Multiplicity Issues in Exploratory Subgroup Analysis

Alex Dmitrienko (Mediana)
Ilya Lipkovich (QuintilesIMS)

Quantitative Methods for Drug Discovery and Development
July 2017

Outline

Exploratory subgroup analysis in clinical drug development

Regulatory considerations

Guideline-driven and principle-driven approaches to subgroup identification

Multiplicity adjustments that support reliable subgroup search strategies

Case study

ATTAIN Phase III development program

Exploratory subgroup analysis

Subgroup analysis

Subgroup analysis approaches

Several classification schemes proposed in clinical trial literature (Varadhan et al., 2013; Lipkovich and Dmitrienko, 2014b)

Simplified classification scheme

Confirmatory subgroup analysis relies on a small set of well defined patient subgroups

Exploratory subgroup analysis focuses on a large set of loosely defined patient subgroups

ATTAIN Phase III program

Two clinical trials in nosocomial pneumonia

Total sample size: 1289 patients

Primary endpoint: All-cause mortality at 28 days

Overall outcome: Negative treatment effect in overall patient population

Exploratory objective

Identify biomarkers that help predict positive treatment response and select patient subgroups with beneficial effect

Exploratory subgroup analysis

FDA guidance

Guidance for Industry: Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products (2012)

EMA guidance

Guideline on The Investigation of Subgroups in Confirmatory Clinical Trials (2014)

Exploratory subgroup analysis

FDA lectures

Invited half-day lecture on SIDES method at CDER (Jan 2013)

EMA meetings

Invited talks to provide an overview of key issues in subgroup analysis at two EMA expert subgroup analysis workshops (London, Nov 2011 and Nov 2014)

Applications of exploratory subgroup analysis

Scenario 1 (positive trial)

Assess consistency of treatment effects across key subgroups

Scenario 2 (positive trial)

Analyze subgroups in a post-hoc manner to (1) exclude a subgroup due to lack of efficacy or (2) focus on a subgroup without safety issues

Add a subgroup with enhanced treatment effect

Scenario 3 (negative trial)

Discover subgroups with enhanced efficacy profile

Applications of exploratory subgroup analysis

Scenario 1 (positive trial)

Consistency assessment

Scenario 2 (positive trial)

Subgroup identification/discovery

Scenario 3 (negative trial)

Subgroup identification/discovery

Scenario 2 (positive trial)

Subgroup identification

Product label is restricted to a subgroup due to lack of effect in the complementary subgroup

PROWESS trial

Xigris trial in patients with severe sepsis

Four subgroups based on predicted risk of mortality: No effect in Subgroups 1 and 2 (low risk) and positive effect in Subgroups 3 and 4 (high risk)

Product label was restricted to Subgroups 3 and 4

Scenario 3 (negative trial)

Subgroup identification

Positive effect is identified in a subgroup (replication of positive effect is required)

PRAISE I trial

Amlodipine trial in patients with with chronic heart failure

Borderline positive effect in overall population and highly beneficial effect in non-ischemic patients

PRAISE II trial was conducted in non-ischemic patients but failed

Guideline-driven and principle-driven approaches to subgroup identification

Post-hoc subgroup identification

Guideline-driven approaches

Multiple sets of guidelines attempt to improve credibility of exploratory subgroup analysis

Checklist with 25 rules (Brookes et al., 2001), checklist with 21 rules (Rothwell, 2005), etc

Main rule: Proceed with caution

Principle-driven approaches

Subgroup identification ought to be based on specific operationalizable principles

Principled subgroup identification

Key idea

Utilize recent developments in machine learning and data mining to pre-specify a principled subgroup exploration approach (Lipkovich, Dmitrienko and D'Agostino, 2017)

Key principles of subgroup identification

Principled subgroup identification

Key principles of subgroup identification

Protect Type I error rate/false discovery rate

Restrict complexity of search space

Perform reliable inferences in selected subgroups

Underlying theme

Multiplicity control and selection bias control

Subgroup identification methods

Global outcome modeling

Virtual Twins method (Foster et al., 2011), Bayesian subgroup search (Xu et al., 2016)

Global treatment effect modeling

Tree-based methods, e.g., Interaction Trees method (Su et al., 2009), GUIDE framework (Loh, 2002; Loh, He and Man, 2015)

Local modeling

Responder Identification method (Kehl and Ulm, 2006), SIDES method (Lipkovich et al., 2011)

SIDES

Subgroup Identification based on Differential Effect Search

Recursive partitioning-based subgroup identification method which provides a multivariate assessment of biomarkers

Family of subgroup search methods

General subgroup search method: SIDES

Enhanced subgroup search method: Two-stage method with biomarker screening (SIDEScreen)

Clinical trial applications

Large multinational Phase III trial in patients with Type 2 diabetes

SIDES method was applied in a retrospective manner to find subgroups with improved efficacy profile (Hardin et al., 2013, *Journal of Diabetes Science and Technology*)

