

1
2
3
4
5
6
7
8
9
10 **Supplemental data**
11

12 **“Sample size determination in group-sequential clinical trials with two co-**
13 **primary endpoints”**
14

15 by Koko Asakura, Toshimitsu Hamasaki, Tomoyuki Sugimoto, Kenichi Hayashi,
16 Scott R Evans and Takashi Sozu
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table S1. Impact of sample size recalculation on the power and Type I error rate. The planned MSS per intervention group is calculated to detect the joint effect for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where one interim and one final analysis are to be performed. The critical values are determined by the OF boundary for both endpoints with the LD alpha-spending method. The upper limit of recalculation sample size is $n'_2 = \lambda n_2$ with $\lambda = 1.5$. The number of replications for simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power. (DF-1)

Information time	Power/ Type I error rate	Mean difference (δ_i, δ_s)	Correlation (MSS)	(i) Only decreasing sample size (ASN)	(ii) Only increasing sample Size (ASN)	(iii) Increasing or decreasing sample size (ASN)
0.25	Power	(0.2, 0.2)	0.0 (516)	71.0 (448)	88.9 (640)	79.7 (571)
			0.3 (504)	69.6 (425)	88.5 (616)	77.9 (537)
			0.5 (492)	68.9 (406)	88.1 (596)	76.7 (509)
			0.8 (460)	67.5 (365)	87.4 (549)	74.6 (455)
			0.99 (412)	66.1 (313)	86.9 (484)	72.6 (384)
	Type I error rate	(0.0, 0.0)	0.0 (516)	0.06 (515)	0.06 (578)	0.06 (577)
			0.3 (504)	0.24 (502)	0.24 (573)	0.24 (571)
			0.5 (492)	0.46 (489)	0.46 (565)	0.46 (562)
		(0.0, 0.2)	0.8 (460)	1.10 (453)	1.10 (537)	1.10 (530)
			0.99 (412)	2.17 (401)	2.17 (490)	2.18 (478)
			0.0 (516)	2.17 (509)	2.37 (620)	2.29 (613)
			0.3 (504)	2.39 (493)	2.47 (603)	2.41 (593)
	0.50	(0.0, 0.0)	0.5 (492)	2.46 (479)	2.50 (588)	2.46 (576)
			0.8 (460)	2.50 (446)	2.50 (549)	2.50 (535)
			0.99 (412)	2.50 (399)	2.50 (491)	2.49 (479)
			0.0 (518)	74.0 (448)	91.3 (618)	85.2 (564)
			0.3 (506)	74.0 (429)	90.9 (591)	84.7 (536)
		(0.0, 0.2)	0.5 (492)	73.8 (411)	90.4 (566)	84.2 (511)
			0.8 (460)	73.6 (376)	89.6 (517)	83.2 (463)
			0.99 (410)	73.3 (327)	88.8 (449)	82.0 (399)
			0.0 (518)	0.06 (518)	0.06 (582)	0.06 (582)
			0.3 (506)	0.23 (506)	0.23 (580)	0.23 (580)
	0.75	(0.0, 0.0)	0.5 (492)	0.46 (491)	0.46 (571)	0.45 (571)
			0.8 (460)	1.08 (459)	1.08 (546)	1.09 (545)
			0.99 (410)	2.17 (407)	2.16 (499)	2.17 (496)
			0.0 (518)	2.17 (516)	2.37 (637)	2.33 (635)
			0.3 (506)	2.41 (503)	2.46 (620)	2.46 (617)
		(0.0, 0.2)	0.5 (492)	2.48 (488)	2.50 (602)	2.48 (599)
			0.8 (460)	2.49 (456)	2.49 (563)	2.49 (559)
			0.99 (410)	2.50 (407)	2.51 (502)	2.51 (498)
			0.0 (528)	79.7 (462)	91.9 (557)	91.3 (552)
	Type I error rate	(0.2, 0.2)	0.3 (516)	79.8 (449)	91.6 (538)	91.0 (534)
			0.5 (500)	79.5 (434)	91.1 (518)	90.5 (515)
			0.8 (468)	79.5 (403)	90.3 (480)	89.9 (476)
			0.99 (416)	79.6 (356)	89.4 (422)	89.0 (418)
			0.0 (528)	0.05 (528)	0.05 (594)	0.05 (594)
		(0.0, 0.0)	0.3 (516)	0.22 (516)	0.21 (592)	0.22 (592)
			0.5 (500)	0.45 (500)	0.44 (582)	0.43 (582)
			0.8 (468)	1.07 (467)	1.07 (558)	1.08 (558)
			0.99 (416)	2.17 (415)	2.17 (509)	2.18 (509)
		(0.0, 0.2)	0.0 (528)	2.17 (527)	2.33 (653)	2.32 (653)
			0.3 (516)	2.41 (515)	2.46 (637)	2.46 (637)
			0.5 (500)	2.47 (499)	2.51 (617)	2.49 (617)
			0.8 (468)	2.50 (467)	2.52 (577)	2.49 (577)
			0.99 (416)	2.49 (415)	2.49 (513)	2.50 (513)

