Leveraging multiple endpoints in small clinical trials

Robin Ristl

Section for Medical Statistics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna

EMA FP7 small-populations workshop, March 2017

Reasons to use multiple endpoints in clinical trials in rare diseases

Setting 1: Co-primary endpoints

- Some diseases are multi-faceted and we need to adress multiple endpoints for full characterization
- Main goal: Show efficacy in all co-primary endpoints
- Challenge: With small sample size, the probability to miss the main goal can be large and we need some fallback strategy

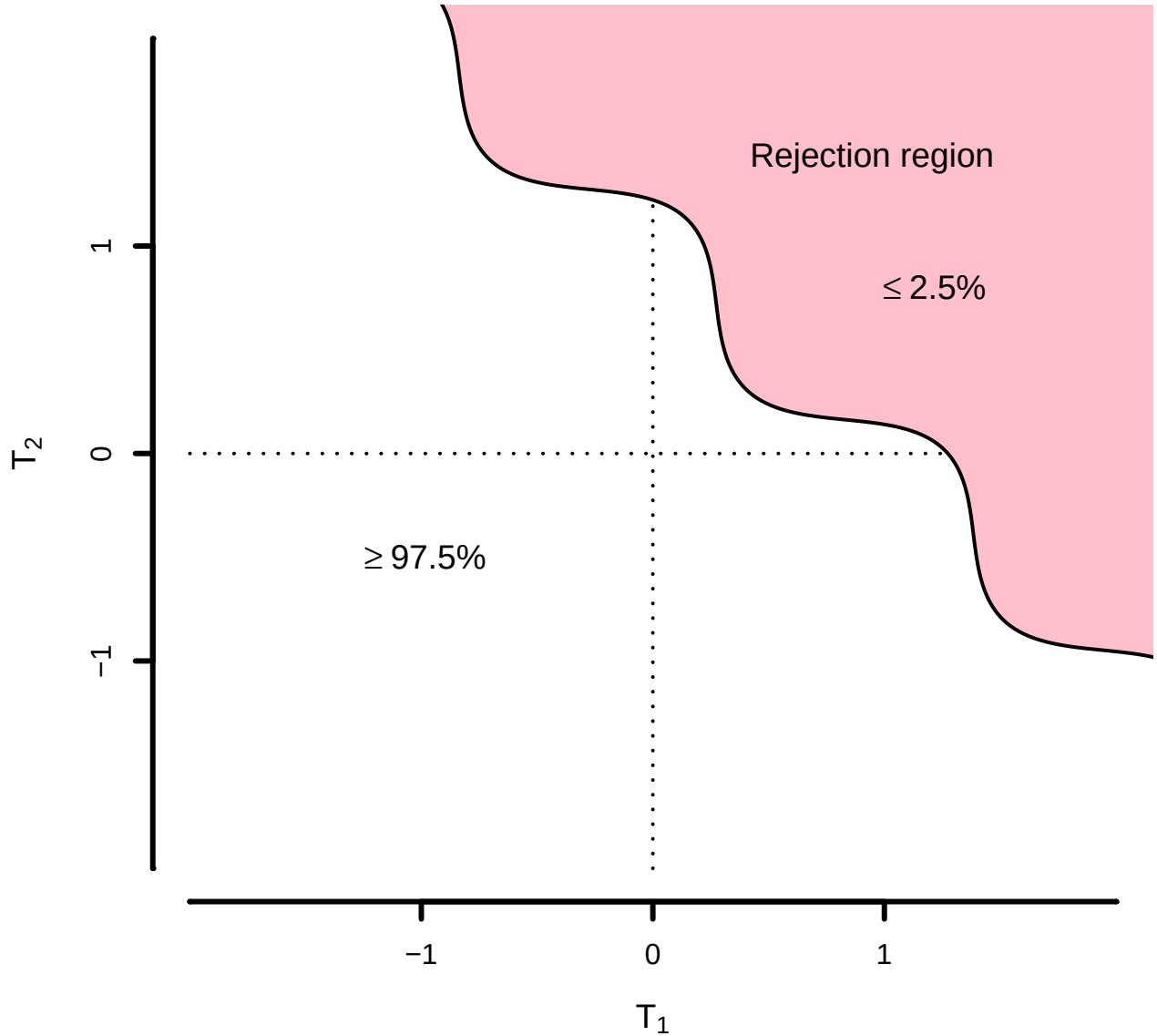
Setting 2: Global tests

- Aim for conclusion of some overall treatment effect
- Use information from multiple endpoints to counteract low information from small sample
- Increased power compared to single endpoint test
- Challenge: Optimize global test for defined alternative

