Type 1 Error Control

Type 1 error:

Saying there is something, when there is nothing.

Typical $\alpha = 0.05$

5 sigma level used in physics:

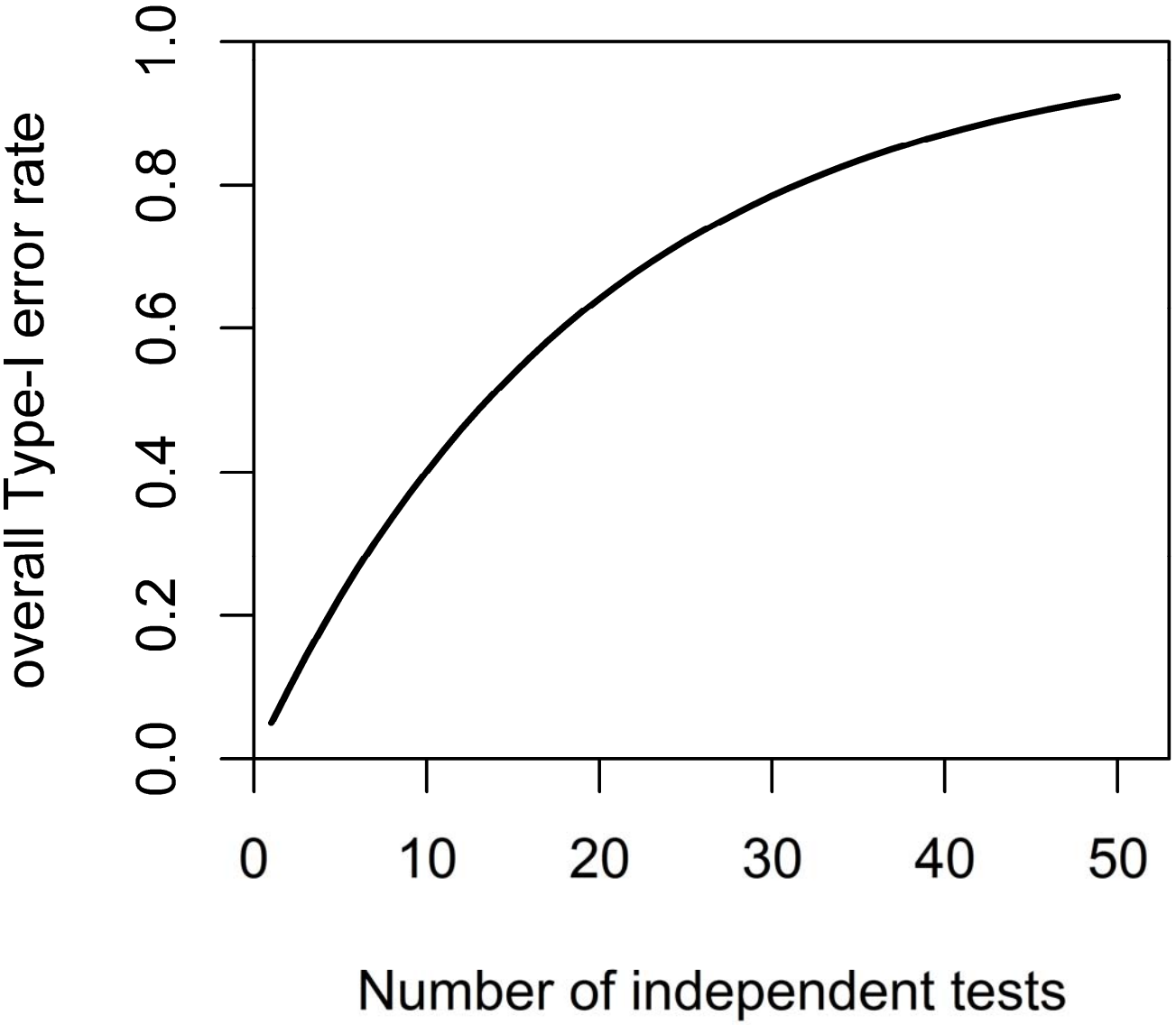
