Supplemental Material to:
Optimized multiple testing procedures for nested sub-populations based on a continuous biomarker

Alexandra Christine Graf[a], Dominic Magirr[b], Alex Dmitrienko[c], and Martin Posch[a,†]

[a] Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria
[b] Advanced Methodology and Data Science, Novartis Pharma AG, Basel, Switzerland
[c] Mediana, Overland Park, Kansas, United States of America

1 Linear Trend Model

Let $Y|X \sim N(a + bx, \sigma)$ and the biomarker values X be uniformly distributed between 0 and 1 as described for the linear trend model. Let q be some threshold and S_q the subgroup of all patients with biomarker value $x < q$. Then the expected value of patients in S_k is

$$\mu = E[Y|X < q] = \int_0^q \frac{1}{q} \int_{-\infty}^{\infty} y\varphi_{(a+bx,\sigma)}(y)dydx = \int_0^q \frac{1}{q}(a + bx)dx = a + b\frac{q}{2}$$

where $\varphi_{(\mu',\sigma')}$ is the density of the normal distribution with mean μ' and standard deviation σ'. The variance of patients in S_q can be calculated by

$$\begin{aligned}
Var[Y|X < q] &= \int_0^q \frac{1}{q} \int_{-\infty}^{\infty} (y - \mu)^2 \varphi_{(a+bx,\sigma)}(y)dydx \\
&= \frac{1}{q} \int_0^q \left[\int_{-\infty}^{\infty} y^2 \varphi_{(a+bx,\sigma)}(y)dy - 2\mu \int_{-\infty}^{\infty} y\varphi_{(a+bx,\sigma)}(y)dy + \mu^2 \int_{-\infty}^{\infty} \varphi_{(a+bx,\sigma)}(y)dy \right] dx \\
&= \frac{1}{q} \int_0^q \left[(a + bx)^2 + \sigma^2 - 2\mu(a + bx) + \mu^2 \right] dx \\
&= \frac{b^2q^2}{12} + \sigma^2
\end{aligned}$$

Setting $a = \beta_0 + \beta_2$ and $b = \beta_1 + \beta_3$ for $U = 1$ (treatment group) and $a = \beta_0$ and $b = \beta_1$ for $U = 0$ (control group) in the linear trend model gives (10) and (11) in the manuscript.

[†] corresponding author: e-mail: martin.posch@meduniwien.ac.at

2 Optimal designs

In the manuscript "Optimized multiple testing procedures for nested sub-populations based on a continuous biomarker" several prior distributions for the cut-off value γ in the step-function model as well as the linear trend model were investigated. For both models, we assumed a prior distribution in form of a beta-distribution with shape parameters (a, b). The results for the shape-parameter $(1, 1)$, $(2, 2)$, $(5, 5)$, $(1, 3)$ and $(3, 1)$ are shown in the manuscript. In the supplemental material we furthermore show the results for shape parameters $(15, 15)$, $(5, 15)$ and $(15, 5)$. The following Figure 1 shows the densities of all investigated priors. The $(1, 1)$-prior is indicating uncertainty over the whole range of γ-values. The $(2, 2)$, $(5, 5)$ and $(15, 15)$-priors are symmetric around 0.5 (Figure 1 (A)) and give more probability to the intermediate γ-values indicating a larger probability of the treatment-effective sub-population with a size half of the full population. The different variances of the beta distribution represent the degree of uncertainty on the true cut-off value γ in the planning phase of a trial. The asymmetric priors $(1, 3)$ and $(5, 15)$ (respectively $(3, 1)$ and $(15, 5)$) give a larger probability for a smaller (respectively larger) sub-population, were the treatment has a positive effect. Note that for both, the linear trend and the step function model, a sample size of $n = 112$ per group was used for the $(3, 1)$ and $(15, 5)$ prior, a sample size of $n = 336$ per group for the $(1, 3)$ and $(5, 15)$ prior as well as $n = 168$ per group for the symmetric priors.

