

Excerpt of Results From the Report Genetic Risk Analysis and Epistasis

July 20, 2015

General Formula for the Odds Ratio

The OR is a measure of how much more likely an individual with a 1 at the first locus in their genotype is to have a specified condition than an individual with a 0. $OR = \frac{P(C=1\text{-and-}g_1=1)P(C=0\text{-and-}g_1=0)}{P(C=1\text{-and-}g_1=0)P(C=0\text{-and-}g_1=1)}$.

1 Block Case

Theorem For a biallelic l -locus system, assume that an individual will have a condition iff the individual has more than one 1 in their genotype. The following is a general formula for the Odds Ratio

$$\frac{[(1 - (p^{l-1})) (1 - p)] [(1 - p)(l - 1)(p)^{l-1} + p^l]}{[p - (1 - p)(l - 1)p^{l-1} - p^l] [(1 - p)(p)^{l-1}]}$$

Proof We know that for l loci, $P(C = 0) = p^l + l(1 - p)(p)^{l-1}$. We also know that for a 1-block model, there is only one way an individual can have a 1 at the first locus and not have the condition, so $P(C = 0\text{-and-}g_1 = 1) = (1 - p)(p)^{l-1}$.

Similarly, we know the possible ways for an individual to not have the condition and not have a 1 at the first locus. So $P(C = 0\text{-and-}g_1 = 0) = (1 - p)(l - 1)(p)^{l-1} + p^l$. This is because in order to not have the condition, an individual may have a 0 at all loci, or may have one 1 at any of $l - 1$ loci ($(l - 1)$ because we have already determined that the individual cannot have a 1 at the first locus).

We know that we can use the definitions of probabilities to solve for the following:

$$P(C = 1\text{-and-}g_1 = 1) = P(C = 1|g_1 = 1)P(g_1 = 1) =$$

$$[1 - P(C = 0|g_1 = 1)]P(g_1 = 1) = \left(1 - \frac{P(C = 0\text{-and-}g_1 = 1)}{P(g_1 = 1)}\right) P(g_1 = 1).$$

We know from our previous computation what this expression is in terms of p and l . So, plugging these in, we get $P(c = 1 - \text{and} - g_1 = 1) = \left(1 - \frac{(1-p)(p^{l-1})}{(1-p)}\right) (1-p) = (1 - (p^{l-1})) (1-p)$.

Similarly, by definition, $P(C = 1 - \text{and} - g_1 = 0) = \left(1 - \frac{P(C=0 - \text{and} - g_1=0)}{P(g_1=0)}\right) P(g_1 = 0)$. We know these values in terms of p and l , so plugging these in, we get: $P(C = 1 - \text{and} - g_1 = 0) = \left(1 - \frac{(1-p)(l-1)p^{l-1} + p^l}{p}\right) p = p - (1-p)(l-1)p^{l-1} - p^l$.

Plugging into the definition of OR, we get

$$OR = \frac{P(C = 1 - \text{and} - g_1 = 1)P(C = 0 - \text{and} - g_1 = 0)}{P(C = 1 - \text{and} - g_1 = 0)P(C = 0 - \text{and} - g_1 = 1)} = \frac{[(1 - (p^{l-1})) (1-p)] [(1-p)(l-1)(p^{l-1} + p^l)]}{[p - (1-p)(l-1)p^{l-1} - p^l] [(1-p)(p)^{l-1}]}$$

2 Block Case

Theorem For a biallelic l -locus, 2 block system (each block with $l/2$ loci), an individual will have a condition iff they have more than one 1 in either or both of the blocks in the genotype. The following is a general formula for the Odds Ratio

$$\frac{[(1-p) - (1-p)(p)^{l-1} - (l/2)(1-p)^2(p)^{l-2}] [p^l + (l-1)(1-p)(p)^{l-1} + (l/2-1)(l/2)(1-p)^2(p)^{l-2}]}{[p - p^l - (l-1)(1-p)(p)^{l-1} - (l/2-1)(l/2)(1-p)^2(p)^{l-2}] [(1-p)(p)^{l-1} + (l/2)(1-p)^2(p)^{l-2}]}$$