Table S2. Impact of sample size recalculation on the power and Type I error rate. The planned MSS per intervention group is calculated to detect the joint effect for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where one interim and one final analysis are to be performed and the critical values are determined by the OF boundary for both endpoints with the LD alpha-spending method. The upper limit of recalculation sample size is $n'_2 = \lambda n_2$ with $\lambda = 1.5$. The number of replications for simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power. (DF-2)

Information time	Power/ Type I error rate	Mean difference (δ_i, δ_s)	Correlation (MSS)	(i) Only decreasing sample size (ASN)	(ii) Only increasing sample Size (ASN)	(iii) Increasing or decreasing sample size (ASN)
0.25	Power	(0.2, 0.2)	0.0 (516)	71.0 (449)	88.9 (640)	79.8 (573)
			0.3 (504)	69.7 (426)	88.4 (616)	77.9 (537)
			0.5 (492)	68.8 (406)	88.1 (596)	76.8 (512)
			0.8 (460)	67.4 (366)	87.4 (549)	74.6 (455)
			0.99 (412)	66.1 (312)	86.9 (484)	72.6 (385)
	Type I error rate	(0.0, 0.0)	0.0 (516)	0.06 (515)	0.06 (578)	0.06 (577)
			0.3 (504)	0.24 (502)	0.24 (573)	0.24 (571)
			0.5 (492)	0.46 (489)	0.46 (565)	0.46 (561)
		(0.0, 0.2)	0.8 (460)	1.09 (453)	1.10 (537)	1.10 (530)
			0.99 (412)	2.17 (401)	2.17 (489)	2.17 (478)
			0.0 (516)	2.17 (509)	2.36 (620)	2.29 (613)
			0.3 (504)	2.39 (493)	2.47 (604)	2.42 (593)
	0.50	(0.0, 0.0)	0.5 (492)	2.46 (479)	2.50 (588)	2.46 (575)
			0.8 (460)	2.50 (446)	2.50 (549)	2.50 (535)
			0.99 (412)	2.50 (399)	2.50 (492)	2.50 (479)
			0.0 (518)	74.4 (448)	91.3 (618)	85.7 (563)
			0.3 (506)	74.3 (428)	91.0 (591)	85.1 (536)
		(0.0, 0.2)	0.5 (492)	74.0 (411)	90.5 (566)	84.4 (512)
			0.8 (460)	73.6 (376)	89.7 (516)	83.4 (462)
			0.99 (410)	73.4 (327)	88.8 (449)	82.1 (398)
			0.0 (518)	0.06 (518)	0.06 (582)	0.06 (582)
			0.3 (506)	0.24 (506)	0.24 (580)	0.24 (580)
	0.75	(0.0, 0.0)	0.5 (492)	0.46 (491)	0.46 (571)	0.46 (570)
			0.8 (460)	1.09 (459)	1.09 (546)	1.09 (544)
			0.99 (410)	2.17 (407)	2.18 (498)	2.18 (496)
			0.0 (518)	2.21 (516)	2.40 (637)	2.37 (635)
			0.3 (506)	2.43 (503)	2.47 (620)	2.46 (617)
		(0.0, 0.2)	0.5 (492)	2.49 (488)	2.50 (603)	2.48 (599)
			0.8 (460)	2.50 (456)	2.50 (563)	2.49 (559)
		0.99 (410)	2.50 (407)	2.50 (502)	2.50 (499)	

1
2
3
4
5
6
7
8 **Table S3.** Impact of sample size recalculation on the power and Type I error rate. The planned MSS per
9 intervention group is calculated to detect the joint effect for two endpoints with the overall power of 80% at the
10 one-sided significance level of 2.5%, where three interims and one final analysis are to be performed at equal time
11 intervals and the critical values are determined by the OF boundary for both endpoints with the LD alpha-spending
12 method. The upper limit of recalculation sample size is $n_2' = \lambda n_2$ with $\lambda = 1.5$. The number of replications for
13 simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power. (DF-1)

