APPENDICES

Appendix A: Treatment switching probabilities

Table A1 presents the probability of switching for different patient groups at different time-points in Scenarios 1-4 and 9-12. Higher group numbers represent higher values for that group (that is, 'time to progression group' 0 are the control group patients that had time-to-progression times in the lowest 33.3% of the control group). Note however that these groups only refer to patients who became 'at-risk' of switching – that is, those control group patients that survived for longer than 21 days. Hence the lowest 33% represent the lowest third of the at-risk group, not the control group as a whole. Switching could happen at the three consultations immediately following disease progression, with the probability of switching declining in each consultation.

Table A1: Probability of treatment switch by prognostic groups and consultation – Good prognosis more likely to switch. Scenarios 1-4 and 9-12

Consultation 1 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.06	0.11	0.18
	1	0.17	0.29	0.42
	2	0.35	0.52	0.65
Consultation 2 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.05	0.09	0.15
	1	0.14	0.25	0.36
	2	0.30	0.46	0.60
Consultation 3 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.03	0.06	0.10
	1	0.09	0.17	0.26
	2	0.21	0.35	0.49

In Scenario 1 the mean switching proportion in the control group across the 1,000 simulations was 39.5%, which was equivalent to 57.5% of control group patients who became at-risk of switching – i.e. those that experienced disease progression.

Table A2 presents the probability of switching for different patient groups at different time-points in Scenarios 17-20 and 25-28. In these scenarios, poor prognosis patients were more likely to switch. In Scenario 17 the mean switching proportion in the control group across the 1,000 simulations was 38.7%, which was equivalent

to 56.2% of control group patients who became at-risk of switching – i.e. those that experienced disease

progression.

Table A2: Probability of treatment switch by prognostic groups and consultation – Poor prognosis more likely to switch. Scenarios 17-20 and 25-28

Consultation 1 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.80	0.67	0.55
	1	0.53	0.36	0.25
	2	0.20	0.11	0.07
Consultation 2 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.76	0.62	0.49
	1	0.47	0.31	0.21
	2	0.17	0.09	0.06
Consultation 3 (post progression)		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.67	0.50	0.38
	1	0.36	0.22	0.14
	2	0.11	0.06	0.04

All probabilities in Table A1 and Table A2 were decreased when investigating lower switching scenarios (i.e. in

Scenarios 5-8, 13-16, 21-24, 29-32. Probabilities were adjusted in Scenarios 33-96 in order to maintain similar

average switch proportions when survival distributions were altered to allow for varying levels of treatment

effect time dependency.

Appendix B: Scenario parameter values

In Table B1, values for each variable in Scenario 1 are quoted, as are alternative values for the 16 base scenarios.

Table B1: Simulated scenarios – Parameter values and alternatives tested

Variable	Value (Scenario 1)	Alternative Values
Sample size	500 (2:1 randomisation)	-
Number of prognosis groups (prog)	2	-
Probability of good prognosis	0.5	-
Probability of poor prognosis	0.5	-
Maximum follow-up time	1.5 years	-
Impact of bad prognosis on survival	Log hazard ratio = 0.3	-
Survival time distribution	Weibull parameters: Mix 1: Shape parameter 0.00001 　　　Scale parameter 0.00001 Mix 2: Shape parameter 2.0 　　　Scale parameter 0.8 p = 0.5 (mix parameter)	Weibull parameters to represent a more severe disease with more censoring: Mix 1: Shape parameter 0.00004 　　　Scale parameter 0.00004 Mix 2: Shape parameter 2.0 　　　Scale parameter 1.5 p = 0.5 (mix parameter)
Progression free survival	Overall survival time multiplied by a value from a beta distribution with shape parameters (5,10) – this implies the assumption that time to progression is 33% of OS. This is not an important assumption – time to progression is only included because we model a situation where switching cannot occur before disease progression	-
Baseline treatment effect (note this is not the true treatment effect as this does not take into account the effect of the treatment that occurs through the time-dependent confounder, biomarker level, or the time-dependent part of the treatment effect, η)	Baseline log hazard ratio in scenarios that include an additional time-dependent effect = -1.30	Alter log hazard ratio to -1.10 to maintain treatment effect with more severe disease Alter log hazard ratio to -0.35 to represent a smaller treatment effect Alter log hazard ratio to -0.65 to maintain smaller treatment effect with more severe disease
Biomarker intercept	Calculated using a normal distribution with mean of 20 and standard deviation of 1. Increased by 2.5 in patients who are in the poor prognosis group.	-
Biomarker value progression over time	As demonstrated by Equation (2). $\beta_2 = -0.02$ to represent that the biomarker value increases more slowly in the experimental group, and $\beta_1 = 0.04$ to indicate that the biomarker value increases over time	-
Impact of biomarker value on overall survival	As demonstrated by Equation (5). Increased biomarker value increases the risk of death. The strength of this relationship depends on the variable α, which equals 0.01 in Scenario 1	$\alpha = 0$ in scenarios with a constant treatment effect
Impact of biomarker value on treatment effect	Because treatment reduces the progression of the biomarker value and increased biomarker values increase the risk of death, the treatment has an additional effect through the biomarker.	All scenarios include a time-dependent treatment effect in the experimental group. However, in selected scenarios the treatment effect received by switchers equals the average treatment

	The strength of this relationship depends on the variable α, which equals 0.01 in Scenario 1	effect in the experimental group, satisfying the 'common treatment effect' assumption. In addition, $\alpha = 0$ in scenarios with a constant treatment effect
Time-dependent portion of treatment effect, η	$\eta = 0.003$ to generate a reduction in the treatment effect over time	All scenarios include a time-dependent treatment effect in the experimental group. However, in selected scenarios the treatment effect received by switchers equals the average treatment effect in the experimental group, satisfying the 'common treatment effect' assumption $\eta = 0$ in scenarios with a constant treatment effect $\eta = 0.006$ in scenarios with a stronger treatment effect time dependency
Assumed frequency of consultations	One every 3 weeks (21 days)	-
Probability of switching treatment over time	As shown in Table A1. This results in a switching proportion of approximately 40% in Scenario 1	Test a low switching scenario where all probabilities are decreased – to an extent where approximately 20% of control group patients switch.
Prognosis of switching patients	As shown in Table A1. This makes switching more likely in good prognosis patients, via a mechanism that takes into account both time to progression and biomarker value at progression	As shown in Table A2. This makes switching more likely in poor prognosis patients, via a mechanism that takes into account both time to progression and biomarker value at progression
Treatment effect in switching patients	Equal to baseline treatment effect multiplied by ω. Set ω such that treatment effect received by switching patients is 80% of the average effect received by experimental group patients in base scenarios.	Alter ω such that the "common treatment effect" assumption holds – the treatment effect received by switching patients equals 100% of the average effect received by experimental group patients.

