Investigating Association Using Surrogate Marker Methodology

Abel Tilahun

Interuniversity Institute for Biostatistics and statistical Bioinformatics
Universiteit Hasselt , Diepenbeek, Belgium

Non-Clinical Statistics Leuven 2008

Outline

- **Definition**
  - *Clinical endpoint:* A characteristic or variable that reflects how a patient feels or functions, or how long a patient survives.
  - *Surrogate Endpoint:* A biomarker intended to substitute for a clinical endpoint.
- **Motivation**
  - Time of producing the study results
  - Cost of the study
  - Convenience for the patient
- **Objective** To predict the clinical outcome using the surrogate endpoint

- Consider the following pair of models:

$$\begin{aligned} S_j &= \mu_S + \alpha Z_j + \varepsilon_{S_j} \\ T_j &= \mu_T + \beta Z_j + \varepsilon_{T_j} \end{aligned}$$

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ \sigma_{TS} & \sigma_{TT} \end{pmatrix}$$

- Buyse and Molenberghs (1998) suggested the use of the adjusted association.
- Then, the *adjusted association*, denoted R^2 can be computed as:

$$R^2 = \frac{\sigma_{ST}^2}{\sigma_{SS}\sigma_{TT}}$$

- Consider the following generalized linear models for some link function:

$$g_T\{E(T_j)\} = \mu_T + \beta Z_j, \quad (1)$$

$$g_T\{E(T_j|S_j)\} = \theta_0 + \theta_1 Z_j + \theta_2 S_j \quad (2)$$

- Alonso *et al* (2005) used information theory to quantify the association using the likelihood reduction factor *LRF* given by

$$LRF = 1 - \exp\left(-\frac{G^2}{n}\right)$$

where G^2 denotes the log-likelihood ratio test statistic n is the sample size.

- Consider the following bivariate model :

$$\begin{aligned} T_{jk} &= \mu_T + \alpha Z_j + f(t_{jk}) + \varepsilon_{Tjk} \\ S_{jk} &= \mu_S + \beta Z_j + f(t_{jk}) + \varepsilon_{Sjk} \end{aligned}$$

$$\Sigma = \begin{pmatrix} \Sigma_{TT} & \Sigma_{TS} \\ \Sigma_{ST} & \Sigma_{SS} \end{pmatrix}$$

- In some practical settings Σ can be modeled as the Kronecker product of two matrices Galecki (1994)

$$\Sigma = \begin{pmatrix} d_{aa} & d_{ab} \\ d_{ba} & d_{bb} \end{pmatrix} \bigotimes R$$

- R can assume any structure such as an $AR(1)$, CS or any general variance covariance matrix as:

- Alonso *et al.* (2004) have suggested two measures of associations:
- Variance Reduction Factor (VRF)

$$VRF = \frac{\text{tr}(\Sigma_{TT}) - \text{tr}(\Sigma_{T|S})}{\text{tr}(\Sigma_{TT})}$$

where $\Sigma_{T|S}$ denotes the conditional variance-covariance matrix of T_{jk} given S_{jk}, i.e., $\Sigma_{T|S} = \Sigma_{TT} - \Sigma_{TS}\Sigma_{SS}^{-1}\Sigma_{ST}$

- R^2_Λ takes the following format

$$R^2_\Lambda = 1 - \frac{|\Sigma|}{|\Sigma_{TT}| \cdot |\Sigma_{SS}|}$$

Properties of VRF

1. VRF ranges between zero and one
2. $VRF = 0$ if and only if the two outcomes are independent
3. $VRF = 1$ if and only if there exists a deterministic relationship
4. $VRF = R^2$ in the cross-sectional setting.

Properties of R^2_Λ

1. R^2_Λ ranges between zero and one
2. $R^2_\Lambda = 0$ if and only if the two outcomes are independent
3. $R^2_\Lambda = 1$ if only if there exist a and b so that $a^T \varepsilon_{S_{jk}} = b^T \varepsilon_{T_{jk}}$ with probability one
4. $R^2_\Lambda = R^2$ in the cross-sectional setting.

