Methods: Mind the Gap

Webinar Series

# Alpha Spending for Clinical Trials


Presented by:

James F. Troendle, Ph.D.

NIH National Heart, Lung, and Blood Institute



NIH National Institutes of Health

Office of Disease Prevention

# Alpha Spending for Clinical Trials

James F. Troendle

Office of Biostatistics Research

*Division of Intramural Research*

*NHLBI, NIH*

Mind the Gap Webinar

September 7, 2023

# Presentation Outline

# Clinical Trial Context

## Multiple Primary Aims

- Multiple Primary Trial Aims
  - Multiple Primary Endpoints
  - Multi-Arm Trials ($\geq 3$)

Examples:

- Multiple Primary Endpoints in a 2 Arm Trial
  - Control of Systolic Blood Pressure
  - Control of Diastolic Blood Pressure
- 3 Arm Trial (A,B,C=control)
  - A vs. C
  - B vs. C
  - A vs. B

- Things not Addressed in This Presentation
  - Trials with FDA Oversight
  - Secondary Aims
  - Monitoring trials as data accumulate (alpha spending functions)

# Familywise Error

**Familywise Error (FWE)**
The familywise error of a collection of null hypotheses (family) is the probability that any true null hypothesis in the collection is rejected.

**Strong Control**
Strong control of the FWE means that the maximum FWE is $\leq \alpha$.

Note the maximum means that any configuration of true or false null hypotheses in the family must be considered.

**For This Presentation**
The family for a clinical trial are the null hypotheses corresponding to the primary aims.

# Familywise Error

**Weak Control**

Weak control of the FWE means that the FWE is $\leq \alpha$ when all null hypotheses are true.

Note strong control of FWE implies weak control.

**Other Errors**

False Discovery Rate (FDR). (1995 Benjamini and Hochberg, JRSS-B 57:289-300)

False Discovery Proportion (FDP) and Number of False Discoveries. (2004 Korn et al., JSPI 124:379-398)

k-FWER. (2005 Lehmann and Romano, Annals of Stat. 33:1138-1154)

# Familywise Error

**For This Presentation**

Methods will have strong control of the FWE (most stringent) over the family of primary aim hypotheses.

# Bonferroni

**Notation**

Let the desired significance level be denoted $\alpha$.

Let the null hypotheses be $H_1, H_2, \ldots, H_K$.

Let the corresponding p-values be $p_1, p_2, \ldots, p_K$.

**Bonferroni Procedure**

Reject each $H_j$ with $p_j \leq \alpha/K$.

Note the Bonferroni procedure strongly controls the FWE.

# Stepwise Extensions

**Notation**

Let the ordered null hypotheses be $H_{(1)}, H_{(2)}, \ldots, H_{(K)}$.

Let the corresponding ordered p-values be $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(K)}$.

**Holm Procedure**

Sequentially reject $H_{(j)}$ if $p_{(j)} \leq \alpha/(K + 1 - j)$ and all previous $H_{(j)}$ have been rejected.

(1979 Holm, Scand. J. Stat. 6:65-70)

Note the Holm procedure strongly controls the FWE (intuitively either a true hypothesis is already rejected or else there are one fewer true hypotheses to consider after each rejection).

# Stepwise Extensions

**Holm Procedure With Logical Restrictions**
Sequentially reject $H_{(j)}$ if $p_{(j)} \leq \alpha/R_j$ and all previous $H_{(j)}$ have been rejected.

$R_j$ = max # untested true nulls at step $j$

(1986 Shaffer, JASA 81:826-831)

Note the Holm procedure with logical restrictions strongly controls the FWE (uses same reasoning).

# Stepwise Extensions

**Step-down with Resampling**

Many variations on Holm's procedure exist, including step-up methods that are sequentially accepting (Holm's is sequentially rejecting) and methods that exploit correlation, but they all require some assumption.

Most useful is perhaps Westfall and Young's method that exploits correlation through resampling (**PROC MULTTEST** in SAS, package **NRejections** in R).

(1993 Westfall and Young, Resampling-Based Multiple Testing)

Note that for multiple outcomes being compared between two groups, the necessary assumption (subset-pivotality) is automatically met so strongly controls the FWE asymptotically.

# Serial Gatekeeper

Order hypotheses according to a pre-trial specified order.

