Genomic Biomarkers for a Categorical Response Variable in Early Drug Development Microarray Experiments

Suzy Van Sanden[1], Ziv Shkedy[1], Tomasz Burzykowski[1],
Hinrich Göhlmann[2], Willem Talloen[2], Luc Bijnens[2]

NCS, September 2008, Leuven

[1] Universiteit Hasselt, Center for Statistics, Agoralaan, gebouw D, B-3590 Diepenbeek, Belgium
[2] Johnson & Johnson, PRD, Turnhoutseweg 30, B-2340 Beerse, Belgium

Overview

- Introduction
- Joint Modeling Approach - Cont. Case
- Case-Study
- Joint Modeling Approach - Binary Case
- Biomarker Selection using BW-criterion
- Results
- Discussion & Conclusion

Introduction

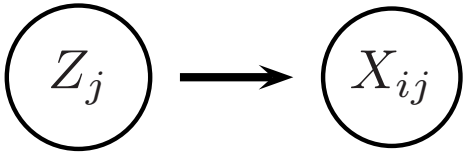
- Microarray: tools to measure the gene expression for a large number of genes at the same time

- Genomic biomarker: expression of a gene that causes a certain response (disease) or is associated with a response

  $\Longrightarrow$ **indicator** for the response

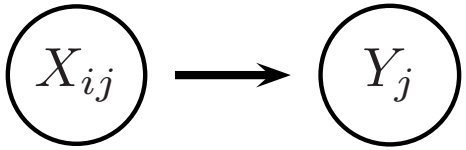
Introduction

- Microarray experiment:
  - Z_j: treatment of subject j
  - X_{ij}: gene-expression for gene i of subject j



$\Longrightarrow$ Detect genes that are differentially expressed

- Microarray biomarker experiment:
  - X_{ij}: gene-expression for gene i of subject j
  - Y_j: response of subject j



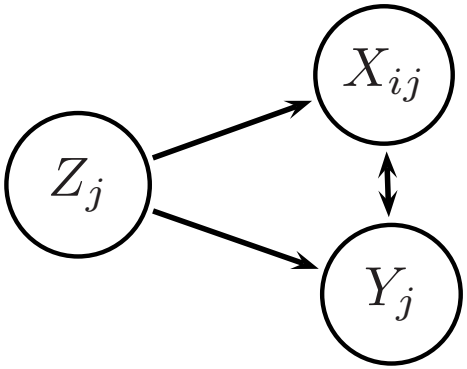
$\Longrightarrow$ Detect genes that can be used to predict the response

Introduction

- Biomarkers in early drug development studies:

(Shkedy *et al*., 2008)

  - Z_j: treatment of subject j
  - Y_j: response of subject j
  - X_{ij}: gene-expression for gene i of subject j



- Asses effect of treatment on response of interest by using information on expression levels of a group of genes

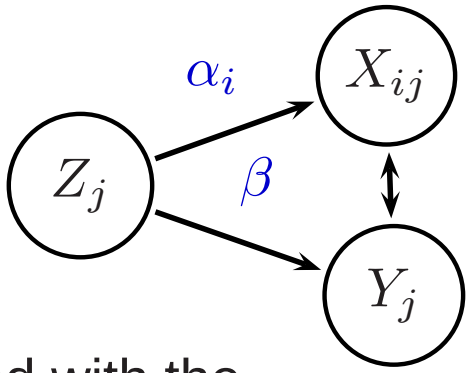
$\Longrightarrow$ Detect genes influenced by treatment and/or correlated with the response

Joint Modeling Approach - Cont. Case

- Joint model for gene-expression and continuous response:

(Shkedy *et al*., 2008)

$$\begin{pmatrix} X_{ij} \\ Y_j \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_i + \alpha_i Z_j \\ \mu_Y + \beta Z_j \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_i} & \sigma_{X_i Y} \\ \sigma_{X_i Y} & \sigma^2_Y \end{pmatrix}\right]$$



- Prognostic biomarker: Gene-expression is correlated with the response, after adjustment for treatment

  $\Longrightarrow$ correlation coefficient $\rho_i = \frac{\sigma_{X_i Y}}{\sigma_{X_i}\sigma_Y} \neq 0$

- Therapeutic biomarker: Gene-expression is affected by treatment and predictive for effect of treatment on response

  $\Longrightarrow \beta \neq 0$ and $\alpha_i \neq 0$

- Prognostic/therapeutic biomarker

Case-Study with Categorical Response

- Toxicology study on rats
- Treatment (Z_j): 3 treatment - 1 control group
- 25 animals per group (100 in total)
- Response (Y_j): Toxicity measurements (4 levels)
- Gene expression data (X_{ij}):
  - $\approx$ 31000 genes
  - only for 38 animals (about 10 per group)

Case-Study with Categorical Response

- Number of rats for different toxicity levels:

	Treatment				
Toxicity	C	T1	T2	T3	
none (0)	10	1	0	0	11
low (1)	0	3	0	1	4
medium (2)	0	6	5	3	14
high (3)	0	0	3	6	9
	10	10	8	10	38

$\Rightarrow$ Toxicity seems to depend on treatment

$\Rightarrow$ Problem of sparse data!

