

# Adaptive vs Group Sequential Designs: Some Trade-Offs

Cyrus R. Mehta(†) and Anastasios A. Tsiatis(‡)

(†) Cytel Software Corporation and Harvard University

(‡) North Carolina State University

ENAR-2002

Tampa, Florida

# Motivating Example: Cui, Hung and Wang (Biometrics, 1999)

- Phase III clinical trial for prevention of myocardial infarct in patients undergoing coronary artery bypass graft surgery
- Design parameters:
  - Placebo MI rate,  $\pi_c = 22\%$ . Treatment MI rate,  $\pi_t = 11\%$ . Looking for an improvement of  $\delta = \pi_t - \pi_c = -11\%$
  - One sided test with  $\alpha = 0.025$  and 95% power to detect  $\delta = -11\%$
  - One interim look at midcourse
- Required up-front commitment: 579 patients (both arms)

## But when the interim look was taken:

- The drop in MI was only half the design specification.  
 $\hat{\pi}_c \approx 22\%$ , and  $\hat{\pi}_t \approx 16.5\%$ ; i.e.,  $\hat{\delta} \approx -6.5\%$ , whereas the study was powered for  $\delta = -11\%$
- Sponsor was concerned that if these estimates accurately reflected the true parameters, the trial would only have 39% power
- At the time there was no valid procedure to increase the sample size without inflating the type-1 error.
- Eventually the trial failed to show statistical significance

# Could this trial have been saved?

Yes! An adaptive approach might have saved this trial

- Basic idea of adaptive trials:
  - increase the sample size if the observed data suggest loss of power
  - at the same time make appropriate adjustments to preserve the type-1 error
- Extensive Literature: Bauer and Kohne (1994), Proschan and Hunsberger (1995), Cui, Hung and Wang (1999), Shen and Fisher (1999), Lemacher and Wassmer (1999), Liu and Chi (2001), and many more

## Is there a down side to adaptive designs?

Yes! They are not as efficient as conventional group sequential trials that don't adapt.

- **General Result:** Tsiatis and Mehta (Biometrika, 2003) show that **any adaptive procedure** for increasing the sample size can be improved uniformly by a classical group sequential design. (Tsiatis will present this theorem in Session # 60, at 4:15 PM, Today.)
- **Specific Result:** For the myocardial infarct example we will show that we could design a **substantially** better trial within the classical group sequential framework than by the adaptive method of Cui, Hung and Wang (CHW).

## The CHW adaptive method

- Two-stage trial is planned for:  
 $n_1$  observations/arm at stage 1,  
 $n_{\max}$  observations/arm overall.
- At stage 1 compute  $f = \frac{\delta}{\hat{\delta}}$
- If  $f > f_0$ , increase the final sample size to

$$n_{\max}^* = f^2 n_{\max}$$

## Usual test statistic is inappropriate for final look

- At the final stage the usual test statistic is proportional to

$$T = \frac{\sum_{i=1}^{n_{\max}^*} X_{ti}}{n_{\max}^*} - \frac{\sum_{i=1}^{n_{\max}^*} X_{ci}}{n_{\max}^*} = \hat{\pi}_t - \hat{\pi}_c$$

- But this statistic will not preserve the type-1 error because we have used the stage 1 data to increase the sample size from  $n_{\max}$  to  $n_{\max}^*$

## Downweight the final test statistic

$$T = \sqrt{\frac{n_1}{n_{\max}}} \sum_{i=1}^{n_1} \frac{X_{it} - X_{ic}}{\sqrt{n_1}} + \sqrt{\frac{n_{\max} - n_1}{n_{\max}}} \sum_{i=n_1+1}^{n_{\max}^*} \frac{X_{it} - X_{ic}}{\sqrt{(n_{\max}^* - n_1)}}$$

- If we used  $n_{\max}^*$  instead of  $n_{\max}$  in the “weights” for the two terms,  $T$  would be proportional to  $\hat{\pi}_t - \hat{\pi}_c$ . Every patient would count equally in the value of  $T$
- But instead, by using the above weighted sum, the contribution of the stage 1 patients counts more than the contribution of the stage 2 patients.
- As a result the type-1 error is preserved despite increasing the sample size from  $n_{\max}$  to  $n_{\max}^*$

