

## APPENDIX

Table 1A. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.2$ ,  $r=0.5$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.2912 0.3801 0.4745	0.3270 0.4111 0.5168	0.0363 0.0106	-1.4598	-1.4863
60	0.2912 0.3801 0.4745	0.2918 0.4072 0.4818	0.0117 0.0053	-1.4598	-1.4559
80	0.2912 0.3801 0.4745	0.2853 0.3850 0.4689	-0.0022 0.0040	-1.4598	-1.4585
100	0.2912 0.3801 0.4745	0.3004 0.3838 0.4741	0.0042 0.0028	-1.4598	-1.4550
120	0.2912 0.3801 0.4745	0.2924 0.3844 0.4815	0.0041 0.0024	-1.4598	-1.4556
140	0.2912 0.3801 0.4745	0.3041 0.3761 0.4800	0.0047 0.0021	-1.4598	-1.4413

Table 1B. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.2, r=0.8$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.1694 0.3028 0.4444	0.2241 0.3348 0.4787	0.0403 0.0092	-1.3138	-1.3157
60	0.1694 0.3028 0.4444	0.2090 0.3045 0.4600	0.0189 0.0053	-1.3138	-1.3077
80	0.1694 0.3028 0.4444	0.1821 0.2795 0.4455	-0.0032 0.0029	-1.3138	-1.3060
100	0.1694 0.3028 0.4444	0.1648 0.3063 0.4487	0.0010 0.0019	-1.3138	-1.3184
120	0.1694 0.3028 0.4444	0.1797 0.2926 0.4675	0.0077 0.0021	-1.3138	-1.3445
140	0.1694 0.3028 0.4444	0.1747 0.2927 0.4357	-0.0045 0.0010	-1.3138	-1.3270

Table 1C. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.5, r=0.2$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.4412 0.5079 0.5787	0.4640 0.5340 0.6261	0.0321 0.0158	-1.6804	-1.6620
60	0.4412 0.5079 0.5787	0.4718 0.5346 0.5996	0.0261 0.0104	-1.6804	-1.7003
80	0.4412 0.5079 0.5787	0.4575 0.5483 0.6060	0.0280 0.0070	-1.6804	-1.6841
100	0.4412 0.5079 0.5787	0.4500 0.5382 0.6026	0.0210 0.0070	-1.6804	-1.6982
120	0.4412 0.5079 0.5787	0.4452 0.5199 0.5861	0.0078 0.0044	-1.6804	-1.6572
140	0.4412 0.5079 0.5787	0.4572 0.5189 0.5969	0.0151 0.0040	-1.6804	-1.7035

Table 1D. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.5, r=0.5$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.2912 0.3801 0.4745	0.3325 0.4437 0.5144	0.0483 0.0124	-1.4598	-1.4786
60	0.2912 0.3801 0.4745	0.3009 0.3918 0.5149	0.0206 0.0091	-1.4598	-1.4404
80	0.2912 0.3801 0.4745	0.3029 0.3929 0.4820	0.0107 0.0050	-1.4598	-1.4645
100	0.2912 0.3801 0.4745	0.2988 0.3880 0.4882	0.0097 0.0032	-1.4598	-1.4597
120	0.2912 0.3801 0.4745	0.2987 0.3979 0.5113	0.0207 0.0037	-1.4598	-1.4598
140	0.2912 0.3801 0.4745	0.3010 0.3892 0.4783	0.0075 0.0024	-1.4598	-1.4829

Table 1E. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.5$ ,  $r=0.8$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.1694 0.3028 0.4444	0.2070 0.3348 0.4725	0.0325 0.0079	-1.3138	-1.3364
60	0.1694 0.3028 0.4444	0.1943 0.3153 0.4456	0.0129 0.0028	-1.3138	-1.3325
80	0.1694 0.3028 0.4444	0.1880 0.3086 0.4672	0.0157 0.0034	-1.3138	-1.3149
100	0.1694 0.3028 0.4444	0.1757 0.3135 0.4380	0.0035 0.0021	-1.3138	-1.3335
120	0.1694 0.3028 0.4444	0.1750 0.3100 0.4497	0.0060 0.0024	-1.3138	-1.3128
140	0.1694 0.3028 0.4444	0.1733 0.3008 0.4554	0.0043 0.0017	-1.3138	-1.3289

