Multiple testing in a Partitioning framework

Pär Karlsson
AstraZeneca, Sweden
BASS XVII 2010-11-10

Abstract

- Multiple testing procedures are typically described as stepwise procedures. Recently several graphical descriptions have been suggested. These approaches are suitable for describing a multiplicity procedure, but not necessary for proving that the procedure controls the family wise error rate. However, the fundamentals behind multiple testing are simple if viewed in a partitioning framework. In this framework you divide the set of hypotheses into new sets of disjoint hypotheses. If each of the disjoint hypotheses is tested at a specific level, alpha, and an original hypothesis is rejected if all disjoint hypotheses that it contains are rejected, then the family wise error rate of the procedure is alpha. In the partitioning framework the seemingly different procedures of Bonferroni-Holm and Hochberg-Hommel differs only in one single aspect. The partitioning framework can also be used when combining different multiple testing procedures, such as gate keeping procedures.

Content

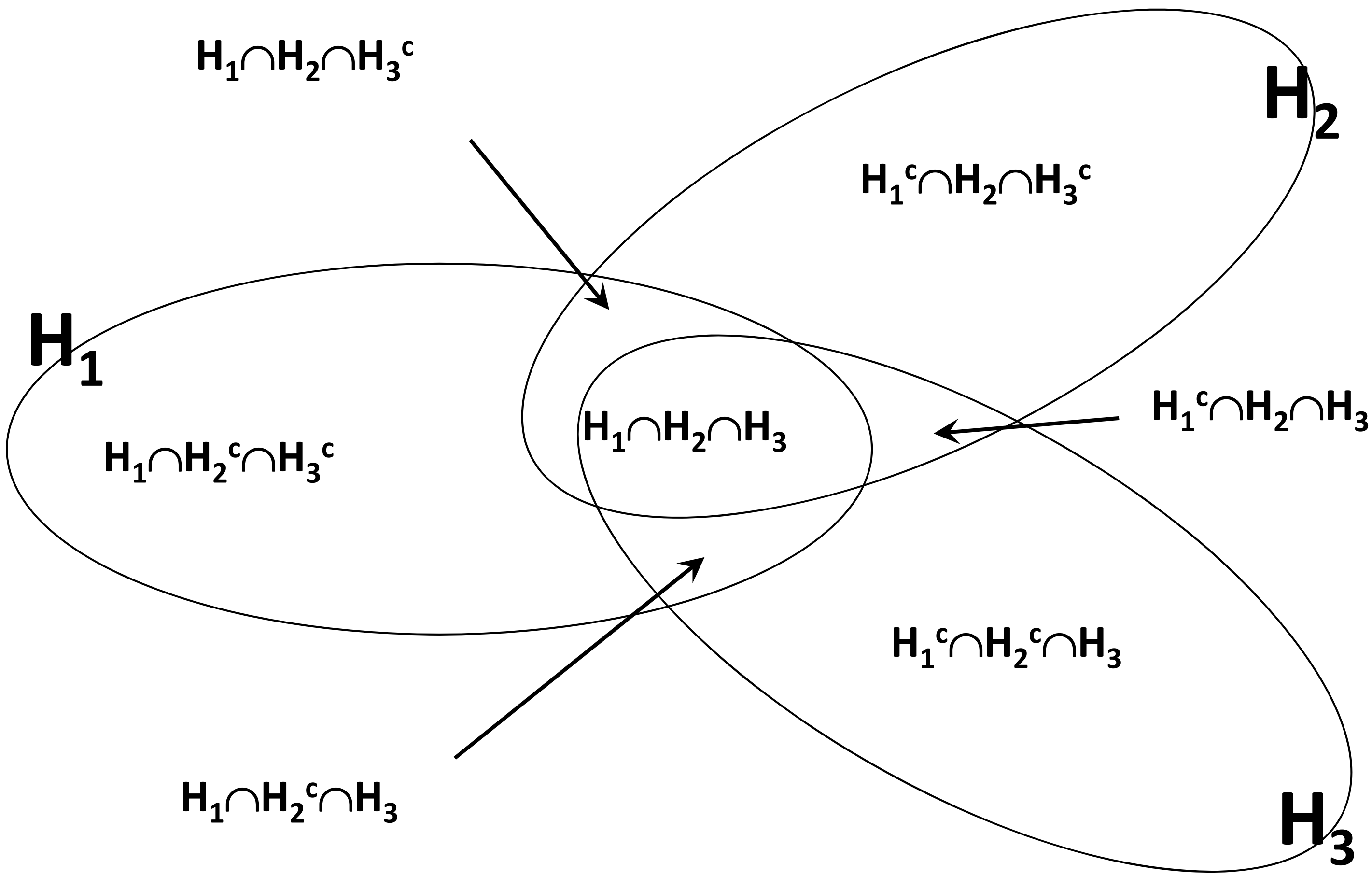
1. **Introduction and definitions**
2. **Observation**
3. **Partitioning principle**
4. **Examples**
    1. Alpha exhausting
    2. Logical restrictions
    3. Combination of tests
    4. Consonant
    5. Inferential restrictions
5. **Combination of multiple comparison procedures**
    1. Adding hypothese to an existing procedure
    2. Parallel gate keeping
6. **Summary**

Multiplicity - observation

- If two tests of the same hypothesis are performed then multiplicity must be taken into consideration
  - e.g. Testing if two samples comes from the same distribution with both a t-test and a Wilcoxon test

- If two tests of two disjoint hypotheses are performed then there are no multiplicity issues
  - e.g. Confidence intervals can be viewed as the set of not rejected hypotheses

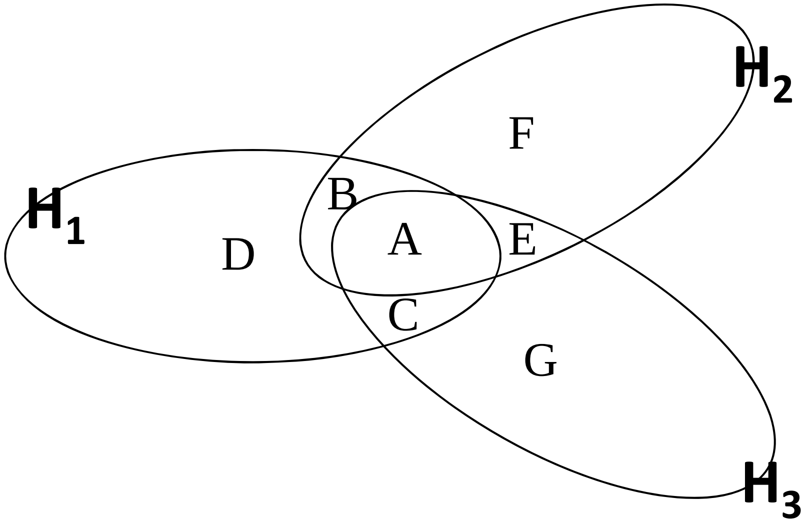
Partitioning of 3 null hypotheses - Venn diagram



In the following examples, the hypotheses, H_1, H_2, and H_3, are tested with significance tests resulting in the p-values p_1, p_2 and p_3, respectively

Partitioning of 3 null hypotheses - Table

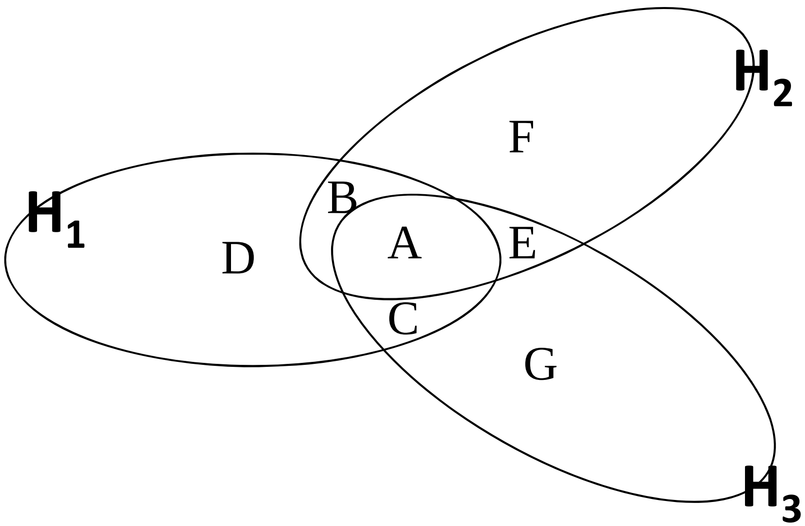
	disjoint hypotheses		
A	H_1	H_2	H_3
B	H_1	H_2	$H^c{}_3$
C	H_1	$H^c{}_2$	H_3
D	H_1	$H^c{}_2$	$H^c{}_3$
E	$H^c{}_1$	H_2	H_3
F	$H^c{}_1$	H_2	$H^c{}_3$
G	$H^c{}_1$	$H^c{}_2$	H_3



- Venn diagrams are useful for 3, but for more hypotheses a table is more useful

Partitioning test 3 null hypotheses

	disjoint hypotheses			Test
A	H_1	H_2	H_3	a
B	H_1	H_2	H^c_3	b
C	H_1	H^c_2	H_3	c
D	H_1	H^c_2	H^c_3	d
E	H^c_1	H_2	H_3	e
F	H^c_1	H_2	H^c_3	f
G	H^c_1	H^c_2	H_3	g