$\alpha = 0.0000003$

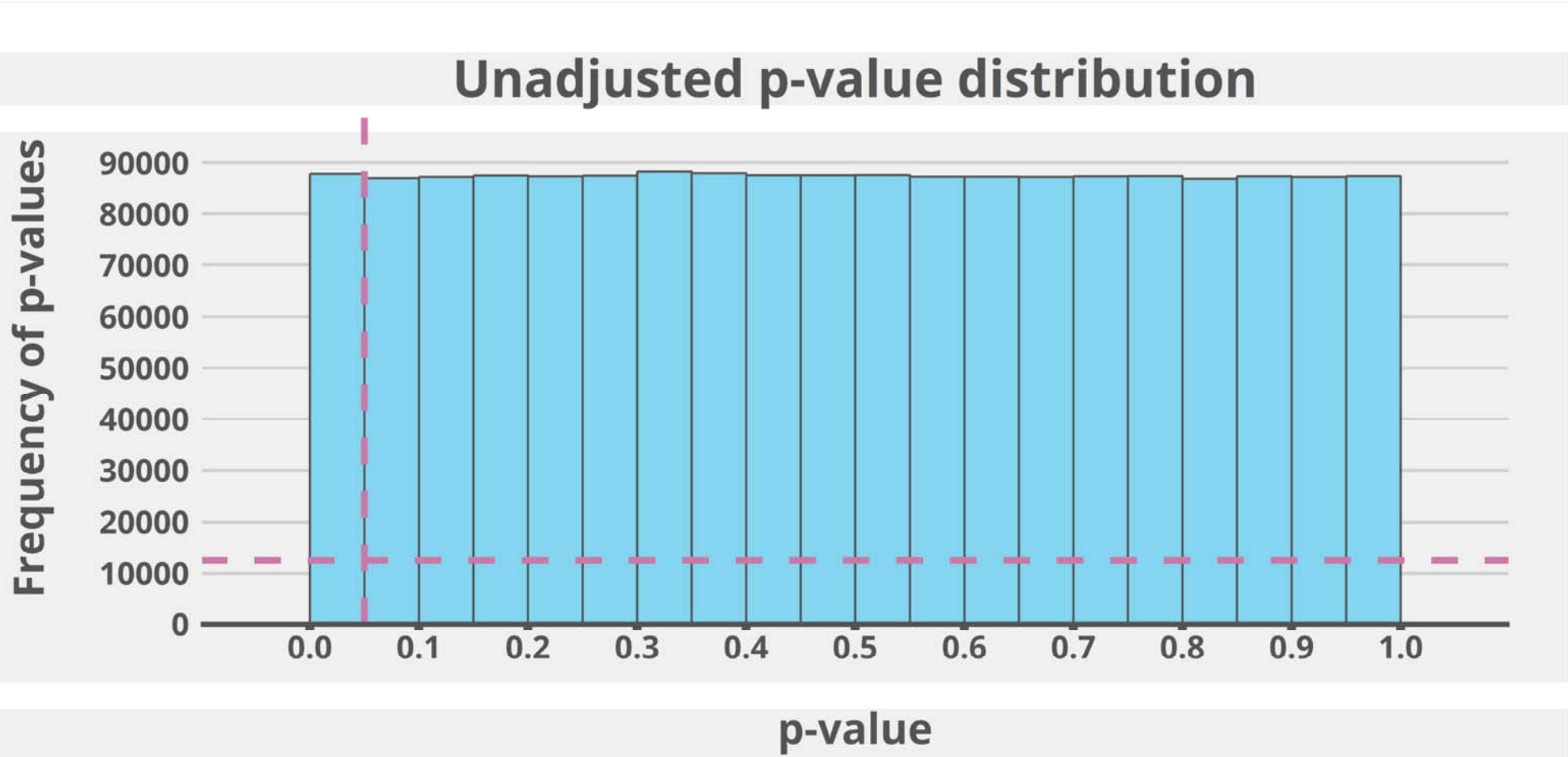
With multiple tests, the Type 1 error rate will inflate.

In a 2x2x2 ANOVA, there are 7 tests.

