

Multiplicity Correction in a Group-Sequential Oncology Trial Including Subgroup Analyses and Multiple Primary Endpoints: Case Study

European Statistical Meeting on Oncology and Survival Analysis

Brussels, 17 November 2017

Agnes Balogh
Bristol-Myers Squibb

Disclaimer

The case study presented here is an actual BMS study. However,

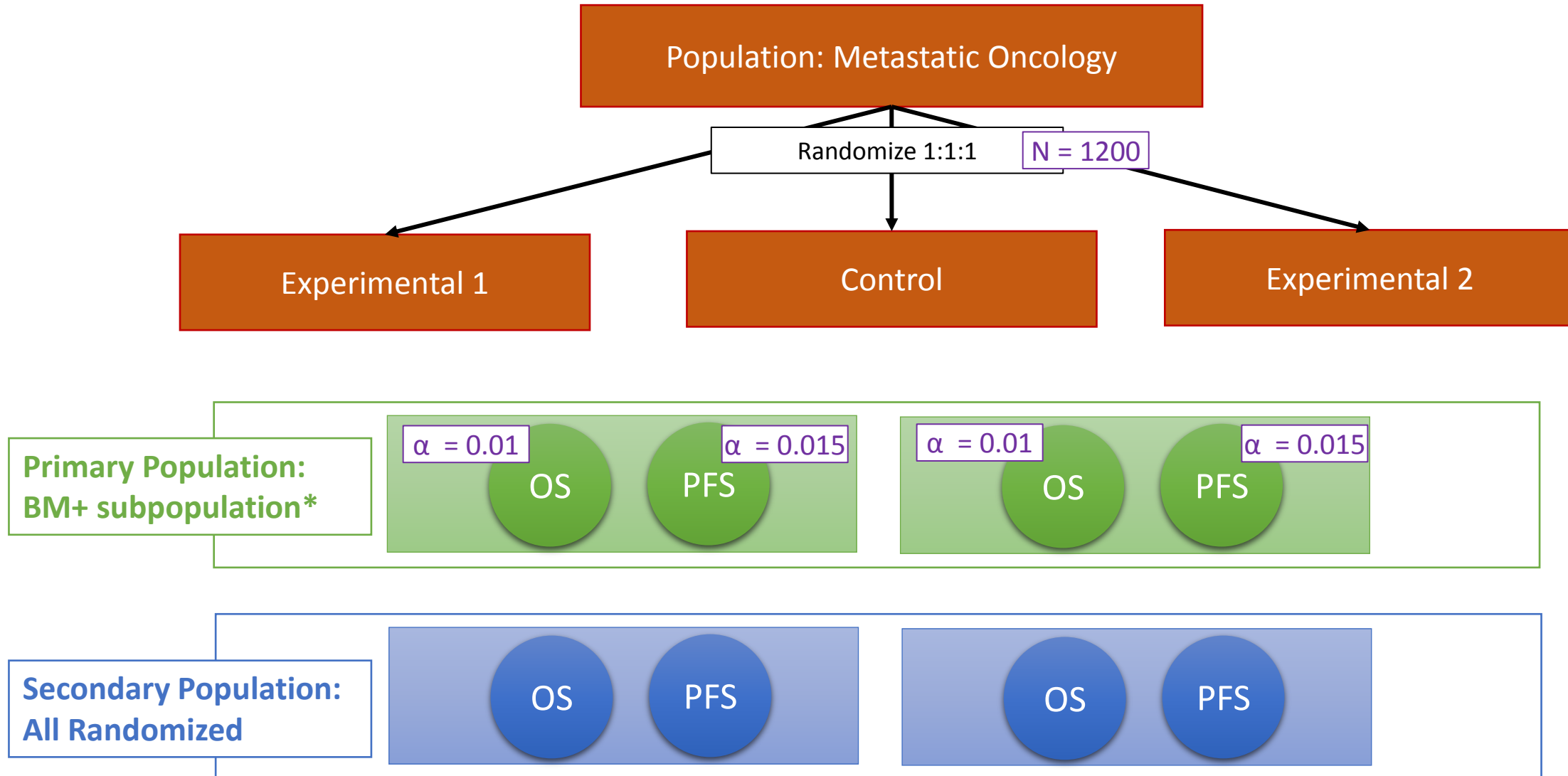
- Study number, disease setting, drug names will not be provided;
- Numerical parameters of the statistical design are not the actual ones (rather hypothetical)

Outline

1. Study Design
2. Testing strategy
3. Concluding remarks

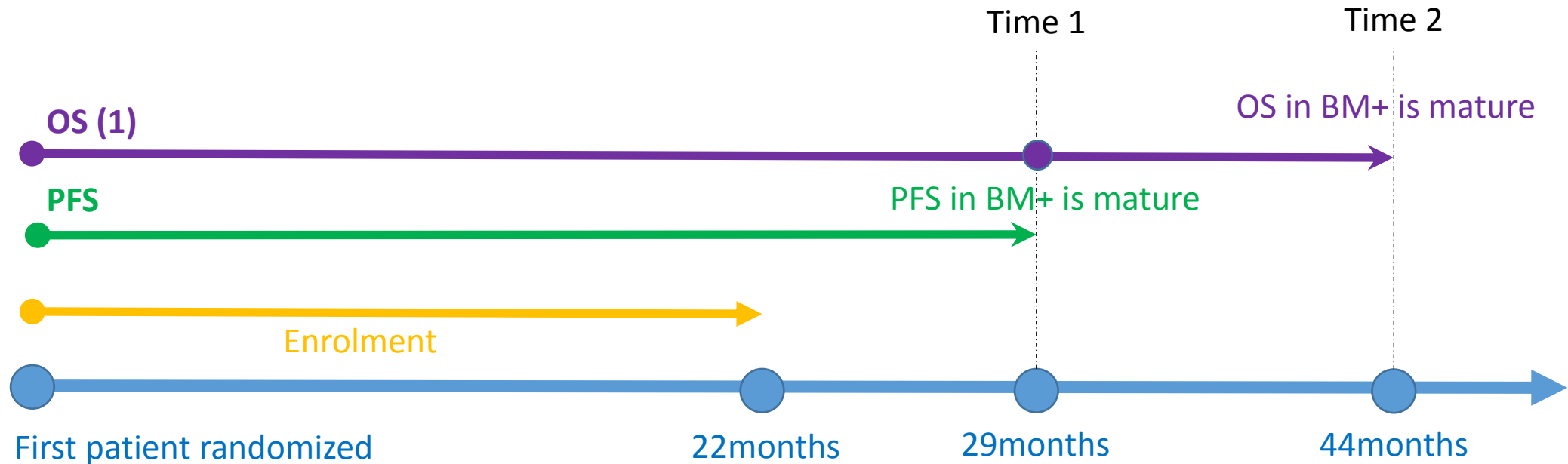
1. Study Design

Hypothetical Study Design



* About 40% of the population: prognostic and predictive

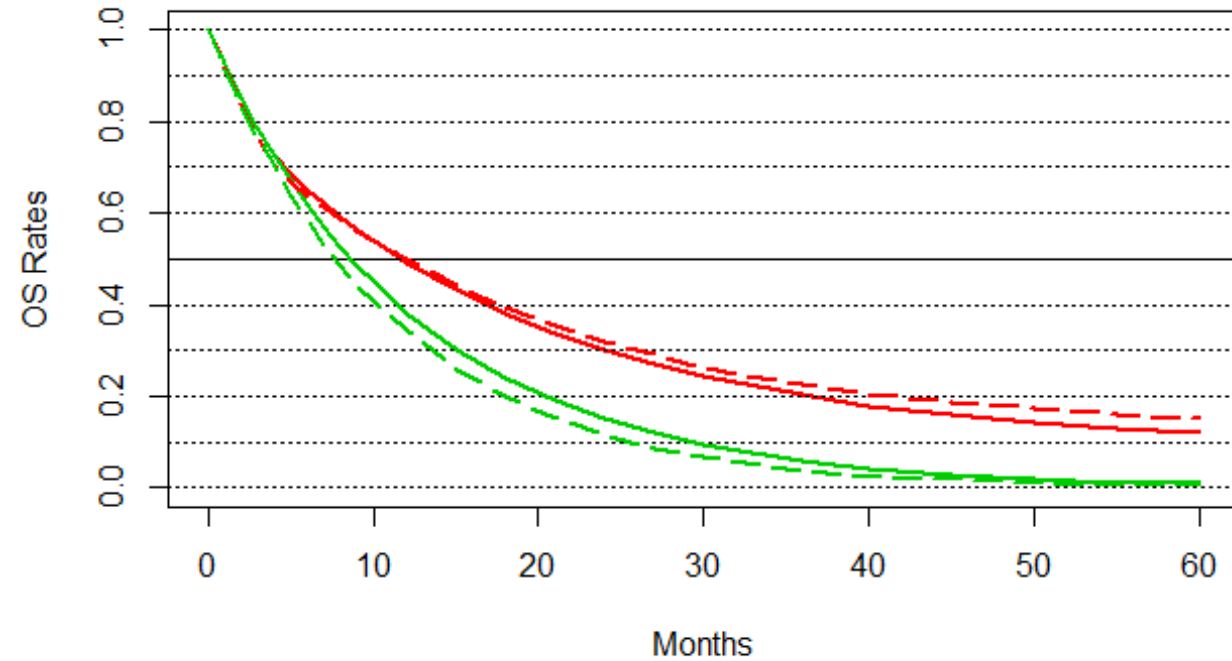
Group-Sequential Nature of the Design



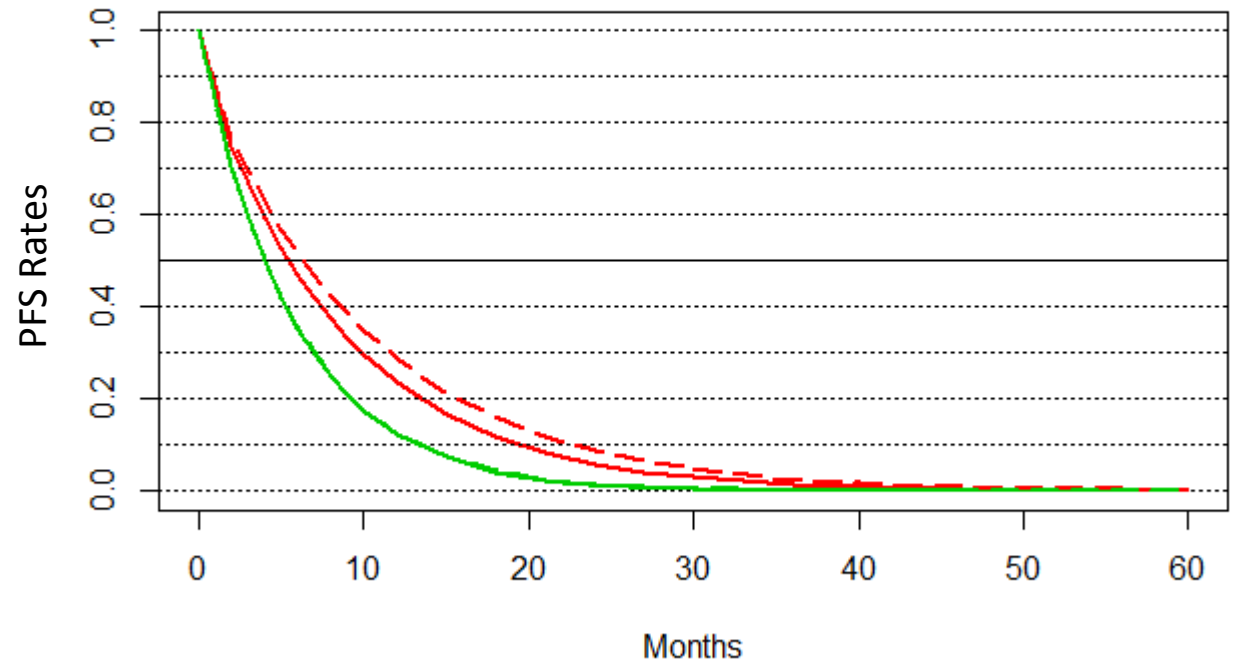
(1) Lan-DeMets α spending function used: for the two BM+ OS endpoints, O'Brien and Fleming; for the two All Randomized OS endpoints, Pocock type of boundary.

