Novel multiple testing procedures for structured study objectives and families of hypotheses – a case study

Guenther Mueller-Velten

Novartis Pharma AG

EMA Workshop on Multiplicity Issues in Clinical Trials
London, 16 November 2012

Acknowledgement: Frank Bretz, Bjoern Holzhauer, Willi Maurer

Outline

- Introduction of graphical approaches to multiple testing
- Case study
  - Background and clinical considerations
  - Resulting multiple testing procedure
  - Interim analysis
- Summary and conclusions

Graphical approaches to multiple testing

Motivation

Increasing complexity of confirmatory trial designs

- Multiple treatment arms, multiple primary and secondary endpoints, interim analyses
- Designing a valid multiple testing strategy with desired properties is a cross-functional exercise and may involve several iterations
- Clinical, regulatory and business requirements need to be translated into a statistical testing procedure in a transparent and understandable way

Graphical approaches to multiple testing

Motivation (cont'd)

Graphical approaches provide a framework to

- Tailor advanced multiple test procedures to structured families of hypotheses
- Visualize complex decision strategies in an efficient and easily communicable way, and
- Ensure strong Type I error rate control (Bretz et al., 2009; Burman et al., 2009)

Notation

- Null hypotheses $H_1, \ldots, H_k$
- Initial allocation of the significance level $\alpha = \alpha_1 + \ldots + \alpha_k$.
- Unadjusted p-values $p_1, \ldots, p_k$

Notation

- Null hypotheses $H_1, \ldots, H_k$
- Initial allocation of the significance level $\alpha = \alpha_1 + \ldots + \alpha_k$.
- Unadjusted p-values $p_1, \ldots, p_k$

"α propagation"

If a hypothesis H_i can be rejected at level α_i (i.e. $p_i \leq \alpha_i$), reallocate its level α_i to the remaining, not yet rejected hypotheses (according to a prefixed rule) and continue testing with the resulting α levels.

Graphical approaches to multiple testing

Conventions

Graphical approaches to multiple testing

Conventions

1. Hypotheses $H_1, \ldots, H_k$ represented as nodes

Graphical approaches to multiple testing

Conventions

1. Hypotheses $H_1, \ldots, H_k$ represented as nodes

2. Split of significance level α as weights $\alpha_1, \ldots, \alpha_k$

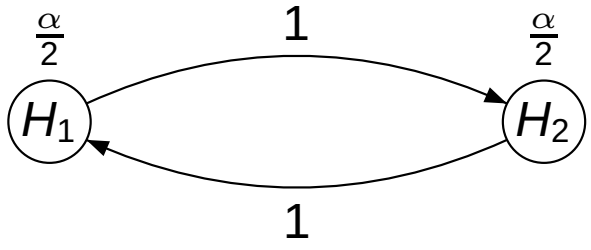
Graphical approaches to multiple testing

Conventions

1. Hypotheses $H_1, \ldots, H_k$ represented as nodes

2. Split of significance level α as weights $\alpha_1, \ldots, \alpha_k$

3. "α propagation" through weighted, directed edges



Case study

Introduction and background

- Randomized double-blind event driven outcome trial in stable post myocardial infarction (MI) patients
- Three doses of a new therapy vs. placebo on top of standard of care
  - No validated surrogate available for dose-finding prior to Phase III
- Primary endpoint is composite of CV death, MI or stroke
- Two key secondary endpoints targeting additional label claims, thus included in multiple testing procedure
  - Extended composite endpoint including hospitalization for unstable angina requiring urgent unplanned revascularizations
  - New onset Type 2 diabetes among patients with pre-diabetes at baseline
- Additional multiplicity due to efficacy interim analyses

Case study

Clinical considerations

- Primary endpoint is essential to establish efficacy of the respective dose group. Key secondary objectives target additional label claims for doses that have established efficacy based on the primary endpoint.
  - Successiveness: Do not reject a secondary hypothesis without having rejected the associated primary hypothesis. (Maurer et al., 2011)
- Benefit risk considerations: Higher doses potentially more efficacious and lower doses generally safer
  - Allow testing of lower doses regardless of efficacy at higher doses.
  - Unequal split of significance level.

Case study

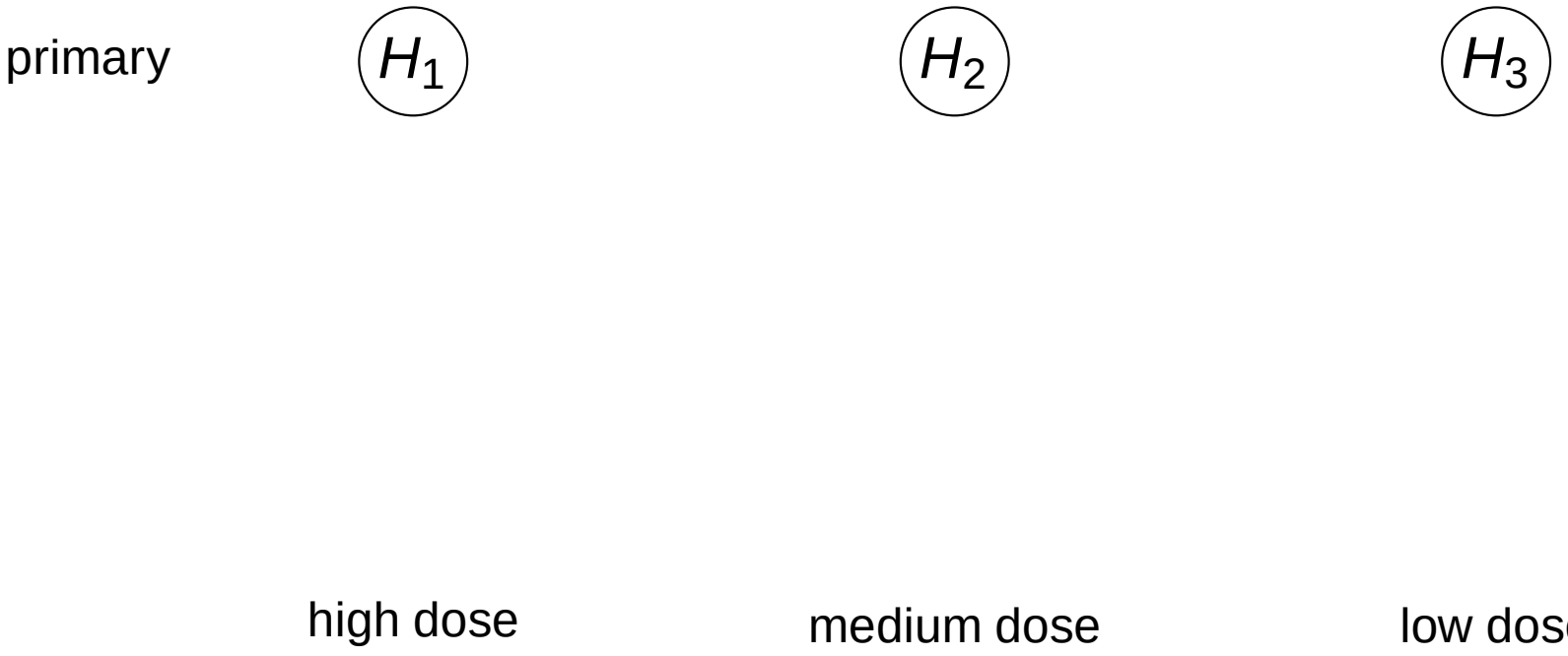
Structured family of hypotheses

- Four-armed trial comparing
  - Three dose levels $+$ standard of care
  - Placebo $+$ standard-of-care
- Three endpoints
  - Primary endpoint: composite of CV death, MI or stroke
  - Two key secondary endpoints