Type 1 error rate:

$1-(0.95)^{7}=30\%$


overall Type-I error rate
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50
Number of independent tests


Unadjusted p-value distribution
Frequency of p-values
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

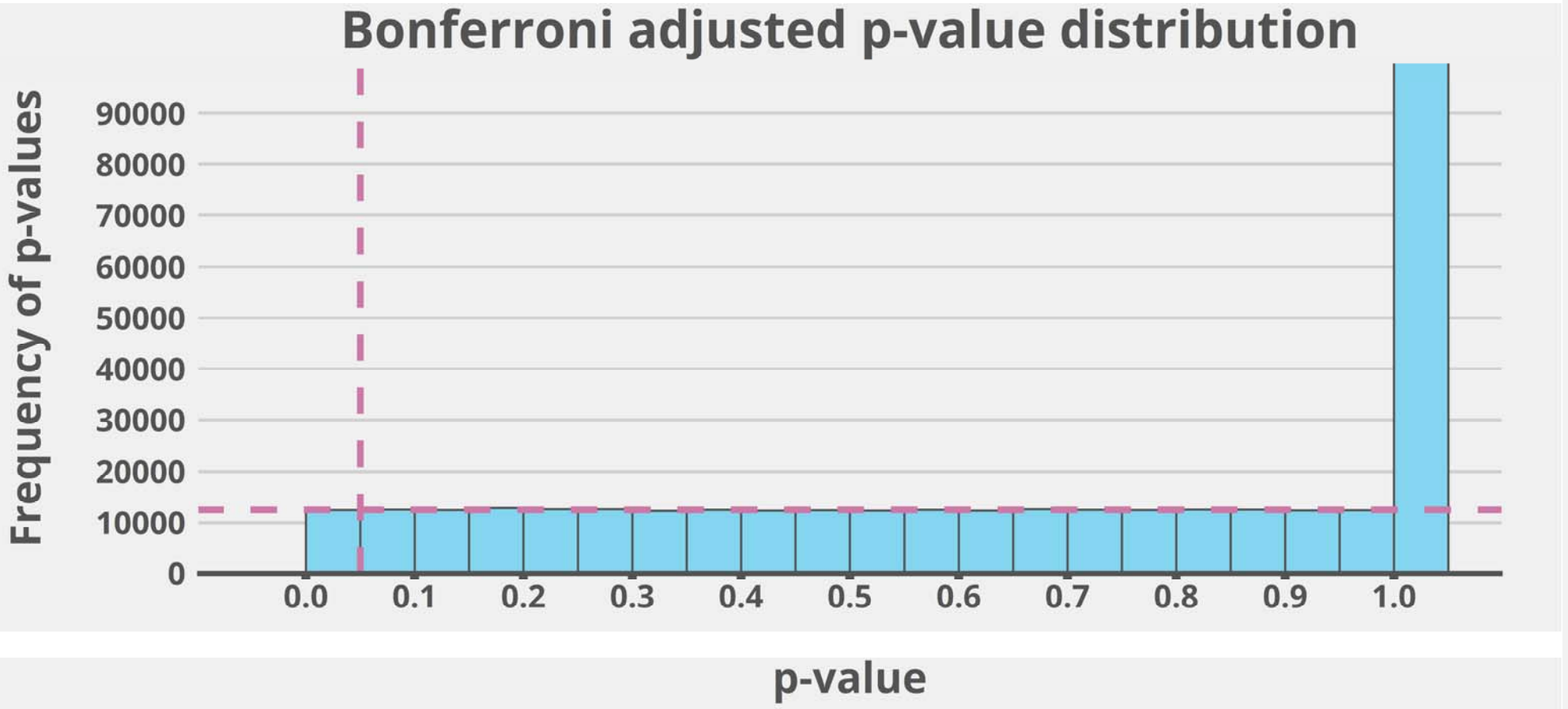
But you can control Type 1 error rate inflation due to multiple comparisons

Bonferroni Correction

(Actually, the Dunn Correction)

$$\frac{\alpha}{number\ of\ tests}$$

($p \times$ number of tests)


Bonferroni adjusted p-value distribution
Frequency of p-values
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

