Supplementary Materials for "Adaptive Hybrid Control Design for Comparative Clinical Trials with History Control Data"

Beibei Guo[1], Glen Laird[2], Yang Song[2], and Ying Yuan[3]

[1]*Department of Experimental Statistics, Louisiana State University*

Baton Rouge, LA 70803, U.S.A.

[2]*Vertex Pharmaceuticals, Boston*

MA 02210, U.S.A.

[3]*Department of Biostatistics, The University of Texas MD Anderson Cancer Center*

Houston, TX 77030, U.S.A.

email: yyuan@mdanderson.org

Table S1: Continuous outcome: Type I error, power, relative bias of $\hat{\tau}_c$, percentages of early stopping due to superiority and futility for the two-stage AHCC design with interim sample size 60. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial; % n save is the percentage of sample size saved for the trial by using AHCC design leveraging historical data compared with a standard two-stage RCT.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.050	99.9	93.8	193.7	3.1	-0.022	0.0	0.2
10	0.053	99.9	87.0	186.9	6.5	-0.038	0.0	0.2
20	0.054	99.9	78.3	178.2	10.8	-0.023	0.0	0.1
30	0.052	99.9	69.2	169.1	15.4	-0.008	0.0	0.1
40	0.052	99.8	59.7	159.5	20.1	-0.001	0.0	0.2
50	0.049	99.7	51.5	151.2	24.3	-0.018	0.0	0.2
60	0.049	99.2	43.4	142.6	28.6	-0.008	0.0	0.2
70	0.048	98.4	38.2	136.5	31.7	-0.003	0.0	0.2
80	0.049	97.4	35.0	132.4	33.7	-0.002	0.0	0.2
90	0.049	96.9	34.0	130.9	34.5	0.000	0.0	0.2
100	0.049	96.6	33.7	130.3	34.8	-0.004	0.0	0.2
110	0.047	96.5	33.5	130.0	34.9	-0.001	0.0	0.1
125	0.049	96.4	33.5	129.9	35.0	0.000	0.0	0.1
140	0.045	96.3	33.5	129.9	35.0	-0.004	0.0	0.2

				$\tau = 0.905$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.822	99.1	93.1	192.2	3.5	-0.021	1.2	0.1
10	0.860	98.4	85.9	184.3	7.5	-0.042	2.3	0.0
20	0.860	98.2	77.0	175.2	12.0	-0.025	2.5	0.0
30	0.863	98.0	68.1	166.1	16.6	-0.014	2.8	0.0
40	0.869	97.6	58.9	156.5	21.4	0.003	3.3	0.0
50	0.901	97.1	50.6	147.7	25.8	-0.027	4.0	0.0
60	0.915	96.3	43.0	139.3	30.1	-0.021	4.4	0.0
70	0.928	95.4	37.8	133.2	33.1	-0.006	4.4	0.0
80	0.943	93.9	34.8	128.7	35.4	-0.008	5.5	0.0
90	0.958	93.4	33.9	127.3	36.1	-0.005	5.4	0.0
100	0.964	92.7	33.4	126.1	36.7	-0.015	6.1	0.0
110	0.975	92.3	33.3	125.6	36.9	-0.013	6.5	0.0
125	0.981	91.8	33.3	125.0	37.2	-0.019	7.1	0.0
140	0.987	91.5	33.2	124.8	37.3	-0.022	7.5	0.0