Phase III trials in patients with pneumonia

Retrospective analysis of 26 biomarkers to identify subgroups with enhanced efficacy (Dmitrienko et al., 2015, *Applied Statistics in Biomedicine and Clinical Trials Design*)

Multiplicity adjustments in subgroup identification

ATTAIN Phase III program

Biomarkers

Candidate set included 26 biomarkers (labelled X_1 through X_{26})

Large set of candidate biomarkers creates a vast search space

Important biomarkers

X_1: Patient's age

X_2: APACHE II score

X_{11}: Creatinine clearance

ATTAIN Phase III program

Exploratory subgroup analysis

Identify biomarkers that help predict positive treatment response

Select subgroups of patients who are likely to experience a beneficial treatment effect

Multiplicity control

Role of multiplicity control

Local and global multiplicity adjustments to support reliable subgroup identification strategies

Local multiplicity adjustments

Adjustments aimed at reducing selection bias due to biomarkers with a large number of values

Global multiplicity adjustments

Adjustments aimed at preserving overall Type I error rate (probability of incorrect subgroup discovery when no promising subgroups exist)

Local multiplicity adjustments

Selection bias

Biomarkers with a large number of values have an advantage over biomarkers with a few values

Selection bias adjustments have been studied in the context of recursive partitioning algorithms (Loh and Shih, 1997; Hothorn et al., 2006; Zeileis et al., 2008)

Šidák-based adjustment to help create a level playing field by penalizing biomarkers with a larger number of values (Lipkovich and Dmitrienko, 2014b)

Global multiplicity adjustments

Overstated statistical significance

Treatment effect in selected patient subgroups tends to be highly statistically significant

Global multiplicity adjustments enable a reliable assessment of statistical significance in individual subgroups

Multiplicity adjustment

Resampling-based multiplicity adjustments are recommended

Permutation-based multiplicity adjustments

Null distribution

Remove treatment effect from all subsets of overall population by using permutation methods (treatment labels and outcomes are randomly permuted)

Generate subgroups from null distribution

Apply SIDES method to generate subgroups from the null distribution in each permuted data set

Select the best subgroup and evaluate treatment effect

Adjusted p-values

Selected patient subgroups

S_i, $i = 1, \ldots, k$, Subgroups

p_i, Treatment effect p-value in Subgroup S_i

Multiplicity-adjusted p-values

Consider m permuted data sets

q_j, $j = 1, \ldots, m$, Treatment effect p-value in the best subgroup in the jth permuted data set

$$\tilde{p}_i = \frac{1}{m} \sum_{j=1}^{m} I(q_j \leq p_i), \; i = 1, \ldots, k$$

Global multiplicity adjustments

Multiplicity adjustment

Multiplicity correction is a penalty that increases the treatment effect p-values

Burden of multiplicity

Total number of generated subgroups (size of search space) has direct impact on the degree of multiplicity adjustment

Complexity control leads to a smaller set of final subgroups and reduces multiplicity burden/multiplicity penalty

Complexity control

Main objective

Reduce the size of search space to facilitate the assessment of clinical relevance and reduce multiplicity burden

Subgroup pruning rules

Growth restrictions (rules for child subgroup selection)

