Gatekeeping Procedures

- Gatekeeping procedure is a multiple testing procedure to address:

  Multiplicity issues: multiple endpoints, dose-control comparisons, objectives, different time points

  – With **hierarchical structure** of hypotheses: hypotheses often can be grouped into families to reflect hierarchical nature

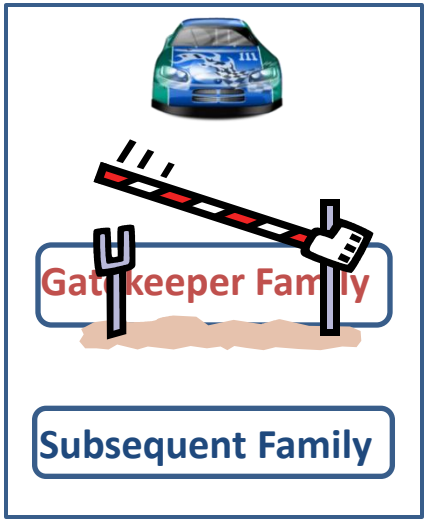
	Low dose vs. PBO	Middle dose vs. PBO	High dose vs. PBO
Primary Endpoint (BRS)	$H_1^P: \theta_1^P \le 0$	$H_2^P: \theta_2^P \le 0$	$H_3^P: \theta_3^P \le 0$
Secondary Endpoint (OF)	$H_1^S: \theta_1^S \le 0$	$H_2^S: \theta_2^S \le 0$	$H_3^S: \theta_3^S \le 0$

$F_1 = \{H_1^P, H_2^P, H_3^P\}$

$F_2 = \{H_1^S, H_2^S, H_3^S\}$

  – With some **pre-specified rules/conditions**:

  Rules/Conditions can be due to regulatory requirement, company position, or other reason



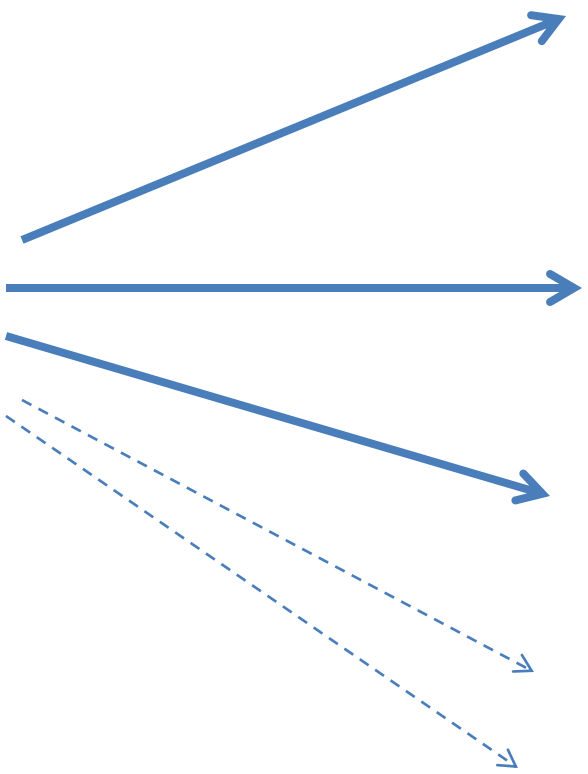
Gatekeeping Procedures

Multiplicity Problems

+

Hierarchical Structure

Pre-specified Rules/Conditions



Serial Gatekeeping Procedures

Parallel Gatekeeping Procedures

Tree Structured Gatekeeping Procedures

Fallback Procedures

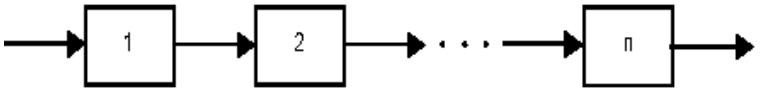
Other methods

Gatekeeping Procedures: Serial vs. Parallel

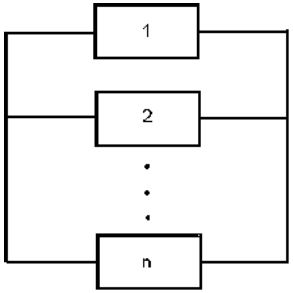
- **Serial:** Proceed to next family **only if all** hypotheses are rejected in gatekeeper family
  - e.g., Alzheimer's disease, 2 primary endpoints are generally required: Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog) and Clinical Global Impression of Change (CGIC)
  - sometimes too stringent
- **Parallel:** Proceed to next family if **at least one** hypothesis is rejected in gatekeeper family

