American Journal of Epidemiology
Vol. 152, No. 5
Printed in U.S.A.

Estimating Genotype Relative Risks in Case-Parental Control Studies: An Optimal Weighting Approach

Wen-Chung Lee and Chin-Hao Chang

The "case-parental control study" is a novel study design. It can quantify the relations between candidate genes and disease risk. Previous authors have proposed noniterative methods for estimating "genotype relative risks" (GRRs) in case-parental control studies. Here the authors propose yet another noniterative method. The new method is simple. It involves taking certain weighted averages with weights chosen according to one's educated guess about the likely values of the true GRRs. Monte Carlo simulation shows that the new estimators are approximately unbiased and that they have smaller variances than the previous estimators. *Am J Epidemiol* 2000;152:487–92.

case-control studies; epidemiologic methods; genotype; risk

Searching for disease susceptibility genes is the interest of geneticists and epidemiologists alike (1, 2). To this end, an association approach aimed at examining and quantifying the relation between candidate genes and disease risk can be useful. In particular, a novel study design, the "case-parental control" design (3–5), has come to much attention recently. It is a variant of a type of case-control study, the case-only study (6). The case-parental control study design requires the case group and their parents only, and can do without a control group entirely. This not only makes the design more cost-efficient but also eliminates the possibility of bias from inappropriate selection of controls whose genetic backgrounds differ systematically from those of cases (3–6).

Khoury (K) (3), Flanders and Khoury (FK) (4), and more recently Sun et al. (SFYK) (5) have proposed noniterative methods for estimating "genotype relative risks" (GRRs) in case-parental control studies. Their methods produce GRRs that are approximately unbiased, but the variances (instability) of different methods vary (K > FK > SFYK). In this paper, we propose yet another noniterative method of estimating GRRs. The new method is simple. It involves taking certain weighted averages with weights chosen according to one's educated guess about the likely values of the true GRRs. In other words, the method incorporates a priori information to optimize the estimation. We demonstrate using Monte Carlo simulation that the resulting estimators are approximately unbiased and have smaller variances than the K and FK estimators and even the SFYK estimator.

Received for publication May 7, 1999, and accepted for publication September 30, 1999.

Abbreviations: CPG, conditional on parental genotype; FK, Flanders and Khoury; GRR, genotype relative risk; K, Khoury; SFYK, Sun, Flanders, Yang, and Khoury.

From the Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan, Republic of China.

Reprint requests to Dr. Wen-Chung Lee, Graduate Institute of Epidemiology, National Taiwan University, No. 1, Jen-Ai Road, 1st Sec., Taipei, Taiwan, Republic of China (e-mail: wenchung@ha.mc.ntu.edu.tw).

BACKGROUND AND BASIC NOTATION

Assume that a candidate locus for disease susceptibility has alleles of M and N, with M the mutant (or high risk) allele and N the normal allele. In the case-parental control design, all newly diseased subjects (or a random sample of them) are recruited. These case subjects and their parents are genotyped at the candidate locus. There are three possible genotypes for each subject: NN, MN, and MM, denoted by $g = 0$, 1, and 2, respectively, corresponding to the number of mutant alleles in the genotype. Parental matings of the types NN × NN ($g = 0 \times g = 0$), NN × MM ($g = 0 \times g = 2$), and MM × MM ($g = 2 \times g = 2$) produce only one offspring genotype and are not informative. These noninformative "case-parent triads" are excluded. The other triads can be cross-tabulated according to the genotypes of case subjects and the genotypes of hypothetical control subjects carrying the nontransmitted parental alleles. We let m_{ij} ($i, j = 0$, 1, 2; $1 \leq i + j \leq 3$) denote the number of such informative case-parent triads, with $g = i$ in the case (transmitted) subject and $g = j$ in the control (nontransmitted) subject. Note that these notations are in accordance with those of Sun et al. (5).

Our interest here is to estimate the GRRs, based on the m_{ij}. Two such GRRs can be defined to describe the differential susceptibility to disease for the three genotypes. We let r_1 represent the relative risk for individuals with $g = 1$ versus those with $g = 0$ and r_2 the relative risk of $g = 2$ versus $g = 1$. Note that we did not use a common baseline to define the two GRRs. The r_1 and r_2 defined in this way characterize more succinctly the possible action modes of the candidate gene. If both are close to 1, we conclude that the gene is probably not associated with the disease. If $r_1 > 1$ and $r_2 \approx 1$, an autosomal dominant gene is suspected. If r_1 is

close to 1 but r_2 is not, the gene possibly conforms to an autosomal recessive model. If $r_1 \approx r_2 > 1$, a "gene-dose effect" (additive or codominant model) is demonstrated, in which disease susceptibility increases precisely in proportion to the number of mutant alleles involved.

