Multi-stage Gatekeeping Procedures with Clinical Trial Applications

Alex Dmitrienko Eli Lilly and Company

Ajit C. Tamhane* Northwestern University

*Research partially supported by grants from NHLBI and NSA

Outline

Gatekeeping testing strategies

General multi-stage gatekeeping procedures

Dmitrienko, Tamhane and Wiens (2007)

Truncated multiple tests

Clinical trial applications

Multiple objectives

Clinical trials with primary and secondary objectives

Product labels typically focus on primary findings

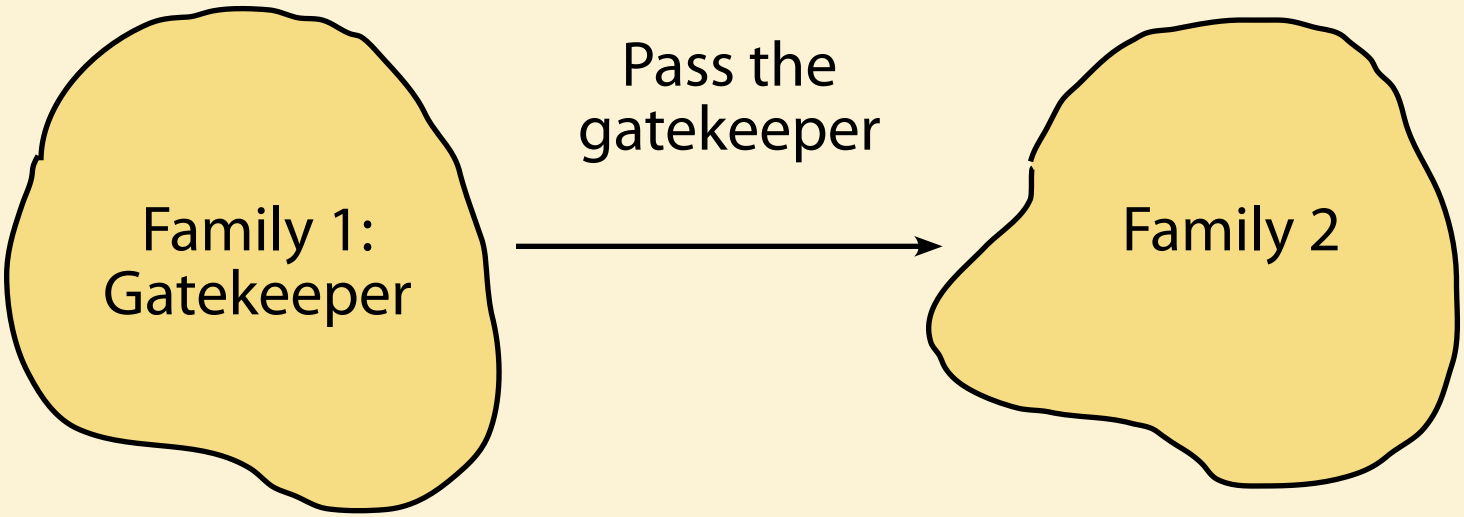
Secondary analyses (secondary endpoints or subgroup analyses) provide much useful information to prescribing physicians, patients, hospital administrators, etc

Gatekeeping testing strategies

Account for the hierarchical structure of multiple analyses

Control the familywise error rate (FWER)

Gatekeeping strategies



Parallel gatekeeper

At least one significant result in the gatekeeper to proceed to the next family of analyses

Clinical trial example

Acute lung injury trial

Dmitrienko, Offen and Westfall (2003)

Family 1 (parallel gatekeeper)

Number of ventilator-free days (Endpoint P_1)

28-day all-cause mortality (Endpoint P_2)

Family 2

Number of ICU-free days (Endpoint S_1)

Quality of life (Endpoint S_2)

Weights

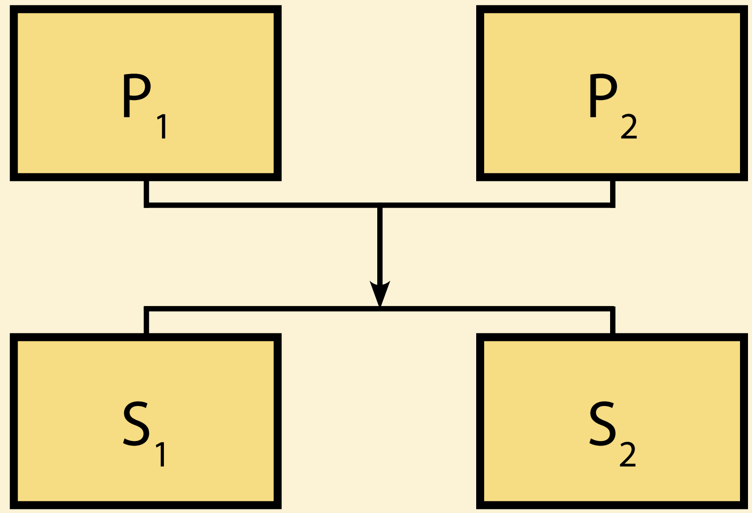
Endpoints are equally weighted within each family

