Article:
Latimer, N.R. orcid.org/0000-0001-5304-5585, Abrams, K.R., Lambert, P.C. et al. (2 more authors) (2018) Assessing methods for dealing with treatment switching in clinical trials: A follow-up simulation study. Statistical Methods in Medical Research, 27 (3). pp. 765-784. ISSN 0962-2802

https://doi.org/10.1177/0962280216642264

Appendix A: Treatment switching probabilities

Table A1 presents the probability of switching for different patient groups at different time-points in the base scenario. 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.

Table A1: Probability of treatment switch by prognostic groups and consultation

Consultation 1		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.10	0.18	0.28
	1	0.25	0.40	0.54
	2	0.40	0.57	0.70
Consultation 2		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.08	0.15	0.24
	1	0.21	0.35	0.48
	2	0.35	0.52	0.65
Consultation 3		Biomarker group at progression		
		0	1	2
Time to progression group	0	0.05	0.10	0.16
	1	0.14	0.25	0.37
	2	0.25	0.40	0.54

In the base scenario the mean switching proportion in the control group across the 1,000 simulations was 43.60%, which was equivalent to 58.26% of control group patients who became at-risk of switching – i.e. those that experienced disease progression and had a 'choice' covariate value of '1'. This proportion of switching led to an increase in the average HR based on an ITT analysis from 0.51 to 0.60, reflecting the beneficial effect on survival of switching from the control group onto the experimental treatment. All probabilities in Table A1 were decreased (increased) when investigating lower (higher) switching scenarios.

Appendix B: Scenario parameter values

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

Table B1: Simulated scenarios – Parameter values and alternatives tested

Variable	Value (Scenario 1)	Alternative Values
Sample size	500 (2:1 randomisation)	300 (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	-
Choice covariate (probability of value of '1')	0.8	-
Multiplication of OS survival time due to bad prognosis group	Log hazard ratio = 0.5	-
Survival time distribution	Weibull parameters: Mix 1: Shape parameter 2.1 Scale parameter 1.8 Mix 2: Shape parameter 0.5 Scale parameter 0.1 p = 0.7 (mix parameter)	Weibull parameters to represent a less severe disease with more censoring: Mix 1: Shape parameter 2.1 Scale parameter 1.5 Mix 2: Shape parameter 0.5 Scale parameter 0.05 p = 0.25 (mix parameter) Gompertz parameters: Mix 1: Shape parameter -1.6 Scale parameter 0.15 Mix 2: Shape parameter 2.2 Scale parameter 0.5 p = 0.3 (mix parameter) Gompertz parameters to represent a less severe disease with more censoring: Mix 1: Shape parameter -1.6 Scale parameter 0.1 Mix 2: Shape parameter 2.2 Scale parameter 0.4 p = 0.75 (mix parameter)
Progression free survival	Overall survival time multiplied by a value from a beta distribution with shape parameters (10,10) – this implies the assumption that time to progression is approximately half 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 = -0.75	Alter log hazard ratio to -0.35 to represent a smaller treatment effect
Biomarker intercept	Calculated using a normal distribution with mean of 20 and standard deviation of 1. Increased by 10 in patients who are in the poor prognosis group.	-

Biomarker value progression over time	As demonstrated by Equation (4). $\beta_2 = -8$ to represent that the biomarker value increases more slowly in the experimental group, and $\beta_1 = 15$ to indicate that the biomarker value increases over time	-
Impact of biomarker value on overall survival	As demonstrated by Equations (6) and (7). Increased biomarker value increases the risk of death. The strength of this relationship depends on the variable α, which equals 0.02 in Scenario 1	-
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. The strength of this relationship depends on the variable α, which equals 0.02 in Scenario 1	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
Time-dependent portion of treatment effect, η	η =0.3 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
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 44% in Scenario 1	Test a low switching scenario where all probabilities are decreased – to an extent where approximately 20% of control group patients switch. Test a very high switching scenario where all probabilities are increased – to an extent where approximately 94% of "at-risk" 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	-
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: Overview of simulation scenarios

Table C1 presents key details associated with each of the scenarios simulated. Scenarios 1-16 are the base scenarios using a 2-component mixture Weibull baseline hazard function. Scenarios 17-32 replicate these but incorporate a reduced sample size. Scenarios 33-64 replicate Scenarios 1-32 using a 2-component mixture Gompertz baseline hazard function. Scenarios 65-68 are additional scenarios investigating the impact of extreme switching proportions.