Appendix C: Scenario settings

Scenario	Severity of disease	Relative treatment effect reduction in switchers	Switch proportion	Treatment effect	Switcher prognosis	Time-dependency of treatment effect
1	Moderate	20%	High	High	Good	Moderate
2	High	20%	High	High	Good	Moderate
3	Moderate	0%	High	High	Good	Moderate
4	High	0%	High	High	Good	Moderate
5	Moderate	20%	Moderate	High	Good	Moderate
6	High	20%	Moderate	High	Good	Moderate
7	Moderate	0%	Moderate	High	Good	Moderate
8	High	0%	Moderate	High	Good	Moderate
9	Moderate	20%	High	Low	Good	Moderate
10	High	20%	High	Low	Good	Moderate
11	Moderate	0%	High	Low	Good	Moderate
12	High	0%	High	Low	Good	Moderate
13	Moderate	20%	Moderate	Low	Good	Moderate
14	High	20%	Moderate	Low	Good	Moderate
15	Moderate	0%	Moderate	Low	Good	Moderate
16	High	0%	Moderate	Low	Good	Moderate
17	Moderate	20%	High	High	Poor	Moderate
18	High	20%	High	High	Poor	Moderate
19	Moderate	0%	High	High	Poor	Moderate
20	High	0%	High	High	Poor	Moderate
21	Moderate	20%	Moderate	High	Poor	Moderate
22	High	20%	Moderate	High	Poor	Moderate
23	Moderate	0%	Moderate	High	Poor	Moderate
24	High	0%	Moderate	High	Poor	Moderate
25	Moderate	20%	High	Low	Poor	Moderate
26	High	20%	High	Low	Poor	Moderate
27	Moderate	0%	High	Low	Poor	Moderate
28	High	0%	High	Low	Poor	Moderate
29	Moderate	20%	Moderate	Low	Poor	Moderate
30	High	20%	Moderate	Low	Poor	Moderate
31	Moderate	0%	Moderate	Low	Poor	Moderate
32	High	0%	Moderate	Low	Poor	Moderate
33	Moderate	20%	High	High	Good	Zero
34	High	20%	High	High	Good	Zero
35	Moderate	0%	High	High	Good	Zero
36	High	0%	High	High	Good	Zero
37	Moderate	20%	Moderate	High	Good	Zero
38	High	20%	Moderate	High	Good	Zero
39	Moderate	0%	Moderate	High	Good	Zero
40	High	0%	Moderate	High	Good	Zero
41	Moderate	20%	High	Low	Good	Zero
42	High	20%	High	Low	Good	Zero
43	Moderate	0%	High	Low	Good	Zero
44	High	0%	High	Low	Good	Zero
45	Moderate	20%	Moderate	Low	Good	Zero
46	High	20%	Moderate	Low	Good	Zero
47	Moderate	0%	Moderate	Low	Good	Zero
48	High	0%	Moderate	Low	Good	Zero
49	Moderate	20%	High	High	Poor	Zero
50	High	20%	High	High	Poor	Zero
51	Moderate	0%	High	High	Poor	Zero
52	High	0%	High	High	Poor	Zero
53	Moderate	20%	Moderate	High	Poor	Zero
54	High	20%	Moderate	High	Poor	Zero
55	Moderate	0%	Moderate	High	Poor	Zero
56	High	0%	Moderate	High	Poor	Zero

57	Moderate	20%	High	Low	Poor	Zero
58	High	20%	High	Low	Poor	Zero
59	Moderate	0%	High	Low	Poor	Zero
60	High	0%	High	Low	Poor	Zero
61	Moderate	20%	Moderate	Low	Poor	Zero
62	High	20%	Moderate	Low	Poor	Zero
63	Moderate	0%	Moderate	Low	Poor	Zero
64	High	0%	Moderate	Low	Poor	Zero
65	Moderate	20%	High	High	Good	Strong
66	High	20%	High	High	Good	Strong
67	Moderate	0%	High	High	Good	Strong
68	High	0%	High	High	Good	Strong
69	Moderate	20%	Moderate	High	Good	Strong
70	High	20%	Moderate	High	Good	Strong
71	Moderate	0%	Moderate	High	Good	Strong
72	High	0%	Moderate	High	Good	Strong
73	Moderate	20%	High	Low	Good	Strong
74	High	20%	High	Low	Good	Strong
75	Moderate	0%	High	Low	Good	Strong
76	High	0%	High	Low	Good	Strong
77	Moderate	20%	Moderate	Low	Good	Strong
78	High	20%	Moderate	Low	Good	Strong
79	Moderate	0%	Moderate	Low	Good	Strong
80	High	0%	Moderate	Low	Good	Strong
81	Moderate	20%	High	High	Poor	Strong
82	High	20%	High	High	Poor	Strong
83	Moderate	0%	High	High	Poor	Strong
84	High	0%	High	High	Poor	Strong
85	Moderate	20%	Moderate	High	Poor	Strong
86	High	20%	Moderate	High	Poor	Strong
87	Moderate	0%	Moderate	High	Poor	Strong
88	High	0%	Moderate	High	Poor	Strong
89	Moderate	20%	High	Low	Poor	Strong
90	High	20%	High	Low	Poor	Strong
91	Moderate	0%	High	Low	Poor	Strong
92	High	0%	High	Low	Poor	Strong
93	Moderate	20%	Moderate	Low	Poor	Strong
94	High	20%	Moderate	Low	Poor	Strong
95	Moderate	0%	Moderate	Low	Poor	Strong
96	High	0%	Moderate	Low	Poor	Strong

Appendix D: Overview of simulation scenarios

Table D1 presents key details associated with each of the scenarios simulated. Scenarios 1-16 are the base scenarios. Scenarios 17-32 replicate these but with poor prognosis patients more likely to switch treatments. Scenarios 33-64 replicate Scenarios 1-32 but with a constant treatment effect over time. Scenarios 65-96 replicate Scenarios 1-32 but with a stronger time dependency of the treatment effect.

The true area under the curve (restricted mean survival time (RMST) at 548 days) unconfounded by treatment switching is presented, along with the average treatment effect in terms of a hazard ratio (HR) and an acceleration factor (AF). These were estimated by generating scenario data for 1,000,000 patients without applying switching – RMST was estimated directly from this data, Cox models were used to estimate the HR, and an RPSFTM under no switching was used to estimate the AF. The HR and AF represent only an approximation of the true treatment effect as the proportional hazards and constant acceleration factor assumptions do not hold. In terms of a hazard ratio, the average treatment effect varied between 0.54 and 0.81.

The proportion of control group patients that switched, averaged across the 1000 simulations that made up each scenario, is also presented. The switching proportion varied between 17% and 57% of all control group patients. Switching proportions are probabilistic and are reliant on other characteristics. Table D1 also presents the switching proportion as a percentage of the control group patients that became 'at-risk' of switching. In our simulations control group patients could only switch treatments if they were alive at their first 'consultation' at 21 days and if their disease progressed before the end of the simulated follow-up. The switching proportion as a percentage of patients that became at-risk of switching is higher than when it is measured as a percentage of all control group patients – it ranged from 21% to 58%. We estimated the proportion of patients who became at risk of switching in each scenario by collecting data on the number of patients for whom disease progression was observed in each simulation and taking the mean. This is approximate, but appropriately indicative for our purposes.

Table D1 also presents details on whether the treatment effect was assumed to be 'common' – that is, whether the treatment effect received by switchers was the same as the average treatment effect received by patients initially randomised to the experimental group. To provide further information on the strength of any violations in the common treatment effect assumption we also include details on the treatment effect size received by switchers. We also indicate the extent to which the treatment effect changed over time.

Table D1 also presents details on the mean proportion of patients that were censored in each scenario – that is, the proportion for whom death was not observed. This varied between 10% and 58%.

Table D1: Overview of simulated scenarios

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Switcher prognosis	Strength of treatment effect time dependency	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	RMST (Control group)	RMST (Exp group)	HR	AF								
1	357.46	430.07	0.57	1.53	39.48%	57.48%	49.28%	Good	Moderate	No	1.42	80%
2	228.38	322.24	0.56	1.85	55.36%	57.52%	25.53%	Good	Moderate	No	1.68	80%
3	357.46	430.07	0.57	1.53	39.51%	57.25%	49.66%	Good	Moderate	Yes	1.53	100%
4	228.38	322.24	0.56	1.85	55.24%	57.40%	25.81%	Good	Moderate	Yes	1.85	100%
5	357.46	430.07	0.57	1.53	17.32%	25.14%	48.08%	Good	Moderate	No	1.42	80%
6	228.38	322.24	0.56	1.85	23.80%	24.72%	24.64%	Good	Moderate	No	1.68	80%
7	357.46	430.07	0.57	1.53	17.13%	24.83%	48.24%	Good	Moderate	Yes	1.53	100%
8	228.38	322.24	0.56	1.85	23.65%	24.56%	24.77%	Good	Moderate	Yes	1.85	100%
9	357.46	391.12	0.81	1.19	39.54%	57.37%	40.12%	Good	Moderate	No	1.15	80%
10	228.38	269.20	0.78	1.30	55.27%	57.40%	17.96%	Good	Moderate	No	1.24	80%
11	357.46	391.12	0.81	1.19	39.47%	57.26%	40.14%	Good	Moderate	Yes	1.19	100%
12	228.38	269.20	0.78	1.30	55.30%	57.46%	18.14%	Good	Moderate	Yes	1.30	100%
13	357.46	391.12	0.81	1.19	17.20%	25.02%	39.66%	Good	Moderate	No	1.15	80%
14	228.38	269.20	0.78	1.30	23.91%	24.85%	17.70%	Good	Moderate	No	1.24	80%
15	357.46	391.12	0.81	1.19	17.05%	24.83%	39.77%	Good	Moderate	Yes	1.19	100%
16	228.38	269.20	0.78	1.30	23.89%	24.82%	17.73%	Good	Moderate	Yes	1.30	100%
17	357.46	430.07	0.57	1.53	38.74%	56.23%	48.05%	Poor	Moderate	No	1.42	80%
18	228.38	322.24	0.56	1.85	51.18%	53.17%	24.23%	Poor	Moderate	No	1.68	80%
19	357.46	430.07	0.57	1.53	38.47%	55.75%	48.25%	Poor	Moderate	Yes	1.53	100%
20	228.38	322.24	0.56	1.85	51.25%	53.24%	24.44%	Poor	Moderate	Yes	1.85	100%
21	357.46	430.07	0.57	1.53	16.75%	24.34%	47.33%	Poor	Moderate	No	1.42	80%
22	228.38	322.24	0.56	1.85	19.89%	20.66%	23.82%	Poor	Moderate	No	1.68	80%
23	357.46	430.07	0.57	1.53	16.93%	24.47%	47.22%	Poor	Moderate	Yes	1.53	100%
24	228.38	322.24	0.56	1.85	20.01%	20.79%	23.92%	Poor	Moderate	Yes	1.85	100%
25	357.46	391.12	0.81	1.19	38.49%	55.93%	39.51%	Poor	Moderate	No	1.15	80%