OS, PFS: Underlying Survival Curves

OS: BM+ vs. All Randomized



PFS: BM+ vs. All Randomized



Legend:

- - - - : BM+ treatment

_____ : All treatment

- - - - : BM+ control

_____ : All control

of Events in All 2 Arms vs in Control Arm

Analysis will be conducted when:

- Time 1: PFS in BM+ is mature
- Time 2: OS in BM+ is mature

Question: should maturity be defined based on the # of events in

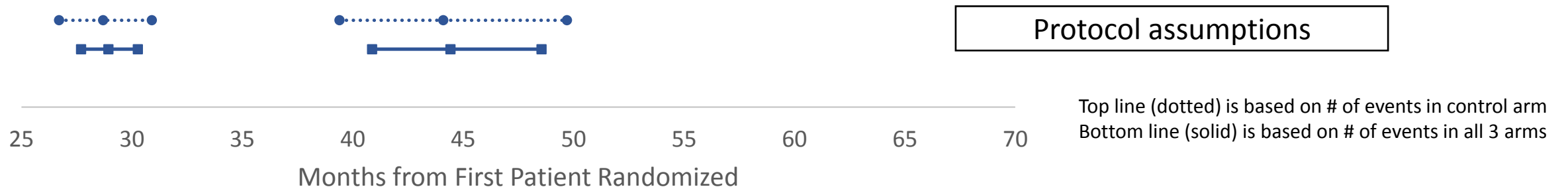
1. All 3 arms in total OR
2. The control arm (requires independent group)

Statistical properties were investigated under 4 scenarios:

<u>Scenario</u>	<u>Details</u>
Protocol Assumptions	
#1 OS, PFS: Exp2 no treatment effect	OS and PFS: no treatment effect in Experimental2 arm
#2 OS, PFS: Exp2 better	OS and PFS treatment effect (delay, cure rate, HR in non-cure patients) improved by 30% in Experimental2 arm
#3 Drop-out in all 3 arms	5% drop-out rate in all 3 arms

of Events in Total vs Control: Simulation Results

Time of Analysis (90% CI) under Different Scenarios



of Events in Total vs Control: Statistical power

- For all 8 endpoints, across all scenarios: power is never lower if timing is later
- Largest difference: for Exp1 in scenario#1 (OS BM+: 11%, OS All: 13%, PFS BM+: 5%)

3. Testing Strategy

A. Base Case

B. Graphical Approach: Introduction

C. Chosen testing strategy:

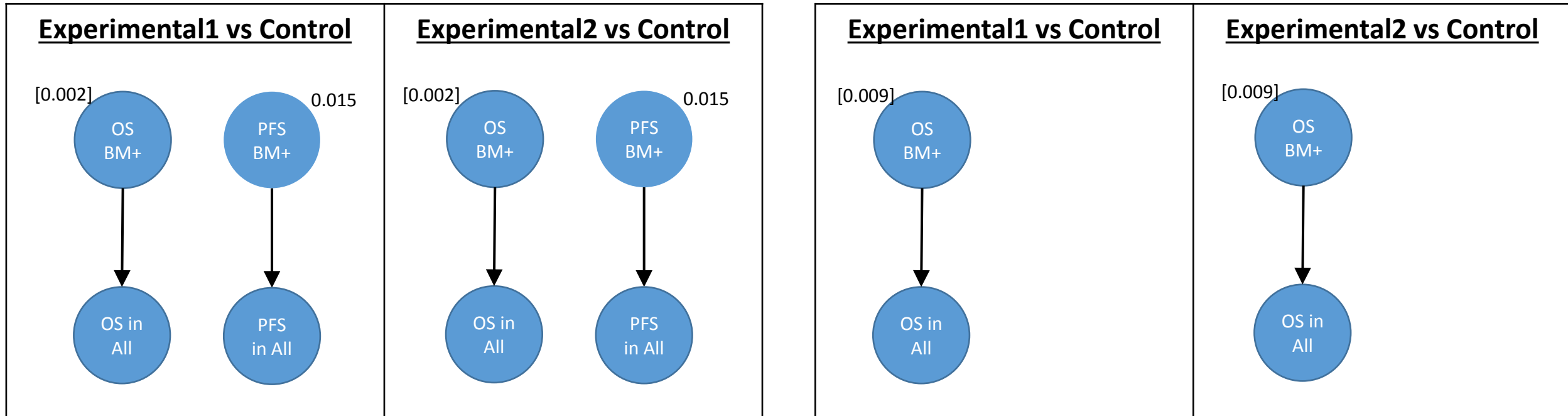
- Illustration of Sequentially Rejective Procedure
- Power Gain?

D. Alternative testing strategies

Testing Strategy: Base Case

Time 1

Time 2

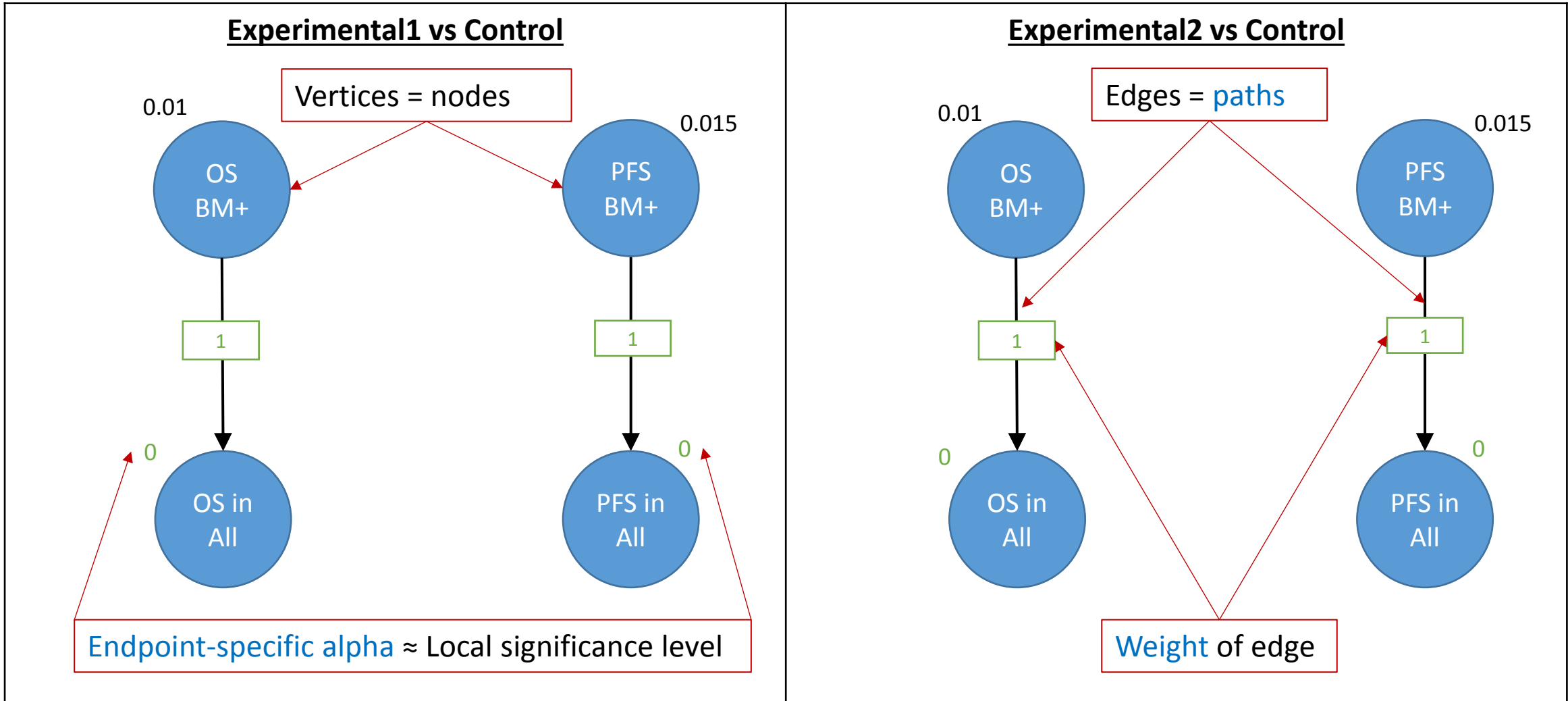


Possibility for Improvement: Graphical Approach

- Historical overview:
 - Introduced in 2009 by Bretz et al. (and Burman et al.)
 - 2013: extension to group-sequential trials (Maurer, Bretz)
 - 2017: Supported by the draft FDA guidance (*Multiple Endpoints in Clinical Trials*)
- Closed test procedure => Strong control of family-wise error rate
- A method to depict / develop / communicate a strategy consisting of Bonferroni-based sequential methods

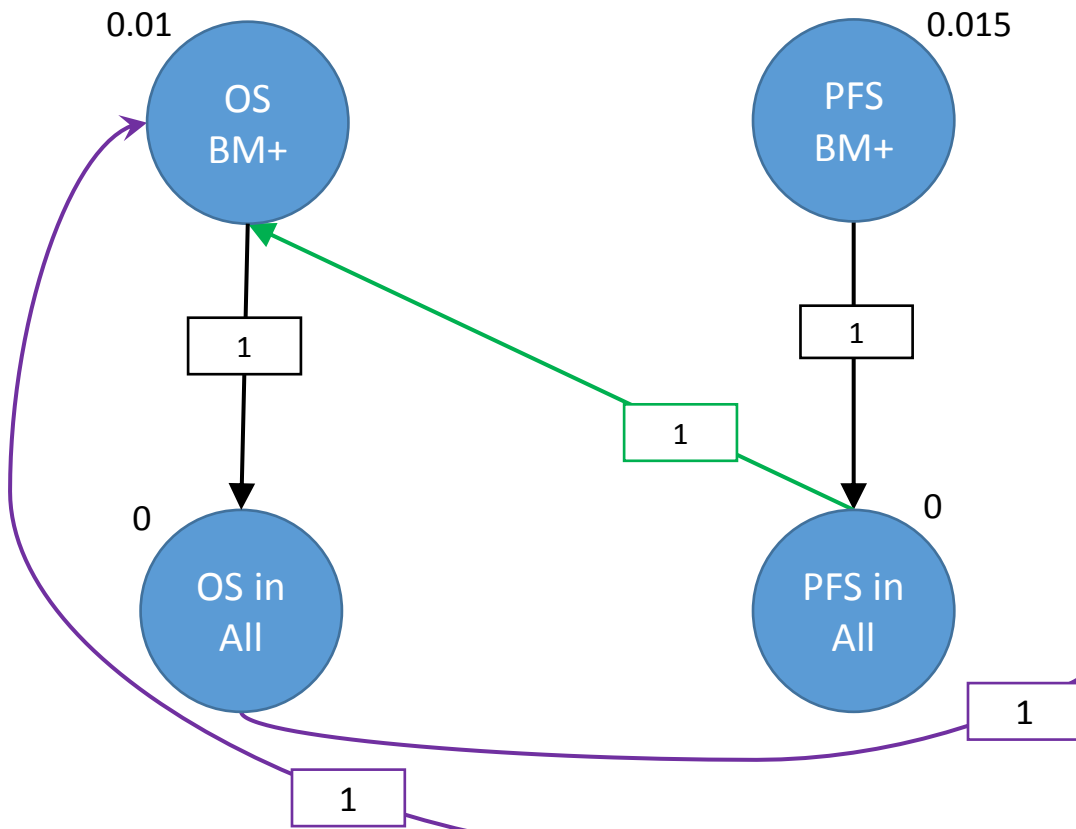
Additional consideration: PFS is not analysed at Time2