Interim analysis	Power/ Type I error rate	Mean difference (δ_1, δ_2)	Correlation (MSS)	(i) Only decreasing sample size (ASN)	(ii) Only increasing sample size (ASN)	(iii) Increasing or decreasing sample size (ASN)
1st	Power	(0.2, 0.2)	0.0 (528)	71.3 (459)	89.1 (655)	80.4 (586)
			0.3 (516)	70.2 (435)	88.6 (631)	78.6 (551)
			0.5 (504)	69.5 (416)	88.3 (611)	77.4 (523)
			0.8 (468)	67.9 (373)	87.4 (558)	75.4 (464)
			0.99 (420)	67.0 (319)	87.0 (494)	73.7 (394)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.05 (527)	0.05 (591)	0.05 (591)
			0.3 (516)	0.21 (514)	0.22 (587)	0.22 (585)
			0.5 (504)	0.44 (501)	0.44 (579)	0.44 (576)
			0.8 (468)	1.07 (461)	1.07 (547)	1.07 (540)
			0.99 (420)	2.17 (409)	2.17 (499)	2.17 (488)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.07 (521)	2.30 (635)	2.21 (628)
			0.3 (516)	2.35 (506)	2.45 (619)	2.38 (608)
			0.5 (504)	2.44 (491)	2.48 (603)	2.45 (591)
			0.8 (468)	2.50 (454)	2.50 (559)	2.50 (545)
			0.99 (420)	2.51 (407)	2.51 (502)	2.51 (489)
2nd	Power	(0.2, 0.2)	0.0 (528)	74.4 (454)	91.4 (628)	85.5 (571)
			0.3 (516)	74.3 (435)	91.1 (600)	85.1 (543)
			0.5 (504)	74.4 (418)	90.9 (578)	84.8 (520)
			0.8 (468)	73.8 (380)	89.8 (524)	83.6 (468)
			0.99 (420)	74.0 (332)	89.2 (458)	82.6 (405)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.05 (528)	0.05 (594)	0.05 (593)
			0.3 (516)	0.21 (516)	0.22 (592)	0.22 (591)
			0.5 (504)	0.43 (503)	0.43 (585)	0.44 (585)
			0.8 (468)	1.07 (467)	1.08 (555)	1.08 (554)
			0.99 (420)	2.17 (417)	2.15 (511)	2.15 (508)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.12 (526)	2.32 (649)	2.28 (647)
			0.3 (516)	2.39 (513)	2.44 (633)	2.43 (630)
			0.5 (504)	2.47 (500)	2.48 (618)	2.48 (614)
			0.8 (468)	2.51 (464)	2.50 (573)	2.50 (570)
			0.99 (420)	2.49 (416)	2.51 (514)	2.51 (511)
3rd	Power	(0.2, 0.2)	0.0 (528)	79.4 (455)	91.8 (549)	91.1 (545)
			0.3 (516)	79.6 (437)	91.6 (525)	91.0 (522)
			0.5 (504)	79.9 (422)	91.3 (508)	90.7 (503)
			0.8 (468)	79.5 (387)	90.4 (464)	89.8 (459)
			0.99 (420)	79.8 (342)	89.7 (407)	89.2 (403)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.05 (528)	0.05 (594)	0.05 (594)
			0.3 (516)	0.22 (516)	0.22 (592)	0.22 (592)
			0.5 (504)	0.43 (504)	0.43 (587)	0.43 (587)
			0.8 (468)	1.09 (467)	1.08 (558)	1.08 (558)
			0.99 (420)	2.17 (419)	2.15 (515)	2.16 (514)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.14 (527)	2.30 (653)	2.30 (653)
			0.3 (516)	2.41 (514)	2.45 (637)	2.47 (636)
			0.5 (504)	2.48 (502)	2.49 (622)	2.49 (621)
			0.8 (468)	2.50 (466)	2.50 (577)	2.52 (577)
			0.99 (420)	2.49 (419)	2.48 (518)	2.50 (518)