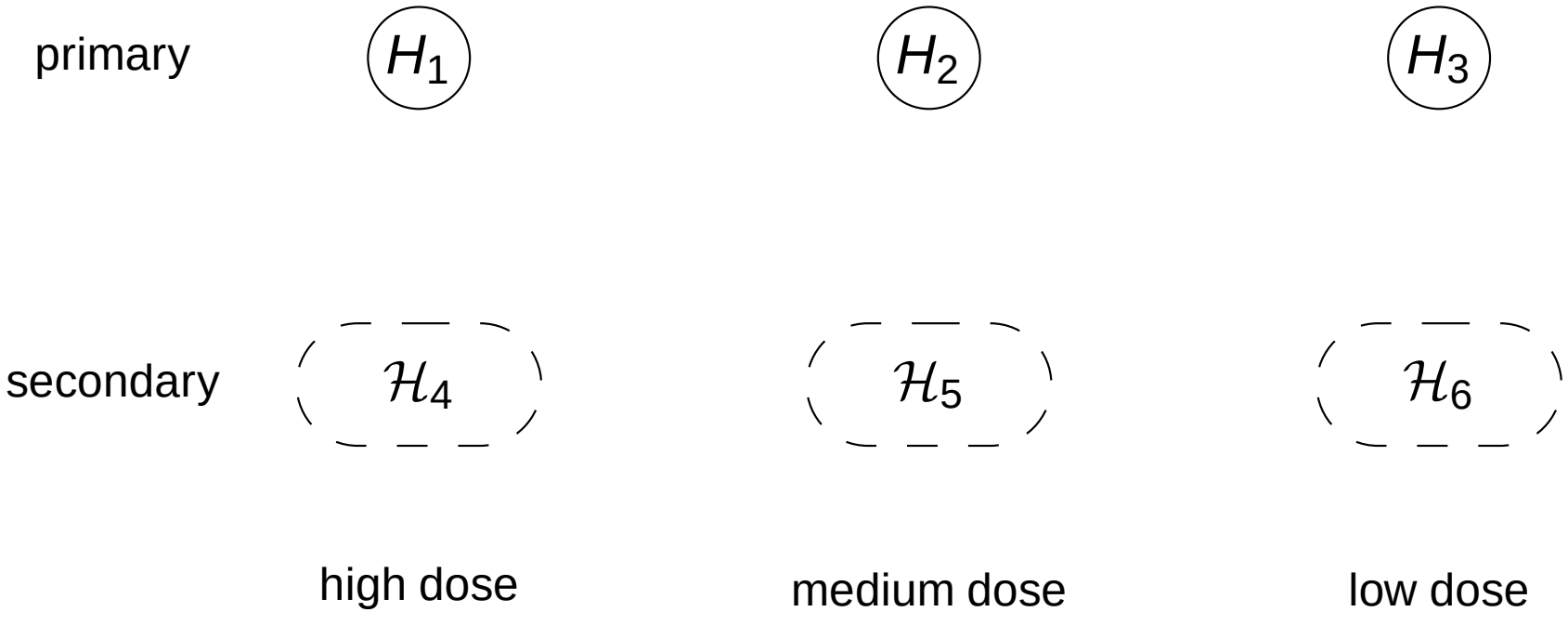
$\Rightarrow$ Nine hypotheses

- Three doses (low, medium, high)
- Three endpoints per dose

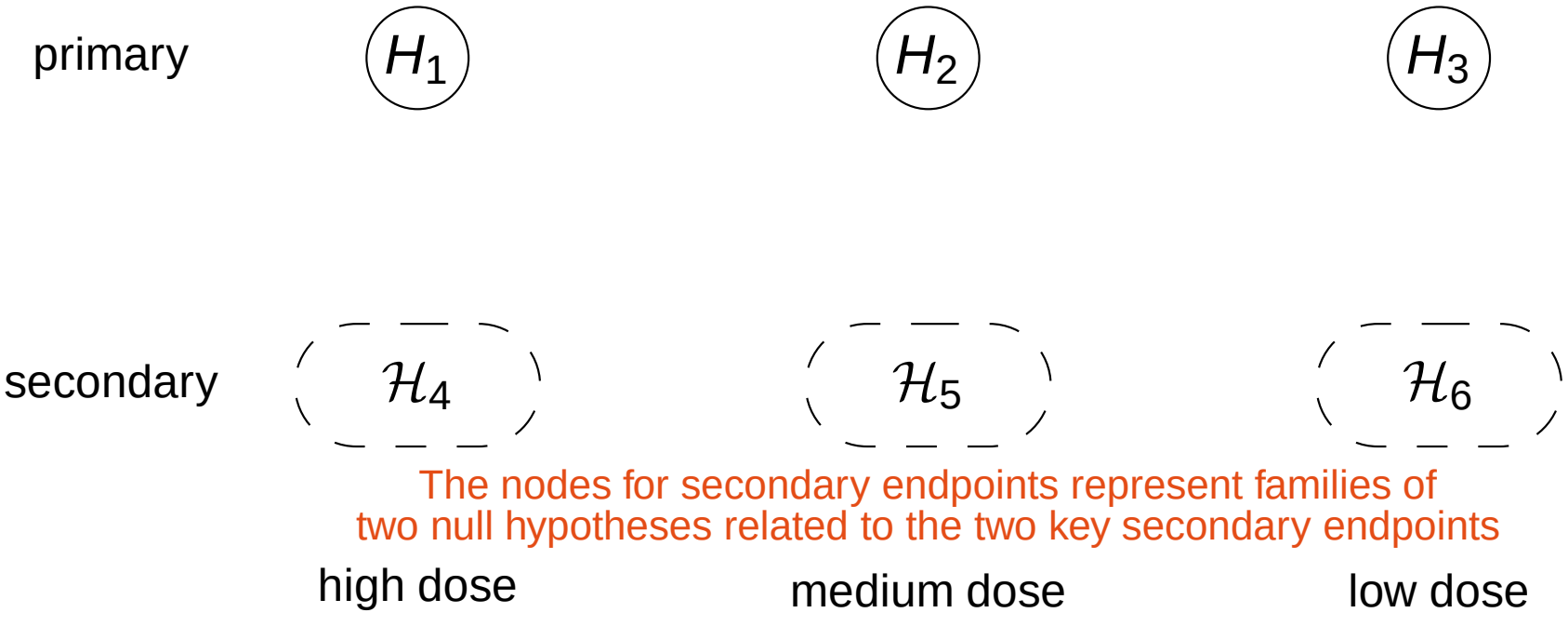
Tailored multiple test procedure



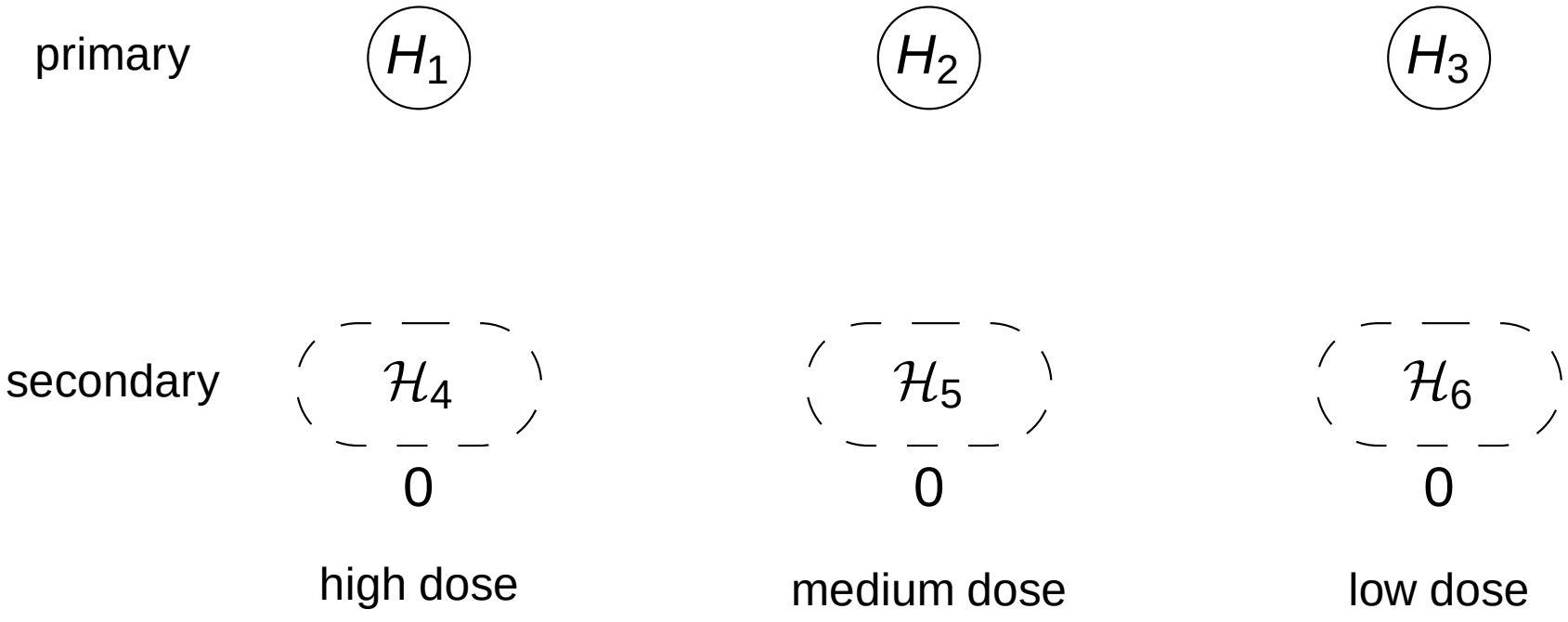