Graphical Approach: Introduction

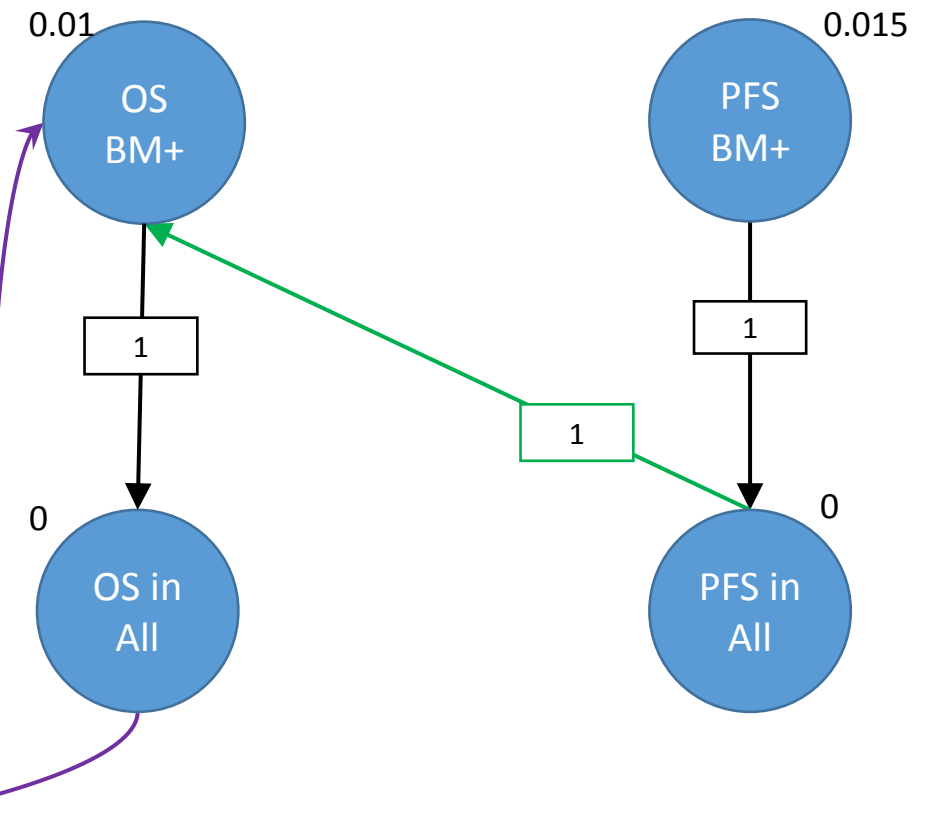


Improving Testing Strategy: Add Edges

Experimental1 vs Control



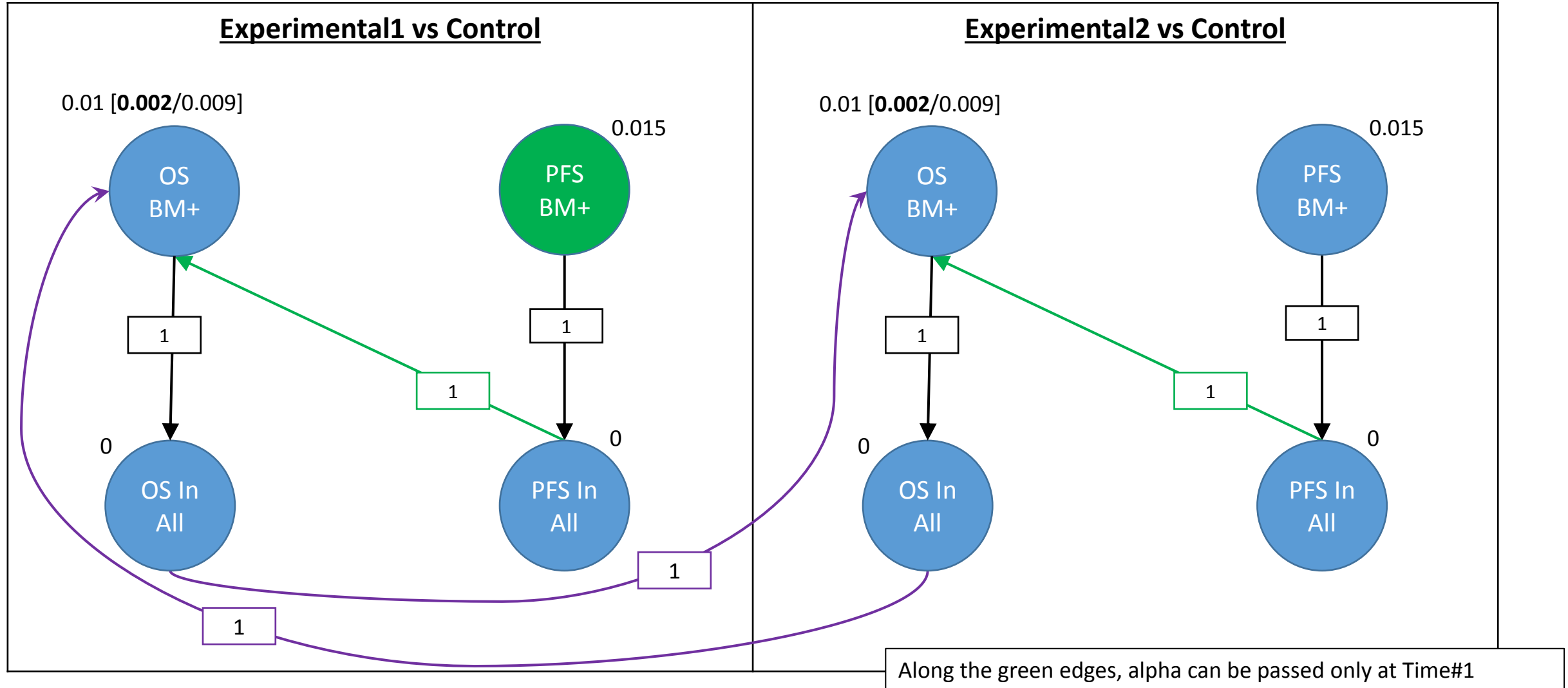
Experimental2 vs Control



Along the green edges, alpha can be passed only at Time1

Illustration of Sequentially Rejective Procedure: Step 1a

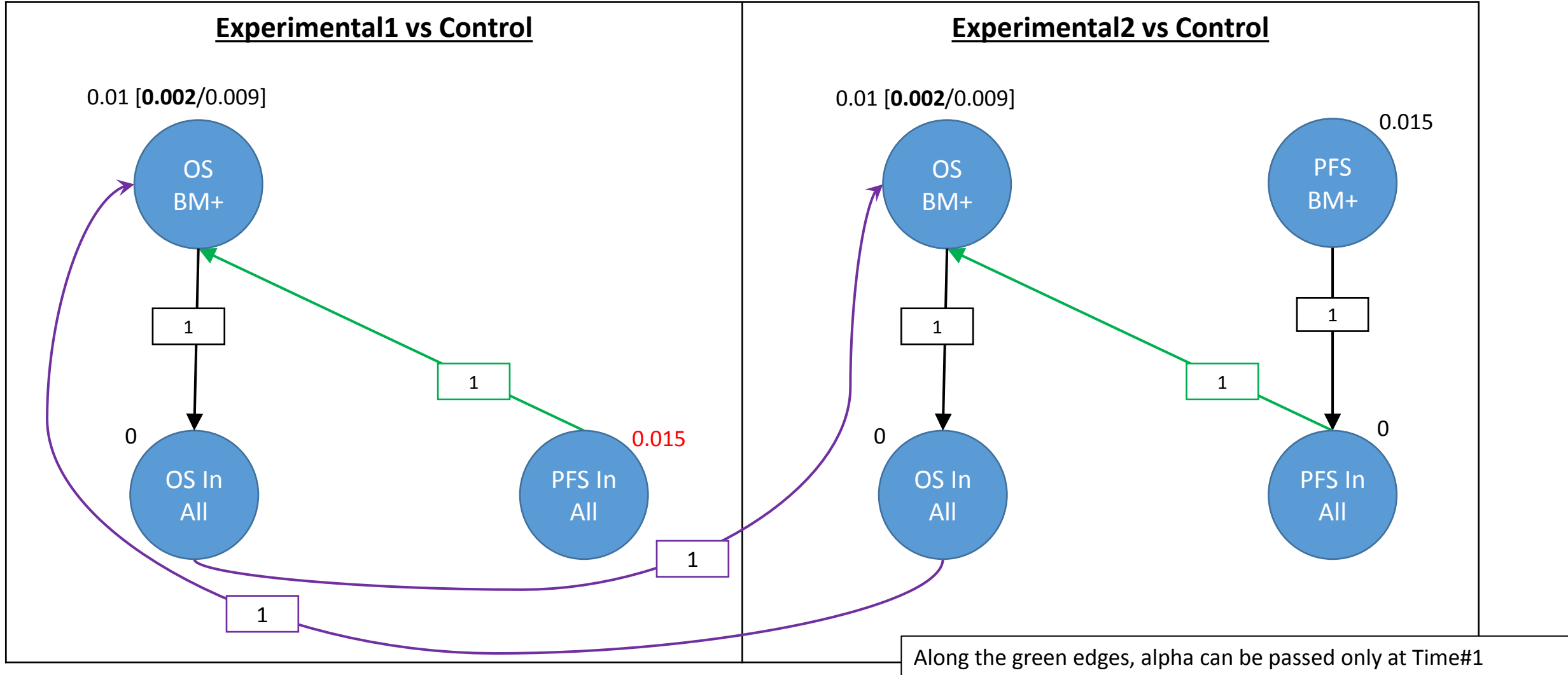
Time 1



For OS, information fraction assumed: BM+ 70% (O'Brien-Fleming); AC 75% (Pocock)

Illustration of Sequentially Rejective Procedure: Step 1b

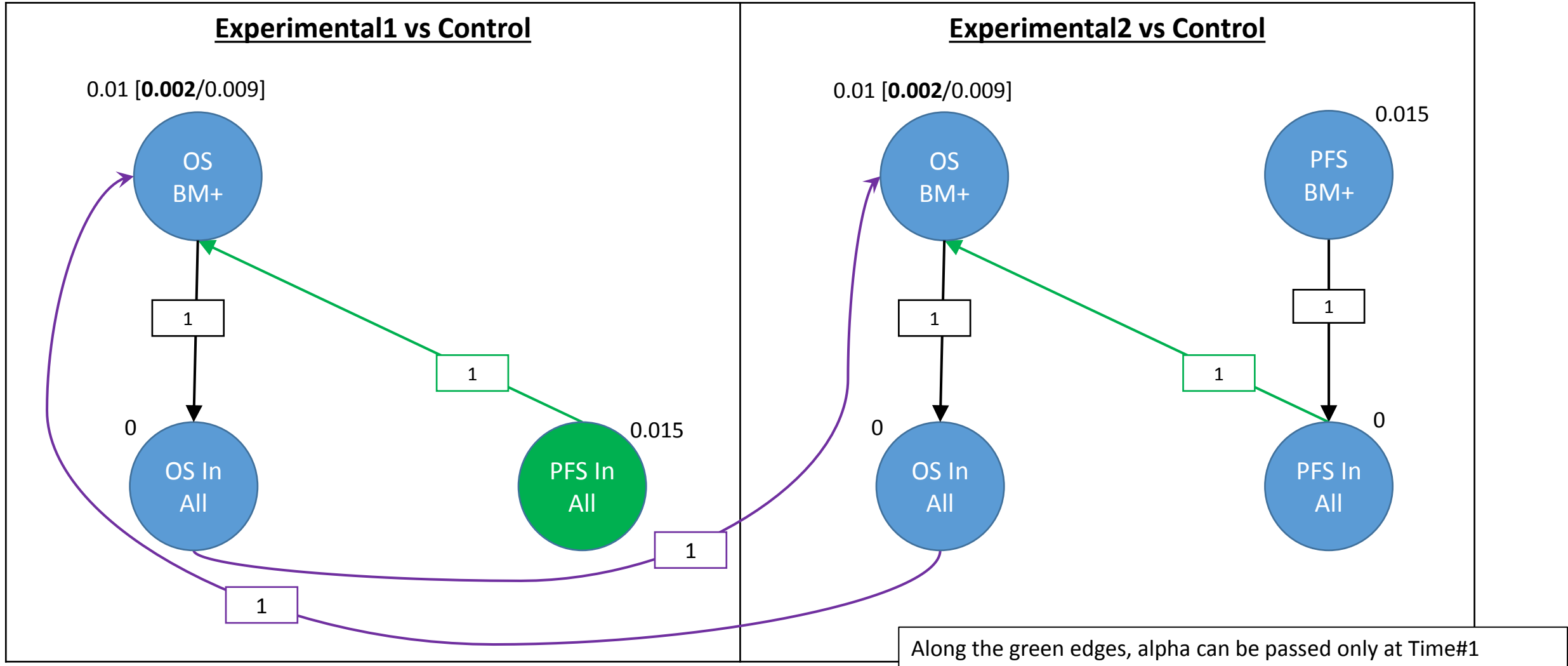
Time 1



For OS, information fraction assumed: BM+ 70% (O'Brien-Fleming); AC 75% (Pocock)

Illustration of Sequentially Rejective Procedure: Step 2a

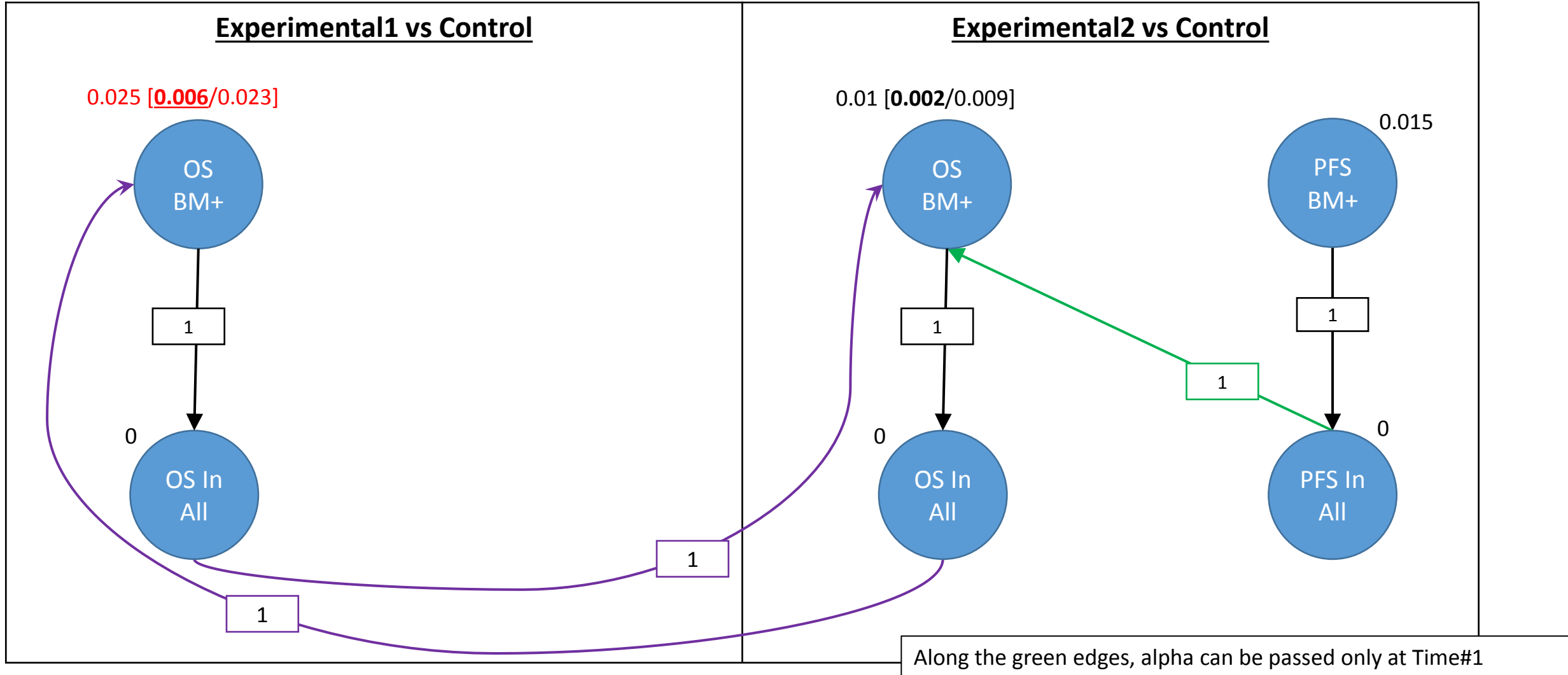
Time 1



For OS, information fraction assumed: BM+ 70% (O'Brien-Fleming); AC 75% (Pocock)

