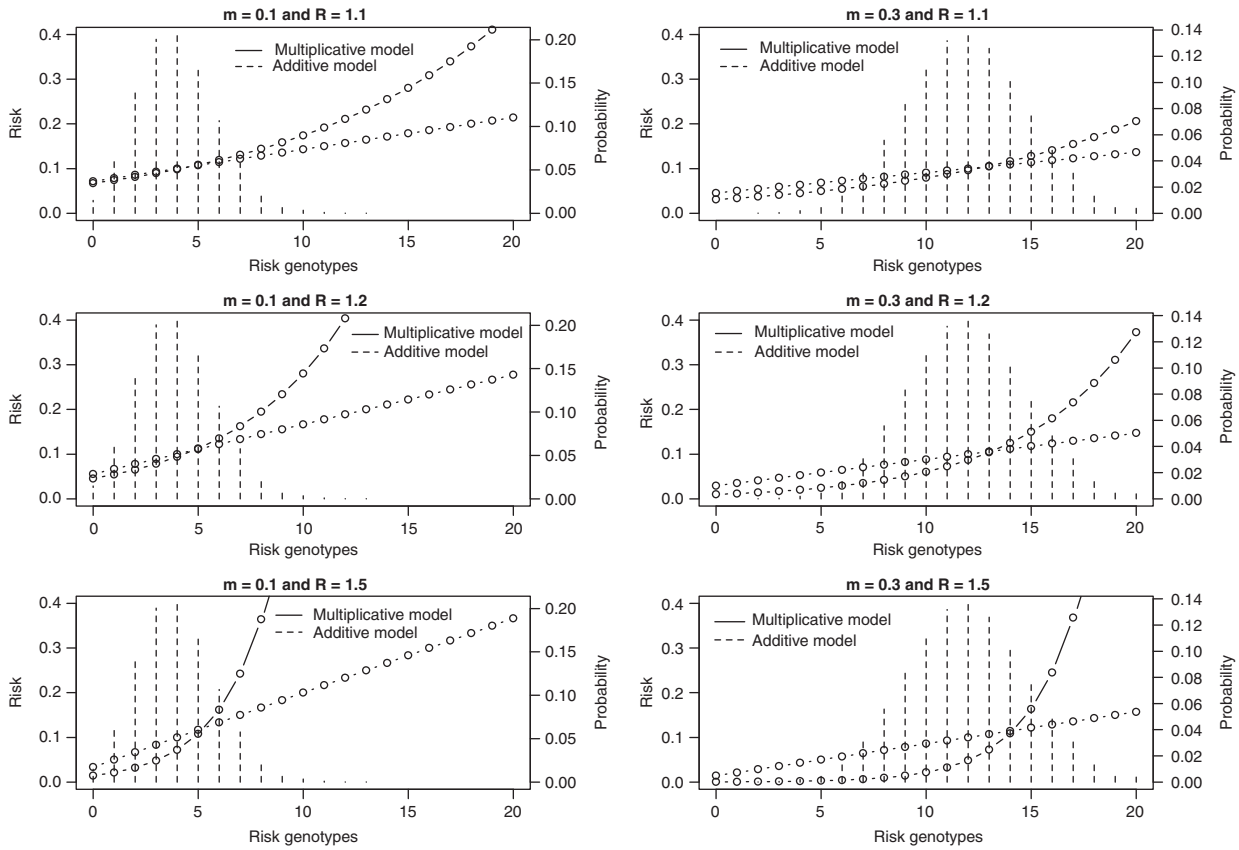

Figure 1 Disease risks for up to 20 risk genotypes when genomic profiles are based on 40 genetic variants with genotype frequency (*m*) 0.1 and 0.3 and relative risk (*R*) 1.1, 1.2 and 1.5 for multiplicative and additive models. Disease risk=10%. Vertical lines give the probability distribution of risk genotypes in the population.

