



# **Across the Multiverse... Statistically Speaking**

Addressing Multiplicity in Clinical Trials

23-Mar-2026



---

# Background

- Research studies are expensive; therefore, one can easily argue that it's both costly and inefficient to design a research study to answer only one question.
- Consequently, research studies are designed to obtain sufficient data to evaluate several questions.
- Evaluating several questions may result in problems.

---

# Type I Error

- Type I error is the probability of rejecting the null hypothesis when the null hypothesis is true.
- i.e. The probability of observing a “positive” result in the study when one does not exist in the overall population
- Also referred to as a “false positive”
- Type I error is often of importance, as one does not want to falsely reject the null hypothesis.

# Type I Error

- Probability of committing a type I error =  $1 - (1 - \alpha)^k$ 
  - $\alpha$  = Significance level for a single test
  - $k$  = Number of independent tests or comparisons performed
- Type I error often increases when multiple statistical tests are performed.

---

# The Potential Problem: Example

- Consider a clinical trial where 2 primary endpoints are of interest.
- Further, assume that both endpoints are independent and each is evaluated at an 0.05 level.
- The overall type I error will be calculated for the evaluation of both endpoints.

# The Potential Problem: Example

- Overall type I error rate =  $1 - (1 - \alpha)^k$   
=  $1 - (1 - 0.05)^2$   
= 0.0975
- Though both endpoints were evaluated at an 0.05 level, the overall type I error rate after evaluating both hypotheses is 9.75%, nearly double the planned type I error rate.

---

# What is Multiplicity?

- Multiplicity is the increased probability of committing a type I error when several statistical tests are performed.
- When many tests are performed, it's possible that at least one may be considered significant due to chance.
- This can lead to misleading conclusions.
  - Researchers may falsely conclude that a treatment is effective or a factor is an important risk factor for a disease.



**What can be done to address multiplicity?**

---

---

# Addressing Multiplicity

- Multiplicity may be addressed by controlling one of the following:
  - Family-wise error rate (FWER)
  - False discovery rate (FDR)

Cytel



# **Family-wise Error Rate (FWER)**

---

# Family-Wise Error Rate (FWER)

- FWER is the probability of making at least one type I error when conducting multiple hypothesis tests within a set or group of hypotheses, known as a “family”.
- Groups may be based on similarities between endpoints, clinical importance, etc.
- FWER methods adjust significance level used to evaluate each endpoint.

---

# Family-Wise Error Rate (FWER)

- There are several methods to control FWER.
- The following will be discussed in this presentation:
  - Bonferroni adjustment
  - Holm adjustment
  - Hochberg adjustment

01

# Bonferroni Adjustment

---



# Bonferroni Adjustment

- All endpoints evaluated using the same critical value.
- Critical value =  $\frac{FWER}{\text{Number of } p\text{-values}}$
- P-values lower than the critical value are considered statistically significant.
- Conservative approach

---

# Bonferroni Adjustment: Example

- Assume that 5 endpoints were evaluated.
- Resulting p-values: 0.0592, 0.0140, 0.0160, 0.0080, 0.0352.
- FWER = 0.05.

# Bonferroni Adjustment: Example

$$\text{Critical Value} = \frac{FWER}{\text{Number of } P\text{-values}} = \frac{0.05}{5}$$

| Result | P-value | Critical Value | Significant? |
|--------|---------|----------------|--------------|
| 1      | 0.0592  | 0.01           | ?            |
| 2      | 0.0140  | 0.01           | ?            |
| 3      | 0.0160  | 0.01           | ?            |
| 4      | 0.0080  | 0.01           | ?            |
| 5      | 0.0352  | 0.01           | ?            |

# Bonferroni Adjustment: Example

Compare each p-value to the corresponding critical value

| Result | P-value | Critical Value | Significant? |
|--------|---------|----------------|--------------|
| 1      | 0.0592  | 0.01           | ?            |
| 2      | 0.0140  | 0.01           | ?            |
| 3      | 0.0160  | 0.01           | ?            |
| 4      | 0.0080  | 0.01           | ?            |
| 5      | 0.0352  | 0.01           | ?            |

# Bonferroni Adjustment: Example

Significant if p-value < critical value

| Result | P-value | Critical Value | Significant? |
|--------|---------|----------------|--------------|
| 1      | 0.0592  | 0.01           | No           |
| 2      | 0.0140  | 0.01           | No           |
| 3      | 0.0160  | 0.01           | No           |
| 4      | 0.0080  | 0.01           | Yes          |
| 5      | 0.0352  | 0.01           | No           |

# 02



## Holm Adjustment

---

# Holm Adjustment

- P-values are arranged in ascending order.
- A rank is assigned to each p-value.
  - Rank 1 = smallest p-value, Rank 2 = second smallest p-value, etc.
- A critical value is calculated for each p-value, based on the rank of the p-value.
- Critical value = 
$$\frac{FWER}{(\text{Number of } p\text{-values}) + 1 - (\text{Rank of } p\text{-value})}$$

---

# Holm Adjustment

- P-values are evaluated sequentially.
  - Begin with the p-value ranked 1.
  - Identify the first instance where the p-value is greater than the critical value.
- All p-values of rank lower than the last evaluated p-value are considered statistically significant.