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Switcher prognosis	Strength of treatment effect time dependency	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	RMST (Control group)	RMST (Exp group)	HR	AF								
26	228.38	269.20	0.78	1.30	51.31%	53.27%	17.48%	Poor	Moderate	No	1.24	80%
27	357.46	391.12	0.81	1.19	38.58%	55.93%	39.50%	Poor	Moderate	Yes	1.19	100%
28	228.38	269.20	0.78	1.30	51.33%	53.33%	17.48%	Poor	Moderate	Yes	1.30	100%
29	357.46	391.12	0.81	1.19	16.95%	24.57%	39.23%	Poor	Moderate	No	1.15	80%
30	228.38	269.20	0.78	1.30	19.94%	20.73%	17.25%	Poor	Moderate	No	1.24	80%
31	357.46	391.12	0.81	1.19	16.80%	24.40%	39.36%	Poor	Moderate	Yes	1.19	100%
32	228.38	269.20	0.78	1.30	19.96%	20.72%	17.30%	Poor	Moderate	Yes	1.30	100%
33	400.80	456.60	0.55	1.40	53.34%	55.36%	56.76%	Good	Zero	No	1.32	80%
34	315.41	390.14	0.55	1.40	55.34%	55.94%	33.29%	Good	Zero	No	1.32	80%
35	400.80	456.60	0.55	1.40	53.41%	55.42%	57.07%	Good	Zero	Yes	1.40	100%
36	315.41	390.14	0.55	1.40	55.46%	56.08%	34.08%	Good	Zero	Yes	1.40	100%
37	400.80	456.60	0.55	1.40	23.63%	24.49%	55.69%	Good	Zero	No	1.43	80%
38	315.41	390.14	0.55	1.40	24.44%	24.70%	32.03%	Good	Zero	No	1.63	80%
39	400.80	456.60	0.55	1.40	23.58%	24.47%	55.84%	Good	Zero	Yes	1.40	100%
40	315.41	390.14	0.55	1.40	24.47%	24.73%	32.23%	Good	Zero	Yes	1.40	100%
41	400.80	423.33	0.80	1.13	53.50%	55.52%	47.17%	Good	Zero	No	1.10	80%
42	315.41	343.95	0.80	1.13	55.42%	56.01%	23.04%	Good	Zero	No	1.10	80%
43	400.80	423.33	0.80	1.13	53.70%	55.69%	47.46%	Good	Zero	Yes	1.13	100%
44	315.41	343.95	0.80	1.13	55.57%	56.18%	23.33%	Good	Zero	Yes	1.13	100%
45	400.80	423.33	0.80	1.13	23.54%	24.43%	46.78%	Good	Zero	No	1.17	80%
46	315.41	343.95	0.80	1.13	24.29%	24.55%	22.56%	Good	Zero	No	1.20	80%
47	400.80	423.33	0.80	1.13	23.50%	24.40%	46.93%	Good	Zero	Yes	1.13	100%
48	315.41	343.95	0.80	1.13	24.41%	24.67%	22.63%	Good	Zero	Yes	1.13	100%
49	400.80	456.60	0.55	1.40	55.88%	58.00%	57.085	Poor	Zero	No	1.43	80%
50	315.41	390.14	0.55	1.40	56.75%	57.37%	32.43%	Poor	Zero	No	1.63	80%
51	400.80	456.60	0.55	1.40	55.94%	58.00%	57.51%	Poor	Zero	Yes	1.40	100%

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Switcher prognosis	Strength of treatment effect time dependency	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	RMST (Control group)	RMST (Exp group)	HR	AF								
52	315.41	390.14	0.55	1.40	56.74%	57.37%	32.73%	Poor	Zero	Yes	1.40	100%
53	400.80	456.60	0.55	1.40	26.71%	27.74%	55.86%	Poor	Zero	No	1.43	80%
54	315.41	390.14	0.55	1.40	26.37%	26.65%	31.31%	Poor	Zero	No	1.63	80%
55	400.80	456.60	0.55	1.40	26.72%	27.74%	55.96%	Poor	Zero	Yes	1.40	100%
56	315.41	390.14	0.55	1.40	26.10%	26.39%	31.43%	Poor	Zero	Yes	1.40	100%
57	400.80	423.33	0.80	1.13	55.96%	58.05%	47.30%	Poor	Zero	No	1.17	80%
58	315.41	343.95	0.80	1.13	56.70%	57.32%	22.69%	Poor	Zero	No	1.20	80%
59	400.80	423.33	0.80	1.13	55.86%	57.97%	47.50%	Poor	Zero	Yes	1.13	100%
60	315.41	343.95	0.80	1.13	56.79%	57.41%	22.59%	Poor	Zero	Yes	1.13	100%
61	400.80	423.33	0.80	1.13	26.69%	27.69%	46.83%	Poor	Zero	No	1.17	80%
62	315.41	343.95	0.80	1.13	26.22%	26.51%	22.31%	Poor	Zero	No	1.20	80%
63	400.80	423.33	0.80	1.13	26.77%	27.76%	46.88%	Poor	Zero	Yes	1.13	100%
64	315.41	343.95	0.80	1.13	26.18%	26.45%	22.24%	Poor	Zero	Yes	1.13	100%
65	357.46	438.20	0.56	1.52	39.58%	57.45%	45.15%	Good	Strong	No	1.41	80%
66	228.38	333.63	0.54	1.85	55.47%	57.62%	19.03%	Good	Strong	No	1.68	80%
67	357.46	438.20	0.56	1.52	39.54%	57.32%	45.63%	Good	Strong	Yes	1.41	100%
68	228.38	333.63	0.54	1.85	55.59%	57.70%	19.43%	Good	Strong	Yes	1.85	100%
69	357.46	438.20	0.56	1.52	17.17%	24.93%	44.03%	Good	Strong	No	1.41	80%
70	228.38	333.63	0.54	1.85	23.79%	24.73%	18.36%	Good	Strong	No	1.68	80%
71	357.46	438.20	0.56	1.52	17.20%	24.90%	44.20%	Good	Strong	Yes	1.52	100%
72	228.38	333.63	0.54	1.85	23.72%	24.64%	18.44%	Good	Strong	Yes	1.85	100%
73	357.46	405.74	0.78	1.21	39.63%	57.58%	35.57%	Good	Strong	No	1.17	80%
74	228.38	273.39	0.81	1.23	55.43%	57.61%	11.09%	Good	Strong	No	1.19	80%
75	357.46	405.74	0.78	1.21	39.47%	57.31%	35.75%	Good	Strong	Yes	1.21	100%
76	228.38	273.39	0.81	1.23	55.47%	57.60%	11.18%	Good	Strong	Yes	1.23	100%
77	357.46	405.74	0.78	1.21	17.15%	24.97%	35.15%	Good	Strong	No	1.17	80%

