GEE for Mean and Association (Continued)

Model:

$$\begin{aligned} g(\mu_{ij}) &= \mathbf{X}_{ij}^T \boldsymbol{\beta} \\ ln(\nu_{ijk}) &= \mathbf{Z}_{ijk}^T \boldsymbol{\alpha} \end{aligned}$$

where

$$\begin{aligned} \mu_{ij} &= E(Y_{ij}) = Pr[Y_{ij} = 1] \\ \nu_{ijk} &= OR(Y_{ij}, Y_{ik}) \end{aligned}$$

e.g. $ln(\nu_{ijk}) = \alpha_0 + \alpha_1 |t_{ij} - t_{ik}|^{-1}$, no constraint on α's.

GEE2

$$
\begin{aligned}
\mathbf{U}_\beta(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} \frac{\partial \boldsymbol{\mu}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0 \\
\mathbf{U}_\alpha(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \sum_{i=1}^{m} \frac{\partial \boldsymbol{\eta}_i^T}{\partial \boldsymbol{\alpha}} \mathbf{H}_i^{-1}(\mathbf{W}_i - \boldsymbol{\eta}_i) = 0
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{V_i} &= \text{working covariance of } \mathbf{Y_i} \\
\mathbf{W}_i &= (R_{i1}R_{i2}, R_{i2}R_{i3}, \cdots, R_{i,n_i-1}R_{i,n_i})^T \\
\boldsymbol{\eta}_i &= E(\mathbf{W}_i) \\
\mathbf{H}_i &= \text{working covariance of } \mathbf{W_i}
\end{aligned}
$$

GEE2: Remarks

- One can set $\boldsymbol{H}_i = diag\{var(R_{i1}R_{i2}), \cdots var(R_{i,n_i-1}R_{in_i})\}$then $\boldsymbol{\eta}_i$and $\mathbf{H}_i$are fully determined by the marginal mean $\boldsymbol{\mu}_i$and the marginal odds ratios $\boldsymbol{\nu_i}$.

- The exact covariance of $\mathbf{W}_i$ depends on higher order moments of $\mathbf{Y}_i$.

- In principle, we could continue modeling the higher order moments... until we modeled them all...

GEE2: $\boldsymbol{\eta}_i$ and $\boldsymbol{\mu}_i$, $\boldsymbol{\nu}_i$

Question: How is $\boldsymbol{\eta}_i$ related to $\boldsymbol{\mu}_i, \boldsymbol{\nu}_i$?

$$
\begin{aligned}
\eta_{ijk} &= E(W_{ijk}) = \frac{E(Y_{ij}Y_{ik}) - \mu_{ij}\mu_{ik}}{\{\mu_{ij}(1-\mu_{ij})\mu_{ik}(1-\mu_{ik})\}^{1/2}} \\
&= \frac{\mu_{ijk} - \mu_{ij}\mu_{ik}}{\{\mu_{ij}(1-\mu_{ij})\mu_{ik}(1-\mu_{ik})\}^{1/2}} \\
\mu_{ijk} &= Pr[Y_{ij} = Y_{ik} = 1] \\
&= \begin{cases} \frac{1-(\mu_{ij}+\mu_{ik})(1-\nu_{ijk})-\{[1-(\mu_{ij}+\mu_{ik})(1-\nu_{ijk})]^2-4(\nu_{ijk}-1)\nu_{ijk}\mu_{ij}\mu_{ik}\}^{\frac{1}{2}}}{2(\nu_{ijk}-1)}, & \text{if } \nu_{ijk} \neq 1 \\ \mu_{ij}\mu_{ik}, & \text{if } \nu_{ijk} = 1 \end{cases}
\end{aligned}
$$