Illustration of Sequentially Rejective Procedure: Step 2b

Time 1



For OS, information fraction assumed: BM+ 70% (O'Brien-Fleming); AC 75% (Pocock)

Illustration of Sequentially Rejective Procedure: Step 3a

Time 2

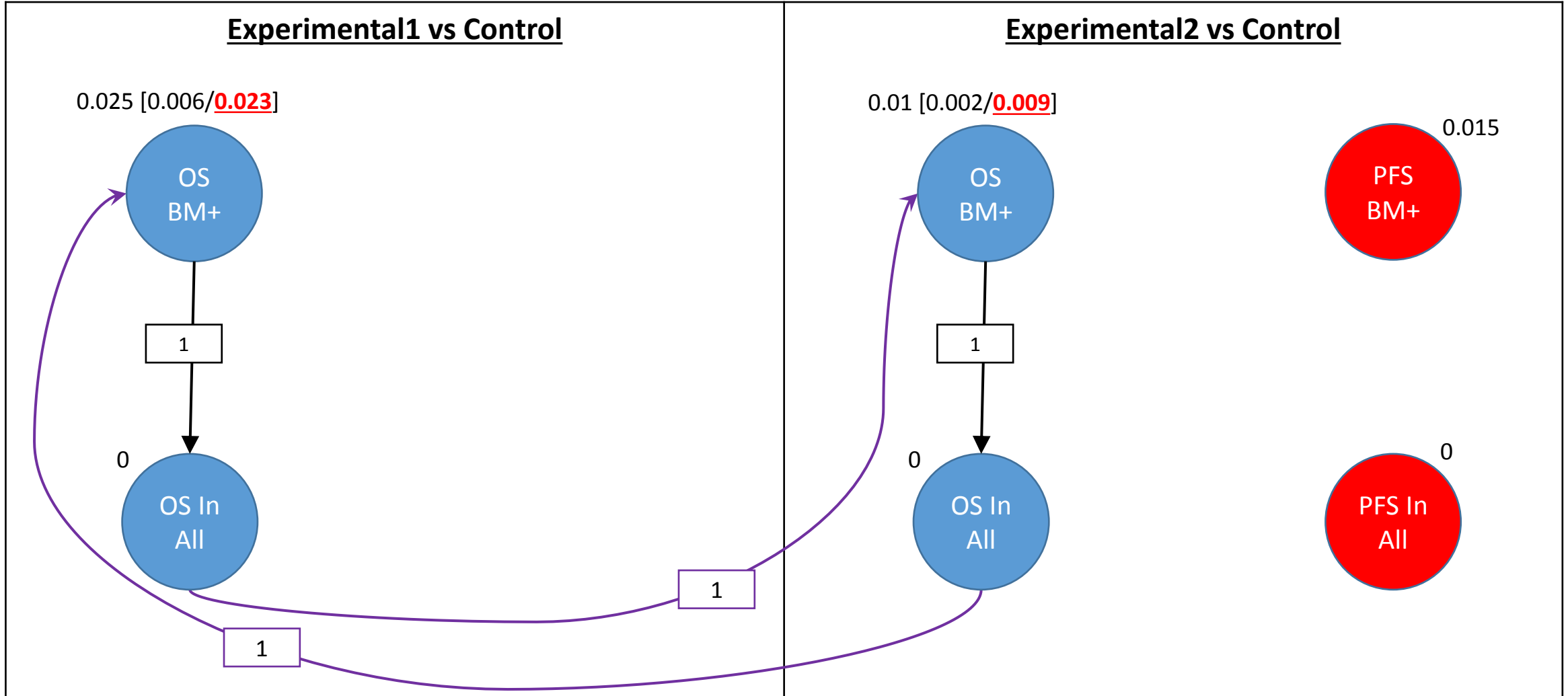


Illustration of Sequentially Rejective Procedure: Step 3b

Time 2

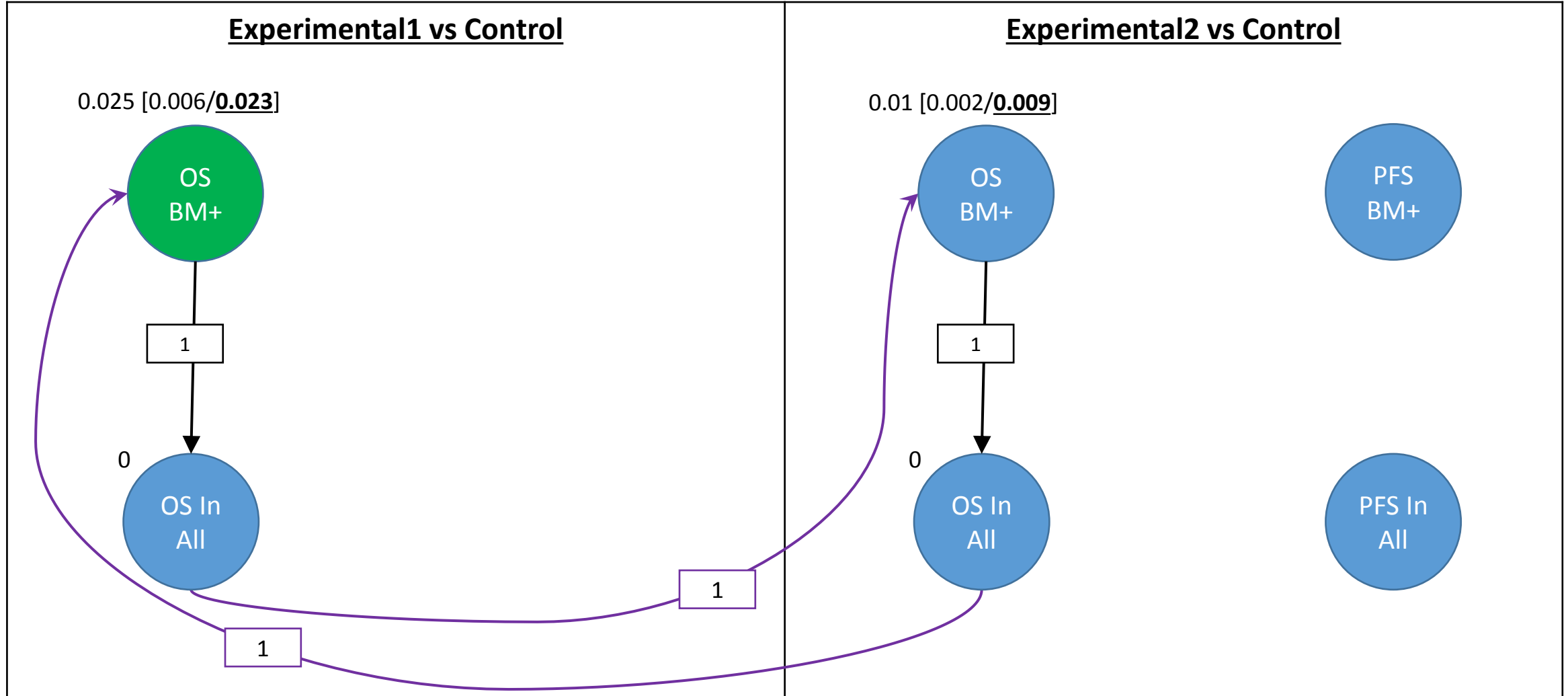
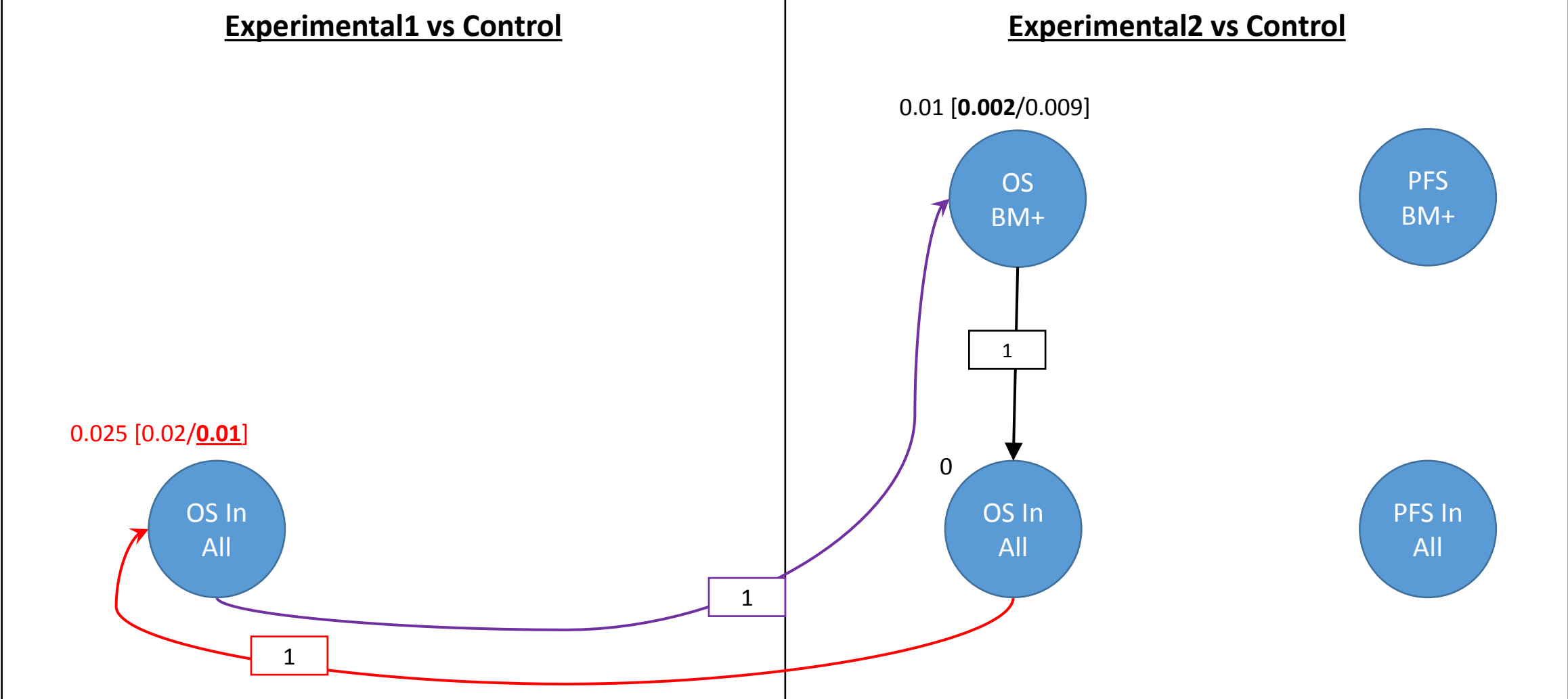


