

Supplemental Appendix to OPTIM-ARTS – An adaptive phase II open platform trial design with application to a metastatic melanoma study

In the main manuscript, we report design operating characteristics under two experimental scenarios for the true values of the ORR for the four treatment arms:

- Scenario A: true ORRs = 0.07, 0.10, 0.25, and 0.30.
- Scenario B: true ORRs = 0.10, 0.15, 0.25, and 0.35.

In this appendix, we report simulations under six additional scenarios:

- Scenario C: true ORRs = 0.10, 0.10, 0.10, and 0.10.
- Scenario D: true ORRs = 0.05, 0.07, 0.10, and 0.10.
- Scenario E: true ORRs = 0.25, 0.25, 0.25, and 0.25.
- Scenario F: true ORRs = 0.25, 0.28, 0.30, and 0.35.
- Scenario G: true ORRs = 0.07, 0.10, 0.25, and 0.25.
- Scenario H: true ORRs = 0.07, 0.15, 0.25, and 0.25.

In this appendix, for each of the scenarios C, D, E, F, G, and H, we have:

1. Two tables showing simulated operating characteristics for design $\mathfrak{D}1$ (our main proposed OPTIM-ARTS design with Bayesian monitoring in Part 1 and shrinkage estimation for treatment selection for Part 2, as described in Sec. 2.2 of the main manuscript):
 - First table displays simulated decision probabilities (%) for the four treatment arms at different IAs during Part 1. See also Table 2 in the main manuscript for similar results under scenarios A and B.
 - Second table displays additional operating characteristics during Part 1: overall decision probabilities for the treatment arms and treatment sample sizes, and the operating characteristics for Part 2: treatment sample sizes for Part 2 and total (Part 1 and Part 2 combined), and power of the final analysis. See also Table 3 in the main manuscript for similar results under scenarios A and B.
2. A figure comparing operating characteristics for the four different designs: $\mathfrak{D}1$, $\mathfrak{D}2$, $\mathfrak{D}3$, and $\mathfrak{D}4$ (see Table 4 in the main manuscript for a detailed description of these designs).
 - The figure shows for each treatment arm the power of the final test based on combined data from Parts 1 and 2, and the average sample size. See also Figures 4 and 5 for similar results under scenarios A and B, respectively.

Overall, the main findings for scenarios C, D, E, F, G, and H are consistent with those for scenarios A and B.

Specifically, for design $\mathfrak{D}1$ (cf. all Tables below):

- In Part 1, the decision probabilities and the sample size per arm are reflective of the underlying treatment efficacy: very efficacious (or very poor) treatments are identified with high probability and small sample size.
- The type I error rate is well controlled at the nominal 2.5% level (cf. unconditional characteristics for treatments with true $\text{ORR} \leq 0.10$).
- The average total (Parts 1 and 2 combined) sample size per arm ranges from 13 for $\text{ORR} = 0.05$ to about 47–52 for $\text{ORR} \geq 0.25$, across the considered scenarios.

As regards the comparison of designs $\mathfrak{D}1$, $\mathfrak{D}2$, $\mathfrak{D}3$, and $\mathfrak{D}4$ (cf. Figures S.1.3 – S.6.3):

- All designs maintain the type I error rate at the nominal 2.5% level, but with different average sample sizes per arm.
- No design stands out as “uniformly best” in terms of power/sample size tradeoff. Higher power naturally comes at the expense of a larger average sample size.
- Design $\mathfrak{D}1$ has, on average, lower sample size for futile arms (with true $\text{ORR} \leq 0.10$) than design $\mathfrak{D}4$, while both $\mathfrak{D}1$ and $\mathfrak{D}4$ maintain the type I error rate.
- Design $\mathfrak{D}1$ is less powerful and generally has lower average total sample size per arm compared to design $\mathfrak{D}4$ for treatment arms with $\text{ORR} \geq 0.25$. However, $\mathfrak{D}1$ looks quite attractive from the cost-efficiency perspective for treatments with $\text{ORR} \geq 0.30$.

Scenario C: True ORRs = 0.10, 0.10, 0.10, 0.10

Table S1.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario C</i>			
	Decision	TRT1 (0.10)	TRT2 (0.10)	TRT3 (0.10)	TRT4 (0.10)
IA1	Expand	6.9	7.4	7.13	N/A
	Terminate	73.42	73.47	73.58	N/A
	Continue	19.68	19.13	19.29	N/A
IA2	Expand	9.76	9.72	9.54	N/A
	Terminate	8.57	8.27	8.43	N/A
	Continue	1.35	1.14	1.32	N/A
IA3	Expand	1.24	1.08	1.16	7.12
	Terminate	0.09	0.05	0.25	73.73
	Continue	0.02	0.01	0.01	19.15
IA4	Expand	0.02	0.01	0.01	11.07
	Terminate	0	0	0	8.08
	Continue	0	0	0	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario C: True ORRs = 0.10, 0.10, 0.10, 0.10