The true area under the curve (restricted mean survival at 1.5 years) unconfounded by treatment switching is presented, along with the average treatment effect in terms of a hazard ratio (calculated using a Cox model) and an acceleration factor (calculated using a Weibull model). These were estimated by fitting Cox and Weibull models to scenario data generated for 1 000 000 patients without applying switching. This represents only an approximation of the true treatment effect as the proportional hazards assumption does not hold. In terms of a hazard ratio, the average treatment effect varied between 0.51 and 0.77.

The proportion of control group patients that switch, averaged across the 1000 simulations that made up each scenario, is also presented. The switching proportion varied between 5% and 70% of all control group patients. Scenarios 5-8, 13-16, 21-24, 29-32 and corresponding Gompertz-based scenarios (37-40, 45-48, 53-56 and 61-64) were designed to result in moderately low levels of switching, although these levels are probabilistic and are reliant on other characteristics. Scenarios 65-68 investigated very high switching proportions. Table C1 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, if their disease progressed before the end of the simulated follow-up, and if they had a 'choice' covariate value of '1'. 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 22% to 95%. This is particularly important to consider for observational-based approaches such as IPCW as these methods are reliant upon differentiating between the patient characteristics of switchers and non-switchers and applying inverse probability weightings based upon these characteristics. This can only be achieved by comparing the patients who were at risk of switching treatments and this will become increasingly difficult at the extremes – either when almost all patients switch, or when very few patients switch. The IPCW formulates a 'pseudo population' whose survival times are based upon those of uncensored patients (those who remain 'un-switched'), and thus if there are very few of these patients high weightings will be applied which could lead to

bias. We estimated the proportion of patients who become at risk of switching in each scenario by collecting data on the number of patients for whom disease progression was observed in each simulation. We then calculated the mean for this value across the scenario, and multiplied this by 0.8, representing the proportion of patients who had a 'choice' covariate value of '1'. This is approximate, but appropriately indicative for our purposes.

Table C1 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. In scenarios 9-16, 25-32, 41-48 and 57-64 the 'common treatment effect' assumption held. To provide further information on the strength of the time-dependent effect in each scenario we also include details on the treatment effect size received by switchers.

Table C1 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 13% and 56%.

Table C1: Overview of simulated scenarios

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Sample size	Data generating model	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	Restricted mean (Control group)	Restricted mean (Exp group)	HR	AF								
1	0.56	0.79	0.51	1.54	43.60%	58.26%	13.59%	500	Weibull	No	1.43	80%
2	0.99	1.20	0.52	1.78	30.03%	60.33%	55.80%	500	Weibull	No	1.63	80%
3	0.64	0.74	0.76	1.22	43.09%	61.20%	15.03%	500	Weibull	No	1.17	80%
4	0.99	1.08	0.77	1.25	30.52%	61.13%	46.60%	500	Weibull	No	1.20	80%
5	0.56	0.79	0.51	1.54	17.77%	23.78%	13.42%	500	Weibull	No	1.43	80%
6	0.99	1.20	0.52	1.78	12.88%	25.86%	55.21%	500	Weibull	No	1.63	80%
7	0.64	0.74	0.76	1.22	18.18%	25.87%	15.15%	500	Weibull	No	1.17	80%
8	0.99	1.08	0.77	1.25	13.24%	26.55%	46.58%	500	Weibull	No	1.20	80%
9	0.56	0.79	0.51	1.54	43.63%	58.27%	13.74%	500	Weibull	Yes	1.54	100%
10	0.99	1.20	0.52	1.78	30.04%	60.53%	56.35%	500	Weibull	Yes	1.78	100%
11	0.64	0.74	0.76	1.22	42.86%	60.90%	15.09%	500	Weibull	Yes	1.22	100%
12	0.99	1.08	0.77	1.25	30.66%	61.63%	46.82%	500	Weibull	Yes	1.25	100%
13	0.56	0.79	0.51	1.54	17.78%	23.76%	13.52%	500	Weibull	Yes	1.54	100%
14	0.99	1.20	0.52	1.78	12.86%	25.86%	55.53%	500	Weibull	Yes	1.78	100%
15	0.64	0.74	0.76	1.22	18.00%	25.61%	15.04%	500	Weibull	Yes	1.22	100%
16	0.99	1.08	0.77	1.25	13.14%	26.44%	46.70%	500	Weibull	Yes	1.25	100%
17	0.56	0.79	0.51	1.54	43.65%	58.34%	13.48%	300	Weibull	No	1.43	80%
18	0.99	1.20	0.52	1.78	29.90%	60.27%	55.81%	300	Weibull	No	1.63	80%
19	0.64	0.74	0.76	1.22	43.08%	61.18%	15.00%	300	Weibull	No	1.17	80%
20	0.99	1.08	0.77	1.25	30.35%	60.91%	46.69%	300	Weibull	No	1.20	80%
21	0.56	0.79	0.51	1.54	17.82%	23.81%	13.45%	300	Weibull	No	1.43	80%
22	0.99	1.20	0.52	1.78	12.95%	26.03%	55.34%	300	Weibull	No	1.63	80%
23	0.64	0.74	0.76	1.22	18.22%	25.90%	15.05%	300	Weibull	No	1.17	80%
24	0.99	1.08	0.77	1.25	13.07%	26.43%	46.74%	300	Weibull	No	1.20	80%
25	0.56	0.79	0.51	1.54	43.58%	58.25%	13.64%	300	Weibull	Yes	1.54	100%