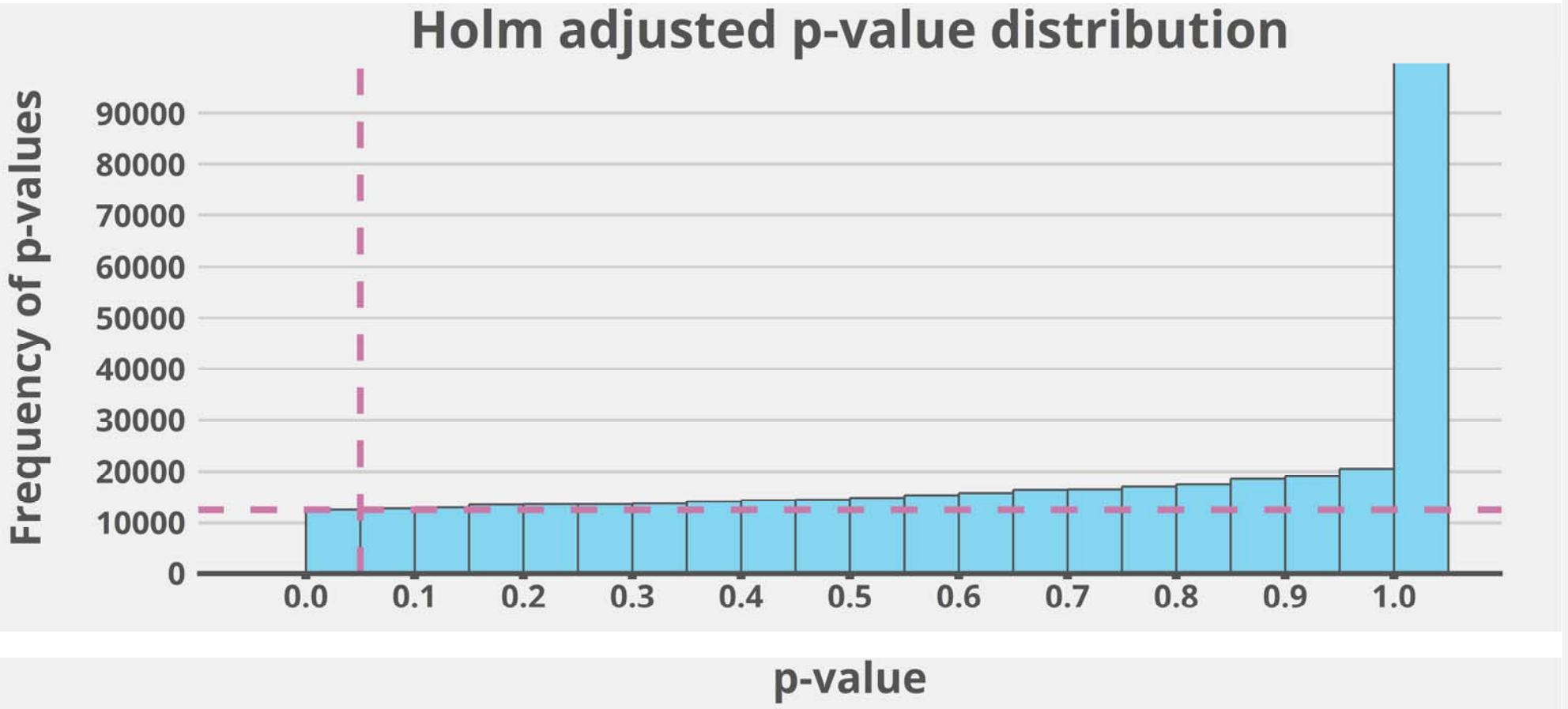
Familywise error rate:

Control for all tests which lead you to say there is something, when there is nothing


YOU DON'T NEED THAT BONFERRONI, GIRL.
THERE'S NO COMPARISON.

