A Group-Sequential Design to Test Efficacy and Inefficiency in Two Subgroups

Darja Tutschkow[1], Astrid Dempfle[2] & Nina Timmesfeld[1]

[1]Institute of Medical Biometry and Epidemiology,
Philipps-University Marburg
[2] Institute of Medical Informatics and Statistics,
Christian Albrechts-University Kiel

Contents

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Example

- trial where the benefit for the M^- was overlooked
- study led by the National Surgical Adjuvant Breast and Bowel Project and the North Central Cancer Treatment Group (Romond et al., 2005)
- effect of Trastuzumab for HER2-positive breast cancer patients
- only HER2-positive patients were included in the trial
- some of initially HER2-positve patients, appeared to be HER2-negative
- subsequently tested HER2-negative patients
- „benefit of adjuvant Trastuzumab may not be limited to patients with HER2 amplification“ (Paik, Kim & Wolmark, 2008)

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-
- study design to test for superiority or inefficiency of a new therapy in both subgroups

Motivation

- stratified medicine $\rightarrow$ tailored therapies for patient subgroups based on biomarkers
- predictive biomarkers $\rightarrow$ effect of a therapy depends on biomarker status
- biomarkers identified as predictive in retrospective or exploratory analyses
- 'issue of multiplicity' $\rightarrow$ risk of false positive findings
- biomarker-negative subgroup (M^-) not included in later phase III trail $\rightarrow$ no statistically confirmed evidence of inefficiency in M^-
- study design to test for superiority or inefficiency of a new therapy in both subgroups

Group-sequential Design for Both Subgroups

1. Assumptions: $X_{Ai}^j \sim \mathcal{N}(\mu_A^j, \sigma^2)$ iid, $X_{Bi}^j \sim \mathcal{N}(\mu_B^j, \sigma^2)$ iid with known σ^2 and $j \in \{+, -\}$
2. Hypotheses:

$$H_0^{j,S} : \delta^j \leq 0 \text{ vs. } H_1^{j,S} : \delta^j > 0 \text{ (Superiority)}$$
$$H_0^{j,I} : \delta^j \geq \Delta \text{ vs. } H_1^{j,I} : \delta^j < \Delta \text{ (Inefficiency)}$$

   where $\delta^j := \mu_A - \mu_B$ (difference in treatment effects in subgroup M^j) and $\Delta > 0$ (inefficiency margin)
3. Restrictions:
   1. predefined number of interim analyses
   2. equal amount of patients in each subgroup for each interim analysis

Group-sequential Design for Both Subgroups

1. Assumptions: $X_{Ai}^j \sim \mathcal{N}(\mu_A^j, \sigma^2)$ iid, $X_{Bi}^j \sim \mathcal{N}(\mu_B^j, \sigma^2)$ iid with known σ^2 and $j \in \{+, -\}$
2. Hypotheses:

$$H_0^{j,S} : \delta^j \leq 0 \text{ vs. } H_1^{j,S} : \delta^j > 0 \text{ (Superiority)}$$
$$H_0^{j,I} : \delta^j \geq \Delta \text{ vs. } H_1^{j,I} : \delta^j < \Delta \text{ (Inefficiency)}$$

   where $\delta^j := \mu_A - \mu_B$ (difference in treatment effects in subgroup M^j) and $\Delta > 0$ (inefficiency margin)

3. Restrictions:
   1. predefined number of interim analyses
   2. equal amount of patients in each subgroup for each interim analysis

Group-sequential Design for Both Subgroups

1. Assumptions: $X_{Ai}^j \sim \mathcal{N}(\mu_A^j, \sigma^2)$ iid, $X_{Bi}^j \sim \mathcal{N}(\mu_B^j, \sigma^2)$ iid with known σ^2 and $j \in \{+, -\}$
2. Hypotheses:

$$H_0^{j,S} : \delta^j \leq 0 \text{ vs. } H_1^{j,S} : \delta^j > 0 \text{ (Superiority)}$$
$$H_0^{j,I} : \delta^j \geq \Delta \text{ vs. } H_1^{j,I} : \delta^j < \Delta \text{ (Inefficiency)}$$

   where $\delta^j := \mu_A - \mu_B$ (difference in treatment effects in subgroup M^j) and $\Delta > 0$ (inefficiency margin)
3. Restrictions:
   1. predefined number of interim analyses
   2. equal amount of patients in each subgroup for each interim analysis

Hierarchical Testing

- first intermin analysis:
  - testing superiority in M^+ and inefficiency in M^-
  - if one of either hypotheses is rejected, test it in the other subgroup
- following interim analyses:
  - testing superiority in M^+ and inefficiency in M^- as long as no hypothesis is rejected
  - testing both hypothesis in a subgroup if a hypothesis got rejected in the other subgroup