<u>Analogous to Reliability Theory</u>

- **Serial:** Similar to a system with basic elements connected in series and strength of the system depends on each element

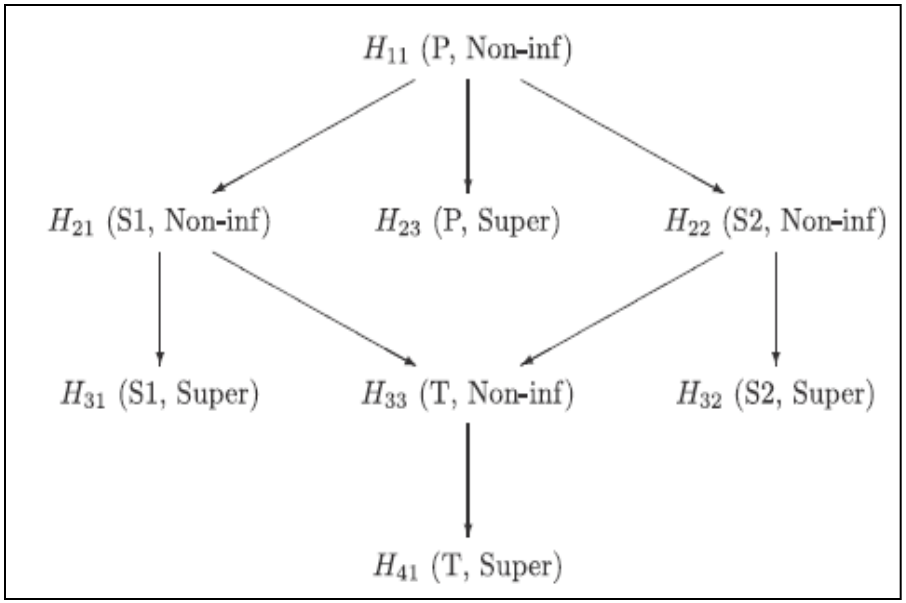


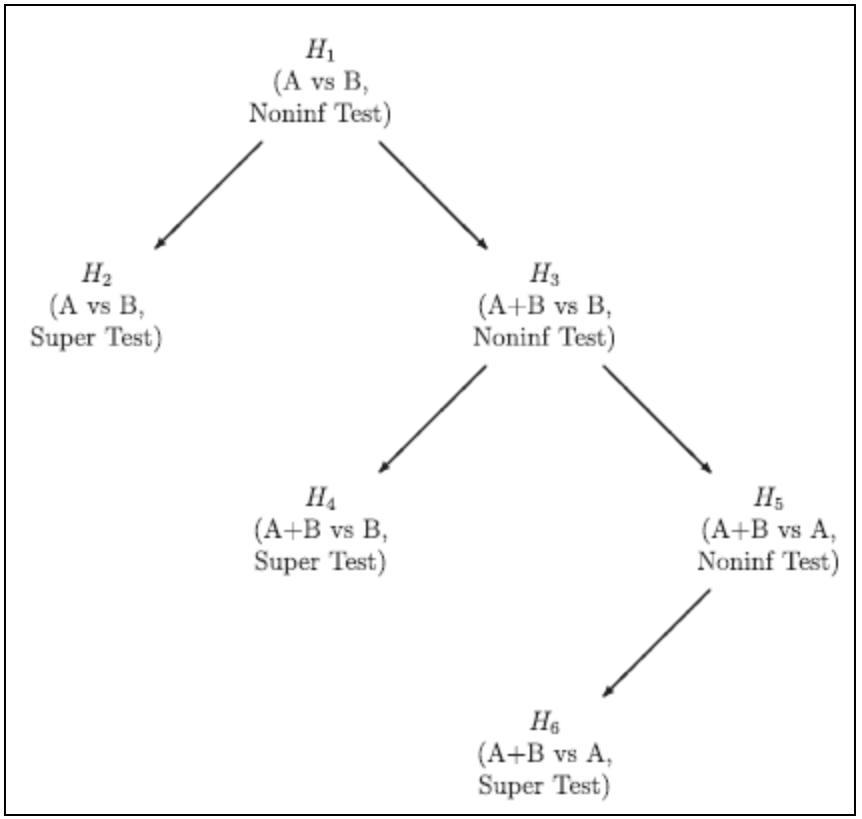
- **Parallel:** Similar to a system with elements connected in parallel



