

Assess effect measure modification: Main effects models, product interaction terms and the constancy assumption

- Generalized linear models can be used to estimate adjusted effect measures. Like pooling and the Mantel-Haenszel method, model-based estimates assume constancy. In this section of the lab, we will begin by testing the constancy assumption using likelihood ratio tests (LRT).
- We assess effect modification by comparing two models, such as:
 - Main effects only $P(\text{death}) = \beta_0 + \beta_1(X_1) + \beta_2(X_2)$
 - Main effects with interaction terms $P(\text{death}) = \beta_0 + \beta_1(X_1) + \beta_2(X_2) + \beta_3(X_1 * X_2)$
- We evaluate the difference between the models using a) likelihood ratio test or b) the test statistic for the interaction term estimate β_3 (more on this topic below).
- The main effects model assumes constancy (i.e., the association between the outcome and X_1 is the same for all levels of X_2 (and the converse)).

Consider the main effects model: $\text{Risk}(\text{death}) = \beta_0 + \beta_1(\text{bord5}) + \beta_2(\text{male})$:

- The RD for high birth order vs. low birth order among male children is:
 $\text{RD}_{\text{bord5}=1 \text{ vs. } \text{bord5}=0} = [\beta_0 + \beta_1(1) + \beta_2(1)] - [\beta_0 + \beta_1(0) + \beta_2(1)] = \beta_1(1)$
- The RD for high birth order vs. low birth order among female children is:
 $\text{RD}_{\text{bord5}=1 \text{ vs. } \text{bord5}=0} = [\beta_0 + \beta_1(1) + \beta_2(0)] - [\beta_0 + \beta_1(0) + \beta_2(0)] = \beta_1(1)$
- The RD for male children vs. female children among those with high birth order is:
 $\text{RD}_{\text{male}=1 \text{ vs. } \text{male}=0} = [\beta_0 + \beta_1(1) + \beta_2(1)] - [\beta_0 + \beta_1(1) + \beta_2(0)] = \beta_2(1)$
- The RD for male children vs. female children among those with low birth order is:
 $\text{RD}_{\text{male}=1 \text{ vs. } \text{male}=0} = [\beta_0 + \beta_1(0) + \beta_2(1)] - [\beta_0 + \beta_1(0) + \beta_2(0)] = \beta_2(1)$
- In contrast, a product interaction model relaxes the constancy assumption (i.e., the association between the outcome and X_1 is the different among levels of X_2 and the converse)

Consider the interaction term model: $\text{Risk}(\text{death}) = \beta_0 + \beta_1(\text{bord5}) + \beta_2(\text{male}) + \beta_3(\text{bord5} * \text{male})$

- The RD for high birth order vs. low birth order among male children is:
 $\text{RD}_{\text{bord5}=1 \text{ vs. } \text{bord5}=0} = [\beta_0 + \beta_1(1) + \beta_2(1) + \beta_3(1*1)] - [\beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0*1)] = \beta_1(1) + \beta_3(1*1)$
- The RD for high birth order vs. low birth order among female children is:
 $\text{RD}_{\text{bord5}=1 \text{ vs. } \text{bord5}=0} = [\beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(1*0)] - [\beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0*0)] = \beta_1(1)$
- The RD for male children vs. female children among those with high birth order is:
 $\text{RD}_{\text{male}=1 \text{ vs. } \text{male}=0} = [\beta_0 + \beta_1(1) + \beta_2(1) + \beta_3(1*1)] - [\beta_0 + \beta_1(1) + \beta_2(0) + \beta_3(1*0)] = \beta_2(1) + \beta_3(1*1)$
- The RD for male children vs. female children among those with low birth order is:
 $\text{RD}_{\text{male}=1 \text{ vs. } \text{male}=0} = [\beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0*1)] - [\beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0*0)] = \beta_2(1)$

Note : The model including the interaction term $\text{bord5} * \text{male}$ is a **saturated model** because we obtain all 4 possible estimates from the included variables ($2 \times 2 = 4$ for all levels of bord5 by male) by including 4 terms in the model (β_0 counts as a term). The previous model without the interaction term yields only 2 estimates because of the constancy constraint.

If the constancy assumption is valid, adding the product interaction term contributes very little additional information to the model. Consequently, the coefficient for the product interaction term (β_3) will be very small, so that $\beta_1(1) + \beta_3(1*1)$ is approximately equal to $\beta_1(1)$ and the p-value of the interaction term will be non-significant.

In addition, the model that includes the interaction term will not “fit” the data substantially better than the model with the lower order terms only, so that there will be very little difference in the likelihood of the observed data between the two models. The null hypothesis for a likelihood ratio test is that the likelihood of the observed data is the same for the two models being compared; therefore, a small p-value by the LRT indicates that the data are not consistent with the homogeneity assumption.

In general the coefficient p-value will be similar to the LRT p-value