Exact Confidence Intervals for Adaptive Group Sequential Trials

Fifth International Conference

On Multiple Comparison Procedures

Vienna, Austria

July 8-11, 2007

Cyrus R. Mehta

President, Cytel Inc.

email: mehta@cytel.com – web: www.cytel.com – tel: 617-661-2011

Collaborators:

- **Joint research with University of Vienna investigators:**

  Werner Brannath

  Martin Posch

  Peter Bauer

- **We thank Aniruddha Deshmukh for programming support**

Background

Two broad approaches are available for making mid-course changes to an on-going clinical trial

- **Combining p-values from different stages**
  (Brannath, Bauer, Posch, JASA, 2002)
  Utilizes the property that a p-value is uniformly distributed under the null hypothesis

- **Extending group sequential inference**
  (Müller and Schäfer , Biometrics, 2001)
  Utilizes the independent increments structure of the sequentially computed maximum likelihood statistic

Advantages of the Group Sequential Approach

- Connects to a well-developed, mature field of statistical inference, widely used in practice
- Offers a very wide selection of efficacy and futilty stopping rules through a variety of published spending functions
- Can be implemented with standard group sequential software
- Power computations are available without resorting to simulations

Parameter Estimation

- **Müller and Schäfer (2001) addressed the problem of hypothesis testing only**
- **The associated problem of parameter estimation has not so far been addressed.**
- **This is the topic of the present paper**

General Statement of Problem

- Given a K-look group sequential design to detect H_0: $\delta = 0$
- At some look $L < K$ we wish to make one or more of the following data dependent changes to the study design:
  - Change the sample size
  - Change the spending function
  - Change the number or spacing of the remaining looks
  - Restrict the eligibility criteria
- How can we do this without inflating the type-1 error?
- How can we estimate the parameter δ at the end of the trial?

The Muller and Schafer Principle

- Muller and Schafer (Biometrics, 2000) showed how you can make any data dependent change in an on-going group sequential trial and still preserve the overall type-1 error
- All you have to do is preserve the conditional type-1 error of the remaining portion of the trial.
- Other investigators with a similar idea are Proschan and Hunsberger (Biometrics, 1995) and Denne (SIM, 2001)

Adaptive Test of $H_0: \delta \leq 0$

- Let $b_1, b_2, \ldots b_K$ be the stopping boundaries of the K-look group sequential trial such that

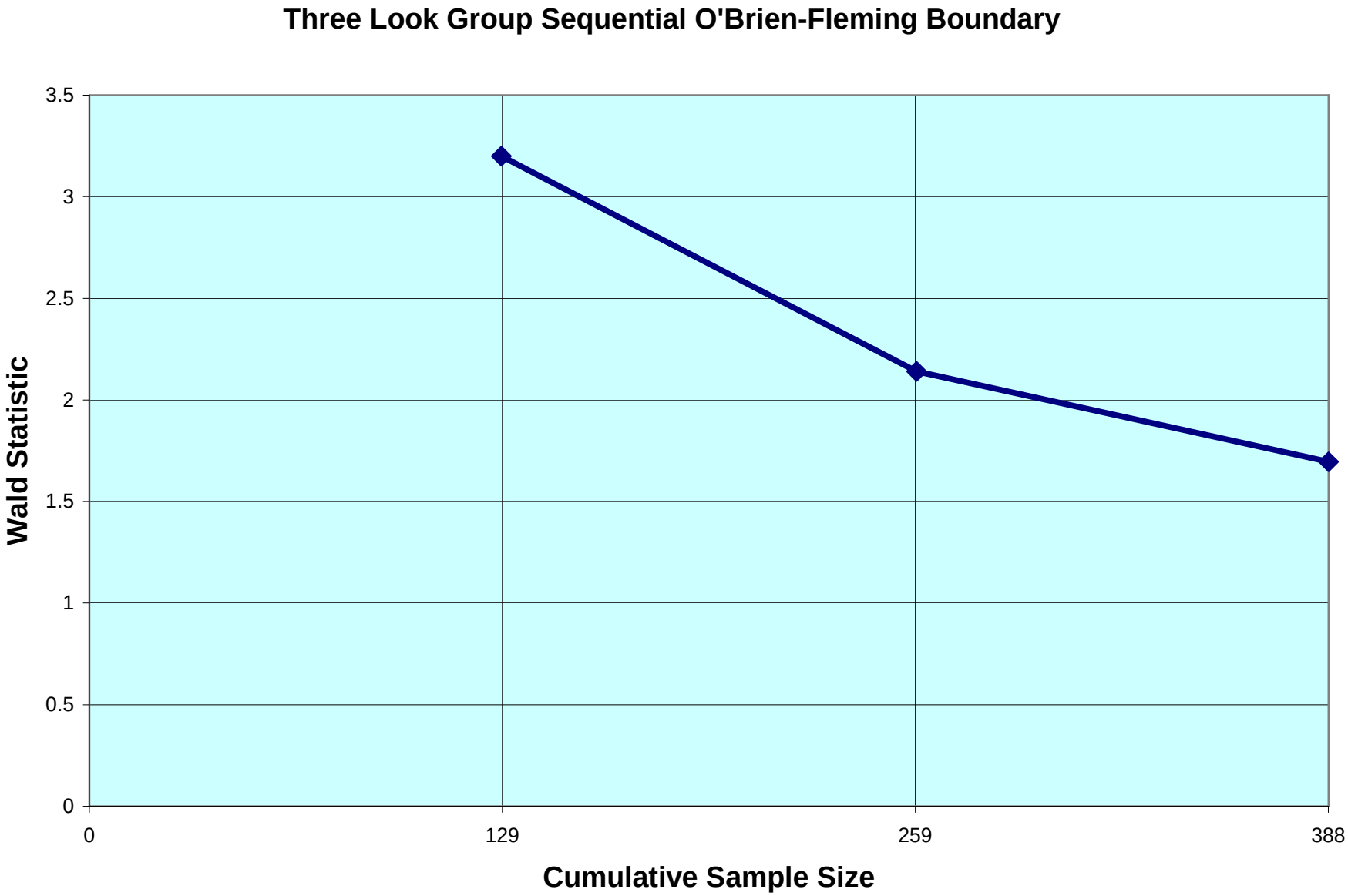
$$P_0(\bigcup_{j=1}^{K}\{Z_j \geq b_j\}) = \alpha$$

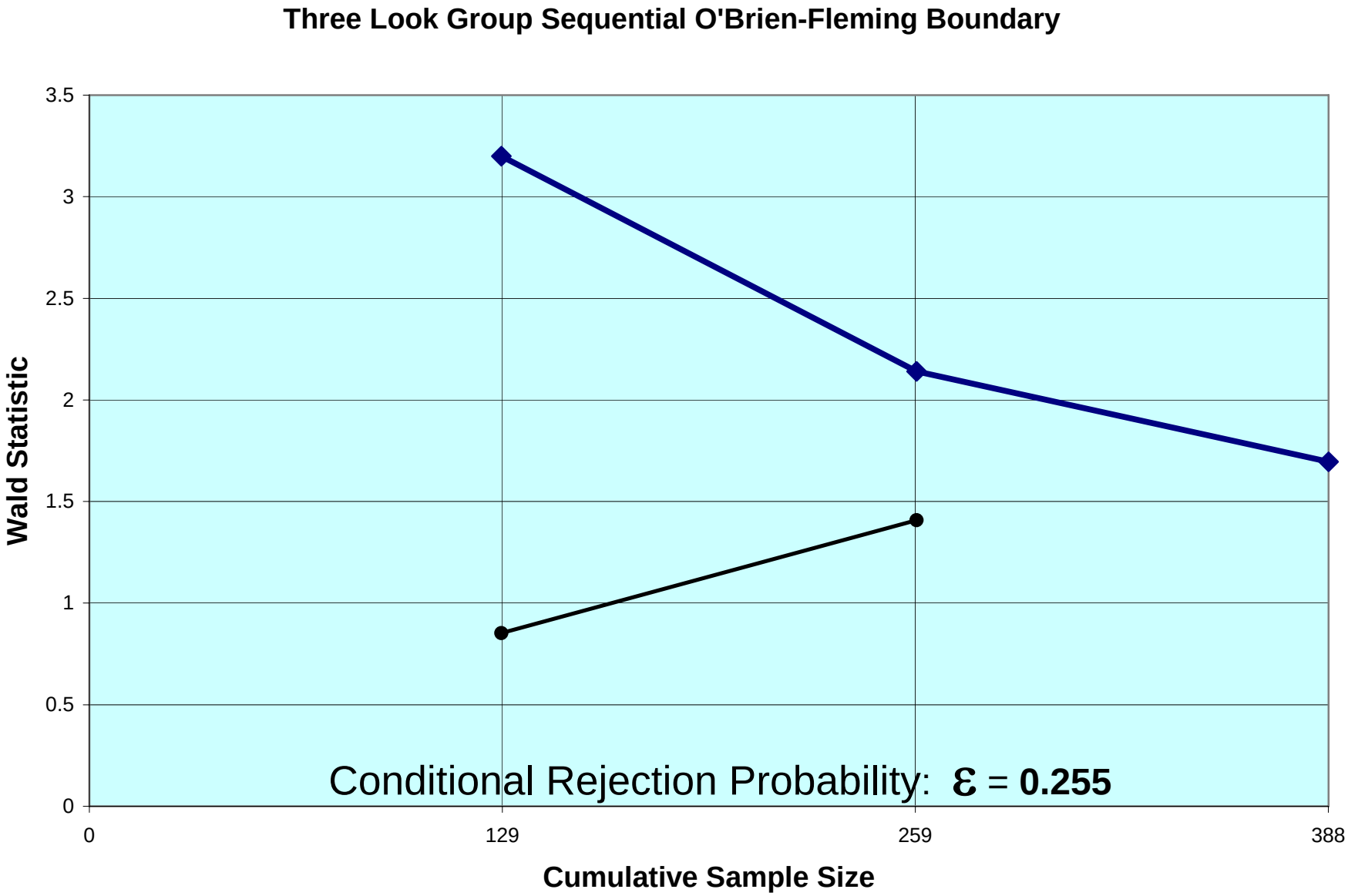
  where $Z_j = \hat{\delta}_j/\text{se}(\hat{\delta}_j)$ is the Wald statistic

- To make an adaptive change at some look L, you must compute the conditional probability of a type-1 error

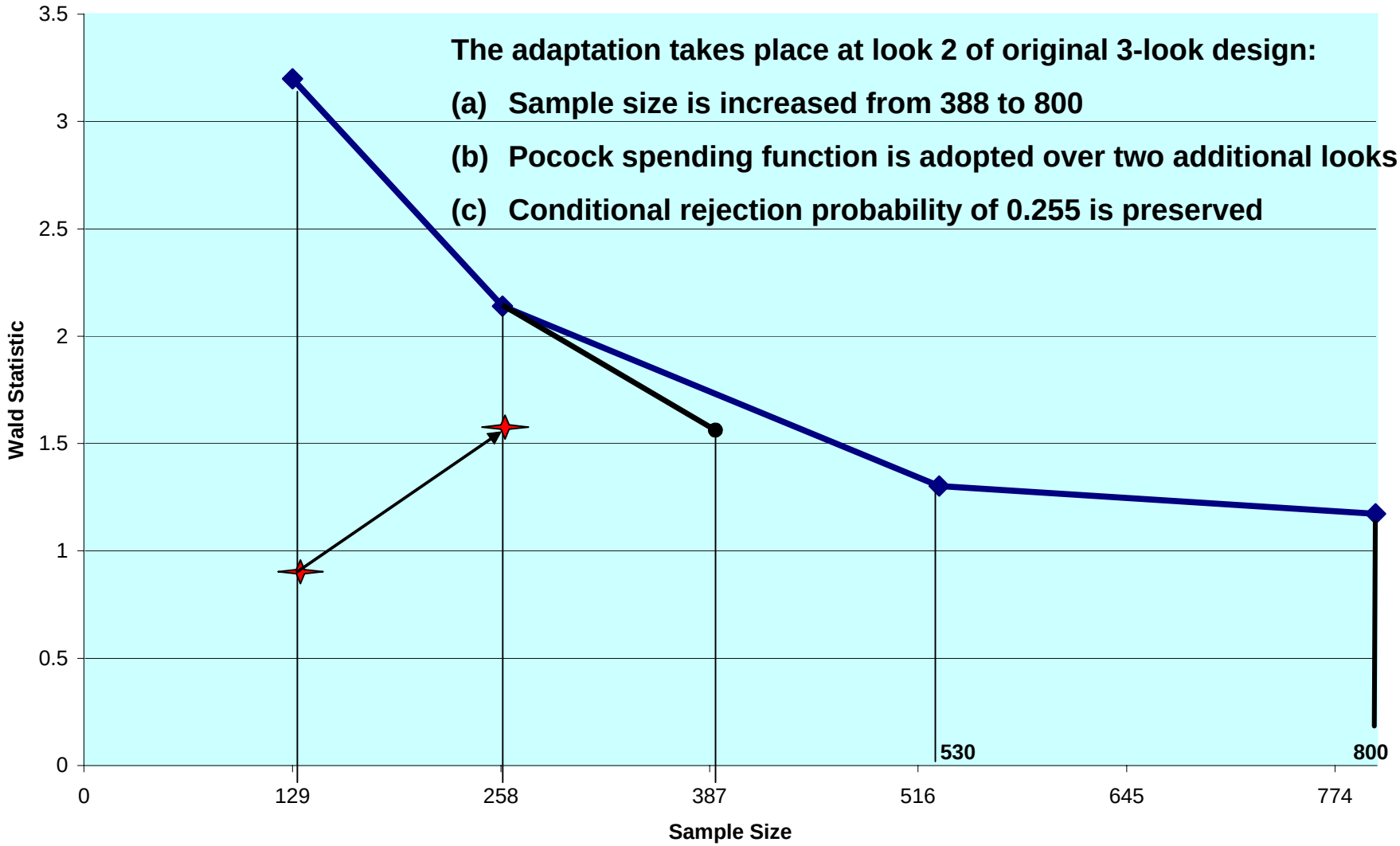
$$\epsilon = P_0(\bigcup_{j=L+1}^{K}\{Z_j \geq b_j\}|z_L)$$