Table S2: Binary outcome: Type I error, power, relative bias, percentages of early stopping due to superiority and futility for the two-stage AHCC design with interim sample size 60. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial; % n save is the percentage of sample size saved for the trial by using AHCC design leveraging historical data compared with a standard two-stage RCT.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.056	99.9	93.9	193.8	3.1	-0.001	0.0	0.1
10	0.054	99.9	87.0	186.8	6.6	0.000	0.0	0.2
20	0.052	99.9	78.3	178.2	10.9	0.001	0.0	0.1
30	0.056	99.9	69.1	169.0	15.5	0.000	0.0	0.1
40	0.055	99.9	60.0	159.9	20.1	0.000	0.0	0.2
50	0.051	99.7	51.5	151.2	24.4	-0.001	0.0	0.1
60	0.053	99.1	43.5	142.6	28.7	-0.002	0.0	0.2
70	0.054	98.3	38.0	136.3	31.8	0.001	0.0	0.2
80	0.051	97.5	35.1	132.7	33.7	0.000	0.0	0.2
90	0.047	96.9	34.1	131.1	34.5	0.000	0.0	0.1
100	0.052	96.6	33.7	130.3	34.9	0.000	0.0	0.2
110	0.051	96.5	33.6	130.0	35.0	-0.001	0.0	0.1
125	0.050	96.4	33.5	129.9	35.1	0.000	0.0	0.2
140	0.049	96.4	33.5	129.9	35.0	0.001	0.0	0.1
				$\tau = 1.417$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.792	98.9	92.9	191.8	3.8	0.001	1.6	0.0
10	0.824	98.7	86.1	184.8	7.3	-0.001	1.8	0.0
20	0.821	98.5	77.2	175.7	11.8	-0.001	2.1	0.0
30	0.827	98.6	68.5	167.1	16.2	-0.001	2.0	0.0
40	0.829	98.7	59.3	158.0	20.7	0.001	1.8	0.0
50	0.823	98.6	51.1	149.8	24.8	-0.002	1.7	0.1
60	0.833	98.0	43.3	141.3	29.1	-0.002	1.8	0.1
70	0.860	97.1	38.0	135.0	32.2	-0.001	2.1	0.0
80	0.873	96.2	35.1	131.2	34.1	0.000	2.1	0.0
90	0.883	95.6	34.0	129.6	35.0	-0.001	2.1	0.0
100	0.896	95.2	33.6	128.8	35.4	-0.002	2.3	0.0
110	0.894	95.0	33.5	128.5	35.5	-0.001	2.3	0.0
125	0.915	94.9	33.4	128.3	35.6	-0.002	2.4	0.0
140	0.918	94.9	33.4	128.4	35.6	-0.002	2.3	0.0

Table S3: Continuous outcome: Type I error, power at $\tau = 0.905$, relative bias, percentages of early stopping due to superiority and futility of $\hat{\tau}_c$ for the 2-stage AHCC design with interim sample size 100. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial; % n save is the percentage of sample size saved for the trial by using AHCC design leveraging historical data compared with a standard 2-stage RCT.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.055	96.1	89.1	185.2	3.8	-0.018	0.3	7.6
10	0.052	95.9	81.3	177.1	7.9	-0.026	0.3	8.0
20	0.050	95.9	73.4	169.3	12.0	-0.012	0.2	8.0
30	0.052	96.3	66.3	162.6	15.5	-0.008	0.2	7.1
40	0.053	95.6	58.9	154.4	19.7	0.002	0.2	7.7
50	0.051	95.0	54.8	149.8	22.2	-0.005	0.2	7.6
60	0.049	93.9	52.9	146.8	23.7	-0.002	0.2	8.4
70	0.050	93.8	52.4	146.2	24.0	0.001	0.2	7.9
80	0.047	93.8	52.3	146.1	24.1	0.005	0.1	7.8
90	0.051	93.7	52.3	146.1	24.1	0.005	0.2	7.7
100	0.048	93.7	52.3	146.0	24.1	0.003	0.2	7.9
110	0.051	93.7	52.3	146.0	24.1	0.004	0.2	7.8
125	0.047	93.6	52.3	146.0	24.1	0.004	0.2	7.9
140	0.046	93.6	52.3	145.9	24.2	0.005	0.1	8.1

				$\tau = 0.905$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.818	90.0	84.0	174.1	5.2	-0.017	19.2	0.7
10	0.855	87.6	75.6	163.1	11.1	-0.034	24.4	0.5
20	0.855	87.0	68.9	155.9	15.1	-0.009	25.7	0.4
30	0.863	85.8	62.6	148.4	19.1	0.001	28.0	0.3
40	0.873	84.1	56.7	140.9	23.2	0.011	30.8	0.3
50	0.889	82.6	53.5	136.1	25.8	0.004	33.0	0.3
60	0.896	81.1	52.1	133.1	27.5	0.012	35.1	0.2
70	0.916	80.1	51.6	131.8	28.2	0.016	36.6	0.2
80	0.934	79.1	51.5	130.6	28.8	0.016	38.7	0.1
90	0.942	77.7	51.5	129.2	29.6	0.010	41.5	0.1
100	0.953	77.2	51.4	128.6	29.9	0.013	42.7	0.1
110	0.961	76.6	51.4	128.0	30.3	0.013	44.0	0.1
125	0.971	75.1	51.3	126.4	31.1	0.006	47.1	0.1
140	0.978	74.0	51.3	125.3	31.7	0.008	49.2	0.1