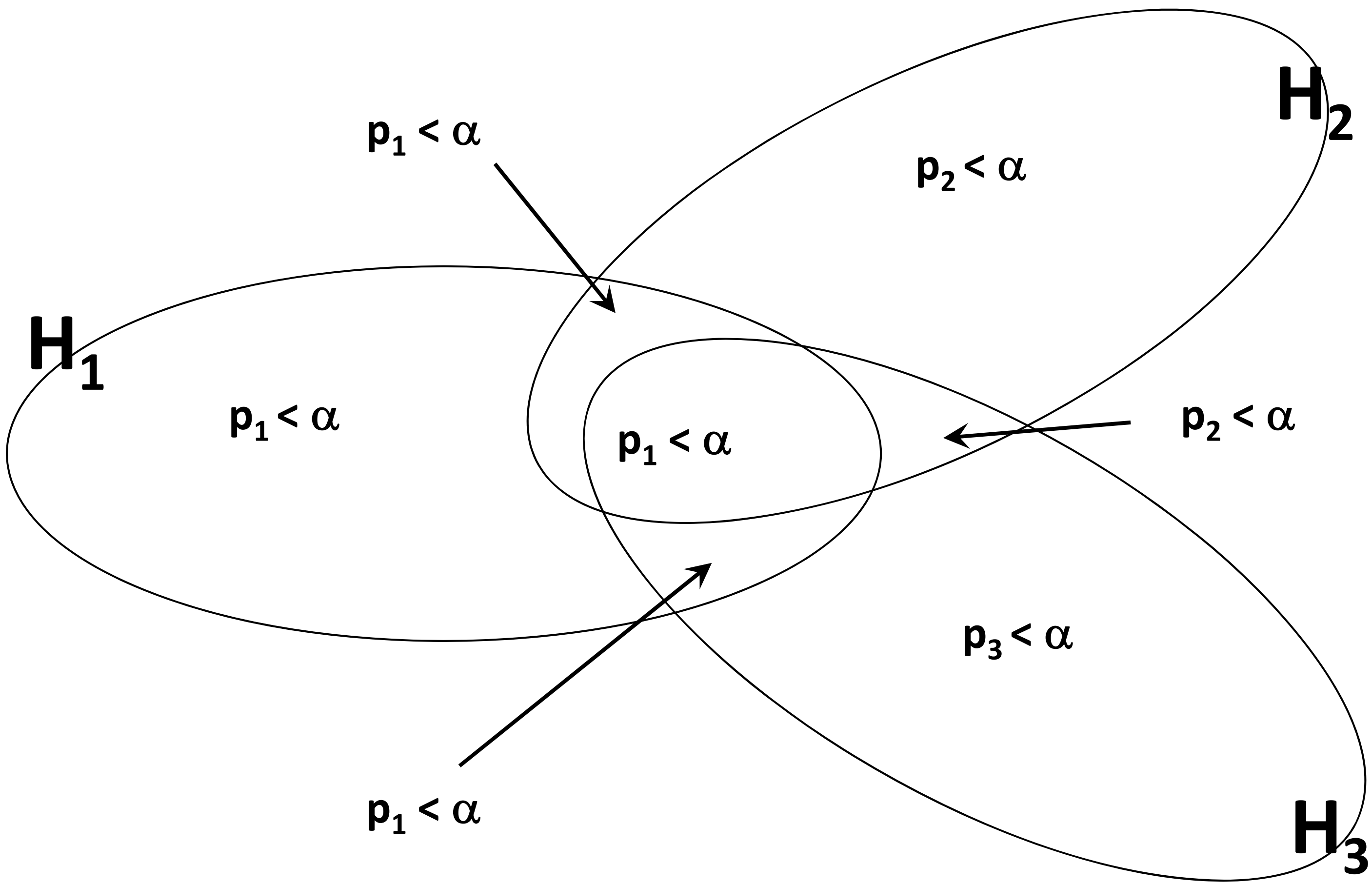
- Each disjoint hypothesis is tested by a significance test
  - Any α-level significance test can be used
- Reject H_1 if {A, B, C, D} all are rejected
- Reject H_2 if {A, B, E, F} all are rejected
- Reject H_3 if {A, C, E, G} all are rejected
- FWER (family wise error rate) is α

Partitioning principle vs. Close Test Procedure

- **Partitioning principle:**
- Given a set of hypotheses, H_1, H_2, ..., H_n, we construct the 2^n-1 disjoint sub hypotheses, $H'_I=(\cap_{i\in I}H_i)\cap(\cap_{i\notin I}H^c_i)$, for all index sets $I\subseteq\{1, 2, ..., n\}$
- Each H'_I is tested, at level α, with a significance test T'_I (can be chosen arbitrarily)
- Reject H_k if all H'_I with $k\in I$ are rejected
- All H'_I are disjoint so only one rejection region per disjoint sub hypothesis

- **Close test procedure:**
- Given a set of hypotheses, H_1, H_2, ..., H_n, we construct the 2^n-1 nested sub hypotheses, $H''_I=\cap_{i\in I}H_i$, for all index sets $I\subseteq\{1, 2, ..., n\}$
- Each H''_I is tested , at level α, with a significance test T''_I (can be chosen arbitrarily)
- Reject H_k if all H''_I with $k\in I$ are rejected
- The sub hypotheses H''_I are not disjoint, so special arguments related to the combinations of rejection regions are required

Example - Fix sequence of tests



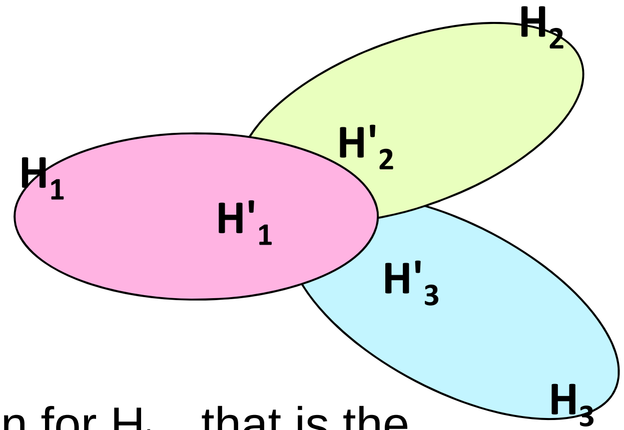
The hypotheses are tested in the following sequence: first H_1, then H_2 and last H_3.

The rejection rules of the disjoint hypotheses are given in the figure

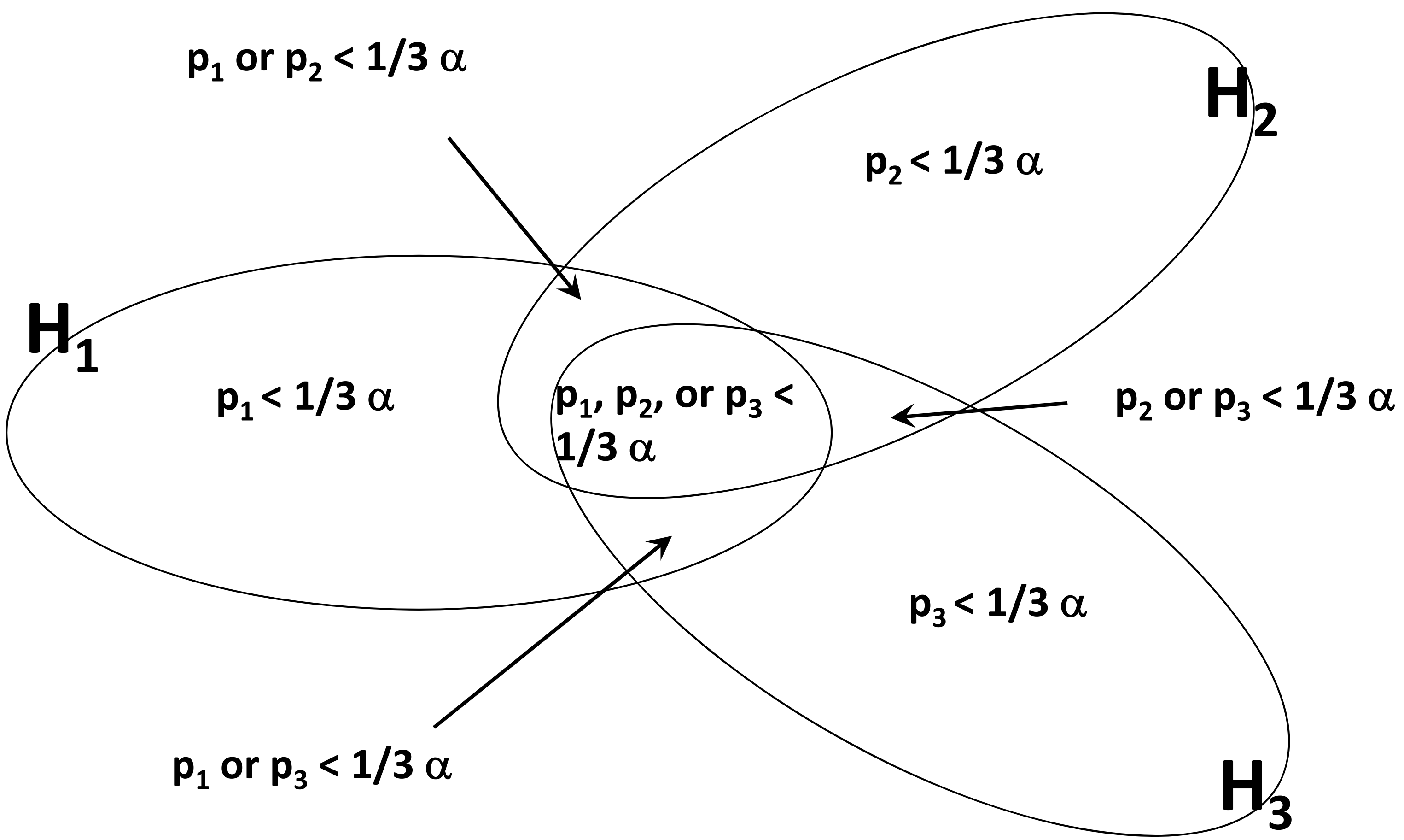
Note: In order to reject H_2 both p_1 and p_2 need to be $<\alpha$. H_3 is rejected if all three p-values are $<\alpha$