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Switcher prognosis	Strength of treatment effect time dependency	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	RMST (Control group)	RMST (Exp group)	HR	AF								
78	228.38	273.39	0.81	1.23	23.79%	24.71%	10.83%	Good	Strong	No	1.19	80%
79	357.46	405.74	0.78	1.21	17.24%	24.98%	35.18%	Good	Strong	Yes	1.21	100%
80	228.38	273.39	0.81	1.23	23.79%	24.72%	10.88%	Good	Strong	Yes	1.23	100%
81	357.46	438.20	0.56	1.52	38.60%	56.04%	44.00%	Poor	Strong	No	1.41	80%
82	228.38	333.63	0.54	1.85	50.98%	52.94%	17.87%	Poor	Strong	No	1.68	80%
83	357.46	438.20	0.56	1.52	38.48%	55.89%	44.35%	Poor	Strong	Yes	1.52	100%
84	228.38	333.63	0.54	1.85	51.10%	53.09%	17.99%	Poor	Strong	Yes	1.85	100%
85	357.46	438.20	0.56	1.52	16.89%	24.44%	43.13%	Poor	Strong	No	1.41	80%
86	228.38	333.63	0.54	1.85	20.06%	20.84%	17.47%	Poor	Strong	No	1.68	80%
87	357.46	438.20	0.56	1.52	17.03%	24.69%	43.35%	Poor	Strong	Yes	1.52	100%
88	228.38	333.63	0.54	1.85	19.98%	20.75%	17.54%	Poor	Strong	Yes	1.85	100%
89	357.46	405.74	0.78	1.21	38.66%	56.11%	35.04%	Poor	Strong	No	1.17	80%
90	228.38	273.39	0.81	1.23	51.18%	53.17%	10.59%	Poor	Strong	No	1.19	80%
91	357.46	405.74	0.78	1.21	38.54%	55.91%	35.08%	Poor	Strong	Yes	1.21	100%
92	228.38	273.39	0.81	1.23	51.10%	53.13%	10.59%	Poor	Strong	Yes	1.23	100%
93	357.46	405.74	0.78	1.21	16.78%	24.37%	34.66%	Poor	Strong	No	1.17	80%
94	228.38	273.39	0.81	1.23	20.01%	20.81%	10.43%	Poor	Strong	No	1.19	80%
95	357.46	405.74	0.78	1.21	16.77%	24.34%	34.79%	Poor	Strong	Yes	1.21	100%
96	228.38	273.39	0.81	1.23	19.94%	20.71%	10.46%	Poor	Strong	Yes	1.23	100%

Appendix E: Percentage bias across all scenarios (note different axis scales)