2.1 Number of thresholds

Figure 2 shows the power as a function of the number of thresholds K which are assumed to be equally spaced at $q_K = k/K$ for $k = 1, \ldots, K$. The results are shown for the $(15, 15)$ (solid black line), the $(15, 5)$ (dotted black line) and the $(5, 15)$ prior (dashed black line) for different shape parameters λ for the step function model (SFM) as well as the linear trend model (LTM). As for the prior functions discussed in the manuscript, an adequate number of thresholds seems to be $K = 3$ or $K = 4$.

2.2 Shape parameter λ

Figure 3 shows the power as a function of λ for the SFM and LTM for $K = 2$, 3 and 4 for the $(15, 15)$ (solid black line), the $(15, 5)$ (dotted black line) and the $(5, 15)$ prior (dashed black line). The thresholds were set equally spaced at $q_K = k/K$ for $k = 1, \ldots, K$. Tables 1 and 2 give the optimal values. For comparison, the power when using $\lambda = 0$ is given.

Results of the shown priors are comparable to the results discussed in the manuscript.

2.3 Optimizing the shape parameter and spacing of thresholds

Figure 4 shows the results when optimizing the power simultaneously in both, λ and the thresholds q_k. Here, q_K was set to 1, always including the full population test. The optimal values are shown in more detail in Tables 1 and 2. For comparison, the values for $\lambda = 0$ but optimizing the thresholds q_k are given. Results of the shown priors are comparable to the results discussed in the manuscript. Over the λ-values the power using the optimal spacing are flat for negative λ values. For positive λ values a reduction in power can be seen. The largest gain in power due to the optimization can be seen for the $(5, 15)$ prior as compared to using equal spaced thresholds and $\lambda = 0$.

3 Simulation results for optimal designs

The thresholds and shape parameters were optimized to maximize the power to demonstrate a differential treatment effect in at least one of the subgroups averaged over the respective prior distribution. To investigate the full operating characteristics of the optimal designs, optimizing λ and the thresholds q_k simultaneously, we performed simulation studies. For each scenario we performed 250000 simulation runs. The results are shown in Tables 3, 4 and 5 for $K = 2$, 3 and 4, respectively. The tables show the power to detect a specific subgroup as well as the probability that a specific subgroup is the largest detected subgroup and may therefore be the most interesting one for a sponsor. Note that the probability that the full population (S_K) is the largest detected population is equal to the Power to detect S_K. Due to the chosen prior functions, the power to detect the full population is rather low, being largest for the $(3, 1)$ as well as the $(15, 5)$ prior where a larger effective sub-population was assumed. Note that the slight variations in the power to detect at least one subgroup between Tables 3, 4 and 5 of the supplement compared to Tables 1 and 2 in the manuscript as well as Tables 1 and 2 in the supplement come from the different calculation methods (numerical integration in Tables 1 and 2 of the manuscript and Tables 1 and 2 in the supplement as compared to simulation studies in Tables 3 to 5 of the supplement).

4 Model Misspecifications

In the manuscript we calculated optimal designs under a step function model as well as under a linear trend model. To evaluate the robustness of the optimal designs we simulated power values under model miss-specifications. Especially, we simulated designs under the Step Function Model (SFM) that were optimized under the Linear Trend Model (LTM) and vice versa. For each scenario we performed 250000 simulation runs. Tables 6, 7 and 8 shows the results for $K = 2$, 3 and 4 respectively. Results are shown for using the optimal designs (optimizing λ and the thresholds q_k simultaneously) of the LTM under the SFM and vice versa. The results can be directly compared to the results of the optimal designs in Tables 3, 4 and 5. When using the optimal designs of the LTM under the SFM a loss in the power to detect at least one sub-population up to 15% points for $K = 2$ and up to 11% for $K = 3$ and 4 is observed.

When using the optimal designs of the SFM when in reality the LTM holds, the loss in power is smaller, being up to 10% points for $K = 2$ and up to 7% for $K = 3$ and 6% for $K = 4$.

