Risk parametrization, interpretation, and choice of reference category

A nice feature with the case-parent triad design is that one can estimate the relative risk (RR) associated with the various haplotypes, and this is done in Haplin. As always, when computing relative risks one needs to pick a suitable reference category. Haplotypes by necessity always appear in pairs, and there is a question about how the two haplotypes interact within an individual. This makes the choice of reference category and of interaction model somewhat non-trivial. Perhaps the most straightforward choice would be to pick, say, homozygous individuals with haplotypes h_1h_1 as the reference group, and assign separate parameters to all other possible haplotype combinations h_ih_j. However, when the number of possible haplotypes in an analysis increases there are too many possible haplotype combinations, and it is necessary to constrain the number of parameters used to model risk.

The multiplicative risk model (dose-response)

A model commonly used to control the number of parameters is the multiplicative risk model (dose-response model):

$$P(D|h_ih_j) = RR_i \cdot RR_j \cdot B,$$

where $P(D|h_ih_j)$ denotes the risk of disease for a child with haplotype combination h_ih_j (for single markers the h_is denote alleles), RR_i is the relative risk associated with haplotype i, and B is a baseline risk level. If there are K haplotypes then one of $RR_1, RR_2, \ldots, RR_K$ is redundant. An appropriate choice of reference removes this redundancy. In the following we will discuss the three different ways in which Haplin can choose its reference.

1. **Reference category**

   A standard way of removing the redundancy is to set $RR_1 = 1$, which turns haplotype h_1 into a reference haplotype, and B is then the risk of disease for a h_1h_1 homozygous individual. RR_i is then the relative risk of disease for an individual with haplotype h_i relative to the reference haplotype. Note that in this situation we can talk about a reference haplotype h_1 *and* the corresponding reference group of individuals homozygous in h_1.

   The syntax for choosing "reference category" as the reference method is haplin(....., reference = "ref.cat",...), which automatically chooses

the most frequent haplotype as reference. To pick a specific haplotype as reference, use the syntax, say, haplin(....., reference = 2,...), which picks the second haplotype as reference. Since Haplin removes rare haplotypes during a run (as determined by the threshold argument), you may have to run Haplin once first to see what haplotypes remain in the analysis, and the run it once more with, for instance, reference = 2. Haplin will then use the second of the remaining haplotypes as reference. Note: it is always a good idea to make sure you don't use a rare haplotype as reference, since this will result in wide confidence intervals and uncertain estimates.

2. **Population reference**

   As an alternative to picking a specific haplotype as reference, one can use a population reference, i.e. the baseline risk is the average population risk. The interpretation of the relative risks RR_i is then how much the presence of a single haplotype h_i in an individual increases (or decreases) the disease risk of that individual, relative to the population. Note that in this situation there is no reference haplotype, the reference is the average risk level in the population.

   To make Haplin use "population reference", use haplin(...., reference = "population",...)

3. **Reciprocal reference**

   A third way of choosing reference is to use, for any haplotype, all remaining haplotypes as reference. That is, when estimating the relative risk RR_i associated with haplotype h_i, we use all other haplotypes h_j, $j \neq i$, as reference. The interpretation of RR_i is then, roughly speaking, the relative increase in risk observed when starting with an individual with no h_i haplotype and replacing one of his/her haplotypes with an h_i haplotype.