Example - Fix sequence of tests

- We are testing the hypotheses: H_1, H_2, ..., H_n each at the significance level α
- Fix sequence procedure:
  - Test H_1, if rejected continue if not then stop
  - Test H_2, if rejected continue if not then stop
  - ...
  - Test H_n
- Partitioning proof
  - Construct the disjoint hypotheses
  - $H'_1=H_1$, $H'_2=H_2\backslash H_1$, ... , $H'_n=H_n\backslash(H_1 \cup H_2 \cup ... \cup H_{n-1})$
  - Use the same significance test to H'_k as was given for H_k , that is the rejection regions are the same
  - P(type I error) = $\max_k P(\text{Reject } H'_k \mid H'_k) = \max_k P(\text{Reject } H_k \mid H'_k) \leq \max_k P(\text{Reject } H_k \mid H_k) \leq \alpha$, as H'_k is a subset of H_k
  - In order to reject H_k all disjoint hypotheses that is part of H_k needs to be rejected, that is $H_1 \cup H_2 \cup ... \cup H_k$, so H_k is rejected if all the k first hypotheses are rejected

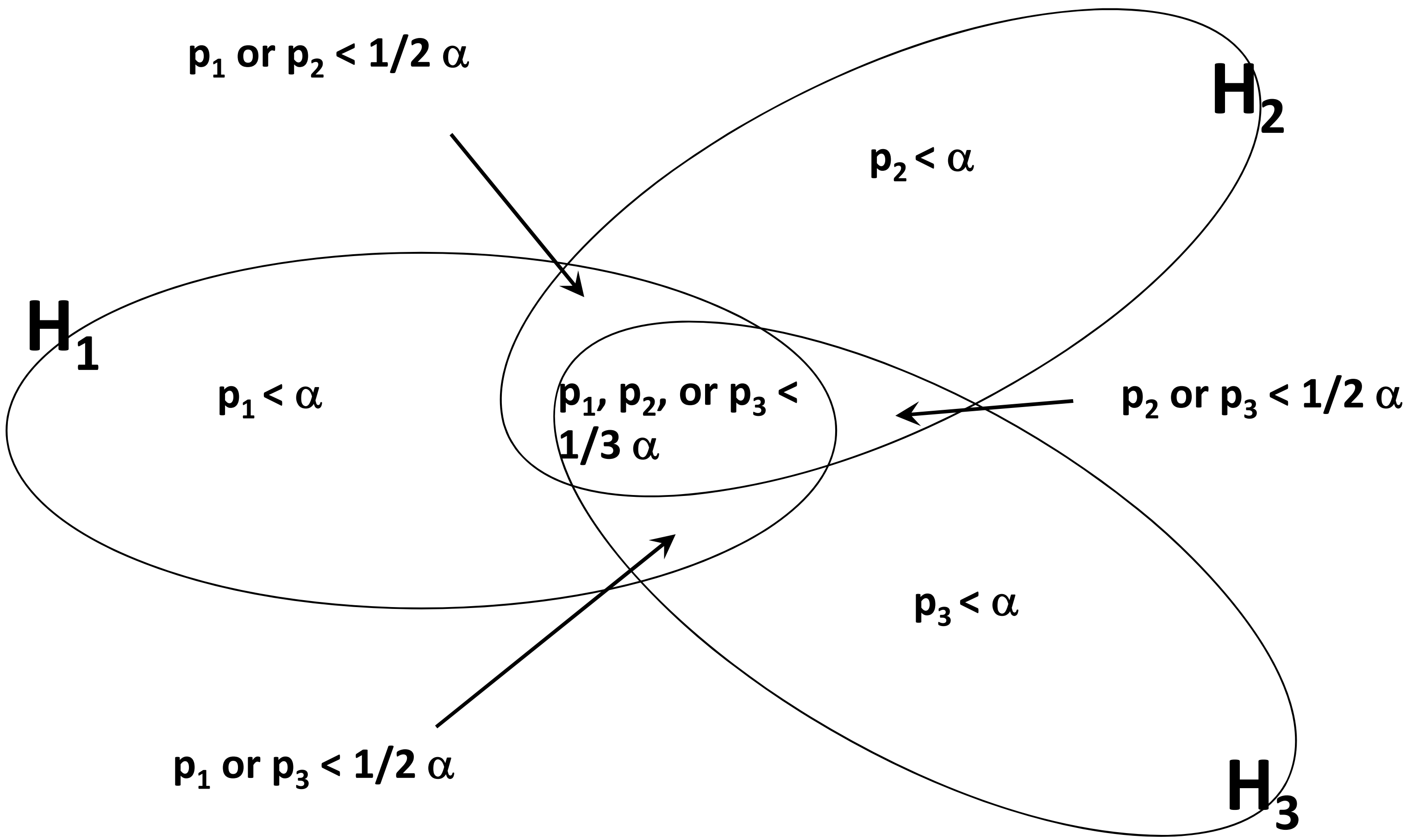


Example - Classic Bonferroni



Note: Only the central disjoint hypothesis is tested on the α level, thus improvements can be made

Example - Bonferroni-Holm



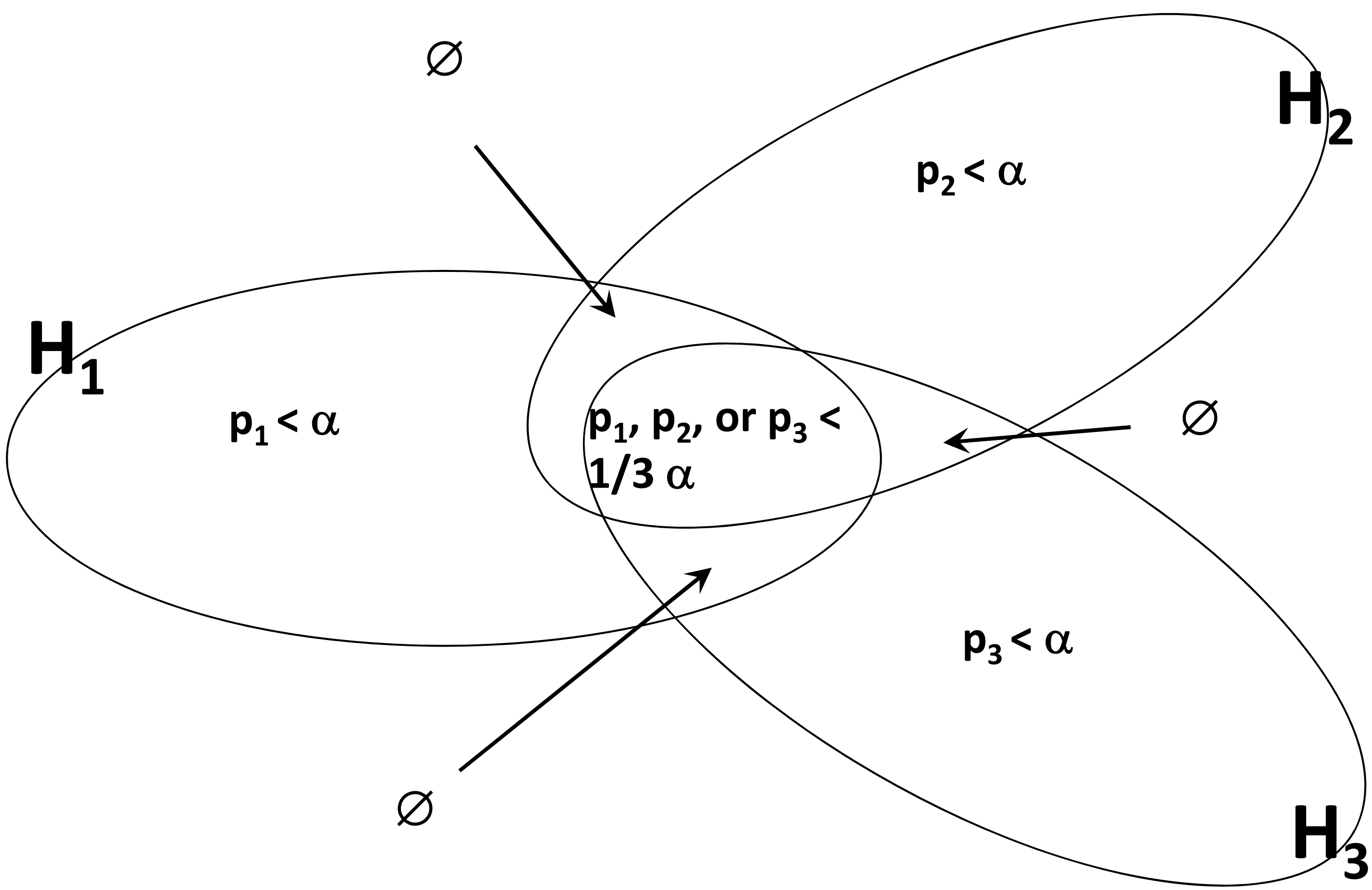
Note: Holm is using all α available, the key concept is **Alpha Exhaustion**

What is sequential in a sequential procedure?