Hierarchical Testing

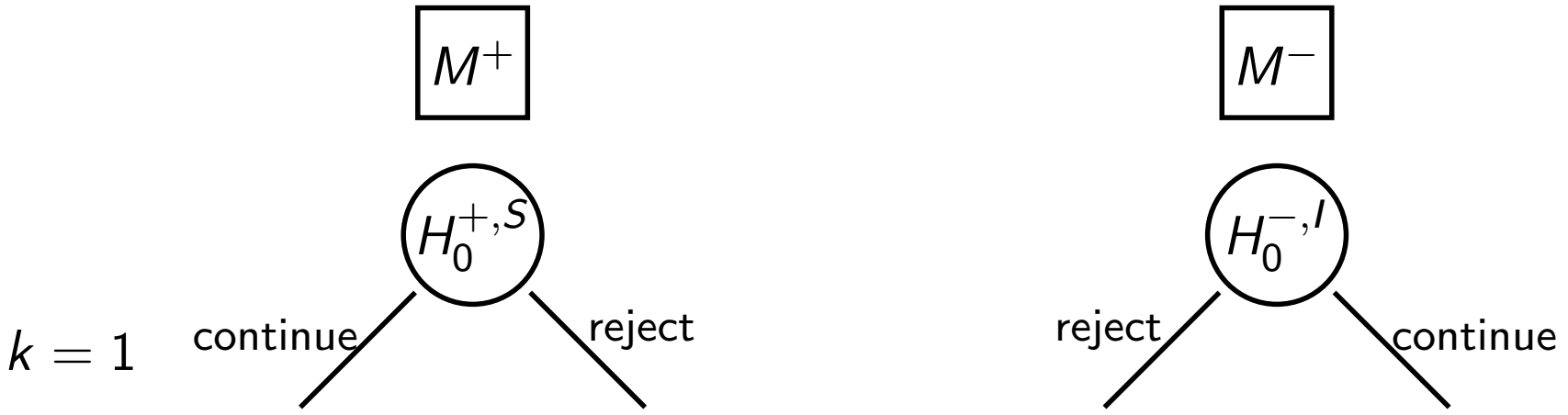
- first intermin analysis:
  - testing superiority in M^+ and inefficiency in M^-
  - if one of either hypotheses is rejected, test it in the other subgroup
- following interim analyses:
  - testing superiority in M^+ and inefficiency in M^- as long as no hypothesis is rejected
  - testing both hypothesis in a subgroup if a hypothesis got rejected in the other subgroup

