## Text S1

## Variance of genetic marker in a case/control sample

Let  be the number of minor alleles at a biallelic marker with minor allele frequency *,* and  be its standardised value. Under Hardy-Weinberg Equilibrium (HWE) the population distribution of  is



Let  be the genetic effect on liability,  be the trait prevalence,  and when liability exceeds , 0 otherwise. The genotype frequencies in cases are







Similarly the genotype frequencies in controls are







The expectation of  in the case/control sample is then



where  is the sampling fraction of cases and  denotes ascertainment, and the mean of  is



The expected variance over a set of markers is obtained by integrating  over the normal distribution of . Values for a range of genetic variances and allele frequencies are given in tables A1 and A2. Except for scenarios in which rare markers explain a high proportion of liability, the variance of  in the case/control sample is close to 1.

Table A1. Expected variance of genetic predictor in a case/control sample with equal number of cases and controls when the population prevalence is 0.1.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | |  | | | |
| 1/10 | 1/100 | 1/1000 | 1/10,000 |
|  | 0.01 | 1.103 | 1.001 | 1.000 | 1.000 |
| 0.05 | 1.023 | 1.000 | 1.000 | 1.000 |
| 0.1 | 1.012 | 1.000 | 1.000 | 1.000 |
| 0.3 | 1.004 | 1.000 | 1.000 | 1.000 |
| 0.5 | 1.003 | 1.000 | 1.000 | 1.000 |

Table A2. Expected variance of genetic predictor in a case/control sample with equal number of cases and controls when the population prevalence is 0.01.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | |  | | | |
| 1/10 | 1/100 | 1/1000 | 1/10,000 |
|  | 0.01 | 1.509 | 1.005 | 1.000 | 1.000 |
| 0.05 | 1.092 | 1.001 | 1.000 | 1.000 |
| 0.1 | 1.045 | 1.000 | 1.000 | 1.000 |
| 0.3 | 1.015 | 1.000 | 1.000 | 1.000 |
| 0.5 | 1.012 | 1.000 | 1.000 | 1.000 |