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Sample size	Data generating model	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	Restricted mean (Control group)	Restricted mean (Exp group)	HR	AF								
26	0.99	1.20	0.52	1.78	30.34%	60.85%	56.14%	300	Weibull	Yes	1.78	100%
27	0.64	0.74	0.76	1.22	42.97%	61.12%	15.11%	300	Weibull	Yes	1.22	100%
28	0.99	1.08	0.77	1.25	30.54%	61.54%	46.85%	300	Weibull	Yes	1.25	100%
29	0.56	0.79	0.51	1.54	17.97%	23.96%	13.46%	300	Weibull	Yes	1.54	100%
30	0.99	1.20	0.52	1.78	13.02%	26.12%	55.58%	300	Weibull	Yes	1.78	100%
31	0.64	0.74	0.76	1.22	18.07%	25.71%	15.17%	300	Weibull	Yes	1.22	100%
32	0.99	1.08	0.77	1.25	13.05%	26.36%	46.86%	300	Weibull	Yes	1.25	100%
33	0.54	0.78	0.51	1.60	40.89%	55.89%	13.82%	500	Gompertz	No	1.48	80%
34	0.99	1.19	0.52	1.77	33.77%	54.62%	55.23%	500	Gompertz	No	1.62	80%
35	0.63	0.74	0.76	1.24	42.71%	59.91%	15.66%	500	Gompertz	No	1.19	80%
36	0.99	1.08	0.77	1.25	36.23%	58.54%	46.17%	500	Gompertz	No	1.20	80%
37	0.54	0.78	0.51	1.60	16.78%	22.91%	13.59%	500	Gompertz	No	1.48	80%
38	0.99	1.19	0.52	1.77	13.64%	22.04%	54.59%	500	Gompertz	No	1.62	80%
39	0.63	0.74	0.76	1.24	18.00%	25.26%	15.77%	500	Gompertz	No	1.19	80%
40	0.99	1.08	0.77	1.25	15.28%	24.70%	46.11%	500	Gompertz	No	1.20	80%
41	0.54	0.78	0.51	1.60	40.82%	55.77%	13.89%	500	Gompertz	Yes	1.60	100%
42	0.99	1.19	0.52	1.77	33.77%	54.59%	55.61%	500	Gompertz	Yes	1.77	100%
43	0.63	0.74	0.76	1.24	42.52%	59.78%	15.78%	500	Gompertz	Yes	1.24	100%
44	0.99	1.08	0.77	1.25	36.24%	58.57%	46.26%	500	Gompertz	Yes	1.25	100%
45	0.54	0.78	0.51	1.60	16.64%	22.74%	13.65%	500	Gompertz	Yes	1.60	100%
46	0.99	1.19	0.52	1.77	13.59%	21.94%	54.84%	500	Gompertz	Yes	1.77	100%
47	0.63	0.74	0.76	1.24	17.98%	25.21%	15.64%	500	Gompertz	Yes	1.24	100%
48	0.99	1.08	0.77	1.25	15.24%	24.66%	46.13%	500	Gompertz	Yes	1.25	100%
49	0.54	0.78	0.51	1.60	40.30%	55.16%	13.67%	300	Gompertz	No	1.48	80%
50	0.99	1.19	0.52	1.77	33.99%	54.89%	55.21%	300	Gompertz	No	1.62	80%
51	0.63	0.74	0.76	1.24	42.60%	59.79%	15.85%	300	Gompertz	No	1.19	80%