- Many multiplicity procedures are stated as sequential procedures
- this does not mean that the data decides what test to do next, only the order of how we search trough the test results

- In the Bonferroni example we are performing 7 significance tests simultaneous
- If the innermost disjoint hypothesis is rejected, then we don't need to look at the disjoint hypotheses which is associated with the lowest p-value

Example - Bonferroni-Holm-Shaffer



For specific sets of hypotheses, some of the disjoint hypotheses might be empty, example:
H_1: $\mu_A=\mu_B$,
H_2: $\mu_A=\mu_C$,
H_3: $\mu_B=\mu_C$,

Note: Empty disjoint hypotheses don't need to be tested. I call such empty disjoint hypotheses for **Logical Restrictions**

What kind of tests could be used?

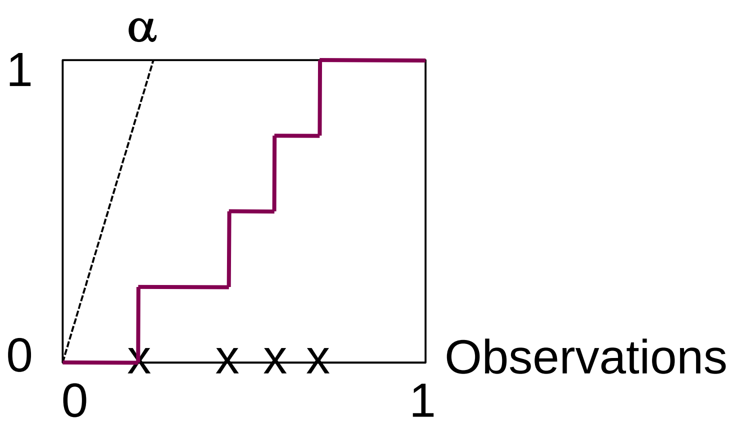
- Tests of the disjoint hypotheses could be based on:

- A function of the marginal tests (combinations of p-values)
  - Language: Marginal tests (p-values) are the tests (p-values) associated with the original hypotheses
  - Example:
    - at least one marginal test rejected, $\min_{k \in I}(p_k)$; alternative notation $p_{(1)}$
    - all marginal test rejected, $\max_{k \in I}(p_k)$; alternative notation $p_{(n)}$
    - at least k marginal tests rejected, $p_{(k)}$
    - use the marginal test with the lowest index in I, $p_{\min\{k:k \in I\}}$
    - combine the marginal p-values with the inverse distribution function, $\Sigma F^{-1}(p_k)$
    - Simes' test, $\min_{k \in I} (np_{(k)}/k)$
    - Dunnett's test, Sidak's test, $\min_{k \in I}(p_k)$

- A separate analyses/model for each disjoint hypotheses

Simes' Test

Empirical distribution function

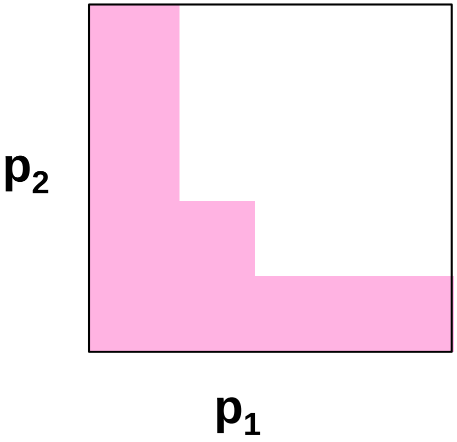


Probability that the edf crosses the dashed line is $\leq \alpha$ if the observations are independent uniformly distributed

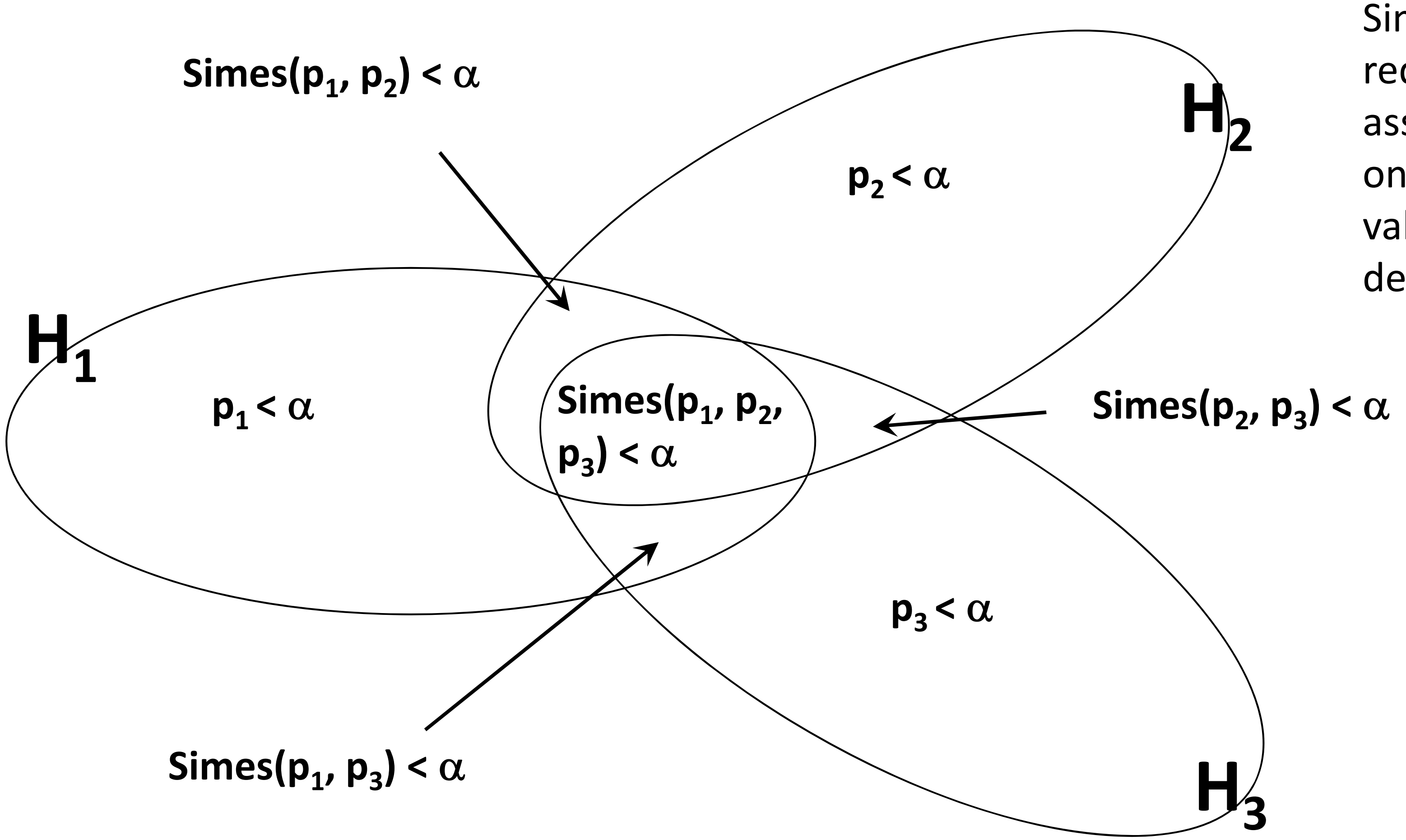
Remarkably:
Result is independent on the number of observations

- Test based on the following function of p-values:
  $\min_k (np_{(k)}/k)$
- If tested at a marginal level of α, that is $\min_k (np_{(k)}/k) < \alpha$
  - If the p-values are independent then the maximum type I error is α
  - Under no assumption about the distribution of the p-values the maximum type I error would be $\alpha\Sigma 1/k$

