**SUPPLEMENTARY MATERIAL**

**Reference to possible one- and two-stage screened selection designs with 2 experimental arms under various response rates for** $k=0 $**and**$ k=0.05$

In each of the following tables, the type 1 error rate α for the A-Hern design (Tables S1a-b) or Simon’s two-stage optimal design (Tables S1c-d) is fixed at 0.1 in each experimental arm, whereas the type 2 error for the A’Hern design/Simon’s two-stage optimal design $β\_{C} $and the selection design $β\_{PW} $is specified at $β=β\_{C}=β\_{PW}=$0.15. Clearly, it is possible to choose different values for alpha and beta. The necessary sample size is derived following the iterative strategy introduced in the Methods section. For the reasons provided in the manuscript, we do not split randomly the ambiguity zone between the two experimental arms but report separately the probability to fall in the ambiguity zone.

For pilot studies, a $π\_{1}-π\_{0} $of 0.15 or 0.20 is commonly the degree of difference targeted (Simon, 1989). We simulated 1 million replications and assumed that $π\_{a} $is the minimum response rate of a promising drug (meaning that$ π\_{a}=π\_{1})$. Randomized Phase II trials often use a one-sided 10% type 1 error. Following the argumentation developed by Freidlin et al (Freidlin *et al.*, 2008), a multiplicity adjustment in multi-arm trials is only required when the corresponding questions are related. Examples of when the comparisons are related would be a trial that evaluates several schedules or doses of an agent versus the control or a trial testing single agents versus combinations of these agents. If, however, several experimental arms are screened for activity in parallel in a single phase II trial because of efficiency or logistical reasons rather than in several separate phase II trials, type 1 error do not need to be adjusted. Indeed, if several separate Phase II trials were conducted, no adjustment would have been made. In particular, if the experimental agents were developed by different companies or use different mechanisms of action, 10% type 1 error in each experimental arm may be appropriate. However, a type 1 error of 20% for the A-Hern design in each arm, as proposed by Yap et al. in their strategy for the sample size computation, may appear too high.

**Table S1a Possible sample size and corresponding overall selection probabilities in the one-stage screened selection design with** $k=0 $**for different response rates based on 1 million replications**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | $$π\_{0}$$ | $$(π\_{a}, π\_{b})$$ | **A’Hern design****r/n** | **α** | **1-β** | **Overall selection probability** |
| **Arm A** | **Arm B** | **No Arm** | **Ambiguity** |
| 1 | 0.05 | (0.20, 0.35) | 4/29 | 0.055 | 0.860 | 0.073 | 0.875 | 0 | 0.051 |
| 2 | 0.10 | (0.30, 0.45) | **6/29** | 0.098 | 0.907 | 0.092 | 0.855 | 0 | 0.053 |
| 3 | 0.15 | (0.30, 0.45) | 10/42 | 0.089 | 0.852 | 0.061 | 0.907 | 0 | 0.032 |
| 4 | 0.20 | (0.40, 0.55) | **10/32** | 0.075 | 0.857 | 0.093 | 0.857 | 0 | 0.050 |
| 5 | 0.25 | (0.40, 0.55) | 18/53 | 0.092 | 0.850 | 0.049 | 0.928 | 0 | 0.023 |
| 6 | 0.30 | (0.50, 0.65) | 14/33 | 0.088 | 0.852 | 0.085 | 0.869 | 0 | 0.046 |
| 7 | 0.35 | (0.50, 0.65) | 26/59 | 0.094 | 0.851 | 0.040 | 0.942 | 0 | 0.019 |
| 8 | 0.40 | (0.60, 0.75) | 19/36 | 0.083 | 0.854 | 0.067 | 0.894 | 0 | 0.039 |
| 9 | 0.45 | (0.60, 0.75) | 33/61 | 0.097 | 0.858 | 0.030 | 0.954 | 0 | 0.016 |
| 10 | 0.50 | (0.70, 0.85) | 22/35 | 0.088 | 0.865 | 0.047 | 0.917 | 0 | 0.036 |
| 11 | 0.55 | (0.70, 0.85) | 38/59 | 0.092 | 0.859 | 0.018 | 0.967 | 0 | 0.013 |
| 12 | 0.60 | (0.80, 0.95) | 22/30 | 0.094 | 0.871 | 0.019 | 0.950 | 0 | 0.031 |
| 13 | 0.65 | (0.80, 0.95) | 36/48 | 0.094 | 0.852 | 0.006 | 0.986 | 0 | 0.009 |

$π\_{0} $is the undesirable response rate, $π\_{a} $is the true response rate in Arm A ($π\_{a}=π\_{1})$, $π\_{b} $is the true response rate in Arm B

r is the minimum number of successes to warrant further investigation, n is the number of patients per experimental arm (with 1:1 allocation ratio)

α is the type 1 error of the A’Hern design, 1-β is the power of the A’Hern design

The A’Hern designs in **bold** (shaded cells) indicate that the sample size of the initial A’Hern design was increased in order to have an overall probability of correctly selecting Arm B of at least 85%.