Scenario	Truth (years)		Average treatment effects		Mean switcher % of total	Mean switcher % of at risk	Mean censoring proportion (%)	Sample size	Data generating model	Common treatment effect?	Treatment effect in switchers (AF)	% of exp group treatment effect
	Restricted mean (Control group)	Restricted mean (Exp group)	HR	AF								
52	0.99	1.08	0.77	1.25	36.45%	58.84%	46.15%	300	Gompertz	No	1.20	80%
53	0.54	0.78	0.51	1.60	16.42%	22.46%	13.64%	300	Gompertz	No	1.48	80%
54	0.99	1.19	0.52	1.77	13.71%	22.13%	54.58%	300	Gompertz	No	1.62	80%
55	0.63	0.74	0.76	1.24	18.01%	25.27%	15.54%	300	Gompertz	No	1.19	80%
56	0.99	1.08	0.77	1.25	15.24%	24.65%	46.10%	300	Gompertz	No	1.20	80%
57	0.54	0.78	0.51	1.60	40.79%	55.73%	13.94%	300	Gompertz	Yes	1.60	100%
58	0.99	1.19	0.52	1.77	33.86%	54.61%	55.51%	300	Gompertz	Yes	1.77	100%
59	0.63	0.74	0.76	1.24	42.54%	59.71%	15.70%	300	Gompertz	Yes	1.24	100%
60	0.99	1.08	0.77	1.25	36.23%	58.52%	46.31%	300	Gompertz	Yes	1.25	100%
61	0.54	0.78	0.51	1.60	16.63%	22.70%	13.59%	300	Gompertz	Yes	1.60	100%
62	0.99	1.19	0.52	1.77	13.72%	22.19%	54.81%	300	Gompertz	Yes	1.77	100%
63	0.63	0.74	0.76	1.24	17.81%	25.05%	15.65%	300	Gompertz	Yes	1.24	100%
64	0.99	1.08	0.77	1.25	15.27%	24.68%	46.06%	300	Gompertz	Yes	1.25	100%
65	0.56	0.79	0.51	1.54	70.20%	93.80%	13.67%	500	Weibull	No	1.43	80%
66	0.99	1.20	0.52	1.78	47.22%	94.45%	56.06%	500	Weibull	No	1.63	80%
67	0.64	0.74	0.76	1.22	66.39%	94.32%	15.15%	500	Weibull	No	1.17	80%
68	0.99	1.08	0.77	1.25	47.11%	94.91%	46.73%	500	Weibull	No	1.20	80%

Appendix D: Percentage bias figures

Figures showing bias across scenarios are presented throughout this Appendix – care should be taken when comparing these because the y-axes use different scales.

Figure D1: Percentage bias (%) across scenarios – ITT

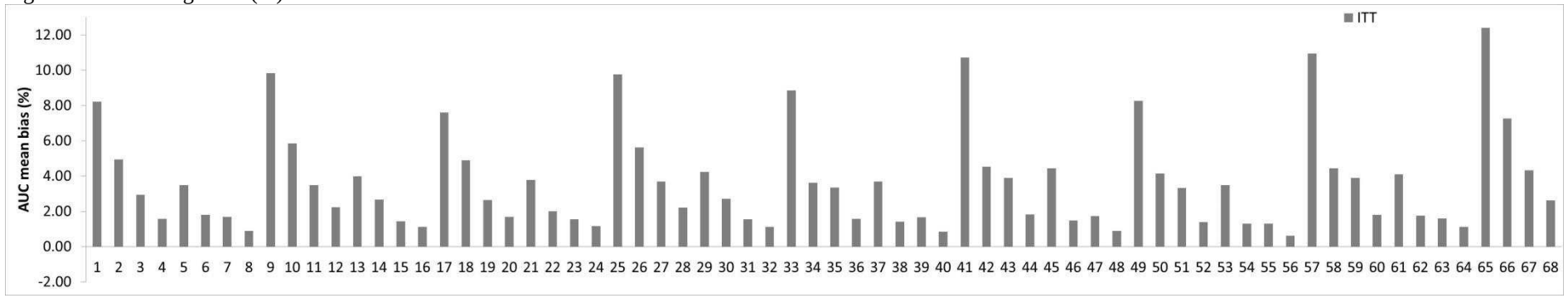


Figure D2: Percentage bias (%) across scenarios – Exclusion and censoring approaches