- The model takes the following format

$$\begin{aligned} C_j &= \mu_C + \alpha Z_j + \varepsilon_{C_j} \\ L_{jk} &= \mu_L + \beta Z_j + f(t_{jk}) + \varepsilon_{L_{jk}} \end{aligned}$$

$$\Sigma = \begin{pmatrix} \sigma_{CC} & \Sigma_{CL} \\ \Sigma_{LC} & \Sigma_{LL} \end{pmatrix}$$

- The VRF and R^2_Λ will take the following expression:

$$VRF_{LC} = \frac{\Sigma_{CL}\Sigma_{LL}^{-1}\Sigma_{CL}}{\sigma_{CC}}$$

$$R^2_{\Lambda_{LC}} = \frac{\Sigma_{CL}\Sigma_{LL}^{-1}\Sigma_{LC}}{\sigma_{CC}}$$

- The model takes the following format

$$\begin{aligned} L_{jk} &= \mu_T + \beta Z_j + f(t_{jk}) + \varepsilon_{Ljk} \\ C_j &= \mu_S + \alpha Z_j + \varepsilon_{Cj} \end{aligned}$$

$$\Sigma = \begin{pmatrix} \Sigma_{LL} & \Sigma_{LC} \\ \Sigma_{CL} & \sigma_{CC} \end{pmatrix}$$

- The VRF and R^2_Λ will take the following

$$VRF_{CL} = \frac{\text{tr}(\Sigma_{LC}\Sigma_{CL})}{\sigma_{CC}.\text{tr}(\Sigma_{LL})}$$

$$R^2_{\Lambda_{CL}} = \frac{\Sigma_{CL}\Sigma_{LL}^{-1}\Sigma_{LC}}{\sigma_{CC}}$$

- Two Normal outcomes
  - Selecting genes as potential biomarkers when the outcome is normally distributed
- Non-normal setting
  - Selecting genes as potential biomarkers when the outcome is non-normally distributed eg. binary , survival e.t.c
- Mixture of Longitudinal and cross-sectional
  - Predicting the final outcome of a longitudinal sequence using earlier measure

- The case study arises from a pre-clinical study involving rats
- The rats were randomly assigned to a treatment or placebo
- They were followed for several minutes in which case several variables were measured
- list of variables
  - Cort: longitudinally measured
  - Activity : Measured cross-sectionally
  - Telemetry : Heart beat and Blood Pressure measured longitudinally.
- Objective: Measure association between the pair of each variable