Tailored multiple test procedure



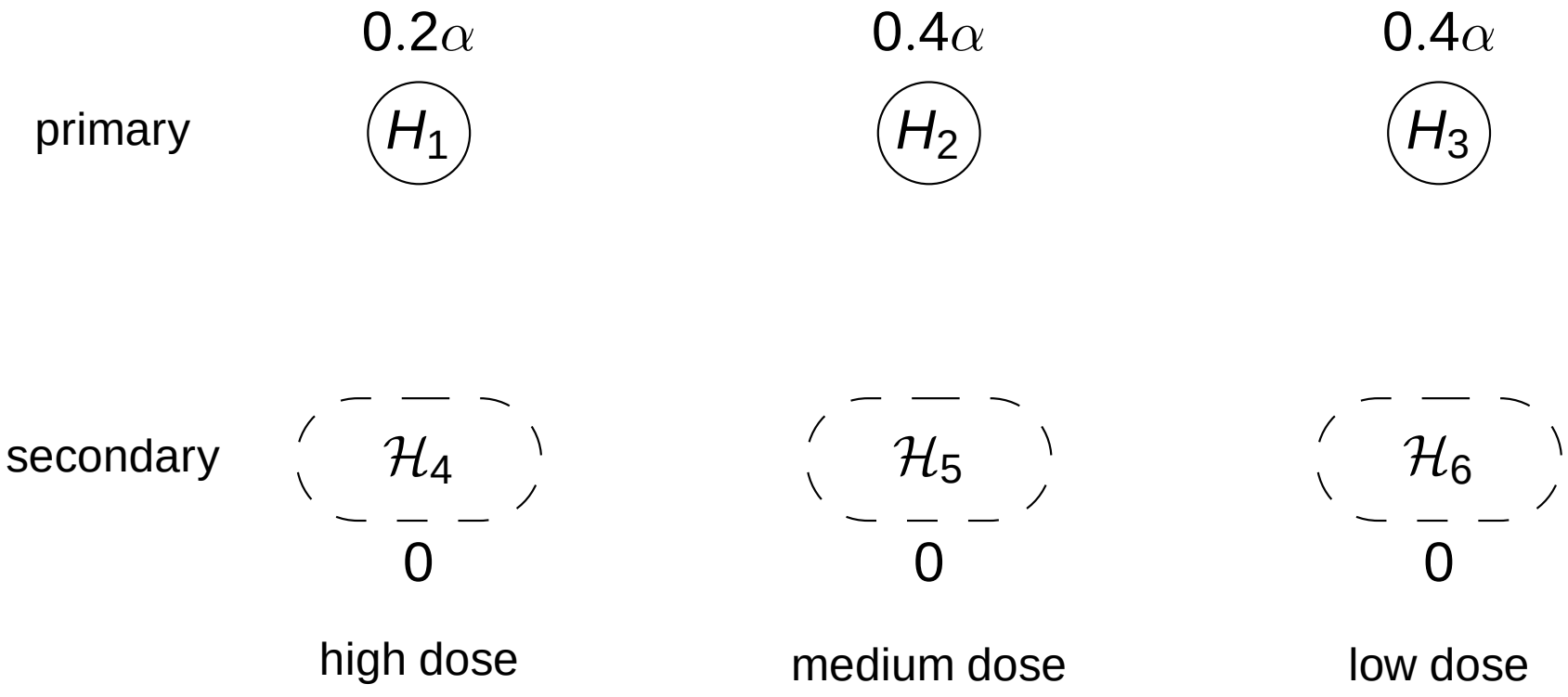
Tailored multiple test procedure



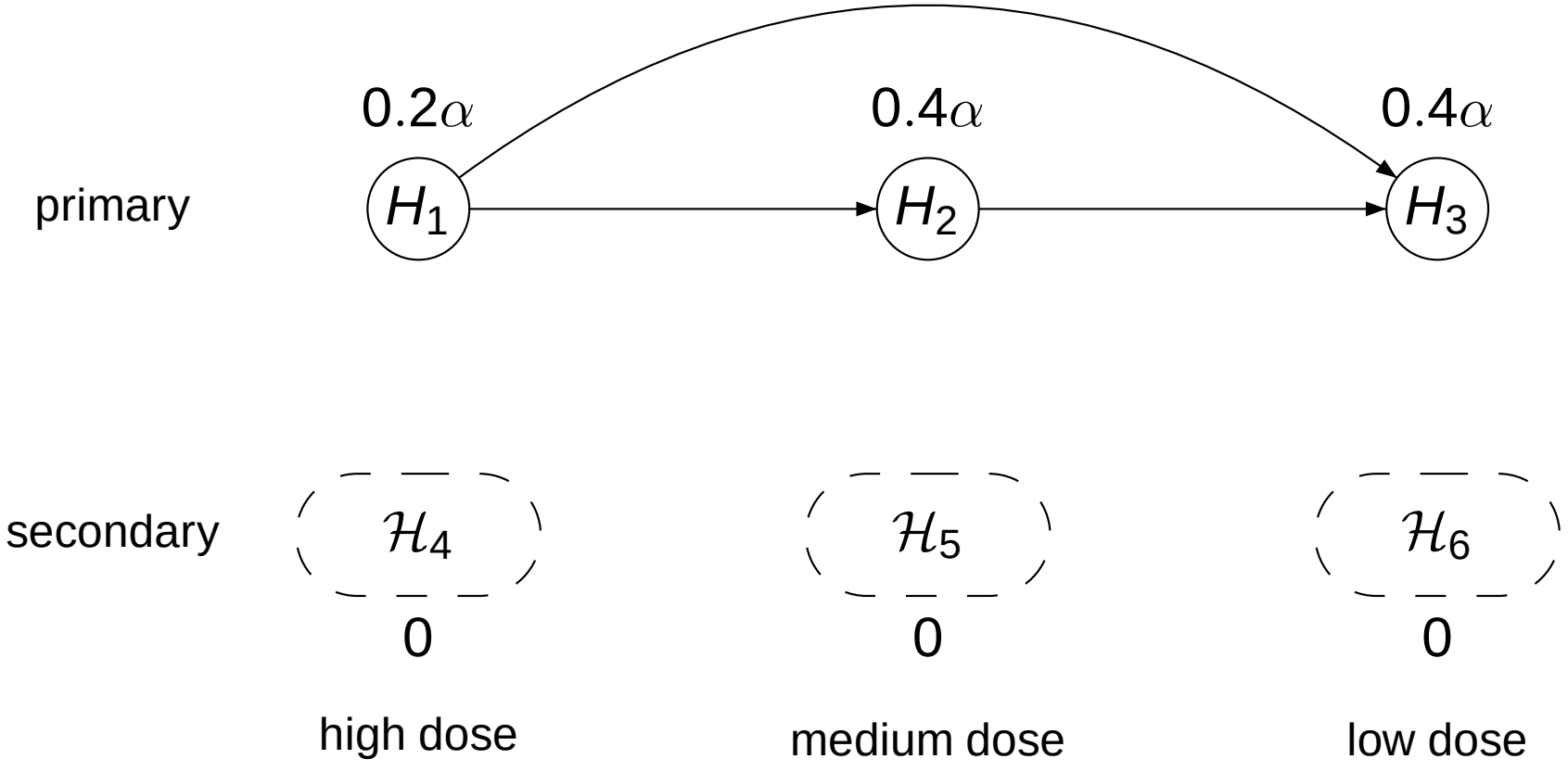
Tailored multiple test procedure



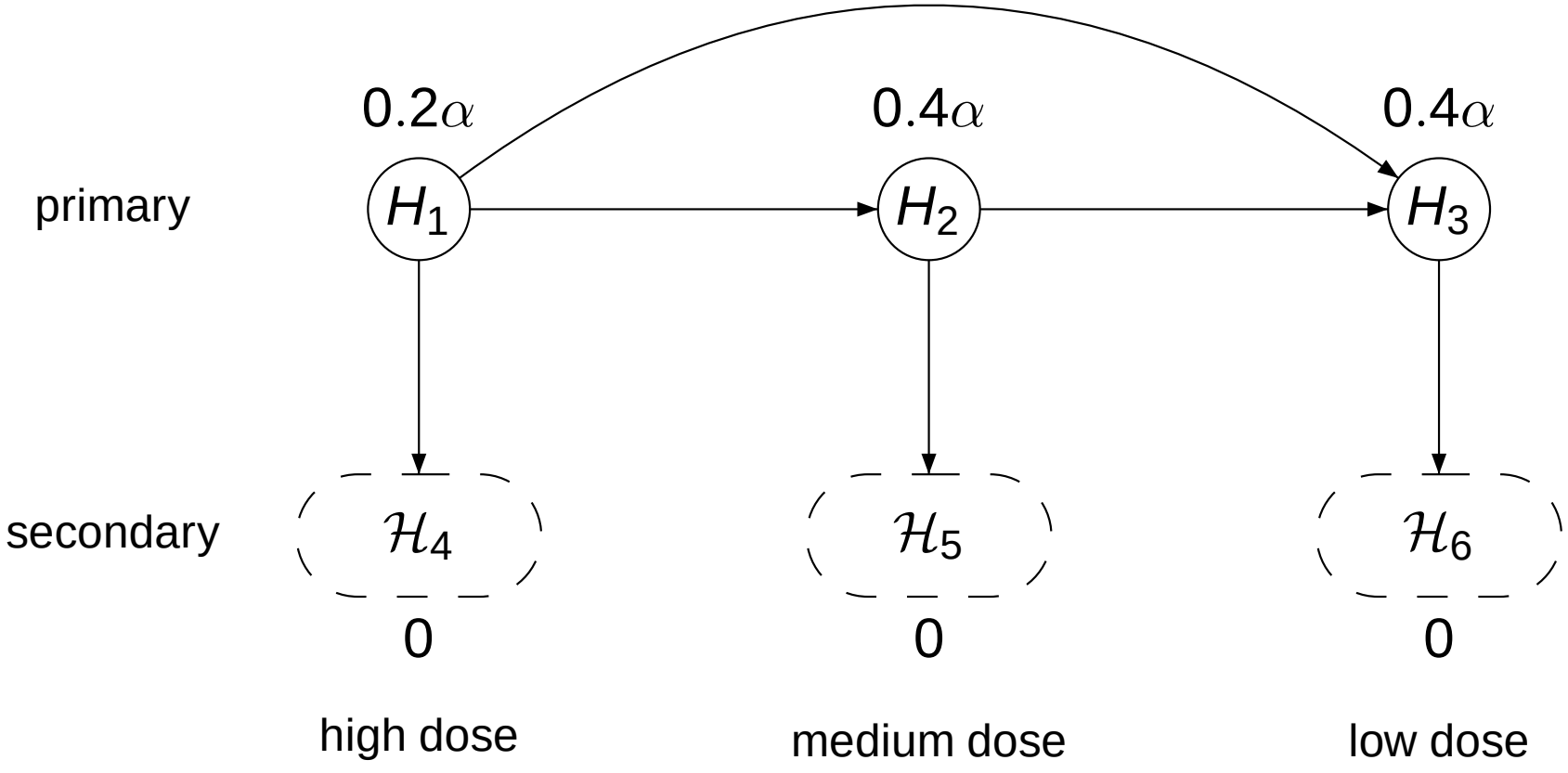
Tailored multiple test procedure