Hierarchical Testing - Example

$k = 1$

$k = 2$

Hierarchical Testing - Example

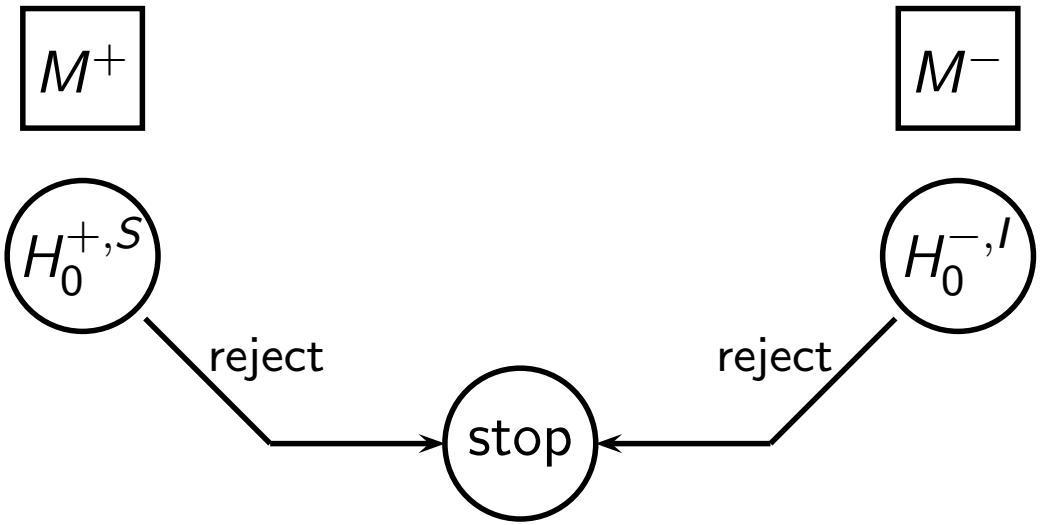


$k = 1$

$k = 2$

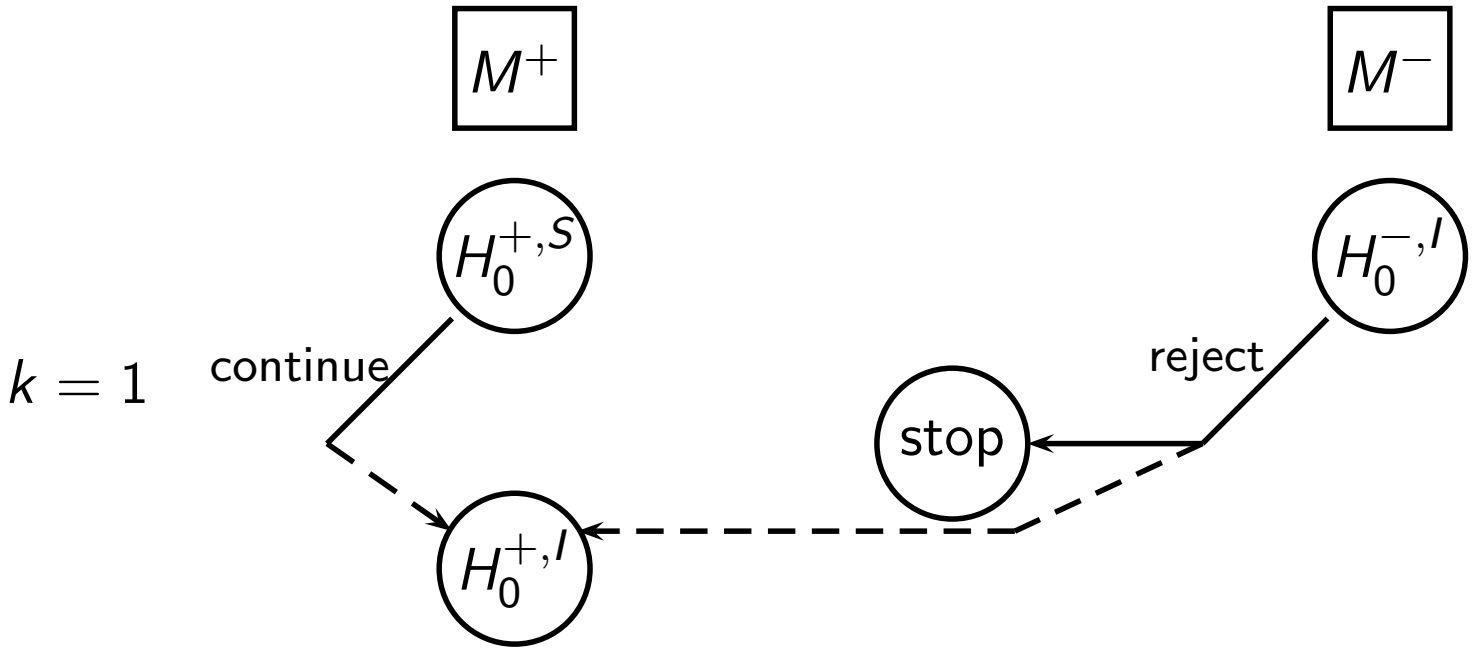
Hierarchical Testing - Example