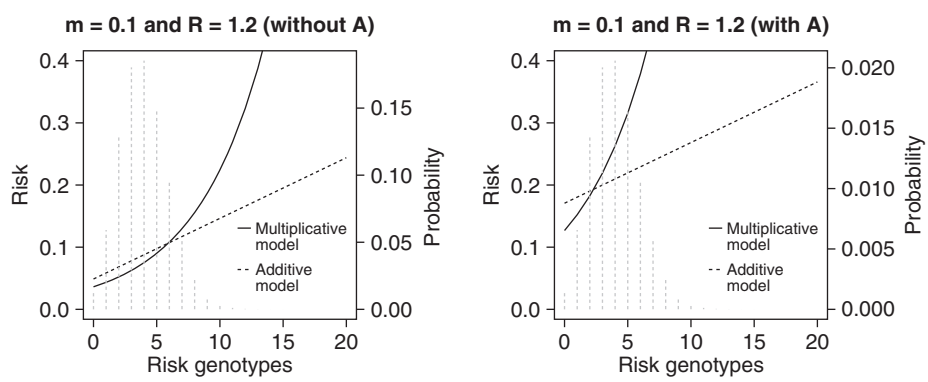

Figure 2 Disease risks for up to 20 risk genotypes when a genomic profile is based on 40 risk genotypes with genotype frequency (m) 0.1 and relative risk (R) 1.2, and another risk genotype, A, with the same genotype frequency and relative risk of 3.5 for multiplicative and additive models. Disease risk=10%. Vertical lines give the probability distributions of risk genotypes for the two sub-populations with and without risk genotype A.

increases to 0.74 and AUC_A increases to only 0.60. This shows that higher relative risks result in larger differences between AUC_M and AUC_A. When genotype frequency is increased from 0.1 to 0.4, keeping relative risk at 1.5, the AUC_M increases from 0.74 to 0.85, but AUC_A declines from 0.6 to 0.55. As seen in Figure 2, there is not much difference between AUC_M and AUC_A for lower values of relative risks (around 1.1) and for lower values of genotype frequencies (less than or equal 10%). For example, when relative risk is 1.1 and genotype frequency is 0.1, the difference between AUC_M and AUC_A is only 0.02. However, when relative risk is 1.5 and genotype frequency is 0.4, the difference between AUC_M and AUC_A is 0.30. AUC_M increases steeply with increasing relative risks and genotype frequency, whereas the increase in AUC_A remains small.

To explore this phenomenon further, we plotted the AUC against genotype frequency for relative risks of 1.2 and 1.5 (Figure 4). The AUC for the multiplicative model increases with increasing genotype frequency up to 50% and then declines. The AUC for the additive model increases with increasing genotype frequency for rare genotypes

