Hierarchical modeling of endpoints of different types with generalized linear mixed models

Christel Faes

I-BioStat, Hasselt University, Belgium

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Introduction

- In many situation, one is confronted with the analysis of longitudinal data
- Often, measurements (or sequences) of both continuous and categorical outcomes are recorded
- Multivariate methods for continuous outcomes are well understood
- Methods for joint continuous and discrete outcomes are less familiar
- Complexity of joint model increases with number of outcomes

Introduction
Bivariate Model
Extension to Higher Dimensions
Discussion

Why Jointly?

- Association structure can be of interest
  - E.g. Relationship between outcome-specific evolutions
- Interest in comparison of trends for different outcomes
  - E.g. testing difference in treatment effect between many outcomes
  - E.g. jointly testing treatment effect on a set of outcomes
- Discriminant analysis, principal component analysis

Christel Faes ROeS 2013

Bivariate Model for Binary and Continuous Longitudinal Outcomes

We want a method that:

- Allows to specify the marginal model for each outcome
- Estimate the bivariate association between continuous and discrete outcomes
- Accounts for the longitudinal effects / longitudinal association
- Is computationally not too complex

Background

Factorization Models

- Models that condition on the continuous outcome

$$f(C, B) = f(C) \times f(B|C)$$

- Models that condition on the binary outcome

$$f(C, B) = f(B) \times f(C|B)$$

- Correlation difficult to characterize with conditional joint models
- Difficult to generalize to three or more endpoints

E.g. Tate (1954), Olkin and Tate (1961), Little and Schluchter (1985), Krzanowsku (1988)
E.g. Catalano and Ryan (1992), Fitzmaurice and Laird (1993)

Background

Direct Specification of Joint Model

- Probit Approach
  - assume normal latent variable
  - use correlation coefficient
- Copula function
  - e.g. Plackett latent variable using odds ratio

GLMM Approach

- allow the link function to change with the nature of the outcomes
- use random effects or residual error to account for the association in the data

Burzykowski et al., 2001; Geys et al., 2001; Faes et al., 2004; de Leon and Wu, 2011

Faes et al. (2006), Gueorguieva and Sanacora (2006), Regan and Catalano (1999)

A General Linear Mixed Model

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\epsilon}_i = \boldsymbol{h}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i) + \boldsymbol{\epsilon}_i$$

with

$$\mathbf{b}_i \sim N(\mathbf{0}, \mathbf{D})$$

- $\boldsymbol{Y}_i = (\mathbf{C}_i', \mathbf{B}_i')'$
- $\boldsymbol{h}(.)$ is allowed to change with nature of outcome
  - Identity link for continuous component
  - Logit link for binary component
- ϵ_i is the residual error structure, of which the variance depends on the mean-variance links of the various endpoints

A General Linear Mixed Model

The approximate variance-covariance matrix of $\mathbf{Y}_i$ is

$$V_i = \mathrm{Var}(\boldsymbol{Y}_i) \simeq \Delta_i Z_i D Z_i' \Delta_i' + \Sigma_i,$$

with

$$\Delta_i = \left(\frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\eta}_i} \right) \Big|_{\boldsymbol{b}_i = \mathbf{0}},$$

and

$$\Sigma_i \simeq \Xi_i^{1/2} A_i^{1/2} R_i(\alpha) A_i^{1/2} \Xi_i^{1/2}.$$

- $R_i(\alpha)$ residual correlation
- Ξ_i diagonal matrix with overdispersion parameters
- A_i diagonal matrix with variances (given $\boldsymbol{b}_i = 0$)

Special cases

Correlations among the measurements can be introduced via:

- Marginal model: correlation via residual variance of $\boldsymbol{Y}_i$ (via Σ_i or R_i)
- Shared random effects model: scale parameter to allow for different variances
- Correlated random effects model: correlation induced via random effects
- Independent random effects and correlated errors model