---

# Holm Adjustment: Example

- Assume that 5 endpoints were evaluated.
- Resulting p-values: 0.0592, 0.0140, 0.0160, 0.0080, 0.0352.
- FWER = 0.05.

# Holm Adjustment: Example

P-values are arranged in **ascending order**

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  |                | ?            |
| 2    | 0.0140  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

Rank the p-values

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  |                | ?            |
| 2    | 0.0140  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

Begin with the first p-value

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  |                | ?            |
| 2    | 0.0140  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

$$\text{Critical Value} = \frac{FWER}{(\text{Number of } p\text{-values}) + 1 - (\text{Rank of } p\text{-value})} = \frac{0.05}{5 + 1 - 1}$$

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  | 0.0100         | ?            |
| 2    | 0.0140  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

Significant if p-value < critical value

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  | 0.0100         | Yes          |
| 2    | 0.0140  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

Continue until first non-statistically significant result

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0080  | 0.0100         | Yes          |
| 2    | 0.0140  | 0.0125         | No           |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0352  |                | ?            |
| 5    | 0.0592  |                | ?            |

# Holm Adjustment: Example

P-values prior to last evaluated p-value are statistically significant

| Rank | P-value | Critical Value | Significant?  |
|------|---------|----------------|---------------|
| 1    | 0.0080  | 0.01           | Yes           |
| 2    | 0.0140  | 0.01           | No            |
| 3    | 0.0160  | 0.01           | Not evaluated |
| 4    | 0.0352  | 0.01           | Not evaluated |
| 5    | 0.0592  | 0.01           | Not evaluated |



03

## Hochberg Adjustment

---



# Hochberg Adjustment

- P-values are arranged in descending order.
- A rank is assigned to each p-value.
  - Rank 1 = largest p-value, Rank 2 = second largest p-value, etc.
- A critical value is calculated for each p-value, based on the rank of the p-value.
  - Critical value =  $\frac{FWER}{Rank\ of\ p-value}$

---

# Hochberg Adjustment

- P-values are evaluated sequentially.
  - Begin with the p-value ranked 1.
  - Identify the first instance where the p-value is less than the corresponding critical value.
- All p-values of rank greater than or equal to the last evaluated p-value are considered statistically significant.

---

# Hochberg Adjustment: Example

- Assume that 5 endpoints were evaluated.
- Resulting p-values: 0.0592, 0.0140, 0.0160, 0.0080, 0.0352.
- FWER = 0.05.

# Hochberg Adjustment: Example

P-values are arranged in **descending order**

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  |                | ?            |
| 2    | 0.0352  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Rank the p-values

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  |                | ?            |
| 2    | 0.0352  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Begin with the first p-value

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  |                | ?            |
| 2    | 0.0352  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

$$\text{Critical Value} = \frac{FWER}{\text{Rank of } p\text{-value}} = \frac{0.05}{1}$$

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  | 0.0500         | ?            |
| 2    | 0.0352  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Significant if p-value < critical value

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  | 0.0500         | No           |
| 2    | 0.0352  |                | ?            |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Continue until first statistically significant result

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  | 0.0100         | No           |
| 2    | 0.0352  | 0.2500         | No           |
| 3    | 0.0160  |                | ?            |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Continue until first statistically significant result

| Rank | P-value | Critical Value | Significant? |
|------|---------|----------------|--------------|
| 1    | 0.0592  | 0.0100         | No           |
| 2    | 0.0352  | 0.2500         | No           |
| 3    | 0.0160  | 0.0167         | Yes          |
| 4    | 0.0140  |                | ?            |
| 5    | 0.0080  |                | ?            |

# Hochberg Adjustment: Example

Last evaluated p-value and all remaining are statistically significant

| Rank | P-value | Critical Value | Significant?  |
|------|---------|----------------|---------------|
| 1    | 0.0592  | 0.01           | No            |
| 2    | 0.0352  | 0.01           | No            |
| 3    | 0.0160  | 0.01           | Yes           |
| 4    | 0.0140  | 0.01           | Not evaluated |
| 5    | 0.0080  | 0.01           | Not evaluated |