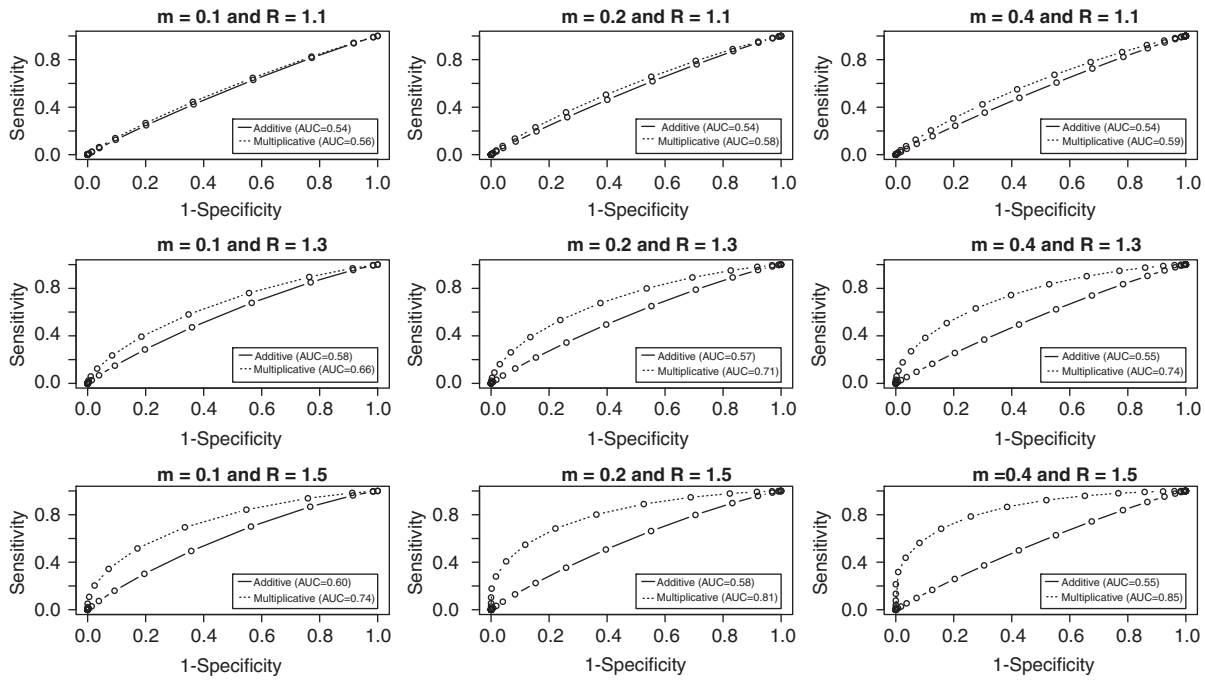

Figure 3 AUC for genomic profiles with risks calculated from additive and multiplicative risk models. Disease prevalence is 10%. m=genotype frequency, R=relative risks.

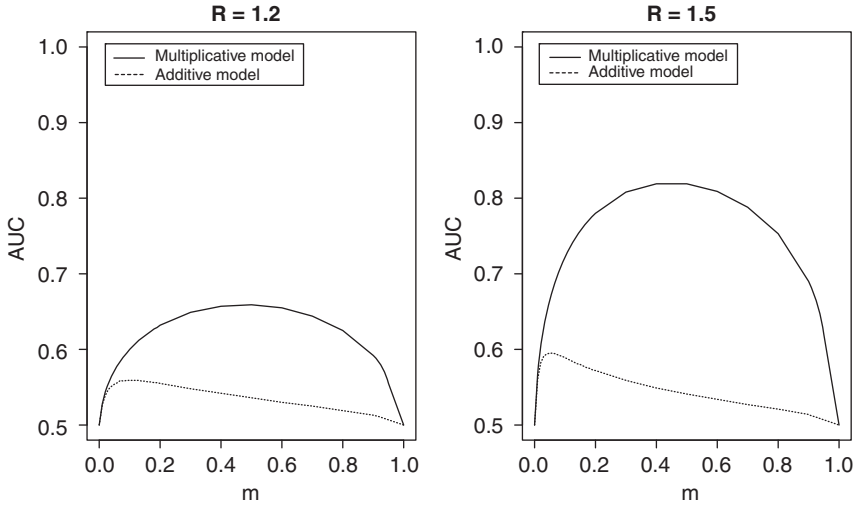

Figure 4 The AUC for multiplicative and additive models as a function of genotype frequency and relative risks. m=genotype frequency, R=relative risk. Disease risk is 10%.

of up to 5–7% range and then declines. Figures 3 and 4 show that, for rare genotypes (less than 10%) and low relative risks (les than 1.2), there is not much difference between the AUC for multiplicative models and that for additive models. When relative risks are around 1.2–1.5, the largest difference in AUC for the two models is achieved when genotype frequency is around 50%. The AUCs for the 41 genetic variants illustrated in Figure 2 are 0.635 and 0.590 for the multiplicative and additive models, respectively. When the genotype frequency is increased to 0.3 for all the 41 risk genotypes, the AUCs for the multiplicative and additive models are 0.690 and 0.588, respectively.