Sample size and treatment effect restrictions

Complexity control

Treatment size restrictions

Child subgroup is selected if $p_C \leq \gamma p_P$

p_C, p_P, Treatment effect p-values in child and parent subgroups

$0 < \gamma \leq 1$, Child-to-parent ratio

Examples

$\gamma = 1$, Liberal child subgroup selection rule

$\gamma \leq 0.5$, Stricter child subgroup selection rule

ATTAIN Phase III program

Subgroup search strategies

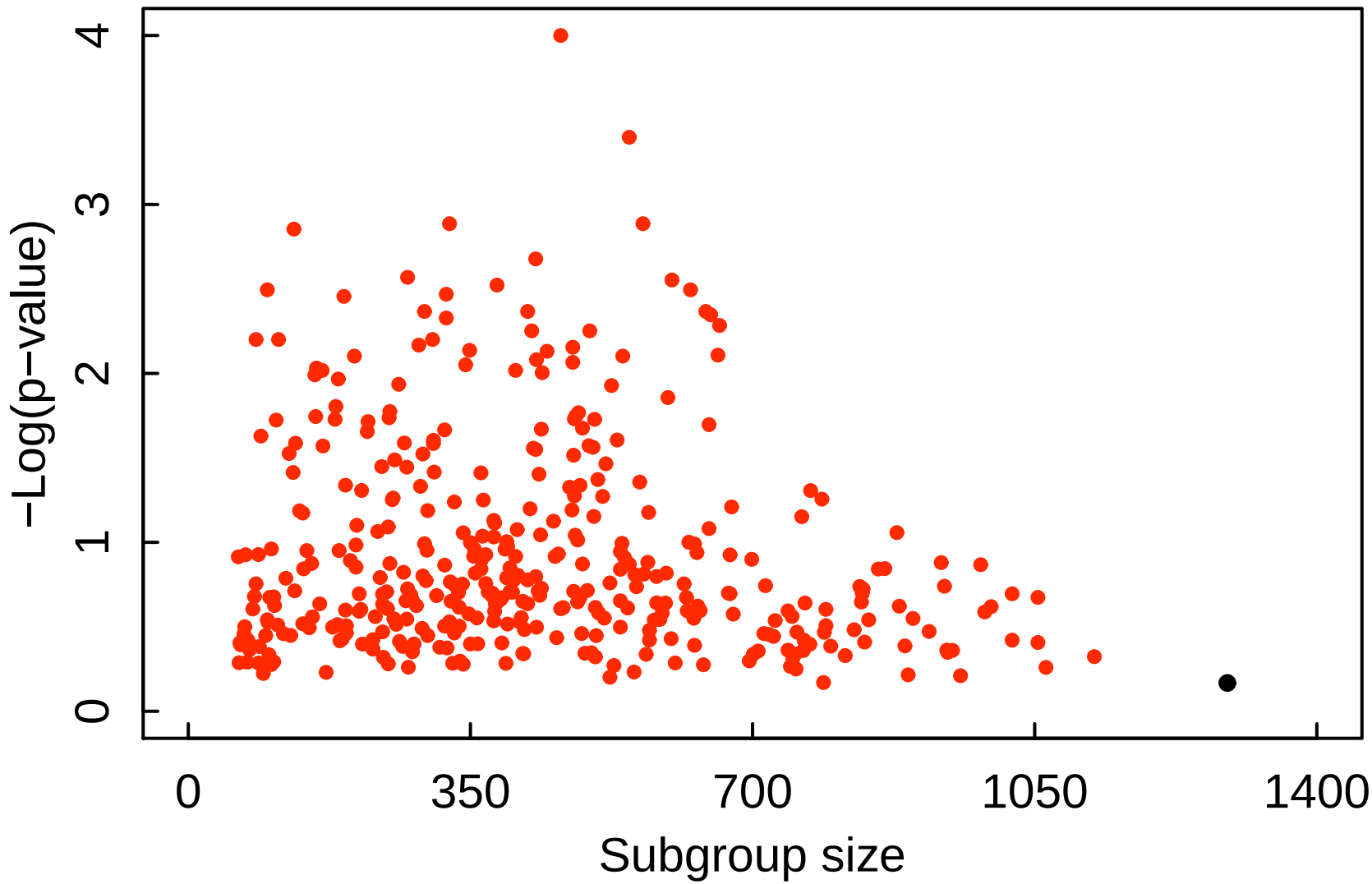
Scenario 1: No complexity control [390 subgroups]

Scenario 2: Complexity control (child-to-parent ratio $\gamma = 1$) [16 subgroups]

Scenario 3: Complexity control (child-to-parent ratio $\gamma = 0.5$) [12 subgroups]

Scenario 4: Complexity control (child-to-parent ratio $\gamma = 0.25$) [3 subgroups]

