Approaches to address multiplicity in pragmatic RCTs: simulation studies to aid with the design and interpretation of RCTs with potential multiplicity concerns

Katie Pike, Barnaby C Reeves, Chris A Rogers·
Bristol Trials Centre, University of Bristol, Bristol, UK

Introduction and aims

The work is part of a project to determine optimal approaches to address multiplicity in pragmatic RCTs, and whether approaches should vary according to research questions and study design. The work reported focuses on simulation studies, which have the following aims:

1) Describe how trial design factors affect Type I/II errors when testing hypotheses

2) Recommend when statistical methods to adjust for multiplicity are advised, and the preferred method(s)

Methods

RCTs were simulated from randomly distributed data with either:
a) Multiple primary outcomes, or
b) Comparing multiple treatments with a control

↓

Trial design factors were varied, including:
- Differences between treatments: zero or 90% underlying power
- Relationship/correlation between outcomes/comparisons
- Number of outcomes/treatment comparisons
- Alpha
- Sample size
- Continuous/binary outcomes
- Variability

↓

Regression models were fitted for each simulation

→

Multiple comparisons within each trial were combined in two ways:
- Disjunctive: **at least one** of the hypotheses must be rejected for the treatment to be declared effective
- Conjunctive: **all** of the hypotheses must be rejected for the treatment to be declared effective

↓

Across all simulations the proportion of trials with the treatment declared effective was calculated:
- Underlying treatment difference zero=estimated overall Type I error
- Underlying treatment difference 90% nominal power=estimated overall power

↓

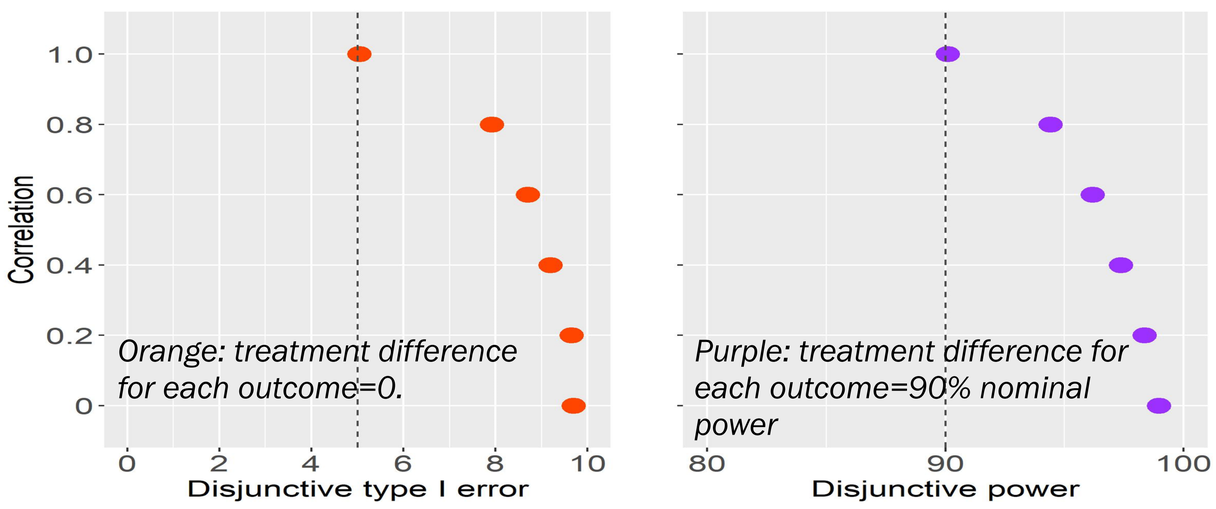
Various multiplicity adjustment methods were applied

Results

The key trial design factors that impact on Type I/II error rates are: 1) how the multiple comparisons are combined, and 2) the correlation/relationship between outcomes or comparisons.

The results presented below are for an RCT with two primary outcomes.

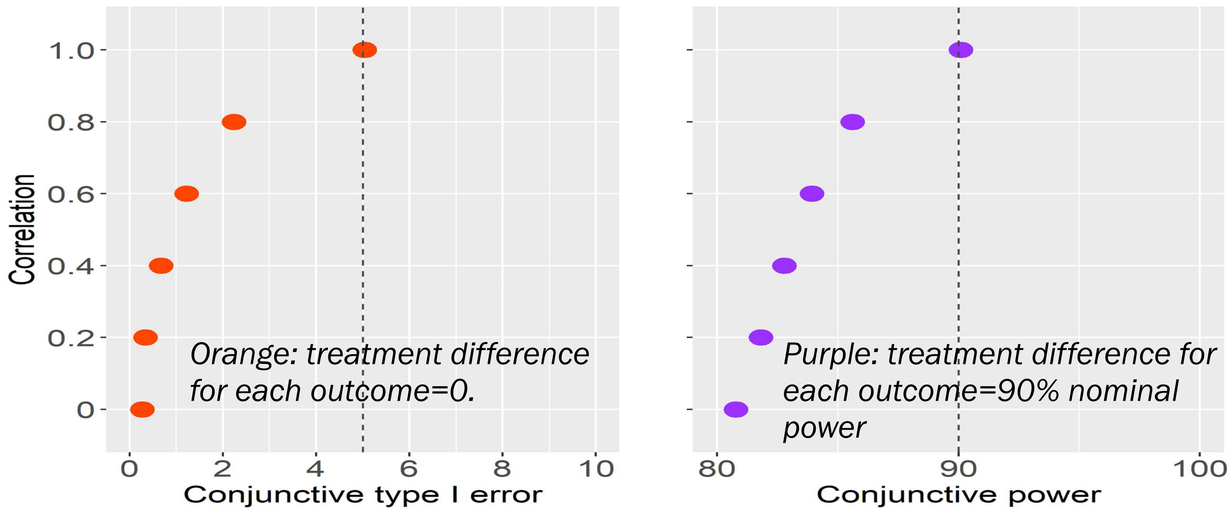
Proportion of trials with hypotheses for at least one outcome rejected:



Type I errors and power are increased above their nominal values, and this increase is greater for weakly correlated outcomes.

Multiplicity adjustment methods should be considered:
- Bonferroni method is conservative
- Hierarchical methods are unsuitable: there would generally will be no hierarchy to the outcomes
- **Hochberg method is recommended**: less conservative than Bonferroni in terms of both Type I error and power

Proportion of trials with hypotheses for both outcomes rejected:



Type I errors and power are decreased above their nominal values, and this decrease is greater for weakly correlated outcomes.

Adjustment for multiplicity is not needed.

However, this reduction in trial (conjunctive) power should be considered when designing trials; an **increase in sample size will be needed**

Discussion

We are currently developing:

Succinct guidance on multiplicity in RCTs, covering: trial design factors to consider and recommended approaches in certain situations

R code to directly calculate overall trial Type I error/power rates, according to trial design factors (in particular correlation between outcomes)

This should help statisticians and methodologists designing future pragmatic RCTs, and guide professionals interpreting RCT findings.

University of BRISTOL | BTC

This project is funded by the NIHR (NIHR Doctoral Research Fellowship, Miss Katie Pike, DRF-2017-10-126). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

SUPPORTED BY NIHR | National Institute for Health and Care Research