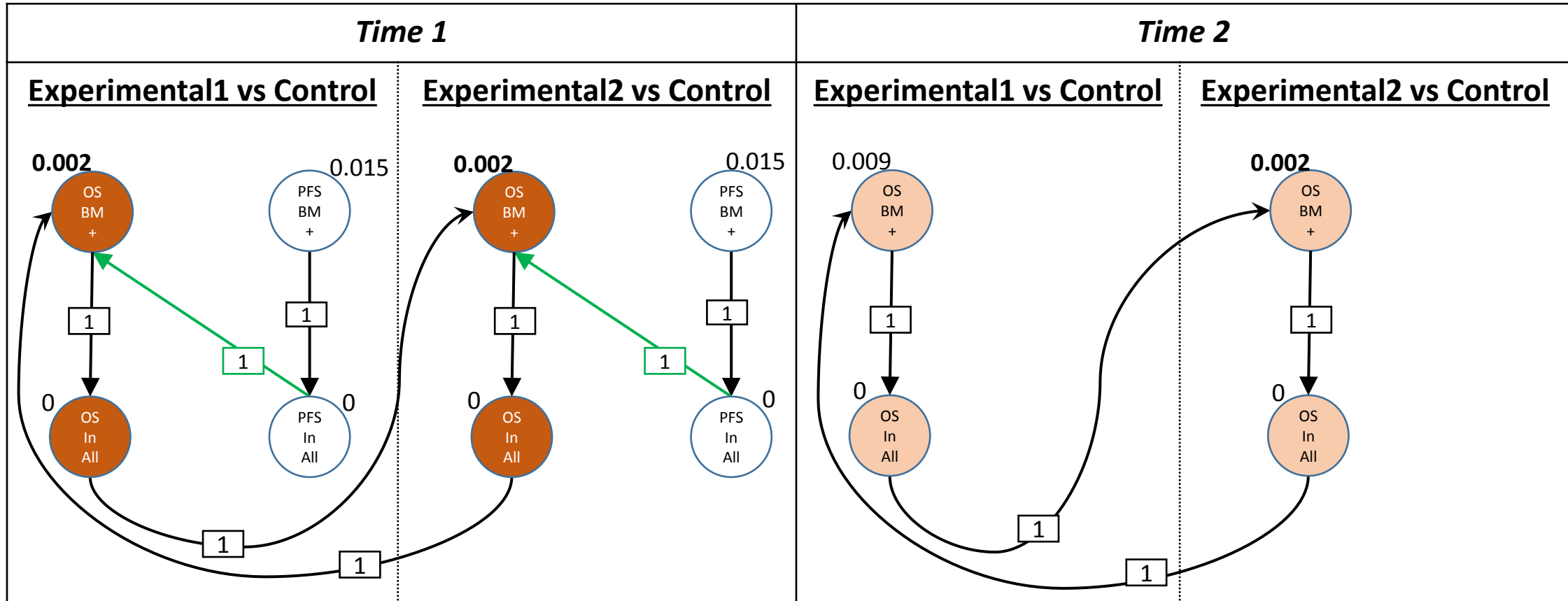
Illustration of Sequentially Rejective Procedure: Step 3c

Time 2



For OS, information fraction assumed: BM+ 70% (O'Brien-Fleming); AC 75% (Pocock)

Change in Power vs. Base Case (under Protocol Assumptions)



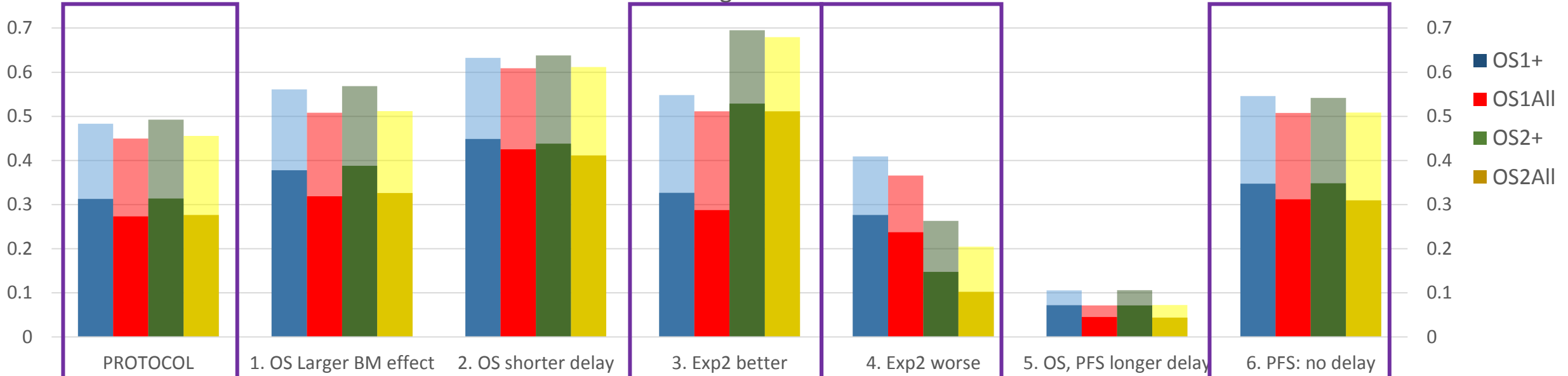
	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+17.1%	0.0%	+17.9%	0.0%
All	+17.6%	0.0%	+17.9%	0.0%

	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+6.6%		+5.7%	
All	+7.7%		+7.3%	

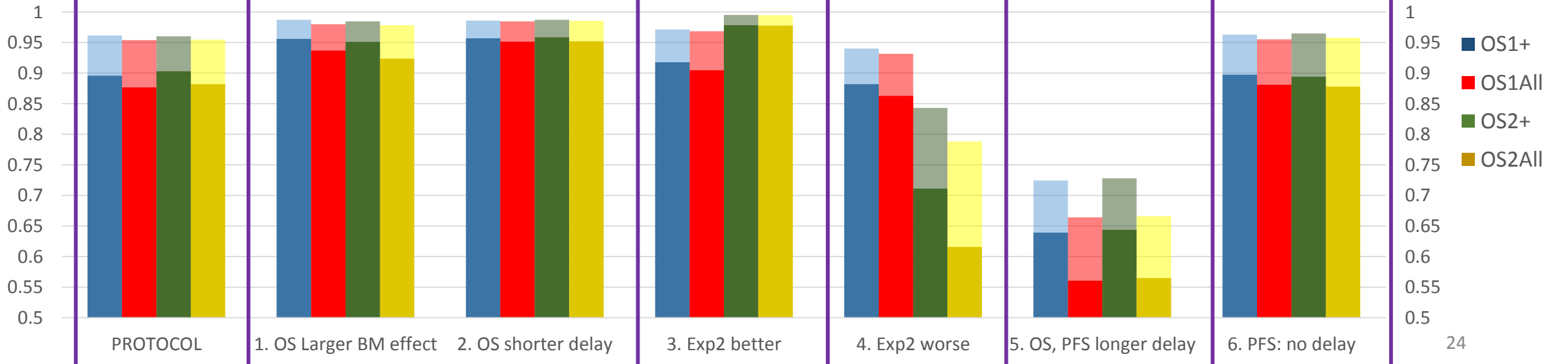
How Does the Power Gain Change under Different Assumptions?

<u>Endpoint Modified</u>	<u>Scenario</u>	<u>Details</u>
OS	#1 OS: larger difference in treatment effect by BM status	HR (for non-cure patients, and after delay) increased by 10% in the BM-, and decreased by 10% in the BM+ population (while keeping HR for all randomized the same). (a)
	#2 OS: shorter delay	10% shorter delay in all 4 OS endpoints (a)
OS, PFS	#3 OS, PFS: Exp2 better	OS and PFS treatment effect (delay, cure rate, HR in non-cure patients) improved by 10% in Experimental2 arm
	#4 OS, PFS: Exp2 worse	OS and PFS treatment effect (delay, cure rate, HR in non-cure patients) worsened by 10% in Experimental2 arm
	#5 OS, PFS: longer delay	OS delay increased with 50%, PFS delay with 300%
PFS	#6 PFS: no delay	No delay in any of the 4 PFS endpoints
Time of analysis is generated based on # of events in all 3 arms.		

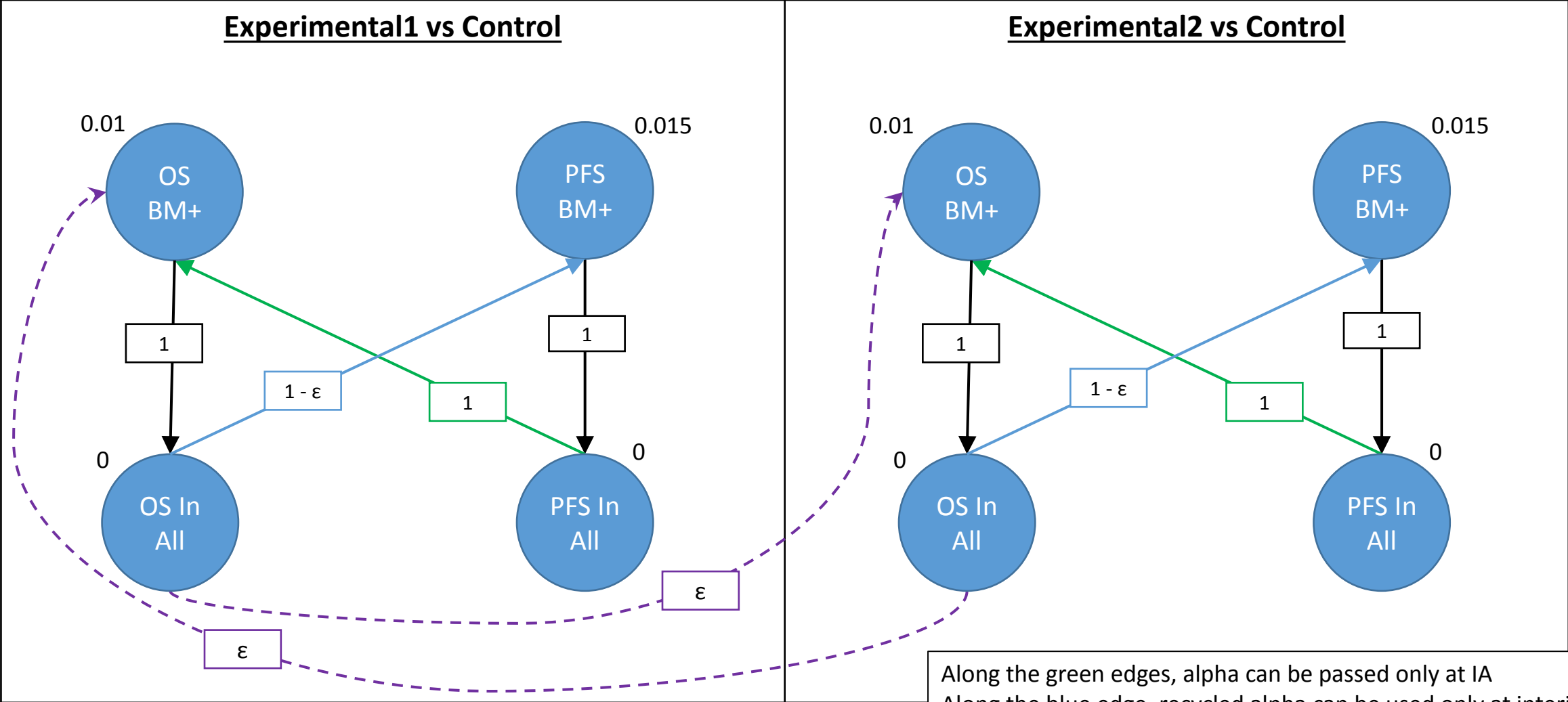
Benefit of Additional Edges in Statistical Power of OS: Time1



Benefit of Additional Edges in Statistical Power of OS: Time2



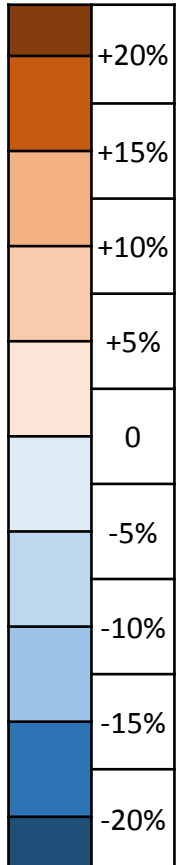
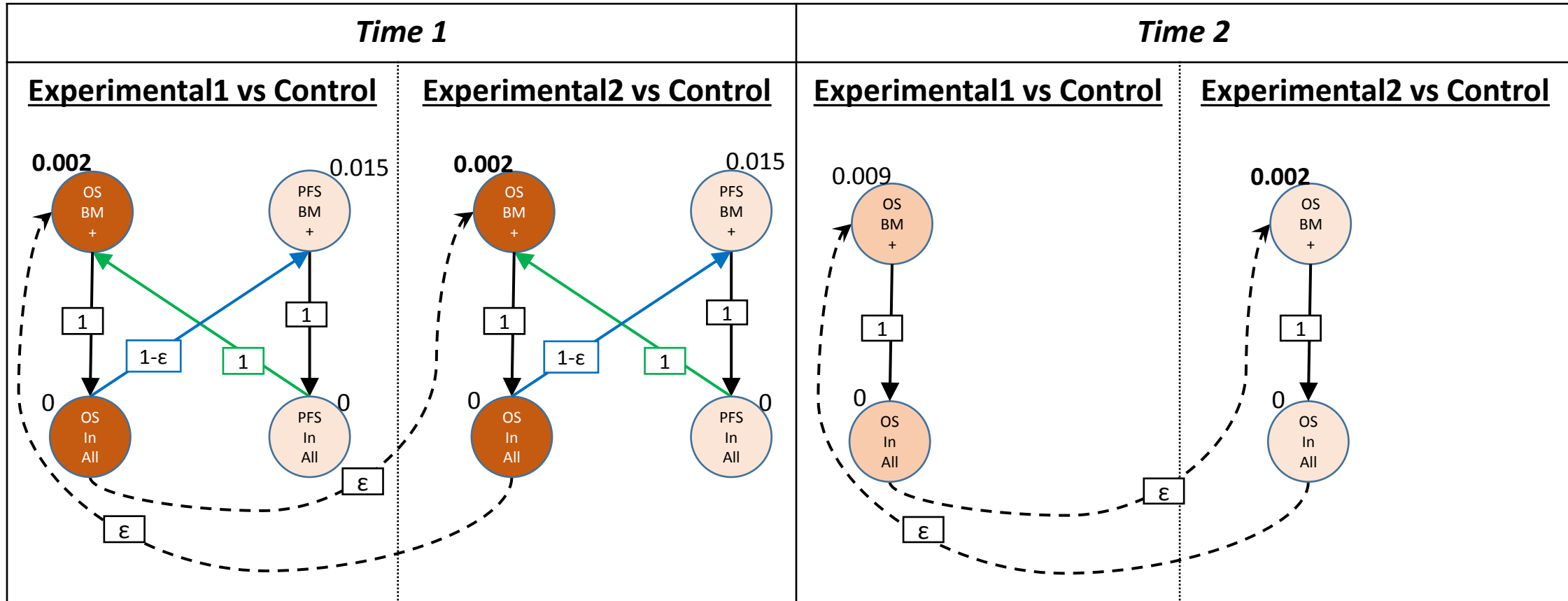
Alternative#1: Secure All Endpoints within Comparison