Acute lung injury trial

Parallel gatekeeping procedure

Family 1 (parallel gatekeeper)

Family 2



Secondary endpoints are tested if at least one primary test is significant

Higher likelihood of detecting treatment effect for secondary endpoints

Acute lung injury trial

Stepwise parallel gatekeeping procedure

Stepwise gatekeeping procedure

Dmitrienko, Tamhane, Wang and Chen (2006)

Bonferroni-based stepwise parallel gatekeeping procedure

Type I error rate control

Familywise error rate (FWER) is controlled in the strong sense at α

Acute lung injury trial

Stepwise parallel gatekeeping procedure

Family 1

Bonferroni test at α level

Endpoint P1, $p_1 \leq \alpha/2$ and Endpoint P2, $p_2 \leq \alpha/2$

Family 2

Penalized Holm test at $\rho\alpha$ level

Rejection gain factor ρ

$\rho=1$ if two significant outcomes in Family 1

$\rho=1/2$ if one significant outcome in Family 1

$\rho=0$ if no significant outcomes in Family 1

Gatekeeping procedures

Extensions

Two-stage procedure

Can more powerful tests be used in Family 1?

Yes but not Holm or Hochberg tests

Can more powerful tests be used in Family 2?

Any FWER-controlling multiple test can be used

General multi-stage procedure

How can powerful multi-stage gatekeeping procedures be constructed?

They can be built recursively starting with a two-stage case

Two-stage case

Two families of hypotheses

$F_1=\{H_{11},\ldots,H_{1n}\}$ and $F_2=\{H_{21},\ldots,H_{2n}\}$

Family 1 test

Error rate function defines the fraction of Type I error rate can be carried over to Family 2

$$e_1(I) = P\left(\text{Reject at least one } H_{1i}, i \in I \mid \bigcap_{i \in I} H_{1i}\right)$$

for any $I \subseteq N = \{1,\ldots,n\}$

Desirable properties

$$e_1(\varnothing) = 0,\ \ e_1(I) \le e_1(J),\ I \subseteq J,\ e_1(N) = \alpha$$

Two-stage case

Gatekeeping procedure

Stage 1

Test F_1 using an FWER-controlling test at $\alpha_1=\alpha$

A_1 is the index set of accepted hypotheses in Family 1

Stage 2

Test F_2 using an FWER-controlling test at

$\alpha_2=\alpha_1-e_1(A_1)=\alpha-e_1(A_1)$

Notes

$\alpha_2=0$ ($\alpha_2=\alpha$) if all hypotheses are accepted (rejected) in Family 1

"Use it or lose it" principle

Two-stage case

Familywise error rate control

Proposition

Two-stage gatekeeping procedure controls FWER at α if the separability condition is met:

$e_1(I) < \alpha$ for any proper subset of N

Proof

Dmitrienko, Tamhane and Wiens (2007)

Special case

Guilbaud (2007) proved this in a special case:

Family 1: Bonferroni test

Family 2: Any FWER-controlling test

Separability condition

Separability condition

A certain fraction of α can be carried over to the next family if one or more hypotheses are rejected

Separable tests

Bonferroni test is separable (satisfies the separability condition) because $e_1(I) = \alpha \, | \, I \, | \, / \, n$ when hypotheses are equally weighted

Holm or Hochberg tests are not separable tests

Example: If one hypothesis is true and others are infinitely false, the probability of rejecting the true hypothesis is α for Holm and Hochberg tests

Separability condition

Truncated tests

Truncated tests

A truncated test is based on a convex combination between a multiple test and Bonferroni test

A truncated test is less powerful than the original test but more powerful than Bonferroni test

A truncated test is separable

Examples

Truncated p-value based tests (Holm, Hochberg and fallback tests), truncated parametric tests (truncated step-down Dunnett test) or truncated resampling-based tests