$k = 1$

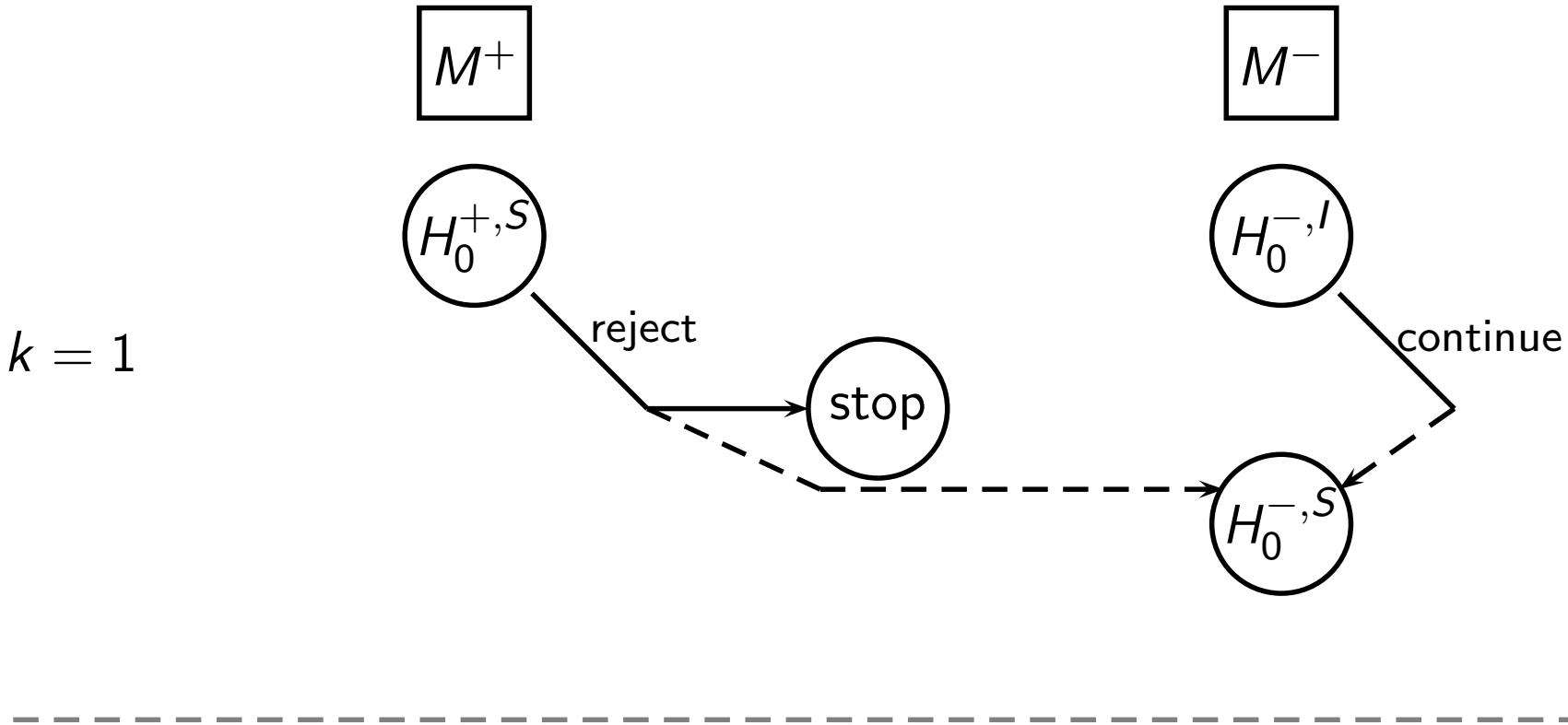


$k = 2$

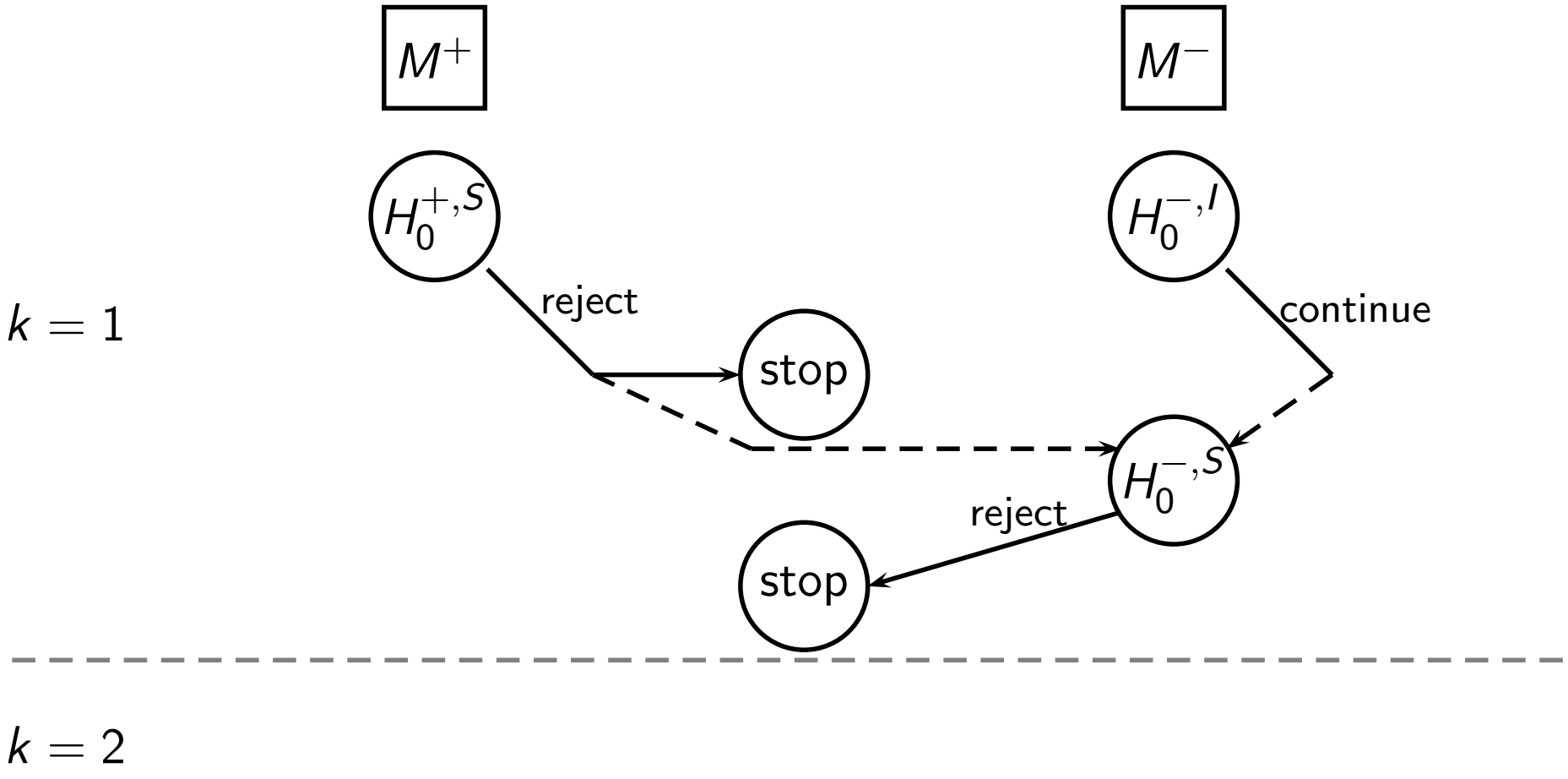
Hierarchical Testing - Example



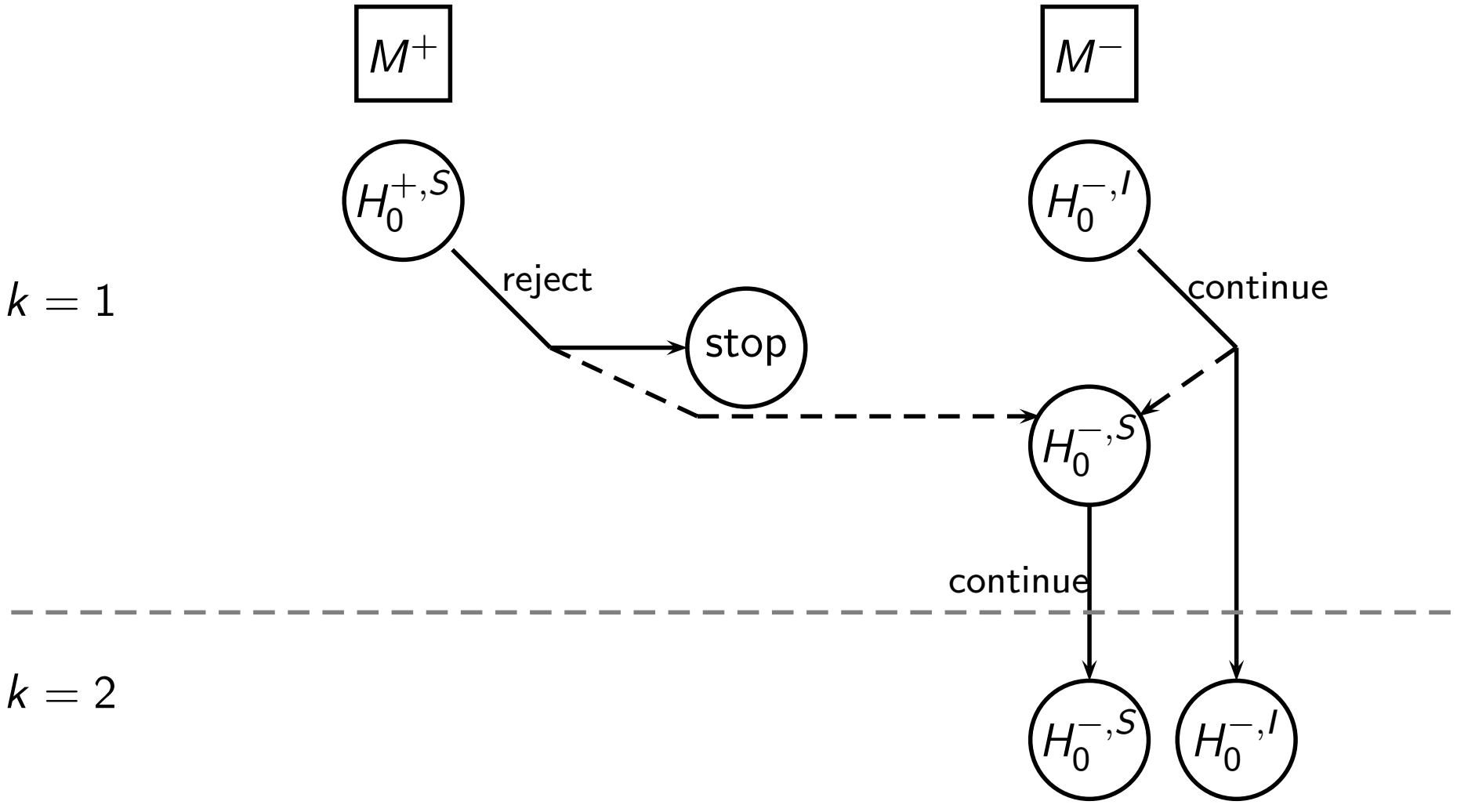
Hierarchical Testing - Example



