Example: Matched Cohort Data

Researchers studied the effect of a new treatment on a fairly frequently occurring skin condition. A total of 79 clinics participated in the study. Each clinic recruited two patients, with one randomized to receive the new treatment, the other receiving placebo. The two patients from a given clinic were matched on various demographic variates, reflecting income, socioeconomic status and baseline general health. Patients evaluated after 30 days post-randomization, then classified based on whether or not their condition improved. Adjustment covariates include age (AGE; recorded in years), SEX, and initial grade of skin condition (1, 2, 3, 4).

(a) The input records are as given below,

```
1 t f 27 0 1  1 p f 32 0 2
```

with the matched pair from each center contained in the same record. Read in the data set, then print off the resulting SAS file.

- See the SAS code

(b) Fit a logistic model which ignores the matching by center. Estimate the treatment effect based on this model.

- Model (patient i in clinic k)

$$\begin{aligned} logit(\pi_{ik}) &= \beta_0 + \beta_1 TRT_{ik} + \beta_2 Female_{ik} \\ &+ \beta_3 AGE_{ik} + \beta_4 Grade_{ik} \end{aligned}$$

$TRT_{ik} = I(NewTreatment)$

$Female_{ik} = I(Female)$

- Estimates: $\widehat{\beta}_1 = 0.8803$, $P - value : 0.0152$

(c) Why might the unmatched analysis from (b) yield biased parameter estimates?

If the true model is

$$\begin{aligned} logit(\pi_{ik}) &= \alpha_k + \beta_1 TRT_{ik} + \beta_2 Female_{ik} \\ &+ \beta_3 AGE_{ik} + \beta_4 Grade_{ik} \end{aligned}$$

and at least one α_k is different from the others, we can have biased results since the model in (b) does not adjust for center (k).

(d) Under what conditions would the parameter estimates from the unmatched analysis be valid?

- If $\alpha_1 = \cdots = \alpha_K$

(e) Fit a model which adjusts for center. Examine the log file, then comment.

- Warning messages. Parameter estimating

procedure is not stable, so the validity of the results is questionable.

- $\widehat{\beta}_1 = 1.4049$, $P - value : 0.0058$

(f) Fit the a logistic model using the conditional likelihood from matched pairs described for cohort studies in the lecture notes.

- $\widehat{\beta}_1 = 0.7024$, $P - value : 0.0511$

(g) Re-fit the conditional logistic regression model using the strata statement. Compare your results with those obtained previously.

- Results are the same.

(h) Why can't we estimate the center effects in this set-up?

- We used the conditional logistic regression, which factors out the center effect (α_k).
- Even if we try other methods, the small sample size in each center (sample size=2) does not allow to stably estimate the center effect.