Correlated random effects model

Assume the following bivariate model:

$$\begin{pmatrix} Y_{i1j} \\ Y_{i2j} \end{pmatrix} = \begin{pmatrix} \alpha_0 + \alpha_1 X_{ij} + b_{i1} \\ \frac{\exp(\beta_0+\beta_1 X_{ij}+b_{i2})}{1+\exp(\beta_0+\beta_1 X_{ij}+b_{i2})} \end{pmatrix} + \begin{pmatrix} \epsilon_{i1j} \\ \epsilon_{i2j} \end{pmatrix}$$

where the random effect b_{i1} and b_{i2} are normally distributed as

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho\tau_1\tau_2 \\ \rho\tau_1\tau_2 & \tau_2^2 \end{pmatrix}\right)$$

and where ϵ_{i1j} and ϵ_{i2j} are independent

Remark: Standard software can be used to obtain parameter estimated for this bivariate model (e.g., by integrating out random effects using SAS-procedure NLMIXED)

Bivariate Continuous/Binary Model

- The correlation among the continuous and binary endpoints is approximately equal to:

$$\rho_{Y_{i1j}Y_{i2j}} \approx \rho \frac{\tau_1}{\sqrt{\tau_1^2+\sigma^2}} \frac{v_{i2j}\tau_2}{\sqrt{v_{i2j}^2\tau_2^2+v_{i2j}}}.$$

- The correlation among the continuous endpoints:

$$\rho_{Y_{i1j}Y_{i1k}} \approx \frac{\tau_1^2}{\tau_1^2+\sigma^2}.$$

- The correlation among the binary endpoints:

$$\rho_{Y_{i2j}Y_{i2k}} \approx \frac{v_{i2j}^2\tau_2^2}{v_{i2j}^2\tau_2^2+v_{i2j}}.$$

Bivariate GLMM Model

In summary:

- Marginal distribution are specified, depending on type of endpoint
- Random effect used to account for longitudinal effect
- Correlation between the random effects used to account for association among the endpoints
- Can be used for any mixture of endpoints:
  - Binary/Continuous
  - Binary/Binary
  - Continuous/Continuous
- Not restricted to a bivariate model

Example: Repeated-Dose Toxicity Study (Irwin's study)

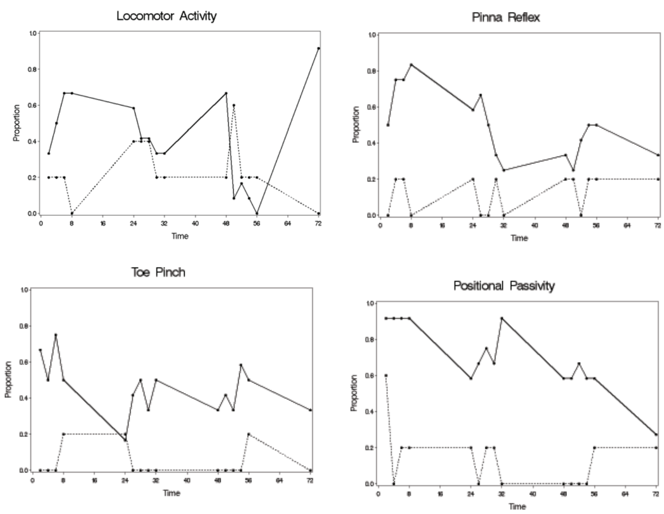
Research Question:
Has the chemical an effect on general activity and behavior?