Figure E1: Percentage bias, Scenarios 1-16
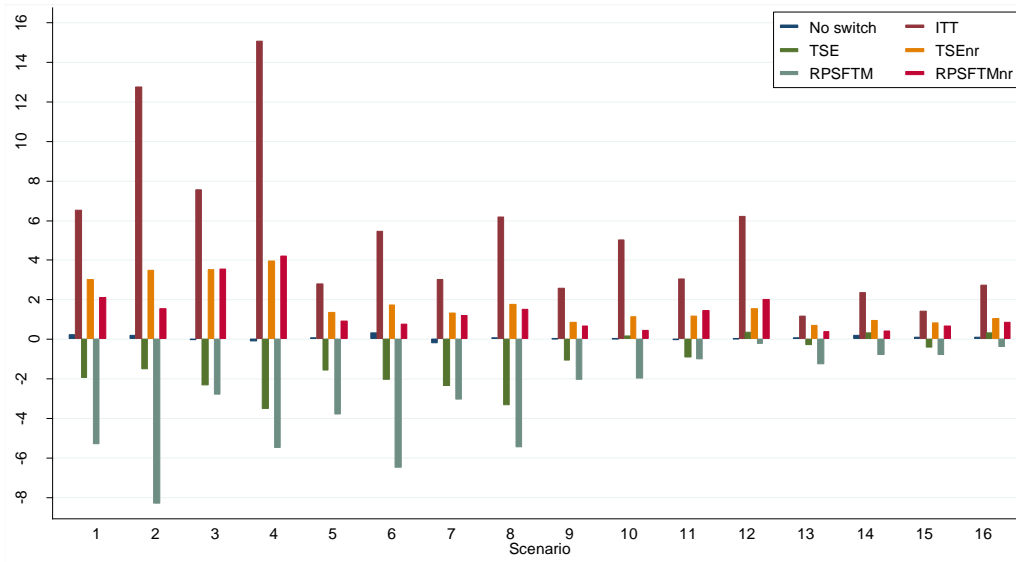
Figure E2: Percentage bias, Scenarios 33-48
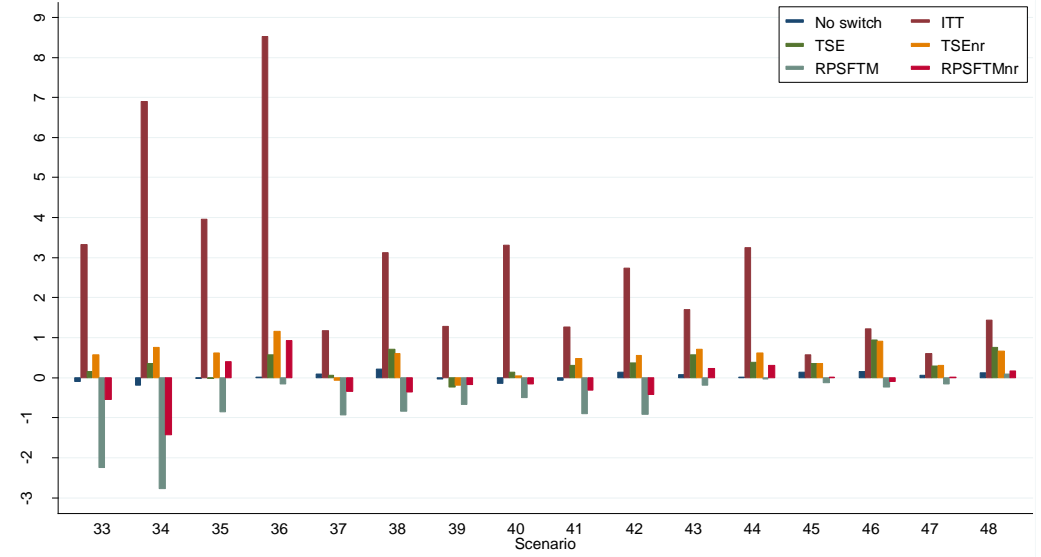
Figure E3: Percentage bias, Scenarios 17-32
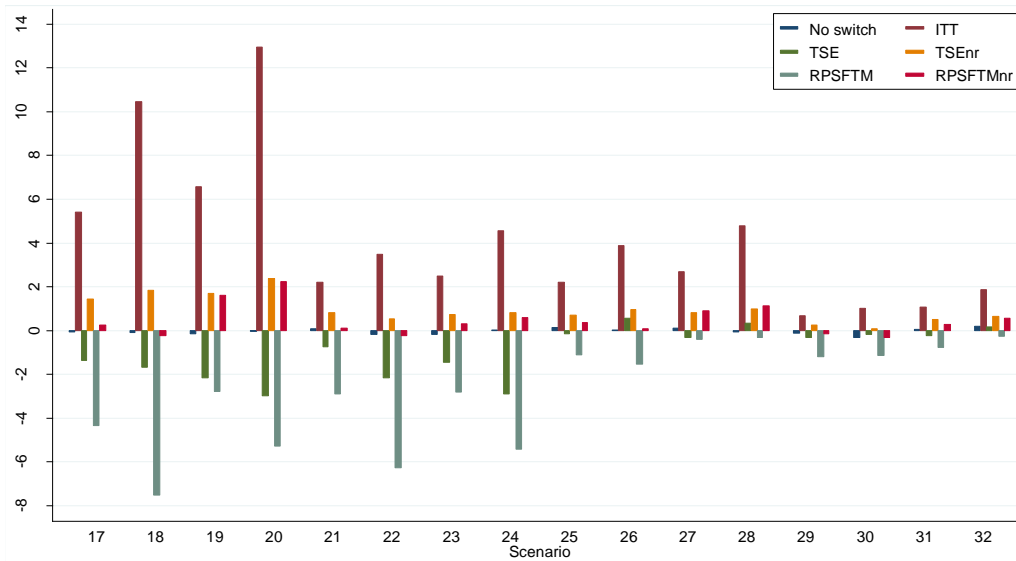
Figure E4: Percentage bias, Scenarios 49-64
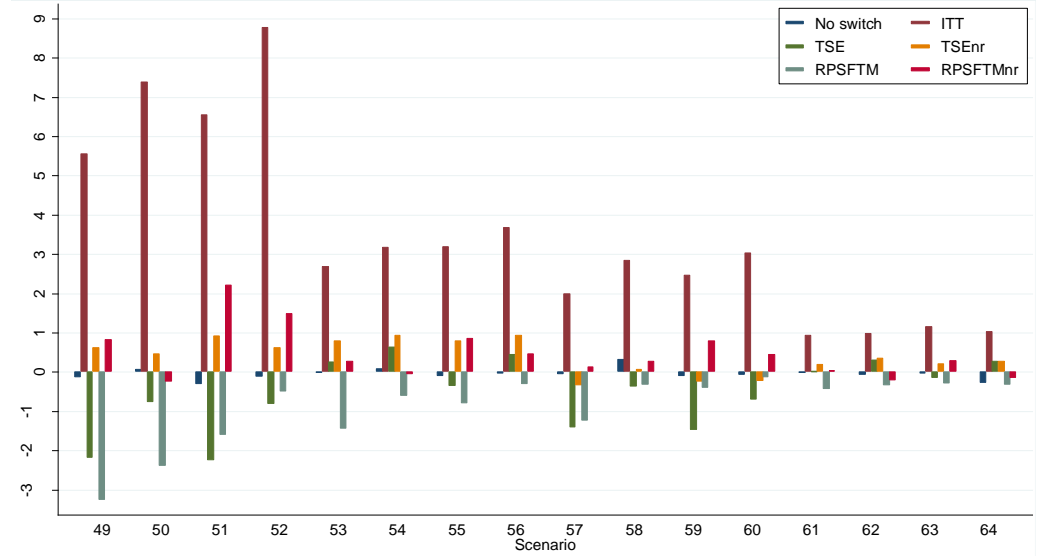
Figure E5: Percentage bias, Scenarios 65-80
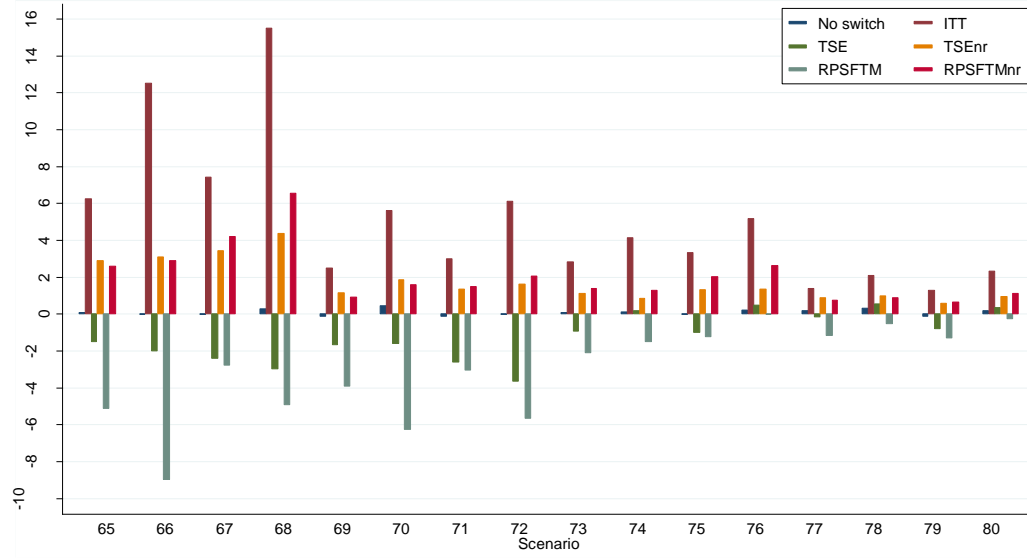
Figure E6: Percentage bias, Scenarios 81-96
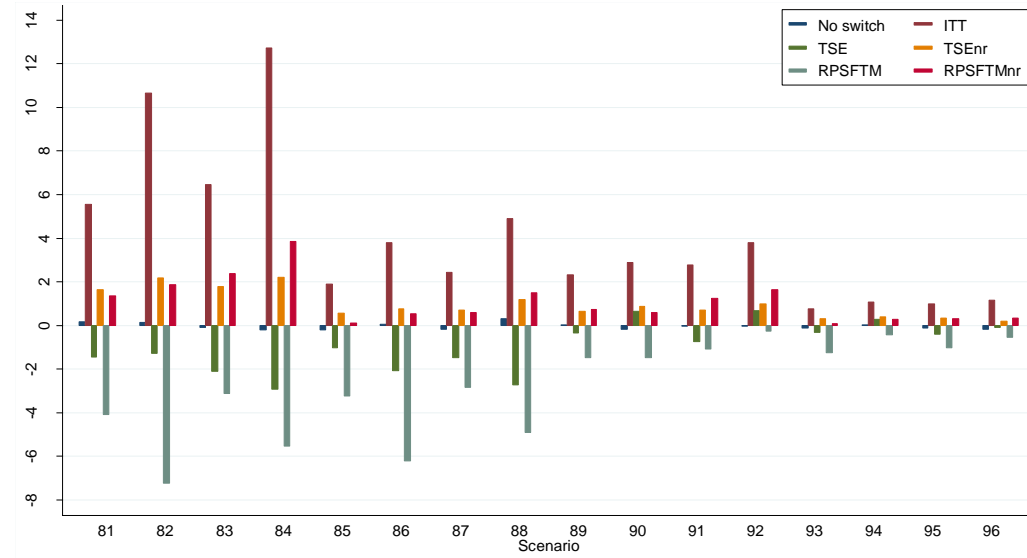
Appendix F: Empirical standard error of percentage bias across all scenarios (note different axis scales)