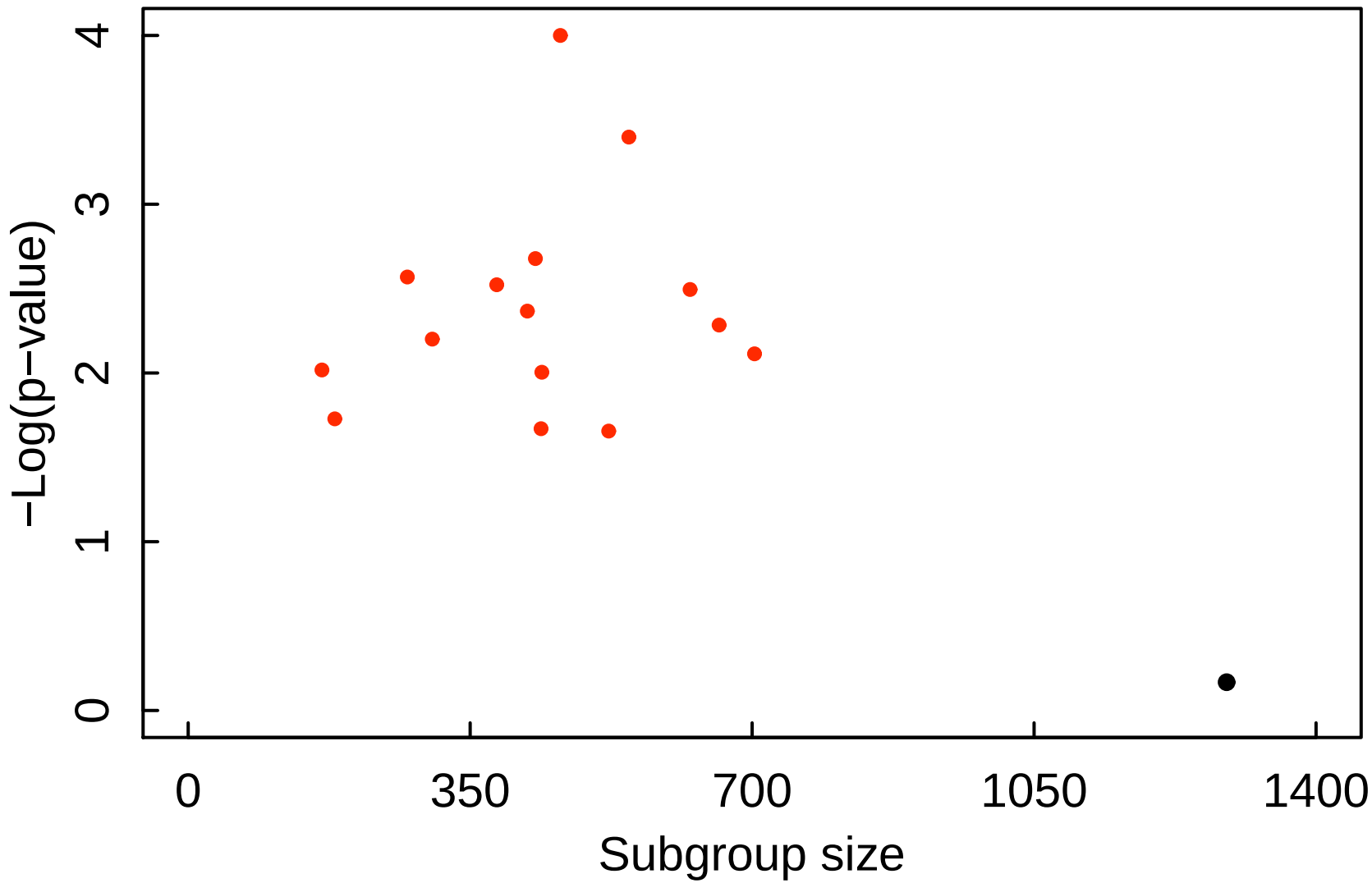
Scenario 1 (390 subgroups)



Red: Patient subgroups

Black: Overall patient population

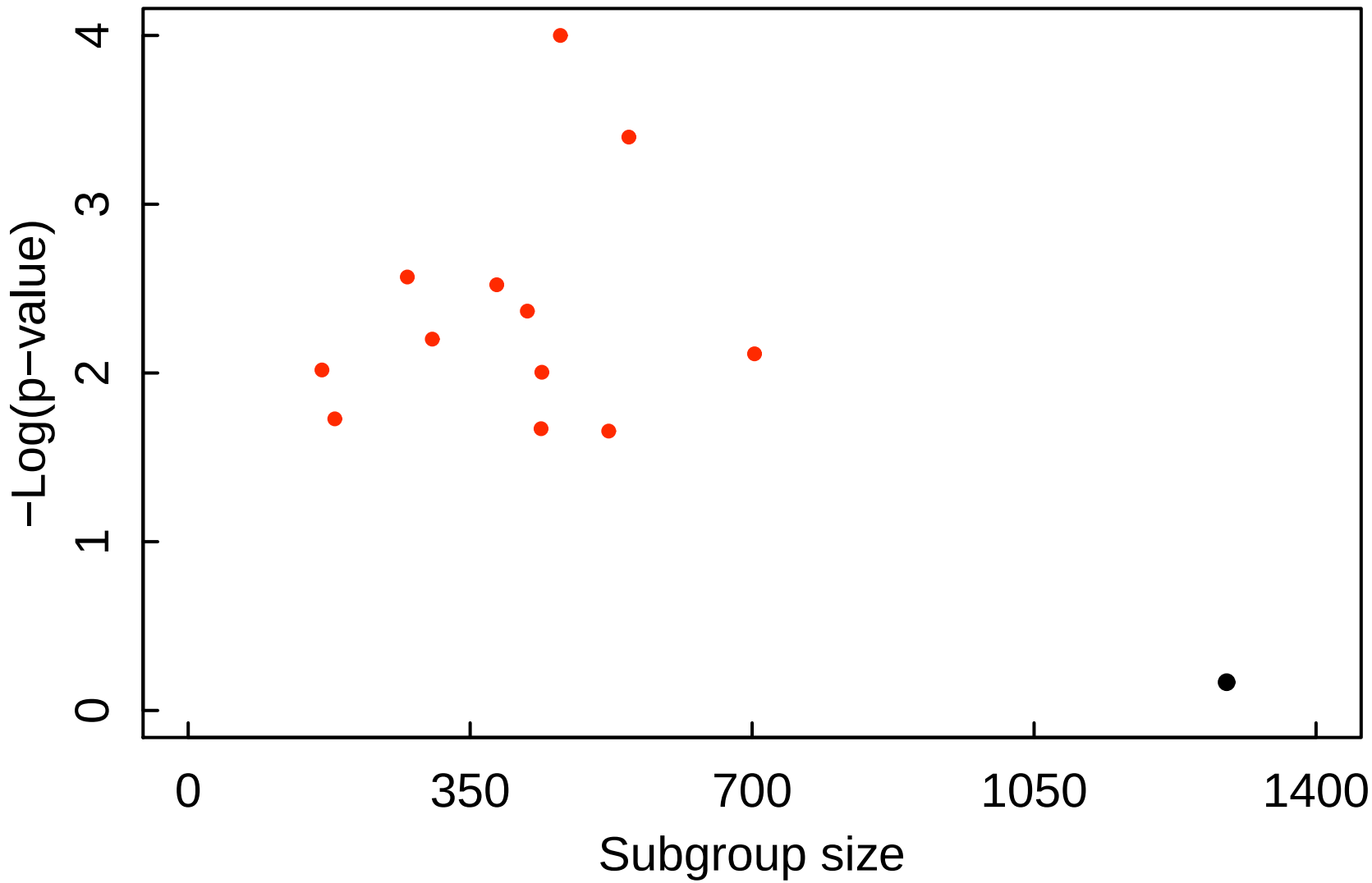
Scenario 2 (16 subgroups)