The K, FK, and SFYK estimators mentioned above have the following general form:

$$\hat{r}_1 = \frac{w_1 m_{10} + m_{11}}{w_1 m_{01} + 2m_{02}}, w_1 \geq 0, \quad (1)$$

and

$$\hat{r}_2 = \frac{w_2 m_{21} + 2m_{20}}{w_2 m_{12} + m_{11}}, w_2 \geq 0. \quad (2)$$

Essentially, $\hat{r}_1$ is a weighted average of two ratios, m_{10}/m_{01} and $m_{11}/(2m_{02})$; and $\hat{r}_2$ is the weighted average of m_{21}/m_{12} and $2m_{20}/m_{11}$. By setting the weighting constants $w_1 = w_2 = 1$, we have the SFYK estimators. By setting $w_1 = (m_{02} + m_{11} + m_{20})/(m_{10} + m_{01})$ and $w_2 = (m_{02} + m_{11} + m_{20})/(m_{12} + m_{21})$, we obtain the FK estimators. By letting w_1 and w_2 both go to infinity, we have the K estimators.

Treating m_{ij} $(i, j = 0, 1, 2; 1 \leq i + j \leq 3)$ as a multinomial random vector and the w_1 and w_2 as fixed constants, the variance of the logarithm of $\hat{r}_1$ and $\hat{r}_2$ can be derived using the delta method (7):

$$\mathrm{Var}(\log \hat{r}_1) \approx \frac{w_1^2 m_{10} + m_{11}}{(w_1 m_{10} + m_{11})^2} + \frac{w_1^2 m_{01} + 4m_{02}}{(w_1 m_{01} + 2m_{02})^2} \quad (3)$$

and

$$\mathrm{Var}(\log \hat{r}_2) \approx \frac{w_2^2 m_{12} + m_{11}}{(w_2 m_{12} + m_{11})^2} + \frac{w_2^2 m_{21} + 4m_{20}}{(w_2 m_{21} + 2m_{20})^2}. \quad (4)$$

TABLE 1. Simulation results* for estimators of genotype relative risk, using an autosomal dominant gene ($r_1 = 5$, $r_2 = 1$), under different gene frequencies ($f = 0.2, 0.5, 0.8$)

Methods†	Log r_1				Log r_2			
	Bias	Variance	Coverage probability of 95% CI‡	Average length of 95% CI	Bias	Variance	Coverage probability of 95% CI	Average length of 95% CI
$f = 0.2$								
K	0.0464	0.1450	0.9629	1.4431	–0.0001	0.0026	0.9495	0.2002
FK	0.0177	0.0542	0.9481	0.8976	–0.0002	0.0053	0.9505	0.2864
SFYK	0.0156	0.0442	0.9494	0.8093	0.0000	0.0023	0.9494	0.1877
Suboptimal								
$r_1 = 1$, $r_2 = 3$	0.0145	0.0418	0.9494	0.7891	0.0000	0.0023	0.9484	0.1857
$r_1 = 3$, $r_2 = 1$	0.0142	0.0414	0.9494	0.7862	0.0000	0.0022	0.9497	0.1853
Optimal	0.0142	0.0414	0.9497	0.7861	0.0000	0.0022	0.9497	0.1853
CPG	0.0147	0.0407	0.9480	0.7790	0.0000	0.0022	0.9491	0.1852
$f = 0.5$								
K	0.0229	0.0663	0.9549	0.9995	0.0000	0.0212	0.9539	0.5685
FK	0.0176	0.0440	0.9493	0.8079	–0.0032	0.0141	0.9483	0.4610
SFYK	0.0195	0.0471	0.9497	0.8342	–0.0027	0.0133	0.9490	0.4484
Suboptimal								
$r_1 = 1$, $r_2 = 3$	0.0173	0.0431	0.9512	0.8024	–0.0023	0.0129	0.9503	0.4412
$r_1 = 3$, $r_2 = 1$	0.0168	0.0426	0.9512	0.7989	–0.0020	0.0127	0.9513	0.4394
Optimal	0.0167	0.0425	0.9516	0.7987	–0.0020	0.0127	0.9513	0.4394
CPG	0.0165	0.0422	0.9519	0.7962	–0.0019	0.0127	0.9505	0.4377
$f = 0.8$								
K	0.0075	0.0162	0.9474	0.4970	–0.0007	0.0897	0.9540	1.1649
FK	0.0073	0.0267	0.9487	0.6377	0.0000	0.0277	0.9477	0.6429
SFYK	0.0050	0.0149	0.9516	0.4794	–0.0006	0.0242	0.9515	0.6072
Suboptimal								
$r_1 = 1$, $r_2 = 3$	0.0053	0.0142	0.9523	0.4670	–0.0005	0.0238	0.9507	0.6011
$r_1 = 3$, $r_2 = 1$	0.0054	0.0142	0.9505	0.4658	–0.0003	0.0237	0.9501	0.5993
Optimal	0.0055	0.0142	0.9506	0.4657	–0.0003	0.0237	0.9501	0.5993
CPG	0.0055	0.0141	0.9504	0.4655	–0.0013	0.0232	0.9513	0.5933

