Comparison of Different Designs for Simultaneous Analysis of Concentration-Response Gene Expression Data

Leonie Schürmeyer[1], Kirsten Schorning[1], Jörg Rahnenführer[1]

[1]Faculty of Statistics, TU Dortmund University, Germany (presenting author)

Understanding the concentration-response relationship of a candidate drug is one of the main goals in toxicology and especially drug development. Therefore, several authors emphasized the importance of optimal design regarding concentration-response experiments. Using classical optimal design approaches significantly enhances the precision of the estimated concentration-response curve or rather specific parameters. Thus, optimal designs can substantially improve comprehension in toxicological context.

We extend classical approaches of optimal design of experiments (see [Pukelsheim] among many others) such that they can be applied for the analysis of concentration-response relationships in the context of gene expression data. The experimental conditions of such data are new challenges in planning those experiments, since all genes are elevated simultaneously. Thousands of concentration-response relationships need to be determined, so the key question is which design works best for the simultaneous analysis. Thereby gene expression data of valproic acid applied to human embryonic stem cells is analyzed to compare different designs [Krug et al.]. First of all genes with biologic activity are evaluated with the Multiple Comparison Procedure and Modelling approach [Bretz et al.]. Simultaneously the true underlying concentration-response relationships are fitted using separate sigmoid Emax models for all active genes. Then locally D-optimal designs are identified for every considered gene. Based on the locally D-optimal design, a summarized design is established using the K-means algorithm. Moreover, a D-optimal design for simultaneous inference is developed by using an appropriate Bayesian D-optimality criterion.
Both developed designs are compared to the design, originally used in the experimental set-up, an equidistant and log-equidistant design with respect to their D-efficiencies. Moreover, a simulation study is conducted to demonstrate the differences of all designs practically. In that simulation study, we also vary the total sample size to investigate its influence on the precision of the model fits.

The results actively demonstrate to support the consideration of using optimal design approaches for gene expression data. Especially the D-optimal design for simultaneous inference and the design developed with K-means perform considerably better than the originally used design and the log-equidistant design, both in terms of the theoretical and the practical comparison. Measured by the root mean squared error (RMSE) the original, equidistant and log-equidistant design have an inferior performance in the simulation study compared to the two other designs. Besides the precision of the model fits highly increases by enlarging the total sample size. Thus, it is recommendable using as many observations as possible, whereupon minimal 27 data points should be used in this analysis. Summarizing, the D-optimal design for simultaneous inference leads to the most exact model fits, consequently, it should be preferred to the other designs investigated.

[Bretz et al.] Bretz, F., Pinheiro, J. C. and Branson, M. (2005): Combining multiple comparisons and modeling techniques in dose-response studies, *Biometrics*, **61**(3), p. 738-748.
[Pukelsheim] Pukelsheim, F. (2006): *Optimal Design of Experiments*, Wiley, New York.
[Krug et al.] Krug, A. K., Kolde, R., Gaspar, J. A., Rempel, E., Balmer, N. V., Meganathan, K., Vojnits, K., Baquie, M., Waldmann, T., Ensenat-Waser, R., Jagtap, S., Evans, R. M., Julien, S., Peterson, H., Zagoura, D., Kadereit, S., Gerhard, D., Sotiriadou, I., Heke, M., Natarajan, K., Henry, M., Winkler, J., Marchan, R., Stoppini, L., Bosgra, S., Westerhout, J., Verwei, M., Vilo, J., Kortenkamp, A., Hescheler, J. R., Hothorn, L., Bremer, S., van Thriel, C., Krause, K. H., Hengstler, J. G., Rahnenfuhrer, J., Leist, M. and Sachinidis, A. (2013): Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach. *Arch Toxicol*, **87**, p. 123–143.