The population attributable fractions (PAF) of the 40 risk genotypes for the additive and multiplicative models are also different.[14] For example, when R=1.2 and G=0.1, the PAF for the additive model is 0.444, whereas the PAF for the multiplicative model is 0.615. When the genotype frequency is increased to 0.2, the PAF values for additive and multiplicative models are 0.547 and 0.792, respectively.

Zheng *et al*[15] studied the genetic predisposition to prostate cancer by examining the association between prostate cancer and five SNPs that map to the three 8q24 loci, to 17q12, and to 17q24.3.

The genotype frequencies of the five SNPs were 0.3, 0.25, 0.07, 0.77, and 0.6, and the relative risks were 1.38, 1.28, 1.53, 1.37, and 1.22. The prevalence of diagnosed prostate cancer in the US adult population is about 1.6%, based on estimates from the National Health Interview Survey. As the true prevalence of prostate cancer is unknown, we assumed an upper bound of 3.2% for the prevalence of prostate cancer. We used the same methods developed in this study for identical risk genotypes to calculate AUC for these five SNPs after adjusting the probabilities calculated in the equations for sensitivity and specificity based on different genotype frequencies and relative risks. The AUCs for the multiplicative model and additive model were 0.569 and 0.541, respectively.

DISCUSSION

In this paper, we compared the clinical discriminative accuracy of a set of 40 polymorphisms in a genomic profile for multiplicative and additive models assuming identical risk and genotype frequency. We showed that multiplicative models yield more extreme risk estimates than additive models. The multiplicative models have higher AUC compared with additive models for the ranges of genotype frequency and relative risk considered in this study.

The difference between clinical discriminative accuracy for multiplicative models and additive models increased with increasing genotype frequency (up to 0.5) and increasing relative risk. There is no difference in clinical discriminative accuracy for the two models when genotype frequency is less than 0.1 and relative risk is around 1.1. The discriminative accuracy for the additive model increases slightly at first and then declines with increasing genotype frequency.

Several researchers have used logistic regression models to calculate the AUC for joint effects of multiple risk genotypes. When using logistic regression to calculate the AUC, a multiplicative (log additive) model for joint effects is assumed. In reality, however, we do not know the true underlying model for joint effects of genetic variants. If the true biology is additive or less than multiplicative, we overestimate the predictive accuracy by using multiplicative risk models, which particularly affects the extreme ends of the risk distribution. On the other hand, but this is hardly practised, if we apply additive models in instances in which the true biology is multiplicative or more than additive, we underestimate the predictive accuracy. If the relative risk is low (around 1.1) and the genotype frequency is less than 0.1, there would not be any difference in AUC for the two models. Therefore, for evaluating the predictive value of genomic profiling, which typically combines multiple weak susceptibility variants, the underlying risk model may not affect the discriminative accuracy so much, but it may affect the estimations of absolute risks.

Lund[16] compared additive and multiplicative models for reproductive risk factors and postmenopausal breast cancer. The author explored a relative risk function ranging from multiplicative to additive by changing the exponent in a power transformation and calculated the goodness-of-fit statistics for different power models. On the basis of this analysis, the author concluded that mathematical models close to being additive fitted slightly better than the multiplicative models. Sample sizes required in the baseline group of a cohort study and a case–control study to detect departure from a multiplicative model in the direction of an additive model and from an additive model in the direction of a multiplicative model for the joint effect of two binary risk factors have been studied.[17]

The mutually adjusted locus relative risks from a multilocus main effects logistic regression will, in principle, be smaller than the individual locus marginal relative risks when the true multilocus interaction model is additive. This may offset the overestimation of risks from the multiplicative interaction assumption.

We considered binary risk genotypes (dominant or recessive models) because the number of binary risk genotypes in the population has a binomial (K, G) distribution, where G is the genotype frequency and K is the number of genetic variants in the genomic profile. For codominant models, the risk of alleles within each locus could be additive or multiplicative. If we assume identical risk alleles and Hardy–Weinberg equilibrium within each locus, it can be shown that the distribution of risk alleles in K loci in the population has a binomial (2K, A) distribution, where A is the allele frequency.[18] Hence, our method can be easily extended to risk alleles when allele risks are either additive or multiplicative within and across biallelic loci.