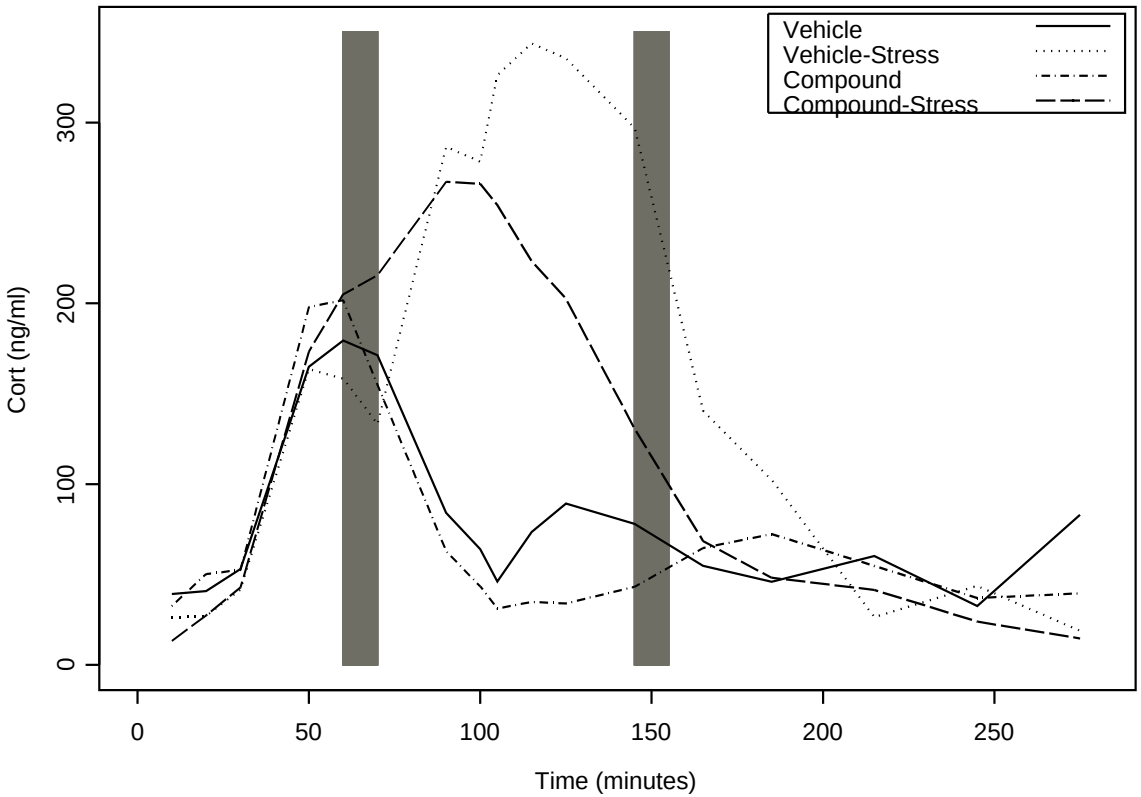

Figure: *Group-specific mean profiles of CORT values, averaged over different treatment periods. The shaded regions indicate the time windows in which activity was measured before and after the stress induction.*

- The case study arises from a depression study involving humans
- Depression level was measured by the Hamilton Depression scale (HAMD score) before and after treatment
- Blood samples were taken from which several genes and metabolites were measured before and after treatment
- We have the case of longitudinal measured outcome and several longitudinally measured biomarkers
- The objective is to select potential gene and metabolite biomarkers
- We use the methods discussed earlier to select potential gene biomarkers for the outcome

- Results for the Behavioral Study

endpoint		unstructured		fract. pol.		pen. splines	
outcome	predictor	*VRF*	R^2_Λ	*VRF*	R^2_Λ	*VRF*	R^2_Λ
Activity	CORT	0.433	0.433	0.372	0.372	0.402	0.402
CORT	Activity	0.060	0.433	0.039	0.372	0.026	0.402
Activity	heart rate	0.807	0.807	0.816	0.816	0.798	0.798
heart rate	Activity	0.119	0.807	0.069	0.816	0.071	0.798
Activity	blood pressure	0.571	0.571	0.586	0.586	0.408	0.408
blood pressure	Activity	0.081	0.571	0.073	0.586	0.011	0.408

- Results for the Biomarker case study

Gene Id	*VRF*	R^2_Λ	*Hcof*0	*Hcof*1	*Gcof*0	*Gcof*1	raw_P
12161	0.7132	0.9177	18.08102	-4.44697	-0.13936	0.446282	0.00001
9806	0.6640	0.8871	-2.04376	63.2346	-0.06813	0.369488	0.00007
4877	0.6627	0.8862	24.78267	133.7798	0.001586	0.271123	0.00008

- There is a strong association between Heart Rate and Activity but moderate relationship between Activity and blood pressure
- CORT has weak association with the Activity
- For the longitudinal outcomes proper modeling should be carried out
- Genes which have strong association were picked by the methods
- The methods can be adopted to different situations in pre-clinical and clinical settings

Collaborators

- Johnson and Johnson Pharmaceutical Research and Development
  - Luc Bijnenes
  - Pim Drinkenburg
  - Helena Geys
  - Van Den Kieboom
  - Leen Raeymaekers
  - Willem Talloen
- I-Biostat
  - Ariel Alonso
  - Dan Lin
  - John Maringwa
  - Geert Molenberghs
  - Ziv Shkedy

THANK YOU!!!