Holm correction

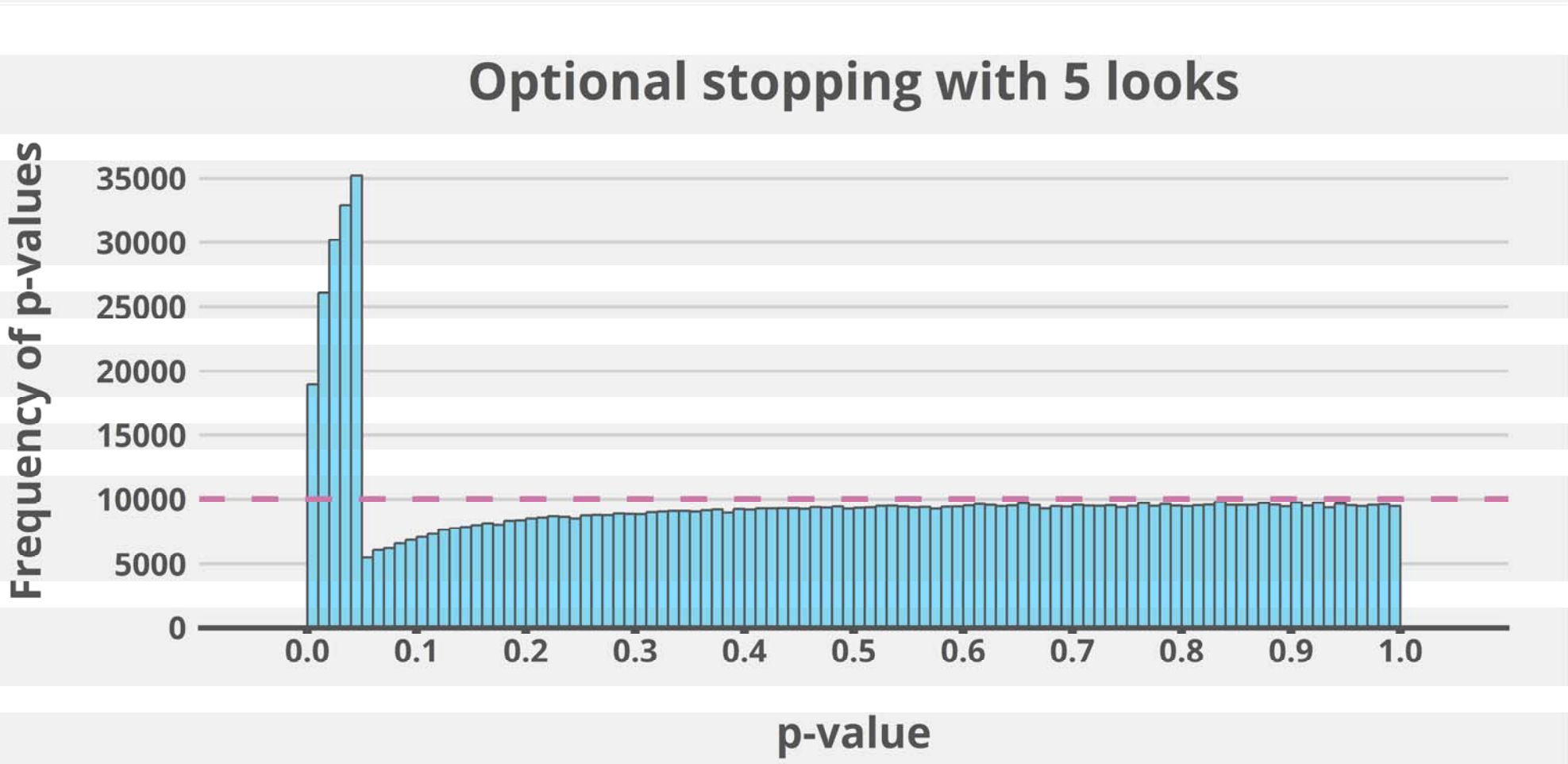
p-value	Inversed rank order	corrected *p*-value
0,001	4	0,004
0,002	3	0,006
0,012	2	0,024
0,032	1	0,032


Holm adjusted p-value distribution
Frequency of p-values
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

Optional stopping: Collecting data until $p < 0.05$ inflates the Type 1 error.

A user of NHST could **always** obtain a significant result through optional stopping.

