## Marginal Poisson Regression Model and GEE

Examples of count data:

- number of panic attacks occurring during six months intervals after receiving treatment
- number of sexual partners in a three months period recorded in an HIV prevention program
- number of infants deaths per month before and after introduction of a prenatal care program

The *Poisson distribution* has been the most commonly used to model count data

$$\begin{aligned} P(Y=y) &= \tfrac{e^{-\mu}\mu^y}{y!} \qquad y = 0,1,2,\ldots \\ E[Y] &= Var(Y) = \mu \\ \text{if } Var(Y) &> E(Y) \Rightarrow \textit{over-dispersed data} \end{aligned}$$



## Epileptic seizures

- Clinical Trail of 59 epileptics
- For each patient, the number of epileptic seizures was recorded during a baseline period of eight weeks
- patient were randomized to treatment with the anti-epileptic drug progabide or placebo
- Number of seizures was then recorded in four consecutive two weeks intervals
- **Question:** is progabide reduces the rate of epileptic seizures?



## Poisson Regression Model

$$\begin{aligned} Y_{ij} &\sim \text{Poisson}(\mu_{ij}) \\ \log \mu_{ij} &= \beta_0 + \beta x_i \end{aligned}$$

- $x_i$ is the treatment indicator
- $\beta$ *describe the change in the log of the population average count per unit change in* $x_i$
- in the progabide example:

  $\exp(\beta)$ represents the *ratio of average seizures rates, measured as the number of seizures per two-week period, for the treated patients compared to the control patients*

$$\frac{E[Y_{ij} \mid x_i = 1]}{E[Y_{ij} \mid x_i = 0]} = \exp(\beta)$$

- If $\beta < 0$ then the treatment if effective relative to the placebo in controlling the seizure rate



**Overdispersed data** – $Var(Y_{ij}) > E[Y_{ij}]$

$$\begin{aligned} Var(Y_{ij}) &= \phi_{ij} E[Y_{ij}] \\ \phi_{ij} &= \phi_1, \text{ if } i \text{ is assigned to a treatment} \\ \phi_{ij} &= \phi_2, \text{ if } i \text{ is assigned to a control} \end{aligned}$$

**Irregular times**: Suppose that the interval times $t_{ij}$, during which the events are observed, *are not the same* for all subjects. The problem can be solved by decomposing the marginal mean $E[Y_{ij}]$ as

$$\begin{aligned} E[Y_{ij}] &= \lambda_{ij} \times t_{ij} \\ \lambda_{ij} &= \text{number of events for unit interval} \\ t_{ij} &= \text{length of time interval} \\ \log E[Y_{ij}] &= \underbrace{\log t_{ij}}_{\text{offset}} + \boldsymbol{x}_{ij}^{'}\boldsymbol{\beta} \end{aligned}$$

## Epileptic seizures

1. Clinical Trail of 59 epileptics
2. 31 patients received a anti-epileptic drug progabide
3. 28 received a placebo
4. patients from the two groups are comparable in terms of age and 8-week baseline seizure counts

### Poisson regression model and GEE method

$$
\begin{aligned}
\log E[Y_{ij}] &= \log t_{ij} + \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} x_{i2} \\
j &= 0, 1, 2, 3, 4 \\
i &= 1, 2, \ldots, 59 \\
t_{ij} &= 8 \text{ if } j = 0 \\
t_{ij} &= 2 \text{ if } j = 1, 2, 3, 4 \\
x_{i1} &= \begin{cases} 1 & \text{if visit } 1, 2, 3, 4 \\ 0 & \text{if baseline} \end{cases} \\
x_{i2} &= \begin{cases} 1 & \text{if progabide} \\ 0 & \text{if placebo} \end{cases}
\end{aligned}
$$



- $x_{i2}$ allow different baseline seizure counts between the treated and placebo groups



## Parameter Interpretation and Results

- $\exp(\beta_1)$ is *the ratio of the average seizure rate after the treatment to before treatment for the placebo group*
- $\beta_3$ is the parameter of interest and represents *the difference in the logarithm of the post-to pre-treatment ratio between the progabide and the placebo groups*. A negative coefficient correspond to a greater reduction (small increase) in the seizure counts for the progabide group
- $\hat{\beta}_3 = -0.10(0.21)$ if patient 207 is included this suggests that there is a very little difference between the treatment and the placebo groups in the change of seizure counts before and after the randomization
- if patient 207 is set aside $\hat{\beta}_3 = -0.30(0.17)$ then there is modest evidence that progabide is favored over the placebo
- $\hat{\phi} = 19.44$ strong overdispersion!
- we have completely ignored correlation within subjects