Supplemental material

Table S1. Operating characteristics of a parallel group trial with 30 patients compared to a cross-over trial with 15 patients for the reference scenario with expected effect size $\delta = 0.5$. The power (10^4 simulation runs) and type I error rate (T1E, 10^5 simulation runs) is given for the unweighted (no w), patient preference weighted (patient w), and treatment effect weighted (effect w) scores. Note that under the null hypothesis the treatment effect weighted and unweighted case coincide.

Design	ρ_0	Method	Power			T1E	
			no w	patient w	effect w	no w	patient w
Parallel group design	0.3	Mean	0.34	0.33	0.44	0.050	0.050
		Kiresuk	0.37	0.35	0.46	0.050	0.050
		GEE	0.39	0.37	0.49	0.056	0.056
Cross-over design $\rho_\epsilon = 0$	0.3	Mean	0.53	0.49	0.70	0.049	0.050
		Kiresuk	0.57	0.53	0.69	0.050	0.051
		GEE 1	0.65	0.58	0.73	0.067	0.070
		GEE 2	0.65	0.59	0.73	0.055	0.055
Cross-over design $\rho_\epsilon = 0.3$	0.3	Mean	0.70	0.64	0.86	0.048	0.049
		Kiresuk	0.74	0.67	0.85	0.050	0.049
		GEE 1	0.81	0.71	0.88	0.066	0.068
		GEE 2	0.82	0.73	0.89	0.054	0.055
Parallel group design	0	Mean	0.43	0.40	0.51	0.049	0.050
		Kiresuk	0.48	0.44	0.56	0.050	0.050
		GEE	0.55	0.50	0.65	0.062	0.062
Cross-over design $\rho_\epsilon = 0$	0	Mean	0.43	0.39	0.58	0.047	0.050
		Kiresuk	0.46	0.42	0.57	0.049	0.050
		GEE 1	0.53	0.47	0.61	0.068	0.070
		GEE 2	0.53	0.47	0.61	0.055	0.056
Cross-over design $\rho_\epsilon = 0.3$	0	Mean	0.53	0.48	0.70	0.047	0.047
		Kiresuk	0.57	0.51	0.69	0.048	0.049
		GEE 1	0.64	0.55	0.73	0.066	0.068
		GEE 2	0.64	0.55	0.73	0.054	0.055

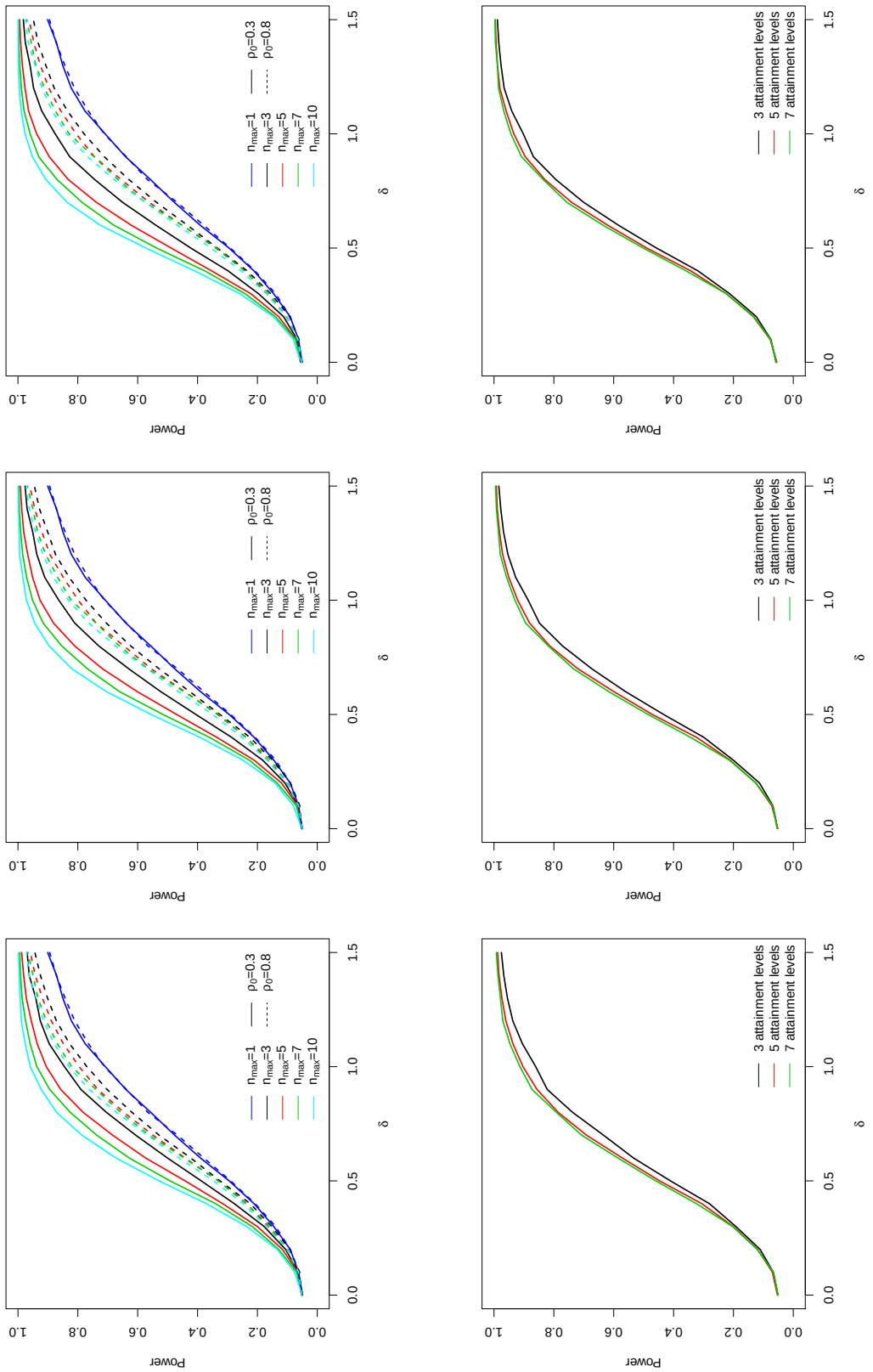

Figure S1. First row:The power for $m = 40$ and different maximum number of goals ($n_{max} = 1, 3, 5, 7, 10$) and correlations $\rho_0 = 0.3, 0.8$: (left) t-test applied to the per-subject means, (middle) standardized per-subject means, and (right) the test based on the GEE model. Second row: The power if discrete attainment scores with $L = 3, 5$, and 7 levels such that $c_j = \phi^{-1}(j/L), j = 1, \ldots, L$ and $\rho_0 = 0.3$ are used: (left) t-test applied to the per-subject means, (middle) standardized per-subject means, and (right) the test based on the GEE model, where $\rho = 0.3$ and $n_{max} = 5$.