Red: Patient subgroups

Black: Overall patient population

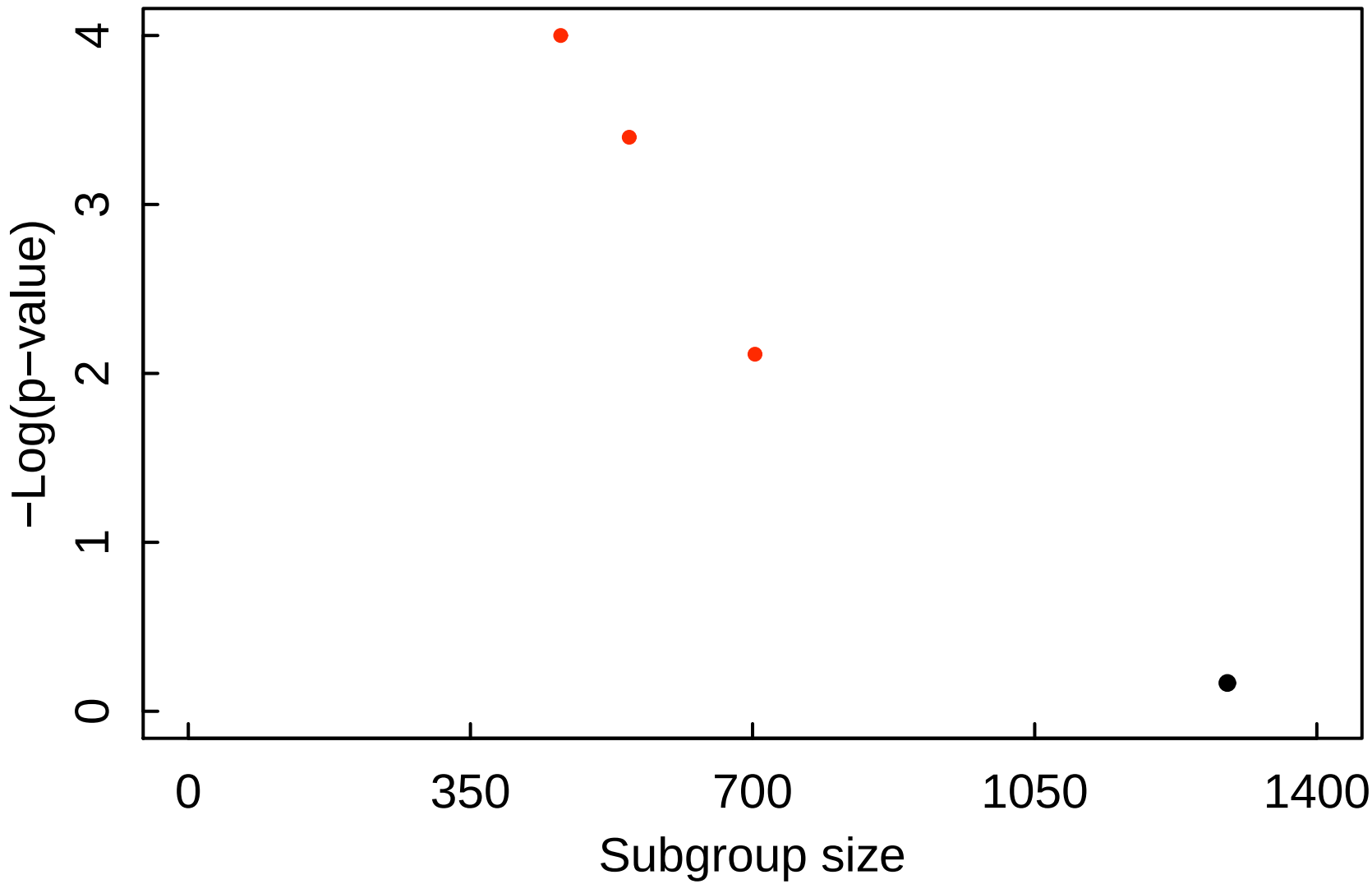
Scenario 3 (12 subgroups)