**Table S1b Possible sample size and corresponding overall selection probabilities in the one-stage screened selection design with** $k=0.05 $**for different response rates based on 1 million replications**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | $$π\_{0}$$ | $$(π\_{a}, π\_{b})$$ | **A’Hern design****r/n** | **α** | **1-β** | **Overall selection probability** |
| **Arm A** | **Arm B** | **No Arm** | **Ambiguity** |
| 1 | 0.05 | (0.20, 0.35) | **5/36** | 0.032 | 0.873 | 0.033 | 0.852 | 0 | 0.115 |
| 2 | 0.10 | (0.30, 0.45) | **9/50** | 0.058 | 0.982 | 0.019 | 0.853 | 0 | 0.129 |
| 3 | 0.15 | (0.30, 0.45) | **12/50** | 0.063 | 0.861 | 0.019 | 0.853 | 0 | 0.128 |
| 4 | 0.20 | (0.40, 0.55) | **15/52** | 0.082 | 0.965 | 0.021 | 0.853 | 0 | 0.126 |
| 5 | 0.25 | (0.40, 0.55) | 18/53 | 0.092 | 0.850 | 0.021 | 0.858 | 0 | 0.122 |
| 6 | 0.30 | (0.50, 0.65) | **21/52** | 0.072 | 0.937 | 0.020 | 0.856 | 0 | 0.124 |
| 7 | 0.35 | (0.50, 0.65) | 26/59 | 0.094 | 0.851 | 0.017 | 0.884 | 0 | 0.100 |
| 8 | 0.40 | (0.60, 0.75) | **20/38** | 0.078 | 0.862 | 0.037 | 0.852 | 0 | 0.110 |
| 9 | 0.45 | (0.60, 0.75) | 33/61 | 0.097 | 0.858 | 0.007 | 0.866 | 0 | 0.128 |
| 10 | 0.50 | (0.70, 0.85) | 22/35 | 0.088 | 0.865 | 0.025 | 0.863 | 0 | 0.112 |
| 11 | 0.55 | (0.70, 0.85) | 38/59 | 0.092 | 0.859 | 0.006 | 0.923 | 0 | 0.071 |
| 12 | 0.60 | (0.80, 0.95) | 22/30 | 0.094 | 0.871 | 0.006 | 0.889 | 0 | 0.104 |
| 13 | 0.65 | (0.80, 0.95) | 36/48 | 0.094 | 0.852 | 0.001 | 0.935 | 0 | 0.064 |

**Table S1c Possible sample size (based on Simon’s two-stage optimal design) and corresponding overall selection probabilities in the two-stage screened selection design with** $k=0 $**for different response rates based on 1 million replications**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | $$π\_{0}$$ | $$(π\_{a}, π\_{b})$$ | **Stage 1****r1/n1** | **Stage 2****r/n** | **α** | **1-β** | **Overall selection probability** |
| **Arm A** | **Arm B** | **No Arm** | **Ambiguity** |
| 1 | 0.05 | (0.20, 0.35) | 1/10 | 4/34 | 0.070 | 0.851 | 0.068 | 0.891 | 0.002 | 0.038 |
| 2 | 0.10 | (0.30, 0.45) | **2/13** | **6/33** | 0.087 | 0.912 | 0.083 | 0.872 | 0.000 | 0.045 |
| 3 | 0.15 | (0.30, 0.45) | 4/20 | 11/50 | 0.095 | 0.851 | 0.050 | 0.926 | 0.000 | 0.023 |
| 4 | 0.20 | (0.40, 0.55) | **3/13** | **10/33** | 0.096 | 0.876 | 0.089 | 0.865 | 0.000 | 0.046 |
| 5 | 0.25 | (0.40, 0.55) | 7/24 | 20/64 | 0.099 | 0.856 | 0.039 | 0.943 | 0.000 | 0.017 |
| 6 | 0.30 | (0.50, 0.65) | 5/14 | 15/37 | 0.095 | 0.853 | 0.077 | 0.883 | 0.001 | 0.038 |
| 7 | 0.35 | (0.50, 0.65) | 12/30 | 29/68 | 0.096 | 0.850 | 0.031 | 0.954 | 0.000 | 0.014 |
| 8 | 0.40 | (0.60, 0.75) | 9/19 | 19/37 | 0.093 | 0.851 | 0.065 | 0.898 | 0.000 | 0.037 |
| 9 | 0.45 | (0.60, 0.75) | 16/32 | 35/66 | 0.099 | 0.852 | 0.026 | 0.961 | 0.000 | 0.013 |
| 10 | 0.50 | (0.70, 0.85) | 9/16 | 23/37 | 0.083 | 0.852 | 0.043 | 0.924 | 0.000 | 0.032 |
| 11 | 0.55 | (0.70, 0.85) | 15/25 | 41/65 | 0.097 | 0.850 | 0.015 | 0.976 | 0.000 | 0.010 |
| 12 | 0.60 | (0.80, 0.95) | 9/13 | 23/32 | 0.095 | 0.852 | 0.017 | 0.957 | 0.000 | 0.027 |
| 13 | 0.65 | (0.80, 0.95) | 18/25 | 41/56 | 0.097 | 0.853 | 0.004 | 0.991 | 0.000 | 0.005 |