* Based on 10,000 simulations.
† K, Khoury's method; FK, Flanders and Khoury's method; SFYK, method of Sun, Flanders, Yang, and Khoury; CPG, conditional on parental genotype method.
‡ CI, confidence interval.

THE OPTIMAL WEIGHTING APPROACH

We show in the Appendix that the variances of the logarithm of $\hat{r}_1$ and $\hat{r}_2$ are minimized, if the weighting constants can be chosen to be $w_1 = 1 + r_1/(r_1 + 1)$ and $w_2 = 1 + 1/(r_2 + 1)$. Estimation of GRRs using these optimized constants will be unbiased and most precise. The above equations imply that the optimal weightings always lie between 1 and 2. This effectively rules out the K estimators, which have weighting constants of infinity, and the FK estimators, which do not guarantee weighting constants that are between 1 and 2. As for the SFYK estimators, we see that they are optimal only when $r_1 = 0$ and $r_2 = \infty$, that is, when the homozygotes (NN and MM) are prone to disease and the heterozygotes (MN) are immune to disease.

In practice, r_1 and r_2 are unknown and the optimal weighting constants cannot be determined in advance. However, one can make an approximate guess about the two GRRs. If previous studies have suggested that the gene under study is only weakly associated with the disease (r_1, $r_2 \approx 1$), we can use $w_1 = w_2 = 1.5$. If the gene is suspected of being autosomal dominant, we assign a suitable value between 1.5 and 2 for w_1 and a value of 1.5 for w_2. If the gene is suspected of being autosomal recessive, we assign 1.5 for w_1 and a value between 1 and 1.5 for w_2. For situations in which the gene is not clearly dominant or recessive, a reasonable choice would be to set w_1 slightly above 1.5 and w_2 slightly below 1.5.

SIMULATION STUDIES

In this section, we perform a simulation study on the statistical properties of the K, FK, SFYK, and optimal weighting methods. For comparison, we also present simulation results for the "conditional on parental genotype" (CPG) method (8). This likelihood-based method requires an iterative algorithm to obtain the estimates, yet standard likelihood theory predicts that it will have maximal stability. To simplify the presentation, we adopt an approach similar to that of Sun et al. (5); i.e., we assume a Hardy-Weinberg equilibrium and random mating in the parental

TABLE 2. Simulation results* for estimators of genotype relative risk, using an autosomal recessive gene ($r_1 = 1$, $r_2 = 5$), under different gene frequencies ($f = 0.2, 0.5, 0.8$)