Truncated Holm test

General form

Single family of hypotheses $H_1,\ldots,H_n$

Ordered p-values $p_{(1)}<\ldots<p_{(n)}$

Ordered null hypotheses $H_{(1)},\ldots,H_{(n)}$

Truncation fraction $0\leq\gamma\leq 1$

Condition	Decision
$p_{(1)}\leq\gamma\alpha/n+(1-\gamma)\alpha/n$	$H_{(1)}$ is rejected
$p_{(2)}\leq\gamma\alpha/(n-1)+(1-\gamma)\alpha/n$	$H_{(2)}$ is rejected
…	…
$p_{(n)}\leq\gamma\alpha+(1-\gamma)\alpha/n$	$H_{(n)}$ is rejected

Truncated Holm test

General form

Properties

Truncated Holm test is equivalent to Bonferroni test if $\gamma=0$ and regular Holm test if $\gamma=1$

Power of truncated Holm test is an increasing function of γ

Error rate function

$$e(I) = (\gamma + (1-\gamma)|I|/n)\alpha$$

Truncated Holm test is separable if $0 \le \gamma < 1$

Acute lung injury trial

Truncated Holm test

Family 1

Ordered p-values, $p_{(1)}<p_{(2)}$, and endpoints, $P_{(1)}$ and $P_{(2)}$

Truncated Holm test at $\alpha_1=\alpha$

Condition	Decision
$p_{(1)}\leq\gamma\alpha_1/2+(1-\gamma)\alpha_1/2$	$P_{(1)}$ is significant
$p_{(2)}\leq\gamma\alpha_1+(1-\gamma)\alpha_1/2$	$P_{(2)}$ is significant

Power in Family 1

An increasing function of γ

Acute lung injury trial

Truncated Holm test

Family 2

Any FWER-controlling multiple test at α_2

$\alpha_2=\alpha$ if two significant outcomes in Family 1

$\alpha_2=\alpha(1-\gamma)/2$ if one significant outcome in Family 1

$\alpha_2=0$ if no significant outcomes in Family 1

Power in Family 2

A decreasing function of γ if one significant outcome in Family 1

Does not depend on γ if two significant outcomes in Family 1

Acute lung injury trial

Truncated Holm test

Two-stage gatekeeping procedure

Familywise error rate, $\alpha=0.05$

Scenario 1

Family 1: Truncated Holm test with $\gamma=0$ (Bonferroni test)

Family 2: Hochberg test

Scenario 2

Family 1: Truncated Holm test with $\gamma=0.5$

Family 2: Hochberg test

Acute lung injury trial

Truncated Holm test

Scenario 1

Family 1: Truncated Holm test with γ=0 (Bonferroni test)

Family 2: Hochberg test

Endpoint	Raw p-value	α	Outcome
P_1	0.031	α_1=0.05	NS
P_2	0.013	α_1=0.05	S
S_1	0.039	α_2=0.025	NS
S_2	0.027	α_2=0.025	NS

Outcome: S (Significant at 0.05), NS (No significant at 0.05)

Acute lung injury trial

Truncated Holm test

Scenario 2

Family 1: Truncated Holm test with γ=0.5

Family 2: Hochberg test

Endpoint	Raw p-value	α	Outcome
P_1	0.031	$\alpha_1=0.05$	S
P_2	0.013	$\alpha_1=0.05$	S
S_1	0.039	$\alpha_2=0.05$	S
S_2	0.027	$\alpha_2=0.05$	S

Outcome: S (Significant at 0.05), NS (No significant at 0.05)

Multi-stage case

General case

Multiple families of hypotheses

Families of null hypotheses, $F_1,\ldots,F_m$

$F_i = \{H_{i1},\ldots,H_{in_i}\}$

Family i (i=1,…,m-1)

