Clinical Trial Designs with Data-Driven Selection of Subgroups

Niewczas, J.[1], Paoletti, X.[2], Michiels, S.[2], Ballarini, N.[1], Posch, M.[1], Koenig, F.[1]

[1]Institute of Medical Statistics, Medical University of Vienna, Vienna, Austria
[2]Service de Biostatistique et d'Epidémiologie & CESP OncoStat, INSERM, Gustave Roussy, Université Paris-11 Saclay, Villejuif, France

Objective

Clinical trials allowing for confirmatory testing of pre-defined subgroups in addition to investigating the treatment effect in the overall study populations have gained popularity [1, 3, 4, 5]. However, often too optimistic assumptions are taken on the actual size of the biomarker defined subgroups in the planning phases of trials. For example, if the observed prevalence is lower than expected, we might experience power loss or a delay in the trial conduct. In oncology trials, there are often several genetically defined subpopulations available. Instead of pre-defining the test of a single subgroup, we investigate clinical trial designs in which the subgroup to be tested is selected based on the data observed in the trial, e.g. by selecting the largest subgroup. The observed data is used to derive multiplicity adjusted tests.

Subgroup Selection Methods

Consider a clinical trial with an experimental treatment being compared to a standard of care, in which there are also some potential candidate subgroups that might be of interest and one of them will be tested at the end of the trial. The aim is to find a subgroup that would have the highest chance to reject the null hypothesis. For simplicity we assume that only one subgroup will be selected to be tested at the end of the trial.
Let F denote the full population and S_k $(k = 1, ..., m)$ the subgroups of interest, whereby here we just consider the situation of disjoint subgroups.
Three data-driven approaches for subgroup selection at the end of the trial are considered:

- the highest observed prevalence,
- the highest product of the observed prevalence and pooled blinded mean,
- the highest pooled blinded variance.

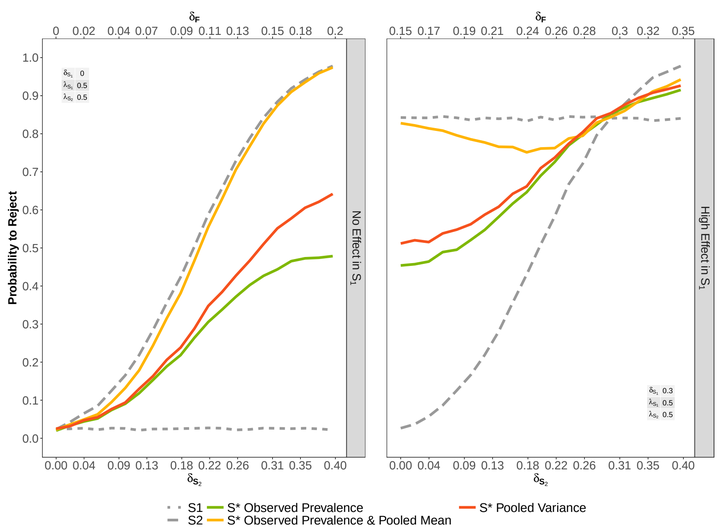
The approaches are later compared to tests with pre-defined subgroups. All these procedures will control the type I error despite the data dependent selection, as under the null these selection rules are independent from the test statistics.

Simulations

Consider a setting with two disjoint subgroups with prevalences: $\lambda_{S_1} = \lambda_{S_2} = 0.5$. Let the total sample size be equal to 800 patients with desired effect in the full population $\delta_F = 0.2$ to achieve power $1 - \beta_F = 0.8$ and with one-sided type I error $\alpha = 0.025$. Consider no and a high effect in S_1: $\delta_{S_1} = (0, 1.5\delta_F)$ and vary the effect δ_{S_2} from 0 to $2\delta_F$. Under such scenario the effect in F varies depending on the effect sizes in the subgroups. Let S^* denote the selected subgroup to be tested at the end of the trial.

Results

In the Figure below one can see the probability to reject the selected subgroup for the three selection approaches compared to testing pre-defined S_1 or S_2 only. It can be seen that data-driven selection procedures are more robust than simply choosing S_1 or S_2 if the effects in pre-selected subgroups are low. The product of pooled mean and observed prevalence outperforms the other two selection procedures.