rejection region



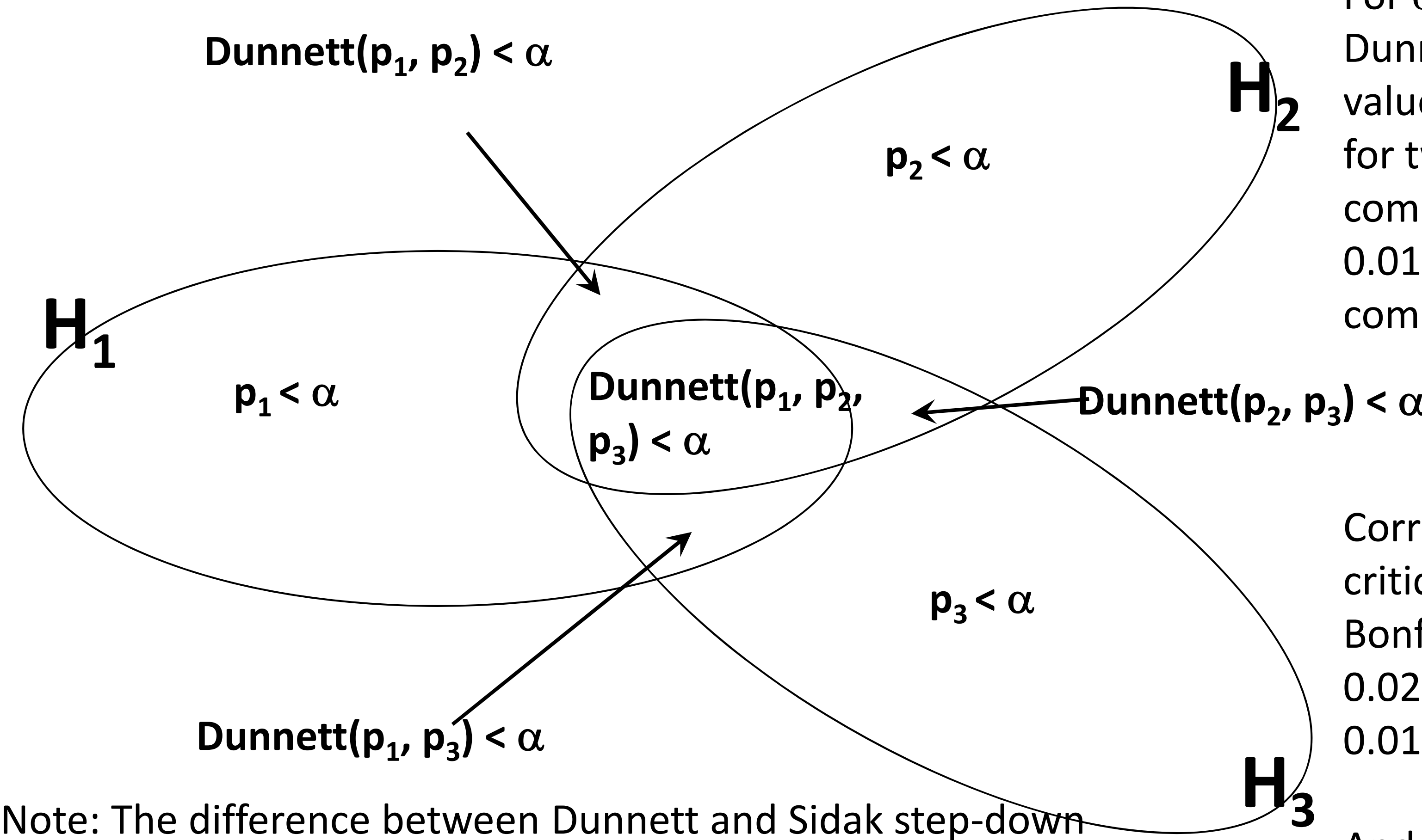
Example - Hochberg-Hommel (based on Simes' test)



Simes' test requires assumption on how the p-values are dependent

Note: The only difference between Hochberg-Hommel and Bonferroni-Holm is how the marginal tests are combined

Example - Dunnett-Tamhane Step-down



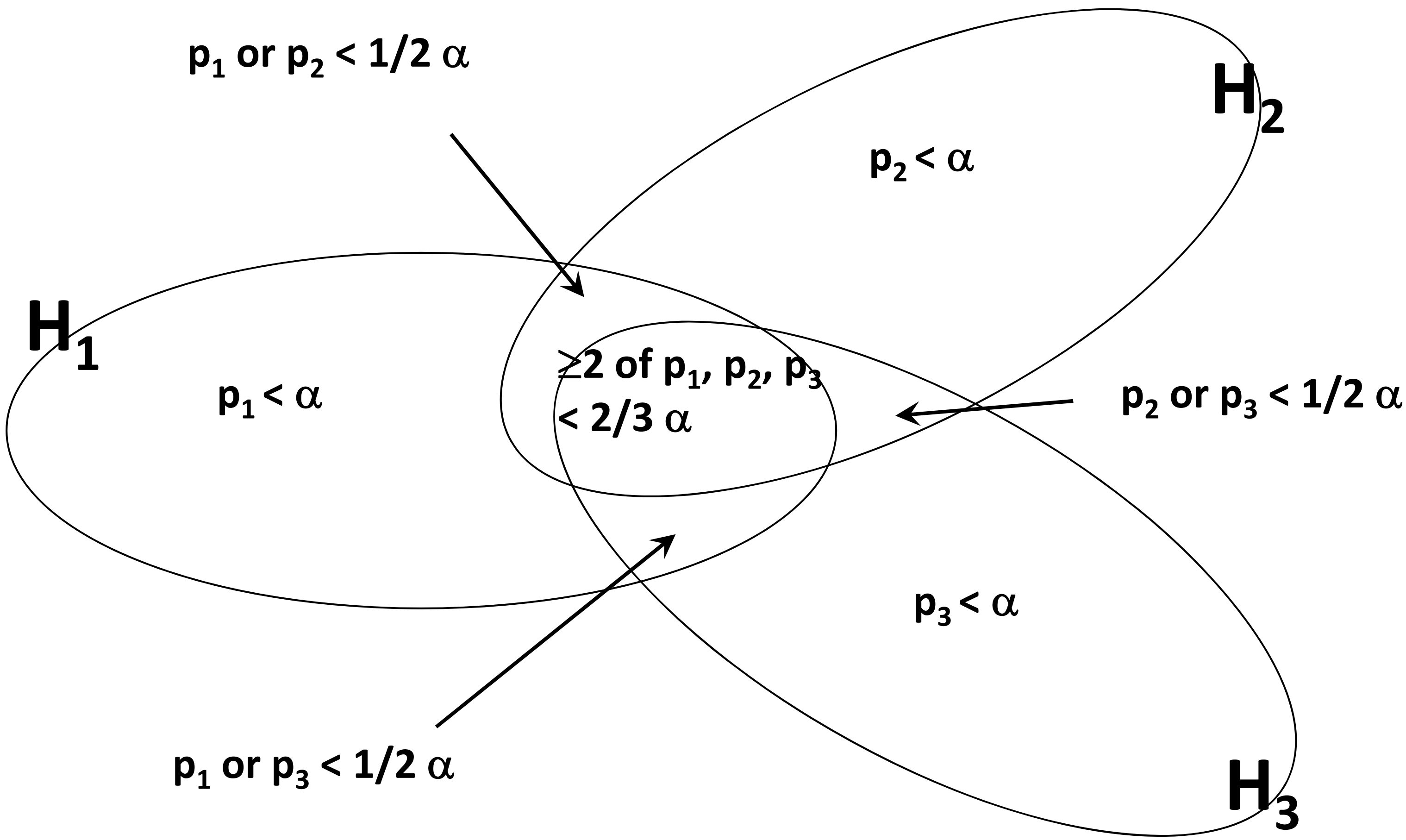
For α=0.05 the Dunnett's critical values are 0.0277 for two comparisons and 0.0196 for three comparisons.

Corresponding critical values for Bonferroni are 0.0250 and 0.0167

And for Sidak: 0.0253 and 0.0170

Note: The difference between Dunnett and Sidak step-down procedures and Bonferroni-Holm is how much you can assume. Dunnett is many groups versus one control and Sidak is when you assume that the p-values are independent

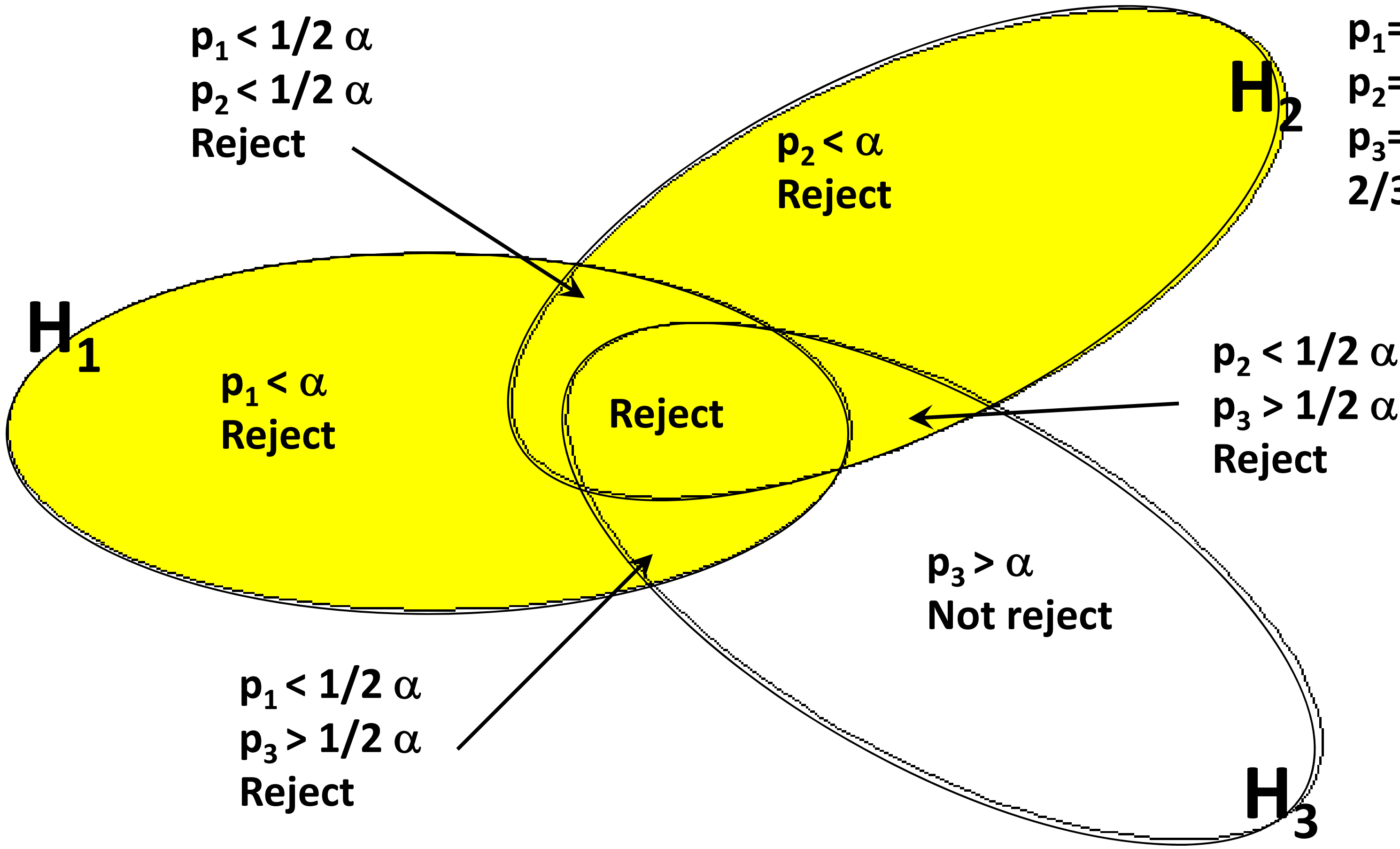
Example - at least 2 of 3 with extended Holm type rule



Note:
P(at least k tests rejected) $\leq$ min{1, ΣP(reject Ti)/k }
follows from the Generalized Chebyshev inequality
P(Y$\geq$k)$\leq$E(Y)/k.