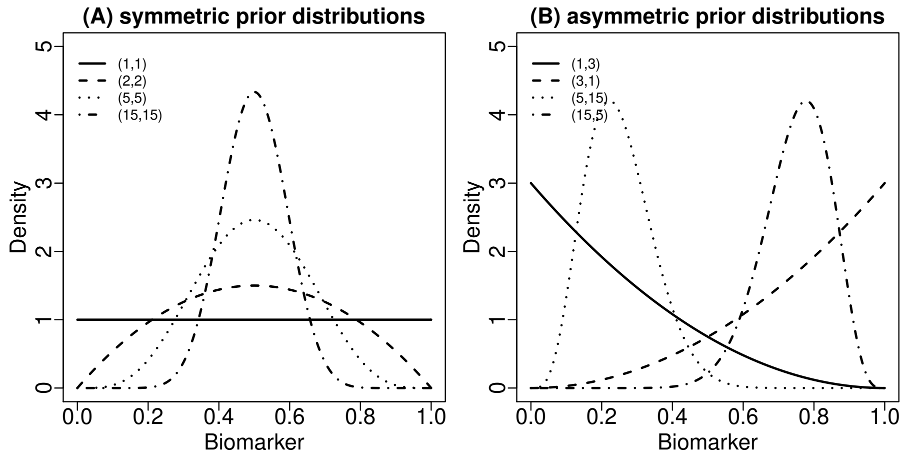

Figure 1: Densities of the different investigated prior distributions. Symmetric (A) and asymmetric (B) priors for different shape parameters.

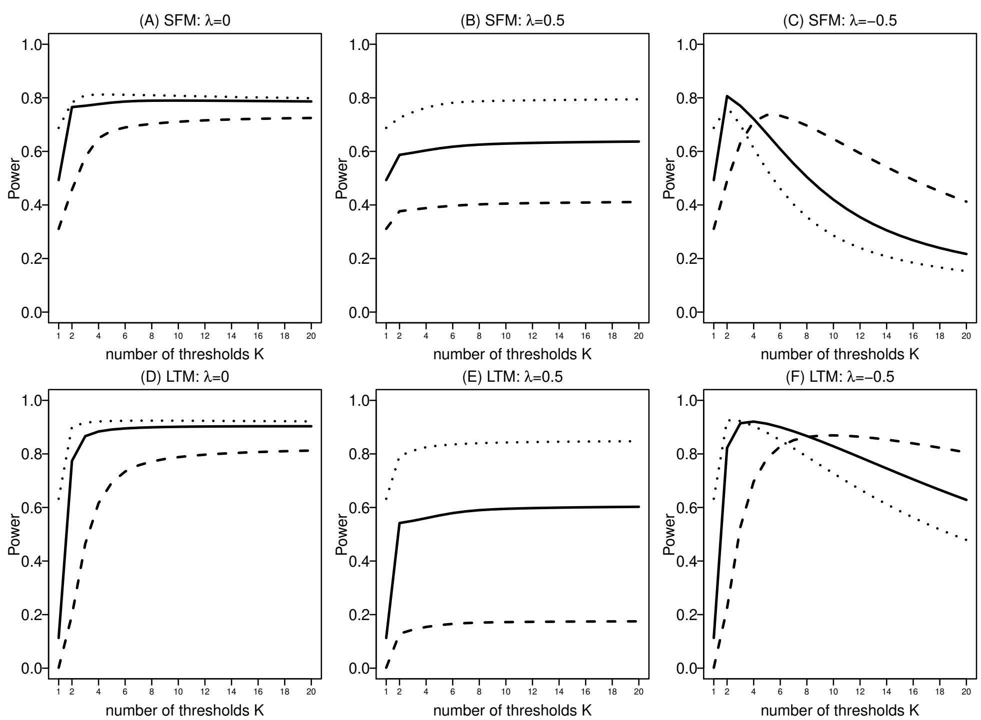

Figure 2: Power as a function of the number of thresholds K for the step function model (SFM, first row) and the linear trend model (LTM, second row). The thresholds were set equally spaced at $q_k = k/K$ for $k = 1, \ldots, K$. The shape parameter λ was set to 0 (first column), 0.5 (second column) and -0.5 (third column). The results are shown for the different prior distributions: (15,15) solid black line, (5,15) dashed black line, (15,5) dotted black line.