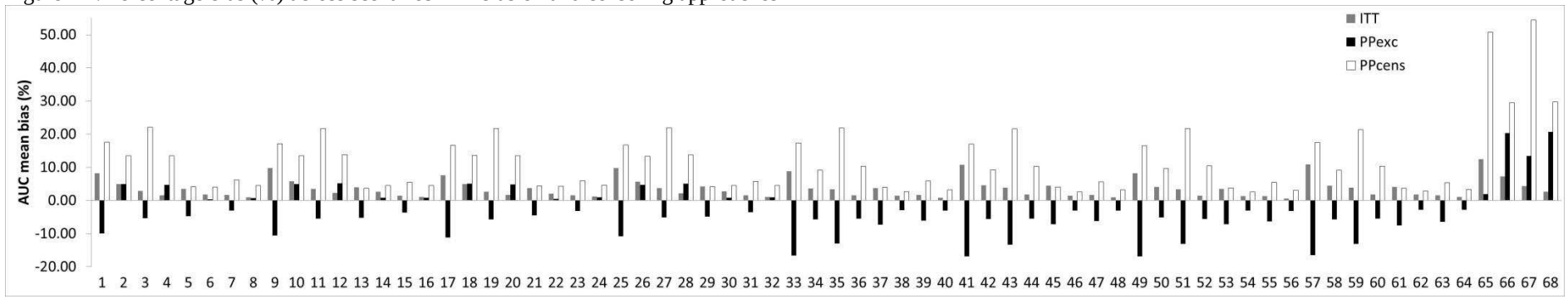


Figure D3: Percentage bias (%) across scenarios – IPCW and IPCWn approaches

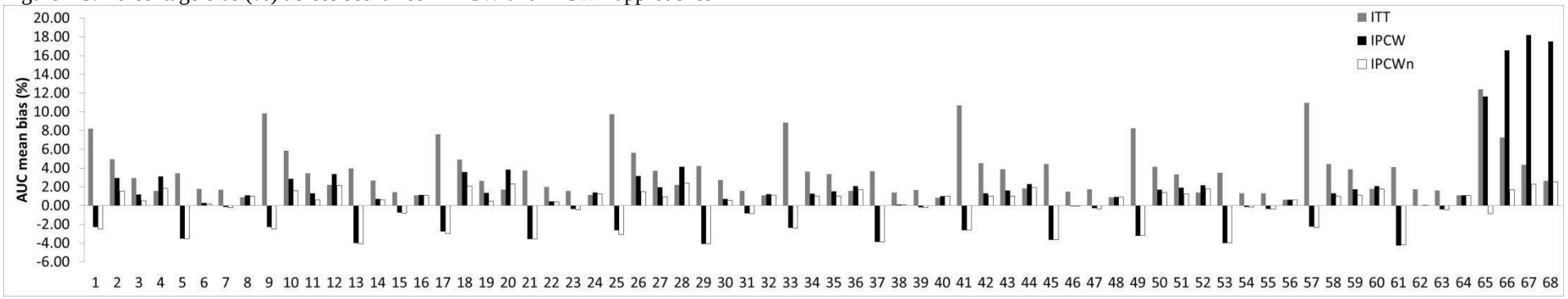


Figure D4: Percentage bias (%) across scenarios – RPSFTM and IPE approaches

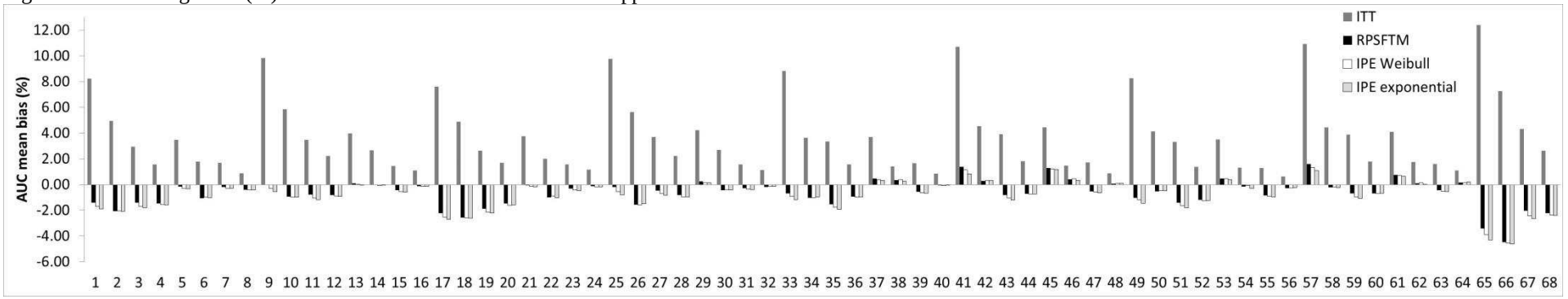
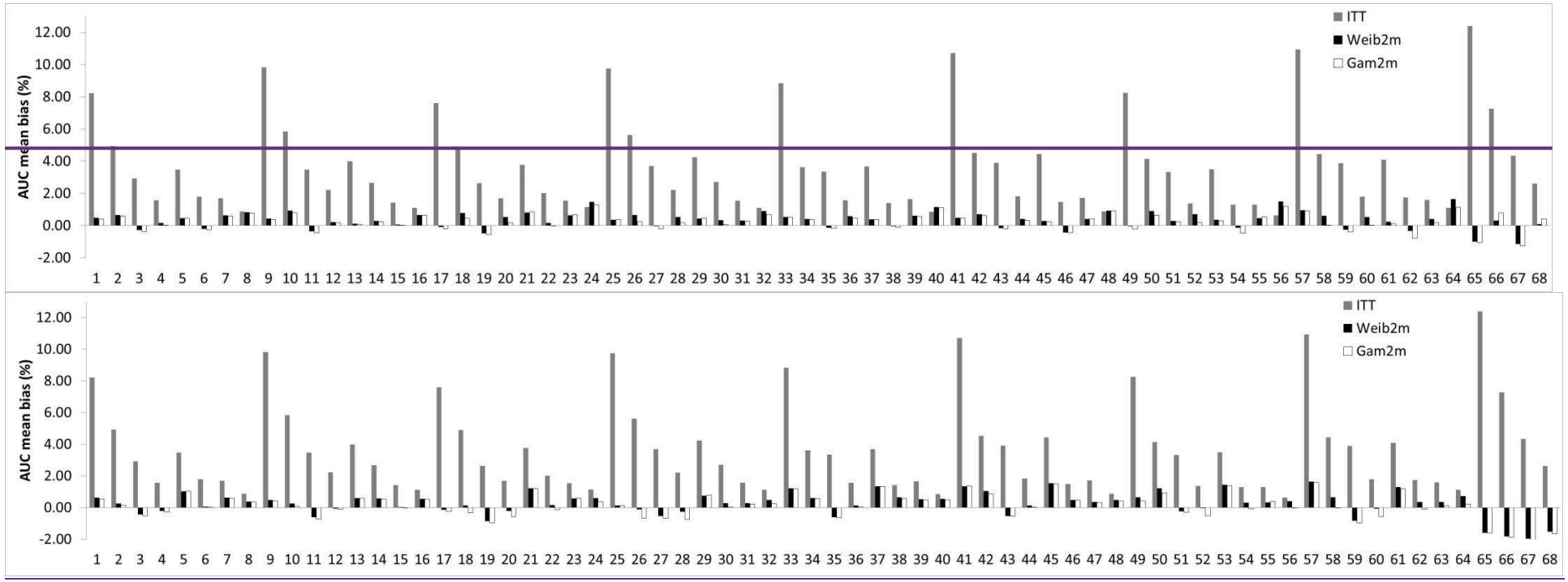


Figure D5: Percentage bias (%) across scenarios – Two-stage approaches


ITT
Weib2m
Gam2m
AUC mean bias (%)
12.00
10.00
8.00
6.00
4.00
2.00
0.00
-2.00
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68

Appendix E: IPCW Convergence

In our simulation study we examined the relative performance of the IPCW adjustment methods according to the convergence of the logistic weighting regressions (for the numerator and denominator of the stabilised weight). Stata provides information on three relevant indicators of the performance of the regression:

a. Whether or not the regression converged
b. The number of completely determined successes
c. The number of completely determined failures