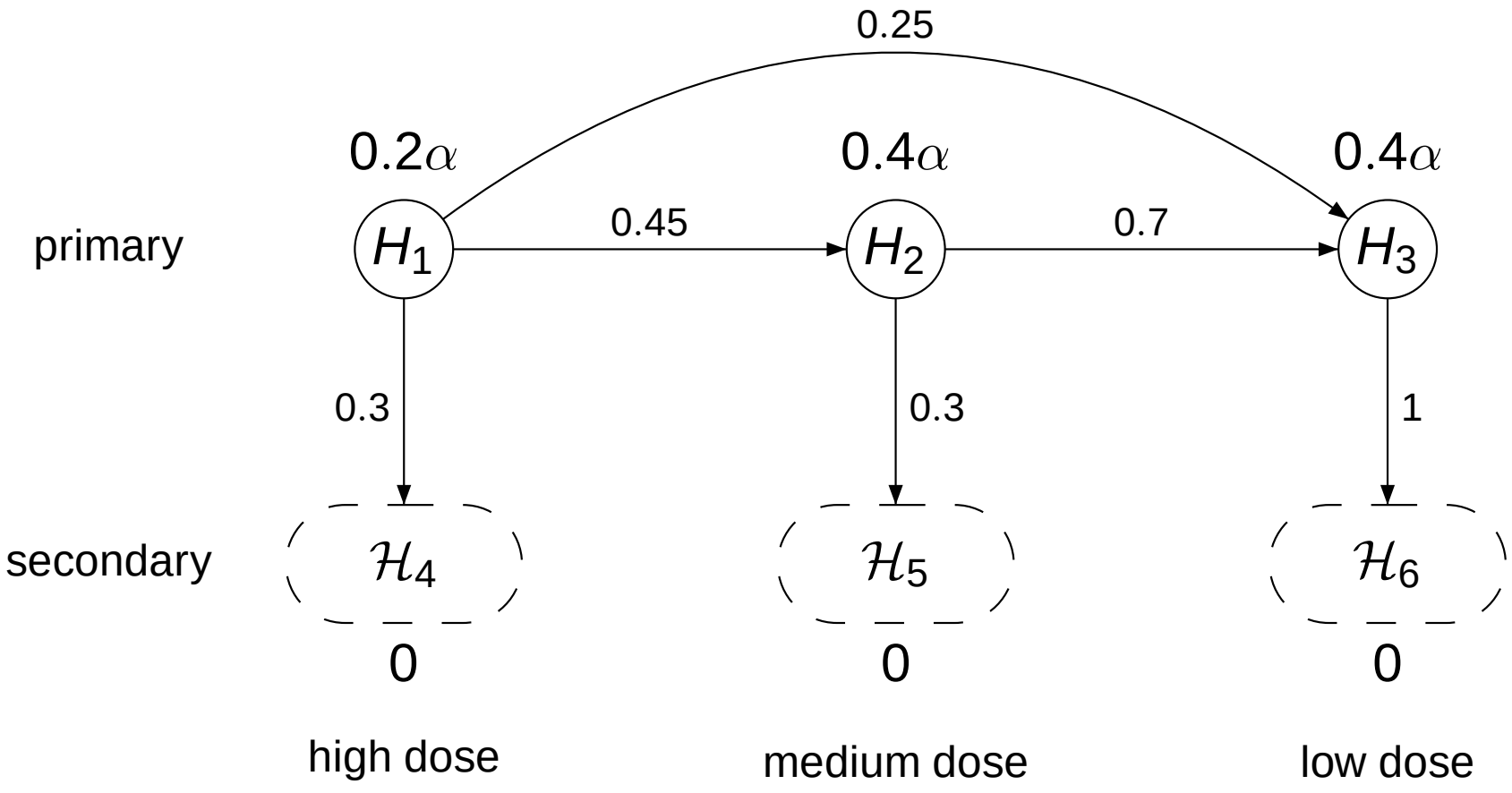
Tailored multiple test procedure



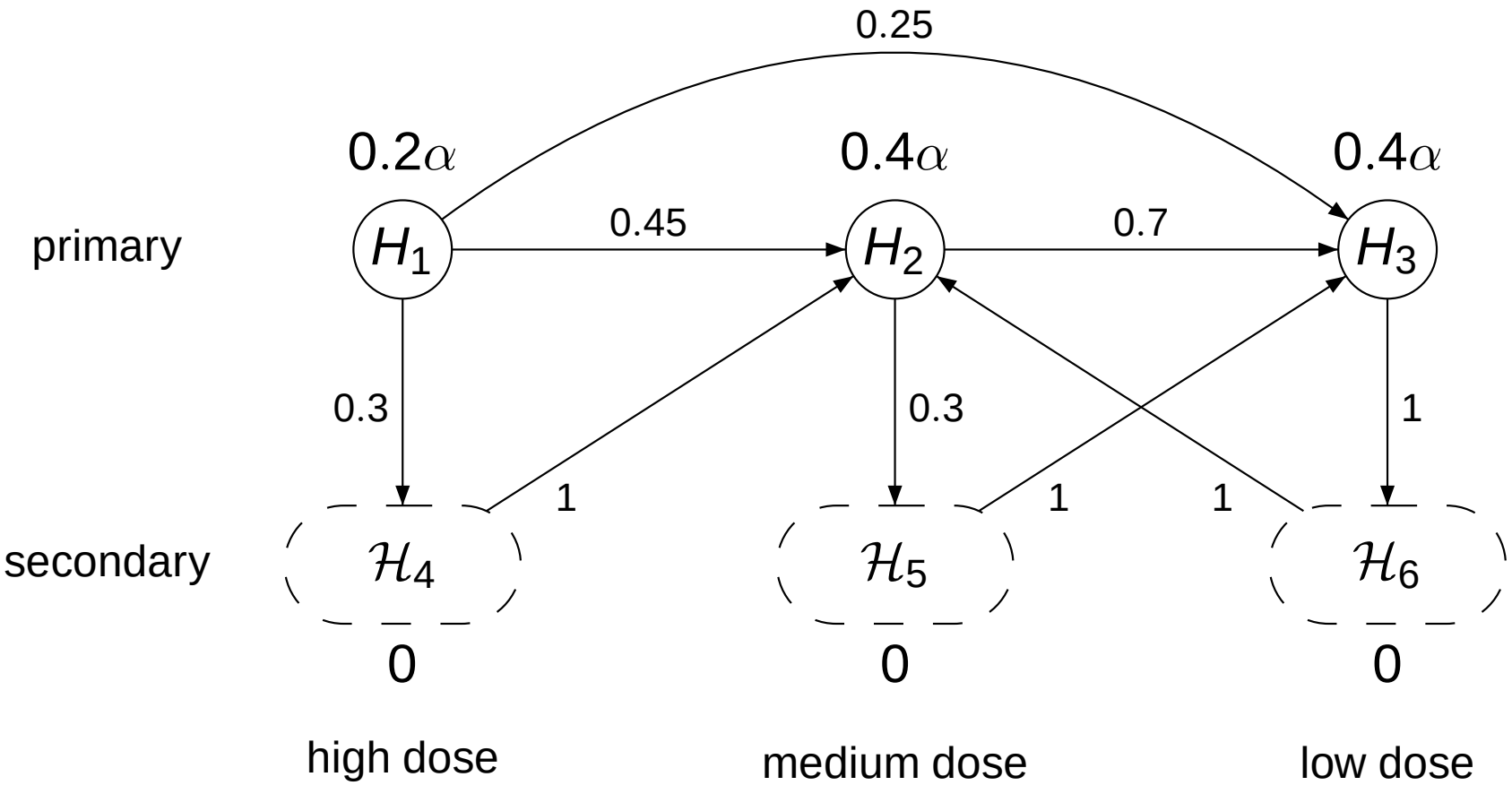
Tailored multiple test procedure



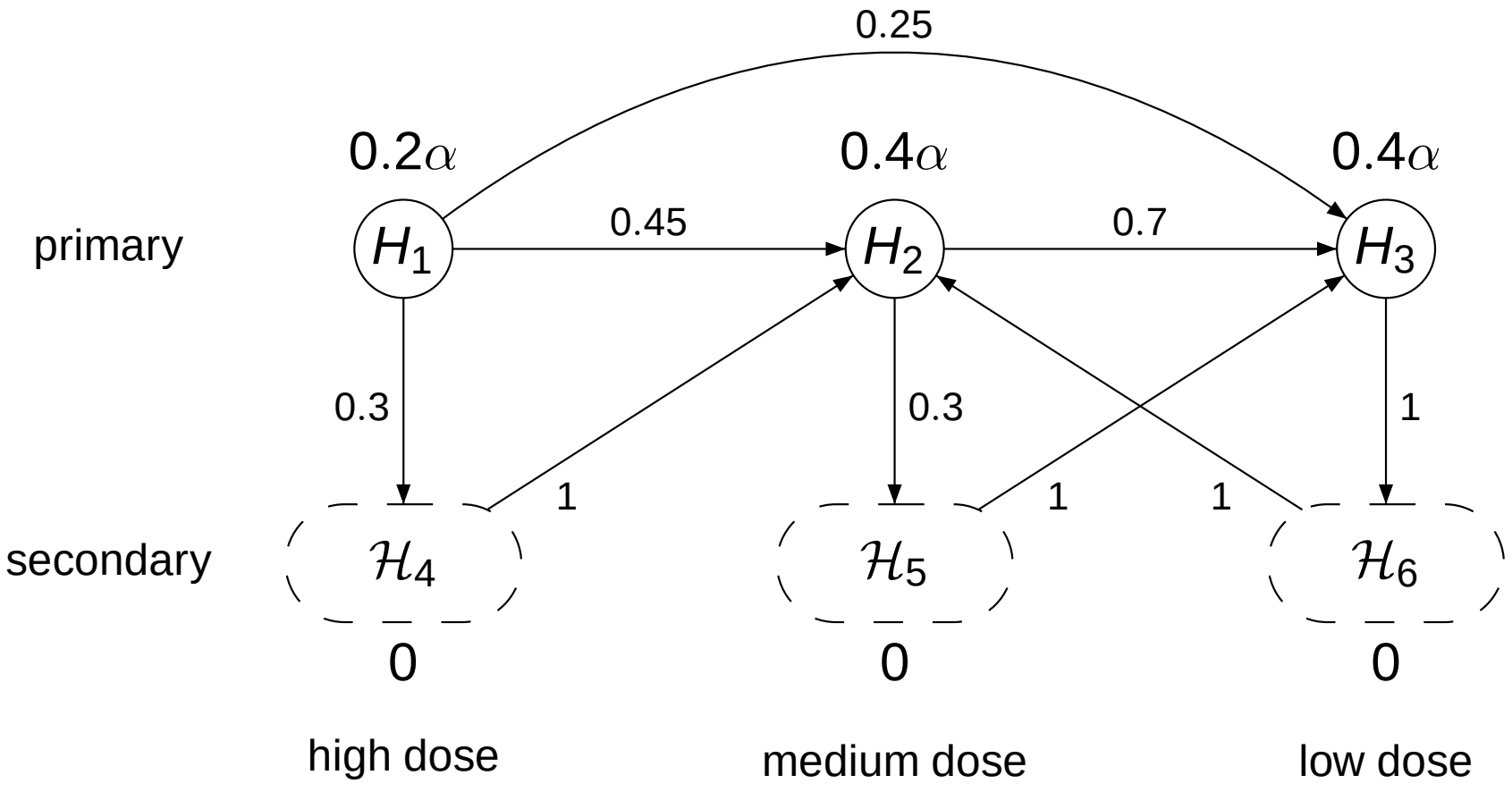
Tailored multiple test procedure