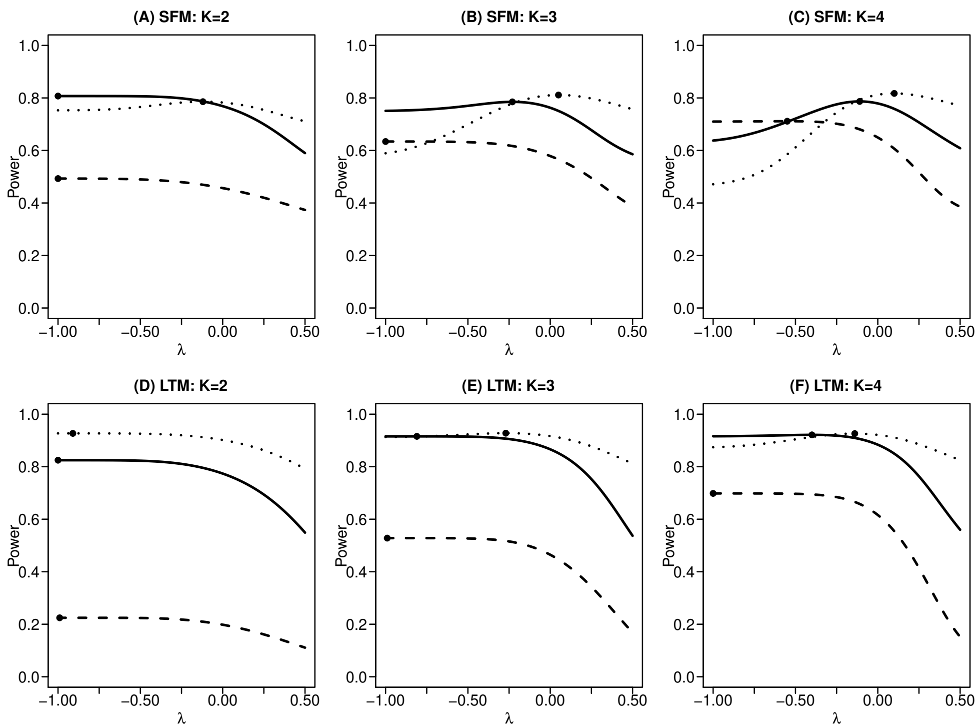

Figure 3: Power as a function of the shape parameter λ for the step function model (SFM, first row) and the linear trend model (LTM, second row) for 3 different prior distributions for γ: (15,15) solid black line, (5,15) dashed line, (15,5) dotted line. The thresholds were set equally spaced at $q_k = k/K$ for $k = 1, \ldots, K$ for $K = 2$ (first column), $K = 3$ (second column) and $K = 4$ (third column). The dots show the optimal power.