Gatekeeping Procedures: Tree structured

- Generalization of serial and parallel gatekeeping
- Decision-making process no longer exhibits a simple sequential structure but rather relies on a decision tree with multiple branches corresponding to individual objectives.





Dmitrienko A, Wiens BL, Tamhane AC, Wang X. *Statistics in Medicine* 2007; **26**: 2465-2478.

One primary endpoint

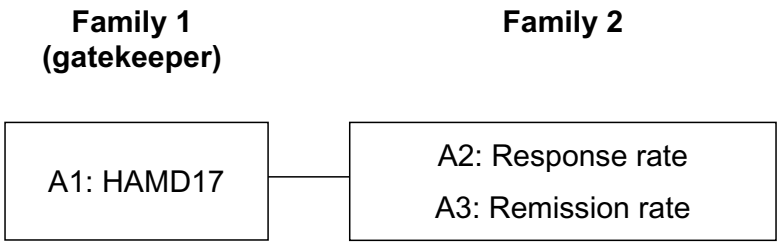
Depression trial

- Single primary endpoint
  - 17-item Hamilton depression rating scale (HAMD17)
  - Successful outcome if the drug is superior to placebo
- Two important secondary endpoints
  - Response rate based on HAMD17
  - Remission rate based on HAMD17

Serial gatekeeping strategy

- Propose including the secondary findings in the product label if the primary endpoint is significant

Serial gatekeeping strategy



Step 1: Primary analysis at α level

- No adjustment for multiplicity

Step 2: Secondary analyses if the primary analysis yielded a significant result

- Stepwise Holm test to adjust for multiplicity within Family 2
- No adjustment for the primary endpoint (memory-less method)

Serial gatekeeping strategy

Endpoint	Raw p	Adjusted p
Primary: HAMD17	0.046	0.046
Secondary: Response rate	0.048	0.048
Secondary: Remission rate	0.021	0.042

Primary and secondary endpoints are significant at 5% level

- Justification for including the secondary endpoints in the product label

Multiple primary endpoints

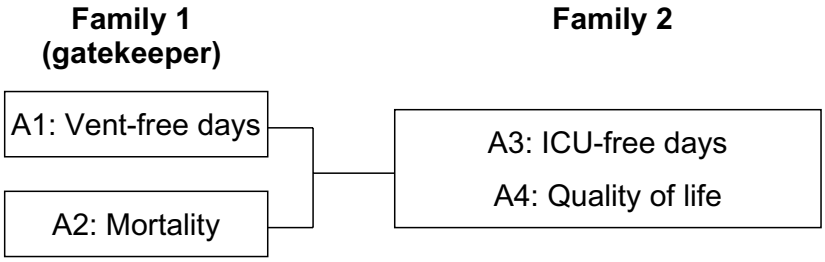
Clinical trial in patients with acute lung injury (ALI)

- Two primary endpoints
  - Number of days patients are off mechanical ventilation (vent-free days)
  - 28-day all-cause mortality rate
  - Successful outcome if the drug is superior to placebo with respect to either endpoint
- Two important secondary endpoints
  - Number of days patients are out of ICU (ICU-free days)
  - Overall quality of life at the end of the study

Parallel gatekeeping strategy

- Propose including the secondary findings in the product label provided at least one primary endpoint is significant

Parallel gatekeeping strategy



Step 1: Primary analysis at overall α level

– Adjustment for multiplicity within Family 1

Step 2: Secondary analyses if at least one primary analysis yielded a significant result

– Adjustment for multiplicity within Family 2 will depend on the number of significant primary outcomes (not memory-less anymore)