Tailored multiple test procedure

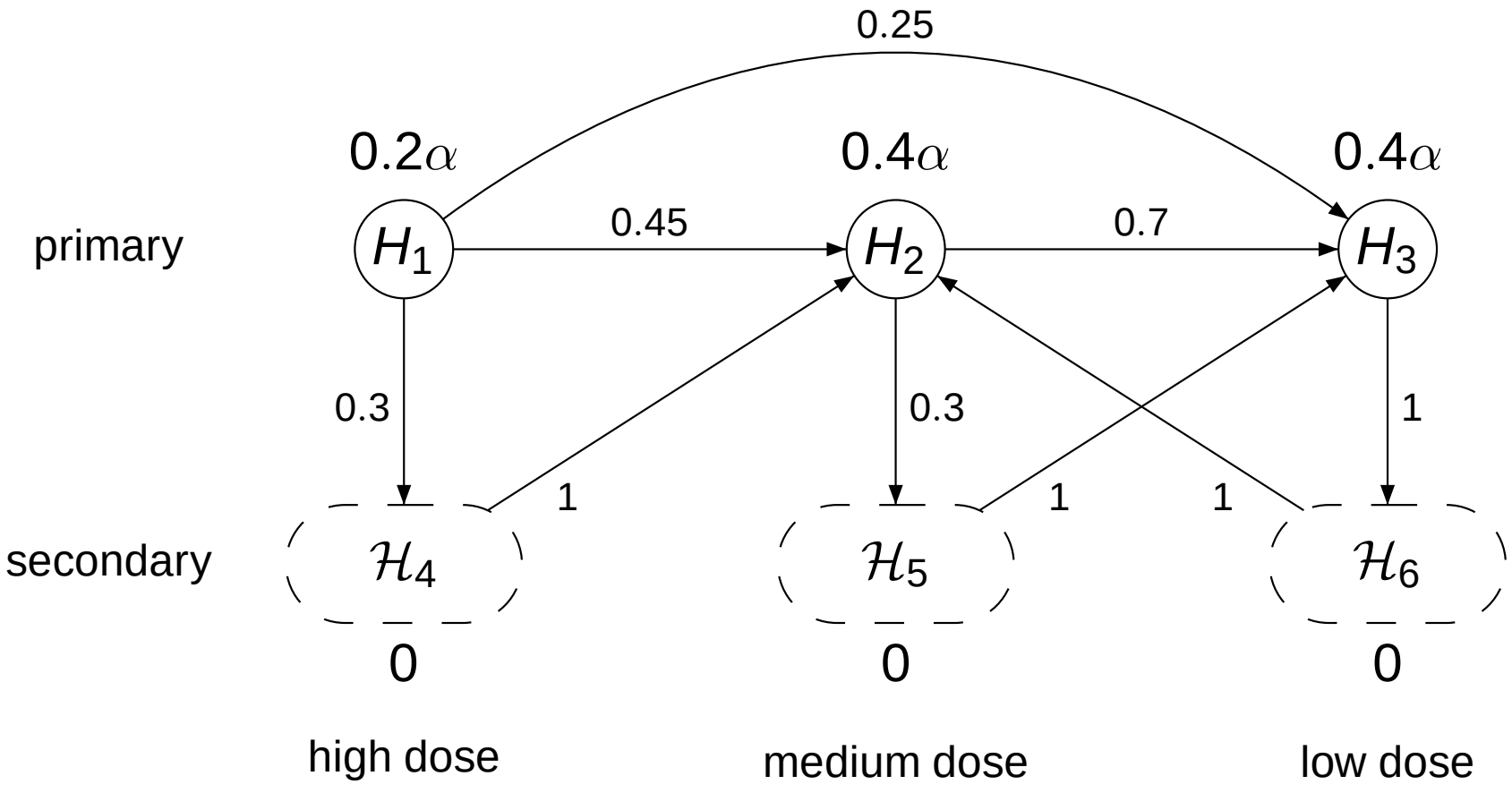


Tailored multiple test procedure



Key secondary endpoints within each dose group will be tested with a weighted Bonferroni-Holm test at the available local significance level