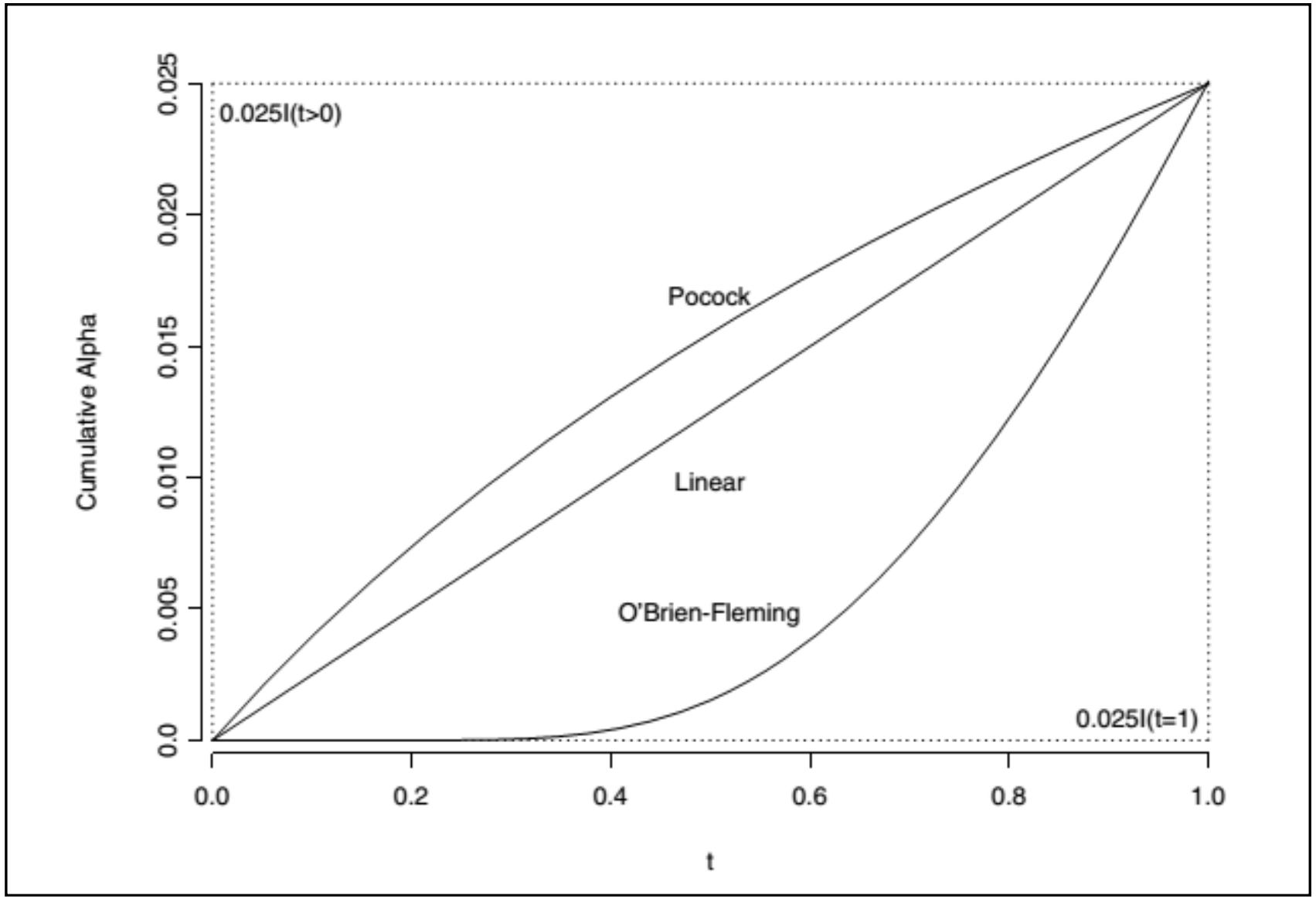
Wagenmakers, 2007


Optional stopping with 5 looks
Frequency of p-values
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

Sequential analysis controls Type 1 error rates (e.g., Pocock correction).