Proof We know that the probability of having the a 1 at the first locus and not having the condition is $P(C = 0 - \text{and} - g_1 = 1) = (1-p)(p)^{(l/2-1)}(p)^{l/2} + (1-p)(p)^{(l/2-1)}(1-p)(l/2)(p)^{l/2-1} = (1-p)(p)^{l-1} + (l/2)(1-p)^2(p)^{l-2}$.

Similarly $P(C = 0 - \text{and} - g_1 = 0) = p^l + p^{l/2}(l-1)(1-p)(p)^{(l/2-1)} + p^{l/2-1}(l/2-1)(l/2)(1-p)^2(p)^{l/2-1} = p^l + (l-1)(1-p)(p)^{l-1} + (l/2-1)(l/2)(1-p)^2(p)^{l-2}$.

We also know that

$$P(C = 1 - \text{and} - g_1 = 1) = P(C = 1 | g_1 = 1)P(g_1 = 1) = [1 - P(C = 0 | g_1 = 1)]P(g_1 = 1) =$$

$$\left[1 - \frac{P(C = 0 - \text{and} - g_1 = 1)}{P(g_1 = 1)}\right] P(g_1 = 1) = \left[1 - \frac{(1-p)(p)^{l-1} + (l/2)(1-p)^2(p)^{l-2}}{(1-p)}\right] (1-p)$$

We can simplify this to $P(C = 1 - \text{and} - g_1 = 1) = (1-p) - (1-p)(p)^{l-1} - (l/2)(1-p)^2(p)^{l-2}$.

Similarly

$$P(C = 1 - \text{and} - g_1 = 0) = \left[1 - \frac{P(C = 0 - \text{and} - g_1 = 0)}{P(g_1 = 0)}\right] P(g_1 = 0) =$$

$$\left[1 - \frac{p^l + (l-1)(1-p)(p)^{l-1} + (l/2-1)(l/2)(1-p)^2(p)^{l-2}}{p} \right] p =$$

$$p - p^l - (l-1)(1-p)(p)^{l-1} - (l/2-1)(l/2)(1-p)^2(p)^{l-2}.$$

So a general formula for the OR can be written as

$$OR = \frac{[P(C = 1 - \text{and} - g_1 = 1)] [P(C = 0 - \text{and} - g_1 = 0)]}{[P(C = 1 - \text{and} - g_1 = 0)] [P(C = 0 - \text{and} - g_1 = 1)]} =$$

$$\frac{[(1-p) - (1-p)(p)^{l-1} - (l/2)(1-p)^2(p)^{l-2}] [p^l + (l-1)(1-p)(p)^{l-1} + (l/2-1)(l/2)(1-p)^2(p)^{l-2}]}{[p - p^l - (l-1)(1-p)(p)^{l-1} - (l/2-1)(l/2)(1-p)^2(p)^{l-2}] [(1-p)(p)^{l-1} + (l/2)(1-p)^2(p)^{l-2}]}$$

Results from OR calculations with different block structures

The following are the results for inheritance and OR, based on a fixed probability with different block divisions of a set amount of loci.

6 Loci-fixed P(C=1)

	1 Block	2 Block	3 Block
Inheritance	25%	25%	25%
OR	≈ 113	≈ 70	≈ 50
P(C=1)-fixed	0.0003	0.0003	0.0003

We can see from these calculations that as we introduce more epistasis by creating more blocks, the OR gets smaller. Despite the fact that the total proportion of individuals with the condition and the inheritance do not change, the OR changes.

Note: For more background and detail, see the report Genetic Risk Analysis and Epistases by Eve Young, supervisor Dr. Kristina Crona.