**Fixed Sequence**
Sequentially reject $H_j$ if $p_j \leq \alpha$ and all previous $H_j$ have been rejected.

(1991 Bauer, Stat. in Med. 10:871-890)

Note the Fixed Sequence procedure strongly controls the FWE.

# Serial Gatekeeper

Order families of hypotheses $F_1, \ldots, F_Q$ according to a pre-trial specified order.

**Serial Gatekeeper**
Sequentially test family $F_j$ if all previous families have been exhaustively rejected.

Note the Serial Gatekeeper procedure will strongly control the FWE if FWE is controlled within each family.

# Multistage Gatekeeper

Order families of hypotheses $F_1, \ldots, F_Q$ according to a pre-trial specified order.

**Multistage Gatekeeper**
Sequentially test family $F_j$ if all previous families have had >= 1 hypothesis rejected.

Note the Multistage Gatekeeper procedure will not in general control FWE.

# Multistage Gatekeeper

Consider two families of hypotheses $F_1$, $F_2$ with $K$ hypotheses in $F_1$.

**Example Multistage Gatekeeper**

Test family $F_1$ at FWE level $\alpha$ using Bonferroni.

Test family $F_2$ at FWE level $\alpha^*$ using Holm's procedure where

$\alpha^* = \alpha R/K$ and

$R$ = # rejected hypotheses in $F_1$.

(2008 Dimitrienko, Tamhane, and Wiens, Biometrical J. 50:667-677)

Note this Multistage Gatekeeper procedure does strongly control FWE.

# Graphical Approach

Consider null hypotheses $H_1, H_2, \ldots, H_K$.

With corresponding p-values $p_1, p_2, \ldots, p_K$.

Pre-trial, determine initial allocations $\alpha_1, \alpha_2, \ldots, \alpha_K$ with $\alpha_1 + \ldots + \alpha_K <= \alpha$.

Pre-trial, determine for each $H_j$ how $\alpha_j$ will be distributed if $H_j$ is rejected.

**This can be described graphically**

# Graphical Approach

**Graphical Procedure** (with $\Sigma\, \alpha_j = \alpha$)
Reject any $H_j$ if $p_j \leq \alpha_j$.

If any hypotheses rejected, distribute $\alpha_j$ and re-test all hypotheses.

(2009 Bretz et al. Stat. in Med. 28:586-604)

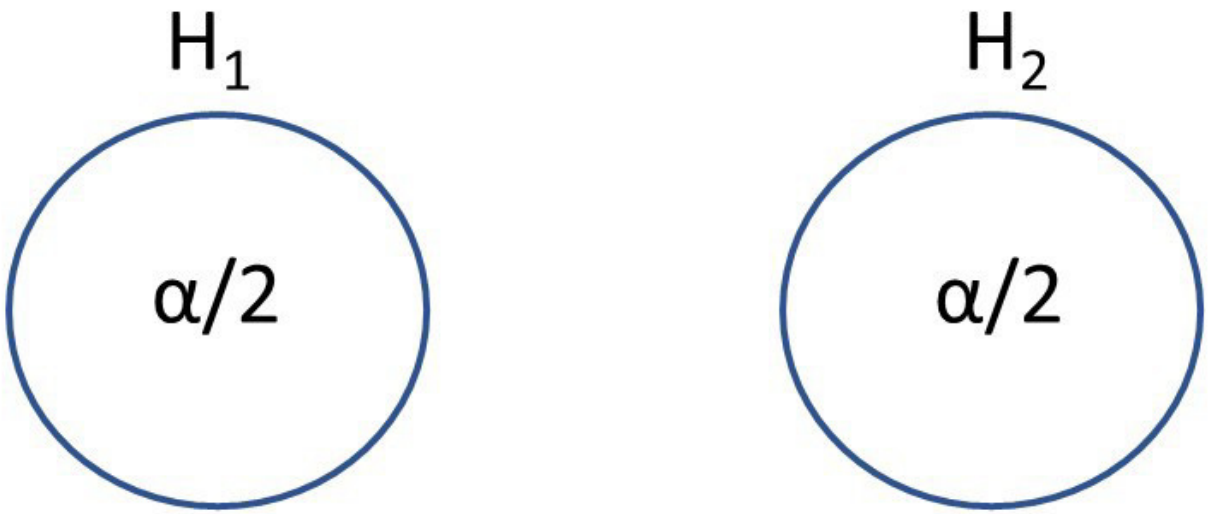
Note any such graphical procedure strongly controls the FWE at $\Sigma\, \alpha_j = \alpha$.