  You can modify the trial in any way at look L but you must preserve ϵ for the modified trial


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Conditional Rejection Probability: ε = 0.255

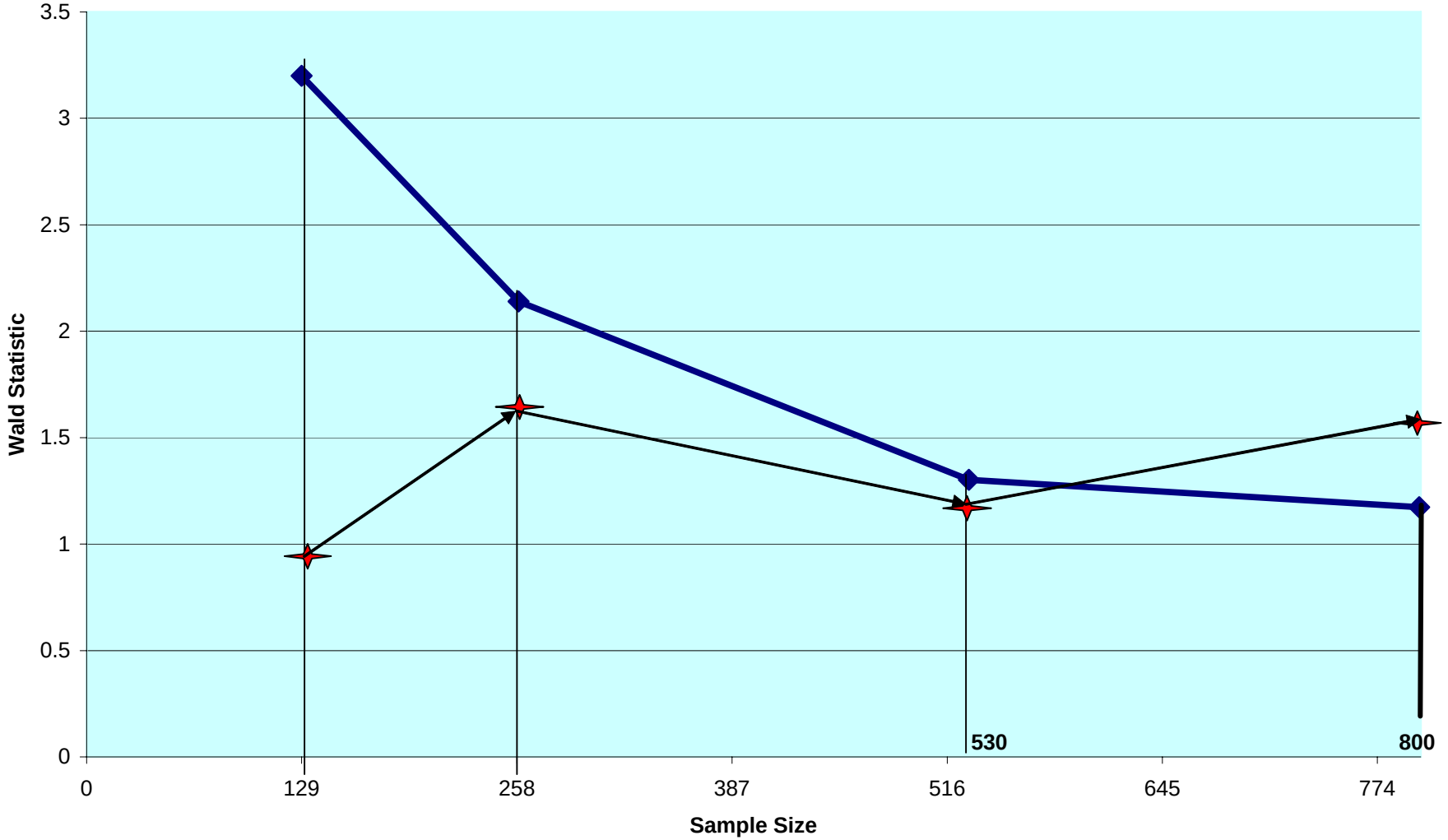
Boundaries of Initial and Adapted Designs
(conditional rejection probabilities of initial and adapted designs are both equal to 0.255)



Testing H_0 at level-α for the Combined Adaptive Trial

- Generate the data for the modified part of the trial, beyond look L
- Combine with the data from the unmodified and modified parts of the trial
- If the modified stopping boundary is crossed then H_0 is rejected
- The conditional probability under H_0,of crossing the modified stopping boundary, given the observed z_L, is ϵ
- The unconditional probability of rejecting H_0, taken over all possible z_L, is α

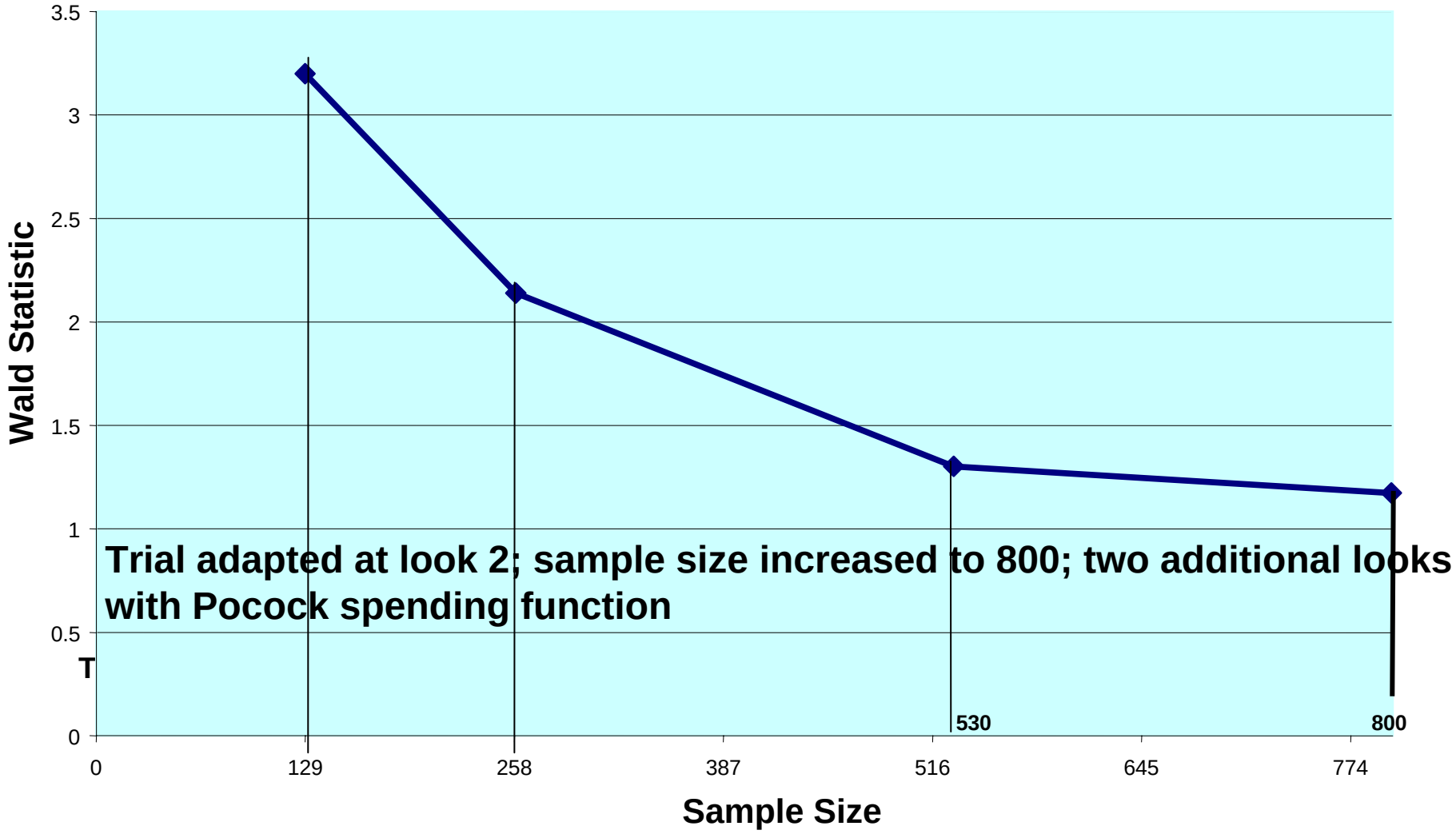
Interim monitoring of the combined trial before and after the adaptive change



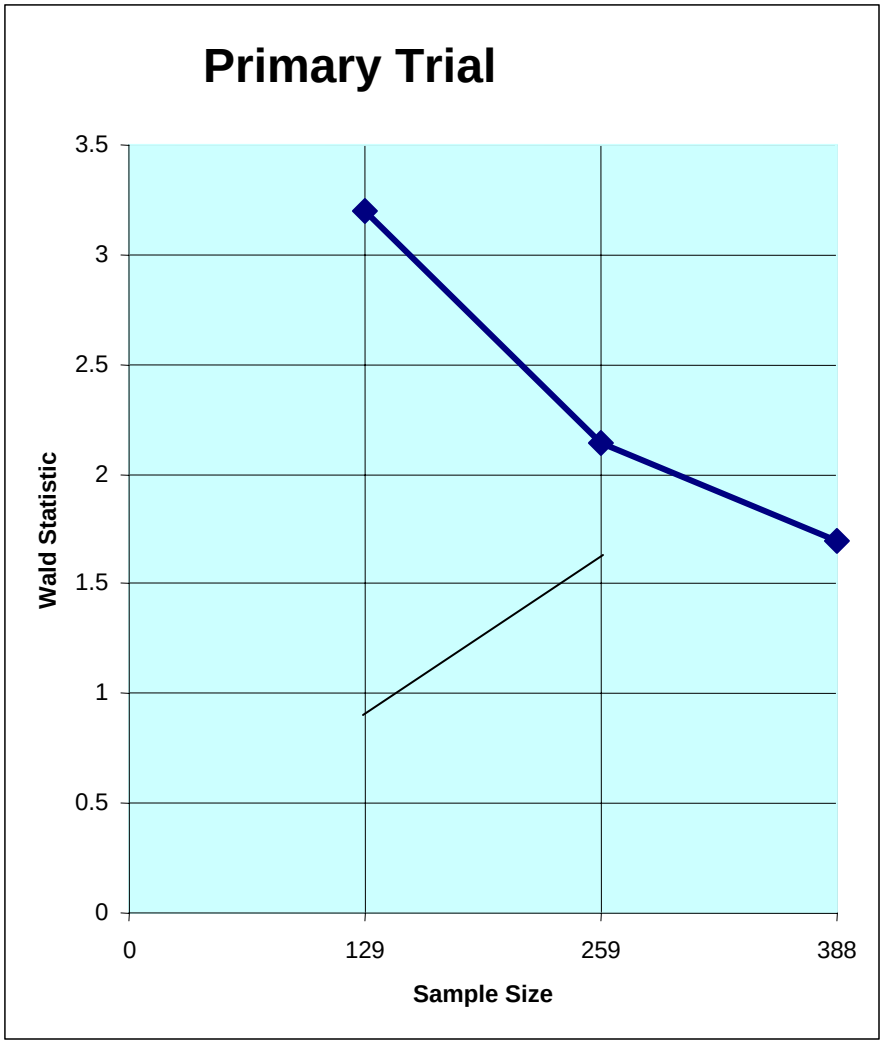
More Convenient Representation of the Combined Adaptive Trial