# Comparing the FWER Approaches

| P-value                    | Bonferroni | Holm | Hochberg |
|----------------------------|------------|------|----------|
| 0.0080                     | ✓          | ✓    | ✓        |
| 0.0140                     | -          | -    | ✓        |
| 0.0160                     | -          | -    | ✓        |
| 0.0352                     | -          | -    | -        |
| 0.0592                     | -          | -    | -        |
| # of P-values<br>Evaluated | 5          | 2    | 3        |

Cytel



# **False Discovery Rate (FDR)**

---

# False Discovery Rate (FDR)

- Controls the proportion of false positive findings among all rejected hypotheses
- Overall type I error may be greater than the FWER method
- Less strict and more powerful than the FWER method
- Accept false positive findings to identify potentially important findings

---

# When to Consider FDR

- Interested in identifying potentially important findings
- When performing large-scale, multiple hypothesis testing
  - Genetics, drug discovery, marketing experiments, exploratory analyses
- When comfortable accepting some false positive results in exchange for potentially identifying more true positive results

# Choosing the FDR Value

- Higher FDR is preferred when interest in identifying more potentially true positive results outweighs fear of observing false positive results.
  - Ex. FDR = 0.10 or 0.20
- A more conservative FDR should be chosen when interested in limiting the number of false positive results.
  - Ex. FDR = 0.05



# Benjamini-Hochberg Method

---

# Benjamini-Hochberg Method

- P-values are arranged in ascending order.
- A rank is assigned to each p-value.
  - Rank 1 = smallest p-value, Rank 2 = second smallest p-value, etc.
- A critical value is calculated for each p-value, based on the rank of the p-value.
- Critical value = 
$$\frac{FDR \times (\text{Rank of } p\text{-value})}{\text{Number of } p\text{-values evaluated}}$$

---

# Benjamini-Hochberg Method

- Identify the largest p-value that is less than or equal to its corresponding critical value.
- All p-values of rank less than or equal to the rank of the chosen p-value will be considered statistically significant.
- Some p-values considered statistically significant may be larger than their corresponding critical value.

# Benjamini-Hochberg Method: Example

- Assume that 20 endpoints were evaluated. The resulting p-values are

|        |        |        |        |
|--------|--------|--------|--------|
| 0.0029 | 0.0532 | 0.1445 | 0.2803 |
| 0.0186 | 0.0560 | 0.1536 | 0.3037 |
| 0.0214 | 0.0818 | 0.1985 | 0.3267 |
| 0.0278 | 0.0843 | 0.2536 | 0.3965 |
| 0.0295 | 0.0997 | 0.2536 | 0.4999 |

- FDR = 0.2

# Benjamini-Hochberg Method: Example

Arrange p-values in **ascending order**

| Rank | P-Value | Rank | P-Value | Rank | P-Value |
|------|---------|------|---------|------|---------|
| 1    | 0.0029  | 8    | 0.0818  | 15   | 0.2536  |
| 2    | 0.0186  | 9    | 0.0843  | 16   | 0.2803  |
| 3    | 0.0214  | 10   | 0.0997  | 17   | 0.3037  |
| 4    | 0.0278  | 11   | 0.1445  | 18   | 0.3267  |
| 5    | 0.0295  | 12   | 0.1536  | 19   | 0.3965  |
| 6    | 0.0532  | 13   | 0.1985  | 20   | 0.4999  |
| 7    | 0.0560  | 14   | 0.2536  |      |         |

# Benjamini-Hochberg Method: Example

Rank the p-values

| Rank | P-Value | Rank | P-Value | Rank | P-Value |
|------|---------|------|---------|------|---------|
| 1    | 0.0029  | 8    | 0.0818  | 15   | 0.2536  |
| 2    | 0.0186  | 9    | 0.0843  | 16   | 0.2803  |
| 3    | 0.0214  | 10   | 0.0997  | 17   | 0.3037  |
| 4    | 0.0278  | 11   | 0.1445  | 18   | 0.3267  |
| 5    | 0.0295  | 12   | 0.1536  | 19   | 0.3965  |
| 6    | 0.0532  | 13   | 0.1985  | 20   | 0.4999  |
| 7    | 0.0560  | 14   | 0.2536  |      |         |

# Benjamini-Hochberg Method: Example

Calculate the critical values

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | ?              | 0.0818  | ?              | 0.2536  | ?              |
| 0.0186  | ?              | 0.0843  | ?              | 0.2803  | ?              |
| 0.0214  | ?              | 0.0997  | ?              | 0.3037  | ?              |
| 0.0278  | ?              | 0.1445  | ?              | 0.3267  | ?              |
| 0.0295  | ?              | 0.1536  | ?              | 0.3965  | ?              |
| 0.0532  | ?              | 0.1985  | ?              | 0.4999  | ?              |
| 0.0560  | ?              | 0.2536  | ?              |         |                |