# Toy Examples

- Multiple Primary Endpoints in a 2 Arm Trial
  - Control of Systolic Blood Pressure $H_1$ : no treatment effect on SBP
  - Control of Diastolic Blood Pressure $H_2$ : no treatment effect on DBP
- 3 Arm Trial (A,B,C=control)
  - A vs. C $H_1$ : no A vs. C difference
  - B vs. C $H_2$ : no B vs. C difference
  - A vs. B $H_3$ : no A vs. B difference

# Multiple Primary Endpoints

## Bonferroni



# Multiple Primary Endpoints

## Holm

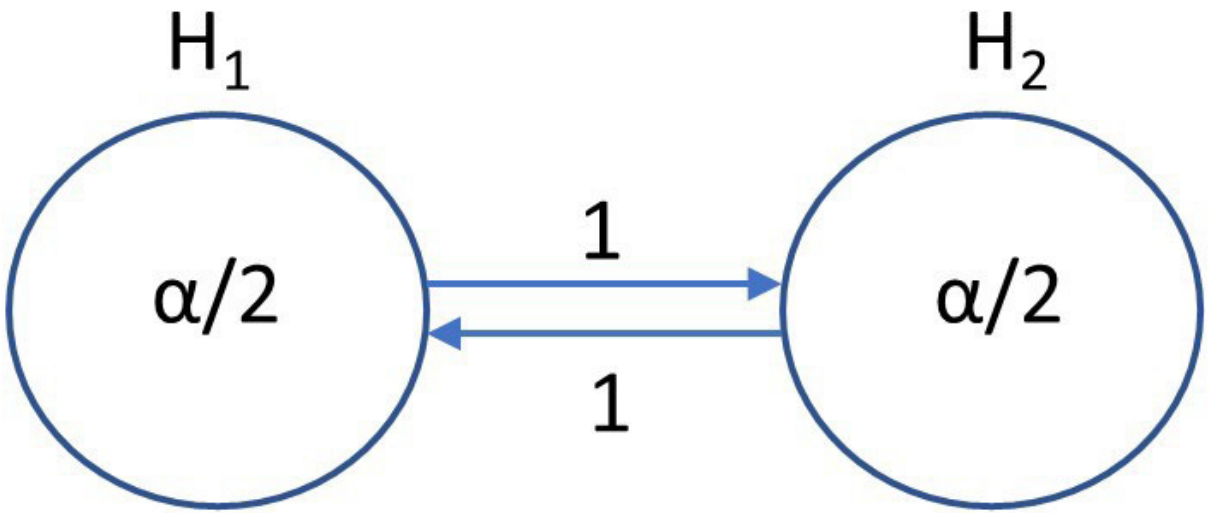
**Holm**

Reject $H_{(1)}$ if $p_{(1)} \leq \alpha/2$.

Reject $H_{(2)}$ if $p_{(2)} \leq \alpha$ and $H_{(1)}$ was rejected.