Case-Study with Categorical Response

- Toxicity variable dichotomized (low level - high level):

	Treatment				
Toxicity	C	T1	T2	T3	
Low toxicity	10	4	0	1	15
High toxicity	0	6	8	9	23
	10	10	8	10	38

$\Rightarrow$ Compare treatment groups 1 and 3

- Logistic regression for effect of treatment on toxicity:
  - reduced dataset (20): no difference ($p=0.1472$)
  - full dataset (50): difference ($p=0.003$)

$\Rightarrow$ Sample-size problem!

Joint Modeling Approach - Binary Case

- Latent continuous variable Y_j^* underlying binary variable Y_j

$$Y_j = \begin{cases} 1 & Y_j^* > 0 \\ 0 & Y_j^* \leq 0 \end{cases}$$

- Joint model for latent outcome Y_j^* and gene-expression X_{ij}:

$$\begin{pmatrix} X_{ij} \\ Y_j^* \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_i + \alpha_i \, Z_j \\ \mu_Y + \beta \, Z_j \end{pmatrix}, \begin{pmatrix} \sigma^2_{X_i} & \sigma_{X_i Y} \\ \sigma_{X_i Y} & \sigma^2_Y \end{pmatrix} \right]$$

- Resulting probit model formulation for Y_j and X_{ij} for gene i:

$$\begin{cases} X_{ij} \sim N(\mu_i + \alpha_i \, Z_j, \sigma^2_{X_i}) \\ Y_j \sim B(p_j) \\ \Phi^{-1}(p_j) = \mu_Y + \beta \, Z_j \end{cases}$$

  - Constraint: σ^2_Y=1
  - $B(p_j)$: Bernoulli distribution
  - $p_j = P(Y_j = 1)$
  - Φ: standard normal cum. dist.

- SAS procedure GLIMMIX

Joint Modeling Approach - Binary Case

- Prognostic biomarker: $\rho_i = \frac{\sigma_{X_i Y}}{\sigma_{X_i}\sigma_Y} \neq 0$
  - Interpretation: correlation coefficient for binary Y_j and X_{ij}
    $\longrightarrow$ correlation between cont. Y_j^* and X_{ij} after correction for treatment
    (Renard *et al.*, 2002)
  - $H_0 : \rho_i = 0$ versus $H_1 : \rho_i \neq 0$ (LR test)
  - Bonferroni correction (5% sign. level): no genes
- Potential therapeutic biomarker: $\alpha_i \neq 0$
  - $H_0 : \alpha_i = 0$ versus $H_1 : \alpha_i \neq 0$ (T-test)
  - Bonferroni correction (5% sign. level): 33 genes

Joint Modeling Approach - Binary Case

- Remarks about the modeling approach in the binary case:
  - Definition of prognostic biomarker?
  - Application is limited:
    - Problems with sparse data
    - Only binary response data (GLIMMIX procedure)

- Remarks about hypothesis testing in the binary case:
  - Advantage: Reduces risk of chance finding
  - Disadvantage: Not necessarily best subset for classification
    - Individual genes $\longleftrightarrow$ Group of genes for classification
    - Too many genes filtered out $\Longrightarrow$ Loss of classification information
    - Too few genes selected $\Longrightarrow$ Not enough to reduce noise
  - Sample size problem: not enough power?

$\Longrightarrow$ Ranking-based approach for biomarker selection

Alternative Approach: BW-criterion

- Biomarker Selection: top p genes with largest BW-ratio

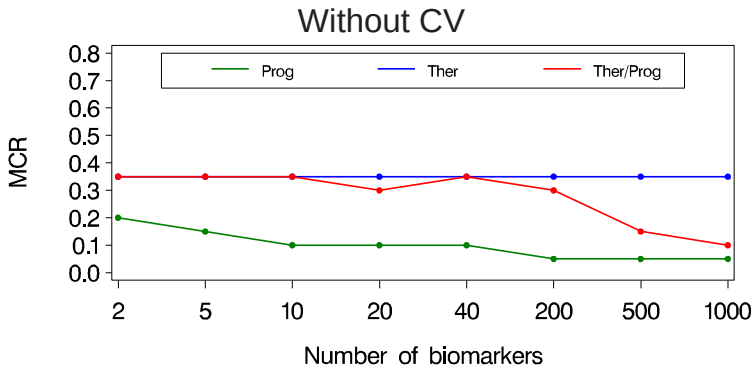
$$\text{BW} = \frac{\text{between-group sum of squares}}{\text{within-group sum of squares}}$$