Red: Patient subgroups

Black: Overall patient population

Scenario 4 (3 subgroups)



Red: Patient subgroups

Black: Overall patient population

Role of complexity control

Subgroup selected in Scenario 4

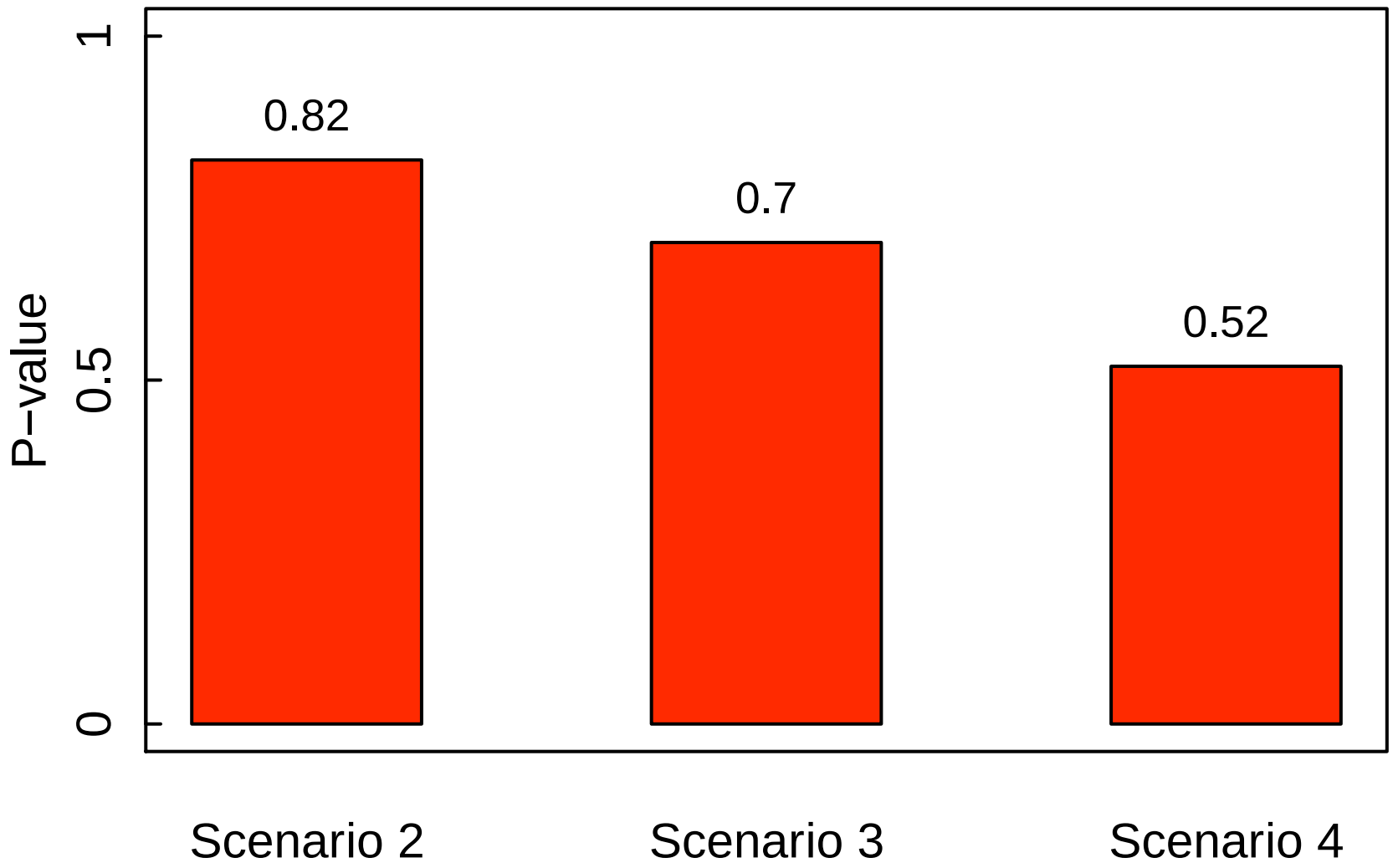
Subgroup $S_{11} = \{X_{11} > 67\}$

Raw treatment effect p-value: $p_{11} = 0.0077$

Permutation-based multiplicity adjustment

Adjusted treatment effect p-values were computed under Scenarios 2, 3 and 4

Adjusted p-values in Subgroup S_{11}



Based on 10,000 permutations

Role of complexity control

Global multiplicity adjustment

Due to a smaller search space, lower multiplicity penalty under Scenario 4 (strict complexity control) compared to Scenario 2 (liberal complexity control)