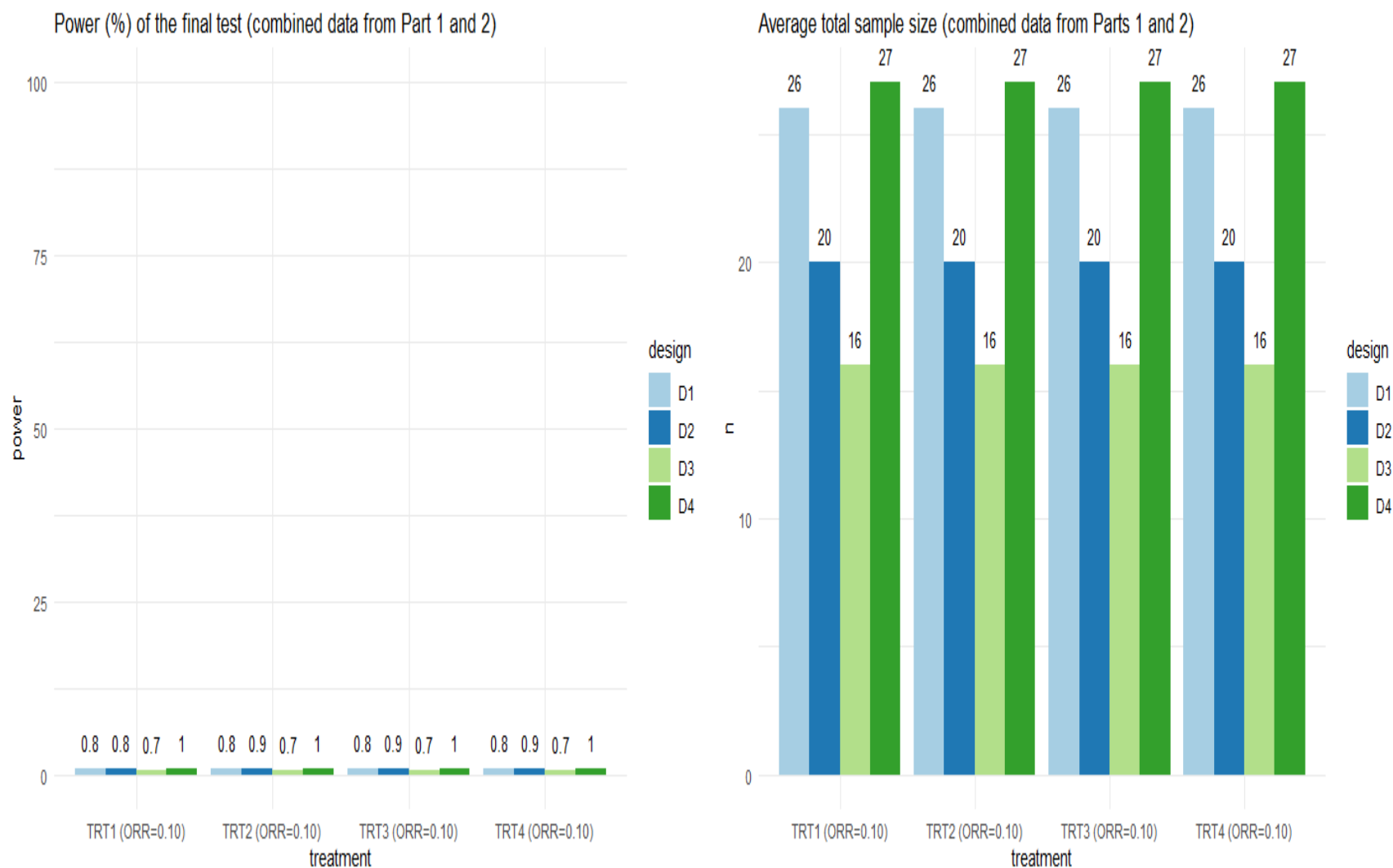
*Table S1.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	<i>Scenario C</i>			
	TRT1 (0.10)	TRT2 (0.10)	TRT3 (0.10)	TRT4 (0.10)
Part 1				
<i>Decision probabilities (%)</i>				
Pr(Expand)	17.92	18.21	17.84	18.19
Pr(Terminate)	82.08	81.79	82.16	81.81
<i>Sample size for Part 1 (n_1)</i>				
Mean (SD) [IQR]	14 (7) [10-10]	14 (7) [10-10]	14 (7) [10-10]	14 (7) [10-10]
Part 2 – Unconditional Characteristics				
<i>Sample size for Part 2 (n_2)</i>				
Mean (SD) [IQR]	12 (26) [0-0]	12 (26) [0-0]	12 (26) [0-0]	12 (26) [0-0]
<i>Total sample size (n_{tot})</i>				
Mean (SD) [IQR]	26 (30) [10-29]	26 (31) [10-29]	25 (30) [10-29]	26 (31) [10-29]
Power (P_{1-2}) (%)	0.8	0.8	0.8	0.8
Part 2 – Conditional Characteristics				
<i>Sample size for Part 2 (n_2)</i>				
Mean (SD) [IQR]	66 (12) [70-70]	66 (12) [70-70]	66 (12) [70-70]	66 (12) [70-70]
<i>Total sample size (n_{tot})</i>				
Mean (SD) [IQR]	88 (16) [80-99]	88 (16) [80-99]	87 (17) [80-99]	88 (17) [80-99]
Power (P_{1-2}) (%)	4.2	4.3	4.4	4.4