Two-look design with no early stopping but possible adaptation at look 1

| <b>East 3: Superiority Trials with Binomial Design for</b> |          |
|------------------------------------------------------------|----------|
| Plan ID                                                    | Plan1    |
| <b>Test Parameters</b>                                     |          |
| 1-Sided or 2-Sided Test                                    | 1-Sided  |
| Significance Level (Alpha)                                 | 0.025    |
| Power (1 - Beta)                                           | 0.95     |
| Assigned Fraction (Treatment)                              | 0.5      |
| <b>Boundary Parameters</b>                                 |          |
| Planned Number of Looks                                    | 2        |
| Spacing of Analysis                                        | Equal    |
| Hypothesis to be Rejected                                  | H0 Only  |
| Boundary Family                                            | SpF(Pub) |
| Boundary to Reject H0                                      | Gm(-12)  |
| Boundary to Reject H1                                      |          |
| <b>Binomial parameters under H1</b>                        |          |
| Proportion Response (Control: n_c)                         | 0.22     |
| Proportion Response (Treatment: n_t)                       | 0.11     |
| <b>Accrual (Subjects)</b>                                  |          |
| Maximum                                                    | 579      |
| Expected Under H0                                          | 579      |
| Expected Under H1                                          | 550      |
| Expected Under H1/2                                        | 577      |

Simulation Results:  $\pi_t = 16.5\%$ . No adaptation at look 1.

**East 3: Simulation of Superiority Sequential Trials (Cui, Hung, Wang Adaptive Method)**

| Design Parameters for Plan1-   |           | Simulation Boundary |          |           | Overall Simulation Results |            |             |             |                   |        |  |
|--------------------------------|-----------|---------------------|----------|-----------|----------------------------|------------|-------------|-------------|-------------------|--------|--|
| 1-Sided or 2-Sided Test        | 1-Sided   | Look                | Boundary | Simulated | Avg.                       | Avg Sample | # Rejecting | # Unable to | Total simulations |        |  |
| Significance Level(Alpha)      | 0.025     | #                   | Info     | Test Stat | Information                | Size       | H0          | reject H0   | Count             | %      |  |
| Power(1-Beta)                  | 0.95      | 1                   | 289.44   | -3.8388   | -1.8832                    | 448.65     | 289.00      | 23          | 23                | 0.23%  |  |
| End Point                      | Binomial  | 2                   | 578.87   | -1.9601   | -1.8751                    | 887.11     | 547.78      | 3635        | 6342              | 99.77% |  |
| Proportion Response(Control)   | 0.22      | 3                   |          |           |                            |            |             |             |                   |        |  |
| Proportion Response(Treatment) | 0.11      | 4                   |          |           |                            |            |             |             |                   |        |  |
| Assigned Fraction              | 0.5       | 5                   |          |           |                            |            |             |             |                   |        |  |
| Planned # of Looks             | 2         | 6                   |          |           |                            |            |             |             |                   |        |  |
| Spacing of Analysis            | Equal     | 7                   |          |           |                            |            |             |             |                   |        |  |
| Hypothesis to be rejected      | H0 Only   | 8                   |          |           |                            |            |             |             |                   |        |  |
| Boundary Family                | SpF(Pub)  | 9                   |          |           |                            |            |             |             |                   |        |  |
| Boundary to Reject H0          | Gm(-12)   | 10                  |          |           |                            |            |             |             |                   |        |  |
| Boundary to Reject H1          |           | 11                  |          |           |                            |            |             |             |                   |        |  |
|                                |           | 12                  |          |           |                            |            |             |             |                   |        |  |
| <b>Design Outputs</b>          |           |                     |          |           | <b>Total</b>               | 886.11     | 547.18      | 3658        | 6342              | 10000  |  |
| Max. Sample Size (Nmax)        | 579       | <b>%</b>            |          |           |                            |            | 36.58%      | 63.42%      |                   |        |  |
| Max. Information (Imax)        | 1074.2115 |                     |          |           |                            |            |             |             |                   |        |  |
| <b>Simulation Parameters</b>   |           |                     |          |           |                            |            |             |             |                   |        |  |
| Proportion Response(Control)   | 0.22      |                     |          |           |                            |            |             |             |                   |        |  |
| Proportion Response(Treatment) | 0.165     |                     |          |           |                            |            |             |             |                   |        |  |
| Use Nmax Till 'L' Looks, L =   | 1         |                     |          |           |                            |            |             |             |                   |        |  |
| Crit. to Rescale Nmax (f_0)    | 1         |                     |          |           |                            |            |             |             |                   |        |  |
| Max. Usable Sample Size        | 539       |                     |          |           |                            |            |             |             |                   |        |  |
| Number of Trials               | 10000     |                     |          |           |                            |            |             |             |                   |        |  |
| Refresh Every 'n' Trials, n =  | 1000      |                     |          |           |                            |            |             |             |                   |        |  |
| Simulation Starting Seed       | Clock     |                     |          |           |                            |            |             |             |                   |        |  |

**Design Outputs**

Max. Sample Size (Nmax): 579  
 Max. Information (Imax): 1074.2115

**Simulation Parameters**

Proportion Response(Control): 0.22  
 Proportion Response(Treatment): 0.165  
 Use Nmax Till 'L' Looks, L = 1  
 Crit. to Rescale Nmax (f\_0): 1  
 Max. Usable Sample Size: 539  
 Number of Trials: 10000  
 Refresh Every 'n' Trials, n = 1000  
 Simulation Starting Seed: Clock

Run    Single Step

Reset    Stop    Help

Simulation Seed =  
 Elapsed Time = 0:00:05

Simulation Results:  $\pi_t = 16.5\%$ . With adaptation at look 1.

