Example: Matched Case-Control Study

A group of oncologists studied women living in a retirement community, to determine if there was an association between the use of estrogen and the incidence of endometrial cancer. Cases were matched to controls who were approximately the same age, had the same marital status, and were living in the same community as the case at the time of the case's diagnosis. In addition to estrogen use, information was also collected from each subject on hypertension, gallbladder disease history, and prescription drug use.

(a) Fit a logistic regression model which accounts for the matching and includes estrogen and all available adjustment covariates.

- Among variables used for the matching, age is available.
- Model (patient i in pair k)

$$\begin{aligned} logit(\pi_{ik}) &= \alpha_k + \beta_1 Estrogen_{ik} + \beta_2 HYP_{ik} \\ &+ \beta_3 Gall_{ik} + \beta_4 Drug_{ik} + \beta_5 Age_{ik} \end{aligned}$$

- Conditional likelihood $(Y_{1k} = 1, Y_{2k} = 0)$

$$L = \prod_{k=1}^{K} L_k(\beta) = \prod_{k=1}^{K} \frac{\exp\{(x_{1k} - x_{2k})'\beta\}}{1 + \exp\{(x_{1k} - x_{2k})'\beta\}}$$

- Fit the logistic regression with the transformed x and y.
  - Unit is the pair
  - $x_k^* = x_{1k} - x_{2k}$
  - $y_k^* = 1$
  - No intercept

(b) Examine the results, paying particular attention to AGE. Given that age is known to be a risk factor for endometrial cancer, do the results of the fitted model make sense?

- Age (β_5): p value = 0.7135
- Age is not significant because it is used for the matching.

(c) Re-fit the model, this time not adjusting for age. Compare the estrogen effects based on this and the previous model.

- With age:

$$\widehat{\beta}_1 = 2.879, \quad \text{p-value} = 0.001$$

- Without age:

$$\widehat{\beta}_1 = 2.814, \quad \text{p-value} = 0.0008$$

- Nearly identical.