Parallel gatekeeping test

Closed testing principle

– Gatekeeping tests are constructed using the closed testing principle

Stepwise representation

– Family 1: Bonferroni test at overall α level

  k is the number of significant outcomes

– Family 2: Stepwise Holm test

  Overall significance level is $\alpha k/2$

  No multiplicity adjustment for the primary endpoints (memory-less method) if both primary endpoints are significant (k=2)

  Penalty if only one primary endpoint is significant (k=1)

  Secondary analyses are not performed if the primary endpoints are not significant (k=0)

ALI clinical trial: Scenario 1

Two significant primary variables

– Significant improvement in the mean number of ventilator-free days and 28-day all-cause mortality

Endpoint	Raw p	Adjusted p
Primary: Vent-free days	0.024	0.027
Primary: Mortality	0.003	0.030
Secondary: ICU-free days	0.026	0.029
Secondary: Quality of life	0.002	0.027

All analyses are significant at 5% level

– Justification for including the secondary endpoints in the product label

ALI clinical trial: Scenario 2

Single significant primary variable

– Significant improvement in 28-day all-cause mortality but not in mean number of ventilator-free day

Endpoint	Raw p	Adjusted p
Primary: Vent-free days	0.084	0.093
Primary: Mortality	0.003	0.030
Secondary: ICU-free days	0.026	0.093
Secondary: Quality of life	0.002	0.040

Primary mortality analysis and secondary quality of life analysis are significant at 5% level

– Justification for including the secondary endpoints in the product label

Dose-ranging study

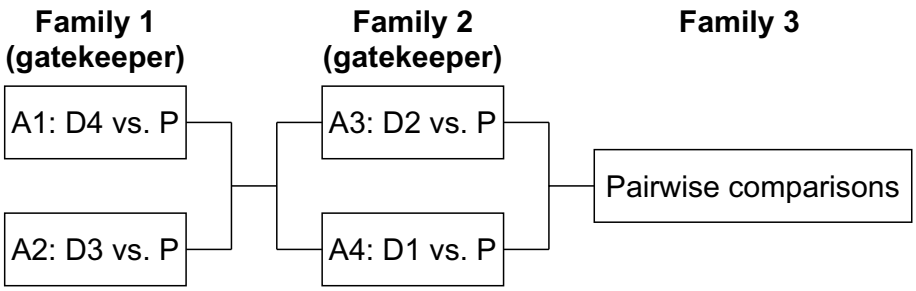
Clinical trial in patients with hypertension

- Four doses of an experimental drug are compared to placebo
  - Doses are labeled as D1, D2, D3 and D4
- Primary endpoint
  - Reduction in diastolic blood pressure

Objectives of the study

- Find the doses with a significant reduction in diastolic blood pressure compared to placebo
- Study the shape of the dose-response curve

Parallel gatekeeping strategy



Step 1: Compare doses D3 and D4 to placebo

Step 2: Compare doses D1 and D2 to placebo if at least one comparison at Step 1 is significant

Step 3: Pairwise dose comparisons if at least one comparison at Step 2 is significant

Parallel gatekeeping strategy

Comparison	Raw p	Adjusted p		
		Gatekeeping procedure	Holm procedure	Dunnett procedure
D4 vs. P	0.0008	0.0016	0.0055	0.0030
D3 vs. P	0.0135	0.0269	0.0673	0.0459
D2 vs. P	0.0197	0.0394	0.0787	0.0656
D1 vs. P	0.7237	1.0000	1.0000	0.9899
D4 vs. D1	0.0003	0.0394	0.0021	
D4 vs. D2	0.2779	1.0000	0.8338	
D3 vs. D1	0.0054	0.0394	0.0324	
D3 vs. D2	0.8473	1.0000	1.0000	

Doses D2, D3 and D4 are significantly different from placebo at 5% level

Comments

Basic gatekeeping framework

– Focused on gatekeeping procedures based on Bonferroni test

More powerful gatekeeping tests

– Based on more powerful tests, e.g., Simes test

– Based on tests accounting for the correlation among the endpoints

Exact parametric tests such as Dunnett test and approximate resampling-based Westfall-Young tests

Software implementation

– SAS macros for performing gatekeeping inferences in *Analysis of Clinical Trials Using SAS* (Chapter 2)