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario C: True ORRs = 0.10, 0.10, 0.10, 0.10

Figure S.1.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4



Scenario D: True ORRs = 0.05, 0.07, 0.10, 0.10

Table S2.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario D</i>			
	Decision	TRT1 (0.05)	TRT2 (0.07)	TRT3 (0.10)	TRT4 (0.10)
IA1	Expand	1.11	2.87	7.13	N/A
	Terminate	91.4	84.32	73.58	N/A
	Continue	7.49	12.81	19.29	N/A
IA2	Expand	1.51	4.78	10.67	N/A
	Terminate	5.79	7.64	7.96	N/A
	Continue	0.19	0.39	0.66	N/A
IA3	Expand	0.17	0.37	0.65	7.07
	Terminate	0.02	0.01	0.01	73.28
	Continue	0	0.01	0	19.65
IA4	Expand	N/A	0.01	N/A	11.45
	Terminate	N/A	0	N/A	8.2
	Continue	N/A	0	N/A	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario D: True ORRs = 0.05, 0.07, 0.10, 0.10

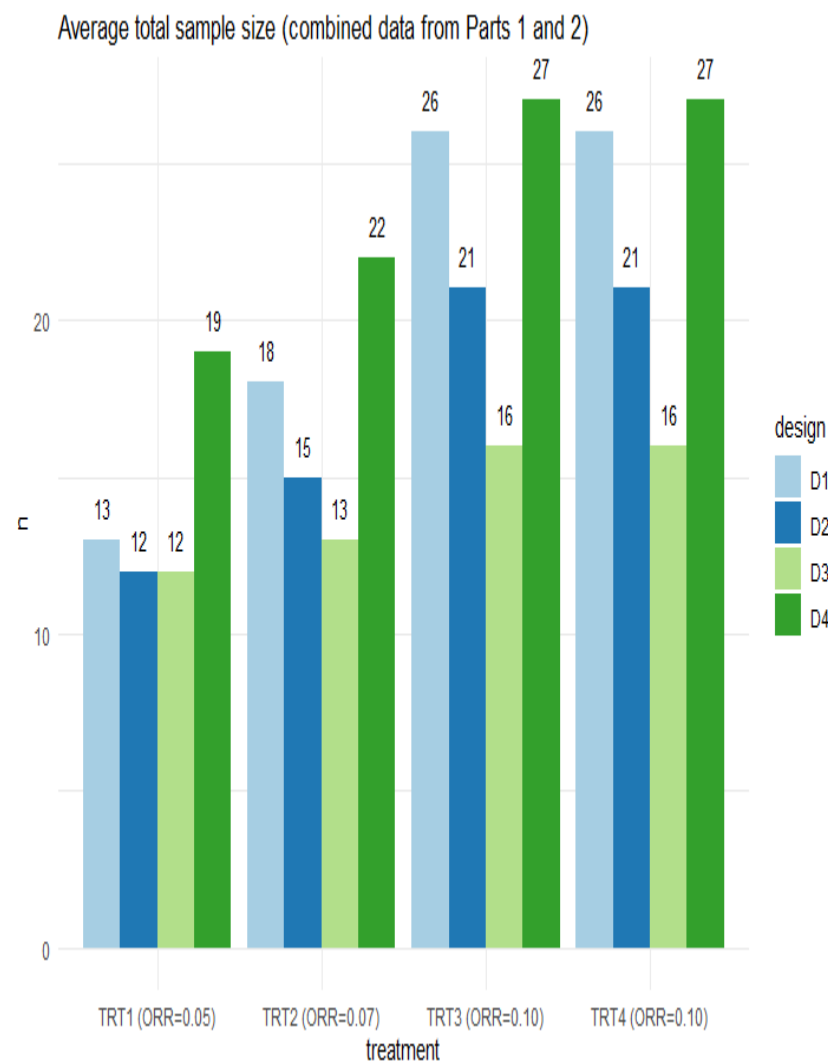
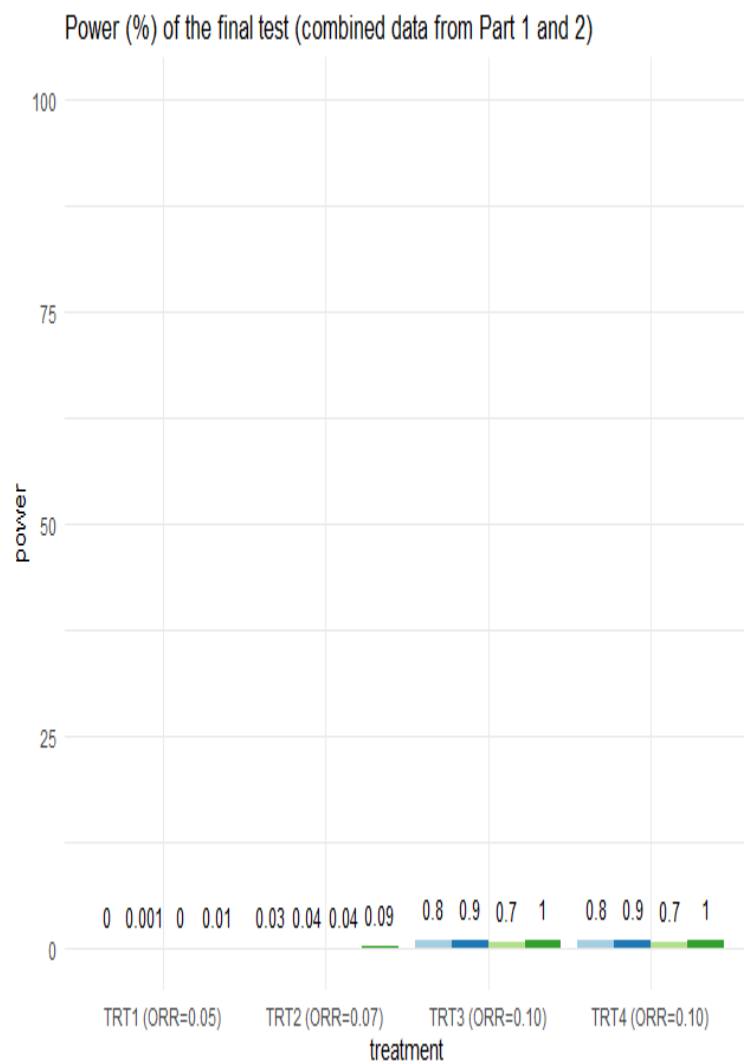
*Table S2.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	Scenario D			
	TRT1 (0.05)	TRT2 (0.07)	TRT3 (0.10)	TRT4 (0.10)
Part 1				
Decision probabilities (%)				
Pr(Expand)	2.79	8.03	18.45	18.52
Pr(Terminate)	97.21	91.97	81.55	81.48
Sample size for Part 1 (n_1)				
Mean (SD) [IQR]	11 (5) [10-10]	12 (6) [10-10]	14 (7) [10-10]	14 (7) [10-10]
Part 2 – Unconditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	2 (11) [0-0]	5 (19) [0-0]	12 (26) [0-0]	12 (26) [0-0]
Total sample size (n_{tot})				
Mean (SD) [IQR]	13 (14) [10-10]	18 (22) [10-10]	26 (31) [10-29]	26 (31) [10-29]
Power (P_{1-2}) (%)	<0.1	<0.1	0.8	0.8
Part 2 – Conditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	69 (6) [70-70]	68 (8) [70-70]	66 (11) [70-70]	66 (11) [70-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	90 (12) [80-99]	91 (14) [80-99]	88 (16) [80-99]	88 (16) [80-99]
Power (P_{1-2}) (%)	<0.1	0.4	4.3	4.4