**East 3: Simulation of Superiority Sequential Trials (Cui, Hung, Wang Adaptive Method)**

| Design Parameters for Plan1-   |           | Simulation Boundary |          |           | Overall Simulation Results |             |             |             |                   |        |
|--------------------------------|-----------|---------------------|----------|-----------|----------------------------|-------------|-------------|-------------|-------------------|--------|
| 1-Sided or 2-Sided Test        | 1-Sided   | Lock                | Boundary | Simulated | Avg.                       | Avg. Sample | # Rejecting | # Unable to | Total simulations |        |
| Significance Level(Alpha)      | 0.025     | #                   | Info     | Test Stat | Information                | Size        | H0          | reject H0   | Count             | %      |
| Power(1-Beta)                  | 0.95      | 1                   | 289.44   | -3.8388   | -0.8229                    | 466.16      | 36          |             | 36                | 0.36%  |
| End Point                      | Binomial  | 2                   | 578.87   | -1.9501   | -3.7923                    | 2841.62     | 7853        | 2101        | 9964              | 99.64% |
| Proportion Response(Control)   | 0.22      | 3                   |          |           |                            |             |             |             |                   |        |
| Proportion Response(Treatment) | 0.11      | 4                   |          |           |                            |             |             |             |                   |        |
|                                |           | 5                   |          |           |                            |             |             |             |                   |        |
| Assigned Fraction              | 0.5       | 6                   |          |           |                            |             |             |             |                   |        |
| Planned # of Looks             | 2         | 7                   |          |           |                            |             |             |             |                   |        |
| Spacing of Analysis            | Equal     | 8                   |          |           |                            |             |             |             |                   |        |
| Hypothesis to be rejected      | H0 Only   | 9                   |          |           |                            |             |             |             |                   |        |
| Boundary Family                | SpF(Pub)  | 10                  |          |           |                            |             |             |             |                   |        |
| Boundary to Reject H0          | Gm(-12)   | 11                  |          |           |                            |             |             |             |                   |        |
| Boundary to Reject H1          |           | 12                  |          |           |                            |             |             |             |                   |        |
| <b>Design Outputs</b>          |           | <b>Total</b>        |          |           | 2833.07                    | 1758.04     | 7899        | 2101        | 10000             |        |
| Max. Sample Size (Nmax)        | 579       | <b>%</b>            |          |           |                            |             | 78.99%      | 21.01%      |                   |        |
| Max. Information (Imax)        | 1074.2115 |                     |          |           |                            |             |             |             |                   |        |
| <b>Simulation Parameters</b>   |           |                     |          |           |                            |             |             |             |                   |        |
| Proportion Response(Control)   | 0.22      |                     |          |           |                            |             |             |             |                   |        |
| Proportion Response(Treatment) | 0.165     |                     |          |           |                            |             |             |             |                   |        |
| Use Nmax Till 'L' Looks, L =   | 1         |                     |          |           |                            |             |             |             |                   |        |
| Crit. to Rescale Nmax (f_0)    | 1         |                     |          |           |                            |             |             |             |                   |        |
| Max. Usable Sample Size        | 3000      |                     |          |           |                            |             |             |             |                   |        |
| Number of Trials               | 10000     |                     |          |           |                            |             |             |             |                   |        |
| Refresh Every 'n' Trials, n =  | 1000      |                     |          |           |                            |             |             |             |                   |        |
| Simulation Starting Seed       | Clock     |                     |          |           |                            |             |             |             |                   |        |

**Design Outputs**

Max. Sample Size (Nmax) 579  
Max. Information (Imax) 1074.2115

**Simulation Parameters**

Proportion Response(Control) 0.22  
Proportion Response(Treatment) 0.165  
Use Nmax Till 'L' Looks, L = 1  
Crit. to Rescale Nmax (f\_0) 1  
Max. Usable Sample Size 3000  
Number of Trials 10000  
Refresh Every 'n' Trials, n = 1000  
Simulation Starting Seed Clock