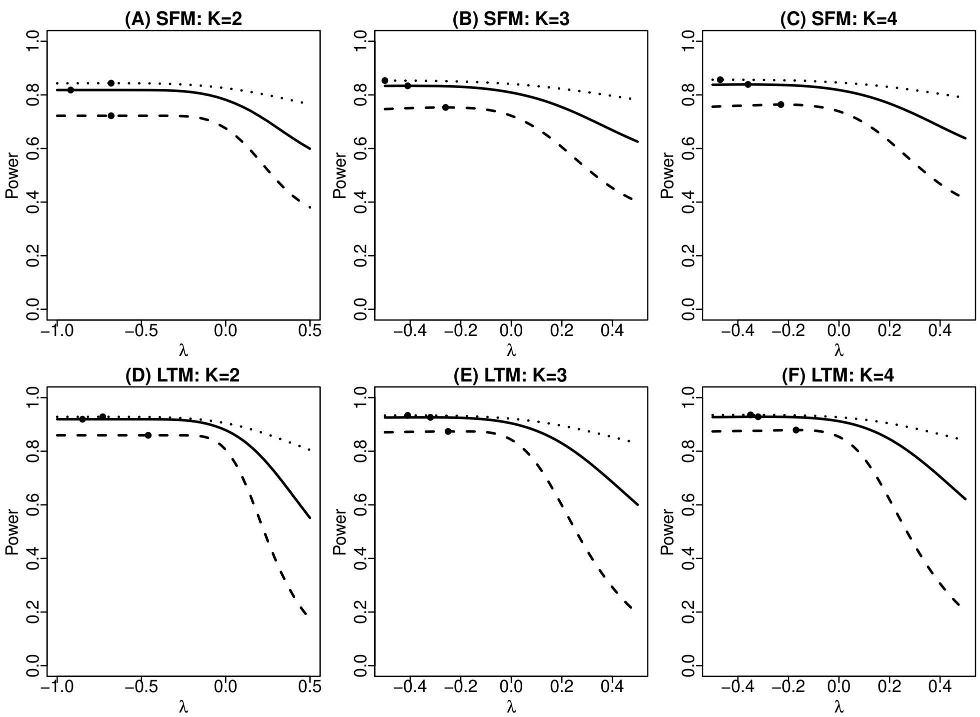

Figure 4: Power as a function of the shape parameter λ when using optimal thresholds q_k for each λ for $K = 2$ (first column), $K = 3$ (second column) and $K = 4$ (third column) for the step function model (SFM, first row) and the linear trend model (LTM, second row) for 3 different prior distributions for γ: (15,15) solid black line, (5,15) dashed line, (15,5) dotted line. The dots show the optimal power.

Table 1: Optimization results for the step function model when optimizing only λ with q_k equally spaced (ES) or both, λ and q_k (OS) for 3 prior distributions. For comparison, the results for setting $\lambda = 0$ are given for equally spaced thresholds and optimized thresholds.

			Prior distributions		
K	Setting		(15,15)	(5,15)	(15,5)
2	ES	Power $\lambda = 0$	0.769	0.457	0.783
		opt λ	-1.00	-1.00	-0.12
		opt Power	0.807	0.493	0.786
	OS	opt q_k for $\lambda = 0$	0.44/1	0.22/1	0.68/1
		opt Power $\lambda = 0$	0.782	0.675	0.825
		opt λ	-0.92	-0.68	-0.68
		opt q_k	0.46/1	0.22/1	0.70/1
		opt Power	0.818	0.722	0.844
3	ES	Power $\lambda = 0$	0.763	0.579	0.810
		opt λ	-0.23	-1.00	0.05
		opt Power	0.785	0.634	0.811
	OS	opt q_k for $\lambda = 0$	0.40/0.52/1	0.16/0.26/1	0.62/0.76/1
		opt Power $\lambda = 0$	0.809	0.722	0.841
		opt λ	-0.41	-0.26	-0.53
		opt q_k	0.42/0.54/1	0.18/0.28/1	0.66/0.78/1
		opt Power	0.834	0.753	0.853
4	ES	Power $\lambda = 0$	0.779	0.649	0.812
		opt λ	-0.11	-0.55	0.10
		opt Power	0.787	0.711	0.817
	OS	opt q_k for $\lambda = 0$	0.36/0.46/0.54/1	0.14/0.20/0.28/1	0.60/0.70/0.78/1
		opt Power $\lambda = 0$	0.818	0.738	0.846
		opt λ	-0.36	-0.23	-0.47
		opt q_k	0.40/0.48/0.56/1	0.16/0.24/0.32/1	0.64/0.74/0.82/1
		opt Power	0.839	0.764	0.857

Table 2: Optimization results for the linear trend model when optimizing only λ with q_k equally spaced (ES) or both, λ and q_k (OS) for 3 prior distributions. For comparison, the results for setting $\lambda = 0$ are given for equally spaced thresholds and optimized thresholds.