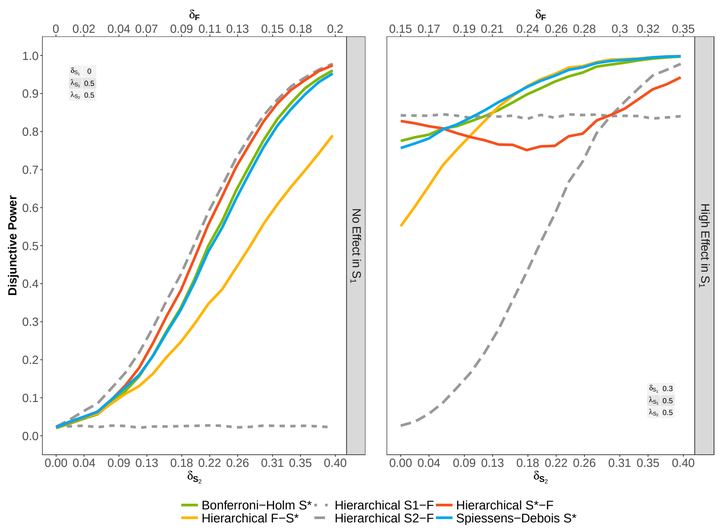


Multiplicity Adjustments

After a subgroup is selected a multiplicity adjustment for testing F and S^* within the closed testing is performed (see [2]). We consider 4 multiple testing procedures:

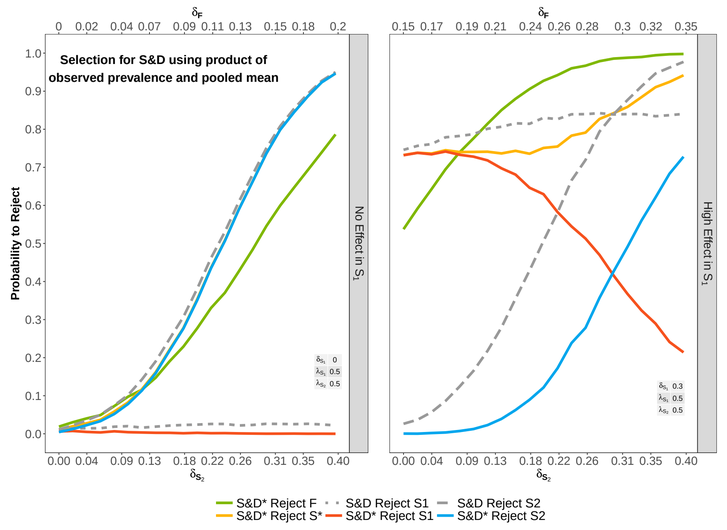
- Bonferroni-Holm,
- Hierarchical procedure where S^* can only be rejected if F is rejected,
- Hierarchical procedure where F can only be rejected if S^* is rejected,
- Method proposed by Spiessens and Debois (2010) using group sequential methodology utilising observed prevalence (S-D).

The four methods and a hierarchical testing procedure with pre-defined selection of either S_1 or S_2 are compared. For S-D approach the significance value for testing F was set to: $\alpha_F = 0.02$. Plot below shows the disjunctive power (power to reject any hypothesis) for all approaches. It can be seen that all hierarchical tests perform well only if the effects in chosen subgroups are large. Whenever the effects in subgroups are different, Bonferroni-Holm and S-D outperform hierarchical procedures and are more robust. As 2% of α was spent on F in S-D approach it has slightly lower power when F does not have a high effect.



Probability to Reject Individual Hypotheses

In the Figure below probabilities to reject individual hypotheses: the full population, any subgroup or individual subgroups are reported for the selection procedure based on observed prevalence and pooled mean for S-D multiplicity adjustment. It can be seen that the power to reject the selected subgroup and the power to reject the full population tend to have values similar to those of pre-defined subgroups with a high effect, allowing one to make the correct selection procedure at a low cost of power loss.



Conclusions

- Data-driven subgroup selection procedures can be a valuable and robust strategy for subgroup selection.
- Method proposed by Spiessens and Debois (2010) is more robust than other approaches (different α allocation can be further investigated).
- If effects in subgroups are different, S-D and Bonferroni-Holm gain a lot of power compared to pre-defined subgroup selection.

References

[1] S. Kropf and J. Läuter. Multiple tests for different sets of variables using a data-driven ordering of hypotheses, with an application to gene expression data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 44(7):789–800, 2002.

[2] R. Marcus, P. Eric, and K. Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.

[3] T. Ondra, A. Dmitrienko, T. Friede, A. Graf, F. Miller, N. Stallard, and M. Posch. Methods for identification and confirmation of targeted subgroups in clinical trials: a systematic review. *Journal of biopharmaceutical statistics*, 26(1):99–119, 2016.

[4] Y. Song and G. Y. Chi. A method for testing a prespecified subgroup in clinical trials. *Statistics in medicine*, 26(19):3535–3549, 2007.

[5] B. Spiessens and M. Debois. Adjusted significance levels for subgroup analyses in clinical trials. *Contemporary clinical trials*, 31 (6):647–656, 2010.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567.

Contact Information

- julia.niewczas@meduniwien.ac.at