Methods†	Log r_1				Log r_2			
	Bias	Variance	Coverage probability of 95% CI‡	Average length of 95% CI	Bias	Variance	Coverage probability of 95% CI	Average length of 95% CI
$f = 0.2$								
K	–0.0041	0.2457	0.9605	1.8762	0.0024	0.0079	0.9506	0.3462
FK	–0.0010	0.1118	0.9438	1.2682	0.0027	0.0143	0.9492	0.4700
SFYK	0.0083	0.0603	0.9517	0.9554	0.0016	0.0065	0.9497	0.3146
Suboptimal								
$r_1 = 1$, $r_2 = 3$	0.0067	0.0589	0.9534	0.9428	0.0017	0.0065	0.9501	0.3142
$r_1 = 3$, $r_2 = 1$	0.0060	0.0593	0.9538	0.9450	0.0018	0.0066	0.9511	0.3153
Optimal	0.0067	0.0589	0.9534	0.9428	0.0017	0.0065	0.9504	0.3141
CPG	0.0113	0.0505	0.9510	0.8674	0.0019	0.0064	0.9492	0.3122
$f = 0.5$								
K	0.0014	0.1108	0.9554	1.2940	0.0197	0.0669	0.9527	0.9976
FK	0.0070	0.0666	0.9527	1.0021	0.0047	0.0374	0.9526	0.7518
SFYK	0.0116	0.0681	0.9499	1.0115	0.0052	0.0348	0.9536	0.7264
Suboptimal								
$r_1 = 1$, $r_2 = 3$	0.0084	0.0650	0.9537	0.9919	0.0058	0.0347	0.9526	0.7246
$r_1 = 3$, $r_2 = 1$	0.0074	0.0652	0.9538	0.9946	0.0065	0.0353	0.9500	0.7296
Optimal	0.0084	0.0650	0.9537	0.9919	0.0056	0.0347	0.9525	0.7242
CPG	0.0098	0.0609	0.9529	0.9628	0.0062	0.0322	0.9517	0.6977
$f = 0.8$								
K	–0.0021	0.0130	0.9492	0.4435	0.0354	0.1407	0.9565	1.4201
FK	0.0012	0.0144	0.9492	0.4660	0.0038	0.0337	0.9462	0.7075
SFYK	–0.0004	0.0114	0.9514	0.4156	0.0022	0.0284	0.9511	0.6557
Suboptimal								
$r_1 = 1$, $r_2 = 3$	–0.0010	0.0111	0.9493	0.4103	0.0022	0.0284	0.9494	0.6547
$r_1 = 3$, $r_2 = 1$	–0.0012	0.0111	0.9501	0.4108	0.0024	0.0287	0.9484	0.6584
Optimal	–0.0010	0.0111	0.9493	0.4103	0.0022	0.0283	0.9502	0.6544
CPG	–0.0008	0.0110	0.9511	0.4086	0.0015	0.0234	0.9501	0.5970

* Based on 10,000 simulations.
† K, Khoury's method; FK, Flanders and Khoury's method; SFYK, method of Sun, Flanders, Yang, and Khoury; CPG, conditional on parental genotype method.
‡ CI, confidence interval.

population in the simulation. We emphasize again that the results should hold even if these conditions are not met. The simulation considers an autosomal dominant gene ($r_1 = 5$, $r_2 = 1$), an autosomal recessive gene ($r_1 = 1$, $r_2 = 5$), and a gene with a gene-dose effect ($r_1 = 5$, $r_2 = 5$) under gene frequencies (f, the prevalence of allele N) of 0.2, 0.5, and 0.8, respectively. For each situation, 10,000 simulations are performed. We chose a sample size (the number of informative case-parent triads) such that the minimum expected value of m_{ij} would be about 10 for each situation.

Table 1 presents the simulation results for the autosomal dominant gene. It can be seen that all of the methods considered yield log GRRs that are approximately unbiased, though the log r_1 estimation is slightly above the true value of log r_1. For the stability of the point estimates, one finds, as expected, that the CPG method has the smallest variance. However, the optimal weighting method as proposed in this paper produces the most stable estimates among the noniterative methods: It has smaller variances than the K and FK methods, and even the SFYK method. Since the true GRRs are unknown in practice and the weighting constants must be based on guesswork, our method may become "suboptimal" when certain incorrect values of r_1 and r_2 are selected. We simulated two cases of incorrect specification: $r_1 = 1$, $r_2 = 3$ and $r_1 = 3$, $r_2 = 1$. The former case corresponds to wrongly assuming a recessive model, and the latter to wrongly assuming a dominant model with underestimated effects. We see that such "suboptimal methods" are still better than the K, FK, and SFYK methods. Table 1 also presents the coverage probabilities and the average lengths of the 95 percent confidence intervals for the various methods. The coverages are close to the nominal 95 percent for all of the methods considered. We notice again that the average length for the optimal weighting method is shortest among the noniterative methods.

Tables 2 and 3 present, respectively, the simulation results for the autosomal recessive gene and the gene with a gene-dose effect. The basic findings are similar to those for the autosomal dominant gene, though the superiority of the optimal weighting method over the SFYK method is not as striking.