In this paper, we clearly show the difference in AUC between additive and multiplicative models for genomic profiles. Our study is limited by the assumption of independence of the genetic variants, and our inability to model gene–gene interactions. More methodological work is needed in this area to detect the joint effect of multiple genetic variants for both additive and multiplicative models when gene–gene interactions are present.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by the Centre for Medical Systems Biology (CMSB) in the framework of The Netherlands Genomics Initiative (NGI). ACJW Janssens was sponsored by the VIDI Grant of The Netherlands Organisation for Scientific Research (NWO).

DISCLAIMER

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry.

1 Morris JK, Wald NJ: Graphical presentation of distributions of risk in screening. *J Med Screen* 2005; **12**: 155–160.
2 Woolas RP, Conaway M, Xu F *et al*: Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. *Gynecol Oncol* 1995; **59**: 111–116.
3 Wald NJ, Morris JK, Rish S: The efficacy of combining several risk factors as a screening test. *J Med Screen* 2005; **12**: 197–201.
4 Skates SJ, Horick N, Yu Y *et al*: Preoperative sensitivity and specificity for early stage ovarian cancer when combining cancer antigen CA-125II, CA 15-3, CA 72-4, and macrophage colony-stimulating factor using mixtures of multivariate normal distributions. *J Clin Oncol* 2004; **22**: 4059–4066.
5 Weedon MN, McCarthy MI, Hitman G *et al*: Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLOS Med* 2006; **3**: e374.
6 Podgoreanu MV, White WD, Morris RW *et al*: Inflammatory gene polymorphisms and risk of postoperative myocardial infarction after cardiac surgery. *Circulation* 2006; **114**: I275–I281.
7 Vaxillaire M, Veslot J, Dina C *et al*: Impact of common type 2 diabetes risk polymorphisms in the DESIR prospective study. *Diabetes* 2008; **57**: 244–254.
8 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.
9 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.
10 Peto J, Mack TM: High constant incidence in twins and other relatives of women with breast cancer. *Nat Genet* 2000; **26**: 411–414.
11 Hanley JA: Receiver operating characteristic (ROC) methodology: the state of the art. *Crit Rev Diagn Imaging* 1989; **29**: 307–335.
12 Pepe MS: Receiver operating characteristic methodology. *J Am Stat Assoc* 2000; **95**: 308–311.

13 Khoury MJ, Newill CA, Chase GA: Epidemiologic evaluation of screening for risk factors: application to genetic screening. *Am J Public Health* 1985; **75**: 1204–1208.
14 Moonesinghe R: A refinement to 'how many genes underlie the occurrence of common complex diseases in the population?'. *Int J Epidemiol* 2006; **35**: 497.
15 Zheng SL, Sun J, Wiklund F *et al*: Cumulative association of five genetic variants with prostate cancer. *N Eng J Med* 2008; **358**: 910–919.
16 Lund E: Comparison of additive and multiplicative models for reproductive risk factors and post-menopausal breast cancer. *Stat Med* 1995; **14**: 267–274.
17 Gonzalez ABD, Cox DR: Additive and multiplicative models for the joint effect of two risk factors. *Biostatistics* 2005; **6**: 1–9.
18 Wray NR, Goddard ME, Visscher PM: Prediction of individual genetic risk to disease from genome-wide association studies. *Genome Res* 2007; **17**: 1520–1528.