Table S4: Binary outcome: Type I error, power at $\tau = 1.417$, relative bias, percentages of early stopping due to superiority and futility of $\hat{\tau}_c$ for the 2-stage AHCC design with interim sample size 100. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial; % n save is the percentage of sample size saved for the trial by using AHCC design leveraging historical data compared with a standard 2-stage RCT.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.052	95.9	89.0	184.9	7.5	-0.001	0.2	8.0
10	0.056	95.9	81.2	177.2	11.3	-0.001	0.2	8.0
20	0.052	95.9	73.5	169.3	15.3	-0.001	0.1	8.1
30	0.051	96.3	66.2	162.5	18.7	0.000	0.1	7.1
40	0.051	95.5	58.9	154.3	22.8	0.000	0.1	8.0
50	0.057	95.0	54.7	149.7	25.1	-0.001	0.1	7.7
60	0.051	94.2	52.9	147.1	26.4	0.000	0.1	7.7
70	0.048	93.9	52.4	146.3	26.8	0.000	0.1	7.7
80	0.050	93.8	52.3	146.1	26.9	0.000	0.1	7.7
90	0.048	93.9	52.3	146.2	26.8	0.000	0.1	7.5
100	0.055	93.8	52.3	146.1	26.9	0.000	0.1	7.8
110	0.050	93.8	52.3	146.1	26.9	0.000	0.1	7.8
125	0.051	93.8	52.3	146.1	26.9	0.000	0.1	7.8
140	0.050	93.6	52.3	145.9	27.0	0.000	0.1	8.1
				$\tau = 1.417$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.795	90.2	84.2	174.4	6.0	0.000	18.8	0.8
10	0.807	89.7	77.0	166.7	10.1	0.000	20.0	0.7
20	0.821	88.5	69.7	158.2	14.7	-0.002	22.3	0.6
30	0.821	88.0	63.3	151.3	18.4	-0.002	23.3	0.6
40	0.822	87.0	57.3	144.4	22.1	-0.001	24.7	0.5
50	0.818	87.1	53.9	141.0	23.9	0.000	23.5	0.4
60	0.834	85.8	52.4	138.2	25.5	-0.001	25.1	0.4
70	0.862	84.3	51.9	136.2	26.5	-0.001	27.7	0.2
80	0.876	84.2	51.8	136.0	26.7	-0.001	27.8	0.3
90	0.876	83.8	51.8	135.6	26.9	-0.001	28.6	0.3
100	0.896	83.2	51.8	134.9	27.2	-0.002	30.0	0.2
110	0.904	83.2	51.7	135.0	27.2	-0.001	29.7	0.3
125	0.904	82.9	51.7	134.6	27.4	-0.001	30.6	0.2
140	0.917	82.4	51.7	134.1	27.7	-0.002	31.7	0.2

Table S5: Continuous outcome: Type I error, power at $\tau = 0.905$, relative bias, percentages of early stopping due to superiority and futility of $\hat{\tau}_c$ the 2-stage AHCC design starting with single-arm design and interim sample size 120. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.055	114.0	69.9	183.9	-3.0	0.008	1.2	24.4
10	0.055	114.0	63.4	177.4	0.6	-0.008	1.1	24.1
20	0.054	114.0	55.9	169.9	4.8	0.003	0.8	26.0
30	0.048	114.0	50.3	164.3	8.0	0.009	0.5	25.8
40	0.046	114.0	42.9	156.9	12.1	0.005	0.4	26.9
50	0.047	114.0	36.8	150.8	15.5	0.009	0.5	28.0
60	0.044	114.0	29.3	143.2	19.8	0.003	0.4	27.8
70	0.048	114.0	22.6	136.6	23.5	0.006	0.4	26.9
80	0.044	114.0	15.6	129.6	27.4	0.001	0.3	27.0
90	0.047	114.0	11.3	125.3	29.8	0.007	0.3	27.3
100	0.046	114.0	8.2	122.2	31.5	0.010	0.3	27.4
110	0.041	114.0	6.7	120.7	32.4	0.010	0.3	26.9
125	0.045	114.0	6.1	120.1	32.7	0.011	0.4	26.6
140	0.045	114.0	6.0	120.0	32.8	0.011	0.3	27.1

				$\tau = 0.905$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.840	114.0	72.7	186.7	-7.2	-0.069	18.6	3.8
10	0.894	114.0	50.8	164.9	5.3	-0.100	40.9	0.8
20	0.911	114.0	36.7	150.7	13.4	-0.043	55.1	0.3
30	0.930	114.0	26.8	140.7	19.2	-0.004	66.3	0.1
40	0.940	114.0	18.2	132.2	24.1	0.033	77.2	0.1
50	0.959	114.0	14.6	128.6	26.1	0.025	82.0	0.0
60	0.968	114.0	10.7	124.7	28.4	0.035	88.0	0.0
70	0.981	114.0	7.8	121.7	30.1	0.060	93.6	0.0
80	0.989	114.0	6.7	120.7	30.7	0.060	96.4	0.0
90	0.993	114.0	6.3	120.3	30.9	0.059	97.4	0.0
100	0.996	114.0	6.1	120.1	31.0	0.061	98.2	0.0
110	0.998	114.0	6.0	120.0	31.1	0.061	98.9	0.0
125	0.999	114.0	6.0	120.0	31.1	0.059	99.5	0.0
140	1.000	114.0	6.0	120.0	31.1	0.061	99.7	0.0