Figure F1: Empirical standard error, Scenarios 1-16

Figure F3: Empirical standard error, Scenarios 33-48

Figure F2: Empirical standard error, Scenarios 17-32
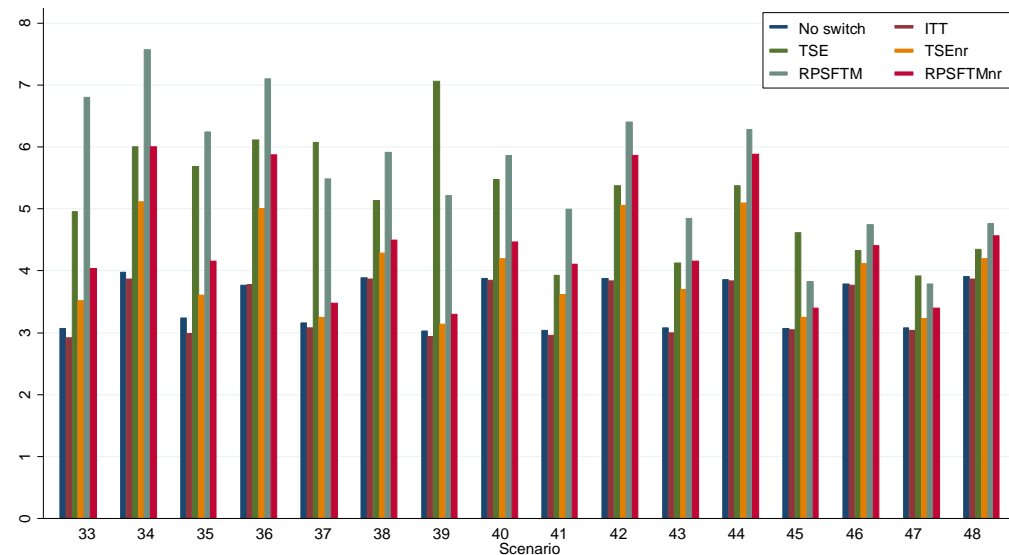
Figure F4: Empirical standard error, Scenarios 49-64
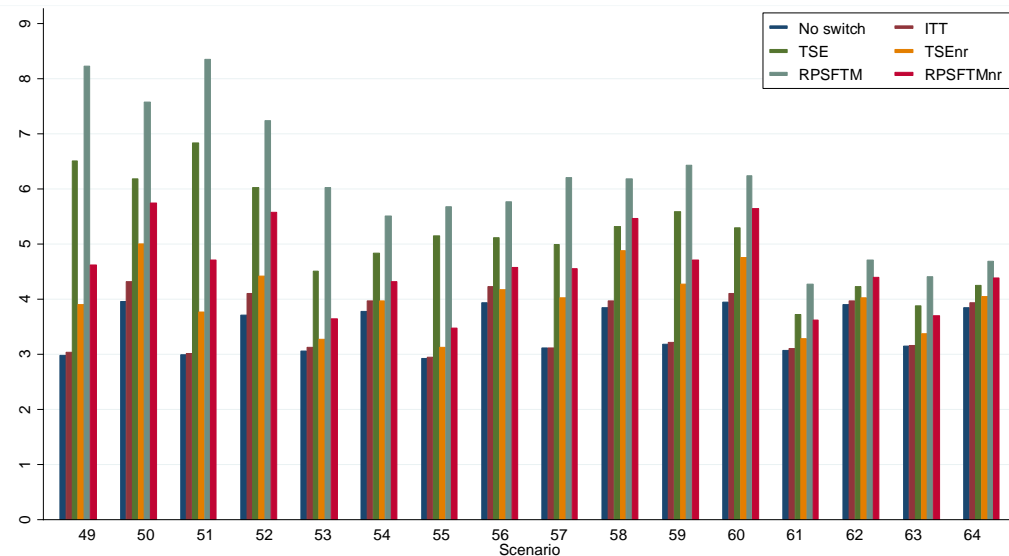
Figure F5: Empirical standard error, Scenarios 65-80
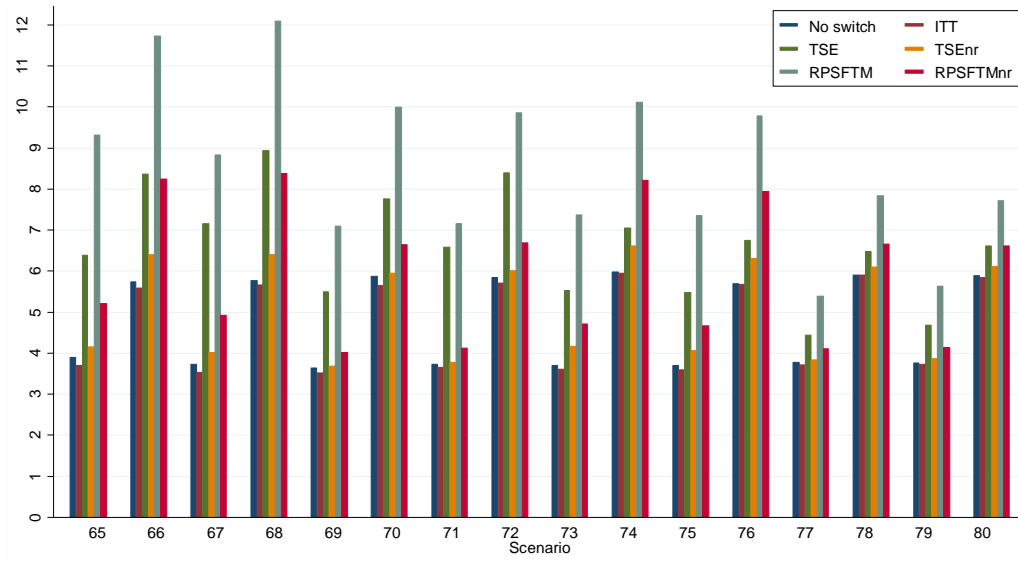
Figure F6: Empirical standard error, Scenarios 81-96
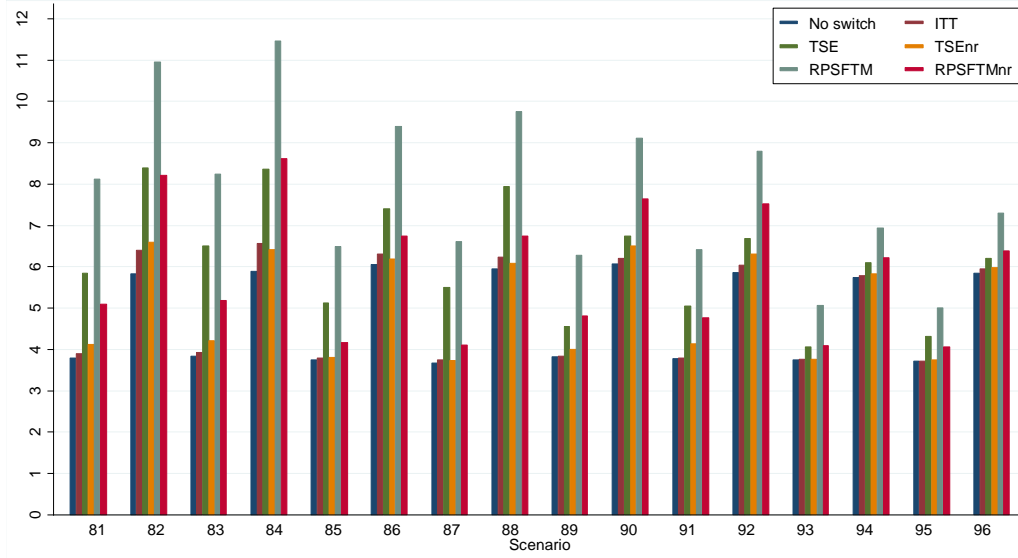
Appendix G: Root mean squared error of percentage bias across all scenarios (note different axis scales)