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario D: True ORRs = 0.05, 0.07, 0.10, 0.10

Figure S.2.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4



Scenario E: True ORRs = 0.25, 0.25, 0.25, 0.25

Table S3.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario E</i>			
	Decision	TRT1 (0.25)	TRT2 (0.25)	TRT3 (0.25)	TRT4 (0.25)
IA1	Expand	47.63	46.7	47.91	N/A
	Terminate	24.08	24.65	24.02	N/A
	Continue	28.29	28.65	28.07	N/A
IA2	Expand	24.23	25.06	24.21	N/A
	Terminate	1.08	0.92	1.32	N/A
	Continue	2.98	2.67	2.54	N/A
IA3	Expand	2.93	2.59	2.53	47.23
	Terminate	0.03	0.05	0	24.54
	Continue	0.02	0.03	0.01	28.23
IA4	Expand	0.02	0.03	0.01	27.36
	Terminate	0	0	0	0.87
	Continue	0	0	0	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario E: True ORRs = 0.25, 0.25, 0.25, 0.25

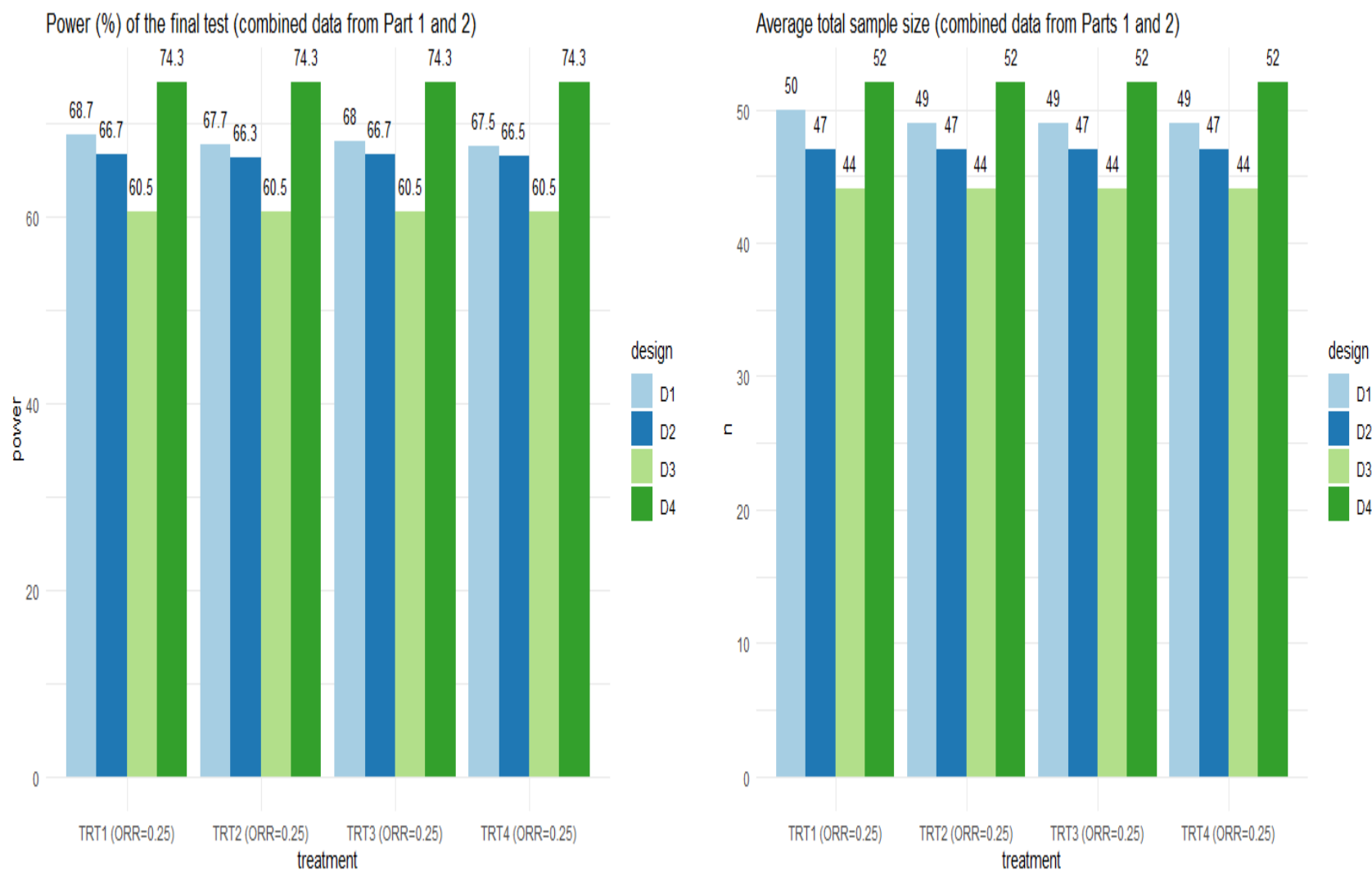
*Table S3.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	Scenario E			
	TRT1 (0.25)	TRT2 (0.25)	TRT3 (0.25)	TRT4 (0.25)
Part 1				
Decision probabilities (%)				
Pr(Expand)	74.81	74.38	74.66	74.59
Pr(Terminate)	25.19	25.62	25.34	25.41
Sample size for Part 1 (n_1)				
Mean (SD) [IQR]	15 (8) [10-29]	15 (8) [10-29]	15 (8) [10-28]	15 (9) [10-29]
Part 2 – Unconditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	34 (28) [0-70]	34 (28) [0-70]	34 (28) [0-70]	33 (28) [0-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	50 (31) [29-80]	49 (31) [27-80]	49 (31) [29-80]	49 (31) [29-80]
Power (P_{1-2}) (%)	68.0	67.7	68.0	67.5
Part 2 – Conditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	46 (23) [20-70]	46 (23) [20-70]	46 (23) [20-70]	45 (23) [20-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	63 (25) [30-80]	62 (25) [30-80]	63 (25) [30-80]	62 (25) [30-80]
Power (P_{1-2}) (%)	91.0	91.0	91.0	90.5

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario E: True ORRs = 0.25, 0.25, 0.25, 0.25

Figure S.3.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4



Scenario F: True ORRs = 0.25, 0.28, 0.30, 0.35

Table S4.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario F</i>			
	Decision	TRT1 (0.25)	TRT2 (0.28)	TRT3 (0.30)	TRT4 (0.35)
IA1	Expand	47.63	55.21	62.54	N/A
	Terminate	24.08	18.24	15.1	N/A
	Continue	28.29	26.55	22.36	N/A
IA2	Expand	25.04	24.14	20.64	N/A
	Terminate	1.01	0.5	0.26	N/A
	Continue	2.24	1.91	1.46	N/A
IA3	Expand	2.19	1.85	1.44	73.26
	Terminate	0.02	0.03	0.01	8.79
	Continue	0.02	0.03	0.01	17.95
IA4	Expand	0.02	0.03	0.01	17.89
	Terminate	0	0	0	0.06
	Continue	0	0	0	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario F: True ORRs = 0.25, 0.28, 0.30, 0.35