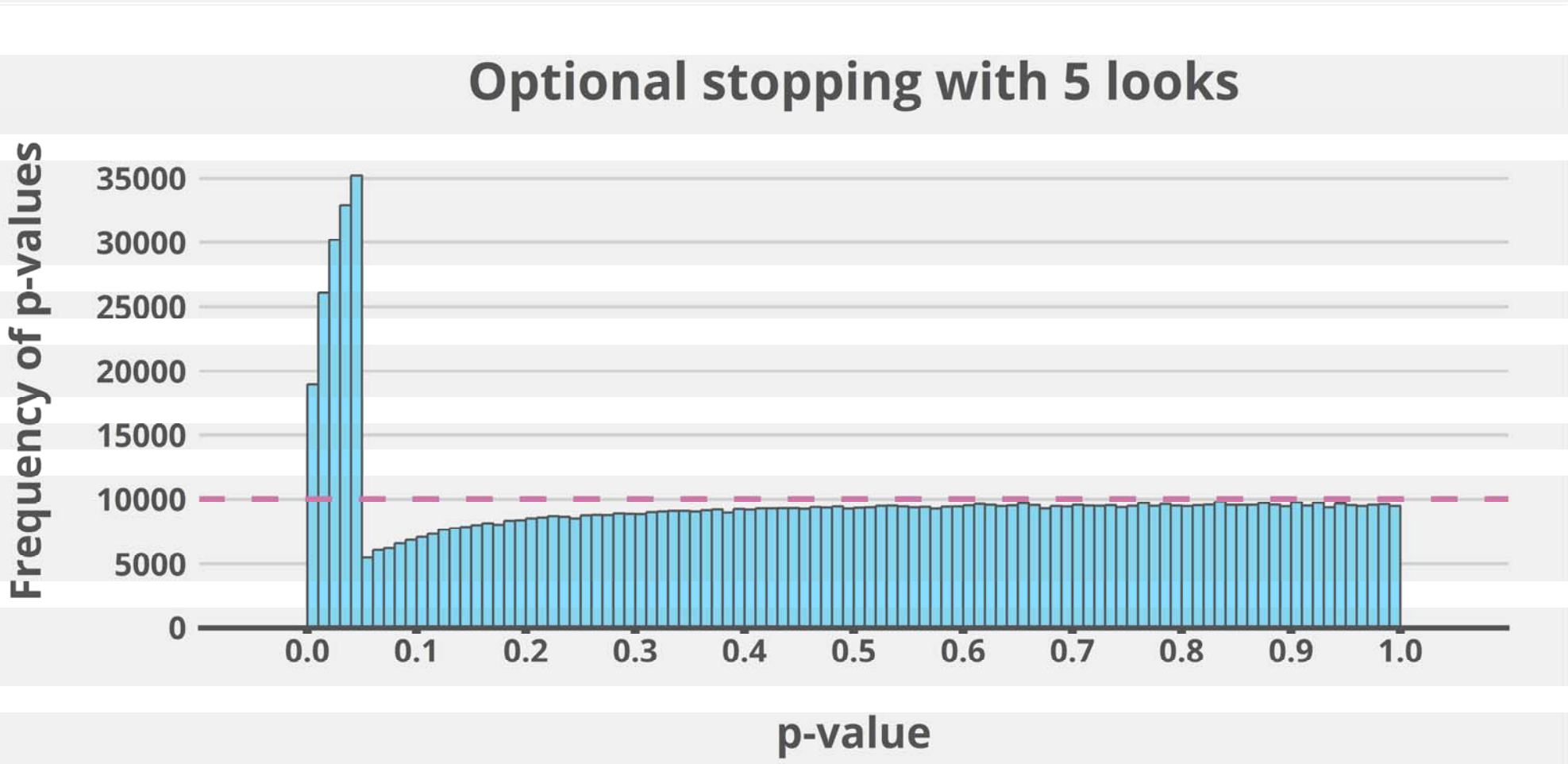
Because of the substantial savings in the expected number of observations effected by the sequential probability ratio test, and because of the simplicity of this test procedure in practical applications, the National Defense Research Committee considered these developments sufficiently useful for the war effort to make it desirable to keep the results out of the reach of the enemy, at least for a certain period of time. The author was, therefore, requested to submit his findings in a restricted report [7] which was dated September, 1943.[3] In this

