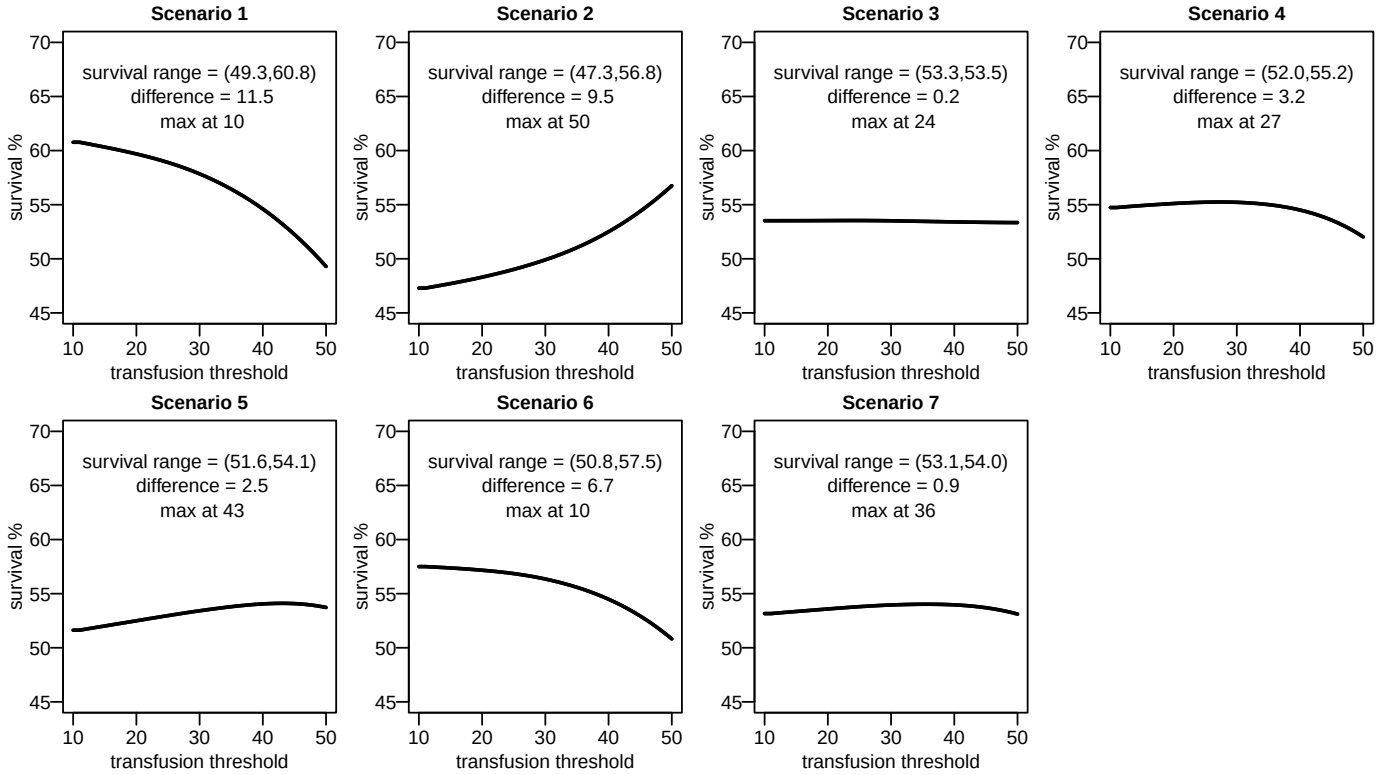
Simulations for T4P

In this appendix we describe and report the results of the latest simulation exercise that informs our choice of statistical model and the adaptive design for T4P. These results support the use of a fractional polynomial model (FP) and the incorporation of elements of adaption. The simulated trials were implemented in a fully Bayesian framework, with analysis via Hamiltonian MCMC implemented via Stan (https://mc-stan.org/). Minimally informative priors were used for all unknown model parameters.

The scenarios for generating the simulated data

As the precise shape of the threshold-survival curve is unknown to us at the design stage of the trial, we generated realisations of T4P from a range of scenarios, with varying optima and differences in mortality across the range of thresholds from <10 to $<50\times10^9/L$. We used test scenarios which have been created by considering the relationships underpinning the threshold-survival curves, and utilising relevant existing data in addition to expert knowledge. These are shown in Figure 1, and include steep and very shallow curves, monotonic increasing and decreasing shapes, and curves with an internal maximum. The survival range, difference between the maximum and minimum survival percentages, and the optimum threshold (i.e. the threshold with the maximum survival percentage) are included on each plot.