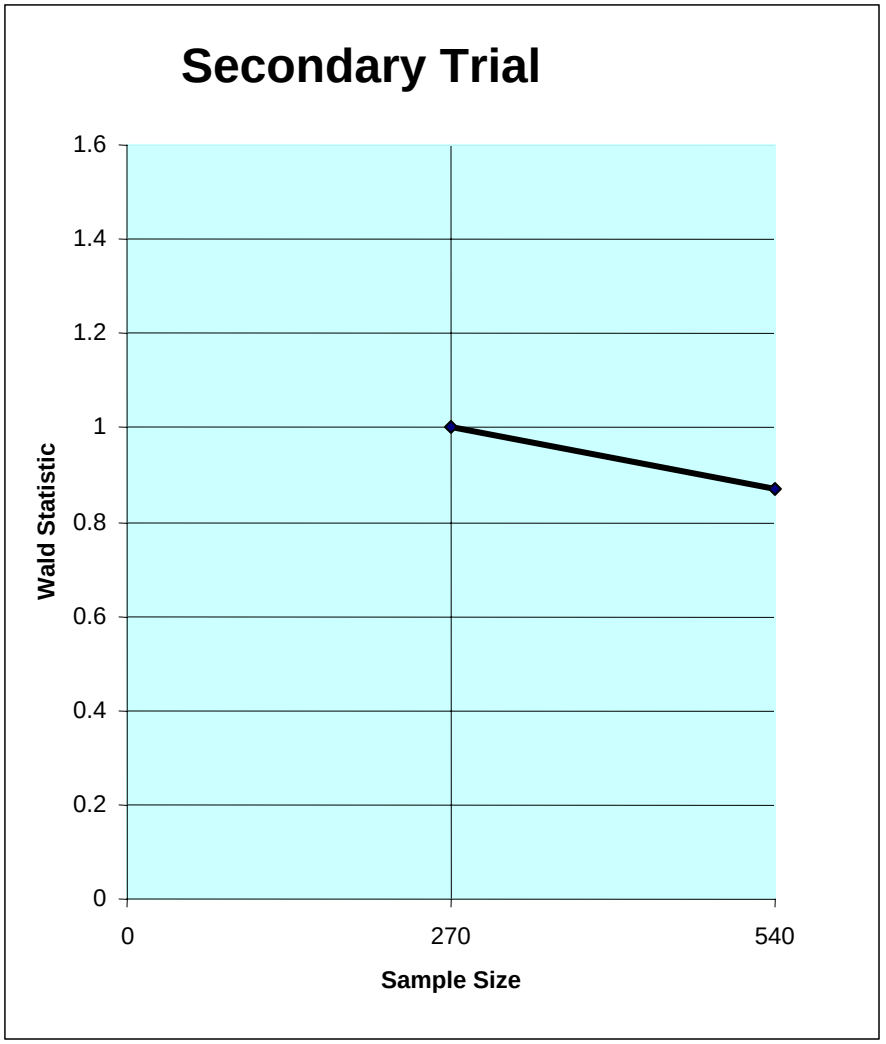
- The combined trial has complicated formulae for the boundaries and test statistics. It is easier to split it into two trials using standard group sequential software
- Think of the initial trial as the primary trial
- Think of the continuation after the adaptive change at look L as a separate and independent secondary trial
- You may use the data from the primary trial to design the secondary trial with any desired sample size, spending function, number and spacing of looks
- However, the secondary trial must have level ϵ

Stopping Boundaries of the Combined Trial after Adaptation



Primary and Secondary Trials Corresponding to Combined Trial





Sample Size 388. OBF Sp Fn.
Conditional rejection prob = 0.255

Sample Size 540. Pocock Sp Fn.
Boundary based on alpha = 0.255

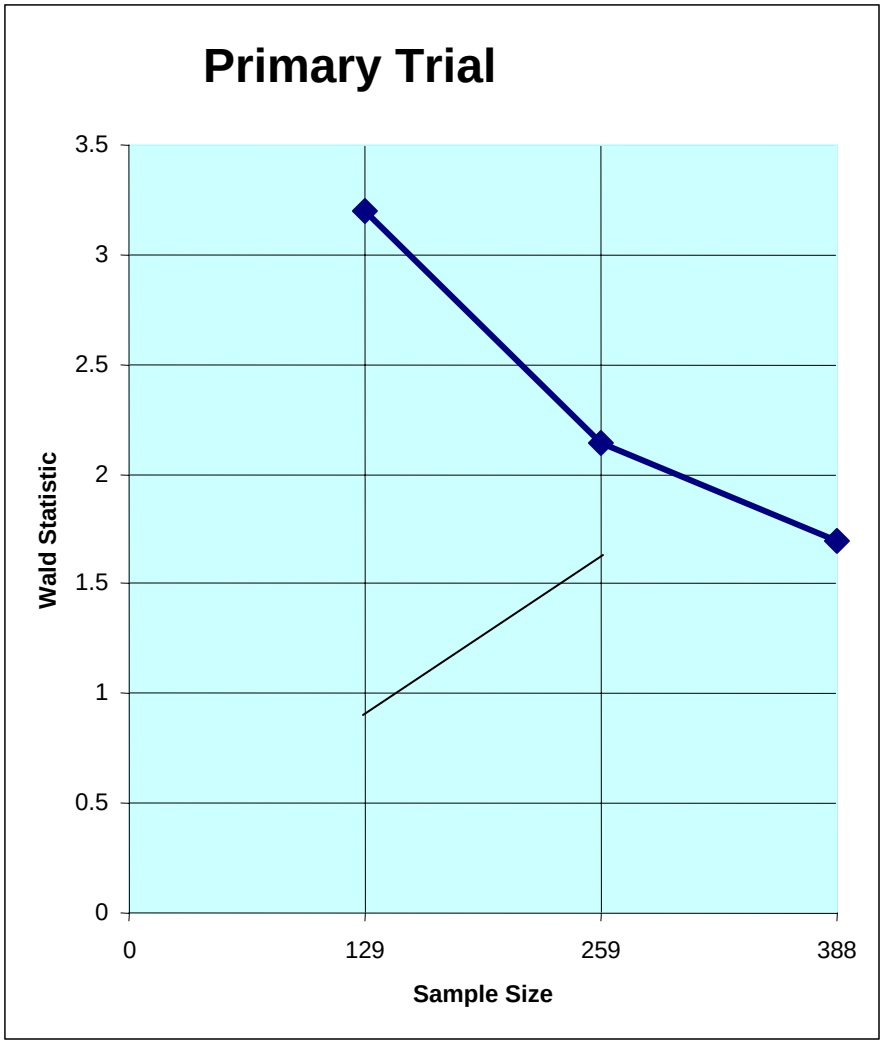
Testing H_0 at Level α with Primary and Secondary Trials

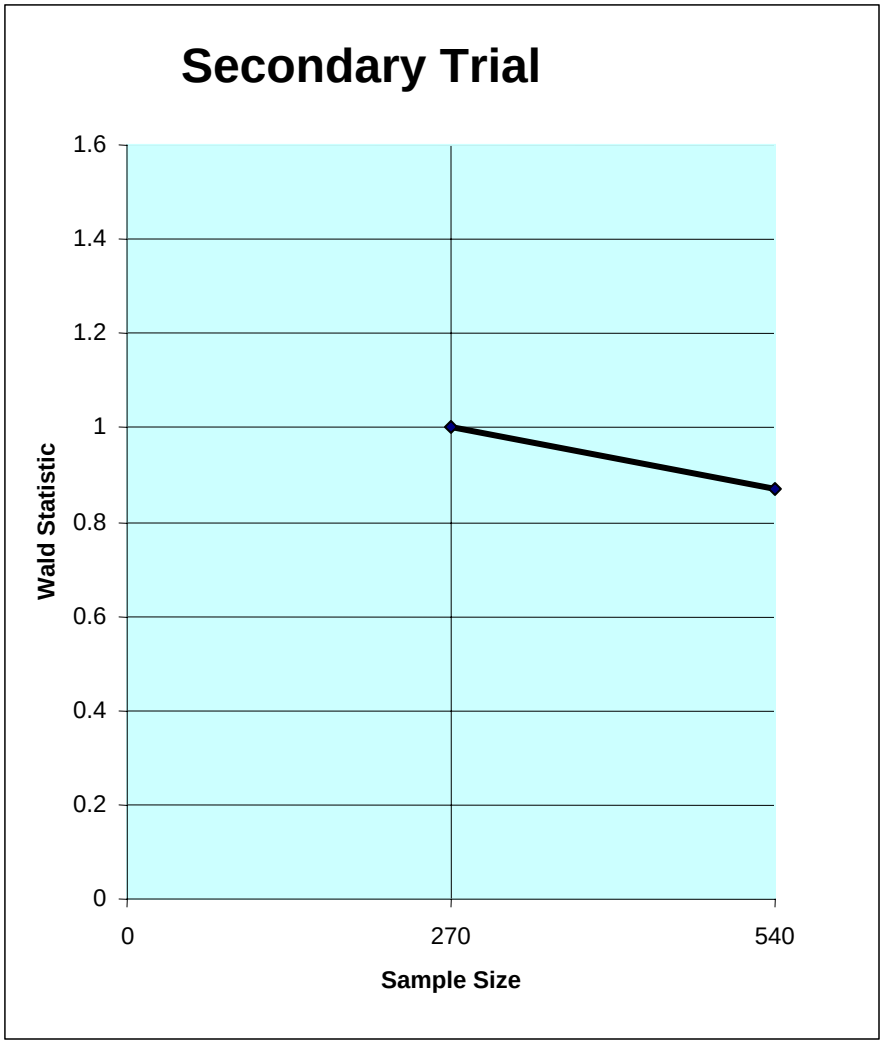
1. At time of adapation of primary trial compute ϵ, the conditional type-1 error if continuing
2. Design secondary trial with $K^{(2)}$ looks and stopping boundaries $b_1^{(2)}, b_2^{(2)}, \ldots b_{K^{(2)}}^{(2)}$ that satisfy

$$P_0(\bigcup_{j=1}^{K^{(2)}} \{Z_j^{(2)} \geq b_j^{(2)}\}) = \epsilon$$

3. Suppose the trial is terminated at look $L^{(2)} \leq K^{(2)}$. Then H_0 is rejected if boundary is crossed ($z_{L^{(2)}}^{(2)} \geq b_{L^{(2)}}^{(2)}$)