Hypothesis tests with multiple endpoints

- E.g. for two endpoints, we consider the following null hypothesis
  - No effect in any endpoint (global null hypothesis)
  - No effect in endpoint 1 (H_1)
  - No effect in endpoint 2 (H_2)
- The probability for any false positive decision must be $\leq \alpha$
- The probabiltiy to identify a true effect (power) should be large
- To find optimized test procedures, we consider general multidimensional rejection regions.

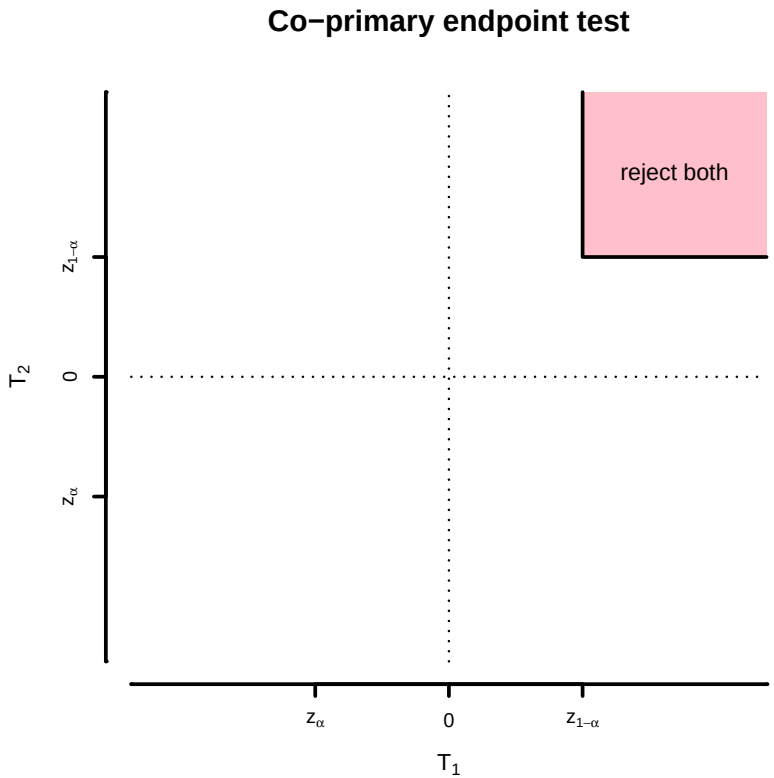
How to find a multidimensional rejection region?



- T_1 and T_2 are the test statistics for endpoints 1 and 2

Setting 1: Co-primary endpoints

- Two endpoints are co-primary, if the main aim is to show an effect for both endpoints.
- Standard co-primary endpoint test:
  - Perform a separate test for each endpoint
  - If both tests are significant at level α, conclude effect in both endpoints
  - Else, no conclusion

