**Supplementary Appendix**

**Simulation results**

1. ***Misspecification of effect size – Two Experimental Treatment (k=2) and A Control***

*Effect size scenario 3* () The behaviour of the five approaches for SSR in this scenario is very similar to that in scenario 2. Therefore we only focus on SSR based on the conditional power approach in the following comparisons. Supplementary Figure 7 shows that the relative performances of the three adaptive designs in this scenario are very similar to that in scenario 2. If no SSR is performed, the power of the FGS design is similar to (when selecting the best treatment) or slightly larger than (when keeping all promising treatments) that of INC design, while the power of FiC is slightly lower. If SSRCP is performed, when selecting the best treatment, all three designs have similar power curves while INC and FGS have much smaller . When keeping all promising treatments, the power of FGS are much higher than that of the other two designs. Note that when keeping all promising treatments, the INC and FiC design still cannot reach the targeted 80% power for even after SSRCP. But it can be seen from Supplementary Figure 8 that the targeted power can be reached for INC and FiC if SSRCP, Ob is performed, with the cost of a larger . As = 1.5 is enough to achieve the desired power, we do not consider larger values of in this scenario. Supplementary Figure 9 compares the difference between the two selection rules. It shows that the power and of the three adaptive designs follow the same pattern as that in scenario 2 (Supplementary Figure 1).

## Misspecification of standard deviation (SD)

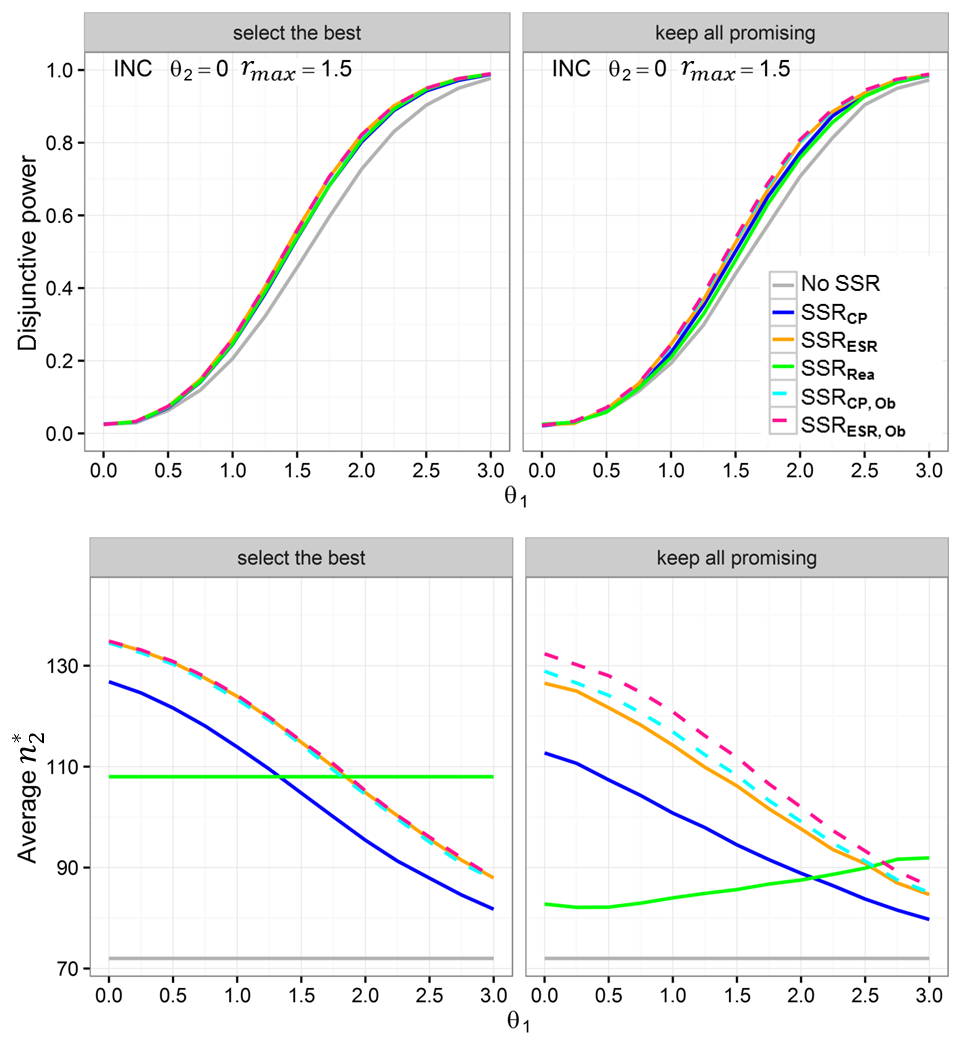
Supplementary Figure 10 shows the disjunctive power for the scenario and when the true underlying SD is different from the planned SD. As expected, the power decreases if the true underlying SD is larger than planned, regardless of the design and treatment selection rule used. For example, if the true underlying SD is 7, the power is only around 70% when both treatments have the planned effect size (). Supplementary Figure 11 shows that performing SSRCP can substantially improve the power and the desired 80% power is reached at . When selecting the best, all three designs have similar power while INC and FGS have smaller . When keeping all promising, FGS has the largest power while INC has similar power to that of FiC with a smaller .