K	Setting		Prior distributions (15,15)	(5,15)	(15,5)
2	ES	Power $\lambda = 0$	0.774	0.198	0.901
		opt λ	-1.00	-1.00	-0.91
		opt Power	0.825	0.224	0.927
	OS	opt q_k for $\lambda = 0$	0.30/1	0.12/1	0.44/1
		opt Power $\lambda = 0$	0.879	0.805	0.905
		opt λ	-0.85	-0.46	-0.73
		opt q_k	0.28/1	0.12/1	0.46/1
		opt Power	0.919	0.859	0.928
3	ES	Power $\lambda = 0$	0.866	0.465	0.916
		opt λ	-0.81	-1.00	-0.27
		opt Power	0.916	0.528	0.928
	OS	opt q_k for $\lambda = 0$	0.22/0.36/1	0.08/0.16/1	0.34/0.54/1
		opt Power $\lambda = 0$	0.904	0.843	0.921
		opt λ	-0.32	-0.25	-0.41
		opt q_k	0.24/0.36/1	0.10/0.16/1	0.40/0.56/1
		opt Power	0.926	0.874	0.934
4	ES	Power $\lambda = 0$	0.884	0.616	0.921
		opt λ	-0.40	-1.00	-0.14
		opt Power	0.921	0.698	0.926
	OS	opt q_k for $\lambda = 0$	0.18/0.28/0.40/1	0.08/0.12/0.18/1	0.30/0.44/0.60/1
		opt Power $\lambda = 0$	0.912	0.854	0.927
		opt λ	-0.32	-0.17	-0.35
		opt q_k	0.24/0.32/0.40/1	0.08/0.12/0.18/1	0.38/0.50/0.62/1
		opt Power	0.929	0.879	0.935

Table 3: Simulation results for the optimal designs of the Step Function Model (SFM) and the Linear Trend Model (LTM) for the different prior distributions for $K = 2$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), the Power to detect Subgroup 2 (S_2, full population) and the probability that S_1 is the largest detected Subgroup. The probability that S_2 is the largest detected Subgroup is equal to the Power of S_2.

Model	Prior	Power at least 1	Power S_1	Power S_2	Prob S_1 largest
SFM	(1,1)	0.642	0.598	0.367	0.275
	(2,2)	0.692	0.670	0.295	0.397
	(5,5)	0.754	0.746	0.189	0.565
	(15,15)	0.818	0.818	0.003	0.815
	(1,3)	0.562	0.545	0.148	0.414
	(5,15)	0.718	0.718	0.000	0.718
	(3,1)	0.783	0.740	0.545	0.238
	(15,5)	0.843	0.840	0.345	0.498
LTM	(1,1)	0.735	0.728	0.220	0.514
	(2,2)	0.805	0.803	0.128	0.677
	(5,5)	0.874	0.874	0.031	0.843
	(15,15)	0.915	0.915	0.000	0.915
	(1,3)	0.666	0.665	0.031	0.634
	(5,15)	0.852	0.852	0.000	0.852
	(3,1)	0.881	0.875	0.374	0.507
	(15,5)	0.924	0.924	0.157	0.767

Table 4: Simulation results for the optimal designs of the Step Function Model (SFM) and the Linear Trend Model (LTM) for the different prior distributions for $K = 3$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), Subgroup 2 (S_2) as well as Subgroup 3 (S_3, full population) and the probability that S_1 as well as S_2 is the largest detected Subgroup. The probability that S_3 is the largest detected Subgroup is equal to the Power of S_3.

Model	Prior	Power at least 1	Power S_1	Power S_2	Power S_3	Prob S_1 largest	Prob S_2 largest
SFM	(1,1)	0.666	0.516	0.530	0.354	0.109	0.203
	(2,2)	0.718	0.599	0.566	0.283	0.134	0.301
	(5,5)	0.780	0.700	0.600	0.211	0.173	0.396
	(15,15)	0.834	0.795	0.670	0.106	0.162	0.566
	(1,3)	0.599	0.469	0.449	0.141	0.140	0.319
	(5,15)	0.751	0.678	0.568	0.036	0.182	0.534
	(3,1)	0.798	0.686	0.671	0.527	0.100	0.172
	(15,5)	0.852	0.817	0.731	0.380	0.114	0.358
LTM	(1,1)	0.758	0.643	0.631	0.227	0.123	0.408
	(2,2)	0.827	0.758	0.699	0.143	0.126	0.558
	(5,5)	0.890	0.846	0.773	0.063	0.116	0.711
	(15,15)	0.924	0.894	0.835	0.017	0.088	0.818
	(1,3)	0.700	0.600	0.562	0.035	0.138	0.527
	(5,15)	0.869	0.826	0.760	0.000	0.109	0.760
	(3,1)	0.892	0.842	0.793	0.385	0.093	0.413
	(15,5)	0.930	0.908	0.842	0.286	0.086	0.558