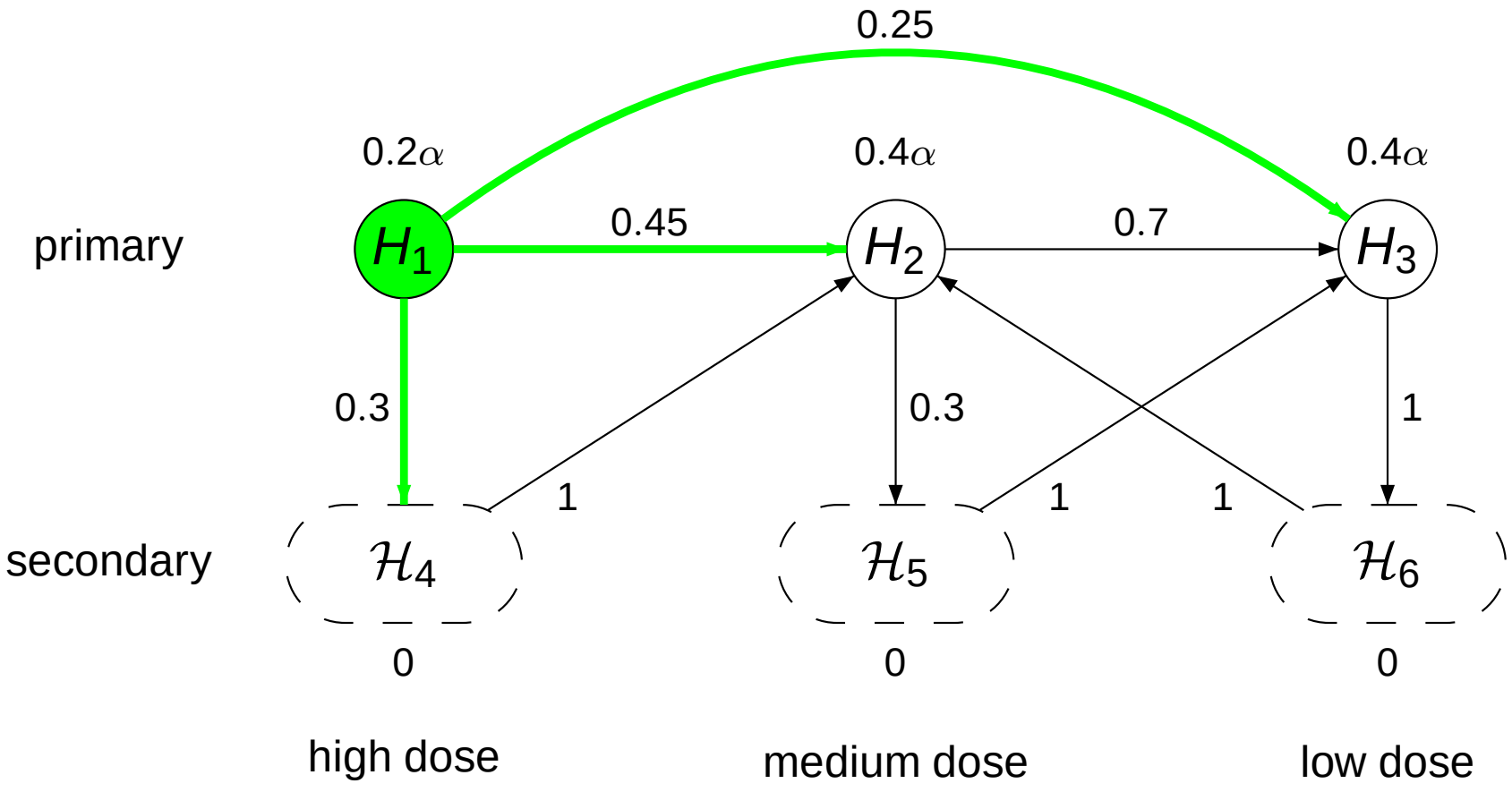
Tailored multiple test procedure



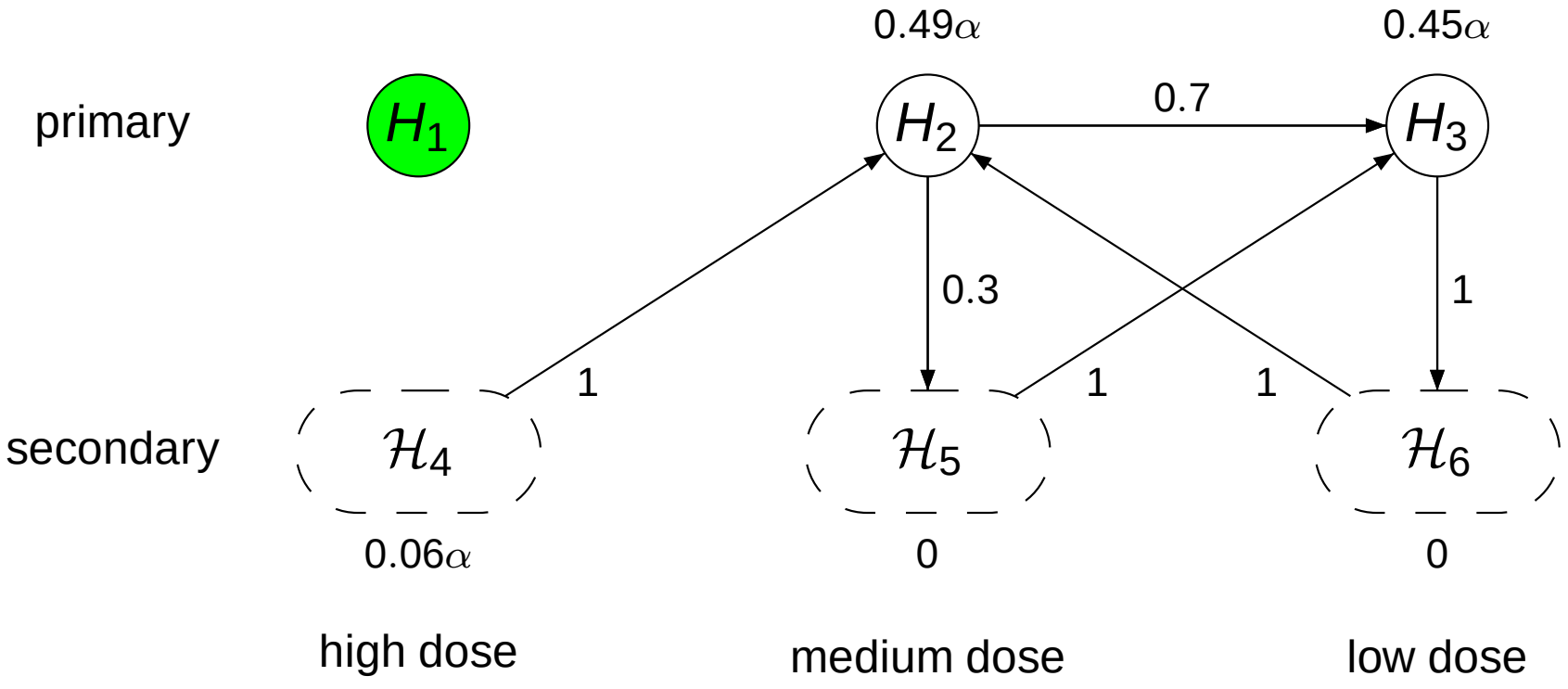
Key secondary endpoints within each dose group will be tested with a weighted Bonferroni-Holm test at the available local significance level

Improvement of the multiple testing procedure by using weighted Dunnett's test for all intersection hypotheses that contain at least two of H_1, H_2 and H_3 for the same endpoint