Table S6: Binary outcome: Type I error, power at $\tau = 1.417$, and relative bias of $\hat{\tau}_c$ the 2-stage AHCC design starting with single-arm design and interim sample size 120. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.107	114.0	61.1	175.1	3.2	0.002	7.1	28.7
10	0.085	114.0	59.1	173.1	4.4	0.003	4.2	26.6
20	0.074	114.0	54.2	168.2	7.0	0.003	3.0	26.2
30	0.066	114.0	48.9	162.9	10.0	0.002	2.2	26.5
40	0.060	114.0	42.6	156.6	13.5	0.000	1.3	26.8
50	0.061	114.0	36.3	150.3	16.9	0.002	1.7	27.1
60	0.059	114.0	29.5	143.4	20.7	0.002	1.1	26.7
70	0.050	114.0	22.2	136.2	24.7	0.003	0.8	27.7
80	0.052	114.0	15.5	129.6	28.4	0.003	0.6	27.4
90	0.051	114.0	11.3	125.3	30.8	0.002	0.7	27.3
100	0.045	114.0	8.2	122.2	32.5	0.002	0.5	26.6
110	0.045	114.0	6.8	120.7	33.3	0.002	0.5	26.6
125	0.049	114.0	6.1	120.1	33.6	0.002	0.4	26.8
140	0.051	114.0	6.0	120.0	33.7	0.002	0.5	26.8

				$\tau = 1.417$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.810	114.1	66.1	180.2	-3.5	-0.004	23.9	6.1
10	0.850	114.0	53.8	167.8	3.6	-0.010	35.5	2.3
20	0.855	114.0	44.6	158.5	8.9	-0.007	42.1	1.4
30	0.855	114.0	37.8	151.8	12.8	-0.003	46.7	1.0
40	0.872	114.0	29.2	143.2	17.7	0.002	54.6	0.4
50	0.860	114.0	27.1	141.1	18.9	-0.001	51.8	0.4
60	0.862	114.0	20.8	134.8	22.6	0.003	56.3	0.3
70	0.892	114.0	13.4	127.4	26.8	0.008	69.2	0.1
80	0.909	114.0	9.8	123.8	28.9	0.009	73.8	0.1
90	0.905	114.0	8.1	122.1	29.8	0.009	73.5	0.1
100	0.924	114.0	6.7	120.7	30.6	0.010	77.5	0.0
110	0.933	114.0	6.3	120.3	30.9	0.010	79.3	0.0
125	0.943	114.0	6.0	120.0	31.0	0.009	82.3	0.0
140	0.954	114.0	6.0	120.0	31.0	0.008	84.3	0.0

Table S7: Continuous outcome: Type I error, power at $\tau = 0.905$, relative bias, percentages of early stopping due to superiority and futility of $\hat{\tau}_c$ for the 2-stage AHCC design with interim sample size 120 when the randomization ratio is 3:1. n_t is the number of patients in the treatment arm; n_c is the number of patients in the concurrent control arm; n is the total number of patients in the trial; % n save is the percentage of sample size saved for the trial by using AHCC design leveraging historical data compared with a standard 2-stage RCT.

				$\tau = 0$				
ESS	Type I error	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.058	133.5	38.6	172.1	3.7	-0.009	0.7	26.8
10	0.062	134.2	32.5	166.7	6.7	-0.034	0.7	25.2
20	0.051	132.0	32.1	164.1	8.2	-0.012	0.6	26.0
30	0.056	131.7	32.2	163.8	8.4	0.012	0.5	26.4
40	0.049	131.6	32.2	163.8	8.4	0.010	0.4	26.7
50	0.049	131.4	32.2	163.6	8.5	0.015	0.5	26.8
60	0.048	131.5	32.2	163.7	8.4	0.014	0.4	26.8
70	0.048	131.4	32.2	163.6	8.5	0.010	0.3	27.0
80	0.052	131.5	32.2	163.7	8.4	0.015	0.4	26.8
90	0.047	131.0	32.2	163.2	8.7	0.015	0.4	27.7
100	0.048	131.6	32.2	163.8	8.4	0.012	0.3	26.6
110	0.051	131.4	32.2	163.5	8.5	0.014	0.5	27.0
125	0.053	131.2	32.2	163.3	8.6	0.013	0.5	27.3
140	0.048	131.1	32.2	163.3	8.7	0.013	0.4	27.5