Figure 1: Scenarios based on existing data and expert clinical opinion



To build these scenarios, ranges of plausible values were agreed by the trial clinicians for four quantities at platelet counts of 10, 20, 30, 40 and $50\times10^9/L$ at the time of a patient's first procedure: (1) absolute mortality caused by transfusion (harms), (2) absolute mortality prevented by transfusion (benefits), (3) overall survival percentage (patients who do and do not recieve a transfusion), and

(4) the proportion of patients transfused in current practice. The first two were informed by clinical experience and recent trial results, the third also took account of Case Mix Programme data and the fourth drew on data from our survey of UK ICU doctors (Shah *et al.*, 2020). Using different profiles of these quantities from within the ranges of values considered plausible, we derived separate survival curves by platelet count for (1) patients who receive a transfusion and (2) patients who do not receive a transfusion. Lastly, we calculated a threshold-survival curve from these two survival-platelet count curves and the proportion of patients at each platelet count at the time of their first procedure (taken from data).

Model comparison

For each of these 7 scenarios, we simulated 200 realisations of the T4P trial using 5 different models:

FP: fractional polynomial with model averaging

GP1: Gaussian process model with a squared exponential kernel

GP2: Gaussian process model with a squared exponential $\times$ linear kernel

RCS3: restricted cubic spline model with 3 knots (15,30,45)

RCS4: restricted cubic spline model with 4 knots (15,25,35,45).

The fractional polynomial (FP) model incorporates model uncertainty via Bayesian Model Averaging (BMA). For the BMA we calculated predictions from 36 possible fractional polynomial models of order 2, and combined these using posterior model probabilities calculated using bridge sampling (Gronau *et al.*, 2017).

Gaussian processes provide a Bayesian non-parametric alternative for defining flexible regression models (Neal, 1999). We investigated two options: GP1 which has a squared exponential kernel and GP2 with a squared exponential $\times$ linear kernel.

The restricted cubic spline (RCS) models are piecewise polynomials with degree 3 defined over adjacent intervals, where the join between two intervals is called a 'knot'. They have continuity and slope constraints at each knot, and linear constraints before the first knot and after last knot. Here we present the results from two RCS models: RCS3 which has 3 knots (fixed at 15, 30 and 45) and RCS4 with 4 knots (fixed at 15, 25, 35 and 45).

Table 1 shows four performance measures:

1. the percentage of times each threshold (in units of 5) is recommended
2. the probability of recommending a threshold within 2% of the true maximum
3. the mean width of the 95% highest probability density interval for the recommended threshold (across realisations)
4. the probability that the 95% highest probability density interval excludes the true maximum.

The 95% highest probability density interval (HDI) is the narrowest interval which contains 95% of the posterior density, such that all points within the interval have a higher probability density than points outside the interval.

The relative performance of the five models is to some extent scenario dependent. The FP model outperforms the others for scenarios where the optimum is at one of the boundaries of the threshold range under consideration, and the GP models do slightly better with the non-montonic shapes. Taken overall, we consider the FP model to be the best option. Hence we will use a FP model for interim analyses and adaption, and for the primary analysis at the end of the trial. As the