- Male rats were dosed during 3 consecutive days
- Randomized in control (5 rats) and treatment group (15 rats)
- Examination of rats at 2, 4, 6, 8 and 24 hours after daily exposure
- A set of observational and interactive measurements are used (Irwin's method)

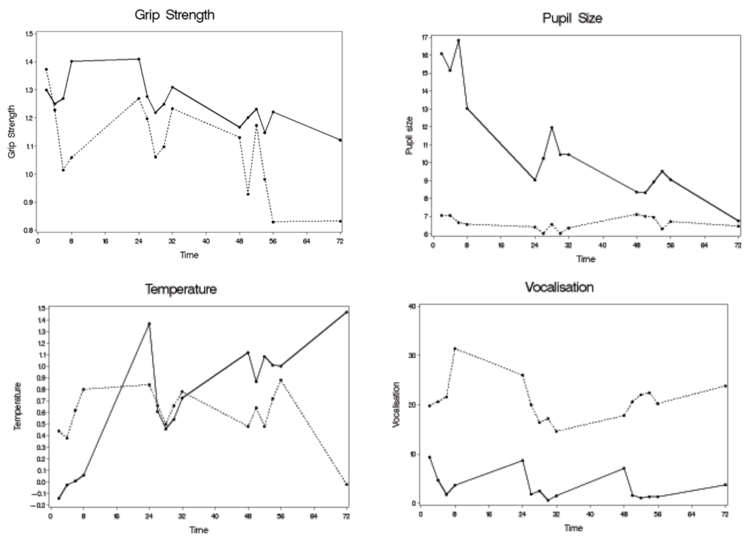
Repeated-Dose Toxicity Study

	Variable	Type	Description	Classification
1	Locomotor Activity	Binary	Characterized by abnormal biting, restlessness, writhing	Behavioral Spontaneous activity
2	Pinna Reflex	Binary	Animal's twitch reflex upon being touched on the auricle	Behavioral Sensoro-motor response
3	Toe Pinch	Binary	Animal's response to pain upon having the toe squeezed	Behavioral Sensoro-motor response
4	Positional Passivity	Binary	Animal's struggle response to sequential handling	Behavioral Motor-affective response
5	Grip Strength	Continuous	Animal's forelimb grip strength	Neurologic Muscle tone
6	Pupil Size	Continuous	Animal's pupil diameter	Autonomic
7	Temperature	Continuous	Animal's body temperature	Autonomic
8	Vocalization	Continuous	Animal's incidence of squeaking during manipulation	Behavioral Motor-affective response

Four Binary Endpoints



Four Continuous Endpoints



Application: Joint Model

$$h_k^{-1}(\mu_{ij}) = \beta_{0k} + \beta_{1k} g_i + \beta_{2k} t_{ij} + \beta_{3k} d_{ij} + \beta_{4k} t_{ij} d_{ij} + \beta_{5k} g_i t_{ij} + \beta_{6k} g_i d_{ij} + b_{ik},$$

with

$$\mathbf{b}_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \dots \\ b_{i8} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho_{12}\tau_1\tau_2 & \dots & \rho_{18}\tau_1\tau_8 \\ \rho_{12}\tau_1\tau_2 & \tau_2^2 & \dots & \rho_{28}\tau_2\tau_8 \\ \dots & \dots & \dots & \dots \\ \rho_{18}\tau_1\tau_8 & \rho_{28}\tau_2\tau_8 & \dots & \tau_8^2 \end{pmatrix} \right\}.$$

- h_k is the identity link in case of a continuous endpoint and the logit link in case of a binary endpoint
- g_i is indicator variable for treatment
- t_{ij} is the time after exposure
- d_{ij} is the day of the experiment

Application: Joint Model

$$h_k^{-1}(\mu_{ij}) = \beta_{0k} + \beta_{1k} g_i + \beta_{2k} t_{ij} + \beta_{3k} d_{ij} + \beta_{4k} t_{ij} d_{ij} + \beta_{5k} g_i t_{ij} + \beta_{6k} g_i d_{ij} + b_{ik},$$

with

$$\mathbf{b}_i = \begin{pmatrix} b_{i1} \\ b_{i2} \\ \dots \\ b_{i8} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & \rho_{12}\tau_1\tau_2 & \dots & \rho_{18}\tau_1\tau_8 \\ \rho_{12}\tau_1\tau_2 & \tau_2^2 & \dots & \rho_{28}\tau_2\tau_8 \\ \dots & \dots & \dots & \dots \\ \rho_{18}\tau_1\tau_8 & \rho_{28}\tau_2\tau_8 & \dots & \tau_8^2 \end{pmatrix} \right\}.$$

- h_k is the identity link in case of a continuous endpoint and the logit link in case of a binary endpoint
- g_i is indicator variable for treatment
- t_{ij} is the time after exposure
- d_{ij} is the day of the experiment

Problem

No convergence!