Along the green edges, alpha can be passed only at IA
 Along the blue edge, recycled alpha can be used only at interim
 Along the dashed edges, alpha is passed only if for the given vertex only dotted outgoing edges remain

Alternative#1:

Change in Power vs. Base Case (under Protocol Assumptions)

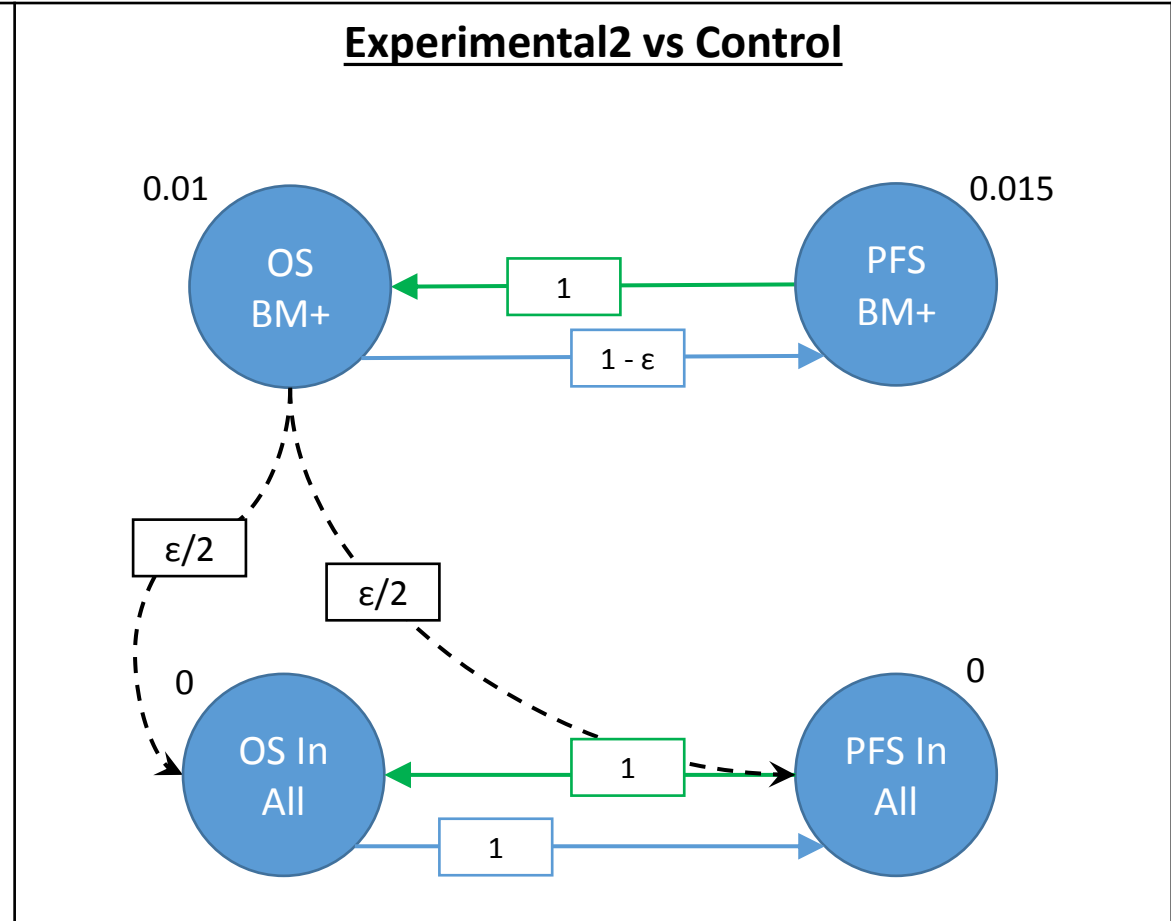
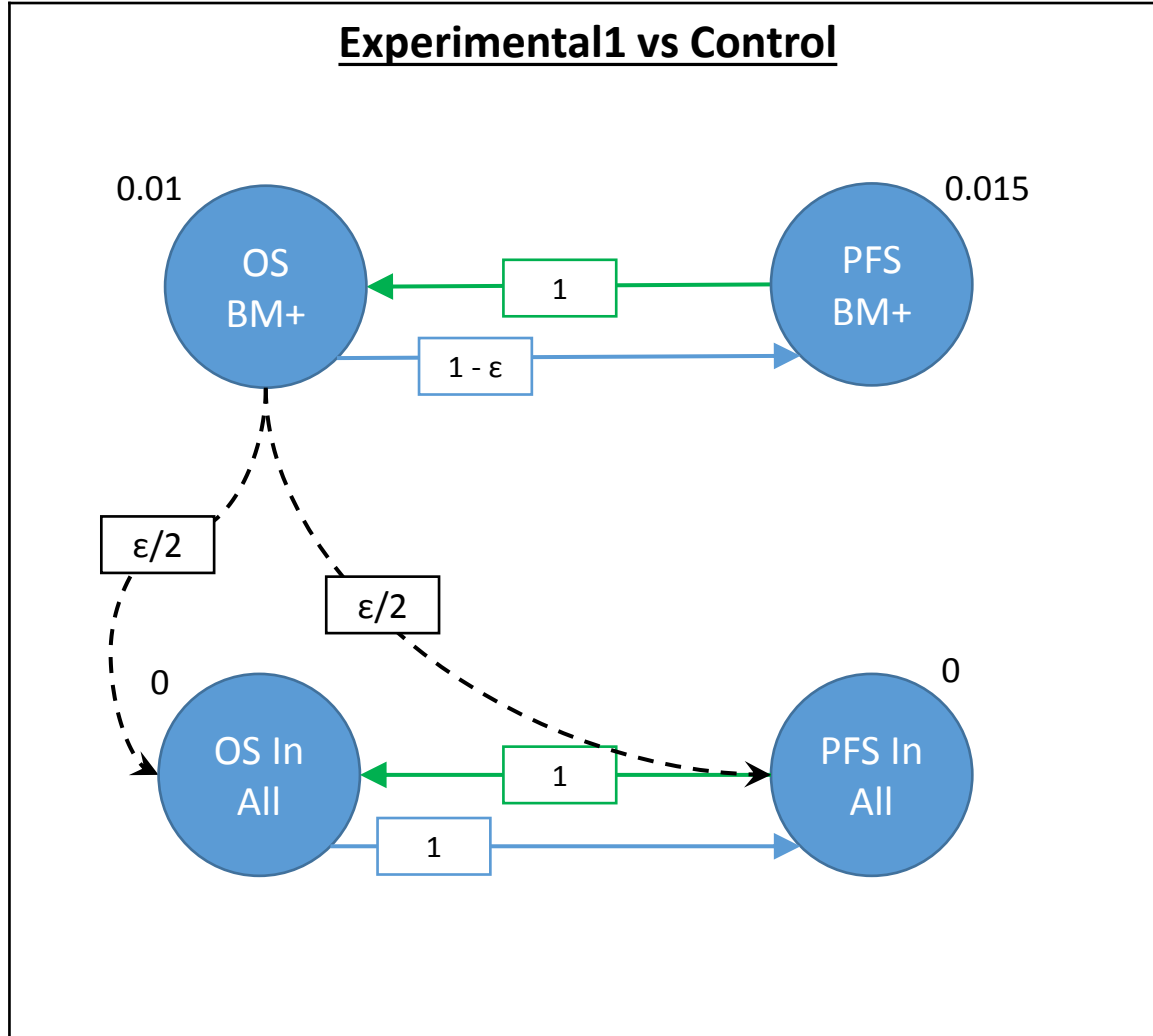


	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 17.0%	+ 1.4%	+ 17.8%	+ 1.7%
All	+ 17.6%	+ 1.6%	+ 17.8%	+ 2.0%

	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 6.4%		+ 5.6%	
All	+ 7.5%		+ 7.1%	

For OS, information fraction assumed: BM+ 70%; AC 75%

Alternative#2: Giving BM+ Maximum Chance

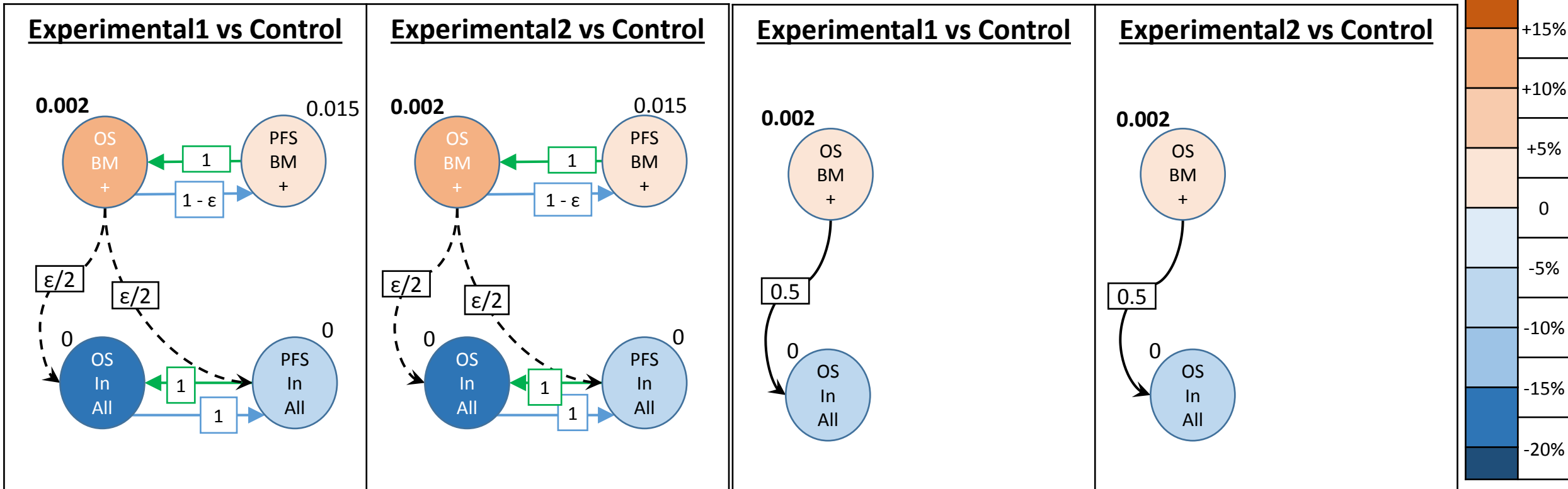


Along the green edges, alpha can be passed only at IA
 Along the blue edge, recycled alpha can be used only at interim
 Along the dotted edges, alpha is passed only if for the given vertex only dotted outgoing edges remain

Alternative#2: Change in Power vs. Base Case (under Protocol Assumptions)

Time 1

Time 2



	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 13.5%	+ 0.8%	+ 13.8%	+ 0.8%
All	- 19.7%	- 8.0%	- 1.9%	- 7.9%

	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 3.7%		+ 3.2%	
All	- 6.7%		- 6.8%	

4. Concluding Remarks

Concluding Remarks: MTP

- Limitation of the results on statistical power:
 - OS and PFS were generated independently
 - Identifying the testing strategy for an actual clinical trial always requires the input of several functions, and the power results represent only one aspect of these considerations.
- Using graphical approach may result in overpowered endpoints. However, this still may be desirable, reasons:
 - Protocol assumptions are complicated => just one detail off can result in power drop (even in the presence of meaningful treatment effect)
 - At an interim analysis, probability of crossing the efficacy boundary increases => gives larger chance for earlier read-out

Thank you for your attention!