Primary and Secondary Trials Corresponding to Combined Trial

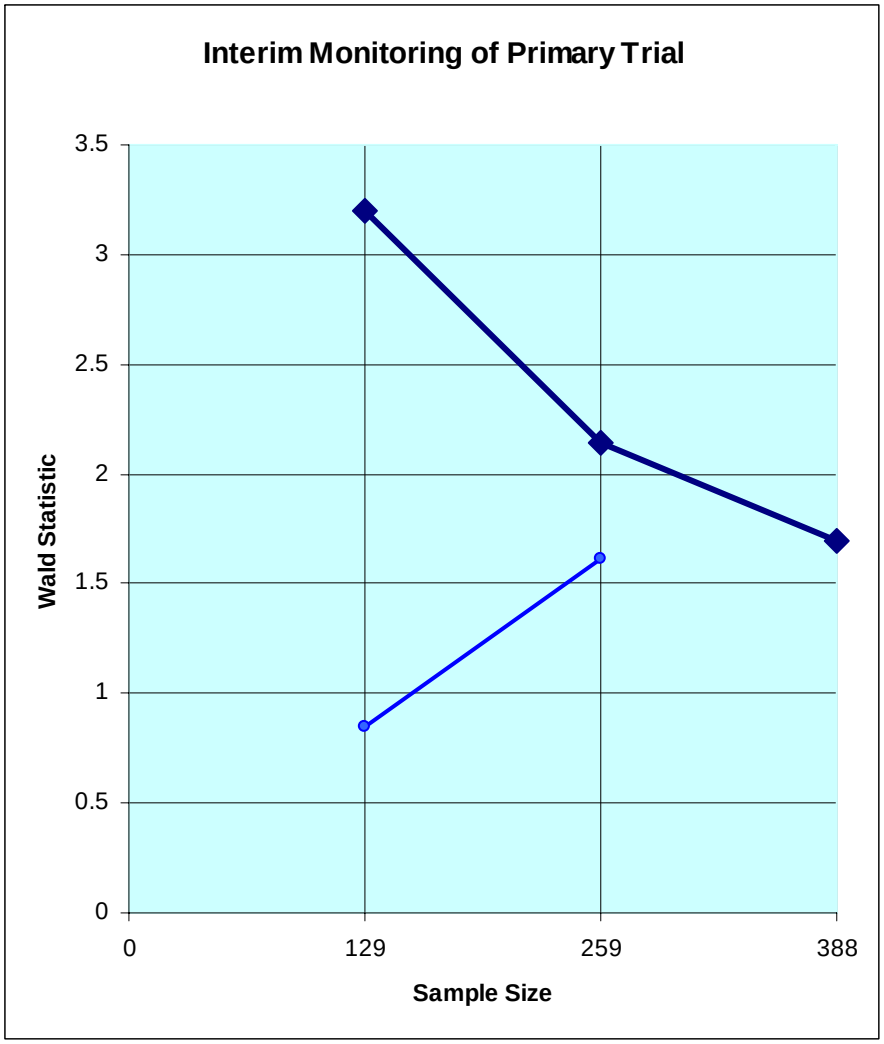


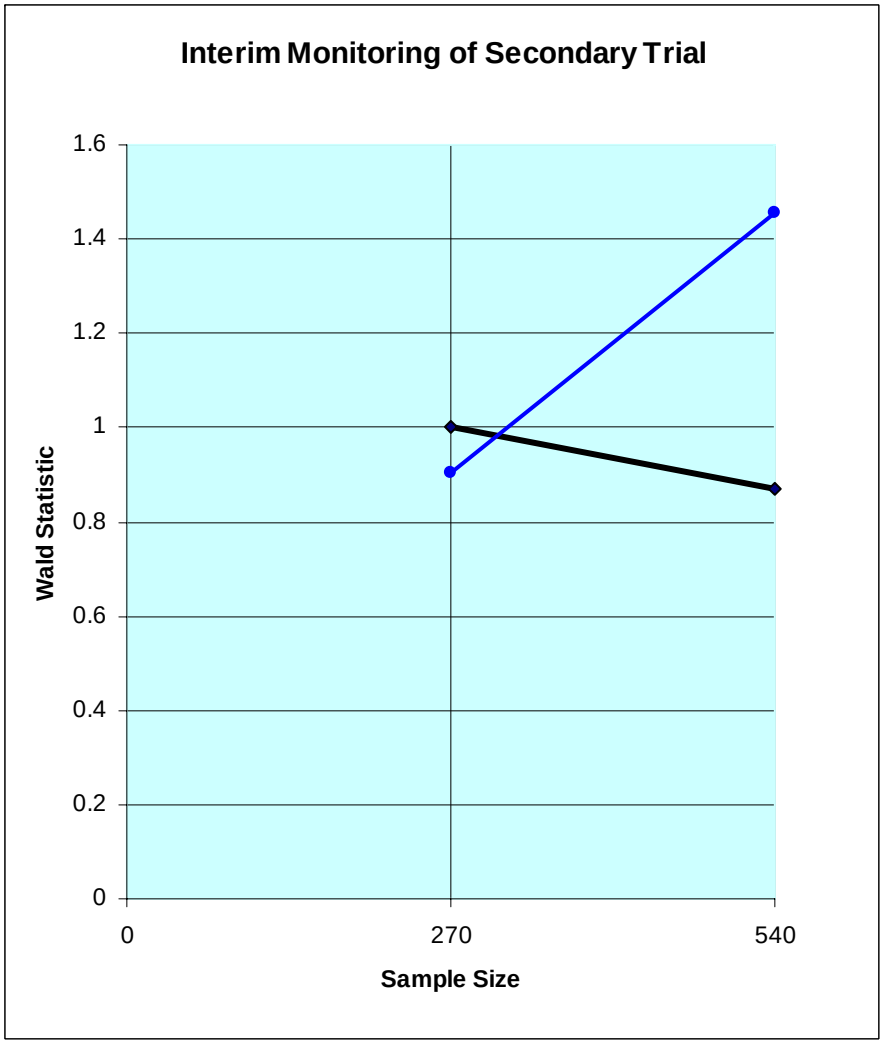


Sample Size 388. OBF Sp Fn.
Conditional rejection prob = 0.255

Sample Size 540. Pocock Sp Fn.
Boundary based on alpha = 0.255

Interim Monitoring of Primary and Secondary Trials





Sample Size 388. OBF Sp Fn.
Conditional rejection prob = 0.255

Sample Size 540. Pocock Sp Fn.
Boundary based on alpha = 0.255

Extension to Confidence Intervals

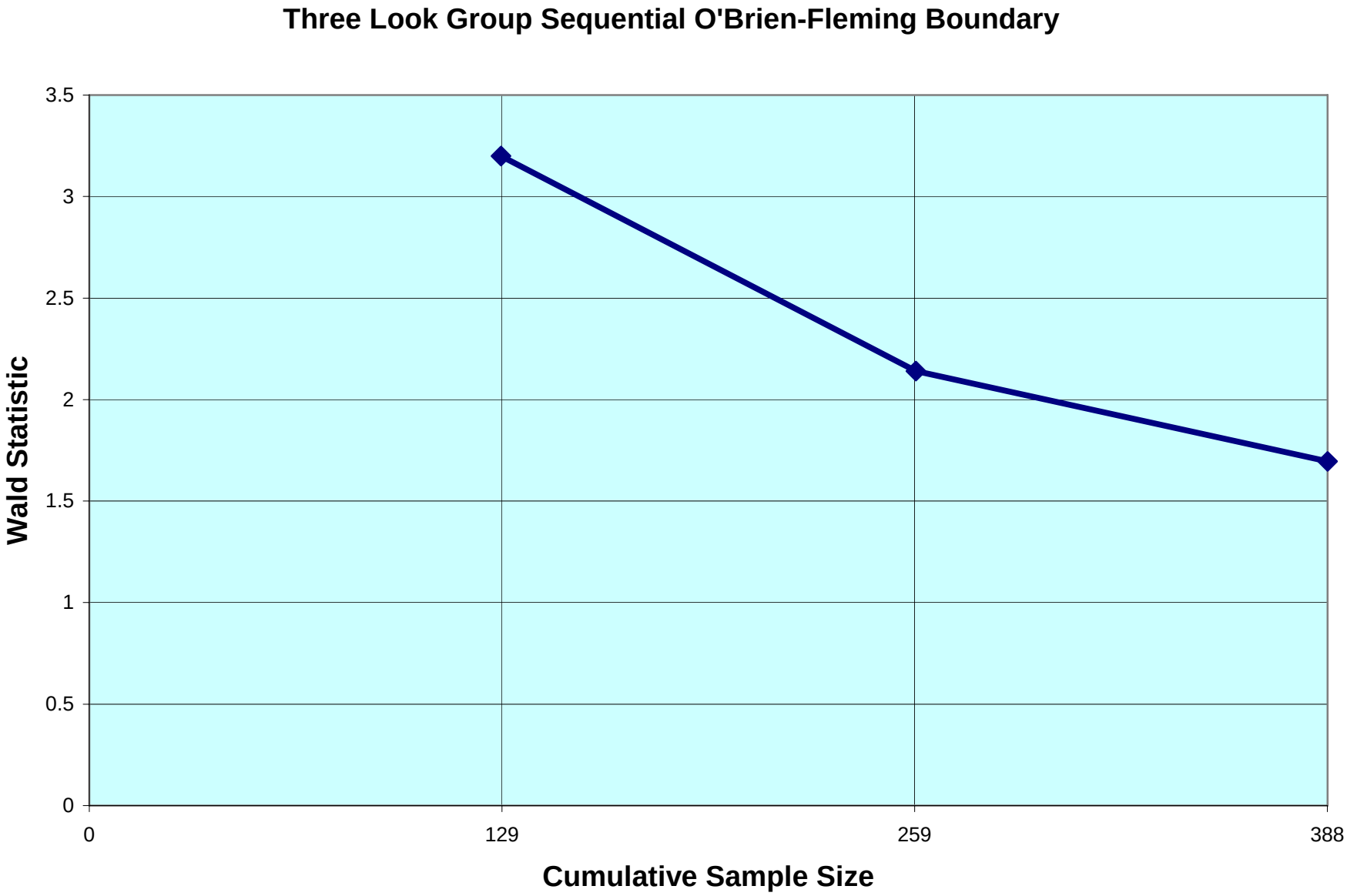
- Test H_h: $\delta \leq h$ against the one sided alternative that $h > 0$ in an adaptive setting
- The $100 \times (1 - \alpha)\%$ confidence set consists of all h for which H_h cannot be rejected at level α
- In practice we search for $\underline{\delta}$, the smallest h for which H_h cannot be rejected at level α, and report $[\underline{\delta}, \infty)$ as the one-sided confidence interval

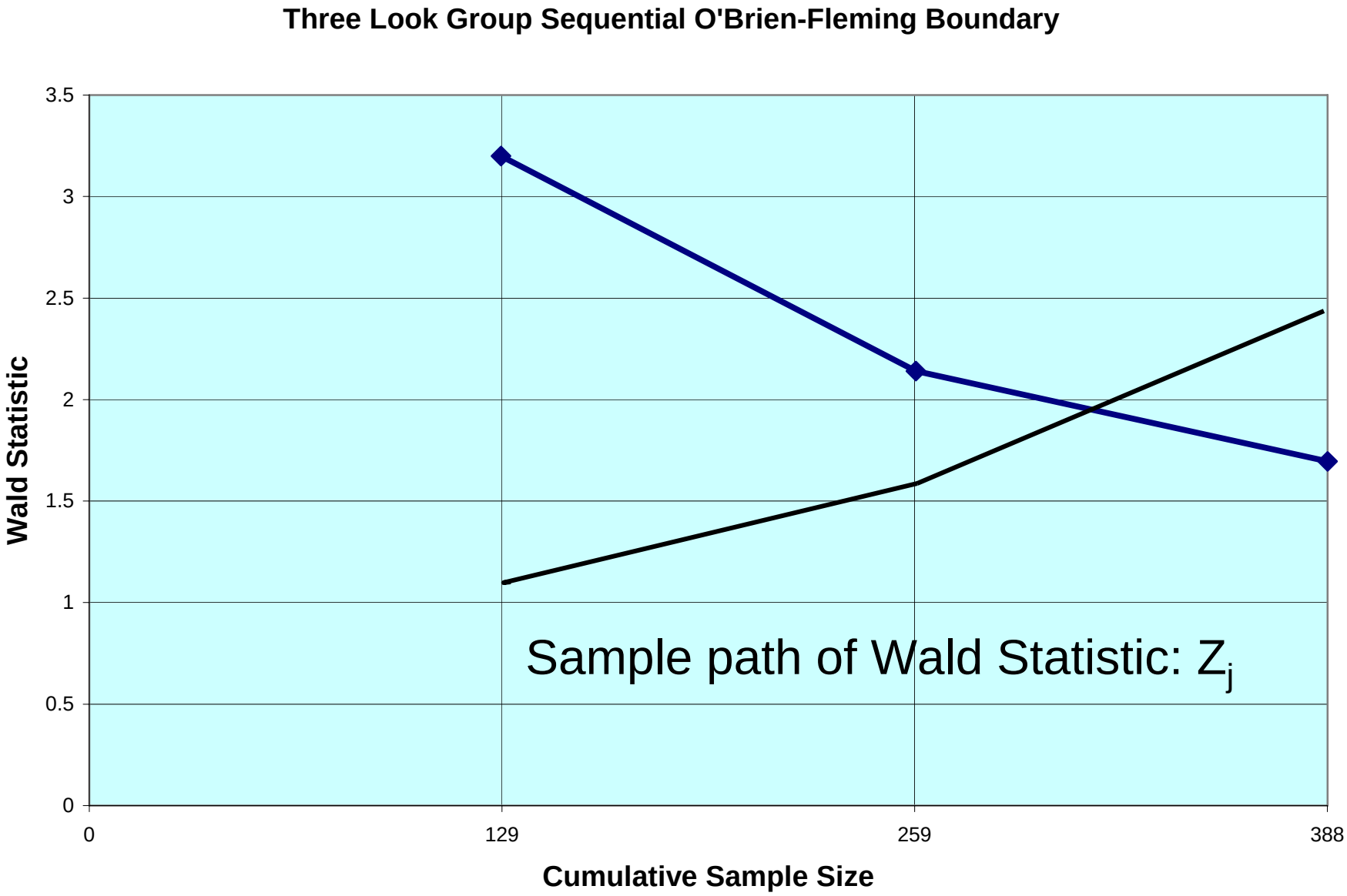
Two Methods to Test H_h: $\delta \leq h$

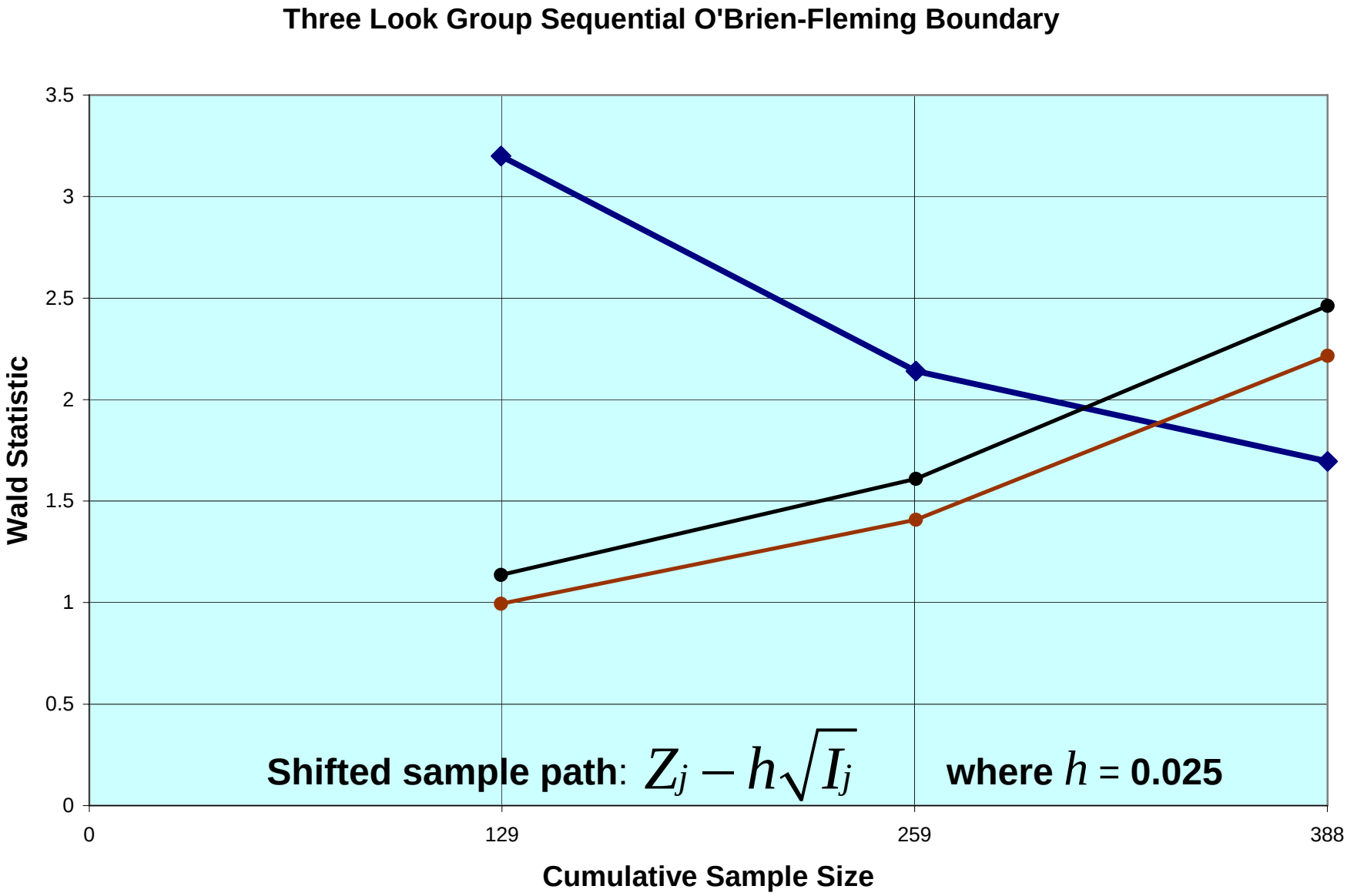
- **RCI Method** Extend the Repeated Confidence Intervals of Jennison and Turnbull (1989) to the adaptive setting
- **SWACI Method** Extend the Stage-Wise Adjusted Confidence Intervals of Tsiatis, Rosner and Mehta (1984) to the adaptive setting