Figure G1: Root mean squared error, Scenarios 1-16
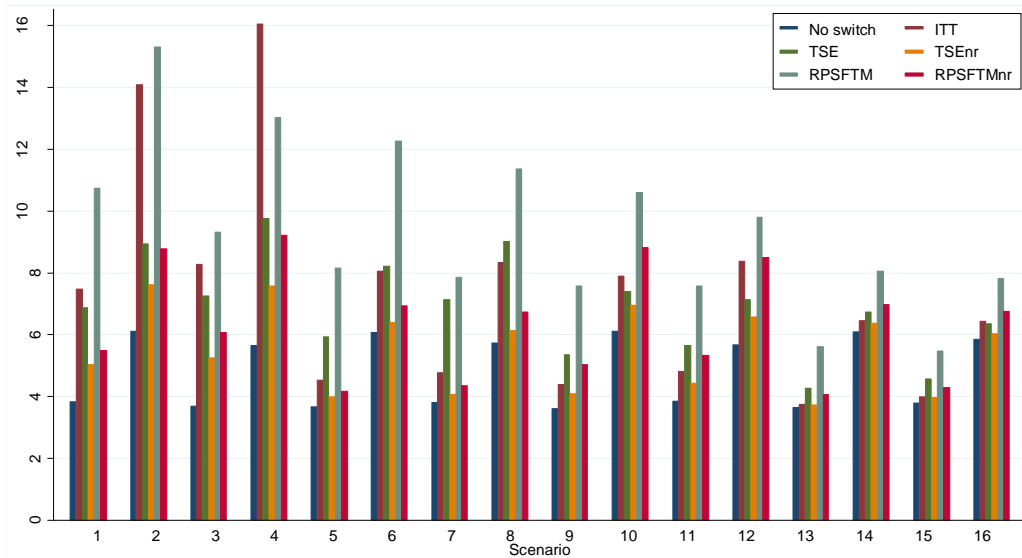
Figure G3: Root mean squared error, Scenarios 33-48
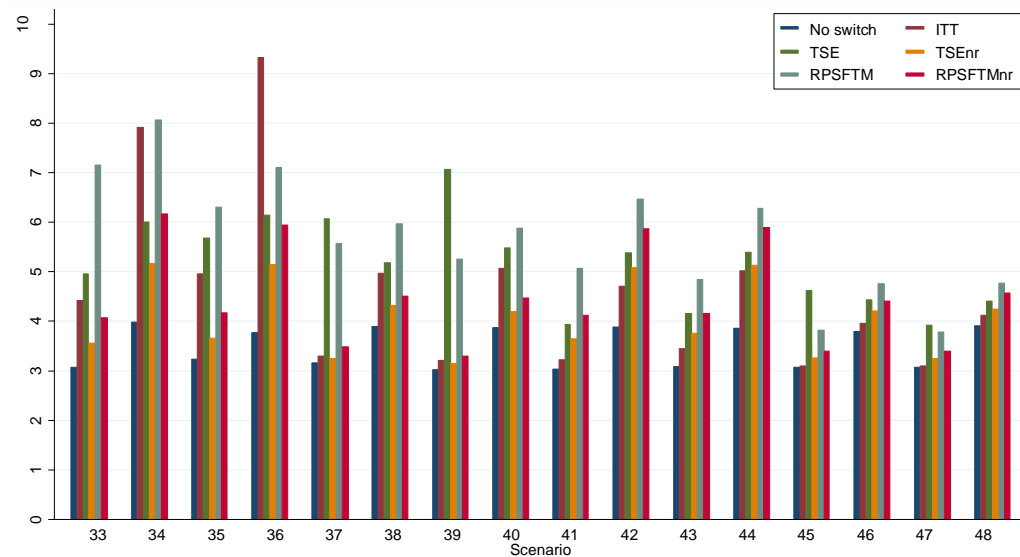
Figure G2: Root mean squared error, Scenarios 17-32
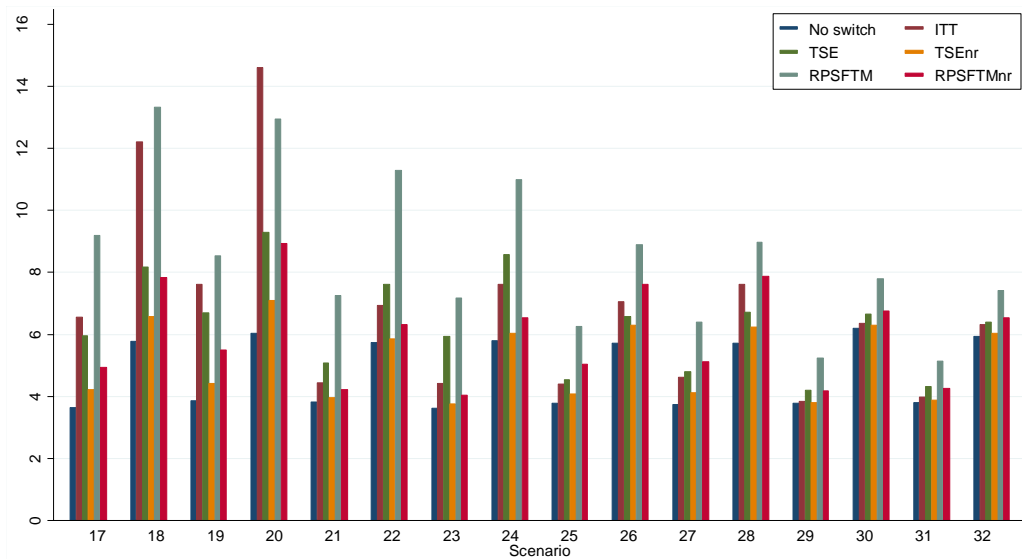
Figure G4: Root mean squared error, Scenarios 49-64
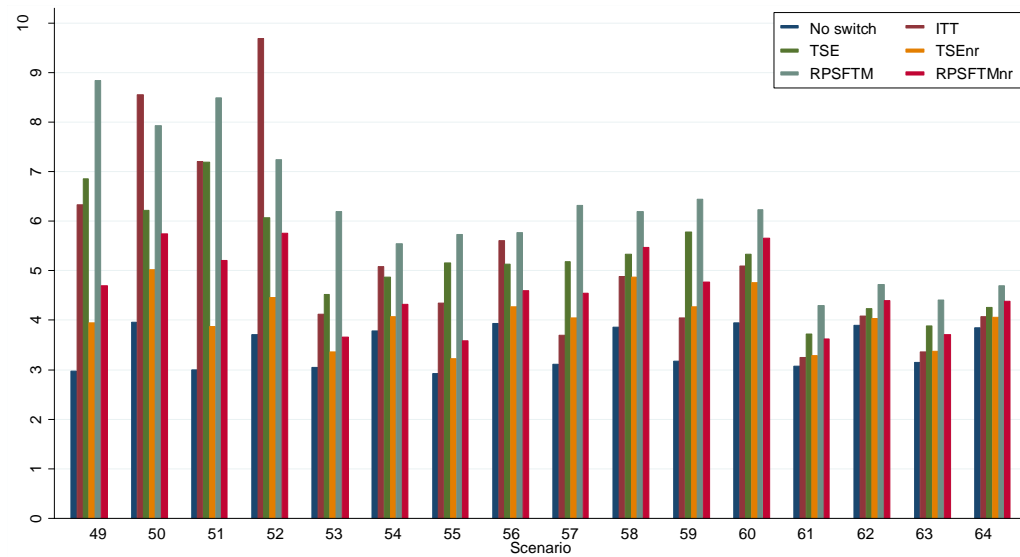
Figure G5: Root mean squared error, Scenarios 65-80
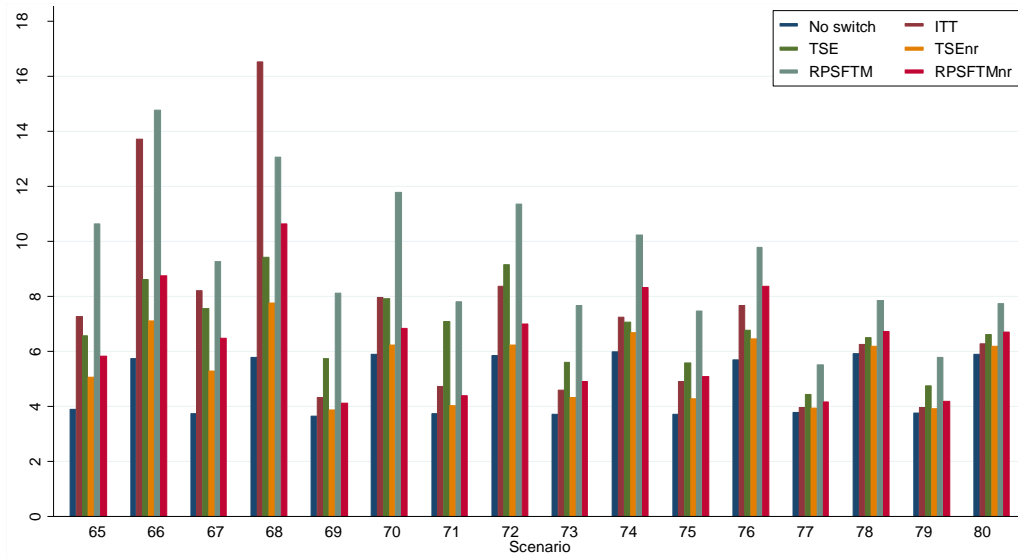
Figure G6: Root mean squared error, Scenarios 81-96
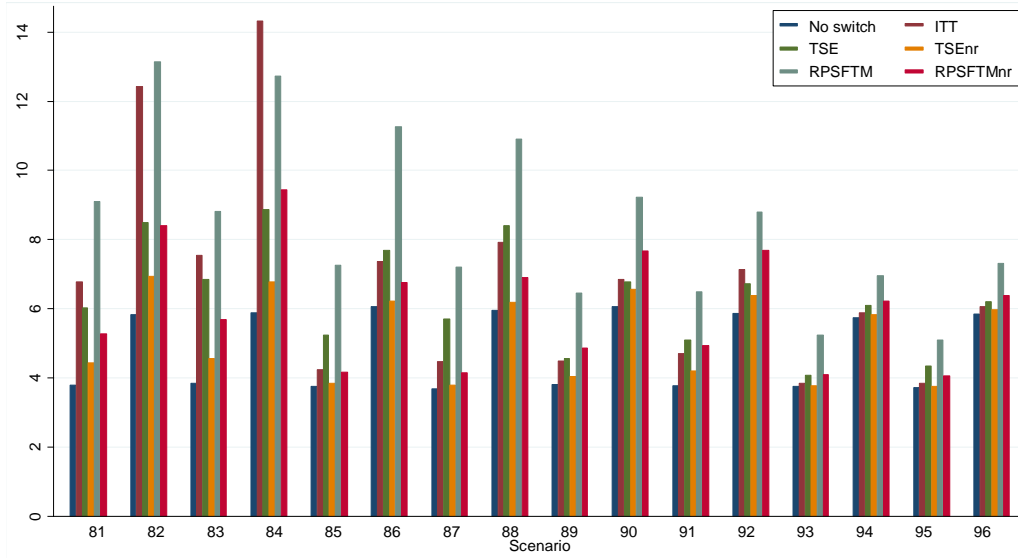
Appendix H: Further analysis of direction of bias associated with informative censoring

Our finding that the informative censoring bias associated with not re-censoring is likely to be associated with positive bias is important, and requires explanation. Our initial hypothesis was that despite our intentions, we had simulated switching in primarily poor prognosis patients even when we intended to simulate switching in primarily good prognosis patients. In the 32 scenarios in which there was a time-dependent treatment effect and a relatively low disease severity only approximately 69% of control group patients experienced disease progression and became eligible to switch – hence, even when switching was intended to be more likely in those with better prognosis, many of the best prognosis patients never became eligible to switch. Therefore, without re-censoring, positive informative censoring bias may be expected.