Table 1F. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.8, r=0.2$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.4412 0.5079 0.5787	0.4967 0.5668 0.6259	0.0539 0.0176	-1.6804	-1.6621
60	0.4412 0.5079 0.5787	0.4940 0.5593 0.6318	0.0525 0.0151	-1.6804	-1.7136
80	0.4412 0.5079 0.5787	0.4698 0.5455 0.5818	0.0231 0.0102	-1.6804	-1.6560
100	0.4412 0.5079 0.5787	0.4671 0.5145 0.6241	0.0260 0.0068	-1.6804	-1.6678
120	0.4412 0.5079 0.5787	0.4523 0.5257 0.5863	0.0122 0.0043	-1.6804	-1.6640
140	0.4412 0.5079 0.5787	0.4568 0.5169 0.5968	0.0142 0.0038	-1.6804	-1.6942

Table 1G. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.8, r=0.5$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.2912 0.3801 0.4745	0.3316 0.4149 0.5215	0.0407 0.0139	-1.4598	-1.4733
60	0.2912 0.3801 0.4745	0.3104 0.3908 0.5034	0.0196 0.0063	-1.4598	-1.4457
80	0.2912 0.3801 0.4745	0.2853 0.4045 0.5030	0.0157 0.0047	-1.4598	-1.4639
100	0.2912 0.3801 0.4745	0.3104 0.3962 0.5014	0.0207 0.0065	-1.4598	-1.4470
120	0.2912 0.3801 0.4745	0.2897 0.3969 0.4872	0.0093 0.0026	-1.4598	-1.4409
140	0.2912 0.3801 0.4745	0.3110 0.3878 0.4668	0.0066 0.0024	-1.4598	-1.4637

Table 1H. True and estimated weights (averaged over 100 simulation samples) to combine three markers for the maximum ratio of mean and SD (efficacy comparison assumed at the end of trial, i.e., time=3, N=sample size for both training and validation data sets,  $\rho =0.8, r=0.8$ )

N	True weights	Estimated weights	Bias (MSE) on weights	True Mean/SD	Estimated Mean/SD
40	0.1694 0.3028 0.4444	0.2223 0.3148 0.4733	0.0312 0.0080	-1.3138	-1.3591
60	0.1694 0.3028 0.4444	0.1991 0.2990 0.4685	0.0166 0.0052	-1.3138	-1.3214
80	0.1694 0.3028 0.4444	0.1882 0.3020 0.4647	0.0127 0.0033	-1.3138	-1.3155
100	0.1694 0.3028 0.4444	0.1609 0.3152 0.4741	0.0112 0.0033	-1.3138	-1.3121
120	0.1694 0.3028 0.4444	0.1719 0.3015 0.4632	0.0066 0.0020	-1.3138	-1.3207
140	0.1694 0.3028 0.4444	0.1912 0.3027 0.4604	0.0125 0.0022	-1.3138	-1.3212



Table 2A. Total sample sizes (averaged over 100 simulated pairs of data sets) required for 80% power to detect specified effect sizes for  $\rho = r = 0.5$  in a 1:1 two-arm clinical trial (weights estimated from training data set with 60 subjects; mean and SD on the combined outcome estimated from either the same data set or another independent sample of the same size. Efficacy comparison assumed at time=3.  $N_i$ =total sample size using  $i$ -th efficacy outcome alone;  $N$  =total sample size using the true weight for the combined outcome;  $NP$  =total sample size using the combined outcome with weights estimated from training data sets and mean/SD estimated from validation data sets;  $NS$  = total sample size using the combined outcome with weights and SD estimated from the same data sets. ES=effect size as a percentage of improvement from the estimated mean for the placebo at  $t=3$ )

ES	$N_1$	$N_2$	$N_3$	$N$	$NP$	$NS$	$NP - NS$
20%	852	720	575	386	399	368	32
25%	545	461	368	247	256	236	20
30%	379	321	256	172	178	164	14
35%	279	236	188	127	131	121	11
40%	214	181	144	97	100	93	8
45%	169	143	114	77	80	73	7
50%	137	116	93	63	65	60	5

Table 2B. Total sample sizes (averaged over 100 simulated pairs of data sets) required for 80% power to detect specified effect sizes for  $\rho = r = 0.5$  in a 1:1 two-arm clinical trial (weights estimated from training data set with 80 subjects; mean and SD on the combined outcome estimated from either the same data set or another independent sample of the same size. Efficacy comparison assumed at time=3.  $N_i$  =total sample size using  $i$ -th efficacy outcome alone;  $N$  =total sample size using the true weight for the combined outcome;  $NP$  =total sample size using the combined outcome with weights estimated from training data sets and mean/SD estimated from validation data sets;  $NS$  = total sample size using the combined outcome with weights and SD estimated from the same data sets. ES=effect size as a percentage of improvement from the estimated mean for the placebo at  $t=3$ )