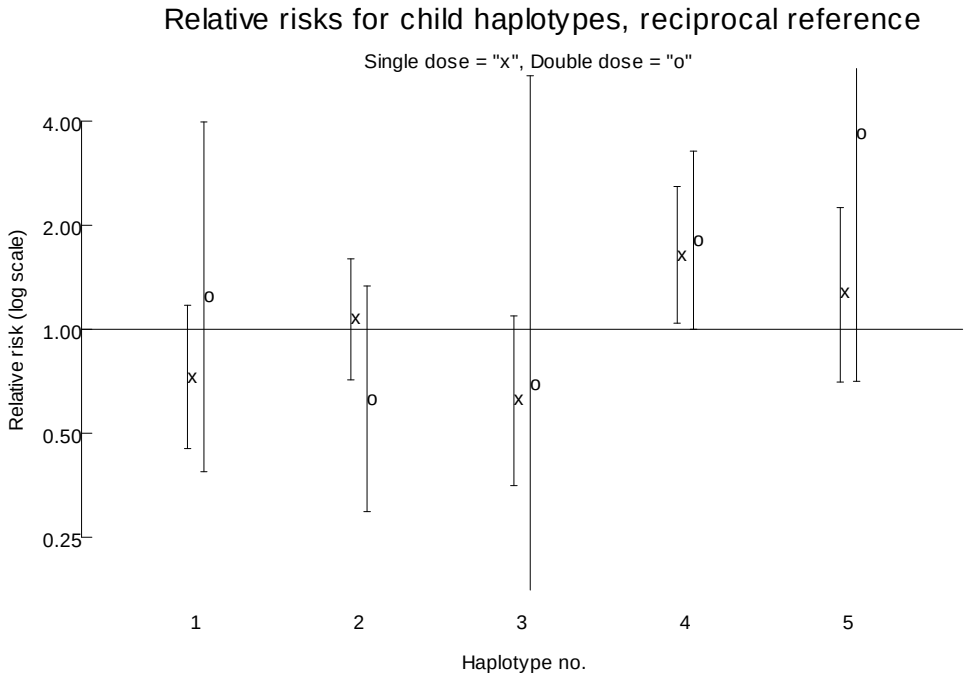
   Since "reciprocal reference" is the default in Haplin, no value is needed for the argument "reference" here.

Of these three, the third is chosen as default in Haplin. While "reference category" may be the most attractive for the multiplicative model, "reciprocal reference" has some advantages when estimating both single- and double-dose risk, as discussed below.

Single- and double-dose risk estimates

The multiplicative parametrization is convenient. However, for some data there is enough information to estimate both single- and double-dose effects, if the homozygotes are sufficiently frequent. HAPLIN incorporates this possibility.

The figure below shows an example plot of the results from HAPLIN.



Reciprocal reference has been used. We see, for instance, that the risk associated with haplotype 4 is (borderline) significantly increased relative to the rest of the haplotypes ("x" = single dose). The homozygotes with haplotype 4 have approximately the same risk as the single dose; this suggests a dominant effect of haplotype 4 ("o" = double dose). If the double dose effect is larger than the single dose effect, there may be a dose-response relationship. If the single dose has no effect, whereas the double dose *has* an effect, there is a suggestion of a recessive effect. This seems to be the case for haplotype 5, where the double dose effect is quite large, whereas the single dose is almost neutral. However, it is often the case that the double-dose estimates have large standard errors since homozygotes may be rare.

It should be kept in mind that dominance- and recessive effects must always be seen in relation to the reference. If only a single haplotype is

deleterious (or beneficial) against the others, HAPLIN should give a fairly good picture of this. More complex interaction patterns between haplotypes may not be completely correctly represented. Also, the power of separating different patterns of inheritance may be small since homozygotes may be rare.

Haplin estimates the double-dose effects by assuming a standard multiplicative model for the single-dose effects (multiplicative interactions between different haplotypes) and then estimating all the double-dose (homozygote) effects by separate parameters. Returning to the issue of choosing reference type, it should be noted that while picking a single reference haplotype may be convenient for the multiplicative model, an issue arises with the double-dose model. In the multiplicative model a reference haplotype corresponds directly to a group of reference individuals, namely those homozygous in the reference haplotype. However, in the double-dose model, picking a single reference haplotype will result in an artificial reference group of individuals having two single doses of the reference haplotype. The reference risk will thus differ from those having a genuine double-dose, i.e. the homozygotes, which is rather confusing. Alternatively, one might pick homozygotes with that specific haplotype as the reference. However, since homozygotes often are rare in the data, this may produce an unstable reference and wide confidence intervals. For this reason, Haplin uses the reciprocal reference as default.