Table 5: Simulation results for the optimal designs of the Step Function Model (SFM) and the Linear Trend Model (LTM) for the different prior distributions for $K = 4$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), Subgroup 2 (S_2), Subgroup 3 (S_3) as well as Subgroup 4 (S_4, full population) and the probability that S_1, S_2 as well as S_3 is the largest detected Subgroup. The probability that S_4 is the largest detected Subgroup is equal to the Power of S_4.

Model	Prior	Power at least 1	Power S_1	Power S_2	Power S_3	Power S_4	Prob S_1 largest	Prob S_2 largest	Prob S_3 largest
SFM	(1,1)	0.676	0.460	0.519	0.485	0.343	0.071	0.100	0.161
	(2,2)	0.729	0.564	0.576	0.498	0.278	0.090	0.126	0.234
	(5,5)	0.790	0.669	0.653	0.539	0.199	0.089	0.153	0.348
	(15,15)	0.840	0.772	0.731	0.619	0.113	0.080	0.139	0.509
	(1,3)	0.614	0.421	0.456	0.386	0.130	0.091	0.131	0.262
	(5,15)	0.763	0.637	0.602	0.476	0.037	0.121	0.165	0.440
	(3,1)	0.804	0.639	0.674	0.634	0.526	0.059	0.087	0.133
	(15,5)	0.857	0.807	0.758	0.662	0.372	0.075	0.113	0.298
LTM	(1,1)	0.769	0.579	0.641	0.586	0.222	0.067	0.113	0.367
	(2,2)	0.834	0.721	0.722	0.622	0.139	0.075	0.137	0.484
	(5,5)	0.894	0.819	0.800	0.702	0.066	0.066	0.125	0.637
	(15,15)	0.926	0.887	0.855	0.783	0.015	0.051	0.092	0.768
	(1,3)	0.709	0.584	0.597	0.506	0.031	0.069	0.134	0.475
	(5,15)	0.876	0.773	0.782	0.710	0.000	0.053	0.112	0.710
	(3,1)	0.896	0.821	0.807	0.736	0.383	0.057	0.097	0.359
	(15,5)	0.934	0.899	0.865	0.792	0.304	0.049	0.091	0.490

Table 6: Model Misspecifications: Simulation results using the optimal designs of the Linear Trend Model under the Step Function Model (SFM) and using the optimal designs of the Step Function Model under the Linear Trend Model (LTM) for the different prior distributions for $K = 2$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), the Power to detect Subgroup 2 (S_2, full population) and the probability that S_1 is the largest detected Subgroup. The probability that S_2 is the largest detected Subgroup is equal to the Power of S_2.

Model	Prior	Power at least 1	Power S_1	Power S_2	Prob S_1 largest
SFM	(1,1)	0.549	0.461	0.242	0.307
	(2,2)	0.573	0.534	0.144	0.429
	(5,5)	0.606	0.601	0.034	0.572
	(15,15)	0.677	0.677	0.000	0.677
	(1,3)	0.460	0.434	0.070	0.390
	(5,15)	0.593	0.593	0.000	0.593
	(3,1)	0.674	0.591	0.351	0.323
	(15,5)	0.718	0.714	0.086	0.631
LTM	(1,1)	0.660	0.660	0.288	0.373
	(2,2)	0.715	0.715	0.204	0.511
	(5,5)	0.806	0.806	0.102	0.705
	(15,15)	0.857	0.857	0.004	0.853
	(1,3)	0.584	0.583	0.046	0.538
	(5,15)	0.751	0.751	0.000	0.751
	(3,1)	0.821	0.818	0.518	0.302
	(15,5)	0.873	0.872	0.425	0.448

Table 7: Model Miss-Specifications: Simulation results using the optimal designs of the Linear Trend Model under the Step Function Model (SFM) and using the optimal designs of the Step Function Model under the Linear Trend Model (LTM) for the different prior distributions for $K = 3$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), Subgroup 2 (S_2) as well as Subgroup 3 (S_3, full population) and the probability that S_1 as well as S_2 is the largest detected Subgroup. The probability that S_3 is the largest detected Subgroup is equal to the Power of S_3.