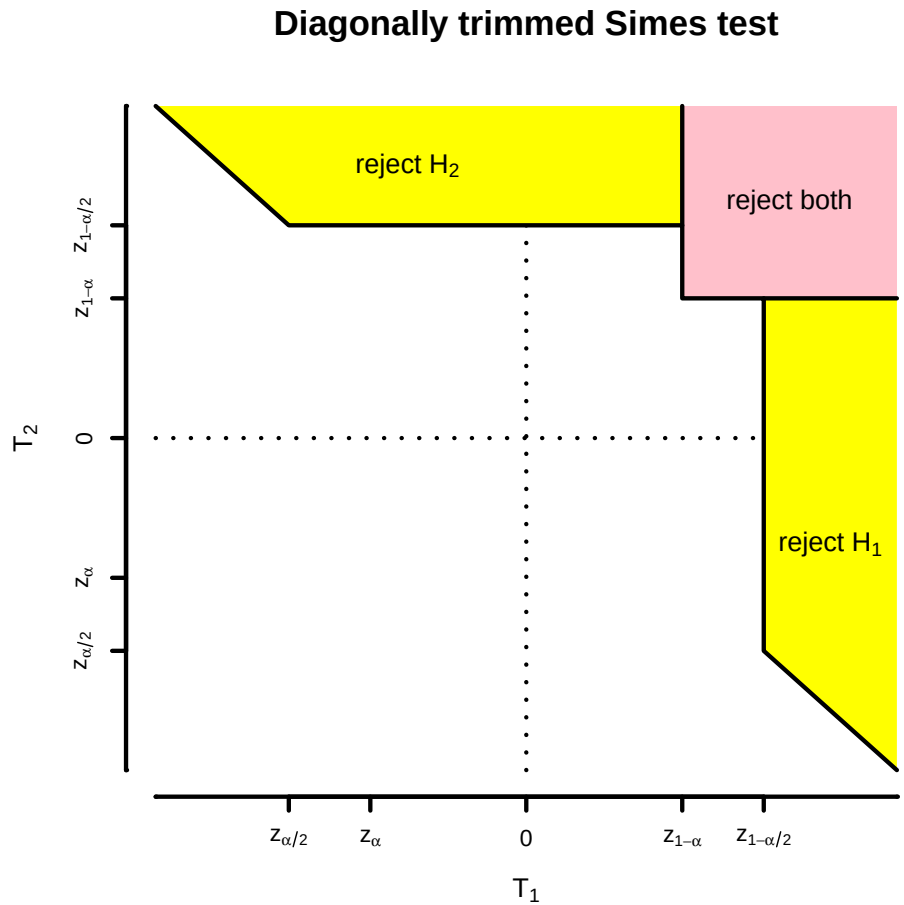


- Most powerful for main aim, but reduces to all or nothing decision.
  - E.g.: Test for H_1 may be highly significant while the test for H_2 is not significant.
  - Then the standard test does not allow for rejection of any null hypothesis.

New method: Fallback tests for co-primary endpoints

- What can we learn if a co-primary endpoint trial does not achieve the main goal to reject all null hypothesis?
- We extend the standard co-primary endpoint test with a fallback option:
- Even if the main goal is not reached, there is the option to reject some null hypothesis.
- Important in small sample situtation, as the probability to miss the main aim may be high

A fallback test for two co-primary endpoints



- Retains power of standard test
- Adds decision rules to claim partial success
- Strict type I error rate control for arbitrarily correlated test statistics (if bivariate normal or t-distributed).

Further results

- A Fallback tests for three endpoints
- Adjusted p-values for the fallback tests
- General approach to combine simple fallback test to a more complex testing procedure

R. Ristl, F. Frommlet, A. Koch, M. Posch, "Fallback tests for co-primary endpoints", Statistics in Medicine, 2016, 35:2669-2686

Setting 2: Global tests for multiple endpoints

- The power to reject a global null hypothesis can be large compared to the power to show some endpoint-specific effect
- Conclusion on endpoint-specific effects requires extension to multiple testing procedure
- Challenge: Find powerful rejection region for global test, control type I error rate
- Proposed solution: Exact tests through multivariate permutation, optimization algorithms to find rejection regions
- We studied in particular optimal exact tests for multiple binary endpoint.