However treatment effect in Subgroup $S_{11} = \{X_{11} > 67\}$ is not even remotely significant under Scenario 4

68-fold increase in treatment effect p-value (from 0.0077 to 0.52)

Role of complexity control

Strict complexity control

Main objective of complexity control is to slow the growth of subgroup trees

Even with the most stringent complexity control/subgroup pruning, subgroup search algorithm is overwhelmed by non-informative (noise) biomarkers

Subgroup pruning does not address the fundamental problem of noise biomarkers

SIDEScreen-based subgroup search

Stage 1

Identify most informative/predictive biomarkers based on variable importance

Stage 2

Apply SIDES subgroup search algorithm to biomarkers selected in Step 1

SIDEScreen procedures

Aggressive pruning rules to reduce the search space and biomarker screening rules to filter out non-informative biomarkers

SIDEScreen method

Variable importance

Variable importance (VI) score quantifies the predictive ability of a biomarker

Fixed biomarker screening rule

Select a fixed number of biomarkers with highest VI scores

Adaptive biomarker screening rule

Selects biomarkers with "significant" VI scores (based on null distribution of the maximum VI score)

ATTAIN Phase III program

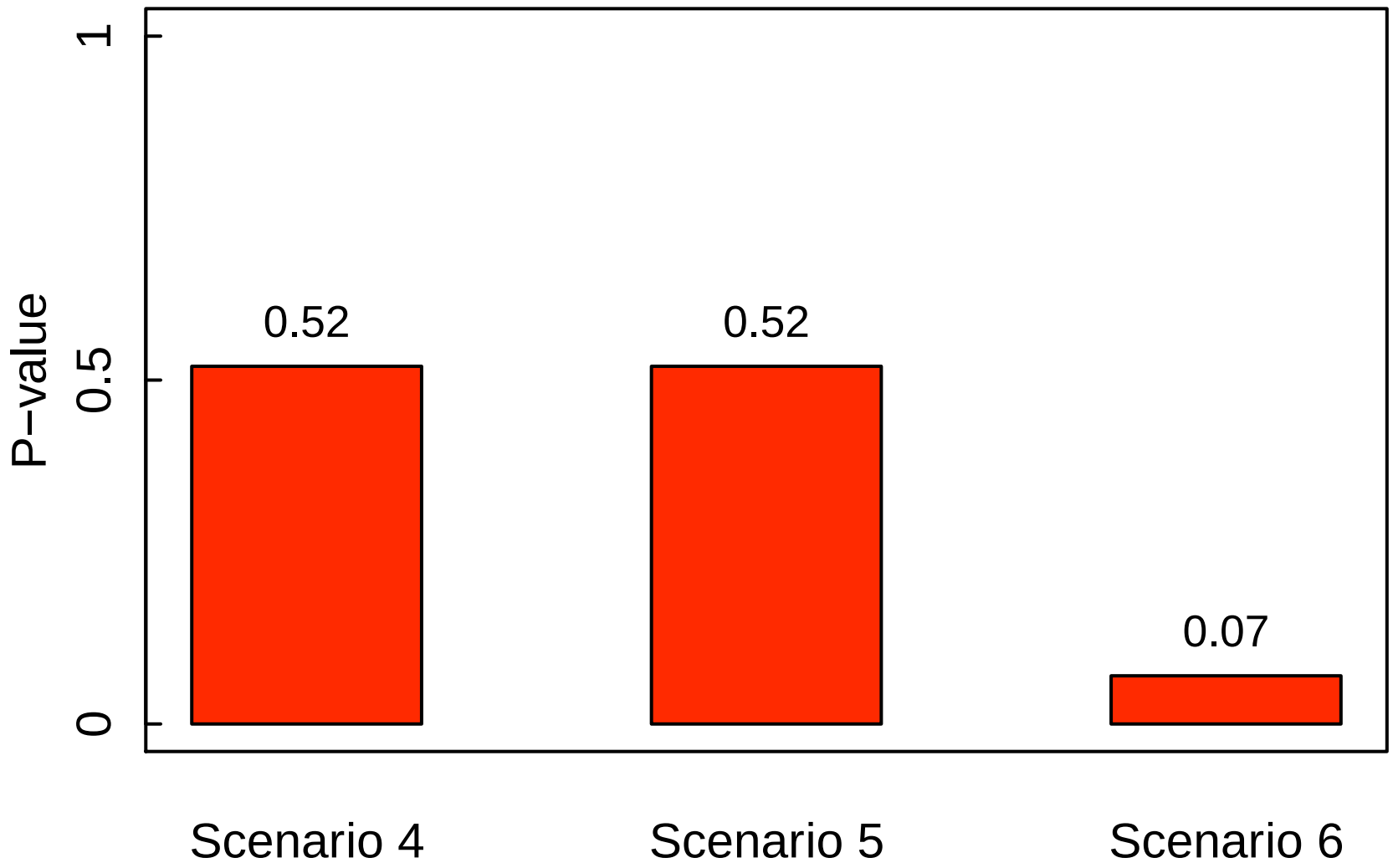
Subgroup search strategies

Scenario 4: SIDES method with complexity control (child-to-parent ratio $\gamma = 0.25$)
[3 subgroups]