TABLE 3. Simulation results* for estimators of genotype relative risk, using a gene with a gene-dose effect ($r_1 = 5$, $r_2 = 5$), under different gene frequencies ($f = 0.2, 0.5, 0.8$)

Methods†	Log r_1				Log r_2			
	Bias	Variance	Coverage probability of 95% CI‡	Average length of 95% CI	Bias	Variance	Coverage probability of 95% CI	Average length of 95% CI
$f = 0.2$								
K	0.0480	0.1522	0.9592	1.4562	0.0003	0.0016	0.9482	0.1558
FK	0.0260	0.0814	0.9463	1.0881	0.0001	0.0031	0.9483	0.2184
SFYK	0.0152	0.0434	0.9535	0.8138	0.0001	0.0013	0.9481	0.1417
Suboptimal								
$r_1 = 3$, $r_2 = 8$	0.0139	0.0407	0.9526	0.7907	0.0001	0.0013	0.9490	0.1414
$r_1 = 8$, $r_2 = 3$	0.0138	0.0407	0.9522	0.7907	0.0001	0.0013	0.9497	0.1414
Optimal	0.0138	0.0407	0.9520	0.7906	0.0001	0.0013	0.9490	0.1414
CPG	0.0143	0.0393	0.9521	0.7758	0.0001	0.0013	0.9488	0.1411
$f = 0.5$								
K	0.0216	0.0163	0.9585	0.9684	0.0054	0.0118	0.9530	0.4271
FK	0.0146	0.0412	0.9551	0.7919	0.0018	0.0070	0.9494	0.3290
SFYK	0.0161	0.0443	0.9502	0.8082	0.0021	0.0064	0.9512	0.3136
Suboptimal								
$r_1 = 3$, $r_2 = 8$	0.0142	0.0398	0.9531	0.7740	0.0021	0.0063	0.9514	0.3128
$r_1 = 8$, $r_2 = 3$	0.0142	0.0397	0.9528	0.7739	0.0023	0.0063	0.9513	0.3128
Optimal	0.0142	0.0397	0.9532	0.7738	0.0022	0.0063	0.9514	0.3127
CPG	0.0140	0.0392	0.9510	0.7690	0.0021	0.0062	0.9541	0.3095
$f = 0.8$								
K	0.0055	0.0161	0.9490	0.4922	0.0148	0.0539	0.9494	0.8886
FK	0.0051	0.0165	0.9480	0.4958	0.0023	0.0128	0.9491	0.4429
SFYK	0.0046	0.0150	0.9470	0.4749	0.0018	0.0113	0.9545	0.4183
Suboptimal								
$r_1 = 3$, $r_2 = 8$	0.0044	0.0142	0.9458	0.4615	0.0018	0.0112	0.9542	0.4176
$r_1 = 8$, $r_2 = 3$	0.0045	0.0142	0.9460	0.4615	0.0018	0.0112	0.9540	0.4177
Optimal	0.0045	0.0142	0.9460	0.4614	0.0018	0.0112	0.9537	0.4175
CPG	0.0045	0.0142	0.9453	0.4610	0.0014	0.0107	0.9528	0.4079

* Based on 10,000 simulations.

† K, Khoury's method; FK, Flanders and Khoury's method; SFYK, method of Sun, Flanders, Yang, and Khoury; CPG, conditional on parental genotype method.

‡ CI, confidence interval.

DISCUSSION

In this paper, we present a new approach to estimating the GRRs for a disease susceptibility gene. The new method produces GRRs that are approximately unbiased and that have variances smaller than those of the other noniterative methods proposed to date. The method requires a priori information about the likely values of the GRRs. However, this is not crucial. To simplify matters and to be on the safe side, one can set w_1 slightly above 1.5 and w_2 slightly below 1.5. The results, though less optimal, are still better than those of the other noniterative methods.

Recently, it was discovered that the likelihood-based CPG method is identical to the use of a log-linear Poisson regression model and that it can therefore be implemented using standard software (9, 10). However, it still requires some computational efforts (feeding the data into a computer, specifying appropriate program codes for analysis, etc.). By contrast, the calculation in our method is so simple that it can be done with pencil and paper. This could be useful for a speedy initial assessment. The estimates derived from our method can also serve as reasonable starting values for Poisson regression analysis.

In this paper, the variance formulae are used to construct the confidence intervals around the point estimates. However, they can also be used to perform hypothesis testing about the GRRs. This implies that the optimal weighting approach as proposed in this paper may lead to an alternative "transmission/disequilibrium test" (11). Such a test is particularly useful when one is dealing with a "marker gene" rather than a "susceptibility gene." In such situations, one's interest lies in whether or not the "marker gene" under study is in linkage disequilibrium with the disease gene, rather than in the magnitude of the GRRs of the marker gene itself. The prospect of an optimal weighting approach to transmission/disequilibrium testing is currently under investigation.