GEE2: Distribution of $\widehat{\beta}$ and $\widehat{\alpha}$

$$
\begin{aligned}
\Rightarrow \quad \mathbf{U}(\delta) &= \sum_i \frac{\partial(\mu_i, \eta_i)^T}{\partial \delta} \begin{pmatrix} \mathbf{V^{-1}}_i & 0 \\ 0 & \mathbf{H}_i^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{W}_i - \boldsymbol{\eta}_i \end{pmatrix} \\
&= \sum_i \mathbf{C}_i^T \mathbf{B}_i \begin{pmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{W}_i - \boldsymbol{\eta}_i \end{pmatrix}
\end{aligned}
$$

where $\boldsymbol{\delta} = (\beta^T, \boldsymbol{\alpha}^T)^T$. Then:

$$\hat{\boldsymbol{\delta}} \sim N(\boldsymbol{\delta}, \mathbf{V}_\delta) \text{ asymptotically}$$

where

$$
\begin{aligned}
\mathbf{V}_\delta &= \frac{1}{m}(\sum_{i=1}^m \mathbf{C}_i^T \mathbf{B}_i \mathbf{C}_i)^{-1} \{\sum_{i=1}^m \mathbf{C}_i^T \mathbf{B}_i \begin{pmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{W}_i - \boldsymbol{\eta}_i \end{pmatrix} \\
&\qquad \begin{pmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{W}_i - \boldsymbol{\eta}_i \end{pmatrix}^T \mathbf{B}_i^T \mathbf{C}_i\} (\sum_{i=1}^m \mathbf{C}_i^T \mathbf{B}_i \mathbf{C}_i)^{-1}
\end{aligned}
$$

Alternative Estimating Equation

An alternative EE:

$$\mathbf{U}(\delta) = \sum_i \frac{\partial(\mu_i, \eta_i)^T}{\partial \boldsymbol{\delta}} cov^{-1} \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{pmatrix} \begin{pmatrix} \mathbf{Y}_i - \boldsymbol{\mu}_i \\ \mathbf{W}_i - \boldsymbol{\eta}_i \end{pmatrix}$$

Now $cov^{-1} \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{pmatrix}$ may not be a block diagonal matrix.

Dis/Advantages

Advantage: If $cov^{-1}\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{pmatrix}$ is correctly specified, then $\hat{\boldsymbol{\delta}}$ would be more efficient and $\mathbf{U}(\delta)$ is optimal.

Disadvantage: If $cov^{-1}\begin{pmatrix} \mathbf{Y}_i \\ \mathbf{W}_i \end{pmatrix}$ is misspecified and contains off-block-diagonal elements, then misspecification of the ν (OR) model would result in a biased estimator of the mean parameter vector $\boldsymbol{\beta}$.

Alternating Logistic Regression (ALR)

Reference: Carey, Zeger, Diggle (1993)

Model:

$$\begin{aligned} g(\mu_{ij}) &= \mathbf{X}_{ij}^T \boldsymbol{\beta} \\ ln(\nu_{ijk}) &= \mathbf{Z}_{ijk}^T \boldsymbol{\alpha} \end{aligned}$$

GEE2 can be fitted as follows:

1. Assign an initial value $\boldsymbol{\alpha}_0$.
2. Use GEE1 to estimate $\boldsymbol{\beta}$ and μ_{ij}.
3. Fit the following offset logistic model:

$$logit\,\Pr[Y_{ij} = 1 | Y_{ik} = 1] = Y_{ik} \mathbf{Z}_{ijk}^T \boldsymbol{\alpha} + \ln \left(\frac{\mu_{ij} - \mu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}} \right)$$

4. Iterate between 2 and 3.

Indonesian Infectious Disease Data (Again)

- Data:
  - $m = 275$ children examined every 3 months up to 6 consecutive quarters
  - Outcome=respiratory infection (Y/N)
  - Covariates=XERO, Gender, Age, Height, Season