Separable FWER-controlling multiple test

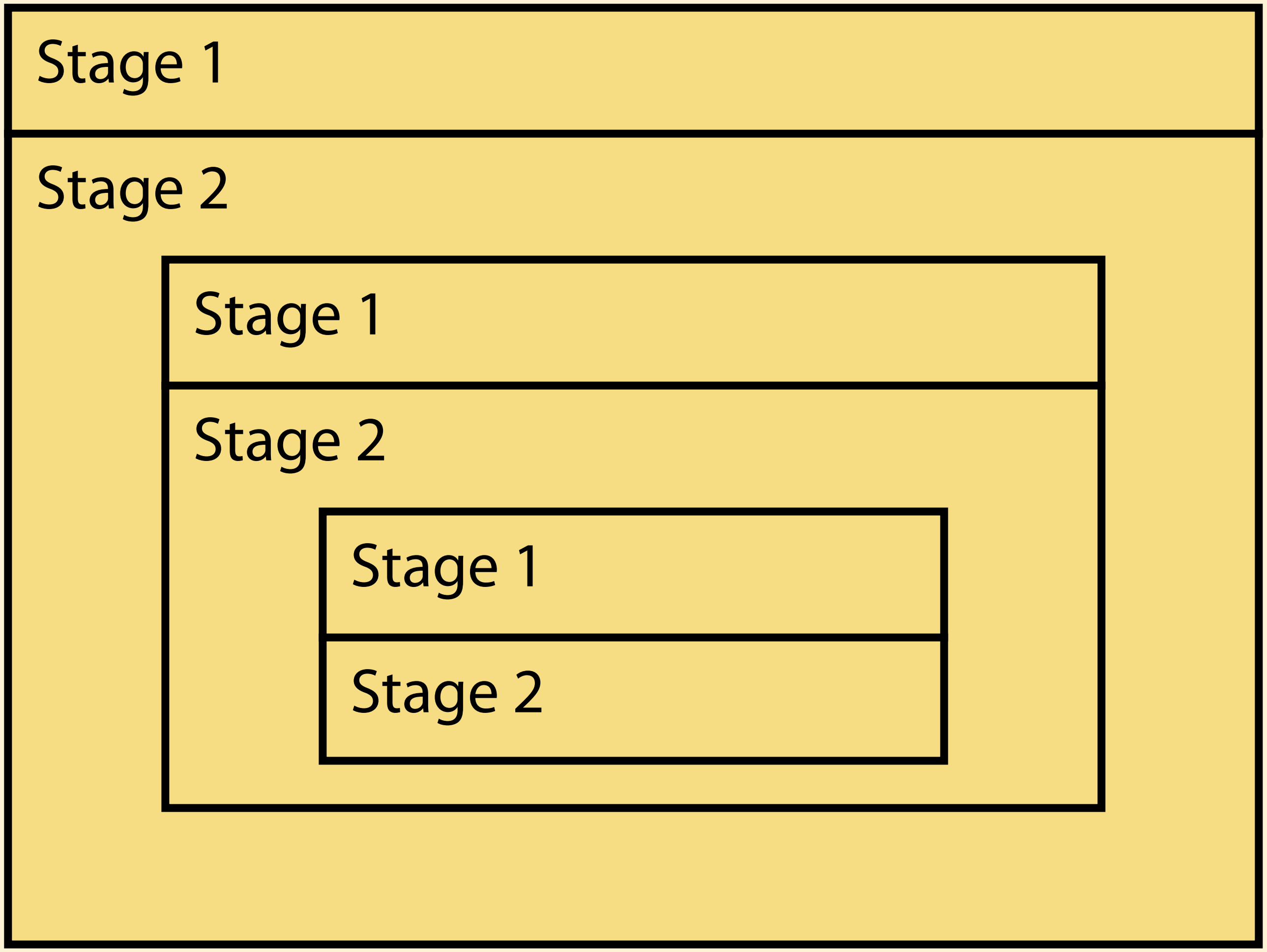
Error rate function $e_i(I),\ I \subseteq N_i = \{1,\ldots,n_i\}$

Family m

Any FWER-controlling multiple test

Multi-stage case

Recursive principle



Multi-stage case

General case

Start

Initialize $\alpha_1 = \alpha$

Family i (i=1,…,m-1)

Separable FWER-controlling multiple test at

$\alpha_i = \alpha_{i-1} - e_{i-1}(A_{i-1})$

A_{i-1} is the index set of accepted hypotheses in Family i-1

Family m

Any FWER-controlling multiple test at

$\alpha_m = \alpha_{m-1} - e_{m-1}(A_{m-1})$

Extensions

Account for importance of individual hypotheses

Procedures based on weighted tests (e.g., weighted Holm or Hochberg tests)

Logical restrictions

Account for logical restrictions among hypotheses

Example: Secondary analyses are restricted to doses at which the primary endpoint is significant

Summary

Gatekeeping testing strategies

Provide justification for including useful secondary findings in the product label

Control the familywise error rate

Multi-stage gatekeeping procedures

Gatekeeping procedures with a simple stepwise structure

Truncated multiple tests

Separable multiple tests that can be used in multi-stage gatekeeping procedures

References

Dmitrienko A, Offen WW, Westfall PH. (2003). Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. Statistics in Medicine. 22, 2387-2400.

Dmitrienko A, Tamhane AC, Wang X, Chen X. (2006). Stepwise gatekeeping procedures in clinical trial applications. Biometrical Journal. 48, 984-991.

Dmitrienko A, Tamhane A, Wiens B. (2007). General multi-stage gatekeeping procedures. Northwestern University. Department of Industrial Engineering and Management Sciences. Working Paper No. 07-06 [http://www.iems.northwestern.edu/content/Papers.asp].

Guilbaud O. (2007). Bonferroni parallel gatekeeping: Transparent generalizations and a short direct proof. Biometrical Journal. In press.