SUPPLEMENTARY MATERIALS

Inference of the presence or absence of bias of odds ratios (ORs) of marker-secondary phenotype association in a case-control study design frequency-matched with respect to the secondary phenotype

Recall that we denoted three binary random variables, *X*, *T*, and *Y*, for the marker, the secondary phenotype, and the primary disease, respectively. Following the proof of [Kraft, 2007; Monsees et al., 2009], we let *A* be another binary random variable for ascertainment indicator, with 1 representing a subject selected to be in the case-control study and 0 representing otherwise. Therefore, for a case-control study where the controls were frequency-matched with the cases on the basis of the secondary phenotype, we assumed that ascertainment is independent of the marker conditional on the primary disease and the secondary phenotype. Therefore, the estimated OR from the data will be:

$$\begin{aligned} OR &= \frac{P(T=1 \mid X=1, A)P(T=0 \mid X=0, A)}{P(T=0 \mid X=1, A)P(T=1 \mid X=0, A)} \\ &= \frac{P(T=1 \mid X=1)P(T=0 \mid X=0)}{P(T=0 \mid X=1)P(T=1 \mid X=0)} \times \psi \quad , \\ &= OR_E \times \psi \end{aligned}$$

where OR_E is the expected OR estimate, and ψ is calculated as

$$\psi = \frac{\sum_Y P(A \mid Y, T=1)P(Y \mid T=1, X=1)}{\sum_Y P(A \mid Y, T=0)P(Y \mid T=0, X=1)} \times \frac{\sum_Y P(A \mid Y, T=0)P(Y \mid T=0, X=0)}{\sum_Y P(A \mid Y, T=1)P(Y \mid T=1, X=0)} .$$

When the primary disease is independent of the marker conditional on the secondary phenotype, ψ will be equal to

$$\psi = \frac{\sum_{Y} P(A \mid Y, T=1) P(Y \mid T=1)}{\sum_{Y} P(A \mid Y, T=0) P(Y \mid T=0)} \times \frac{\sum_{Y} P(A \mid Y, T=0) P(Y \mid T=0)}{\sum_{Y} P(A \mid Y, T=1) P(Y \mid T=1)} = 1.$$

In this situation, $OR = OR_E$. The ascertainments do not have an impact on the OR estimation, and the estimated OR is unbiased. Therefore, in a scenario where the secondary phenotype is associated with the primary disease, but the marker is not associated with the primary disease, the standard estimate of OR relating a binary secondary phenotype and a binary marker is unbiased.