- Models considered:
  - Model 1: GEE1 (cross-sectional age effect: $\beta_1 age_{ij} + \beta_2 age_{ij}^2$)
  - Model 2: GEE2 (cross-sectional age effect)

$$ln \nu_{ijk} = \alpha$$

  - Model 3: GEE2 (distinguish cross-sectional and longitudinal effects, no seasonal effect)

$$\beta_1 age_{i1} + \beta_2 age_{i1}^2 + \beta_3 (age_{ij} - age_{i1}) + \beta_4 (age_{ij} - age_{i1})^2$$

Indonesian Infectious Disease Data: Results

Variable	Model 1	2	3	4
Intercept	-1.47	-2.05	-1.76	-2.21
	(0.36)	(0.21)	(0.25)	(0.32)
Gender	-0.66	-0.49	-0.53	-0.53
	(0.44)	(0.24)	(0.24)	(0.24)
Height for age	-0.11	-0.042	-0.051	-0.048
	(0.041)	(0.023)	(0.025)	(0.024)
Seasonal cosine	-	-0.59	-	-0.54
		(0.17)		(0.21)
Seasonal sine	-	-0.16	-	-0.016
	-	(0.14)	-	(0.18)
Xerophthalmia	0.44	0.50	0.53	0.64
	(1.15)	(0.44)	(0.45)	(0.44)
Age	-0.089	-0.030	-	-
	(0.027)	(0.008)	-	-
Age^2	-0.0026	-0.0010	-	-
	(0.0011)	(0.0004)	-	-
Age at entry	-	-	-0.053	-0.053
	-	-	(0.013)	(0.013)
$(\text{Age at entry})^2$	-	-	-0.0013	-0.0013
	-	-	(0.0005)	(0.0005)
Follow-up time	-	-	-0.19	-0.082
	-	-	(0.071)	(0.099)
$(\text{Follow-up})^2$	-	-	0.013	0.007
	-	-	(0.004)	(0.007)

Indonesian Infectious Disease Data: Cross Sectional

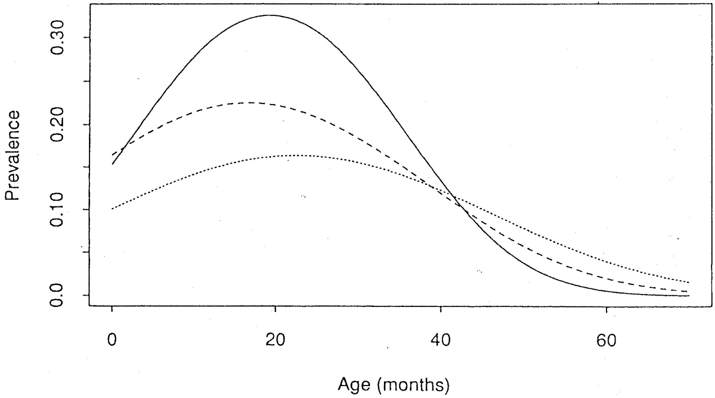

Fig. 8.1. Prevalence of respiratory infection as a function of age for three different models. —— : Model 1; .... : Model 2; – – – : Model 3.

Indonesian Infectious Disease Data: Longitudinal

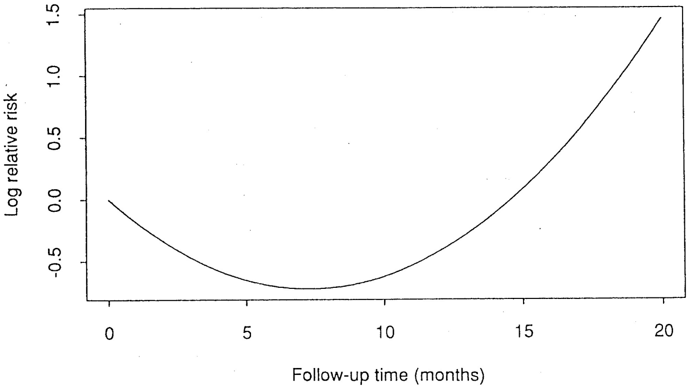

Fig. 8.3. The logarithm of the risk of respiratory infection as a function of follow-up time relative to the risk at an individual's first visit.