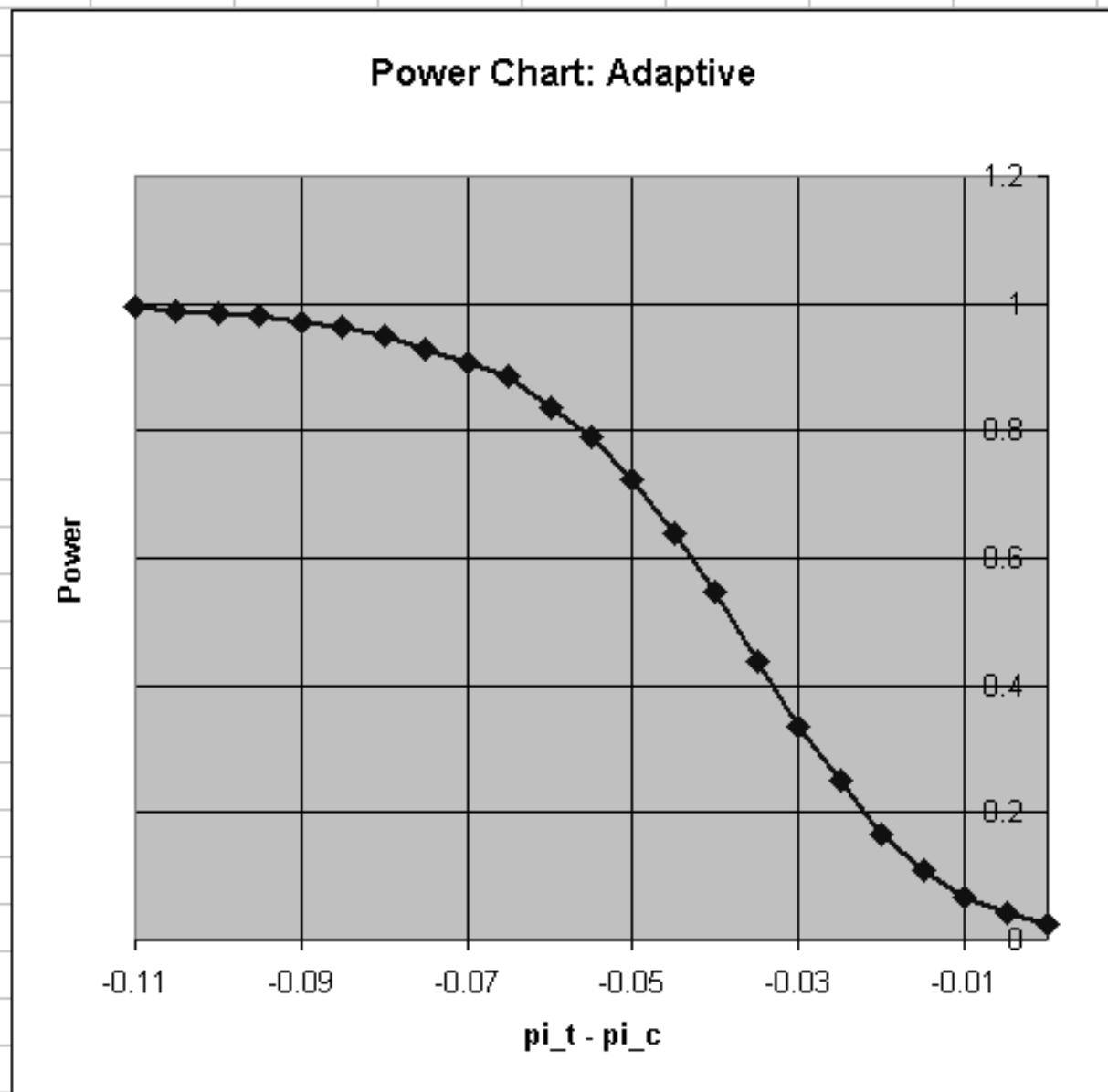
Run Single Step  
Reset Stop Help

Simulation Seed =  
Elapsed Time = 0:00:13

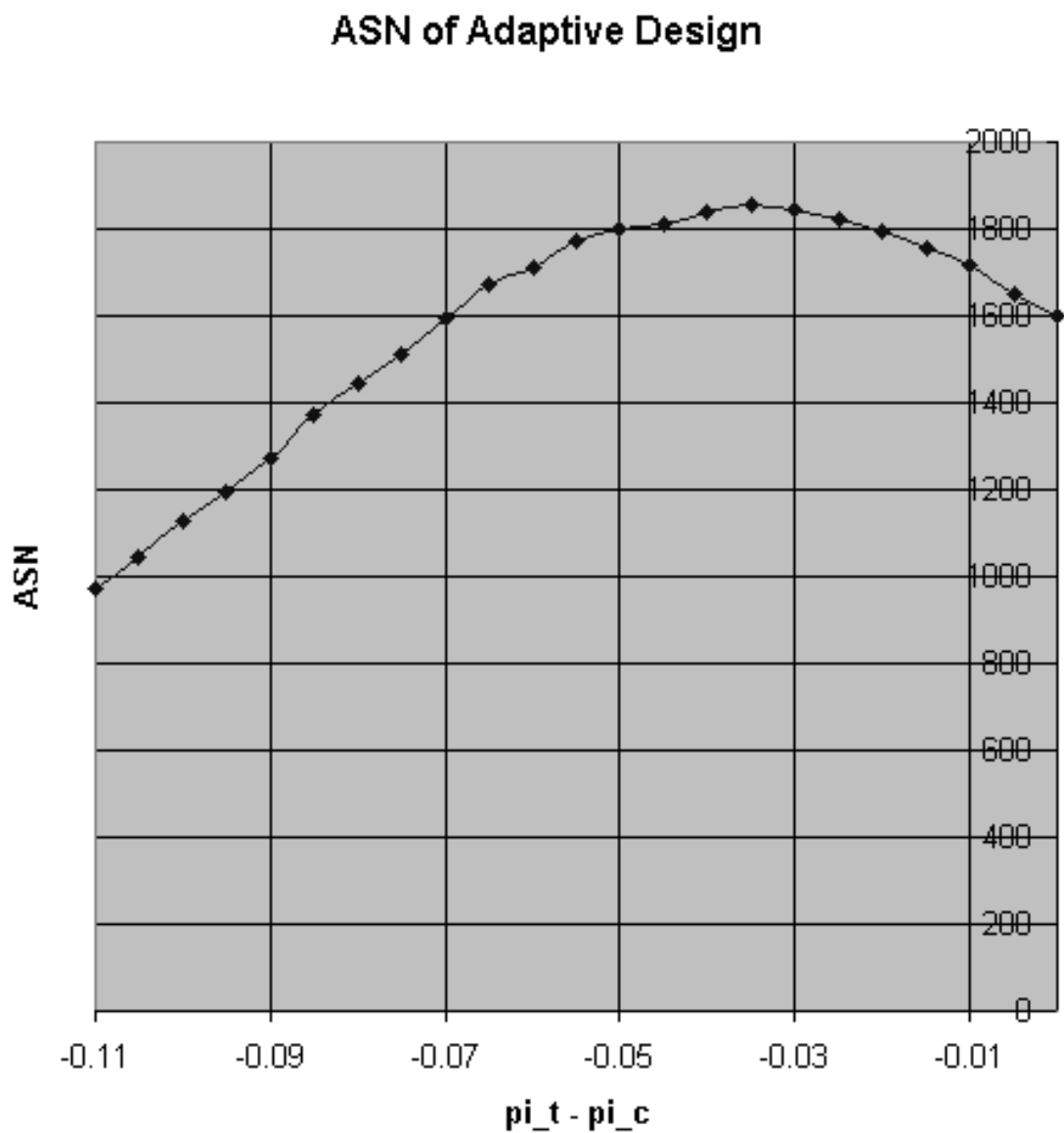
The line graph plots Normal Critical Point (y-axis, -6 to 6) against Sample Size (x-axis, 0 to 800). Two lines are shown: 'Reject H0' (black line with diamond markers) and 'Test Statistic' (grey line with square markers). The 'Reject H0' line starts at approximately (200, -3.8) and rises to (600, -2.0). The 'Test Statistic' line starts at approximately (200, -0.8) and falls to (600, -3.8).

The bar chart shows Frequency (y-axis, 0 to 12000) versus Look # (x-axis, 1 to 12). Two series are shown: 'Reject H0' (black bars) and 'Unable to reject H0' (grey bars). At Look 1, the 'Reject H0' frequency is approximately 10000, and the 'Unable to reject H0' frequency is approximately 36. For Look 2, the 'Reject H0' frequency is approximately 7853, and the 'Unable to reject H0' frequency is approximately 2101. All other looks have zero frequency.

| pi_t-pi_c | power |
|-----------|-------|
| -0.11     | 0.994 |
| -0.105    | 0.989 |
| -0.1      | 0.986 |
| -0.095    | 0.982 |
| -0.09     | 0.972 |
| -0.085    | 0.963 |
| -0.08     | 0.948 |