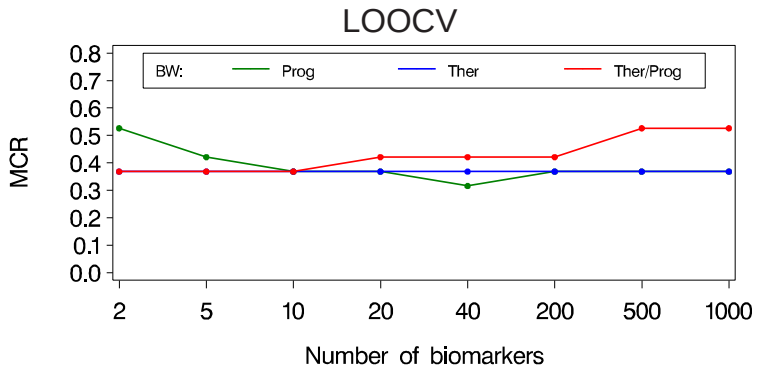
- Choice of grouping variable:
  - Response level ($\text{BW}_{Response}$) $\longrightarrow$ Potential prognostic biomarkers
  - Treatment group (BW_{Treat}) $\longrightarrow$ Potential therapeutic biomarkers
  - Combination ($\text{BW}_{Resp-Treat}$) $\longrightarrow$ Potential therapeutic/prognostic biomarkers

    $\hookrightarrow$ Rank = sum of ranks from $\text{BW}_{Response}$ and BW_{Treat}

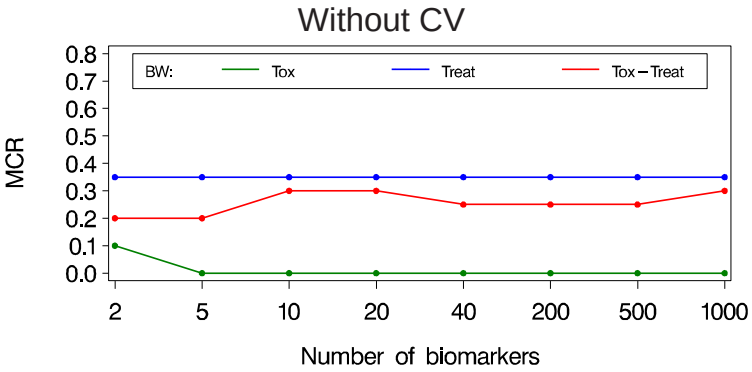
MCR (DLDA) for Toxicology Study

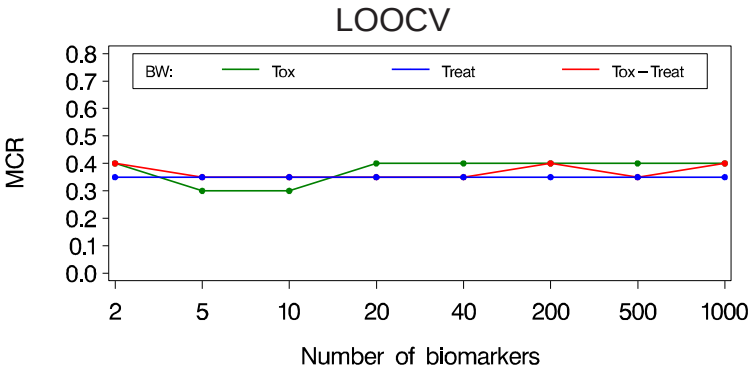
- Toxicity: Low - High, Treatment: T1 - T3 (20 Samples)
- Joint modelling approach:
  - 33 potential therapeutic biomarker: MCR = 0.35
  - Ranking according to p-value:





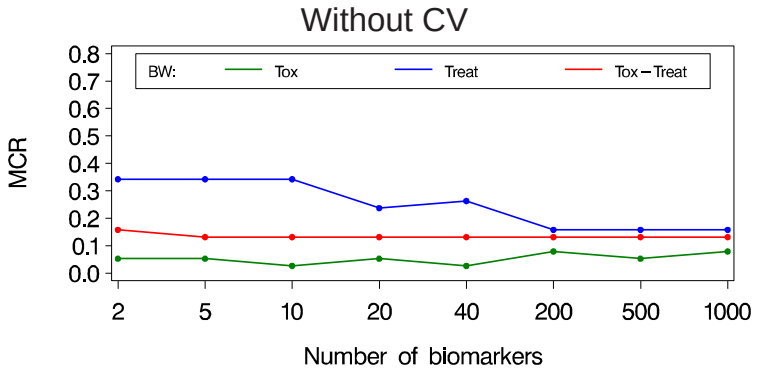
- BW-criterion:

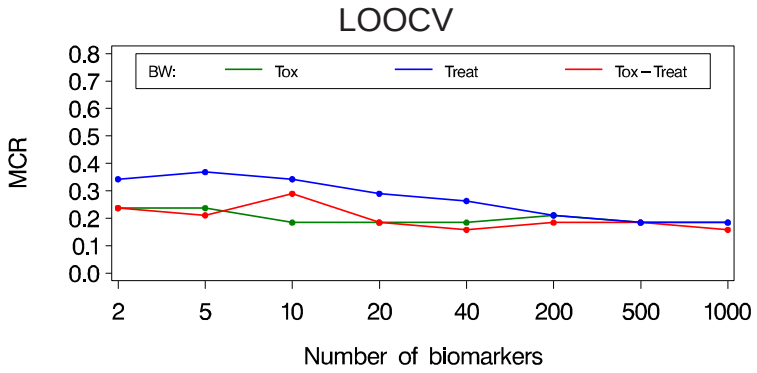




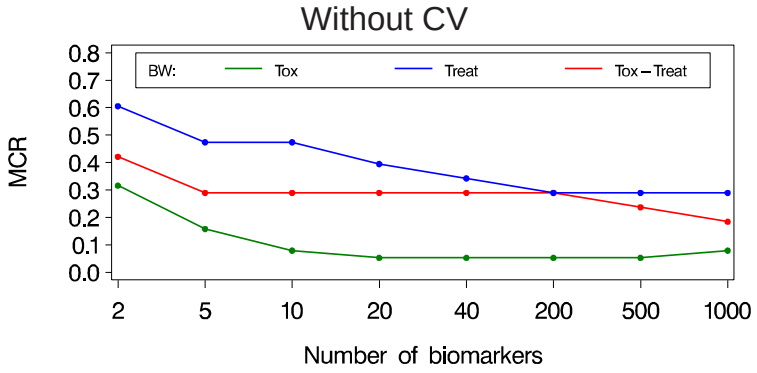
MCR (DLDA) for Toxicology Study

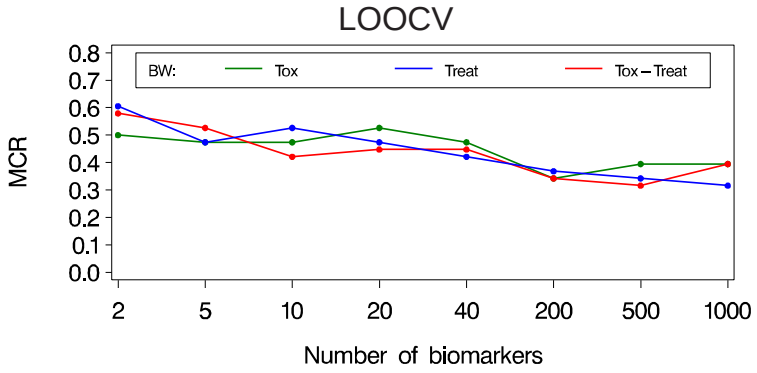
- BW-criterion
- Low - high toxicity – 4 treatment groups (38 samples):





- 4 levels of toxicity – 4 treatment groups (38 samples):





Discussion

- Correspondence (modelling approach – BW-ratio) for therapeutic biomarkers

- Alternative definition of prognostic biomarkers:

  - Model: Linear association between gene-expression and response after correction for treatment

  - BW-ratio: Ability to separate samples between levels of response variable

- How to choose optimal number of biomarkers?

Conclusion

- Two approaches for biomarker selection:
  - Joint-modeling in a binary setting
    - Computationally intensive
    - Problematic for sparse data
    - Definition prognostic biomarker?
  - BW-criterion in a categorical setting
- Detection of biomarkers (subgroup of gene) influenced by treatment (therapeutic) and/or that can discriminate between the response levels (prognostic)

References

- Renard, D., Geys, H., Molenberghs, G., Burzykowski, T., and Buyse, M. (2002) Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical*, **44**, 921–935.

- Shkedy, Z., Lin, D., Molenberghs, G., Göhlmann, H., Talloen, W., and Bijnens, L. (2008) Gene-specific and joint surrogacy in microarray pre-clinical experiments. *Submitted*.

- Van Sanden, S., Shkedy, Z., Burzykowski, T, Göhlmann, H., Talloen, W., and Bijnens, L. (2008) Genomic Biomarkers for a Binary Clinical Outcome in Early Drug Development Microarray Experiments. *Submitted*.