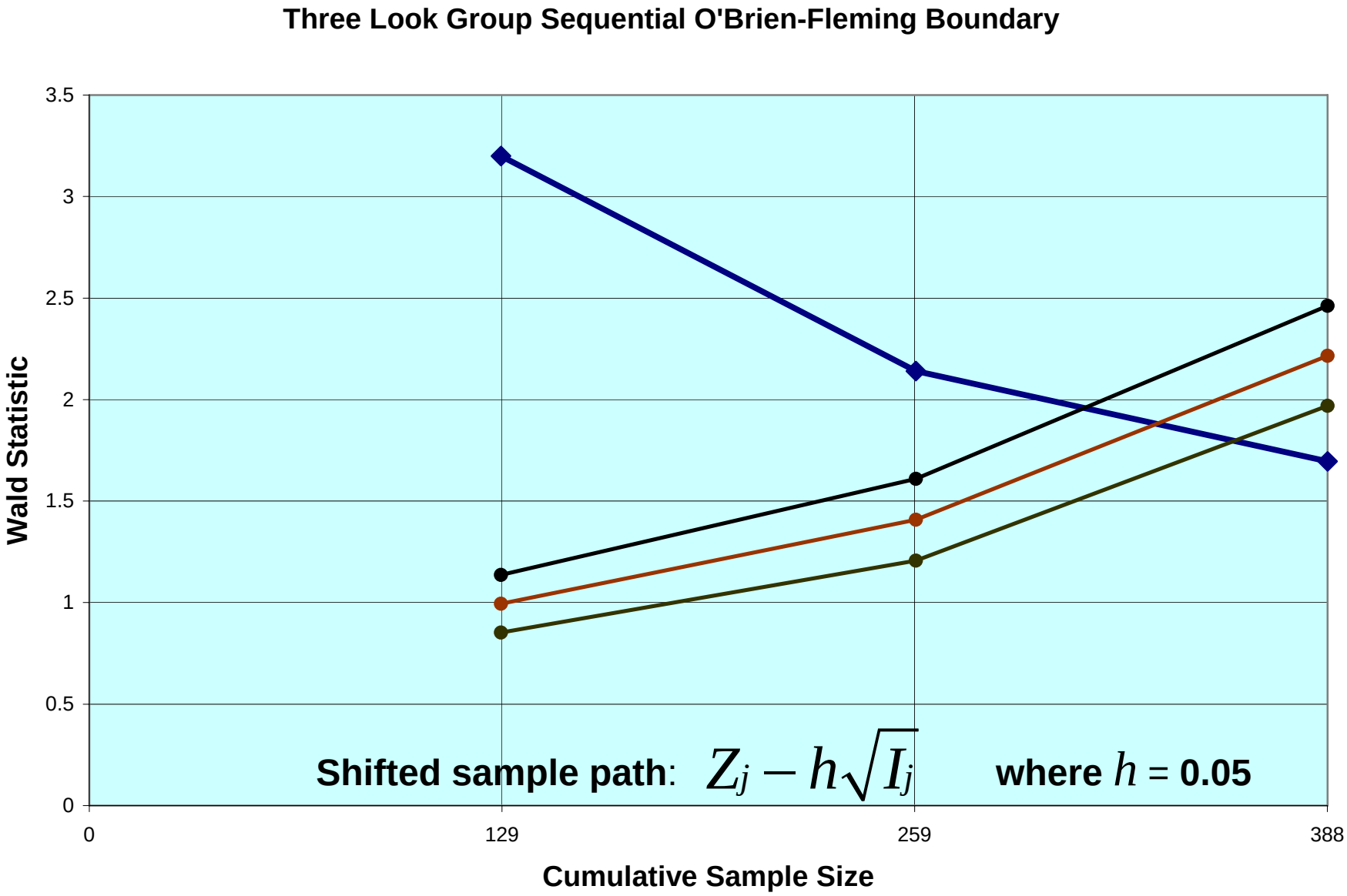
Classical RCI Method

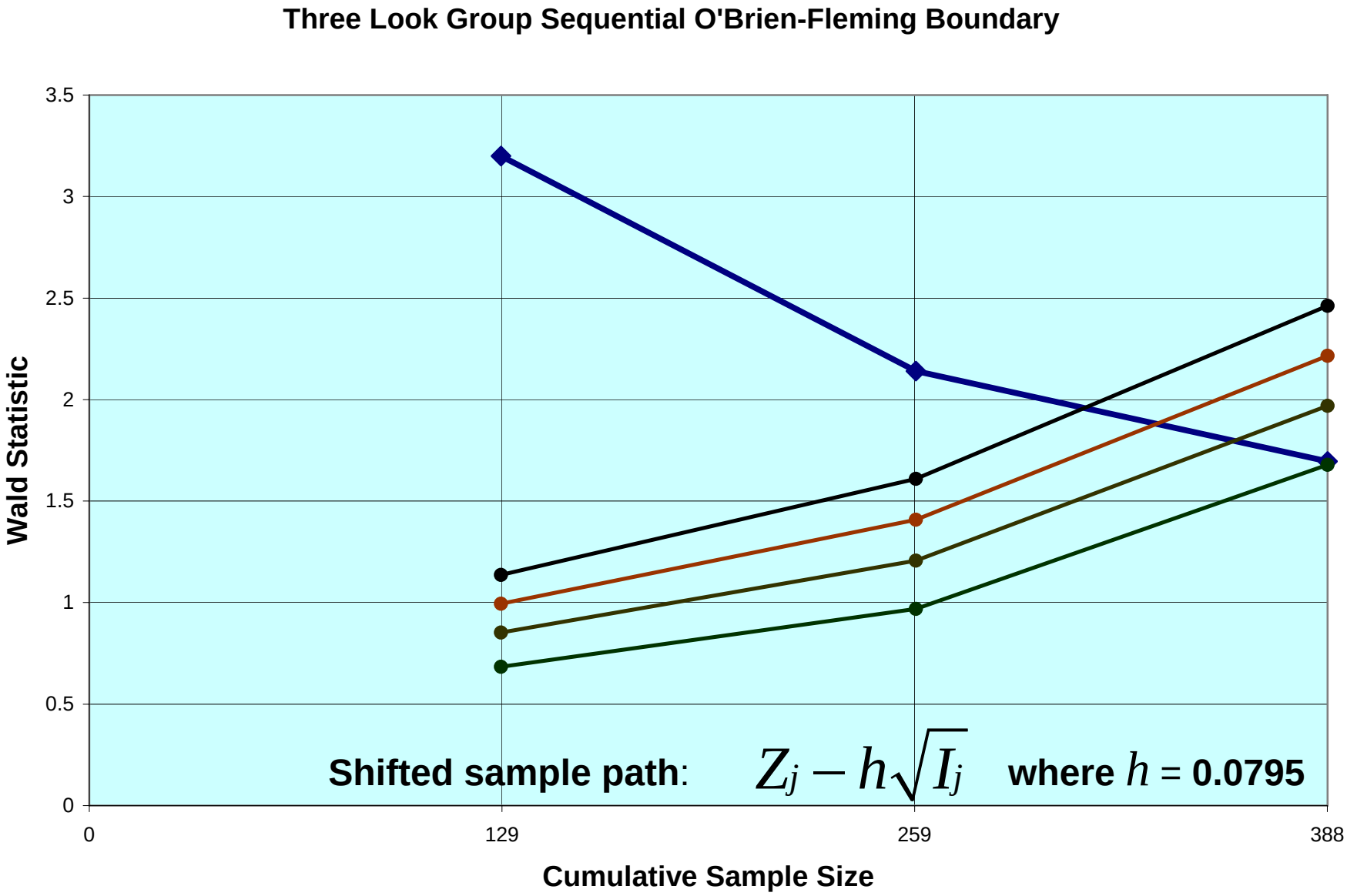
- Shift the data at look j from Z_j to $Z_j - h\sqrt{I_j}$ where I_j is the Fisher information
- Reject H_h if the shifted statistic crosses the boundary
- The one sided RCI is $[\underline{\delta}, \infty)$ where $\underline{\delta}$ is the smallest h at which H_h is accepted


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Sample path of Wald Statistic: Z_j


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Shifted sample path: $Z_j - h\sqrt{I_j}$
where h = 0.025


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Shifted sample path: $Z_j - h\sqrt{I_j}$ where h = 0.05


Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Shifted sample path: $Z_j - h\sqrt{I_j}$ where h = 0.0795

Adaptive RCI Method

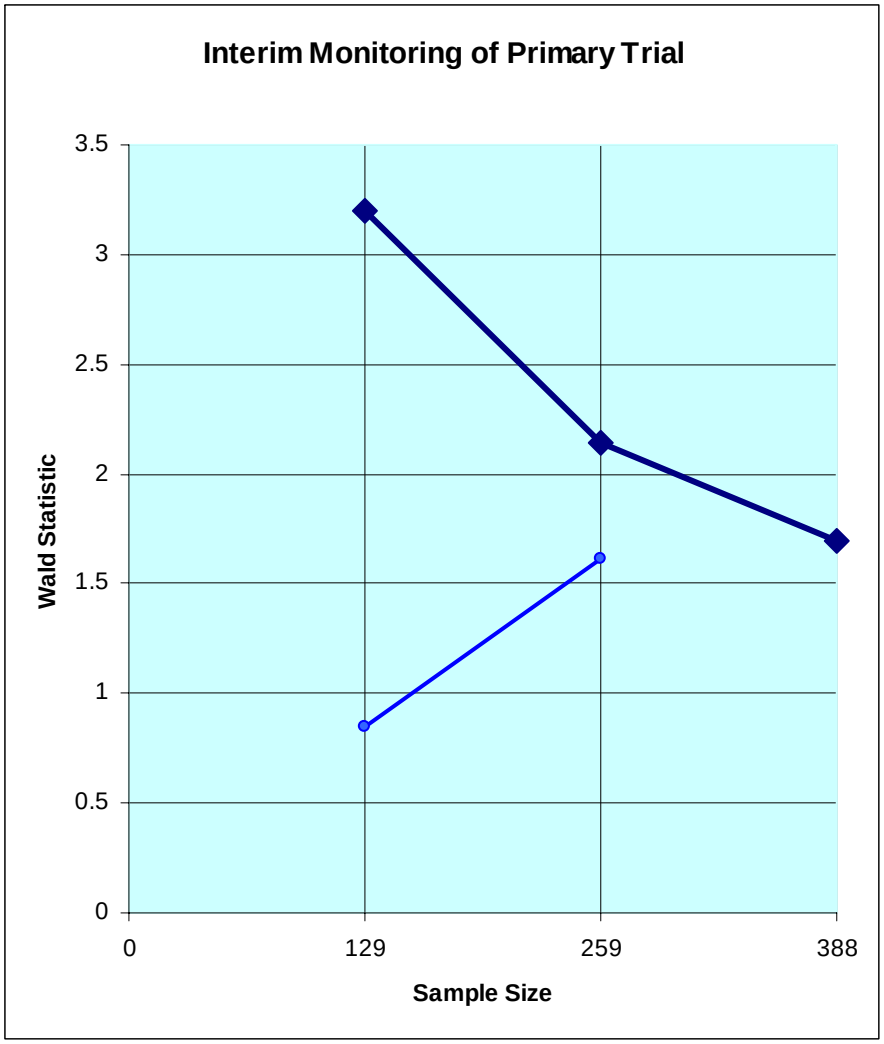
Suppose adaptive change at look L of primary trial with $Z_L = z_L$