| -0.075    | 0.927 |
| -0.07     | 0.907 |
| -0.065    | 0.886 |
| -0.06     | 0.835 |
| -0.055    | 0.791 |
| -0.05     | 0.723 |
| -0.045    | 0.638 |
| -0.04     | 0.546 |
| -0.035    | 0.437 |
| -0.03     | 0.334 |
| -0.025    | 0.252 |
| -0.02     | 0.166 |
| -0.015    | 0.111 |
| -0.01     | 0.067 |
| -0.005    | 0.041 |
| 0         | 0.025 |



| pi_t-pi_c | asn  |
|-----------|------|
| -0.11     | 973  |
| -0.105    | 1045 |
| -0.1      | 1126 |
| -0.095    | 1194 |
| -0.09     | 1274 |
| -0.085    | 1373 |
| -0.08     | 1445 |
| -0.075    | 1509 |
| -0.07     | 1593 |
| -0.065    | 1671 |
| -0.06     | 1709 |
| -0.055    | 1771 |
| -0.05     | 1798 |
| -0.045    | 1812 |
| -0.04     | 1839 |
| -0.035    | 1857 |
| -0.03     | 1847 |
| -0.025    | 1823 |
| -0.02     | 1797 |
| -0.015    | 1757 |
| -0.01     | 1718 |
| -0.005    | 1649 |
| 0         | 1601 |



# Compare the Adaptive and Group Sequential Approaches

**Adaptive Approach:** Start small, then ask for more.

- We designed the trial for 579 patients, to detect a drop in the event rate of 11%.
- But mid-way through the trial, when we saw that the drop was only 6.5% we were willing to increase the sample size.