References

- Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine* 2009; 28: 586-604.
- Burman, C.-F., Sonesson, C. and Guilbaud, O. (2009). A recycling framework for the construction of Bonferroni-based multiple tests. *Statistics in Medicine* **28**, 739–761.
- Bretz F, Posch M, Glimm E, Klinglmueller F, Maurer W, Rohmeyer K. Graphical approaches for multiple comparison procedures using weighted Bonferroni, Simes, or parametric tests. *Biometrical Journal* 2011; 53(6): 894-913.
- Ye, Y., Li, A., Liu, L. and Yao, B. (2013). A group sequential Holm procedure with multiple primary endpoints. *Statistics in Medicine*, 32, 1112-1124
- Maurer, W., and Bretz, F. (2013a), “Graphical Multiple Test Procedures With Memory,” *Statistics in Medicine*, DOI: 10.1002/sim.5711.
- Willi MAURER and Frank BRETZ: Multiple Testing in Group Sequential Trials Using Graphical Approaches. *American Statistical Association Statistics in Biopharmaceutical Research* November 2013, Vol.5, No.4
- Xi D, Tamhane AC. Allocating recycled significance levels in group sequential procedures for multiple endpoints. *Biom J.* 2015 Jan;57(1):90-107.
- Tamhane, A. C., Mehta, C. R., and Liu, L. (2010). Testing a primary and a secondary endpoint in a group sequential design. *Biometrics* **66**, 1174–1184.

Back-up

Basic Parameters of the Study Design

Table 1. Basic Parameters of the Study Design

	Overall Survival			Progression-Free Survival		
	BM+	BM-	All Rand.	BM+	BM-	All Rand.
# randomized, Total (per comparison)	≈ 480 (≈ 320) (c)	≈ 720 (≈ 480) (c)	1200 (800) (fixed)	As OS	As OS	As OS
Significance level (a)	0.01			0.015		
Delayed Effect (b)	4m	5m		1m	1m	
Cure Rate in Experimental (b)	15%	10%		0	0	
HR after Delay Effect (b) (d)	0.6	0.75		0.56	0.75	
Overall HR (b) Time1/ Time2	0.70 / 0.61	0.84 / 0.74	0.77 / 0.69	0.63	0.84	0.72
Median, control/experimental	8m / 12m	9m / 12m	9m / 12m	4m / 6.4m	4m / 4.7m	4m / 5.5m
# events, Total, Time1/ Time2	≈ 285 / 385 (fixed)		≈ 693 / ≈ 963	380 (fixed)		≈ 989
# events per comparison, Time1/ Time2	≈ 198 / ≈ 266		≈ 478 / ≈ 661	≈ 261		≈ 672

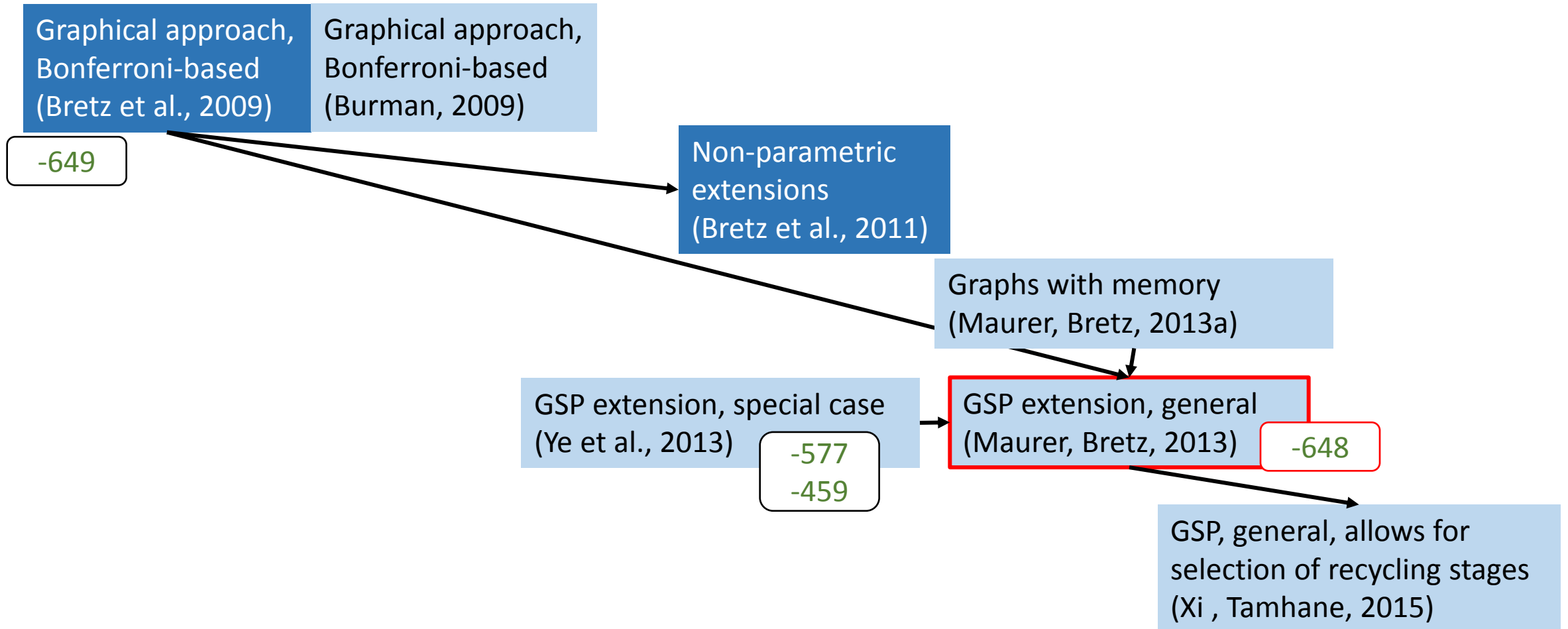
(a) Initially allocated endpoint-specific alpha; used for sample size calculation

(b) For each comparison with the two experimental arms

(c) Based on an expected BM+ proportion of 40%. (Total number randomized is fixed)

(d) As assumed for non-cure patients.

Literature Overview



GSP = group-sequential procedure
Dark blue boxes: referenced by FDA guidance

Multiplicity: History of Regulatory Guidances

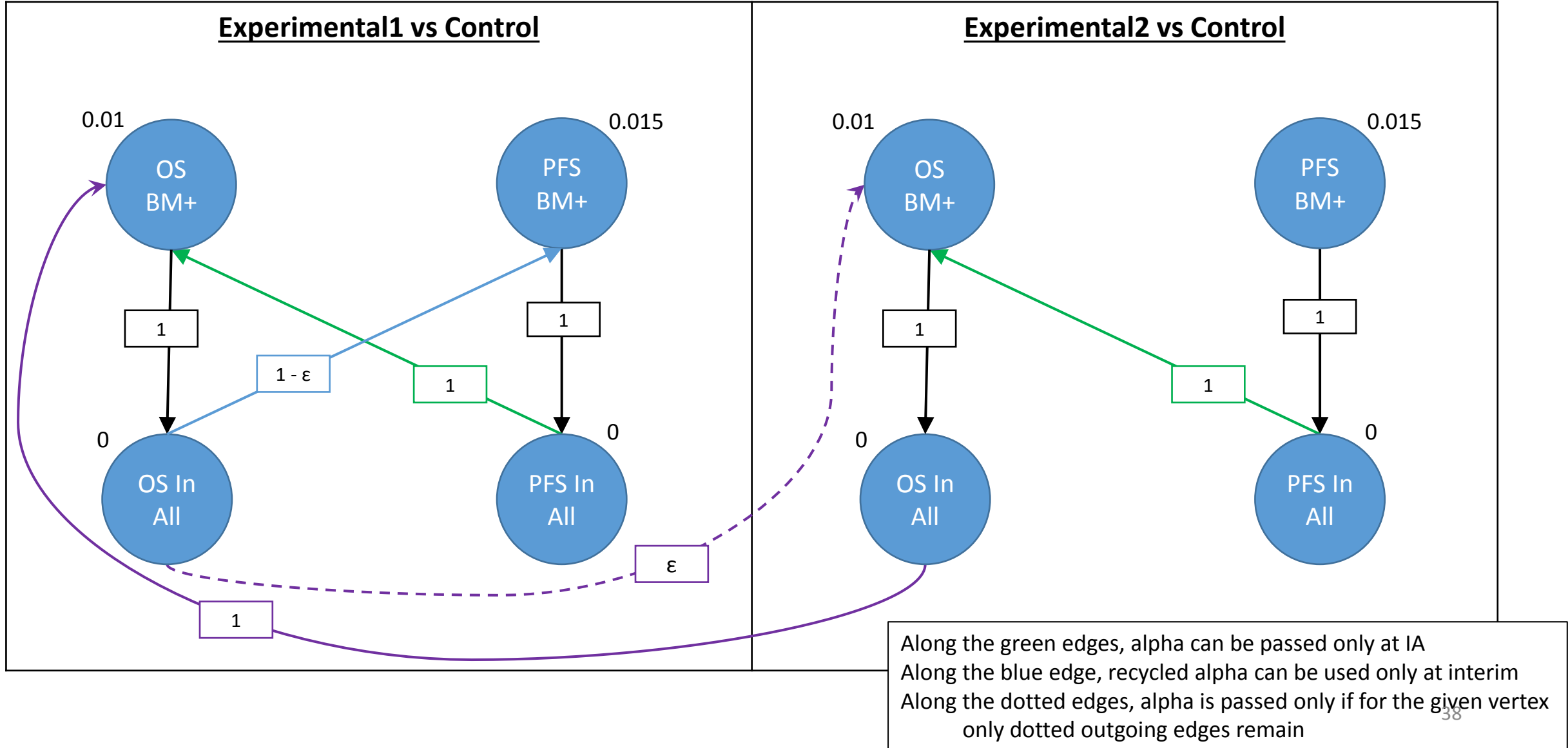
- E9: Statistical Principles for Clinical Trials (1998; Section 5.6)
 - Pre-specification of MTP
- **EMA: Points to Consider on Multiplicity Issues in Clinical Trials (2002)**
 - Strong control of FWER
- EMA: Concept paper on the need for a guideline on multiplicity issues in clinical trials (Draft; 2012) => Nov2012: Workshop
- **EMA: Guideline on multiplicity issues in clinical trials**
- **FDA: Guidance document on multiplicity issues in clinical trials**
- PMDA: no multiplicity guidance (“MTPs do not necessarily match with labelling”)

DRAFT
Released for public
consultation in 2017

Multiplicity: Latest Regulatory Guidance's (2017)

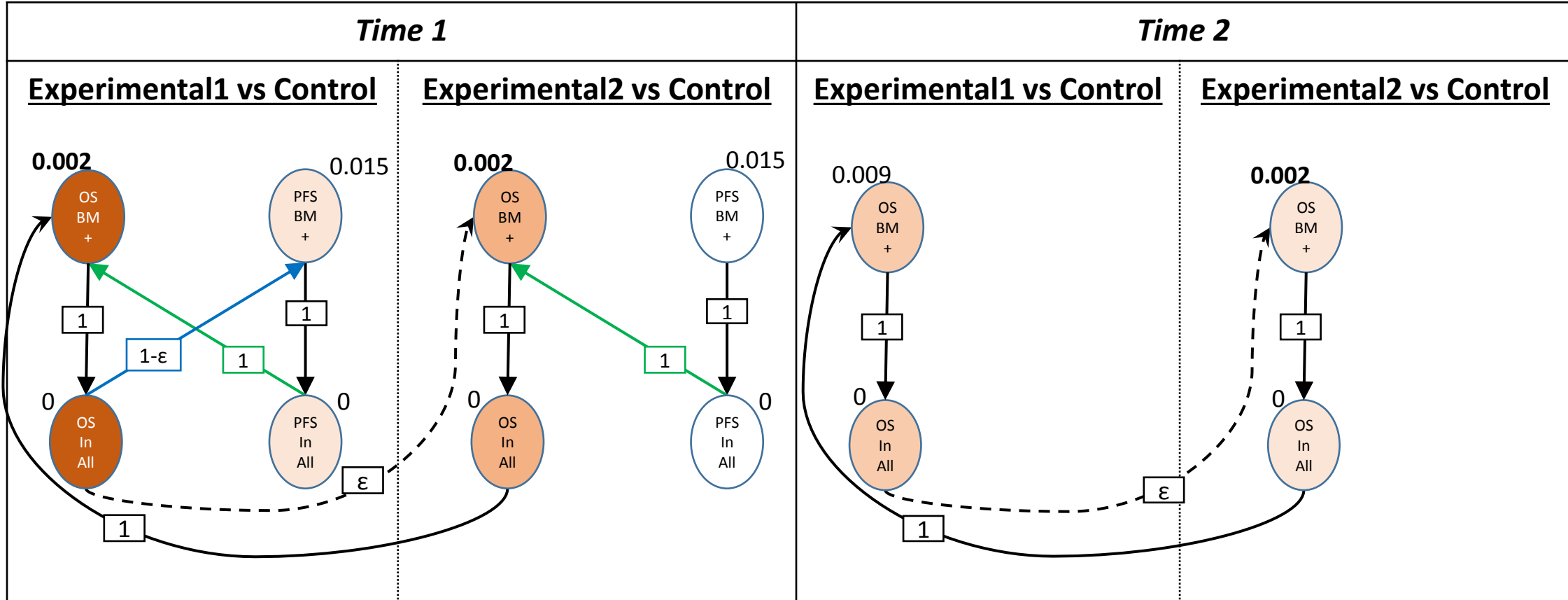
	EMA: Guideline on multiplicity issues in clinical trials	FDA: Guidance document on multiplicity issues in clinical trials
Principles	<ul style="list-style-type: none"> • Family-wise strong control • Pre-specification of multiplicity adjustment 	
Secondary Endpoints	<ul style="list-style-type: none"> • Secondary endpoints expressing supportive evidence: needs no MTP • Secondary endpoints that may become base for additional claims: need MTP <i>(Section 6)</i>	Distinction between different types of secondary endpoints is not clear: all require MTP (<i>Section IIIA</i>)
Scope	Similar (not the same)	
Subgroups	<i>Section 7</i> (also own guidance)	Excluded (another guidance)
Estimation and descriptive statistics	<i>Section 10</i>	Not included
Graphical approach	Not mentioned	Explicitly encouraged