Model	Prior	Power at least 1	Power S_1	Power S_2	Power S_3	Prob S_1 largest	Prob S_2 largest
SFM	(1,1)	0.585	0.298	0.421	0.261	0.082	0.242
	(2,2)	0.619	0.401	0.484	0.177	0.091	0.351
	(5,5)	0.672	0.498	0.571	0.106	0.086	0.481
	(15,15)	0.723	0.581	0.651	0.064	0.067	0.591
	(1,3)	0.508	0.286	0.384	0.093	0.095	0.320
	(5,15)	0.641	0.490	0.552	0.013	0.087	0.541
	(3,1)	0.712	0.490	0.569	0.378	0.067	0.267
	(15,5)	0.752	0.630	0.671	0.238	0.063	0.450
LTM	(1,1)	0.700	0.688	0.540	0.279	0.159	0.262
	(2,2)	0.769	0.759	0.582	0.197	0.186	0.385
	(5,5)	0.835	0.828	0.603	0.102	0.232	0.501
	(15,15)	0.874	0.868	0.701	0.029	0.173	0.672
	(1,3)	0.638	0.629	0.444	0.044	0.195	0.400
	(5,15)	0.800	0.794	0.570	0.000	0.230	0.570
	(3,1)	0.845	0.833	0.694	0.503	0.147	0.195
	(15,5)	0.885	0.877	0.766	0.433	0.116	0.336

Table 8: Model Miss-Specifications: Simulation results using the optimal designs of the Linear Trend Model under the Step Function Model (SFM) and using the optimal designs of the Step Function Model under the Linear Trend Model (LTM) for the different prior distributions for $K = 4$. The table gives the Power to detect at least 1 subgroup, the Power to detect Subgroup 1 (S_1), Subgroup 2 (S_2), Subgroup 3 (S_3) as well as Subgroup 4 (S_4, full population) and the probability that S_1, S_2 as well as S_3 is the largest detected Subgroup. The probability that S_4 is the largest detected Subgroup is equal to the Power of S_4.

Model	Prior	Power at least 1	Power S_1	Power S_2	Power S_3	Power S_4	Prob S_1 largest	Prob S_2 largest	Prob S_3 largest
SFM	(1,1)	0.600	0.241	0.320	0.411	0.253	0.046	0.073	0.229
	(2,2)	0.641	0.341	0.415	0.464	0.170	0.052	0.092	0.327
	(5,5)	0.702	0.440	0.520	0.554	0.120	0.046	0.088	0.448
	(15,15)	0.748	0.564	0.617	0.642	0.057	0.038	0.065	0.587
	(1,3)	0.534	0.269	0.315	0.367	0.074	0.053	0.095	0.312
	(5,15)	0.655	0.380	0.464	0.531	0.035	0.039	0.079	0.502
	(3,1)	0.731	0.444	0.517	0.564	0.380	0.040	0.064	0.248
	(15,5)	0.768	0.587	0.633	0.661	0.263	0.034	0.056	0.415
LTM	(1,1)	0.717	0.691	0.603	0.474	0.272	0.109	0.134	0.202
	(2,2)	0.773	0.755	0.637	0.481	0.192	0.131	0.161	0.290
	(5,5)	0.844	0.828	0.714	0.536	0.101	0.124	0.183	0.436
	(15,15)	0.883	0.871	0.781	0.650	0.030	0.094	0.138	0.621
	(1,3)	0.657	0.633	0.522	0.356	0.042	0.132	0.168	0.314
	(5,15)	0.820	0.807	0.646	0.456	0.000	0.170	0.194	0.456
	(3,1)	0.853	0.830	0.748	0.638	0.499	0.093	0.116	0.144
	(15,5)	0.891	0.880	0.797	0.701	0.420	0.085	0.102	0.284