**Group Sequential Approach:** Ask for more up front, then cut back.

- Start out by designing a trial that has good power to detect a drop of 6.5%.
- But have high probability of early detection and termination if actual drop is 11%.
- By judicious selection of sample size, number of looks, spacing of looks and stopping boundaries, one can improve on the adaptive approach uniformly.

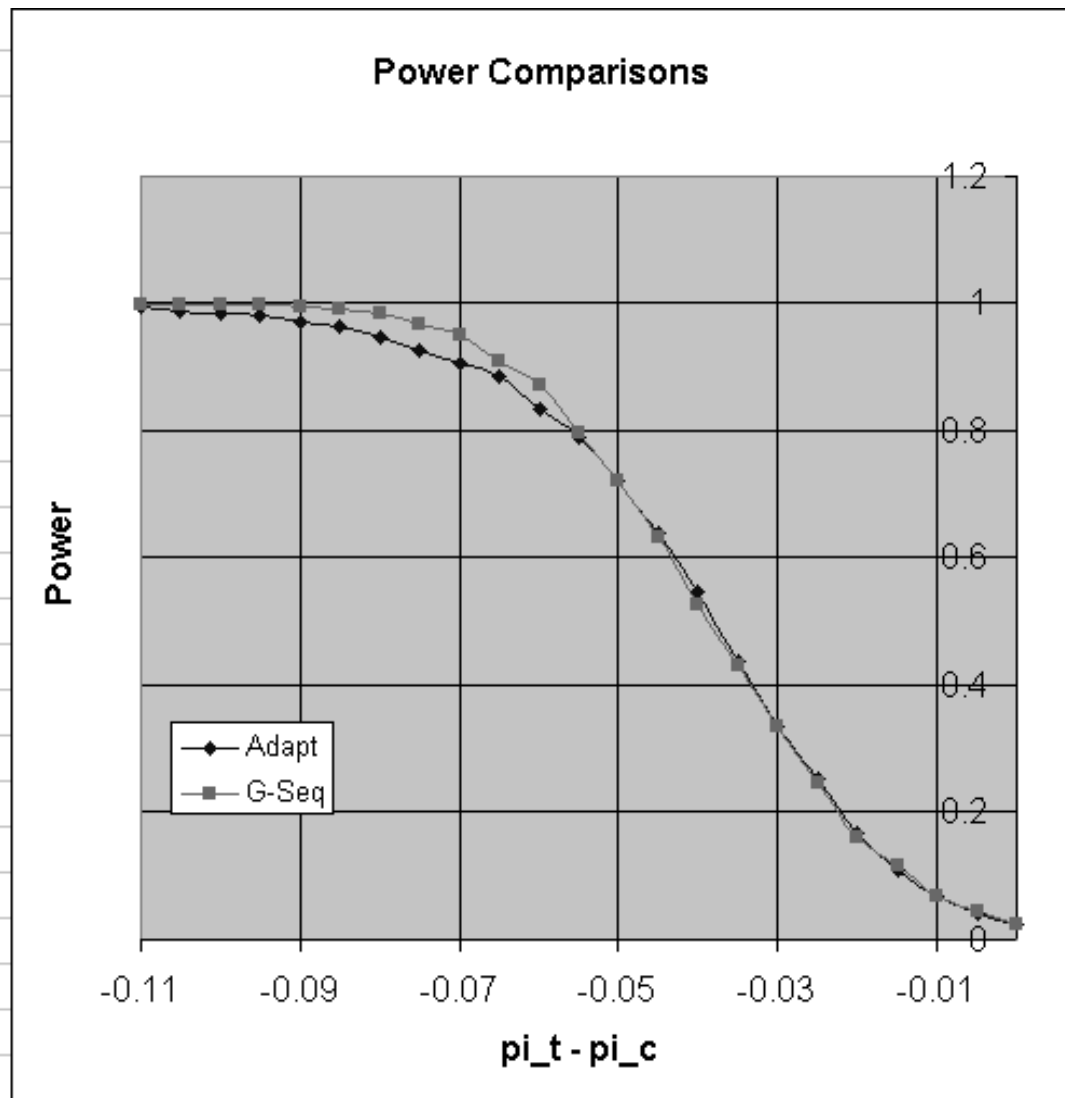
## Tools to decrease ASN of a group sequential design

1. Select both an efficacy and a futility boundary.
2. Add more interim looks.
3. Change spacing of the looks.
4. Increase the concavity of the spending function.