Example - at least 2 of 3 with extended Holm type rule, α=0.05

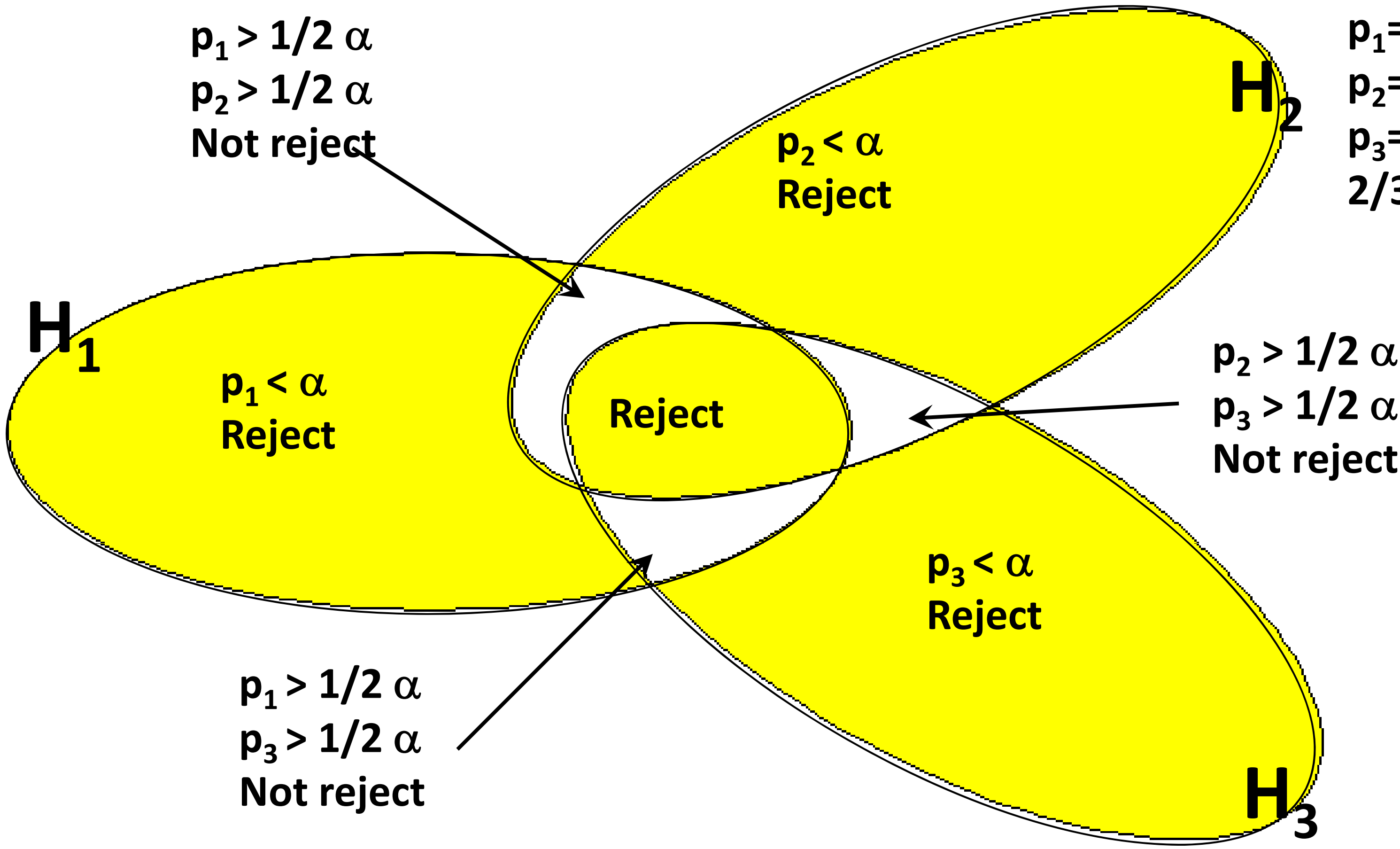
Result:
p_1=0.01
p_2=0.02
p_3=0.71
2/3 α=0.033



Conclusion: H_1 and H_2 can be rejected but not H_3

Example - at least 2 of 3 with extended Holm type rule, α=0.05

Result:
p_1=0.03
p_2=0.03
p_3=0.03
2/3 α=0.033

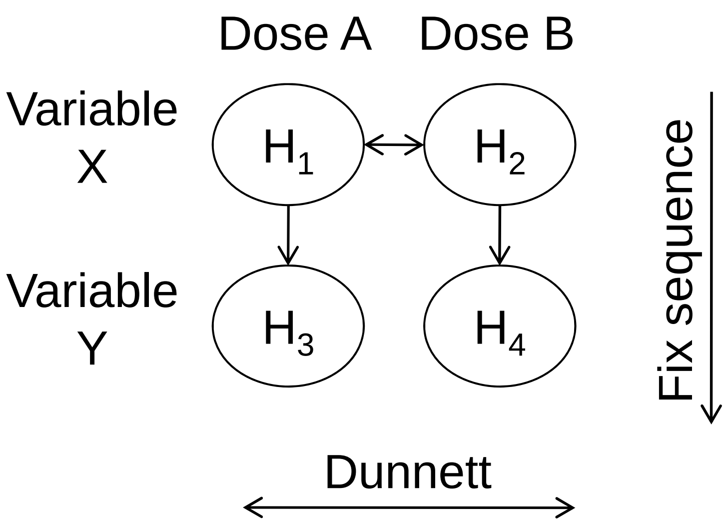


Conclusion: Can not reject any of H_1, H_2, or H_3

Note: This procedure is not **Consonant**, as a rejection of the inner disjoint hypothesis does not guarantees that at least one original hypothesis is rejected

Example - Interlinked Dunnett and fixed sequence

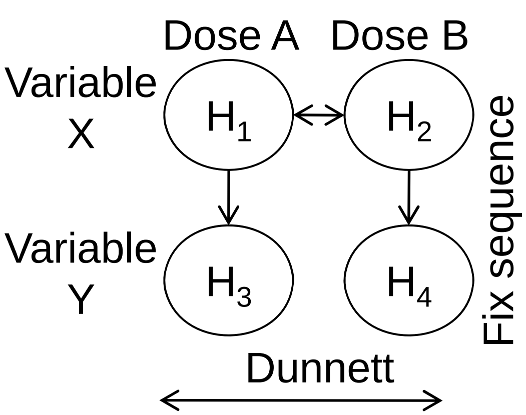
- 2 doses vs. placebo
- 2 variables, one primary and one secondary
- A significant difference for the secondary variable in a dose is only meaningful if there is a significance in the primary variable for that dose.
  I call such restrictions **Inferential restrictions**



Example - Interlinked Dunnett and fixed sequence

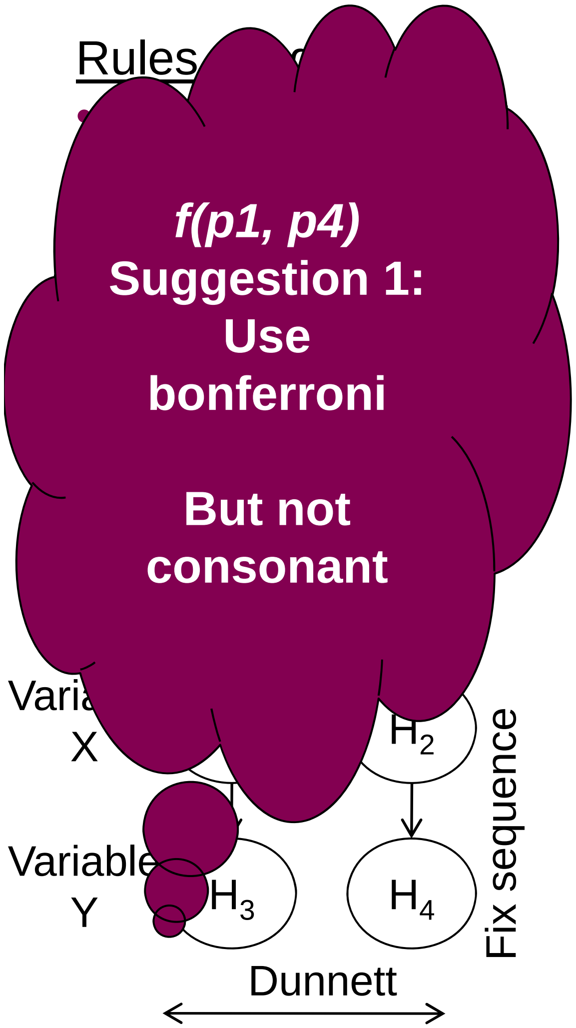
<u>Rules used:</u>

- Dunnett
  - H_1 and H_2
  - H_3 and H_4
- Fix sequence:
  - If H_1 and H_3 then only p_1 used
  - If H_2 and H_4 then only p_2 used