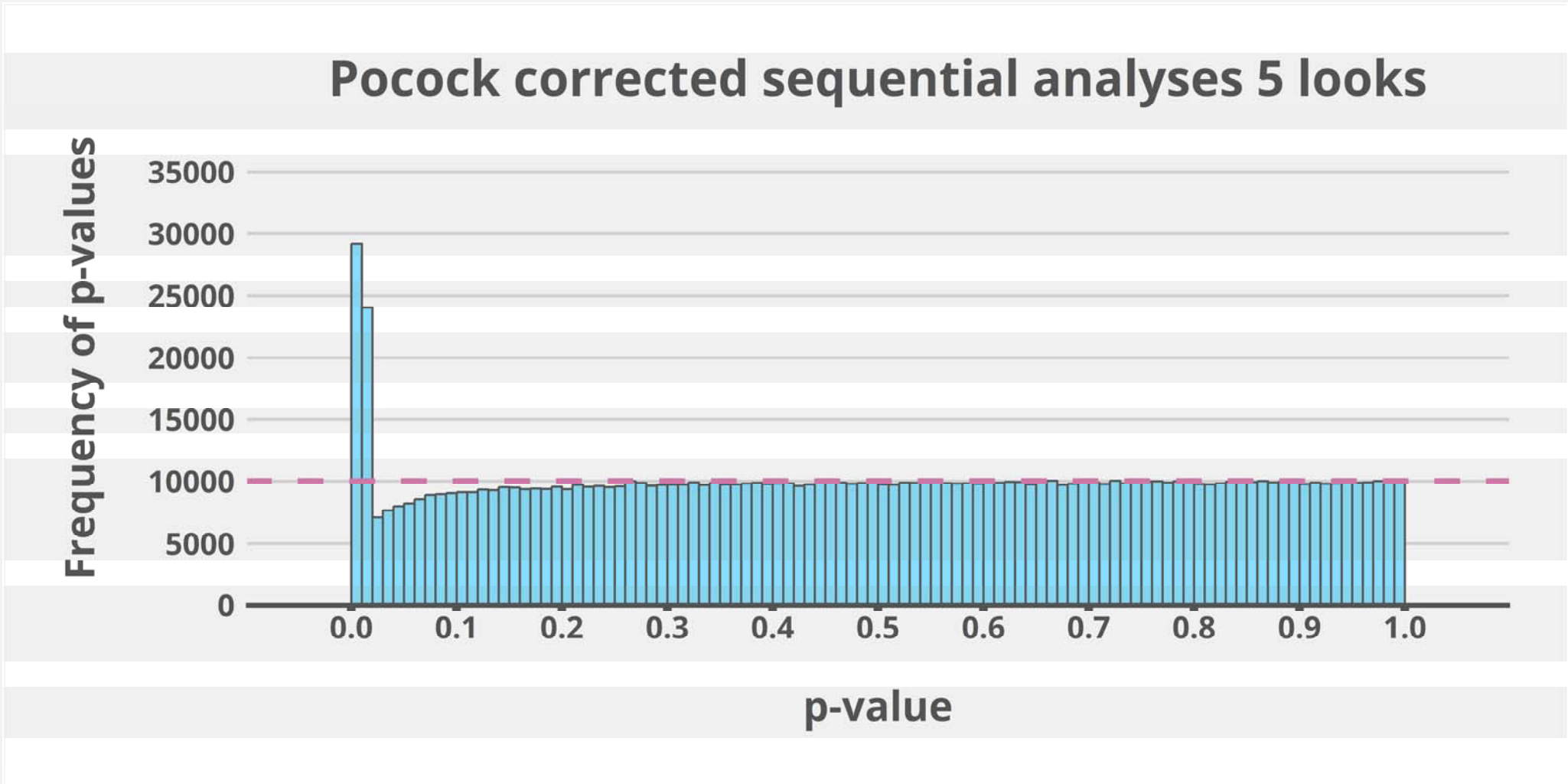
Wald, 1945


0.025I(t>0)
Pocock
Linear
O'Brien-Fleming
0.025I(t=1)
Cumulative Alpha
0.0
0.005
0.010
0.015
0.020
0.025
0.0
0.2
0.4
0.6
0.8
1.0
t

Pocock Boundary

Number of analyses	*p*-value threshold
2	0.0294
3	0.0221
4	0.0182
5	0.0158

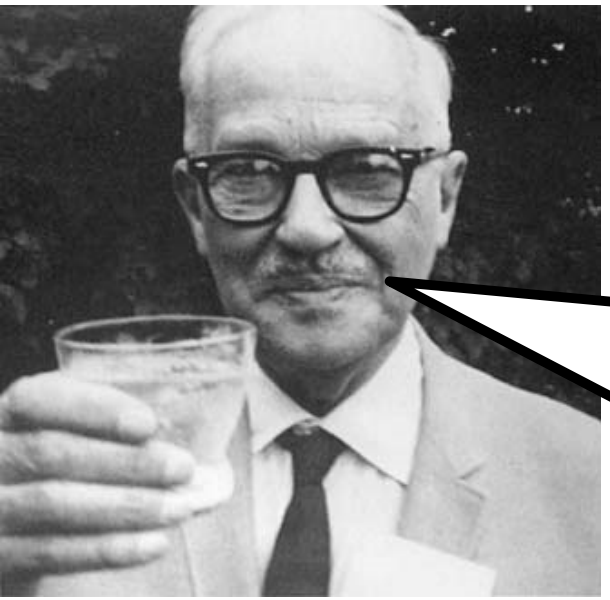

Optional stopping with 5 looks
Frequency of p-values
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value


Pocock corrected sequential analyses 5 looks
Frequency of p-values
35000
30000
25000
20000
15000
10000
5000
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
p-value

Recent approaches control the false discovery rate, not the family-wise error-rate.

Benjamini & Hochberg, 1995

There is nothing special about 5%.



Determining how the balance must be struck should be left to the investigator.

Don’t inflate your Type 1 error rate (you might be fooling yourself).