Problem with Extension to Higher Dimensions

The marginal likelihood contribution for subject i becomes

$$\ell_i(\boldsymbol{\Theta}|\mathbf{Y}_{i1},\dots,\mathbf{Y}_{im}) = \int_{I\!R^m} \prod_{j=1}^{n} f_{ij}(y_{i1j},\dots,y_{imj}|b_i,\boldsymbol{\Theta}) f(\mathbf{b}_i|D) d\mathbf{b}_i.$$

Computational problems often arise when m increases due to the $m-$ dimensional integral, especially when outcomes are of different type.

Pseudo-Likelihood (Composite Likelihood)

Replace the full likelihood contribution for subject i

$$\ell_i(\boldsymbol{\Theta}|\mathbf{Y}_{i1},\ldots,\mathbf{Y}_{im})$$

by the pseudo-likelihood function

$$p\ell_i = \prod_{k=1}^{m-1}\prod_{l=k+1}^{m} \ell_{ikl}(\boldsymbol{\Theta}|\mathbf{Y}_{ik},\mathbf{Y}_{il}),$$

where each contribution ℓ_{ikl} is the likelihood corresponding with a bivariate GLMM model for outcomes k and l.

Remark

Standard software can be used to obtain parameter estimated for this model (e.g., SAS-procedure NLMIXED).

Idea Pseudo-Likelihood

- Replace full likelihood by function that is easier to evaluate
- Estimates all pairwise correlations
- Use of sandwich variance estimator to adjust for potential misspecification
- Pseudo-likelihood estimates are consistent and asymptotically normal
- Loss of efficiency is small for realistic settings
- Pseudo-likelihood ratio test statistic easy-to-compute

Arnold and Strauss (1991)

Geys, Molenberghs and Ryan (1999)

Idea Pseudo-Likelihood

The asymptotic variance-covariance matrix of the estimate $\widehat{\mathbf{\Theta}}$ is given by

$$\mathbf{J}(\mathbf{\Theta})^{-1}\mathbf{K}(\mathbf{\Theta})\mathbf{J}(\mathbf{\Theta})^{-1},$$

where $\mathbf{J}(\mathbf{\Theta})$ is a matrix with elements defined by

$$J_{pq} = -\sum_{k=1}^{K-1}\sum_{k'=k+1}^{K} E\left(\frac{\partial^2 \ln \ell_{ikk'}(\mathbf{\Theta}; \mathbf{Y}_{ik}, \mathbf{Y}_{ik'})}{\partial\theta_p \partial\theta_q}\right),$$

and $\mathbf{K}(\mathbf{\Theta})$ is a symmetric matrix with elements

$$K_{pq} = -\sum_{k=1}^{K-1}\sum_{k'=k+1}^{K} E\left(\frac{\partial \ln \ell_{ikk'}(\mathbf{\Theta}; \mathbf{Y}_{ik}, \mathbf{Y}_{ik'})}{\partial\theta_p}\frac{\partial \ln \ell_{ikk'}(\mathbf{\Theta}; \mathbf{Y}_{ik}, \mathbf{Y}_{ik'})}{\partial\theta_q}\right).$$

Pairwise Approach

$$\ell_i(\mathbf{\Theta}|\mathbf{Y}_{i1}, \ldots, \mathbf{Y}_{im})$$

Pseudo-likelihood Approach

$$p\ell_i = \prod_{k=1}^{m-1}\prod_{l=k+1}^{m} \ell_{ikl}(\mathbf{\Theta}|\mathbf{Y}_{ik}, \mathbf{Y}_{il})$$