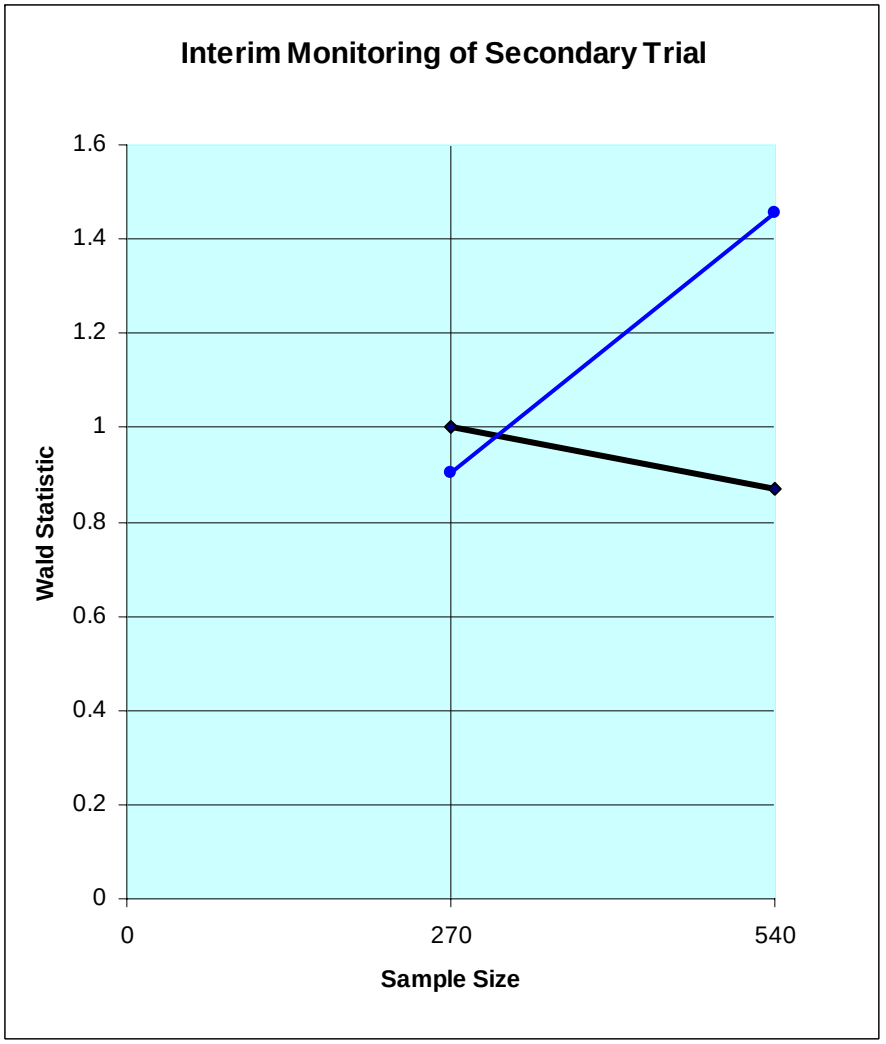
- **H_h is rejected in the primary trial if shifted statistic crosses the boundary**
- **The probability of this event given $Z_L = z_L$ and $\delta = h$ is**

$$\epsilon(h) = P_h \bigcup_{j=L+1}^{K} (Z_j - h\sqrt{I_j} \geq b_j | z_L - h\sqrt{I_L})$$

- **Implement the secondary trial with $\epsilon(h)$ as the type-1 error and reject H_h if the boundary is crossed by the shifted statistic**

Muller and Schafer Adaptive Test of delta = 0

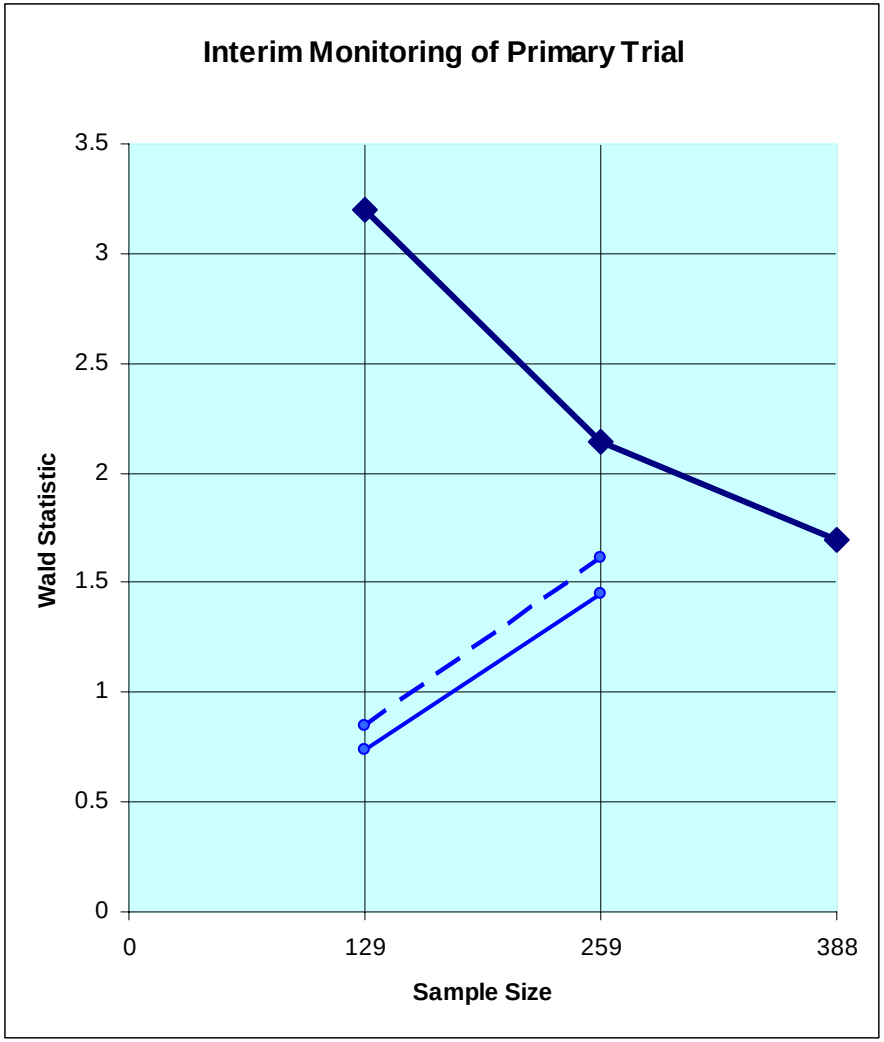




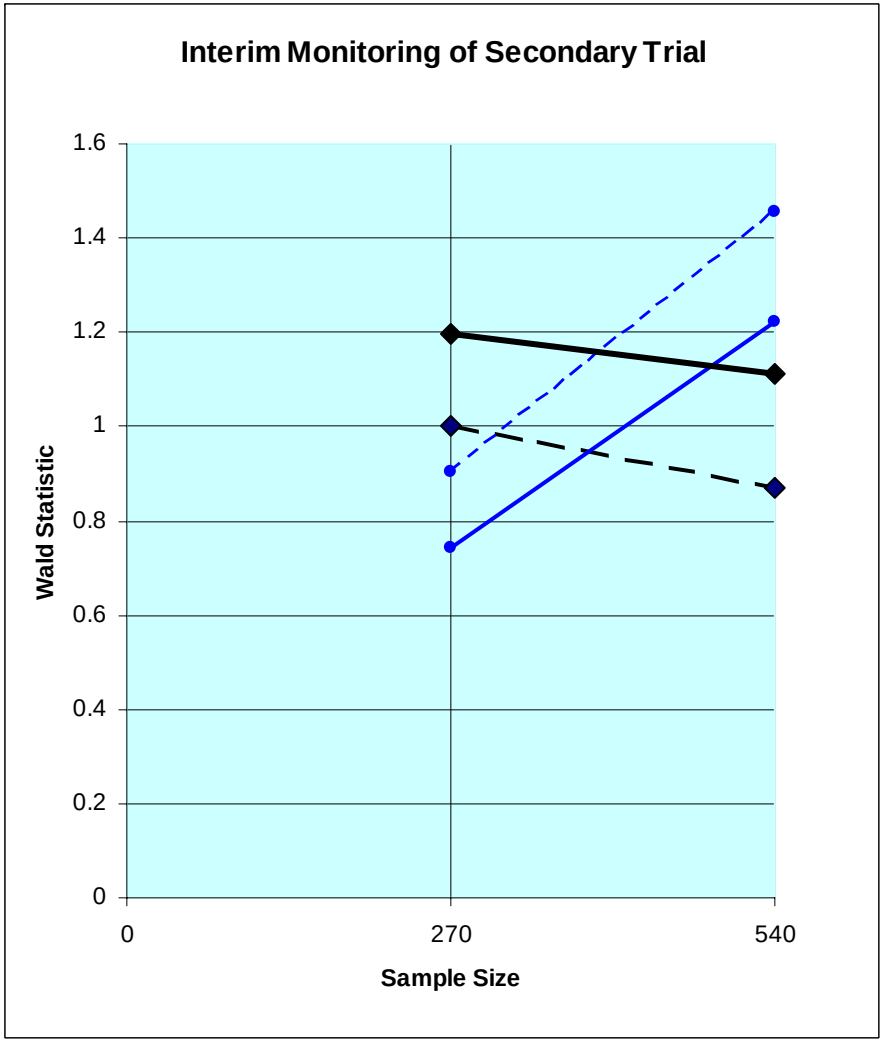
Shift = 0
Conditional rejection prob = 0.255