Example rejection sequence



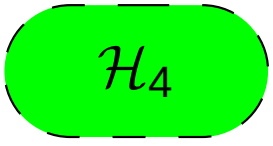
Example rejection sequence



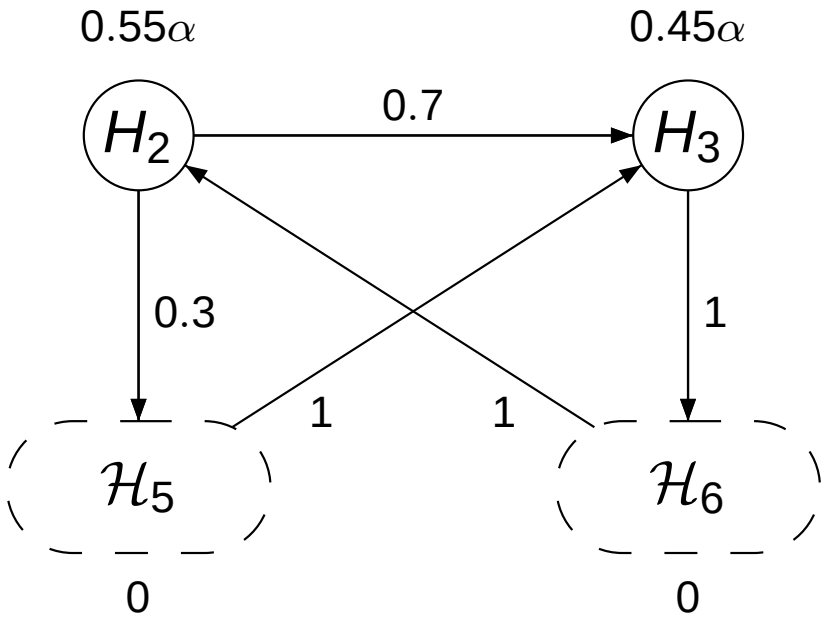
Example rejection sequence

primary

secondary



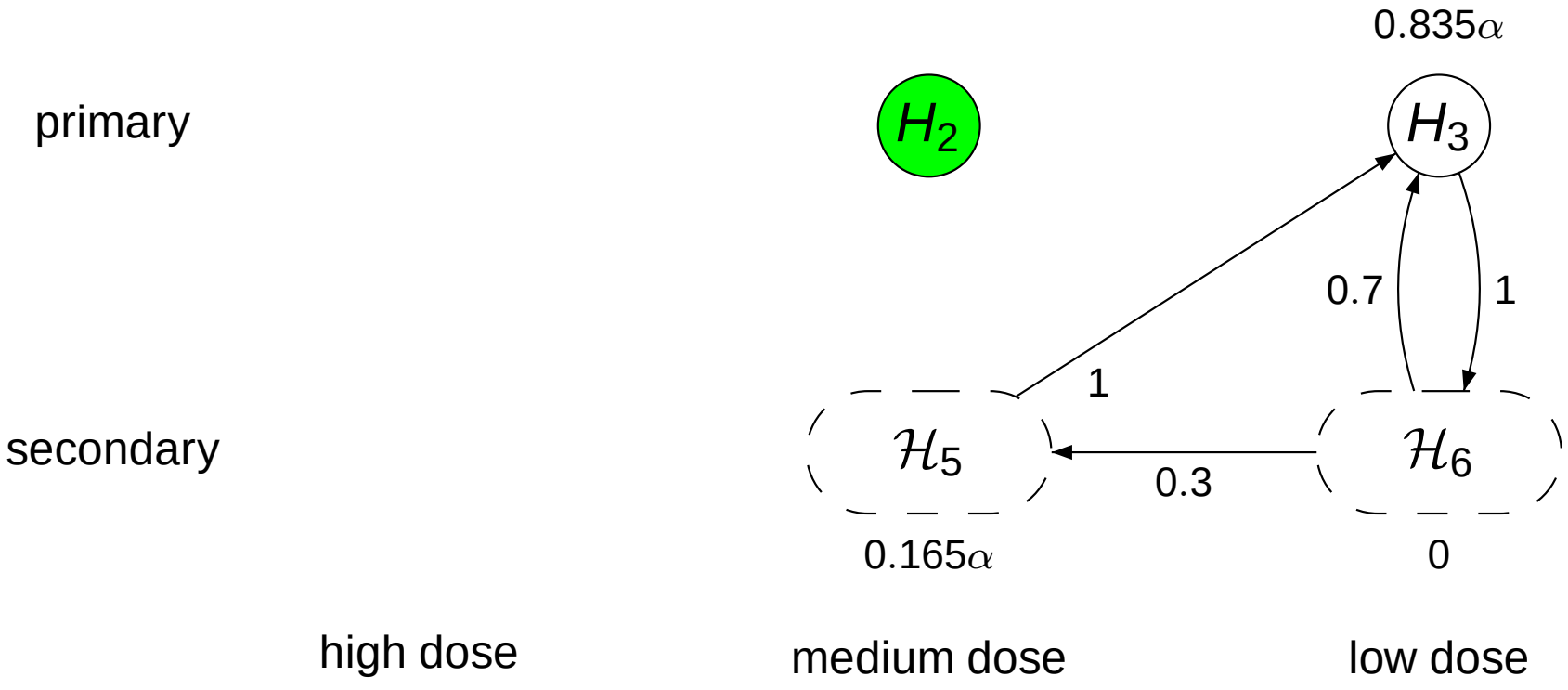
high dose



medium dose

low dose

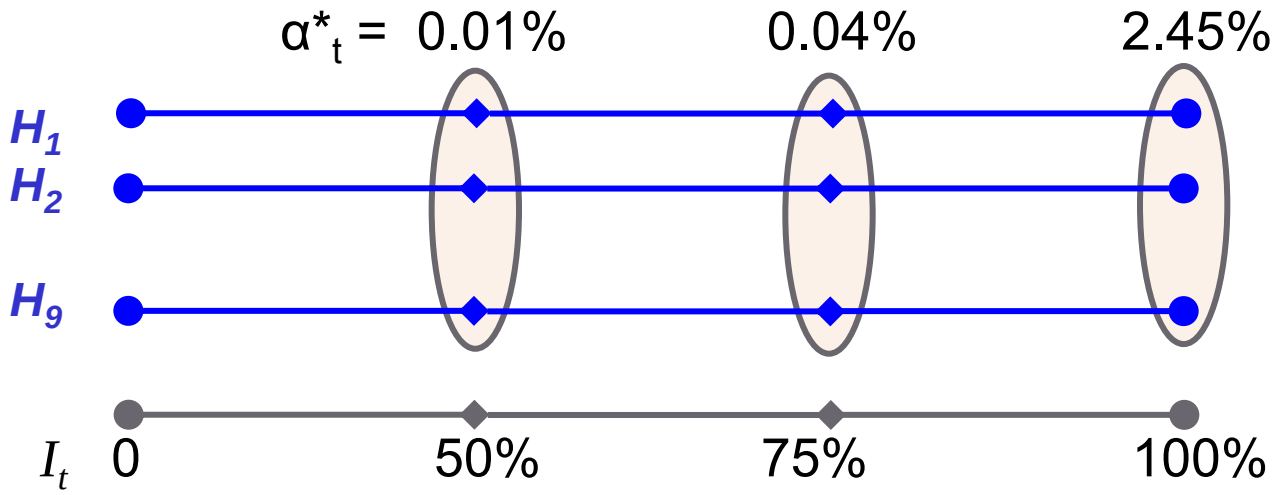
Example rejection sequence



Case study

Interim analyses: Bonferroni inequality for the repeated hypothesis testing

- Interim analyses at 50% and 75% information fraction I_t
  - A fixed Bonferroni split is used, with nominal overall significance levels α^*_t, t = 1, 2, 3, of 0.01% and 0.04% for the first and second efficacy interim analysis, leaving 2.45% for the final analysis.
  - At each of the 3 analyses, the graphical testing procedure exploiting correlations between test statistics will be performed at the respective nominal level α^*_t, controlling the familywise type I error rate at the overall (one-sided) significance level α = 2.5%.



Note: If a primary hypothesis is rejected, the respective secondary hypothesis cannot be tested at the full level α, even if the trial is stopped !

NOVARTIS

Summary and conclusions

- Confirmatory clinical trials are becoming increasingly more complex, often comparing multiple doses or treatments with a control for several primary and secondary endpoints.
- The multiple study objectives are reflected by structured families of hypotheses that are characterized by multiple groups of “parent” primary hypotheses and “descendant” secondary hypotheses.
- Novel graphical approaches for constructing and visualizing complex multiple test procedures with a focus on structured families of hypotheses are well suited to facilitate communication in clinical teams and to provide transparent decision strategies.