Example: Non-24-hour sleep-wake disorder

- Assume a study similar to Lockley et al., Lancet 2015, 386:1754-64
- EP 1: Entrainment (synchronization of the master body clock to the 24-hour day)
- EP 2: Clinical response

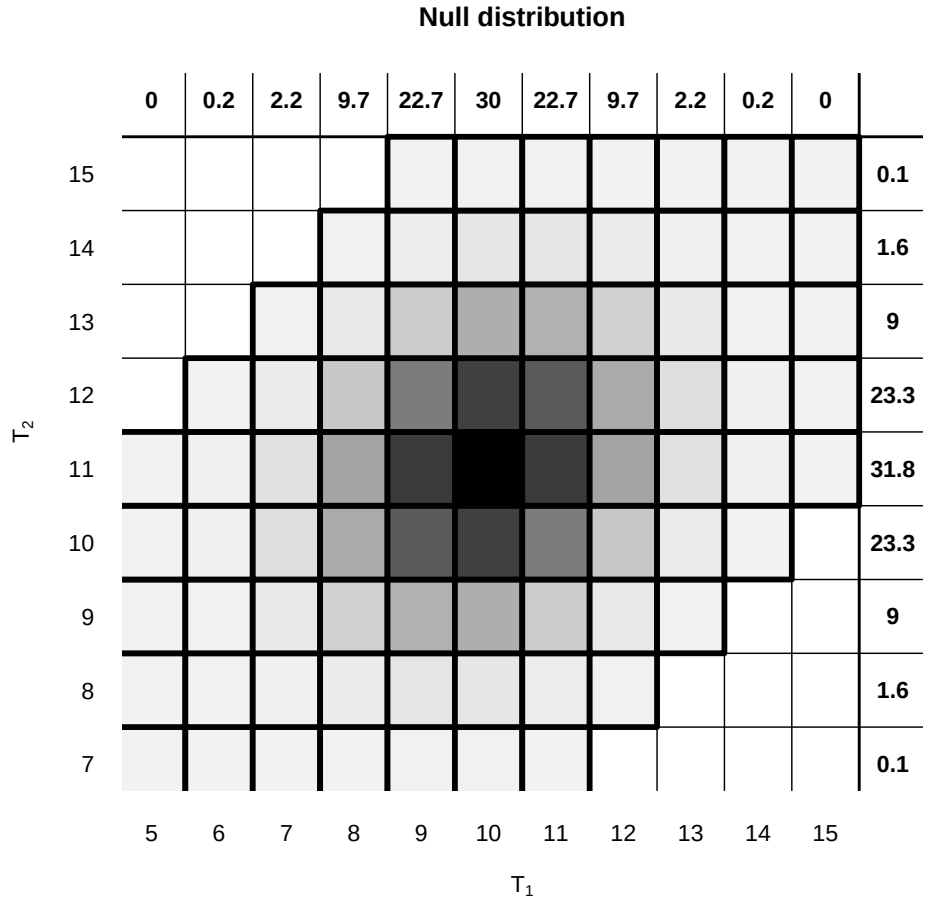
Assumed success rates for planning

	Treatment	Control
Endpoint 1&2	0.35	0.03
Endpoint 1	0.1	0.03
Endpoint 2	0.1	0.03
None	0.45	0.91