Scenario 5: Fixed SIDEScreen method
[6 subgroups]

Scenario 6: Adaptive SIDEScreen method
[1 subgroup]

Adjusted p-values in Subgroup S_{11}



Based on 10,000 permutations

Raw treatment effect p-value: $p_{11} = 0.0077$

SIDEScreen method

Comparison of Scenarios 4 and 6

Standard SIDES method with strict complexity control: 68-fold increase in treatment effect p-value (from 0.0077 to 0.52)

Adaptive SIDEScreen method: 9-fold increase in treatment effect p-value (from 0.0077 to 0.07)

Importance of biomarker screening

Efficient biomarker screening considerably reduces multiplicity burden and leads to lower multiplicity penalty by filtering out noise biomarkers

Summary

Summary

Principled approaches to subgroup identification

Analytic subgroup search procedures for examining all relevant patient subgroups to find subsets of overall population with desirable characteristics

Multiplicity issues

Critical to address multiplicity issues to control the Type I error rate as part of developing reliable subgroup identification strategies

References

References

Bretz, F., Westfall, P.H. (2014). Multiplicity and replicability: Two sides of the same coin. *Pharmaceutical Statistics*. 13, 343-344.

Brookes, S. T., Whitley, E., Peters, T. J., Mulheran, P. A., Egger, M., Davey Smith, G. (2001). Subgroup analyses in randomised controlled trials: Quantifying the risks of false-positives and false-negatives. *Health Technology Assessment*. 5, 1-56.

Dmitrienko, A., Lipkovich, I., Hopkins, A., Li, Y.P., Wang, W. (2015). Biomarker evaluation and subgroup identification in a pneumonia development program using SIDES. *Applied Statistics in Biomedicine and Clinical Trials Design*. Chen, Z., Liu, A., Qu, Y., Tang, L., Ting, N., Tsong, Y. (editors). Springer, New York.

References

Foster, J.C., Taylor, J.M.C., Ruberg, S.J. (2011). Subgroup identification from randomized clinical trial data. *Statistics in Medicine*. 30, 2867-2880.

Hothorn, T., Hornik, K., Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*. 15, 651-674.

Kehl, V., Ulm, K. (2006). Responder identification in clinical trials with censored data. *Computational Statistics and Data Analysis*. 50, 1338-1355.

Lipkovich, I., Dmitrienko, A., Denne, J., Enas, G. (2011). Subgroup identification based on differential effect search (SIDES): A recursive partitioning method for establishing response to treatment in patient subpopulations. *Statistics in Medicine*. 30, 2601-2621.

References

Lipkovich, I., Dmitrienko, A. (2014a). Strategies for identifying predictive biomarkers and subgroups with enhanced treatment effect in clinical trials using SIDES. *Journal of Biopharmaceutical Statistics*. 24, 130-153.

Lipkovich, I., Dmitrienko, A. (2014b). Biomarker identification in clinical trials. *Clinical and Statistical Considerations in Personalized Medicine*. Carini, C., Menon, S., Chang, M. (editors). Chapman and Hall/CRC Press, New York.

Lipkovich, I., Dmitrienko, A., D'Agostino, R.B. (2017). Tutorial in Biostatistics: Data-driven subgroup identification and analysis in clinical trials. *Statistics in Medicine*. 36, 136-196.

References

Loh, W.Y., Shih, Y.S. (1997). Split selection methods for classification trees. *Statistica Sinica*. 7, 815-840.

Loh, W.Y. (2002). Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica*. 12, 361-386.

Loh, W.Y., He, X., Man, M. (2015). A regression tree approach to identifying subgroups with differential treatment effects. *Statistics in Medicine* 34, 1818-1833.

Rothwell, P. M. (2005). Subgroup analysis in randomized controlled trials: Importance, indications, and interpretation. *Lancet*. 365, 176-186.

Su, X., Tsai, C.L., Wang, H., Nickerson, D.M., Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*. 10, 141-158.

References

Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology*. 66, 818-825.

Xu, Y., Trippa, L., Müller, P., Ji, Y. (2016). Subgroup-based adaptive (SUBA) designs for multi-arm biomarker trials. *Statistics in Biosciences*. 8, 159-180.

Zeileis, A., Hothorn, T., Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*.17, 492-514.