disjoint hypotheses				Rejection based on
H_1	H_2	H_3	H_4	Dunnett(p_1, p_2)
H_1	H_2	H_3	H^c_4	Dunnett(p_1, p_2)
H_1	H_2	H^c_3	H_4	Dunnett(p_1, p_2)
H_1	H_2	H^c_3	H^c_4	Dunnett(p_1, p_2)
H_1	H^c_2	H_3	H_4	f(p_1, p_4) *(to be discussed)*
H_1	H^c_2	H_3	H^c_4	p_1
H_1	H^c_2	H^c_3	H_4	f(p_1, p_4) *(to be discussed)*
H_1	H^c_2	H^c_3	H^c_4	p_1
H^c_1	H_2	H_3	H_4	f(p_2, p_3) *(to be discussed)*
H^c_1	H_2	H_3	H^c_4	f(p_2, p_3) *(to be discussed)*
H^c_1	H_2	H^c_3	H_4	p_2
H^c_1	H_2	H^c_3	H^c_4	p_2
H^c_1	H^c_2	H_3	H_4	Dunnett(p_3, p_4)
H^c_1	H^c_2	H_3	H^c_4	p_3
H^c_1	H^c_2	H^c_3	H_4	p_4

Example - Interlinked Dunnett and fixed sequence

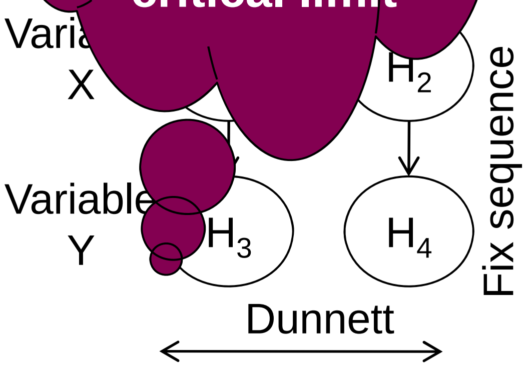


disjoint hypotheses				Rejection based on (α=0.05)
H_1	H_2	H_3	H_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H_3	H^c_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H^c_3	H_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H^c_3	H^c_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H^c_2	H_3	H_4	$p_1<0.025$ or $p_4<0.025$
H_1	H^c_2	H_3	H^c_4	$p_1<0.05$
H_1	H^c_2	H^c_3	H_4	$p_1<0.025$ or $p_4<0.025$
H_1	H^c_2	H^c_3	H^c_4	$p_1<0.05$
H^c_1	H_2	H_3	H_4	$p_2<0.025$ or $p_3<0.025$
H^c_1	H_2	H_3	H^c_4	$p_2<0.025$ or $p_3<0.025$
H^c_1	H_2	H^c_3	H_4	$p_2<0.05$
H^c_1	H_2	H^c_3	H^c_4	$p_2<0.05$
H^c_1	H^c_2	H_3	H_4	$p_3<0.0277$ or $p_4<0.0277$
H^c_1	H^c_2	H_3	H^c_4	$p_3<0.05$
H^c_1	H^c_2	H^c_3	H_4	$p_4<0.05$

Example - Interlinked Dunnett and fixed sequence

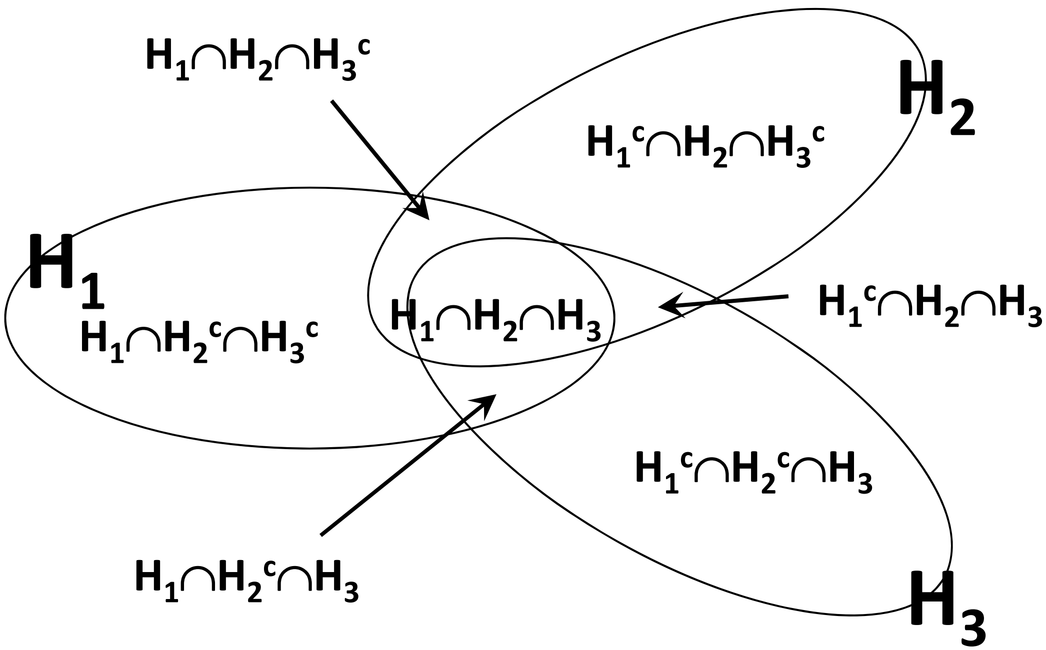
Rules

_f(p1, p4)_
Suggestion 2: Use weighted bonferroni, where the weight of H_1 and H_2 is the same as the Dunnett critical limit

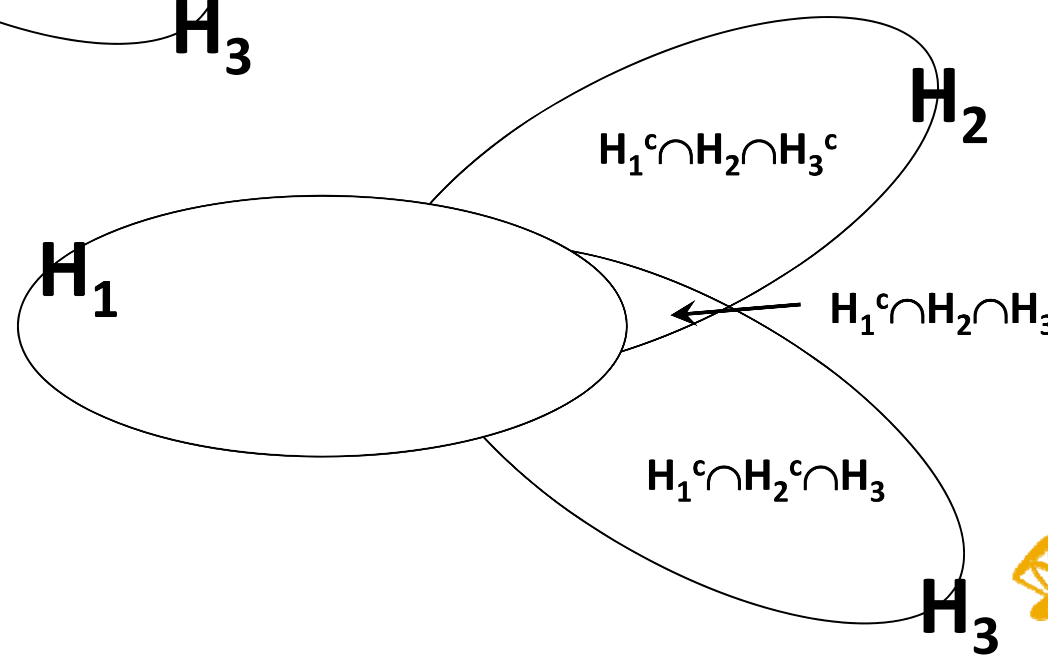


disjoint hypotheses				Rejection based on (α=0.05)
H_1	H_2	H_3	H_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H_3	H^c_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H^c_3	H_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H_2	H^c_3	H^c_4	$p_1<0.0277$ or $p_2<0.0277$
H_1	H^c_2	H_3	H_4	$p_1<0.0277$ or $p_4<0.0223$
H_1	H^c_2	H_3	H^c_4	$p_1<0.05$
H_1	H^c_2	H^c_3	H_4	$p_1<0.0277$ or $p_4<0.0223$
H_1	H^c_2	H^c_3	H^c_4	$p_1<0.05$
H^c_1	H_2	H_3	H_4	$p_2<0.0277$ or $p_3<0.0223$
H^c_1	H_2	H_3	H^c_4	$p_2<0.0277$ or $p_3<0.0223$
H^c_1	H_2	H^c_3	H_4	$p_2<0.05$
H^c_1	H_2	H^c_3	H^c_4	$p_2<0.05$
H^c_1	H^c_2	H_3	H_4	$p_3<0.0277$ or $p_4<0.0277$
H^c_1	H^c_2	H_3	H^c_4	$p_3<0.05$
H^c_1	H^c_2	H^c_3	H_4	$p_4<0.05$