Observed blinded frequencies in the example

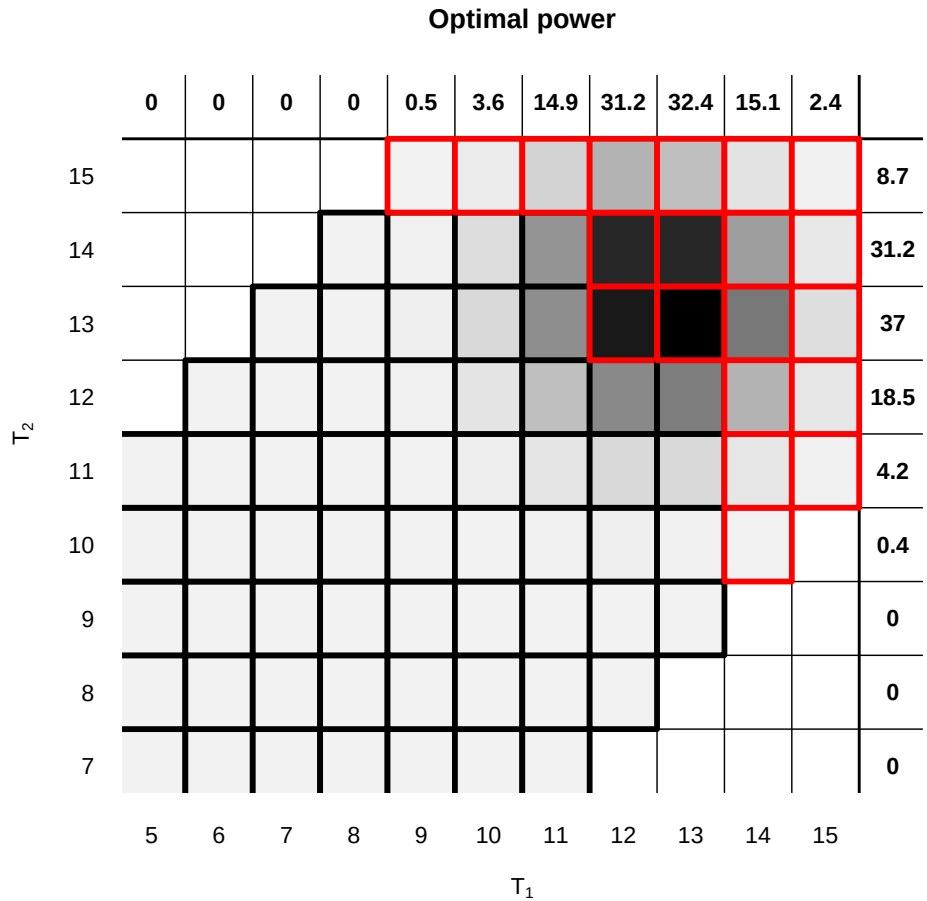
	Treatment	Control	Total
Endpoint 1&2	blinded	blinded	16
Endpoint 1	blinded	blinded	4
Endpoint 2	blinded	blinded	6
None	blinded	blinded	4
Total	15	15	30

Discrete null distribution found through permutation



- Test statistics T_1, T_2 are the number of successes for endpoints 1 and 2 in the treatment group.
- Dark fields correspond to larger probability under the global null hypothesis.

Rejection region with maximal power under the assumed alternative



- Power to reject global null hypothesis in the example
  - Optimal joint permutation test: 81%
  - Fisher exact tests with Bonferroni correction: 61%
  - Single endpoint Fisher exact test: 59%

Further results

- Optimally weighted Bonferroni tests
- Construction of multiple testing procedures
- Adjusted p-values
- Fast greedy algorithm for approximate solution

R. Ristl, X. Dong, E. Glimm, M. Posch, "Optimal exact tests for binary endpoints", arXiv:1612.07561

Points for discussion

1. In co-primary endpoint trials, what is the impact of a partial claim of success on regulatory decision making?

2. In which situation is it sufficient to show a global treatment effect on multiple endpoints?

References

R. Ristl, F. Frommlet, A. Koch, M. Posch, "Fallback tests for co-primary endpoints", *Statistics in Medicine*, 2016, 35:2669-2686

R. Ristl, X. Dong, E. Glimm, M. Posch, "Optimal exact tests for binary endpoints", arXiv:1612.07561

R. Ristl, S. Urach, G. Rosenkranz, M. Posch, "Methods for the analysis of multiple endpoints in small populations: A review", *submitted to Statistical Methods in Medical Research*