Table 1: Operating characteristics for Scenarios 1-7 based on 200 realisations

	Percentage recommended[1]									Power	95% HDI[⋆]	True max
	10	15	20	25	30	35	40	45	50	within 2%[2]	width[3]	excluded[4]
S1: FP	51.0	31.5	17.5	0.0	0.0	0.0	0.0	0.0	0.0	100	11.4	2
S1: GP1	45.0	25.0	21.0	8.5	0.5	0.0	0.0	0.0	0.0	98	17.9	0
S1: GP2	51.5	29.5	14.5	4.5	0.0	0.0	0.0	0.0	0.0	100	17.3	0
S1: RCS3	66.5	1.0	12.0	20.5	0.0	0.0	0.0	0.0	0.0	95	13.8	1
S1: RCS4	64.0	1.0	15.5	16.5	2.5	0.5	0.0	0.0	0.0	96	17.3	0
S2: FP	0.0	0.0	0.0	0.0	0.0	0.5	0.5	0.0	99.0	99	7.0	0
S2: GP1	0.0	0.0	0.0	0.0	0.0	1.0	2.0	10.5	86.5	93	14.7	0
S2: GP2	0.0	0.0	0.0	0.0	0.0	0.0	1.0	10.0	89.0	94	13.8	0
S2: RCS3	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	98.5	98	6.9	0
S2: RCS4	0.0	0.0	0.0	0.0	0.0	1.0	6.5	2.5	90.0	90	10.7	0
S3: FP	19.0	8.5	15.0	11.5	15.5	6.0	5.0	0.5	19.0	100	36.1	6
S3: GP1	3.0	10.0	9.0	15.5	22.5	21.0	12.0	4.5	2.5	100	38.5	2
S3: GP2	3.5	13.0	9.5	15.5	15.5	14.5	16.0	9.0	3.5	100	38.7	2
S3: RCS3	21.5	0.0	2.0	8.5	34.0	13.5	2.5	0.0	18.0	100	34.4	14
S3: RCS4	13.5	0.0	6.5	15.0	7.5	36.0	4.0	0.5	17.0	100	33.9	10
S4: FP	18.5	17.5	21.5	22.5	9.0	4.0	2.0	1.0	4.0	96	29.8	9
S4: GP1	5.5	10.0	15.0	26.0	25.0	12.0	3.5	2.0	1.0	98	35.1	0
S4: GP2	9.0	16.5	15.5	22.0	17.5	10.5	5.0	2.5	1.5	98	36.2	0
S4: RCS3	25.5	0.5	3.5	20.0	42.0	3.5	0.0	0.0	5.0	95	28.1	6
S4: RCS4	12.5	0.5	12.5	26.0	17.5	26.0	0.5	0.0	4.5	96	29.7	4
S5: FP	4.5	4.5	7.0	9.0	9.0	9.0	8.0	6.0	43.0	94	32.3	9
S5: GP1	0.5	1.0	2.5	12.5	13.0	17.5	26.0	14.5	12.5	99	35.6	2
S5: GP2	1.0	1.0	7.0	8.0	10.0	11.5	24.0	19.5	18.0	98	36.0	1
S5: RCS3	4.0	0.5	1.0	5.0	24.5	19.5	4.5	1.0	40.0	96	29.6	8
S5: RCS4	2.5	0.0	0.5	11.5	5.0	38.5	12.5	0.0	29.5	98	29.9	10
S6: FP	35.5	27.5	25.5	8.5	2.0	0.0	0.0	0.0	1.0	99	18.3	5
S6: GP1	23.0	19.0	27.5	18.5	9.5	1.5	0.5	0.5	0.0	99	26.7	2
S6: GP2	29.5	26.5	22.5	13.0	6.0	1.5	0.0	1.0	0.0	99	27.4	2
S6: RCS3	47.0	2.0	8.0	31.5	10.5	0.0	0.0	0.0	1.0	99	19.3	3
S6: RCS4	36.5	0.5	21.0	24.5	9.5	7.0	1.0	0.0	0.0	98	23.0	6
S7: FP	8.0	11.5	18.5	16.0	15.0	4.0	4.5	2.5	20.0	100	35.5	12
S7: GP1	1.0	2.5	16.5	15.0	23.0	21.0	11.0	8.0	2.0	100	38.2	2
S7: GP2	1.5	8.0	13.5	13.0	18.0	19.0	10.5	11.0	5.5	100	38.4	2
S7: RCS3	13.0	1.0	2.5	9.0	43.0	9.0	2.5	0.5	19.5	100	33.6	14
S7: RCS4	5.0	0.0	9.0	15.5	12.0	34.5	5.0	0.5	18.5	100	33.4	6

[⋆] HDI = highest probability density interval; FP = fractional polynomial; GP = Gaussian process; RCS = restricted cubic spline

[1] percentage of times each threshold (in units of 5) is recommended

[2] probability of recommending a threshold within 2% of the true maximum

[3] the mean width of the 95% HDI for the recommended threshold (across realisations)

[4] probability that the 95% HDI excludes the true maximum

GP models favour curves with an internal maximum more than the FP model, we will carry out a sensitivity analysis using a GP model option. The details will be fully specified in the statistical analysis plan (SAP).