Convergence is clearly an issue, and if any successes or failures are completely determined this is a sign of potential hidden collinearity. Across Scenarios 1-32, convergence of the IPCW method occurred in 59.9% of simulations, and convergence of the IPCWn method occurred in 63.8% of simulations. Convergence combined with zero completely determined successes or failures occurred in just 21.0% of simulations for the IPCW method, and 25.1% of simulations for the IPCWn method. Convergence was lower in simulations in which relatively lower proportions of patients switched treatments, and was particularly low (sometimes as low as 4-7% for convergence combined with zero completely determined successes or failures) in scenarios with lower switching proportions combined with a simulated sample size of 300 patients. The lowest level of convergence (irrespective of whether any successes or failures were completely determined) was 33.2%, in Scenario 21. Convergence with no successes or failures completely determined was achieved in 14.4% and 17.5% of simulations for the IPCW and IPCWn methods respectively in Scenarios 17-32 (with lower sample size), compared to 27.6% and 32.6% of simulations in Scenarios 1-16 (with higher sample size).

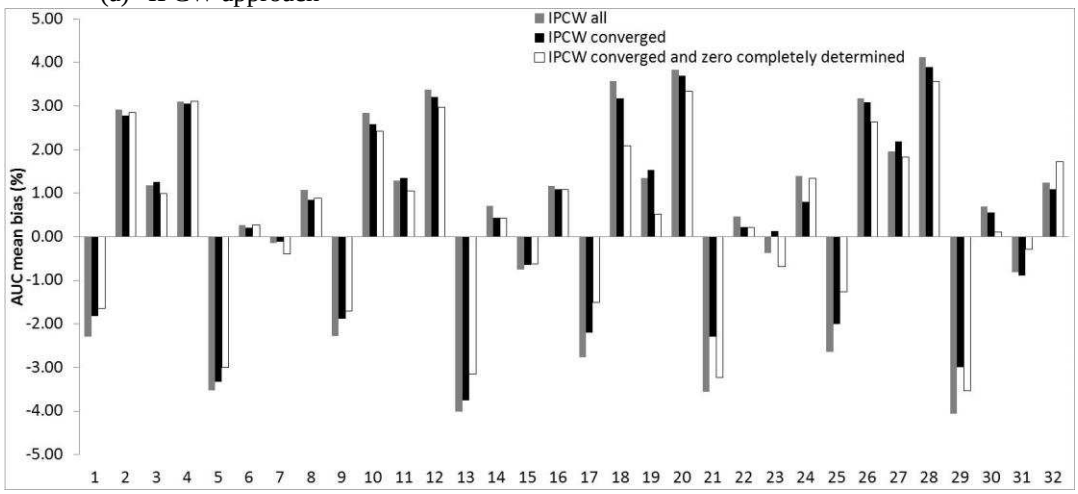
In the results presented in the main report we included all simulations for the IPCW method, since Stata provides coefficient estimates even if regressions fail to converge. However, convergence and possible collinearity are clear problems associated with the IPCW method, and therefore in practice the application of the IPCW method will need to be considered carefully on a case-by-case basis, and models may need to be adapted in order to achieve convergence. Given the high proportions of simulations in which convergence was not achieved, comparisons of the results of the IPCW method according to the extent to which convergence was achieved is problematic. However, we found that both IPCW and IPCWn methods only produced marginally lower levels of bias in instances where full convergence was achieved, compared to instances where full convergence was not achieved. This is demonstrated in Figure E1, which presents percentage bias for the IPCW

and IPCWn analyses across Scenarios 1-32, comparing instances where the analyses converged and instances where they did not.

Stata has strict convergence criteria, and this may explain why the IPCW methods appear to have produced reasonable results even when one or more of the logistic regressions did not converge. In addition, we anticipate that the convergence problems may have resulted from the use of splines within the logistic weighting regressions. To create these splines we used the spbase Stata program, as recommended by Fewell et al,[41] with 5 knots placed according to percentiles of the survival time distribution. However, we believe that generating knots based upon the event time distribution may allow convergence issues to be avoided. As a lack of convergence does not seem to be of key importance in our simulations we do not anticipate that this has had an important impact upon our results, but from a practical perspective models may need to be adapted to achieve convergence.

Figure E1: Percentage bias (%) by IPCW convergence status – Scenarios 1-32

(a) IPCW approach



(b) IPCWn approach