However, findings from scenarios in which there was a higher disease severity, and those in which there was not a time-dependent treatment effect, shed more light on this issue. In these scenarios, substantially more (approximately 96-98%) control group patients experienced disease progression and became eligible to switch treatments, and therefore there was a much clearer distinction in the prognosis of switchers and non-switchers. In some of these scenarios the directions of bias were more as might have been expected, For instance, in Scenarios 33, 34, 37, 38, 41, 42, 45 and 46 in which good prognosis patients were more likely to switch and the common treatment effect assumption did not hold, RPSFTMnr led to negative bias, and hence either the informative censoring led to negative bias (as would be expected), or the positive bias it created did not outweigh the negative bias associated with the violation of the common treatment effect assumption. However, in the scenarios within Scenarios 1-16 in which disease severity was high and the common treatment effect assumption was violated, the RPSFTMnr produced positive bias despite the clearer relationship between switch and prognosis and the negative bias expected due to violations of the common treatment effect assumption.

Closer inspection reveals that in these scenarios the RPSFTMnr did indeed over-estimate the treatment effect in switchers, but still resulted in positive bias in the estimate of mean survival – strongly indicating that the informative censoring led to positive bias. This finding was further supported in scenarios in which there was a constant treatment effect combined with a common treatment effect, in which we would expect to most clearly observe the impact of informative censoring because there are no competing biases (Scenarios 35, 36, 39, 40, 43, 44, 47, 48, 51, 52, 55, 56, 59, 60, 63 and 64) – the RPSFTMnr continued to produce positive bias in the vast majority of scenarios irrespective of the intended prognosis of switchers and despite the relatively clear relationship between prognosis and switch.

It is important to consider why non-re-censored analyses consistently increased estimates of control group mean survival compared to re-censored analyses. Non-re-censored analyses would be expected to lead to bias if participants who are censored at time t are not representative of all participants at risk in their randomised group at time t. Our analyses suggest that switchers who were not observed to die had worse prognosis than non-switchers at the censoring times that emerged after the estimation of counterfactual censoring times using RPSFTM and two-stage adjustment methods. This finding was apparent no matter whether patients with good or poor prognosis were more likely to switch treatments. This may appear surprising in scenarios where patients with good prognosis were more likely to switch. However, consider the situation when any non-switchers survive until the administrative censoring time-point – these patients demonstrably have good prognosis; they have long-term survival despite not receiving a beneficial treatment. Switchers, on the other hand, may have achieved long-term survival because they switched, not only because they had good prognosis. Even if switching was more likely to occur in patients with relatively good prognosis at the time of disease progression (when the switching decision was made), (not) re-censoring has its largest impact at much later time points at which point the prognostic balance between switchers and non-switchers may be substantially different due to the selection of patients that remain alive. At this point of the survival distribution, in situations where the experimental drug extends survival, it is plausible that non-switchers will generally have better prognosis than switchers.

White *et al.* (1999) suggest an empirical approach to demonstrate whether or not censoring is informative in the context of adjusting for treatment switching.[15] Informative censoring can be demonstrated if a variable can be found which predicts both censoring and event time. The authors suggest a plausible candidate to be a treatment history variable that indicates whether or not a patient has switched treatments. A series of time-dependent proportional hazards models can then be fitted, first with time-to-event as the outcome, and then with time-to-censoring as the outcome. Each model should be fitted on the original time scale, on the counterfactual time scale with re-censoring, and on the counterfactual time scale without re-censoring. According to White *et al.*, if the treatment history variable predicts events on all three time scales, but predicts censoring time only on the non-re-censored counterfactual time scale, this is consistent with censoring on the original scale being non-informative, and censoring on the counterfactual scale being informative without re-censoring and non-informative with re-censoring. We conducted this analysis on an example dataset from Scenario 2, in which good prognosis patients were more likely to switch treatments, and indeed found that treatment history predicted time-to-event, but was predictive of censoring time only on the non-re-censored counterfactual scale.

Interestingly, in our analysis and in the analysis presented by White *et al.*, the hazard ratio estimated for the treatment history variable for time-to-event was substantially greater than 1.00, suggesting that a history of having switched treatments increased the hazard of death. In the analysis of White *et al.* it was stated that switchers had worse prognosis than non-switchers. However, in Scenario 2 of the simulation study presented here, better prognosis patients were more likely to switch, and 96% of control group patients experienced disease progression and hence became eligible to switch. This may be because in our simulated scenarios switching could only happen after progression, which is a strong predictor of death. Hence the risk of death after switching was actually higher than the risk of death in non-switchers at the same time-point because some of the non-switchers will not yet have progressed (and despite the fact that the treatment was beneficial and better prognosis patients were more likely to switch). In addition, as discussed previously, at late time-points non-switching long-term survivors may have generally better prognosis than survivors at the same time-point who had switched treatment. An estimated HR>1.00 is consistent with the non-re-censored analysis over-estimating survival: adjusting for switching leads switchers to have earlier censoring times leading to underestimation of the longer term risk of death in the control group.

Our finding that the informative censoring associated with not re-censoring is likely to result in positive bias is important, and is likely to be relevant in any study in which switching only becomes possible after a specified disease-related time-point, where a proportion of patients do not reach this time-point, and in any study where there are long-term survivors who do not switch.

Appendix I: Estimation of hazard ratios

Estimates of the treatment effect in terms of a hazard ratio were typically prone to much higher levels of percentage bias than estimates of restricted mean survival, particularly for methods that incorporated re-censoring when there was a time-dependent treatment effect. For instance, percentage bias in the average hazard ratio estimate of approximately 5% was often associated with very low (less than 1%) percentage bias in the estimate of restricted mean survival. This was particularly notable, in scenarios with a strong time-dependent treatment effect. In these scenarios (Scenarios 65-96) TSE and RPSFTM methods that incorporated re-censoring often produced very high bias relating to estimates of the average hazard ratio (percentage bias usually in the region of 7-30%), but produced relatively low percentage bias with respect to restricted mean survival (often in the region of 1-4%). Methods that did not incorporate re-censoring were much less prone to substantial increases in percentage bias with respect to HR estimation compared to mean survival estimation when there was a time-dependent treatment effect – levels of percentage bias were high, but to a lesser extent.

Methods that incorporated IPCW were also particularly prone to producing high levels of percentage bias with respect to the HR in scenarios where there was a time-dependent treatment effect – this was most serious in scenarios with a high censoring proportion – in which the mean coefficient of variation of the weights was high. Whilst this had a detrimental effect on estimates of restricted mean survival, the impact was not huge, with percentage bias typically increasing by a few percentage points. However, for estimates of the mean HR, percentage bias increased substantially in these scenarios – from around -1 to -5%, to around -15 to -25%.

Occasionally, in scenarios for which a particular method produced low levels of bias, the direction of bias was different for the estimate of the HR compared to the estimate of restricted mean survival. However, in the majority of cases the patterns and directions of bias associated with the different adjustment methods were similar when considering estimates of the average hazard ratio and estimates of restricted mean survival, though the levels of percentage bias were substantially higher when considering hazard ratios.