## Different timing (information fraction ) for interim analysis

The planned sample size per arm is 152, 144 and 146 for and , respectively. Supplementary Figure 12 and 13 demonstrate the impact of the timing for interim analysis on the disjunctive power and average new sample size per arm ( when and . When no SSR is performed, for FGS and INC designs with keeping all promising rule, the timing for interim analysis does not have a significant impact on power, but for other cases it seems the later the interim analysis the higher the power for a fixed value of . When SSRCP, Ob is performed, the power increases with the increase of for a fixed value of . FGS design with keeping all promising rule seems to be more robust to the timing of interim analysis than other methods. Even though the power is usually highest at among the three information fraction considered here, is also the highest. The power curve for is close to that for while is much smaller. Therefore, based on the tradeoff between power and , seems to be a reasonable choice for interim analysis.

Supplementary Table 1. List of abbrevations.

|  |  |  |
| --- | --- | --- |
| Abbreviation | |  |
| INC | Adaptive design based on inverse normal combination function | |
| FiC | Adaptive design based on Fisher's combination function | |
| FGS | Flexible group sequential design | |
| SSR | Sample size re-estimation | |
| CP | Conditional power | |
| ESR | Effect size ratio | |
| FWER | Familywise error rate | |
|  | The ratio of maximum sample size allowed per arm to the planned sample size per arm | |



Supplementary Figure 1. Comparison of SSR methods under INC design when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 4.tif

Supplementary Figure 2. Comparison of SSR methods under FiC design when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 5.tif

Supplementary Figure 3. Comparison of SSR methods under FGS design when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 8.tif

Supplementary Figure 4. Comparison of two selection rules when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 9.tif

Supplementary Figure 5. Comparison of SSR methods under FGS design when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 13.tif

Supplementary Figure 6. Comparison of two selection rules when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 14.tif

Supplementary Figure 7. Comparison of three adaptive designs when SSRCP is performed and



Supplementary Figure 8. Comparison of three adaptive designs when SSRCP, Ob is performed and

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 16.tif

Supplementary Figure 9. Comparison of two selection rules when

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 17.tif

Supplementary Figure 10. Comparison of the power when the true underlying SD were different from the preplanned SD=6

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 18.tif

Supplementary Figure 11. Disjunctive power for the three adaptive designs when the true underlying SD is 7

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 19.tif

Supplementary Figure 12. The impact of the timing for interim analysis on disjunctive power

C:\Users\liy\OneDrive - Washington University in St. Louis\own research\Y manuscript\all figures final2\Figure 20.tif

Supplementary Figure 13. Average new sample size per arm when using different timing for interim analysis