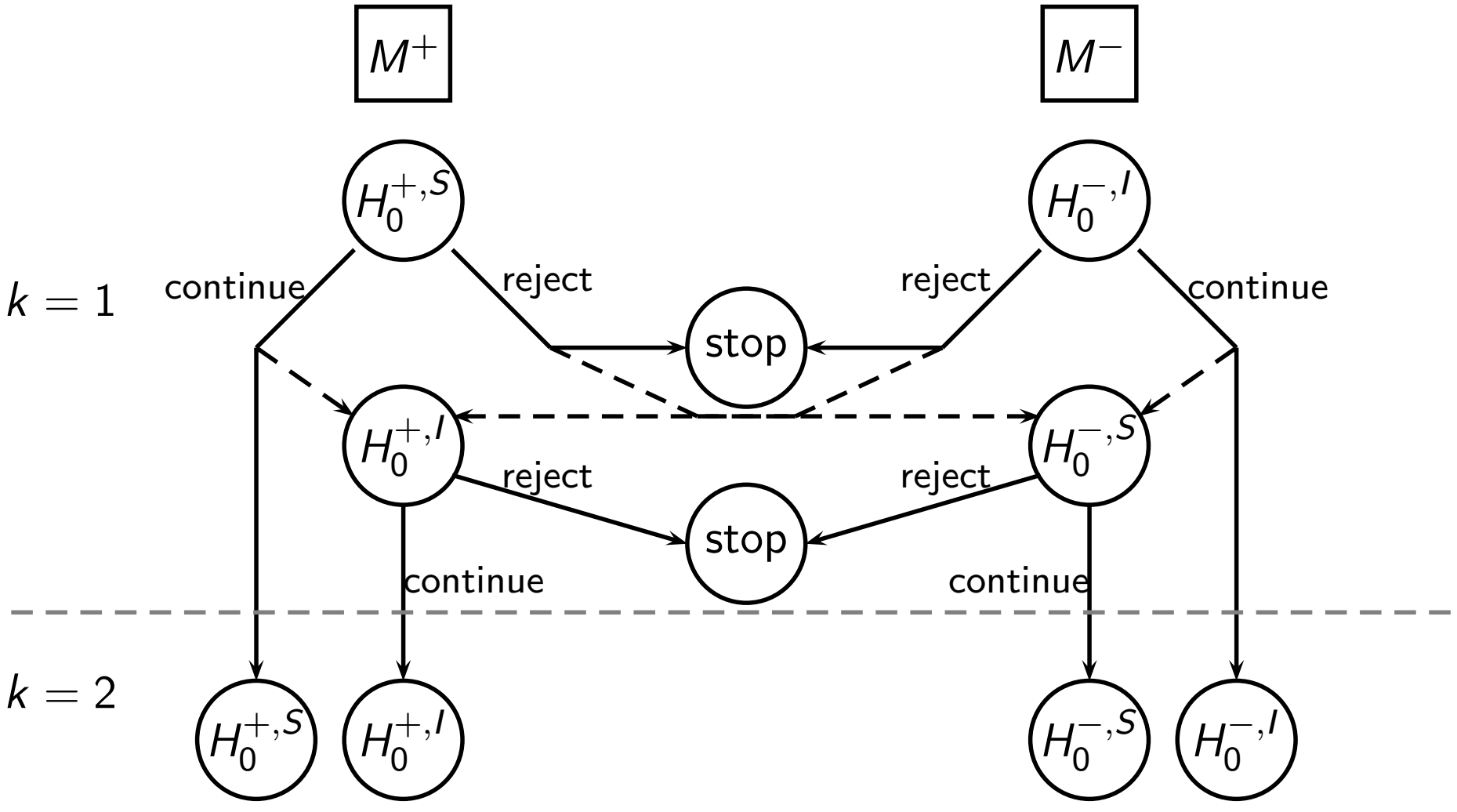
Hierarchical Testing - Example



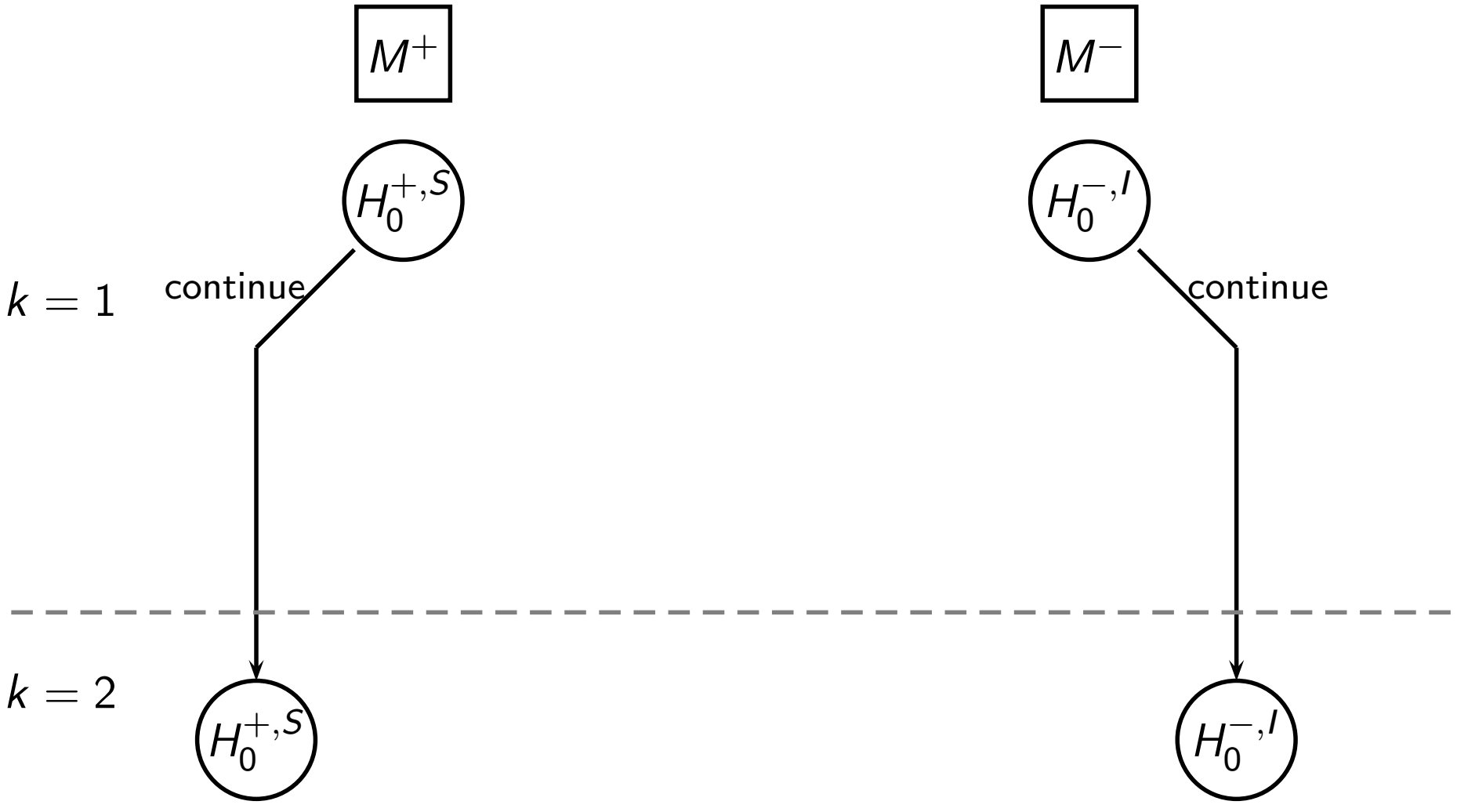
Hierarchical Testing - Example



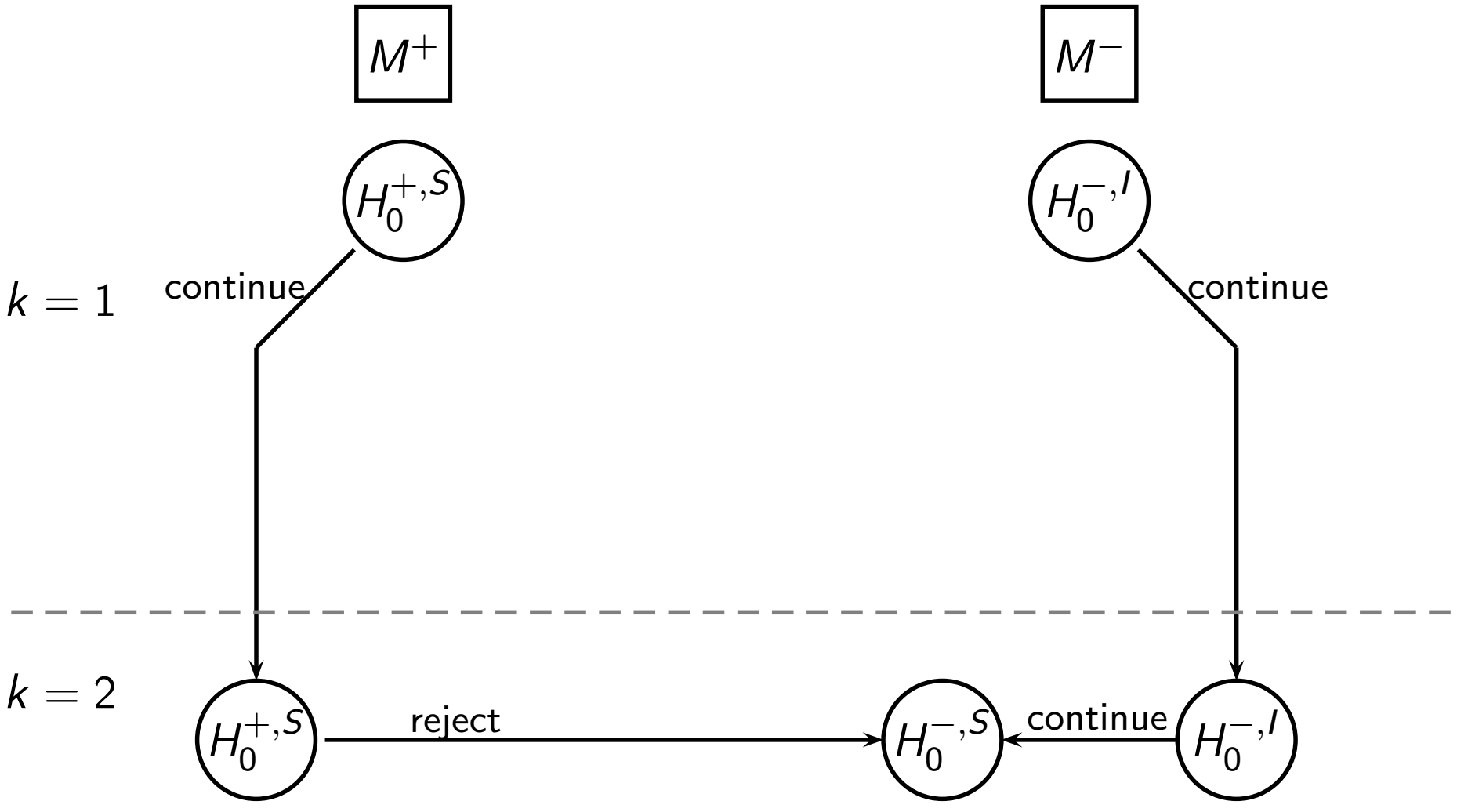
Hierarchical Testing - Example



Hierarchical Testing - Example



Hierarchical Testing - Example



Simulation

- rejection probabilities for different treatment effects in both subgroups
- expected sample size in both subgroups
- computing the number of required patients N^+ and N^- in M^+ and M^- respectively, such that decision at the last analysis $\rightarrow N^+ = N^- = 2 \cdot 72$
- FWER 5 %
- K=2 analyses
- 10.000 repetitions
- $\Delta = 0.5$
- $\sigma^2 = 1$