# Benjamini-Hochberg Method: Example

$$\text{Critical value} = \frac{FDR \times (\text{Rank of } p\text{-value})}{\text{Number of } p\text{-values evaluated}} = \frac{0.2 \times (\text{Rank of } p\text{-value})}{20}$$

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | 0.01           | 0.0818  | 0.08           | 0.2536  | 0.15           |
| 0.0186  | 0.02           | 0.0843  | 0.09           | 0.2803  | 0.16           |
| 0.0214  | 0.03           | 0.0997  | 0.10           | 0.3037  | 0.17           |
| 0.0278  | 0.04           | 0.1445  | 0.11           | 0.3267  | 0.18           |
| 0.0295  | 0.05           | 0.1536  | 0.12           | 0.3965  | 0.19           |
| 0.0532  | 0.06           | 0.1985  | 0.13           | 0.4999  | 0.20           |
| 0.0560  | 0.07           | 0.2536  | 0.14           |         |                |

# Benjamini-Hochberg Method: Example

Identify all p-values less than the corresponding critical value

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | 0.01           | 0.0818  | 0.08           | 0.2536  | 0.15           |
| 0.0186  | 0.02           | 0.0843  | 0.09           | 0.2803  | 0.16           |
| 0.0214  | 0.03           | 0.0997  | 0.10           | 0.3037  | 0.17           |
| 0.0278  | 0.04           | 0.1445  | 0.11           | 0.3267  | 0.18           |
| 0.0295  | 0.05           | 0.1536  | 0.12           | 0.3965  | 0.19           |
| 0.0532  | 0.06           | 0.1985  | 0.13           | 0.4999  | 0.20           |
| 0.0560  | 0.07           | 0.2536  | 0.14           |         |                |

# Benjamini-Hochberg Method: Example

Identify largest p-value that is less than the corresponding p-value

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | 0.01           | 0.0818  | 0.08           | 0.2536  | 0.15           |
| 0.0186  | 0.02           | 0.0843  | 0.09           | 0.2803  | 0.16           |
| 0.0214  | 0.03           | 0.0997  | 0.10           | 0.3037  | 0.17           |
| 0.0278  | 0.04           | 0.1445  | 0.11           | 0.3267  | 0.18           |
| 0.0295  | 0.05           | 0.1536  | 0.12           | 0.3965  | 0.19           |
| 0.0532  | 0.06           | 0.1985  | 0.13           | 0.4999  | 0.20           |
| 0.0560  | 0.07           | 0.2536  | 0.14           |         |                |

# Benjamini-Hochberg Method: Example

All p-values less than or equal to this p-value are considered statistically significant

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | 0.01           | 0.0818  | 0.08           | 0.2536  | 0.15           |
| 0.0186  | 0.02           | 0.0843  | 0.09           | 0.2803  | 0.16           |
| 0.0214  | 0.03           | 0.0997  | 0.10           | 0.3037  | 0.17           |
| 0.0278  | 0.04           | 0.1445  | 0.11           | 0.3267  | 0.18           |
| 0.0295  | 0.05           | 0.1536  | 0.12           | 0.3965  | 0.19           |
| 0.0532  | 0.06           | 0.1985  | 0.13           | 0.4999  | 0.20           |
| 0.0560  | 0.07           | 0.2536  | 0.14           |         |                |

# Benjamini-Hochberg Method: Example

| P-value | Critical Value | P-Value | Critical Value | P-value | Critical Value |
|---------|----------------|---------|----------------|---------|----------------|
| 0.0029  | 0.01           | 0.0818  | 0.08           | 0.2536  | 0.15           |
| 0.0186  | 0.02           | 0.0843  | 0.09           | 0.2803  | 0.16           |
| 0.0214  | 0.03           | 0.0997  | 0.10           | 0.3037  | 0.17           |
| 0.0278  | 0.04           | 0.1445  | 0.11           | 0.3267  | 0.18           |
| 0.0295  | 0.05           | 0.1536  | 0.12           | 0.3965  | 0.19           |
| 0.0532  | 0.06           | 0.1985  | 0.13           | 0.4999  | 0.20           |
| 0.0560  | 0.07           | 0.2536  | 0.14           |         |                |



Cytel



## Conclusions

---

# Conclusion

- Multiplicity may occur when multiple endpoints are evaluated.
- FWER and FDR may be used to account for multiplicity.
- FWER methods may be used to control the type I error rate.
- FDR may be used to identify potentially important results, though some false positive results may also be accepted.

# Authors



**Cralen Davis**  
**Cralen.Davis@Cytel.com**



**Trevano Dean**  
**Trevano.Dean@Cytel.com**

Cytel

**Thank You!**