- Graphical procedures ensure strong control of the overall Type I error rate across all primary and secondary hypotheses.
- Multiple test procedure should be customized based on operating characteristics obtained via clinical scenario simulation.
- Ideally, EMA could provide in its new guidance a harmonized terminology and framework categorizing study objectives and endpoints with respect to their impact on approval and labeling and respective need for type 1 error control.

References

- Bretz, F., Maurer, W., Brannath, W., and Posch, M. (2009) A graphical approach to sequentially rejective multiple test procedures. Statistics in Medicine **28**, 586-604.
- Burman, C.-F., Sonesson, C. and Guilbaud, O. (2009) A recycling framework for the construction of Bonferroni-based multiple tests. Statistics in Medicine **28**, 739–761.
- Kordzakhia G. and Dmitrienko A. (2012) Superchain procedures in clinical trials with multiple objectives. Statist. Med. (early view)
- Maurer, W., Glimm, E., and Bretz, F. (2011) Multiple and repeated testing of primary, co-primary and secondary hypotheses. Statistics in Biopharmaceutical Research **3**, 336-352.
- Rohmeyer, K., Klinglmueller, F., Bornkamp, B. (2012) gMCP: Graph Based Multiple Test Procedures. R package version 0.8-0. URL http://CRAN.R-project.org/package=gMCP

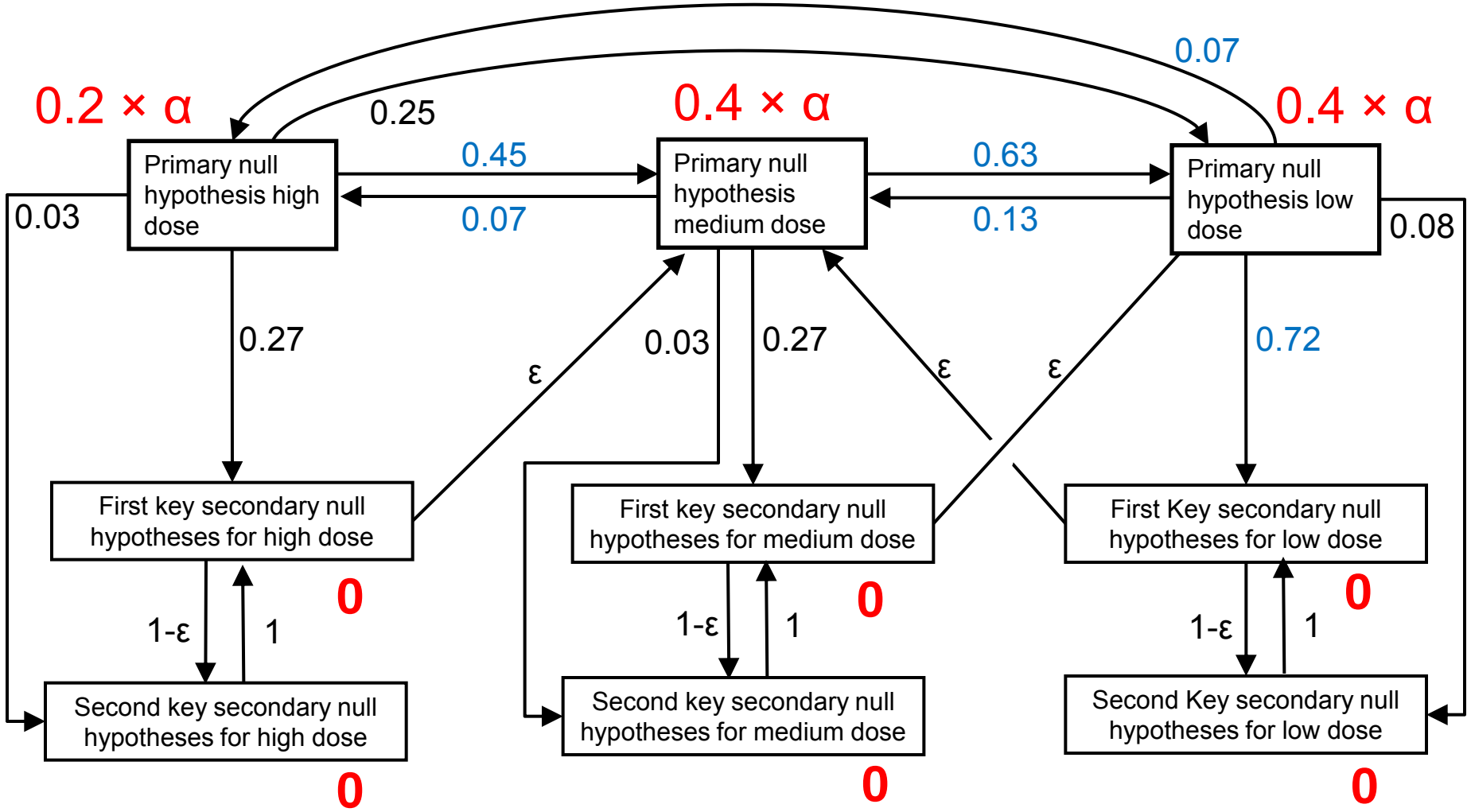
NOVARTIS

Back-up

NOVARTIS

Case study

Improvement of the multiple testing procedure by using weighted Dunnett's test for all intersection hypotheses that contain at least two of H_1, H_2 and H_3 for the same endpoint



NOVARTIS