- maximize the full pseudo-likelihood function at once
- number of parameters to estimate jointly increases with m
- pseudo-likelihood ratio tests can be used

Faes et al. (2007)

Pairwise Approach

$$p\ell_i = \prod_{k=1}^{m-1}\prod_{l=k+1}^{m} \ell_{ikl}(\mathbf{\Theta}_{\mathbf{k},\mathbf{l}}|\mathbf{Y}_{ik}, \mathbf{Y}_{il})$$

- maximize all pairwise likelihood functions separately
- number of pairwise models increases with m
- estimates of joint model derived by taking averages over estimates from the pairwise models

Fieuws and Verbeke (2005)

Model Fitting

- Pseudo-likelihood and pairwise method used to estimate parameters
- Relatively easy to program
- Parameter estimates from the two approaches are very similar

Estimated Correlations

Estimated correlation matrix for random effects

		1	2	3	4	5	6	7	8
1	Locom Act	0.08							
2	Pinna Reflex	0.36	0.29						
3	Toe Pinch	-0.34	-0.58	0.38					
4	Vertical Hind	0.11	0.15	-0.56	0.24				
5	Grip Strength	0.39	-0.23	**0.76**	-0.09	0.25			
6	Pupil Size	0.50	0.25	-0.37	0.40	-0.21	0.40		
7	Temperature	-0.42	-0.43	0.36	-0.55	0.12	**-0.82**	0.37	
8	Vocalization	**-0.69**	-0.32	-0.34	-0.08	**-0.60**	-0.03	0.24	0.53

Values on the diagonal are the intra-class correlations corresponding to each outcome.

Represents relationship among the subject-specific deviations from different outcomes.

Estimated Correlations

Estimated correlation matrix for endpoints

		1	2	3	4	5	6	7	8
1	Locom Act	1							
2	Pinna Reflex	0.02	1						
3	Toe Pinch	-0.02	-0.03	1					
4	Vertical Hind	0.01	0.01	-0.05	1				
5	Grip Strength	0.05	-0.03	**0.08**	-0.02	1			
6	Pupil Size	0.07	0.04	-0.05	0.10	-0.07	1		
7	Temperature	-0.06	-0.07	0.05	-0.13	0.04	**-0.31**	1	
8	Vocalization	**-0.12**	-0.06	-0.05	-0.02	**-0.22**	-0.01	0.11	1

Represents relationship between outcomes-specific evolutions.

Hypothesis Testing

Four types of tests are considered:

- Test whether there is a treatment effect for each response separately
- Test whether there is an overall treatment effect (over all responses)
- Test whether there is a dose effect for groups of variables (such as the sensoro-motor responses and motor-affective responses)
- Test whether there is a difference among variables (such as difference between the sensoro-motor responses pinna reflex and toe pinch)

Hypothesis Testing

Pseudo-likelihood ratio tests are used.

Test	Response	PLR	df	p-value
Treatment effect	Locomotor Activity	5.98	3	0.112
	Pinna Reflex	21.16	3	<0.001
	Toe Pinch	11.67	3	0.009
	Positional Passivity	7.23	3	0.065
	Vocalization	9.77	3	0.021
	Grip Strength	17.80	3	<0.001
	Pupil Size	71.99	3	<0.001
	Temperature	72.25	3	<0.001
Treatment effect	Sensoro-Motor	104.09	6	<0.001
	Motor-affective	335.63	6	<0.001
	Neurologic	453.40	6	<0.001
Equality	Sensoro-Motor	230	3	<0.001
	Motor-affective	2.00	3	0.573
	Neurologic	90.85	3	<0.001

Discussion

- The GLMM approach is very flexible to model longitudinal endpoints of a different type
- Not restricted to binary and continuous endpoints
- The pseudo-likelihood method makes is possible to extend the approach to higher dimensions
- Easy to program