Results

δ^+	δ^-	$H_0^{+,S}$	$H_0^{+,I}$	$H_0^{-,I}$	$H_0^{-,S}$	$\mathbb{E}(N^+)$	$\mathbb{E}(N^-)$
0	Δ	0.025	0.024	0.025	0.024	142	142
0	2Δ	0.025	0	0	0.025	143	143
0	0	0.024	0.963	0.984	0	100	88
Δ	Δ	0.986	0	0.025	0.962	88	100
Δ	0	0.979	0.021	0.980	0.020	88	88
2Δ	0	1	0	0.977	0.023	72	87
2Δ	2Δ	1	0	0	1	72	72

$H_0^{j,S} : \delta^j \leq 0$ vs. $H_1^{j,S} : \delta^j > 0$ (Superiority)

$H_0^{j,I} : \delta^j \geq \Delta$ vs. $H_1^{j,I} : \delta^j < \Delta$ (Inefficiency); $\Delta = 0,5$

Results

δ^+	δ^-	$H_0^{+,S}$	$H_0^{+,I}$	$H_0^{-,I}$	$H_0^{-,S}$	$\mathbb{E}(N^+)$	$\mathbb{E}(N^-)$
0	Δ	0.025	0.024	0.025	0.024	142	142
0	2Δ	0.025	0	0	0.025	143	143
0	0	0.024	0.963	0.984	0	100	88
Δ	Δ	0.986	0	0.025	0.962	88	100
Δ	0	0.979	0.021	0.980	0.020	88	88
2Δ	0	1	0	0.977	0.023	72	87
2Δ	2Δ	1	0	0	1	72	72

$H_0^{j,S} : \delta^j \leq 0$ vs. $H_1^{j,S} : \delta^j > 0$ (Superiority)

$H_0^{j,I} : \delta^j \geq \Delta$ vs. $H_1^{j,I} : \delta^j < \Delta$ (Inefficiency); $\Delta = 0,5$

Results

δ^+	δ^-	$H_0^{+,S}$	$H_0^{+,I}$	$H_0^{-,I}$	$H_0^{-,S}$	$\mathbb{E}(N^+)$	$\mathbb{E}(N^-)$
0	Δ	0.025	0.024	0.025	0.024	142	142
0	2Δ	0.025	0	0	0.025	143	143
0	0	0.024	0.963	0.984	0	100	88
Δ	Δ	0.986	0	0.025	0.962	88	100
Δ	0	0.979	0.021	0.980	0.020	88	88
2Δ	0	1	0	0.977	0.023	72	87
2Δ	2Δ	1	0	0	1	72	72

$H_0^{j,S} : \delta^j \leq 0$ vs. $H_1^{j,S} : \delta^j > 0$ (Superiority)

$H_0^{j,I} : \delta^j \geq \Delta$ vs. $H_1^{j,I} : \delta^j < \Delta$ (Inefficiency); $\Delta = 0,5$

Summary and Outlook

- group-sequantial design to test for superiority and inefficiency for both subgroups for normally distributed data
- next steps:
  1. account for different group sizes
  2. extension for survival data
- at some point: add more flexibility, e.g.
  - start with the full set, switch to hierarchical procedure and the other way around
  - increase or reduce number of interim analyses
  - change test statistic or outcome measure during the course of the trail, etc. $\rightarrow$ CRP-method (Müller & Schäfer)

Summary and Outlook

- group-sequantial design to test for superiority and inefficiency for both subgroups for normally distributed data
- next steps:
  1. account for different group sizes
  2. extension for survival data
- at some point: add more flexibility, e.g.
  - start with the full set, switch to hierarchical procedure and the other way around
  - increase or reduce number of interim analyses
  - change test statistic or outcome measure during the course of the trail, etc. $\rightarrow$ CRP-method (Müller & Schäfer)

Summary and Outlook

- group-sequantial design to test for superiority and inefficiency for both subgroups for normally distributed data
- next steps:
  1. account for different group sizes
  2. extension for survival data
- at some point: add more flexibility, e.g.
  - start with the full set, switch to hierarchical procedure and the other way around
  - increase or reduce number of interim analyses
  - change test statistic or outcome measure during the course of the trail, etc. $\rightarrow$ CRP-method (Müller & Schäfer)

Thank you for your attention! Questions?

Thank you for your attention! Questions?