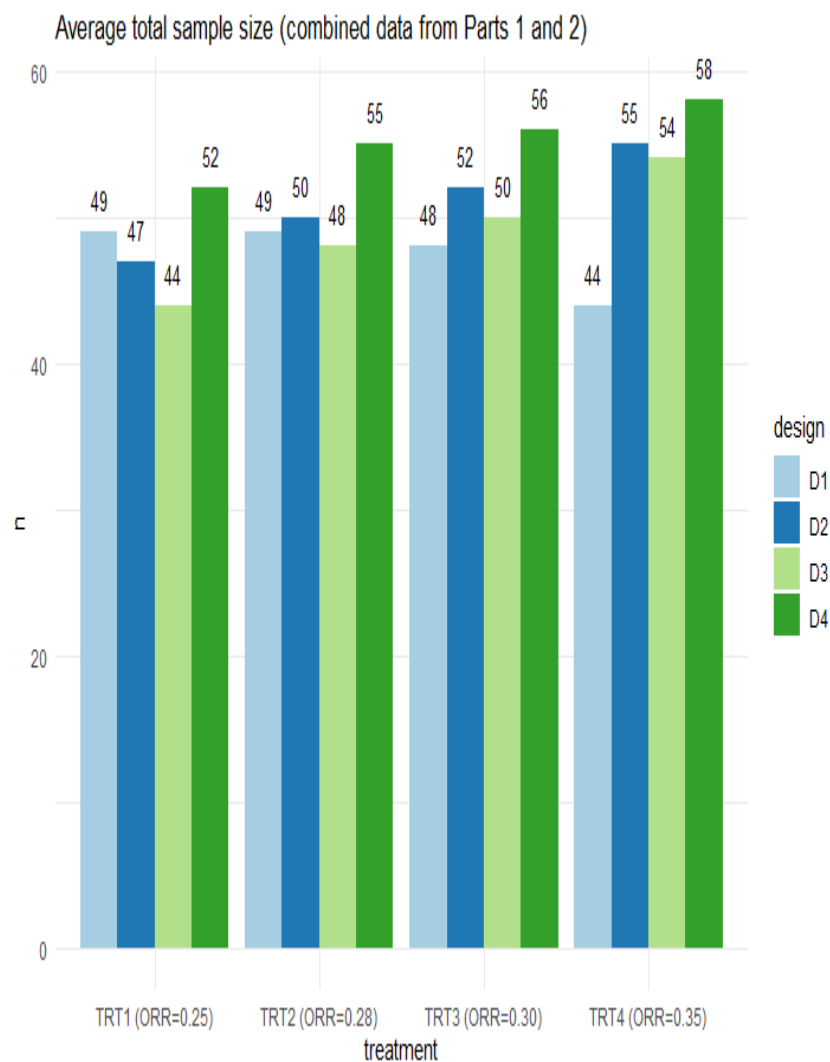
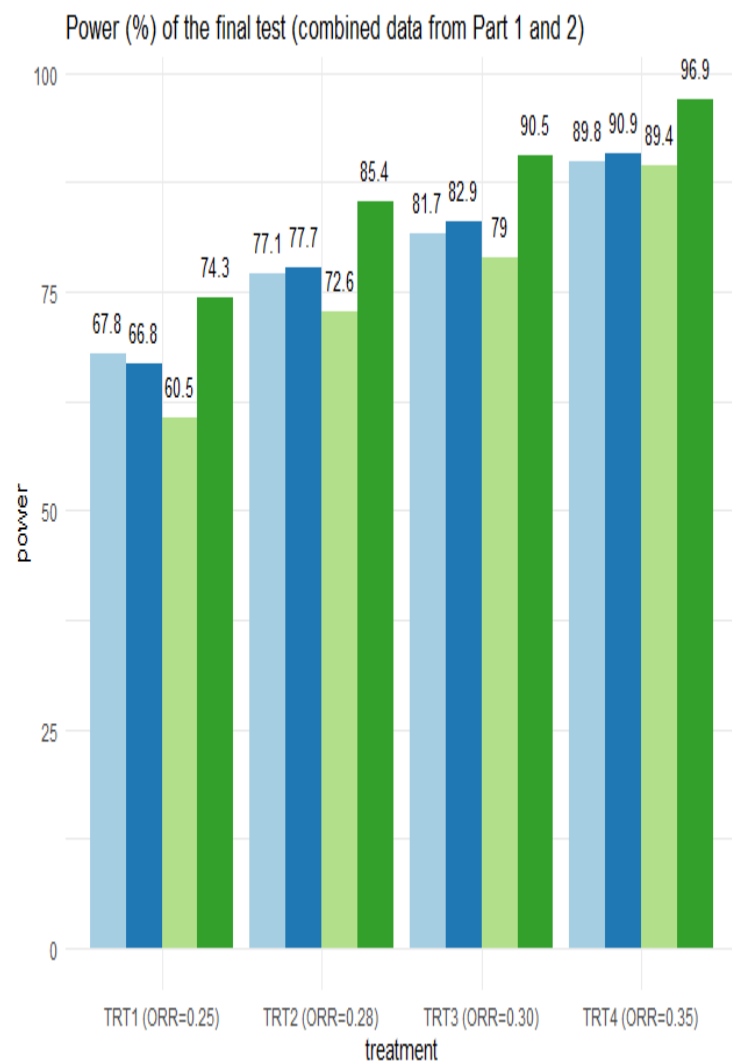
*Table S4.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	Scenario F			
	TRT1 (0.25)	TRT2 (0.28)	TRT3 (0.30)	TRT4 (0.35)
Part 1				
Decision probabilities (%)				
Pr(Expand)	74.88	81.23	84.63	91.15
Pr(Terminate)	25.12	18.77	15.37	8.85
Sample size for Part 1 (n_1)				
Mean (SD) [IQR]	15 (8) [10-29]	15 (8) [10-26]	14 (8) [10-10]	13 (7) [10-10]
Part 2 – Unconditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	34 (28) [0-70]	34 (26) [20-70]	34 (26) [20-70]	31 (22) [20-55]
Total sample size (n_{tot})				
Mean (SD) [IQR]	49 (31) [29-80]	49 (29) [30-80]	48 (27) [30-80]	44 (23) [30-65]
Power (P_{1-2}) (%)	67.8	77.2	81.7	89.8
Part 2 – Conditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	45 (23) [20-70]	41 (23) [20-70]	40 (23) [20-70]	34 (20) [20-55]
Total sample size (n_{tot})				
Mean (SD) [IQR]	62 (25) [30-80]	57 (24) [30-80]	55 (24) [30-80]	47 (21) [30-65]
Power (P_{1-2}) (%)	90.6	95.0	96.6	98.6