				$\tau = 0.905$				
ESS	Power	n_t	n_c	n	% n save	Relative bias	% super	% futil
3	0.741	130.5	38.1	168.6	7.2	-0.045	31.0	1.5
10	0.825	122.7	31.9	154.6	14.9	-0.053	44.5	0.7
20	0.866	118.2	31.4	149.6	17.6	-0.017	50.3	0.4
30	0.896	115.1	31.3	146.4	19.4	0.004	55.8	0.2
40	0.924	111.6	31.1	142.7	21.4	0.020	62.0	0.1
50	0.936	108.9	31.0	139.8	23.0	0.010	66.9	0.1
60	0.958	105.5	30.8	136.3	25.0	0.012	72.8	0.1
70	0.972	103.3	30.7	134.0	26.2	0.024	76.7	0.0
80	0.981	101.1	30.6	131.6	27.5	0.028	80.6	0.0
90	0.988	99.1	30.5	129.6	28.7	0.026	84.1	0.0
100	0.990	98.1	30.4	128.5	29.3	0.030	85.9	0.0
110	0.993	96.6	30.3	127.0	30.1	0.025	88.4	0.0
125	0.997	95.3	30.3	125.6	30.9	0.029	90.7	0.0
140	0.999	93.9	30.2	124.1	31.7	0.028	93.2	0.0

The three scenarios in the sensitivity analyses.

- Scenario 1: $\mu_0 = (-0.85, -0.9, -0.85, 0.1)$, and

$$\Sigma_0 = \begin{pmatrix} 4 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 \\ 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

- Scenario 2: $\mu_0 = (-0.95, -0.7, -0.95, -0.2)$, and

$$\Sigma_0 = \begin{pmatrix} 4 & 1 & 1 & 1 \\ 1 & 4 & 1 & 1 \\ 1 & 1 & 4 & 1 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

- Scenario 3: $\mu_0 = (1.7, -0.5, -0.6, 0.15)$, and

$$\Sigma_0 = \begin{pmatrix} 5 & 2 & 2 & 2 \\ 2 & 5 & 2 & 2 \\ 2 & 2 & 5 & 2 \\ 2 & 2 & 2 & 5 \end{pmatrix}$$

Details of the choice of μ_0 and Σ_0 for historical data.

We set N_0, the number of historical control patients, to be 500. For historical control patients, we also assumed the covariates $\mathbf{x}$ followed a multivariate normal distribution $\phi_p(\mu_0, \Sigma_0)$, where μ_0 and Σ_0 were chosen to achieve varying ESS for historical data with

sample size $N_0 = 500$ when mapped to a current RCT dataset of size 100. (Depending on the amount of information that can be borrowed from the historical control and whether stage I is randomized or single armed, the sample size of the current RCT may fluctuate. For example, for the settings we evaluated in our simulation studies, the sample size of the current RCT ranges from 69.8 to 145.9. The sample size of 100 of the current RCT was an intermediate value.) Specifically, we randomly generated a large number, say 2000, sets of data (for tuning μ_0 and Σ_0). For each dataset, we generated 100 sets of $\mathbf{x}$ from $\phi_p(\mu_1, \Sigma_1)$ and $N_0 = 500$ sets of $\mathbf{x}$ from $\phi_p(\mu_0, \Sigma_0)$, and calculated the ESS of the historical data. The average ESS across the 2000 iterations was taken as ESS for the scenario.

An illustrated example of the trial

Suppose we consider a two-stage trial with targeted sample size $N = 200$. Stage I equally randomizes 60 patients to T and C so each treatment arm receives 30 patients. At stage II, we first calculate ESS_H of the historical control data and perform futility and superiority interim analyses based on O'Brien-Fleming boundary. If the futility stopping boundary is crossed, terminate the trial and conclude that T is futile. If the superiority stopping boundary is crossed, terminate the trial and conclude that T is superior. Otherwise, calculate TESS for the control arm and randomize $70 + \max(n_0, 100 - \text{TESS})$ patients in the ratio of $\big[70\big] : \big[\max(n_0, 100 - \text{TESS})\big]$ to T and C. For example, if TESS is 40, then we randomize 70 patients to T and 60 patients to C. If TESS is 120, then we randomize 70 patients to T and n_0 patient to C.