ES	$N_1$	$N_2$	$N_3$	$N$	$NP$	$NS$	$NP - NS$
20%	806	695	554	370	380	374	6
25%	516	445	355	237	244	240	4
30%	359	309	247	165	170	167	3
35%	264	228	182	121	125	123	2
40%	202	175	139	93	96	94	2
45%	160	138	110	74	76	75	2
50%	130	112	90	60	62	61	1

Table 2C. Total sample sizes (averaged over 100 simulated pairs of data sets) required for 80% power to detect specified effect sizes for  $\rho = r = 0.5$  in a 1:1 two-arm clinical trial (weights estimated from training data set with 100 subjects; mean and SD on the combined outcome estimated from either the same data set or another independent sample of the same size. Efficacy comparison assumed at time=3.  $N_i$  =total sample size using  $i$ -th efficacy outcome alone;  $N$  =total sample size using the true weight for the combined outcome;  $NP$  =total sample size using the combined outcome with weights estimated from training data sets and mean/SD estimated from validation data sets;  $NS$  = total sample size using the combined outcome with weights and SD estimated from the same data sets. ES=effect size as a percentage of improvement from the estimated mean for the placebo at  $t=3$ )

ES	$N_1$	$N_2$	$N_3$	$N$	$NP$	$NS$	$NP - NS$
20%	811	665	574	377	381	366	15
25%	520	426	368	242	244	234	10
30%	361	296	256	168	170	163	7
35%	266	218	188	124	125	120	5
40%	204	167	144	95	96	92	4
45%	161	132	114	75	76	73	3
50%	131	107	93	61	62	59	3

Table 2D. Total sample sizes (averaged over 100 simulated pairs of data sets) required for 80% power to detect specified effect sizes for  $\rho = r = 0.5$  in a 1:1 two-arm clinical trial (weights estimated from training data set with 120 subjects; mean and SD on the combined outcome estimated from either the same data set or another independent sample of the same size. Efficacy comparison assumed at time=3.  $N_i$ =total sample size using  $i$ -th efficacy outcome alone;  $N$  =total sample size using the true weight for the combined outcome;  $NP$  =total sample size using the combined outcome with weights estimated from training data sets and mean/SD estimated from validation data sets;  $NS$  = total sample size using the combined outcome with weights and SD estimated from the same data sets. ES=effect size as a percentage of improvement from the estimated mean for the placebo at  $t=3$ )

ES	$N_1$	$N_2$	$N_3$	$N$	$NP$	$NS$	$NP - NS$
20%	807	658	572	372	377	353	24
25%	517	421	367	239	241	227	15
30%	359	293	255	166	168	158	11
35%	264	215	188	122	124	116	8
40%	202	165	144	94	95	89	6
45%	160	131	114	74	75	71	5
50%	130	106	92	60	61	57	4

Table 2E. Total sample sizes (averaged over 100 simulated pairs of data sets) required for 80% power to detect specified effect sizes for  $\rho = r = 0.5$  in a 1:1 two-arm clinical trial (weights estimated from training data set with 140 subjects; mean and SD on the combined outcome estimated from either the same data set or another independent sample of the same size. Efficacy comparison assumed at time=3.  $N_i$ =total sample size using  $i$ -th efficacy outcome alone;  $N$  =total sample size using the true weight for the combined outcome;  $NP$  =total sample size using the combined outcome with weights estimated from training data sets and mean/SD estimated from validation data sets;  $NS$  = total sample size using the combined outcome with weights and SD estimated from the same data sets. ES=effect size as a percentage of improvement from the estimated mean for the placebo at  $t=3$ )

ES	$N_1$	$N_2$	$N_3$	$N$	$NP$	$NS$	$NP - NS$
20%	809	647	545	362	365	368	-2
25%	518	414	349	232	234	236	-1
30%	360	288	243	161	163	164	-1
35%	265	212	178	119	120	121	0
40%	203	162	137	91	92	93	0
45%	161	129	108	72	73	73	0
50%	130	104	88	59	59	60	0