Table S4. Impact of sample size recalculation on the power and Type I error rate. The planned MSS per intervention group is calculated to detect the joint effect for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where three interims and one final analysis are to be performed at equal time intervals and the critical values are determined by the OF boundary for both endpoints with the LD alpha-spending method. The upper limit of recalculation sample size is $n_2' = \lambda n_2$ with $\lambda = 1.5$. The number of replications for simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power. (DF-2)

Interim analysis	Power/ Type I error rate	Mean difference (δ_i, δ_s)	Correlation (MSS)	(i) Only decreasing sample size (ASN)	(ii) Only increasing sample Size (ASN)	(iii) Increasing or decreasing sample size (ASN)
1st	Power	(0.2, 0.2)	0.0 (528)	71.8 (459)	89.2 (655)	80.6 (585)
			0.3 (516)	70.2 (433)	88.4 (626)	78.7 (547)
			0.5 (504)	69.4 (413)	88.2 (606)	77.4 (519)
			0.8 (468)	68.0 (372)	87.5 (559)	75.4 (463)
			0.99 (420)	67.0 (319)	87.0 (494)	73.6 (394)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.06 (527)	0.06 (591)	0.06 (591)
			0.3 (516)	0.23 (510)	0.23 (582)	0.23 (580)
			0.5 (504)	0.45 (497)	0.45 (574)	0.45 (571)
			0.8 (468)	1.09 (461)	1.09 (546)	1.08 (540)
			0.99 (420)	2.17 (409)	2.17 (499)	2.17 (489)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.17 (521)	2.36 (635)	2.29 (628)
			0.3 (516)	2.39 (502)	2.47 (613)	2.41 (603)
			0.5 (504)	2.45 (488)	2.49 (598)	2.46 (585)
			0.8 (468)	2.50 (454)	2.50 (559)	2.49 (545)
			0.99 (420)	2.50 (407)	2.50 (502)	2.50 (489)
2nd	Power	(0.2, 0.2)	0.0 (528)	74.8 (453)	91.5 (629)	86.0 (571)
			0.3 (516)	74.3 (431)	90.9 (597)	85.1 (540)
			0.5 (504)	74.4 (414)	90.7 (574)	84.8 (516)
			0.8 (468)	74.0 (380)	89.9 (524)	83.7 (467)
			0.99 (420)	73.9 (332)	89.1 (458)	82.7 (405)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.06 (528)	0.06 (594)	0.06 (593)
			0.3 (516)	0.23 (512)	0.23 (587)	0.23 (587)
			0.5 (504)	0.45 (499)	0.45 (581)	0.45 (580)
			0.8 (468)	1.08 (466)	1.08 (555)	1.08 (554)
			0.99 (420)	2.17 (417)	2.16 (511)	2.17 (508)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.21 (526)	2.38 (649)	2.37 (647)
			0.3 (516)	2.42 (509)	2.48 (628)	2.46 (625)
			0.5 (504)	2.47 (496)	2.50 (613)	2.48 (609)
			0.8 (468)	2.51 (464)	2.51 (573)	2.49 (570)
			0.99 (420)	2.49 (416)	2.50 (514)	2.50 (511)
3rd	Power	(0.2, 0.2)	0.0 (528)	79.7 (453)	91.9 (549)	91.1 (544)
			0.3 (512)	79.2 (434)	91.3 (525)	90.7 (518)
			0.5 (500)	79.5 (419)	91.1 (506)	90.5 (500)
			0.8 (468)	79.6 (387)	90.4 (463)	89.7 (460)
			0.99 (420)	79.9 (342)	89.7 (407)	89.1 (403)
	Type I error rate	(0.0, 0.0)	0.0 (528)	0.06 (528)	0.06 (594)	0.06 (594)
			0.3 (512)	0.23 (512)	0.23 (588)	0.23 (588)
			0.5 (500)	0.45 (500)	0.46 (582)	0.45 (582)
			0.8 (468)	1.09 (467)	1.08 (558)	1.08 (558)
			0.99 (420)	2.16 (419)	2.18 (514)	2.16 (514)
	(0.0, 0.2)	(0.0, 0.2)	0.0 (528)	2.23 (527)	2.39 (653)	2.41 (653)
			0.3 (512)	2.44 (510)	2.48 (632)	2.47 (632)
			0.5 (500)	2.48 (498)	2.49 (617)	2.49 (616)
			0.8 (468)	2.51 (466)	2.50 (577)	2.51 (577)
			0.99 (420)	2.51 (419)	2.50 (518)	2.49 (518)