Recurrence



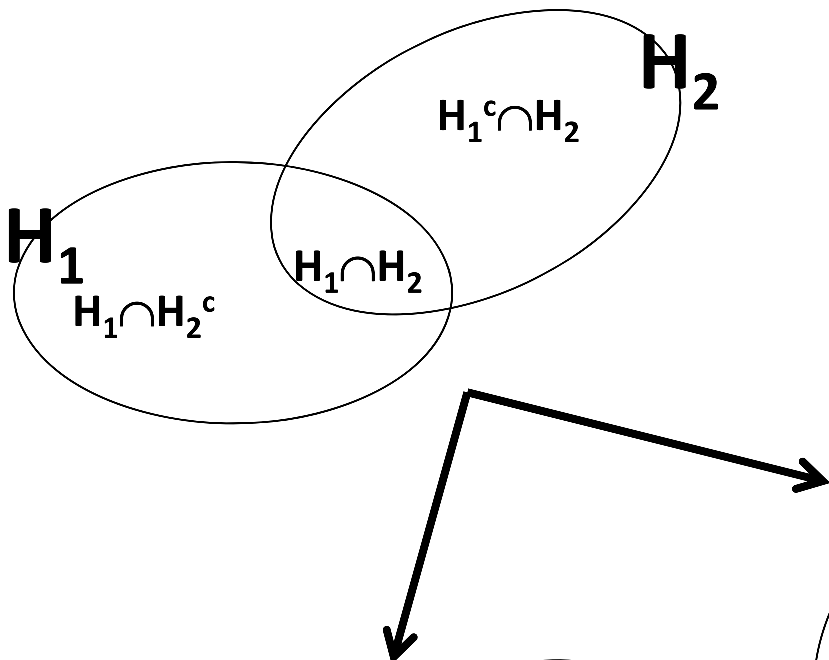
If you have a consonant procedure then if you reject the innermost disjoint hypothesis then you know that one of the marginal hypothesis can be rejected. If you remove this hypothesis, for example H_1, you will have a new multiplicity procedure consisting of H_2, and H_3.

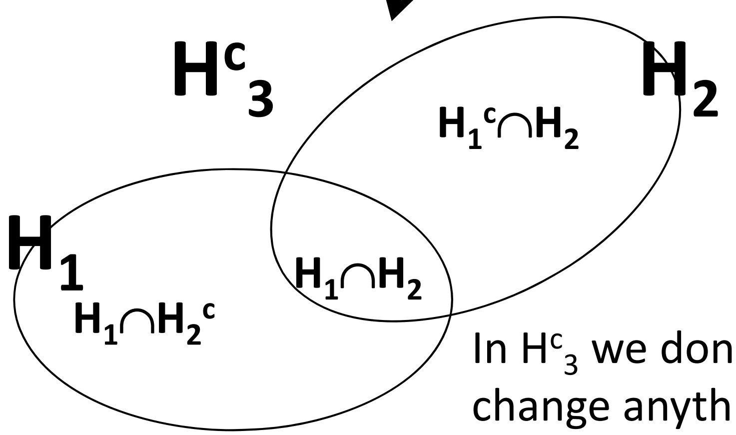


Adding a hypothesis

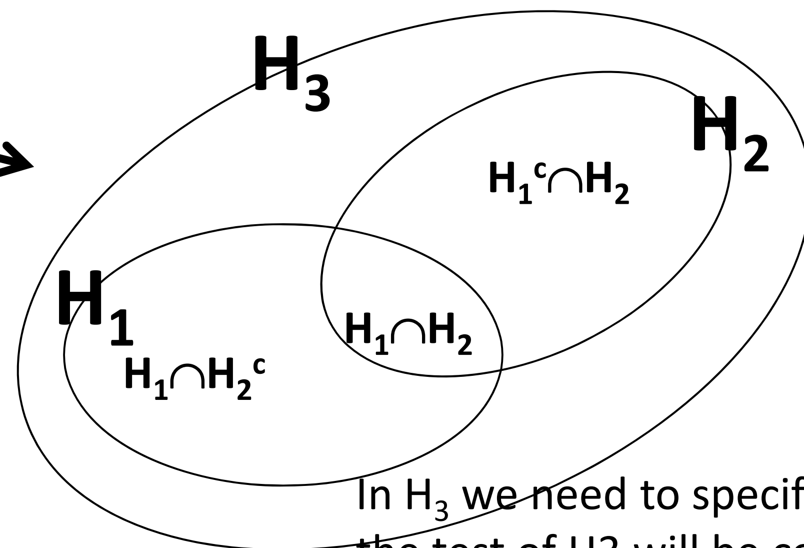
In the same manner as we remove hypotheses we can add hypotheses to a procedure

Here we add H_3 to the existing procedure H_1 and H_2





In H^c_3 we don't need to change anything in the existing procedure



In H_3 we need to specify how the test of H3 will be combined with the existing tests.

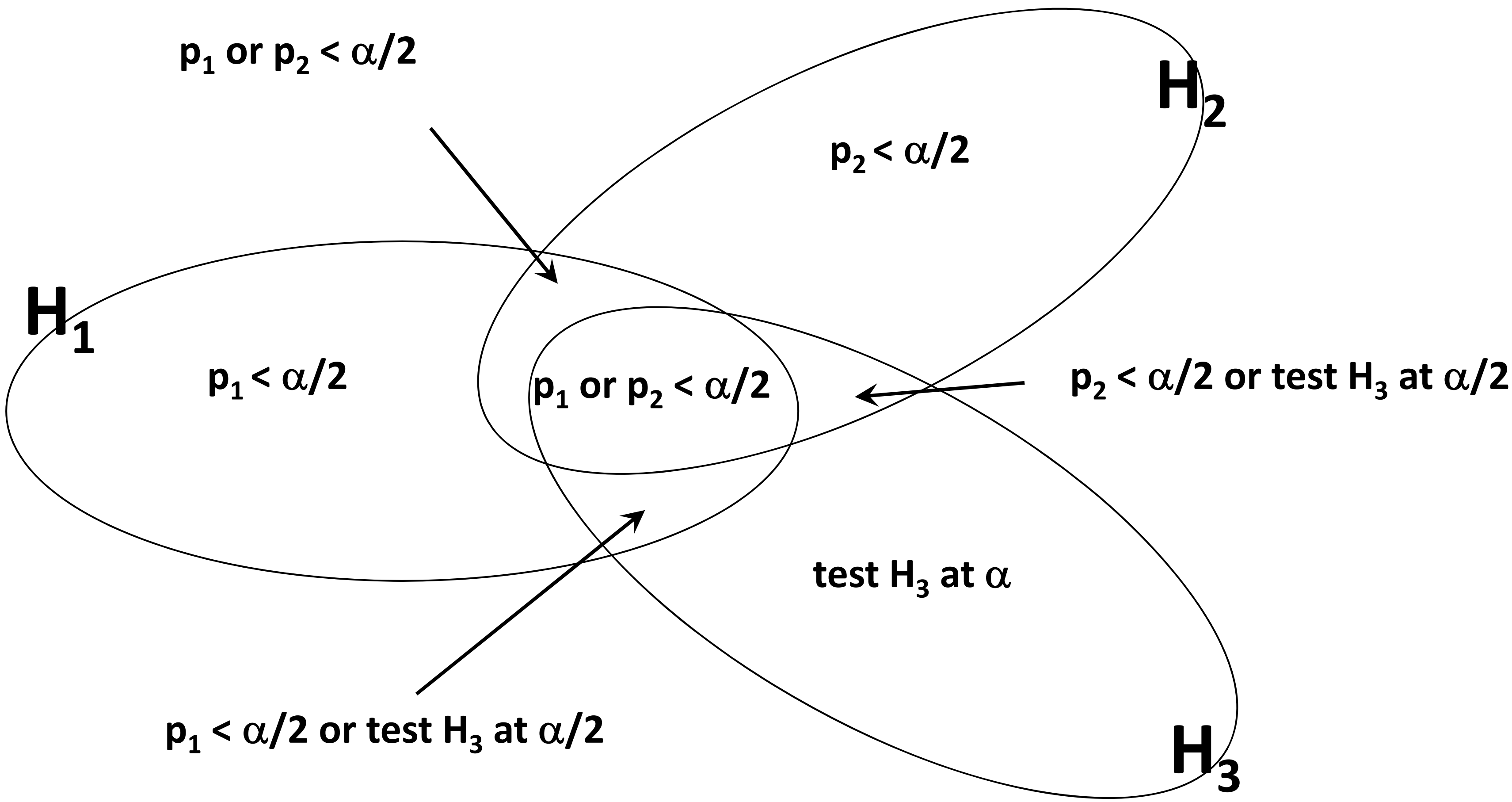
Example: by sequence or bonferroni

Combining two (or more) Multiple Comparison Procedures

- Combining two different sets of Multiple Comparison Procedures (MCP) is done as follows
- The new set of disjoint hypotheses will consist of two types of disjoint hypotheses:
- Those that contain only elements from one of the MPC
  - These disjoint hypotheses will be tested as in the original MPC
- Those that contain elements from two or more MPC
  - Such disjoint hypotheses will be tested by combining the test from the MPCs
  - The combination of tests should be used with methods that preserve the consonance property, e.g. Bonferroni or fixed sequence combinations

Example - Parallel gate keeping

H_1 and H_2 are the gate keepers. If at least one is rejected the H_3 is tested (H_3 could be a MCP)



Note: The procedure is not Alpha Exhaustive

Summary

- Formulating a Multiple Test Procedure in the partitioning framework makes it easy to prove that the procedure controls the type 1 error rate
- Care is needed in making sure that the tests of the disjoint hypotheses are generating a particular sequential description

- Key concepts discussed
  - Alpha exhaustion
  - Consonance
  - Logical restrictions versus Inferential restrictions