# Multiple Primary Endpoints

## Holm



# Multiple Primary Endpoints

## Fixed Sequence

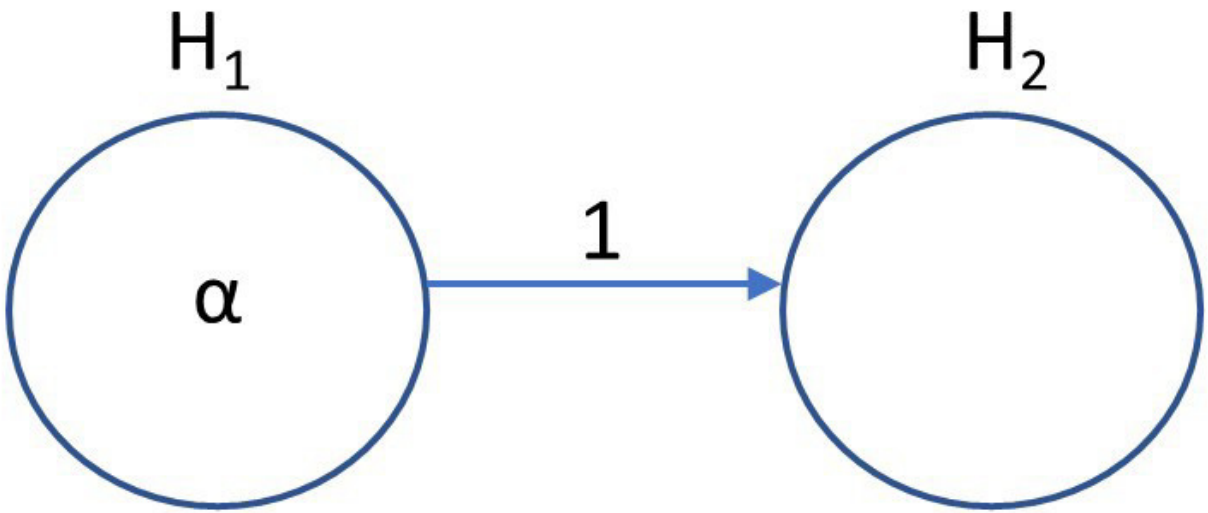
**Fixed Sequence**

Reject $H_1$ if $p_1 \leq \alpha$.

Reject $H_2$ if $p_2 \leq \alpha$ and $H_1$ was rejected.

# Multiple Primary Endpoints

## Fixed Sequence



# 3 Arm Trial

## Holm

- 3 Arm Trial (A,B,C=control)
  - A vs. C $H_1$ : no A vs. C difference
  - B vs. C $H_2$ : no B vs. C difference
  - A vs. B $H_3$ : no A vs. B difference

**Holm**

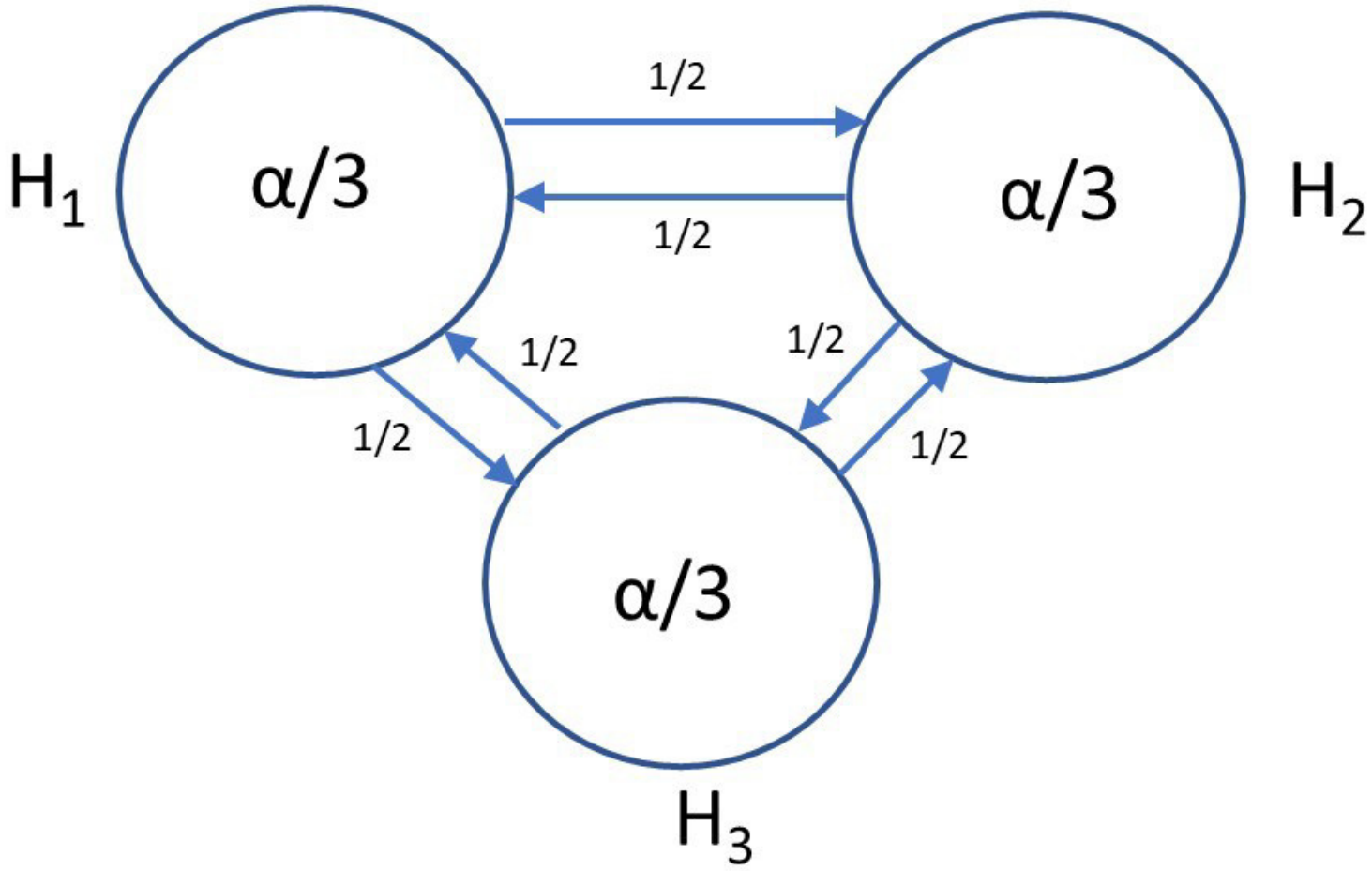
Reject $H_{(1)}$ if $p_{(1)} \leq \alpha/3$.