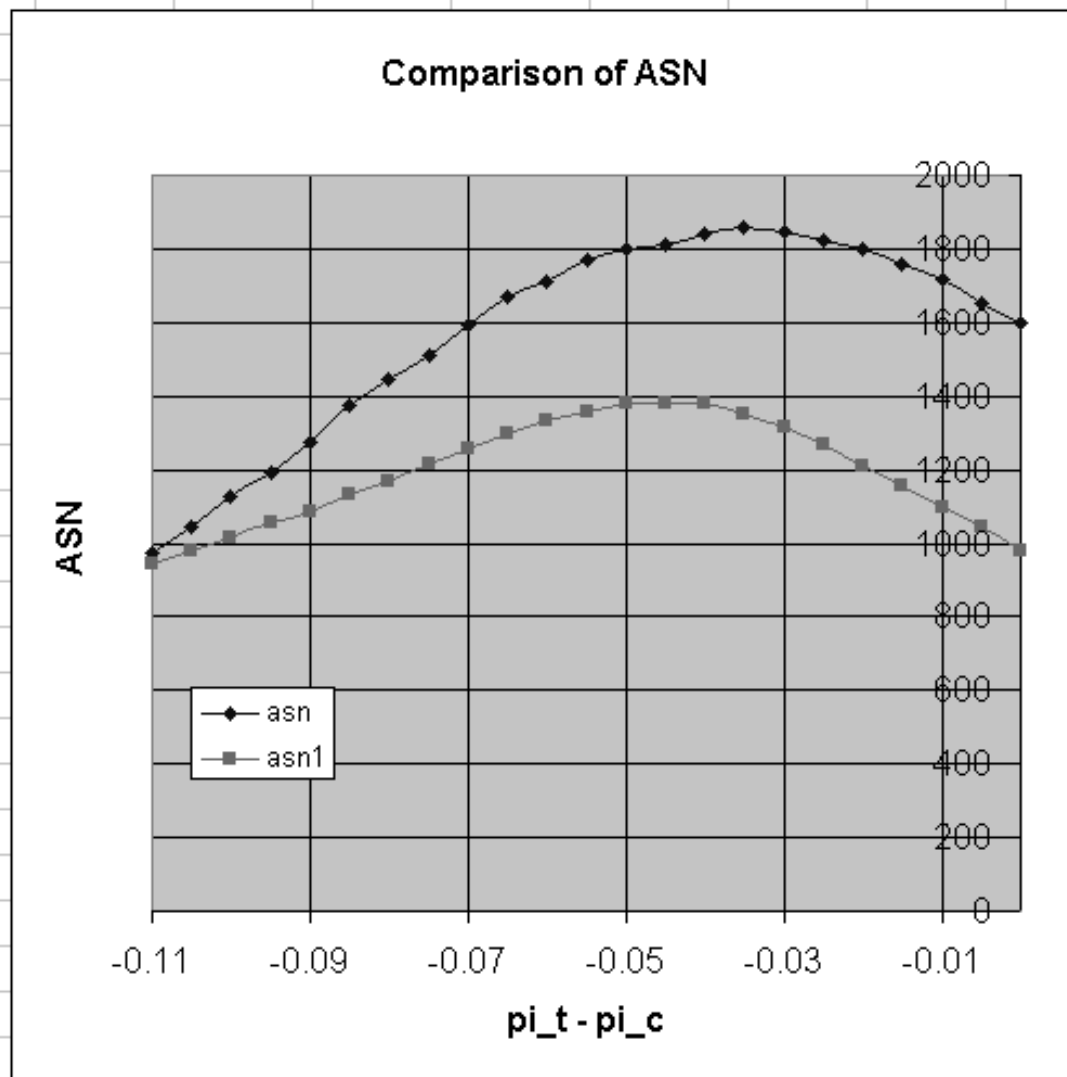
### East 3: Superiority Trials with Binomial Design for EastBook1

| Plan ID                              | Plan1    | Plan2    | Plan3    |
|--------------------------------------|----------|----------|----------|
| <b>Test Parameters</b>               |          |          |          |
| 1-Sided or 2-Sided Test              | 1-Sided  | 1-Sided  | 1-Sided  |
| Significance Level (Alpha)           | 0.025    | 0.025    | 0.025    |
| Power (1 - Beta)                     | 0.95     | 0.95     | 0.7544   |
| Assigned Fraction (Treatment)        | 0.5      | 0.5      | 0.5      |
| <b>Boundary Parameters</b>           |          |          |          |
| Planned Number of Looks              | 2        | 2        | 3        |
| Spacing of Analysis                  | Equal    | Equal    | Unequal  |
| Hypothesis to be Rejected            | H0 Only  | H0 Only  | H0 or H1 |
| Boundary Family                      | SpF(Pub) | SpF(Pub) | SpF(Pub) |
| Boundary to Reject H0                | Gm(-12)  | Gm(-12)  | Gm(0)    |
| Boundary to Reject H1                |          |          | Gm(0)    |
| <b>Binomial parameters under H1</b>  |          |          |          |
| Proportion Response (Control: n_c)   | 0.22     | 0.22     | 0.22     |
| Proportion Response (Treatment: n_t) | 0.11     | 0.165    | 0.165    |
| <b>Accrual (Subjects)</b>            |          |          |          |
| Maximum                              | 579      | 2658     | 1700     |
| Expected Under H0                    | 579      | 2658     | 954      |
| Expected Under H1                    | 550      | 2527     | 1204     |
| Expected Under H1/2                  | 577      | 2651     | 1180     |