Shift = 0
Boundary based on alpha = 0.255
H0: delta = 0 is rejected

Muller and Schafer Adaptive Test of delta = 0.02

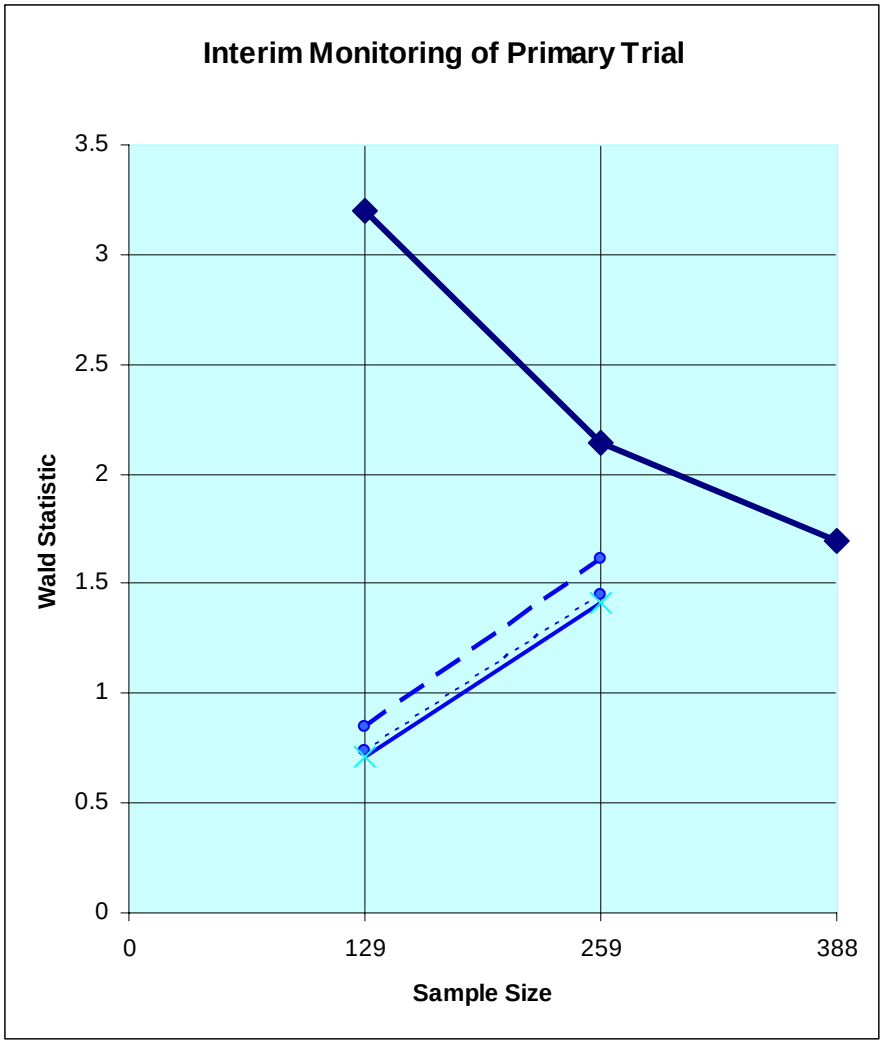


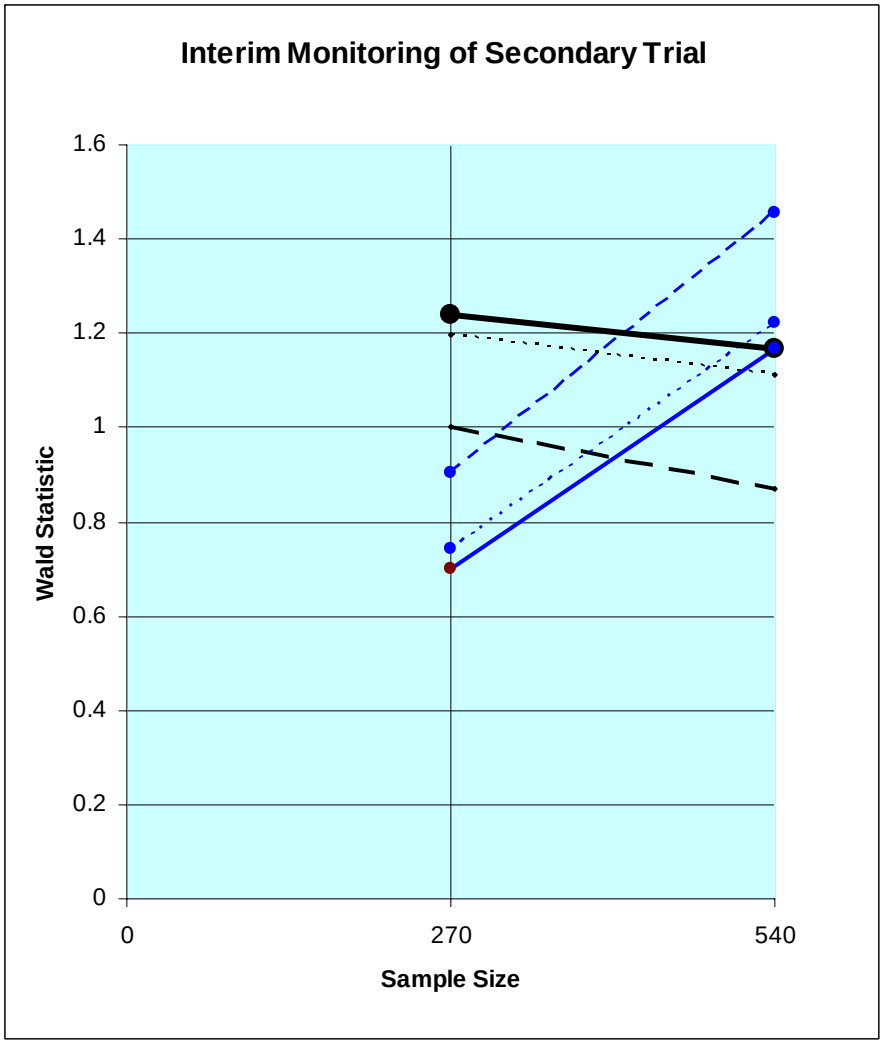
Shift = 0.02
Conditional rejection prob = 0.187



Shift = 0.02
Boundary based on alpha = 0.187
H0: delta = 0.02 is rejected

Muller and Schafer Adaptive Test of delta = 0.0248





Shift = 0.0248
Conditional rejection prob = 0.173

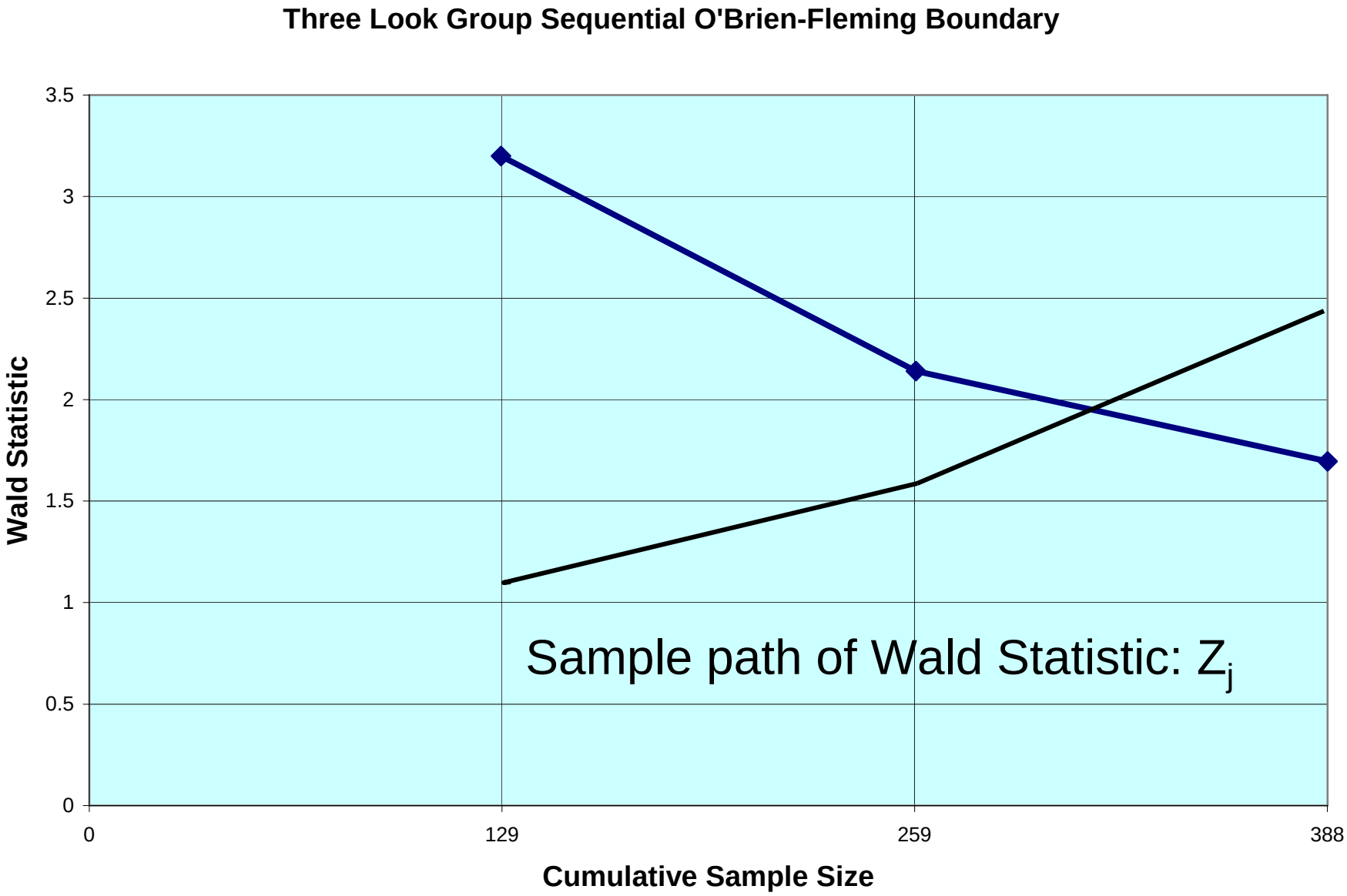
Shift = 0.0248
Boundary based on alpha = 0.173
H0: delta = 0.0248 is **ACCEPTED**

Classical SWACI Method

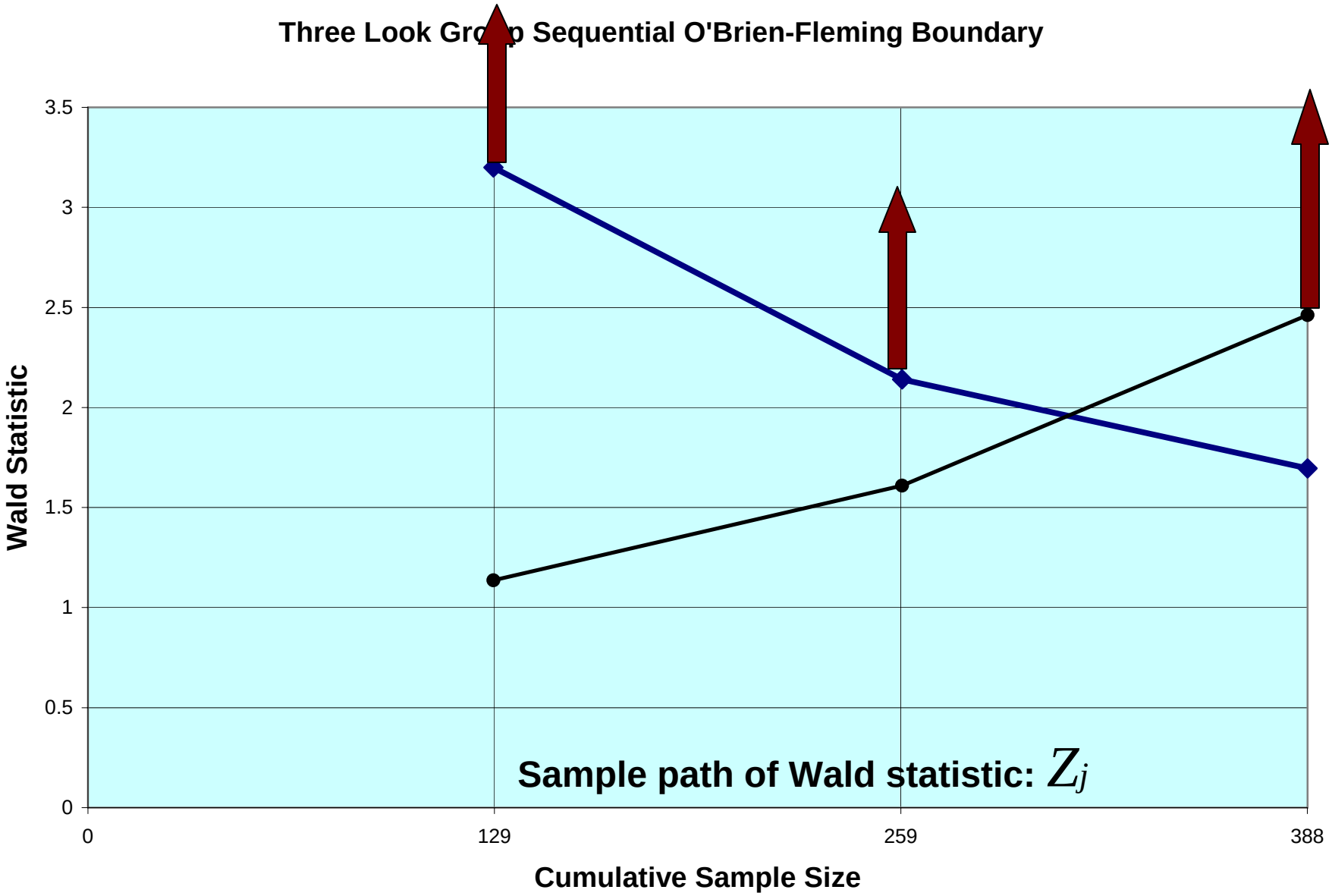
- Suppose the trial terminates at look L
- The p-value for testing H_h is obtained by the stage-wise ordering of the sample space (Tsiatis, Rosner, Mehta, 1984)

$$p(h) = P_h \bigcup_{j=1}^{L-1} \{Z_j \geq b_j\} \cup \{Z_L \geq z_L\}$$

- H_h is rejected if $p(h) \leq \alpha$
- The SWACI is $[\underline{\delta}, \infty)$ where $\underline{\delta}$ is the smallest h at which H_h cannot be rejected

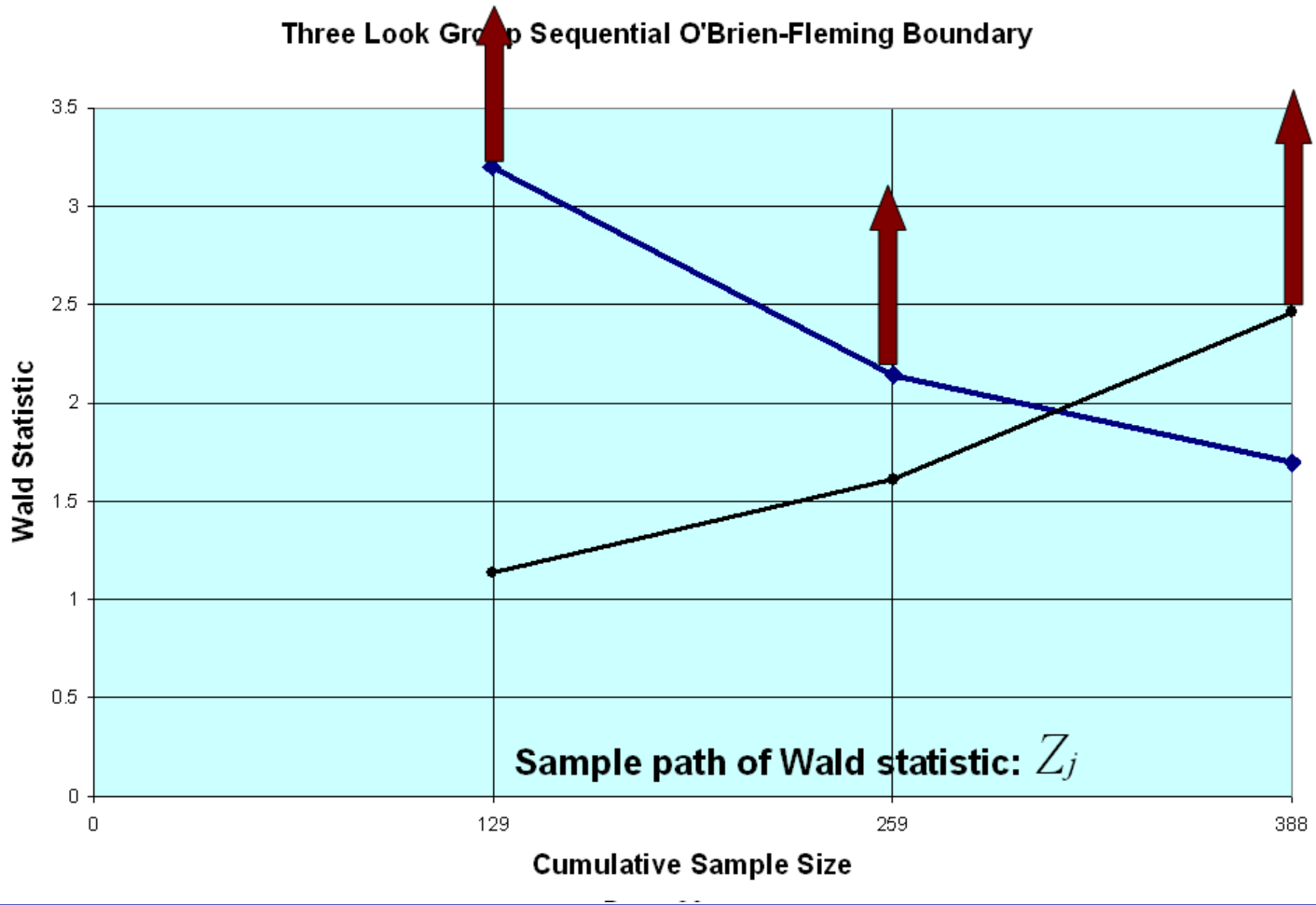

Three Look Group Sequential O'Brien-Fleming Boundary
Wald Statistic
3.5
3
2.5
2
1.5
1
0.5
0
0
129
259
388
Cumulative Sample Size
Sample path of Wald Statistic: Z_j