Reject $H_{(2)}$ if $p_{(2)} \leq \alpha/2$ and $H_{(1)}$ was rejected.

Reject $H_{(3)}$ if $p_{(3)} \leq \alpha$ and $H_{(1)}$ and $H_{(2)}$ were rejected.

# 3 Arm Trial

## Holm



# 3 Arm Trial

## Holm With Logical Restriction

**Holm With Logical Restriction**

Reject $H_{(1)}$ if $p_{(1)} \leq \alpha/3$.

Reject $H_{(2)}$ if $p_{(2)} \leq \alpha$ and $H_{(1)}$ was rejected.

Reject $H_{(3)}$ if $p_{(3)} \leq \alpha$ and $H_{(1)}$ and $H_{(2)}$ were rejected.

Note: **Not able to be described graphically!**

# 3 Arm Trial

## Fixed Sequence

**Fixed Sequence**

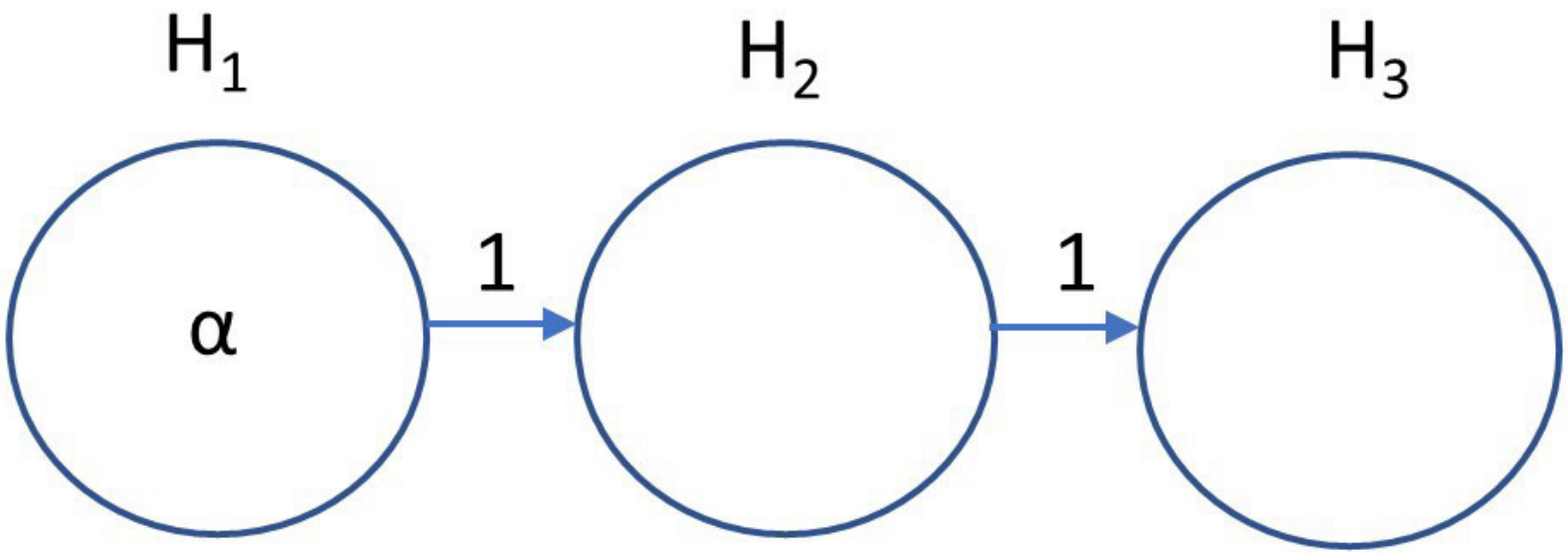
Reject $H_1$ if $p_1 \leq \alpha$.