APPENDIX A
MULTIPLICATIVE MODELS

Let R be the relative risk and m the genotype frequency for each of the K risk genotypes. Let X be the genotype score in the genomic profile of K risk genotypes. Disease frequency p can be expressed as

$$p = \Pr[D^+] = \sum_{c=0}^{K} \Pr[D^+|X=c]\Pr[X=c] \tag{A1}$$

Consider the multiplicative risk model:

$$\Pr[D^+|X=\vec{c}] = IR^c \tag{A2}$$

where I is the risk for individuals not carrying risk variants or the background risk ($I=\Pr[D^+|X=0]$). Substituting $\Pr[X=c] = \binom{K}{i} m^c(1-m)^{(K-c)}$ and (A2) in (A1) gives:

$$\begin{aligned} p &= I\sum_{c=0}^{K}\binom{K}{c} R^c m^c (1-m)^{(K-c)} \\ &= I[Rm+(1-m)]^K \end{aligned} \tag{A3}$$

The sensitivity of the test for the multiplicative model is given by

$$\beta_{\mathrm{Mc}} = \Pr[X \geq c|D^+] = \frac{\sum_{i=c}^{K}\Pr[D^+|X=i]\Pr[X=i]}{p}$$

$$\beta_{\mathrm{Mc}} = \frac{I\sum_{i=c}^{K}\binom{K}{i} R^i m^i (1-m)^{K-i}}{p}$$

Substituting the expression for I from (A3) gives

$$\begin{aligned} &= \frac{\sum_{i=c}^{K}\binom{K}{i} R^i m^i (1-m)^{K-i}}{[Rm+(1-m)]^K} \\ &= \sum_{i=c}^{K}\binom{K}{i}\left(\frac{Rm}{Rm+(1-m)}\right)^i\left(1-\frac{Rm}{Rm+(1-m)}\right)^{K-i} \\ &= \sum_{i=c}^{K}\binom{K}{i}\beta_s^i(1-\beta_s)^{K-i} \end{aligned}$$

where $\beta_S = \frac{Rm}{[Rm+(1-m)]}$ is the sensitivity for screening for individual markers.

The disease risk when the genotype score is c in a genomic profile with K genetic variants using the multiplicative model is given by

$$\Pr[D^+|X=c] = IR^c = \frac{p * R^c}{[Rm+(1-m)]^K} \tag{A4}$$

APPENDIX B
FOR THE ADDITIVE RISK MODEL

$$\Pr[D^+|X=c] = I[1+c(R-1)]$$

and

$$\begin{aligned} p &= \sum_{c=0}^{K} I[1+c(R-1)]\binom{K}{c} m^c(1-m)^{(K-c)} \\ &= I[1+\mathrm{K}m(R-1)] \end{aligned}$$

The sensitivity for the test for the additive model is given by

$$\begin{aligned} \beta_{\mathrm{Ac}} &= \frac{\sum_{i=c}^{K}\Pr[D^+|X=i]\Pr[X=i]}{p} \\ &= \frac{\sum_{i=c}^{K}[1+i(R-1)]\binom{K}{i} m^i(1-m)^{(K-i)}}{[1+Km(R-1)]}, \end{aligned}$$

and α_{Ac} is obtained from Table A1 given below:

$$\alpha_{\mathrm{Ac}} = 1 - \frac{\sum_{i=c}^{K}\binom{K}{i} m^i(1-m)^{(K-i)} - P\beta_{\mathrm{A}}}{1-p}$$

The disease risk when the genotype score is c in a genomic profile with K genetic variants using the additive model is given by

$$\Pr[D^+|X=c] = I[1+c(R-1)] = \frac{p[1+c(R-1)]}{[1+Km(R-1)]}. \tag{B1}$$

Table A1 The relationship of K genetic markers sensitivity and specificity to marker and disease frequencies (additive models)

K Markers	*Disease*		
	Will develop disease	*Will not*	*Total*
More than c of K markers present	$p\beta_{\mathrm{Ac}}$	$(1-\alpha_{\mathrm{Ac}})(1-p)$	$\sum_{i=c}^{K}\binom{K}{i} m^i(1-m)^{(K-i)}$
At most c of K markers present	$(1-\beta_{\mathrm{Ac}})p$	$\alpha_{\mathrm{Ac}}(1-p)$	$1-\sum_{i=c}^{K}\binom{K}{i} m^i(1-m)^{(K-i)}$
Total	p	$1-p$	1