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario F: True ORRs = 0.25, 0.28, 0.30, 0.35

Figure S.4.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4



Scenario G: True ORRs = 0.07, 0.10, 0.25, 0.25

Table S5.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario G</i>			
	Decision	TRT1 (0.07)	TRT2 (0.10)	TRT3 (0.25)	TRT4 (0.25)
IA1	Expand	2.82	7.4	47.91	N/A
	Terminate	85.04	73.47	24.02	N/A
	Continue	12.14	19.13	28.07	N/A
IA2	Expand	4.1	9.54	25.76	N/A
	Terminate	7.3	8.21	1.0	N/A
	Continue	0.74	1.38	1.31	N/A
IA3	Expand	0.68	1.28	1.3	47.25
	Terminate	0.06	0.1	0.01	24.13
	Continue	0	0	0	28.62
IA4	Expand	N/A	N/A	N/A	27.76
	Terminate	N/A	N/A	N/A	0.86
	Continue	N/A	N/A	N/A	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario G: True ORRs = 0.07, 0.10, 0.25, 0.25

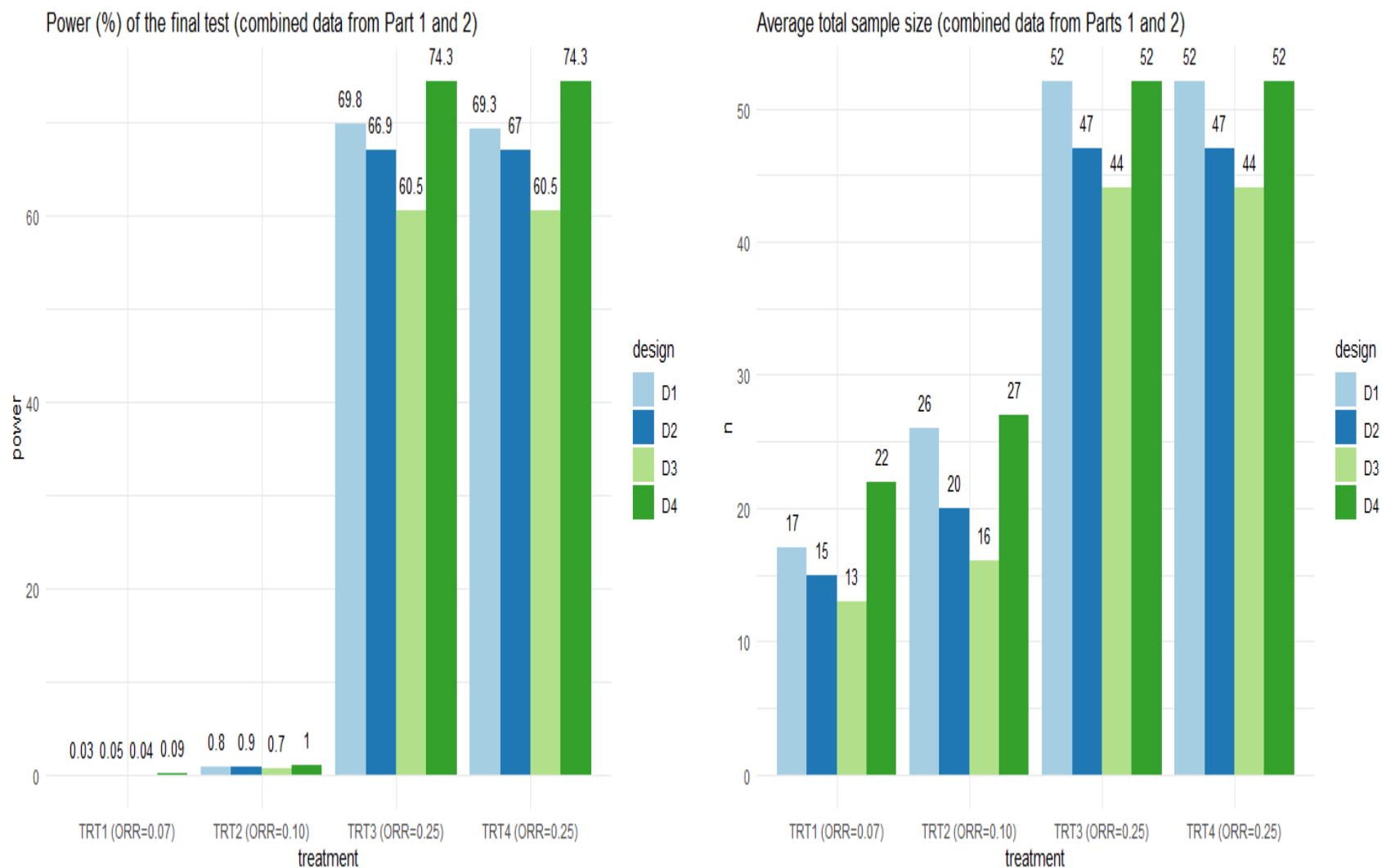
*Table S5.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	Scenario G			
	TRT1 (0.07)	TRT2 (0.10)	TRT3 (0.25)	TRT4 (0.25)
Part 1				
Decision probabilities (%)				
Pr(Expand)	7.6	18.22	74.97	75.01
Pr(Terminate)	92.4	81.78	25.03	24.99
Sample size for Part 1 (n_1)				
Mean (SD) [IQR]	12 (6) [10-10]	14 (7) [10-10]	15 (8) [10-29]	15 (9) [10-29]
Part 2 – Unconditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	5 (18) [0-0]	12 (26) [0-0]	37 (29) [0-70]	36 (29) [20-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	17 (22) [10-10]	26 (30) [10-29]	52 (32) [29-80]	51 (32) [30-80]
Power (P_{1-2}) (%)	<0.1	0.8	69.8	69.3
Part 2 – Conditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	68 (10) [70-70]	66 (13) [70-70]	50 (22) [20-70]	48 (27) [20-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	90 (15) [80-99]	87 (17) [80-99]	66 (24) [45-80]	65 (24) [45-80]
Power (P_{1-2}) (%)	0.4	4.4	93.1	92.3