ACKNOWLEDGMENTS

This study was partly supported by a grant from the Taiwan National Science Council.

REFERENCES

1. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–17.
2. Khoury MJ, Yang Q. The future of genetic studies of complex human diseases: an epidemiologic perspective. Epidemiology 1998;9:350–4.
3. Khoury MJ. Case-parental control method in the search for disease-susceptibility genes. (Letter). Am J Hum Genet 1994; 55:414–15.
4. Flanders WD, Khoury MJ. Analysis of case-parental control studies: method for the study of associations between disease and genetic markers. Am J Epidemiol 1996;144:696–703.
5. Sun F, Flanders WD, Yang Q, et al. A new method for estimating the risk ratio in studies using case-parental control design. Am J Epidemiol 1998;148:902–9.
6. Greenland S. A unified approach to the analysis of case-distribution (case-only) studies. Stat Med 1999;18:1–15.
7. Agresti A. Categorical data analysis. New York, NY: John Wiley and Sons, Inc, 1990.
8. Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. Am J Hum Genet 1993;53:1114–26.
9. Wilcox AJ, Weinberg CR, Lie RT. Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads." Am J Epidemiol 1998;148:893–901.
10. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 1998;62:969–78.
11. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993; 52:506–16.

APPENDIX

To obtain the optimal weighting, we differentiate equation 3 in the text with respect to w_1 and equation 4 with respect to w_2, and equate them to zero. That is,

$$\frac{d[\mathrm{Var}(\log \hat{r}_1)]}{dw_1} = 2 \times \left(\frac{w_1 m_{10} m_{11} - m_{10} m_{11}}{(w_1 m_{10} + m_{11})^3} + \frac{2 w_1 m_{01} m_{02} - 4 m_{01} m_{02}}{(w_1 m_{01} + 2 m_{02})^3} \right) = 0$$

and

$$\frac{d[\mathrm{Var}(\log \hat{r}_2)]}{dw_2} = 2 \times \left(\frac{w_2 m_{11} m_{12} - m_{11} m_{12}}{(w_2 m_{12} + m_{11})^3} + \frac{2 w_2 m_{20} m_{21} - 4 m_{20} m_{21}}{(w_2 m_{21} + 2 m_{20})^3} \right) = 0.$$

After rearrangement, we have

$$w_1 = \frac{m_{10} m_{11} (w_1 m_{01} + 2 m_{02})^3 + 4 m_{01} m_{02} (w_1 m_{10} + m_{11})^3}{m_{10} m_{11} (w_1 m_{01} + 2 m_{02})^3 + 2 m_{01} m_{02} (w_1 m_{10} + m_{11})^3} \quad (5)$$

and

$$w_2 = \frac{m_{11}m_{12}(w_2m_{21} + 2m_{20})^3 + 4m_{20}m_{21}(w_2m_{12} + m_{11})^3}{m_{11}m_{12}(w_2m_{21} + 2m_{20})^3 + 2m_{20}m_{21}(w_2m_{12} + m_{11})^3}. \quad (6)$$

The expectations (E) of the m_{ij} are proportional to the "probability of mating" (the mating probability of any two given genotypes in the population) and the GRRs. That is,

$$E(m_{01}) \propto h_1/2, \qquad E(m_{02}) \propto h_2/4,$$
$$E(m_{10}) \propto h_1r_1/2, \qquad E(m_{11}) \propto h_2r_1/2,$$
$$E(m_{12}) \propto h_3r_1/2, \qquad E(m_{20}) \propto h_2r_2/4,$$
$$E(m_{21}) \propto h_3r_2/2.$$

h_k ($k = 1, 2, 3$) is the probability of mating: h_1 is the probability of mating between genotype MN and genotype NN (MN × NN); h_2 is the probability of MN × MN; and h_3 is the probability of MN × MM. r_1 and r_2 are the GRRs, as defined in the text. Replacing m_{ij} in equations 5 and 6 with its corresponding expected value and rearranging and canceling out h_k, we arrive at

$$w_1 = 1 + \frac{r_1}{r_1 + 1}$$

and

$$w_2 = 1 + \frac{1}{r_2 + 1}.$$