Reject $H_2$ if $p_2 \leq \alpha$ and $H_1$ was rejected.

Reject $H_3$ if $p_3 \leq \alpha$ and $H_1$ and $H_2$ were rejected.

# 3 Arm Trial

## Fixed Sequence



# 3 Arm Trial

## Multistage Gatekeeper

Family 1 = $\{H_1, H_2\}$ Family 2 = $\{H_3\}$

**Multistage Gatekeeper**

**Stage 1:**

Reject $H_1$ if $p_1 \leq \alpha/2$.

Reject $H_2$ if $p_2 \leq \alpha/2$.

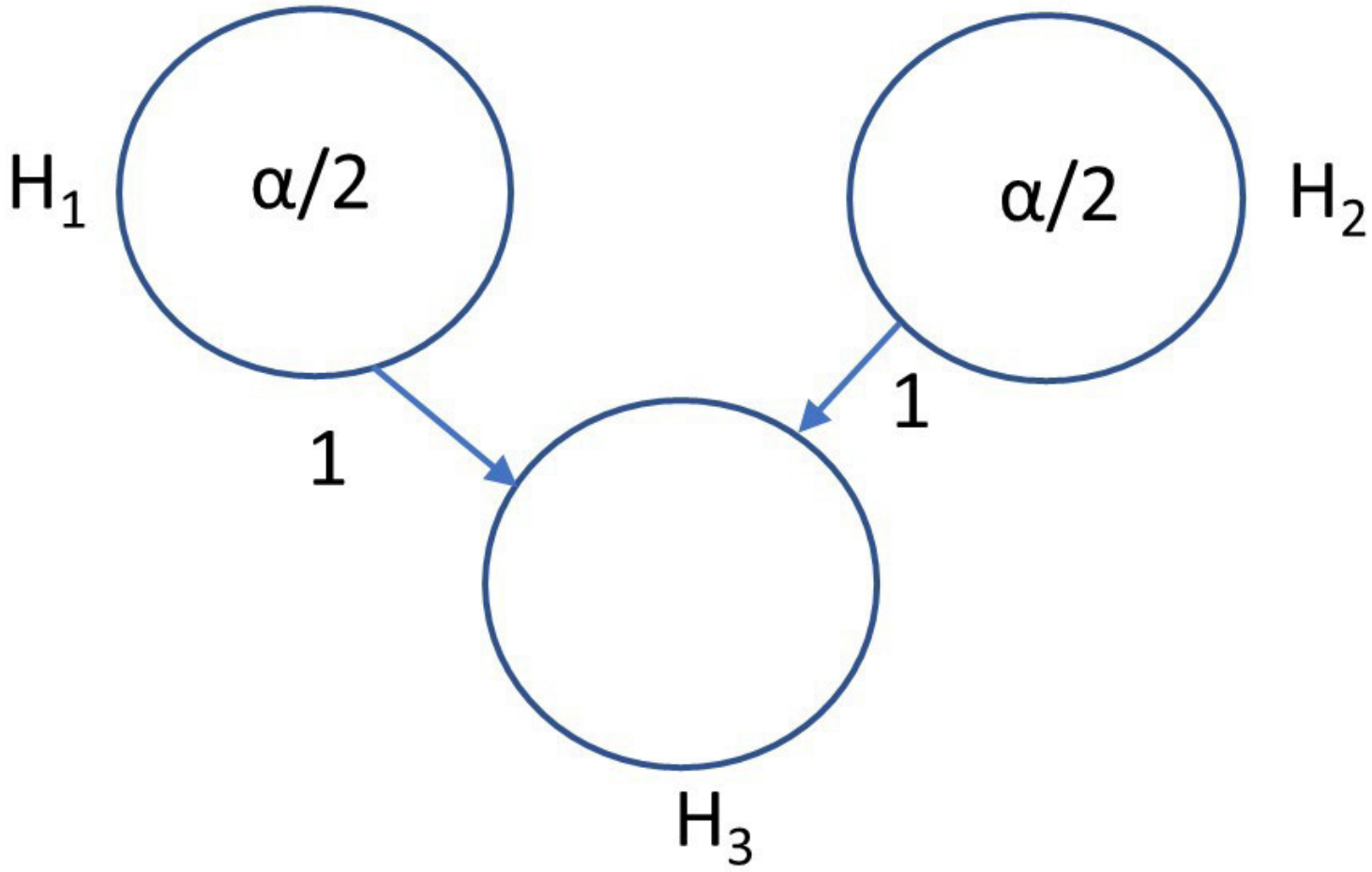
**Stage 2:**

Reject $H_3$ if $p_3 \leq R\alpha/2$.

$R$ = # rejected hypotheses in $F_1$.