Critical Region for the Stage-Wise Adjusted P-Value



Stage-Wise Adjusted P-Value p(h)

Critical region is evaluated under $\delta = h$. SWACI is obtained when $p(h) = \alpha$



Adaptive SWACI Method

Suppose adaptive change at look L of primary trial with $Z_L = z_L$

- **H_h is rejected in primary trial if $p(h) \leq \epsilon$**
- **The probability of this event given $Z_L = z_L$ and $\delta = h$ is**

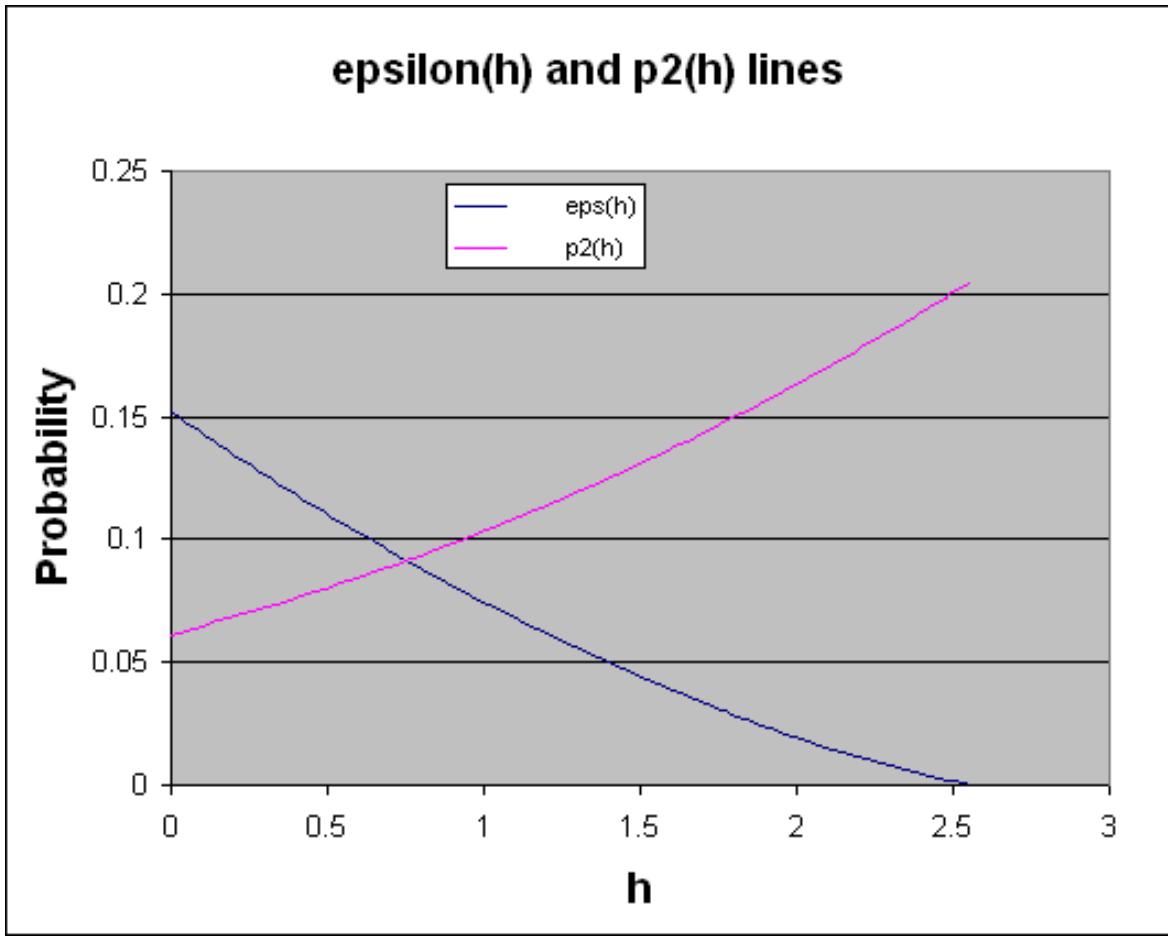
$$\epsilon(h) = P_h \bigcup_{j=L+1}^{K} (p(h) \leq \alpha | z_L)$$

  given $\delta = h$

- **Implement the secondary trial with $\epsilon(h)$ as the type-1 error and reject H_h if**

$$p^{(2)}(h) \leq \epsilon(h)$$

Plots of $\epsilon(h)$ and $p^{(2)}(h)$ vs. h



The curves intersect at $h = 0.752 = \underline{\delta}$, the lower 95% confidence bound

Theorem for Computing $\epsilon(h)$

Define α-absorbing constants $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_{K-1}$ to be such that, for any $k = 1, 2, \ldots K-1$,

$$P_{\delta_k}(\bigcup_{j=1}^{k}\{Z_j \geq b_j\}) = \alpha$$

Further, define $\delta_0 = \infty$ and $\delta_K = -\infty$, so that for every real valued h we can find a unique index $k(h) \equiv k$ such that $\delta_k \leq h < \delta_{k-1}$. For each such h define the 'threshhold boundary value' $b_k(h)$ to be such that

$$P_h(\bigcup_{j=1}^{k-1}\{Z_j \geq b_j\} \cup \{Z_K \geq b_k(h)\}) = \alpha$$

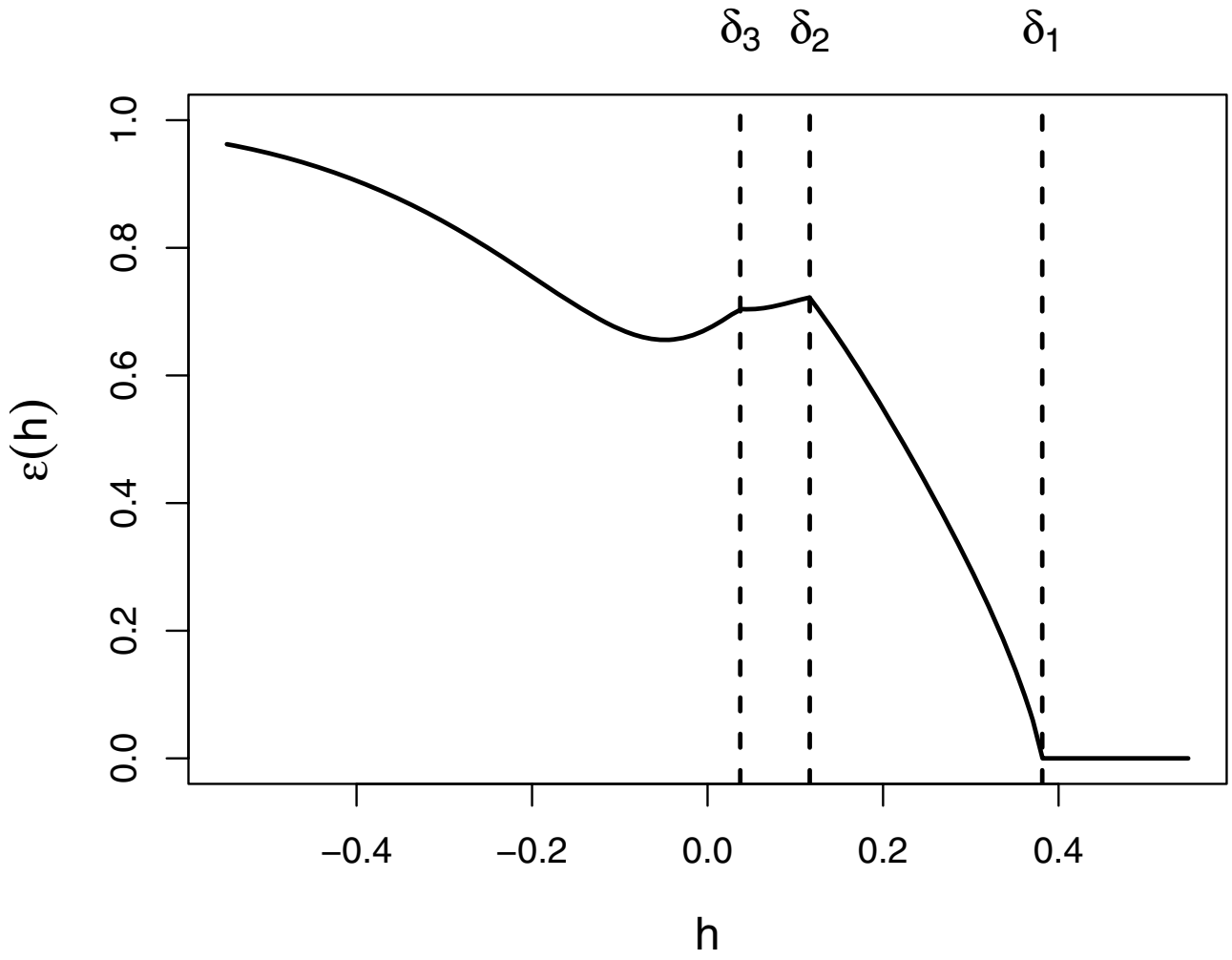
Then

$$\epsilon(h) = \begin{cases} 0 & \text{if } h \geq \delta_L \\ P_h(\bigcup_{j=1}^{k-1}\{Z_j \geq b_j\} \cup \{Z_k \geq b_k(h) | Z_L = z_L\}) & \text{if } \delta_k \leq h < \delta_{k-1} \end{cases}$$

Behavior of $\epsilon(h)$

- Computation of $\epsilon(h)$ is difficult and requires a numerical algorithm
- The function is not guaranteed to be monotone unless adaptation takes place at penultimate look
- Consequently the solution to $\epsilon(h) = p^{(2)}(h)$ may not be unique
- However, extensive simulations involving worst-case scenarios demonstrate that the CI's are exact and the point estimates are median unbiased

Figure 1: An example of non-monotonicity of the conditional rejection probability function $\epsilon(h)$



Results from Extensive Simulations

Table 1: 25,000 simulations; 3-look primary trial; adaptation at look 1; 3-look secondary trial

Group Sequential Design	True δ	Actual Coverage of 95% CI		Median of $\underline{\delta}_{0.5}$	
		SWACI	RCI	SWACI	RCI
LD(OBF)–LD(PK)	0.0	0.9487	0.9495	-0.000185	-0.0897
LD(OBF)–LD(PK)	0.15	0.9509	0.9780	0.1501	0.1331
LD(OBF)–LD(PK)	0.3	0.9506	0.9854	0.2997	0.2309
LD(OBF)–LD(PK)	0.5	0.9496	0.9965	0.5011	0.4474

Table 2: 25,000 simulations; 4-look primary trial; adaptation at look 1; variable-look secondary trial

Group Sequential Design	True δ	Actual Coverage of 97.5% CI		Median of $\underline{\delta}_{0.5}$	
		SWACI	RCI	SWACI	RCI
LD(OBF)–LD(PK)	0.0	0.9742	0.9758	-0.0003	-0.0221
LD(OBF)–LD(PK)	0.15	0.9746	0.9819	0.1495	0.1362
LD(OBF)–LD(PK)	0.3	0.9754	0.9803	0.2985	0.2555
LD(OBF)–LD(PK)	0.5	0.9767	0.9841	0.4965	0.4765

Concluding Remarks: RCI Method

- Has conservative coverage and negatively biased point estimates for group sequential designs
- Has exact coverage and median unbiased point estimates for two-stage designs with no early stopping
- Specializes to the method of Lehmacher and Wassmer (1999) when sample size re-estimation is the only adaptation

Concluding Remarks: SWACI Method

- Conservative coverage and negatively biased point estimates guaranteed in general
- Exact coverage and median unbiased point estimates guaranteed if adaptation takes place at penultimate look
- Extensive simulations demonstrate that in fact exact coverage and median unbiased point estimates are obtained under all conditions
- Only currently available method with this property