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario G: True ORRs = 0.07, 0.10, 0.25, 0.25

Figure S.5.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4



Scenario H: True ORRs = 0.07, 0.15, 0.25, 0.25

Table S6.1: Probabilities (%) of decisions for the four treatment arms at different IAs during Part 1 for design D1.

		<i>Scenario H</i>			
	Decision	TRT1 (0.07)	TRT2 (0.15)	TRT3 (0.25)	TRT4 (0.25)
IA1	Expand	2.82	18.65	47.91	N/A
	Terminate	85.04	55.22	24.02	N/A
	Continue	12.12	26.13	28.07	N/A
IA2	Expand	3.91	18.64	25.29	N/A
	Terminate	7.27	5.5	0.9	N/A
	Continue	0.96	1.99	1.88	N/A
IA3	Expand	0.89	1.94	1.83	47.25
	Terminate	0.07	0.04	0.01	24.13
	Continue	0	0.01	0.04	28.62
IA4	Expand	N/A	0.01	0.04	27.76
	Terminate	N/A	0	0	0.86
	Continue	N/A	0	0	0

Note: Probabilities are not cumulative. TRT4 is introduced into the study after IA2.

Scenario H: True ORRs = 0.07, 0.15, 0.25, 0.25

*Table S6.2: Operating characteristics for the four treatment arms for design D1. **Part 1:** Decision probabilities (%); Sample size for Part 1 (n_1) – mean (SD) [IQR]. **Part 2** (Unconditional—derived across 10,000 simulation runs, and Conditional—derived across simulation runs for which a decision to expand a treatment into Part 2 is made): Sample size for Part 2 (n_2) – mean (SD) [IQR]; Total sample size (n_{tot}) – mean (SD) [IQR]; Power (%) of the final analysis using pooled data from Parts 1 and 2.*

	Scenario H			
	TRT1 (0.07)	TRT2 (0.15)	TRT3 (0.25)	TRT4 (0.25)
Part 1				
Decision probabilities (%)				
Pr(Expand)	7.62	39.24	75.07	74.7
Pr(Terminate)	92.38	60.76	24.93	25.3
Sample size for Part 1 (n_1)				
Mean (SD) [IQR]	12 (6) [10-10]	15 (8) [10-28]	15 (8) [10-29]	15 (9) [10-29]
Part 2 – Unconditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	5 (18) [0-0]	24 (32) [0-70]	37 (29) [20-70]	35 (29) [0-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	17 (22) [10-10]	39 (36) [10-80]	52 (32) [30-80]	51 (32) [29-80]
Power (P_{1-2}) (%)	<0.1	13.2	69.6	68.7
Part 2 – Conditional Characteristics				
Sample size for Part 2 (n_2)				
Mean (SD) [IQR]	67 (11) [70-70]	60 (18) [70-70]	49 (23) [20-70]	47 (23) [20-70]
Total sample size (n_{tot})				
Mean (SD) [IQR]	89 (15) [80-99]	80 (22) [80-99]	66 (29) [45-80]	64 (25) [45-80]
Power (P_{1-2}) (%)	0.5	33.6	92.7	91.9

Note: SD = standard deviation; IQR = Interquartile Range [Q25 – Q75]

Scenario H: True ORRs = 0.07, 0.15, 0.25, 0.25

Figure S.6.3: Power of the final analysis (left plot) and average total sample size (right plot) for the four treatment arms for designs D1, D2, D3, and D4

