Biomarkers for a Categorical Response Variable in early Drug Development Microarray Experiments

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

[1]Universiteit Hasselt, Center for Statistics, Agoralaan, gebouw D, B-3590 Diepenbeek, Belgium

[2]Johnson & Johnson, PRD, Turnhoutseweg 30, B-2340 Beerse, Belgium

Microarrays experiments have become a popular tool to examine the expression of thousands of genes at the same time. In early drug development experiments, the primary interest of researchers lies in finding genes that are differentially expressed when different condition (treatments) are tested. Recently, microarray data have also been considered as a means to select genes capable of serving as a biomarker for a primary response variable. Within this framework, one wants to assess the effect of a treatment on the response of interest by using information about the expression levels of a group of genes.

Shkedy et al. (2007) have considered a joint model for the gene expression and the response in pre-clinical experiments. The model allows to detect differentially expressed genes and to evaluate genes as biomarkers. In their study the response of primary interest was a continuous variable. In this paper we proposed a new model for biomarkers detection and evaluation for categorical response data. In particular, the response of primary interest is toxicity and the aim of the analysis is to find a subset of genes which can serve as biomarkers for toxicity. The method is applied to data containing information about toxicity levels of 38 subjects for whom microarray data for approximately 31000 genes are available.

Several approaches for gene selection are considered, amongst which an extension of the joint model to the categorical setting. We formulate a joint model for the binary response and the gene expression and we have shown that gene selection based on the joint modelling approach is similar to classical selection methods such as the BW ratio approach (Dudoit et al., 2002). The proposed method leads to the identification of a subgroup of genes (biomarkers) which are either treatment related, toxicity related or toxicity and treatment related from which toxicity can be predicted. In particular the new approach of identifying toxicity and treatment related genes allows investigators to focus on those genes that are influenced by the treatment and can discriminate between high or low toxicity levels. The new methodology and the software developed in this study will provide an important tool to investigators in the pharmaceutical industry for detection and prediction of toxicity using information from functional genomic experiments.

Keywords: Biomarkers; Microarray Experiments; Categorical Data; Joint Modelling.