# 3 Arm Trial

## Multistage Gatekeeper



# Nudge Trial

## NHLBI sponsored Nudge Trial

Pragmatic individually randomized trial to test effect of "nudges" on medication adherence in 3 health systems in Colorado.

Four arms (A=Generic Nudge, B=Optimized Nudge, C=Optimized Nudge + Chatbot, D=Usual Care).

Outcome is proportion of days covered (medication supply) over 12 months as determined by prescription refill information in pharmacy records.

# Nudge Trial

## Hypotheses

First Family

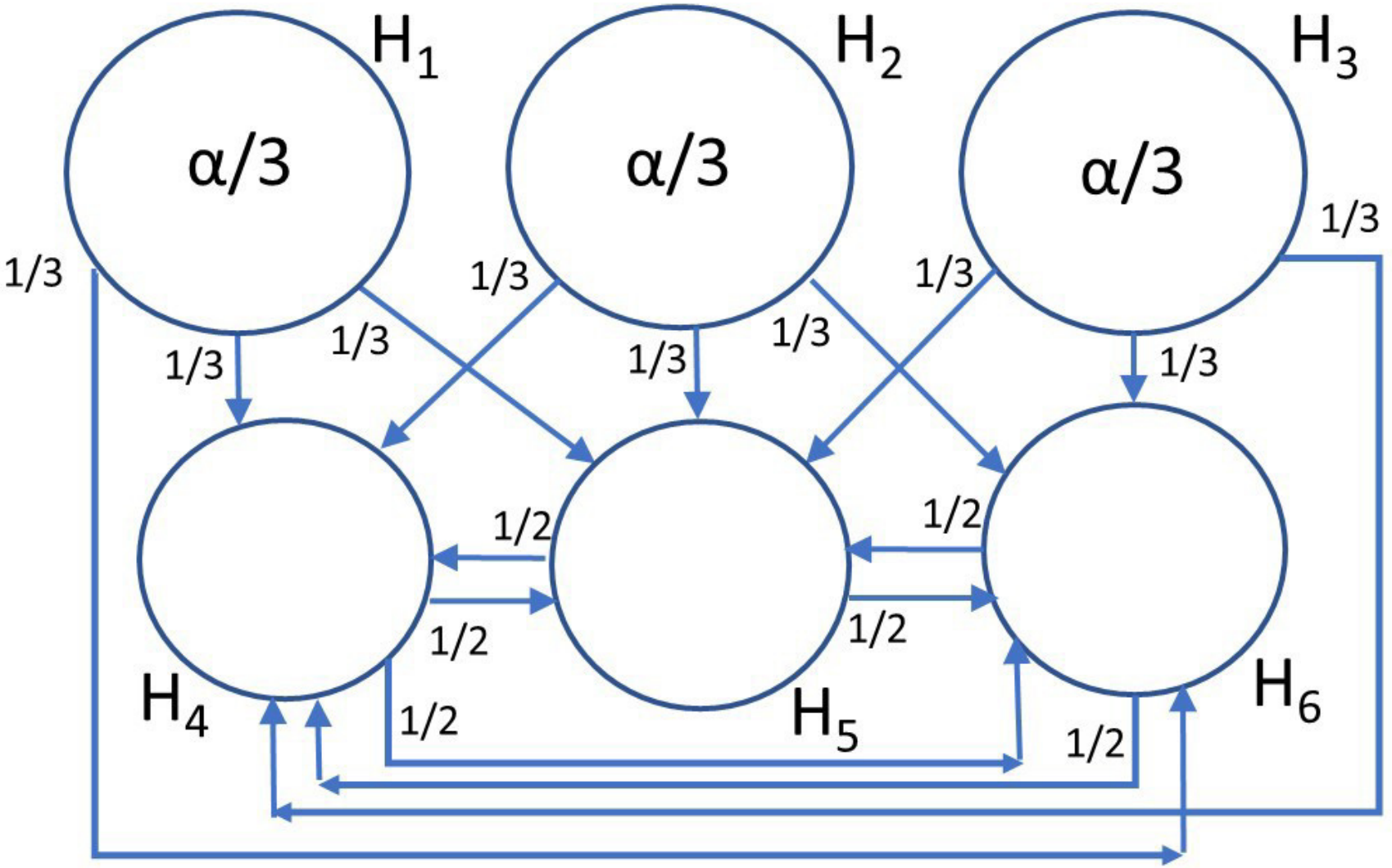
- Hypothesis 1: A vs. D
- Hypothesis 2: B vs. D
- Hypothesis 3: C vs. D