$π\_{0} $is the undesirable response rate, $π\_{a} $is the true response rate in Arm A ($π\_{a}=π\_{1})$, $π\_{b} $is the true response rate in Arm B

r1 is the minimum number of successes for Stage 1 to not stop the trial for futility, n1 is the number of patients required for Stage 1 per experimental arm (with 1:1 allocation ratio)

r is the minimum number of successes to warrant further investigation, n is the number of patients per experimental arm (with 1:1 allocation ratio)

α is the type 1 error of the Simon’s two-stage design, 1-β is the power of the Simon’s two-stage design

The Simon’s two-stage designs in **bold** (shaded cells) indicate that the sample size of the initial Simon’s two-stage design was increased in order to have an overall probability of correctly selecting Arm B of at least 85%.

**Table S1d Possible sample size (based on Simon’s two-stage optimal design) and corresponding overall selection probabilities in the two-stage screened selection design with** $k=0.05 $**for different response rates based on 1 million replications**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Scenario** | $$π\_{0}$$ | $$(π\_{a}, π\_{b})$$ | **Stage 1****r1/n1** | **Stage 2****r/n** | **α** | **1-β** | **Overall selection probability** |
| **Arm A** | **Arm B** | **No Arm** | **Ambiguity** |
| 1 | 0.05 | (0.20, 0.35) | **1/10** | **4/37** | 0.086 | 0.867 | 0.041 | 0.855 | 0.000 | 0.103 |
| 2 | 0.10 | (0.30, 0.45) | **4/30** | **9/56** | 0.090 | 0.987 | 0.016 | 0.878 | 0.000 | 0.106 |
| 3 | 0.15 | (0.30, 0.45) | 4/20 | 11/50 | 0.095 | 0.851 | 0.022 | 0.853 | 0.000 | 0.124 |
| 4 | 0.20 | (0.40, 0.55) | **6/26** | **15/54** | 0.096 | 0.963 | 0.020 | 0.863 | 0.000 | 0.117 |
| 5 | 0.25 | (0.40, 0.55) | 7/24 | 20/64 | 0.099 | 0.856 | 0.012 | 0.855 | 0.000 | 0.132 |
| 6 | 0.30 | (0.50, 0.65) | **7/20** | **20/52** | 0.099 | 0.921 | 0.021 | 0.857 | 0.000 | 0.122 |
| 7 | 0.35 | (0.50, 0.65) | 12/30 | 29/68 | 0.096 | 0.850 | 0.009 | 0.881 | 0.000 | 0.110 |
| 8 | 0.40 | (0.60, 0.75) | **10/21** | **24/49** | 0.098 | 0.894 | 0.017 | 0.856 | 0.000 | 0.127 |
| 9 | 0.45 | (0.60, 0.75) | 16/32 | 35/66 | 0.099 | 0.852 | 0.006 | 0.886 | 0.000 | 0.108 |
| 10 | 0.50 | (0.70, 0.85) | 9/16 | 23/37 | 0.083 | 0.852 | 0.023 | 0.875 | 0.000 | 0.102 |
| 11 | 0.55 | (0.70, 0.85) | 15/25 | 41/65 | 0.097 | 0.850 | 0.003 | 0.909 | 0.000 | 0.088 |
| 12 | 0.60 | (0.80, 0.95) | 9/13 | 23/32 | 0.095 | 0.852 | 0.006 | 0.904 | 0.000 | 0.091 |
| 13 | 0.65 | (0.80, 0.95) | 18/25 | 41/56 | 0.097 | 0.853 | 0.001 | 0.961 | 0.000 | 0.039 |

REFERENCE (SUPPLEMENTARY DOCUMENT)

Freidlin, B. *et al.* (2008) ‘Multi-arm clinical trials of new agents: some design considerations.’, *Clinical cancer research : an official journal of the American Association for Cancer Research*, 14(14), pp. 4368–71. doi: 10.1158/1078-0432.CCR-08-0325.

Simon, R. (1989) ‘Optimal two-stage designs for phase II clinical trials’, *Controlled Clinical Trials*, 10(1), pp. 1–10. doi: 10.1016/0197-2456(89)90015-9.