Alternative#4: Importance of Experimental1 PFS



Alternative#4:

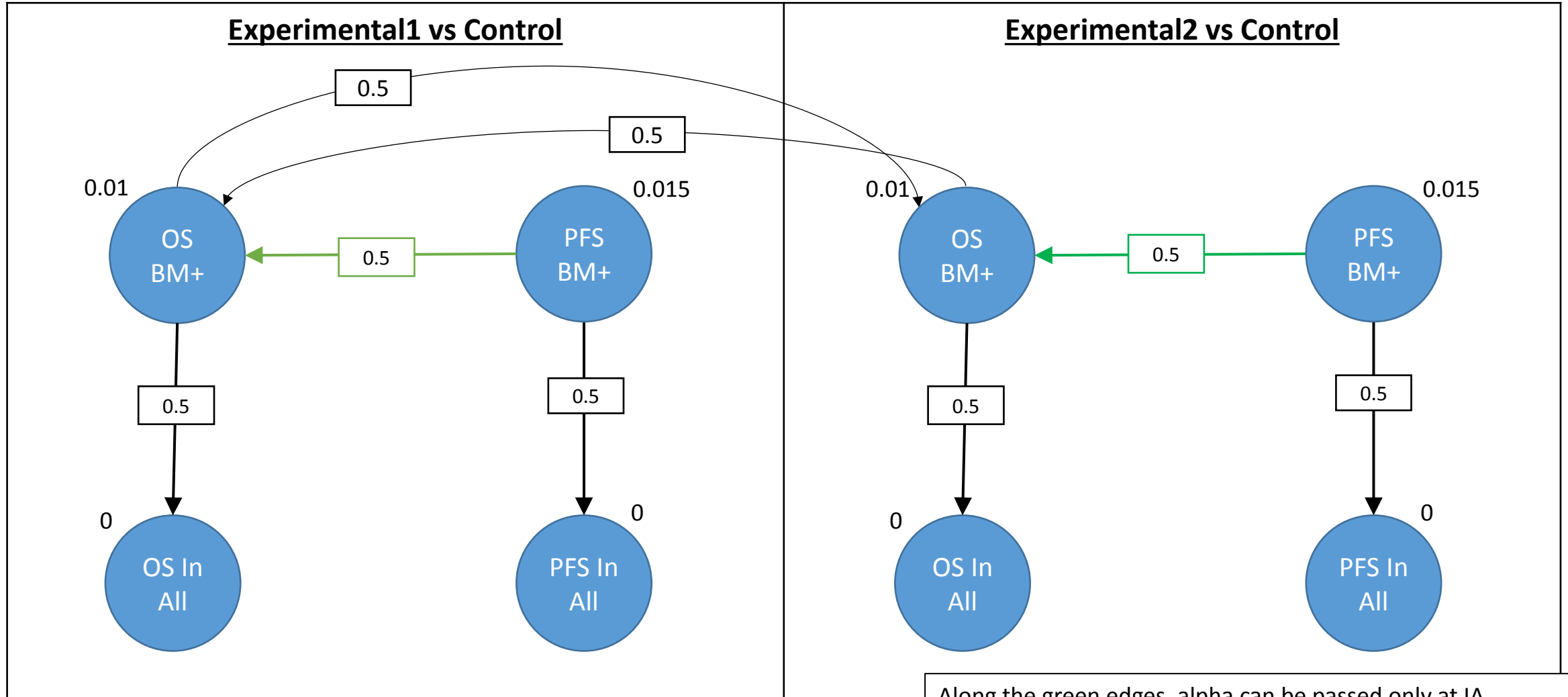
Change in Power vs. Base Case (under Protocol Assumptions)



	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 16.6%	+ 1.3%	+ 13.1%	0
AC	+ 16.2%	+ 1.9%	+ 11.6%	0

	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 5.6%		+ 3.0%	
AC	+ 6.9%		+ 4.0%	

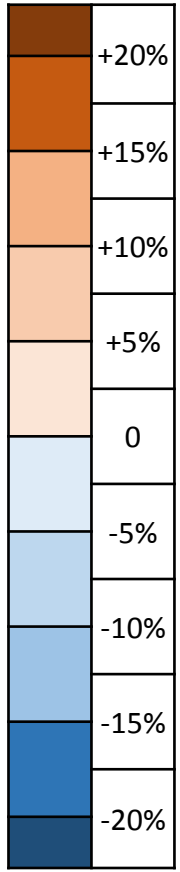
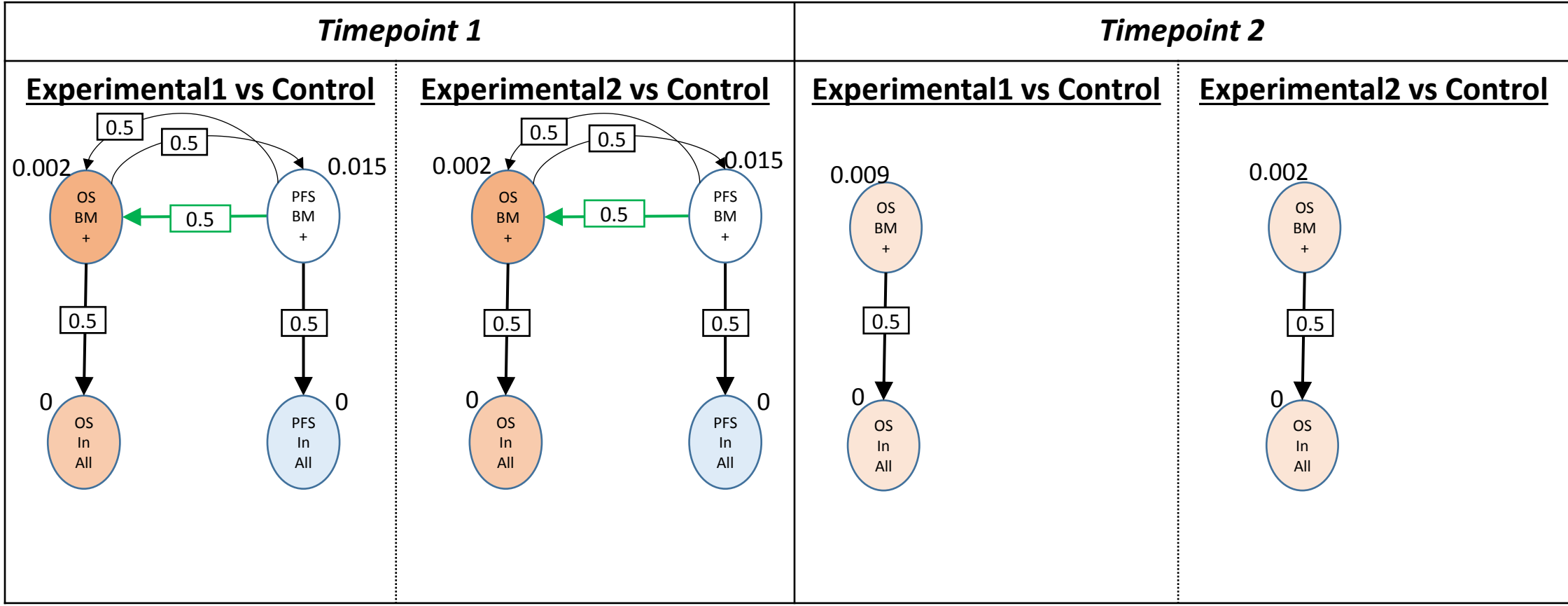
Alternative#3: Importance of both OS BM+



Along the green edges, alpha can be passed only at IA
Along the blue edge, recycled alpha can be used only at interim
Along the dotted edges, alpha is passed only if for the given vertex only dotted outgoing edges remain

Alternative#3:

Change in Power vs. Base Case (under Protocol Assumptions)



	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 10.4%	0.0%	+ 10.5%	0.0%
All	+ 8.7%	- 1.4%	+ 8.4%	- 1.6%

	Exp1 vs. Control		Exp2 vs. Control	
	OS	PFS	OS	PFS
BM+	+ 4.6%	N/A	+ 4.0%	N/A
All	+ 4.9%	N/A	+ 4.2%	N/A

For OS, information fraction assumed: BM+ 70%; AC 75%

B) Final OS Look: Adjusting Significance Levels

- The time of the final OS analysis will be determined by the number of BM+ OS events in the 3 arms or in the control arm.
 - As a result, the number of events for each OS comparison is a random variable by design.
 - Consequence:
 - Time 1: Assume none of the OS endpoints was rejected
 - Time 2: If the # of events for a given comparison is more (less) than expected => the information fraction and nominal significance level used at Time 1 was larger (smaller) than what we should have used
- ⇒ For the final look, critical values for OS will be re-calculated using the number of exactly observed events at the final analysis
- ⇒ The effect of adjustment is limited (always < 0.004 , and, in general, < 0.001)

Illustration: Effect of Final Alpha Adjustment

Experimental1 vs. Control: OS All

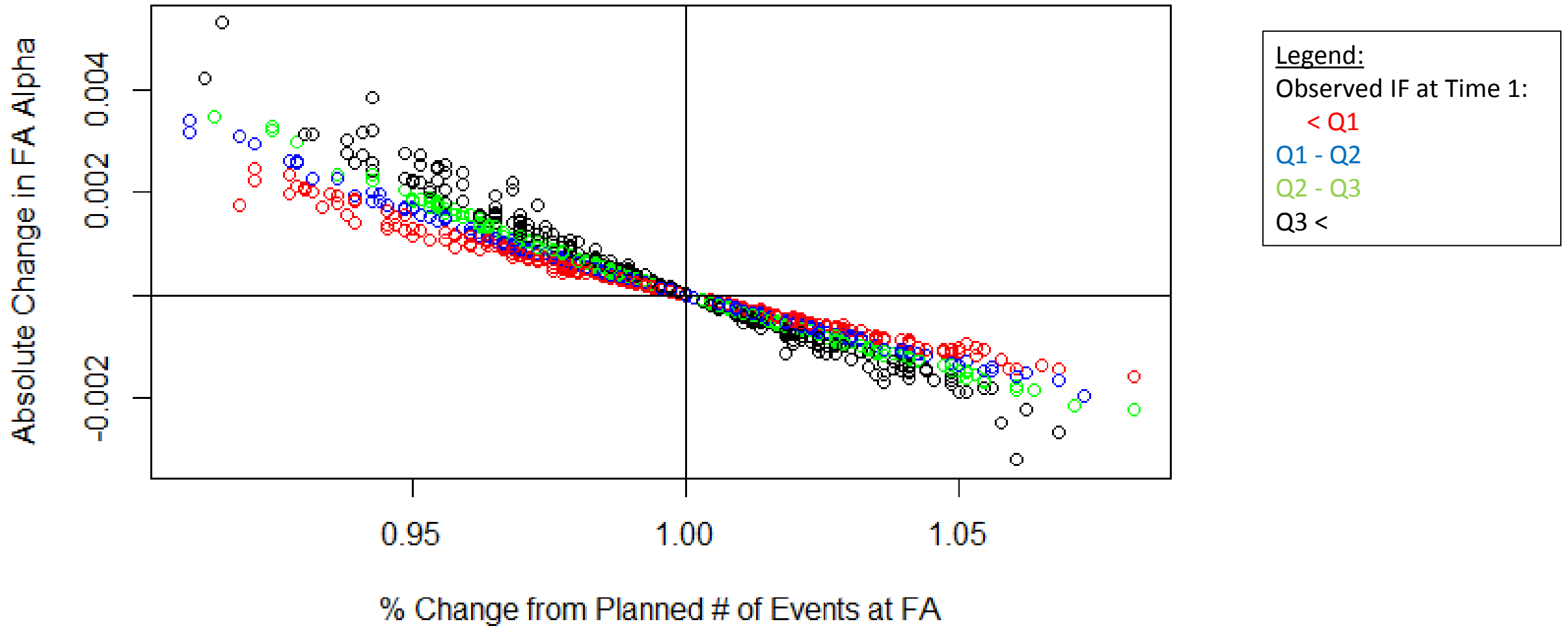


Illustration: Effect of Final Alpha Adjustment

Experimental1 vs. Control: OS All