| pi_t-pi_c | Adapt | G-Seq |
|-----------|-------|-------|
| -0.11     | 0.994 | 0.999 |
| -0.105    | 0.989 | 0.999 |
| -0.1      | 0.986 | 0.999 |
| -0.095    | 0.982 | 0.998 |
| -0.09     | 0.972 | 0.996 |
| -0.085    | 0.963 | 0.991 |
| -0.08     | 0.948 | 0.986 |
| -0.075    | 0.927 | 0.968 |
| -0.07     | 0.907 | 0.949 |
| -0.065    | 0.886 | 0.909 |
| -0.06     | 0.835 | 0.873 |
| -0.055    | 0.791 | 0.796 |
| -0.05     | 0.723 | 0.723 |
| -0.045    | 0.638 | 0.631 |
| -0.04     | 0.546 | 0.527 |
| -0.035    | 0.437 | 0.431 |
| -0.03     | 0.334 | 0.334 |
| -0.025    | 0.252 | 0.246 |
| -0.02     | 0.166 | 0.162 |
| -0.015    | 0.111 | 0.115 |
| -0.01     | 0.067 | 0.068 |
| -0.005    | 0.041 | 0.044 |
| 0         | 0.025 | 0.025 |



| pi_t-pi_c | Adapt | G-Seq |
|-----------|-------|-------|
| -0.11     | 973   | 942   |
| -0.105    | 1045  | 979   |
| -0.1      | 1126  | 1013  |
| -0.095    | 1194  | 1054  |
| -0.09     | 1274  | 1088  |
| -0.085    | 1373  | 1130  |
| -0.08     | 1445  | 1167  |
| -0.075    | 1509  | 1214  |
| -0.07     | 1593  | 1255  |
| -0.065    | 1671  | 1295  |
| -0.06     | 1709  | 1336  |
| -0.055    | 1771  | 1358  |
| -0.05     | 1798  | 1380  |
| -0.045    | 1812  | 1382  |
| -0.04     | 1839  | 1379  |
| -0.035    | 1857  | 1352  |
| -0.03     | 1847  | 1315  |
| -0.025    | 1823  | 1270  |
| -0.02     | 1797  | 1211  |
| -0.015    | 1757  | 1156  |
| -0.01     | 1718  | 1098  |
| -0.005    | 1649  | 1042  |
| 0         | 1601  | 980   |



## Do practical advantages of adaptive designs offset their loss of efficiency?

There has been a great deal of recent interest in adaptive designs. Given that they can always be dominated by standard group sequential trials that are relatively easy to find why should we adopt them? I have heard three reasons given.

## Reason 1: There is great uncertainty about the effect size

Is that a valid reason?

The effect size is never known at the design stage. Nor does it have to be known.

We are only required to specify a **clinically meaningful effect size** for the power calculations.

## Reason 2: The clinically meaningful effect size is a moving target

- At the design stage you felt that the clinically meaningful drop in MI is 11%. It is not worth running the trial to detect a smaller drop.
- But after the trial starts you feel that even a drop in MI of 6.5% is clinically meaningful. Had you known this at the start, you could have designed a better trial with the standard group sequential approach.
- Why this change of heart? We need published discussions, with specific examples, of why the clinically meaningful endpoint is a moving target.

## Reason 3: Budgetary Considerations

You might find it easier to request a small budget at the design stage and ask for supplemental funding after seeing the interim data.

- Could this argument be countered by showing management the simulation and ASN results?
- Could you request supplemental funding for an increasing the sample size without unblinding the interim results to the trial sponsor?
- Again, we need published discussions, with specific examples of how the actual logistics of running an adaptive trial would be handled. Who would make the decision to adjust the sample size? Who would know about the decision? What is the potential for bias?

# Conclusion

- We believe that under certain specific conditions you might indeed want to change your mind about certain design parameters (sample size, end-point, test statistic etc).
- In such situations adaptive schemes serve as an excellent “rescue operation”.
- But these are usually special cases. Real situations where one would want to incorporate the adaptive approach, at the design stage itself, need to be brought out and discussed. There is a heavy trade-off in terms of loss of efficiency.