Second Family

- Hypothesis 4: A vs. B
- Hypothesis 5: A vs. C
- Hypothesis 6: B vs. C

# Nudge Trial

## Multistage Gatekeeper



# BMT CTN 0702 Trial

## NHLBI sponsored BMT Trial

Three arm individually randomized trial of 758 patients with symptomatic multiple myeloma.

Tested (A) tandem AHCT, (B) AHCT+RVD , (C) AHCT (all arms with len).

AHCT: autologous hematopoietic cell transplantation

RVD: Revlimid + Velcade + dexamethasone

len: lenalidomide maintenance

Three arms (A=tandem AHCT, B=AHCT+RVD, C=AHCT).

Outcome was progression free survival (PFS) at 38 months.

# BMT CTN 0702 Trial

## NHLBI sponsored BMT Trial

Bonferroni used to test all 3 pairwise comparisons ($H_1$,$H_2$,$H_3$).

PFS in Arm A: 58.5%
PFS in Arm B: 57.8%
PFS in Arm C: 53.9%

None significantly different from any others with $\alpha = 0.0167$.

Other options: Multistage gatekeeper or just test $H_1$ and $H_2$.

Holm with logical restrictions on ($H_1$,$H_2$,$H_3$).

Resampling methods (Westfall and Young 1993) would provide more power.

# Conclusions

- Simple graphical approach can be used to efficiently test multiple primary hypotheses
- Each hypothesis given initial alpha allocation
- Each hypothesis, if rejected, can recycle their alpha to other hypotheses
- Graphical approach strongly controls FWE at sum of initial alphas
- Multistage Gatekeeper (graphical) to prioritize initial family
- Logical restrictions (not graphical) can be helpful
- Resampling methods (not graphical) can improve power (especially when test statistics are highly correlated)

# References

Benjamini Y. and Hochberg Y. (1995) JRSS-B 57:289-300.
Korn E. et al. (2004) JSPI 124:379-398.
Lehmann EL. and Romano JP. (2005) Annals of Stat. 33:1138-1154.
Holm S. (1979) Scand. J. Stat. 6:65-70.
Shaffer JP. (1986) JASA 81:826-831.
Westfall PH. and Young SS. (1993) Resampling-Based Multiple Testing, Wiley-interscience.
Bauer P. (1991) Stat. in Med. 10:871-890.
Dimitrienko A. et al. (2008) Biometrical J. 50:667-677.
Bretz F. et al. (2009) Stat. in Med. 28:586-604.