Adaption

In the simulation work in preparation for this trial, we investigated two types of adaption (drop thresholds and adapt allocation ratio) and compared with equal allocation to all 5 thresholds throughout the trial. Our aim was to improve the operating characteristics of the T4P trial. We considered adapting the allocation ratio with and without floors, and compared the use of 2, 3 and 4 adaption points.

Based on simulation results, our preferred design uses equal allocation for the first 40% of patients, and has 3 adaption points each of which applies to 20% of patients. The allocation ratio is modified based on the probabilities each of the five thresholds has the maximum probability of survival, and is based on 780 (32.5%), 1260 (52.5%) and 1740 (72.5%) randomised patients (which allows for the primary outcome being 90 days post randomisation). Thresholds 10 and 50 are dropped if their probability of having the maximum probability of survival <1%, but the other thresholds have a minimum ratio set to 5%.

Table 2 shows that using the fractional polynomial model described above, the operating characteristics of our preferred design (RAR) are similar to using equal randomisation throughout (ER). Table 3 displays additional information: the mean number of patients randomised to each threshold and overall mean survival percentage, averaged across 200 realisations. As expected, for the more steeply curved scenarios, RAR has the advantage of randomising more patients to the thresholds with higher survival percentages resulting in a higher overall mean survival percentage — amounting to an increase of 2% for Scenarios 1 and 2.

Presentation

Figure 2 provides two examples of the presentation of the results of the trial, using realisation 1 of scenario 1 and realisation 96 of scenario 6. In the left plot: the green line shows the posterior mean estimate of the threshold-survival curve; the grey shading shows the uncertainty in 10% bands, i.e the lightest grey shading shows the 80% interval between the 0.1 and 0.9 percentiles, the next band is the 60% interval between the 0.2 and 0.8 percentiles, and so on; the blue crosses represent the simulated trial data; and the dashed red line indicates the true threshold-survival curve (available for simulations only). In the right plot: the grey histogram represents the posterior distribution of the optimum threshold. The estimated optimum is the posterior median (black cross in the left plot) and the CrI/95% probability interval is the 95% HDI (black line above threshold-survival curve in the left plot).

References

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., and Steingroever, H. (2017). A tutorial on bridge sampling. *Journal of Mathematical Psychology*, **81**, 80–97.

Table 2: Comparison of ER and RAR using fractional polynomial models based on 200 realisations

	Percentage recommended[1]									Power within 2%[2]	95% HDI* width[3]	True max excluded[4]
	10	15	20	25	30	35	40	45	50			
S1: ER	51.0	31.5	17.5	0.0	0.0	0.0	0.0	0.0	0.0	100	11.4	2
S1: RAR	53.5	34.0	10.5	2.0	0.0	0.0	0.0	0.0	0.0	99	11.7	4
S2: ER	0.0	0.0	0.0	0.0	0.0	0.5	0.5	0.0	99.0	99	7.0	0
S2: RAR	0.0	0.0	0.0	0.5	0.5	0.5	0.0	0.0	98.5	98	6.0	0
S3: ER	19.0	8.5	15.0	11.5	15.5	6.0	5.0	0.5	19.0	100	36.1	6
S3: RAR	18.0	11.0	11.5	13.0	11.5	7.5	6.0	3.5	18.0	100	36.0	2
S4: ER	18.5	17.5	21.5	22.5	9.0	4.0	2.0	1.0	4.0	96	29.8	9
S4: RAR	16.0	18.5	23.0	20.5	11.0	5.5	1.5	0.5	3.5	96	31.3	8
S5: ER	4.5	4.5	7.0	9.0	9.0	9.0	8.0	6.0	43.0	94	32.3	9
S5: RAR	4.5	4.5	9.5	11.0	13.0	14.0	8.0	5.5	30.0	92	34.9	6
S6: ER	35.5	27.5	25.5	8.5	2.0	0.0	0.0	0.0	1.0	99	18.3	5
S6: RAR	38.5	27.5	23.0	9.5	0.5	0.5	0.0	0.0	0.5	99	19.9	4
S7: ER	8.0	11.5	18.5	16.0	15.0	4.0	4.5	2.5	20.0	100	35.5	12
S7: RAR	11.5	6.5	19.5	17.5	13.5	8.5	5.0	3.5	14.5	100	35.5	10

* HDI = highest probability density interval; ER = equal randomisation; RAR = response adaptive randomisation

[1] percentage of times each threshold (in units of 5) is recommended

[2] probability of recommending a threshold within 2% of the true maximum

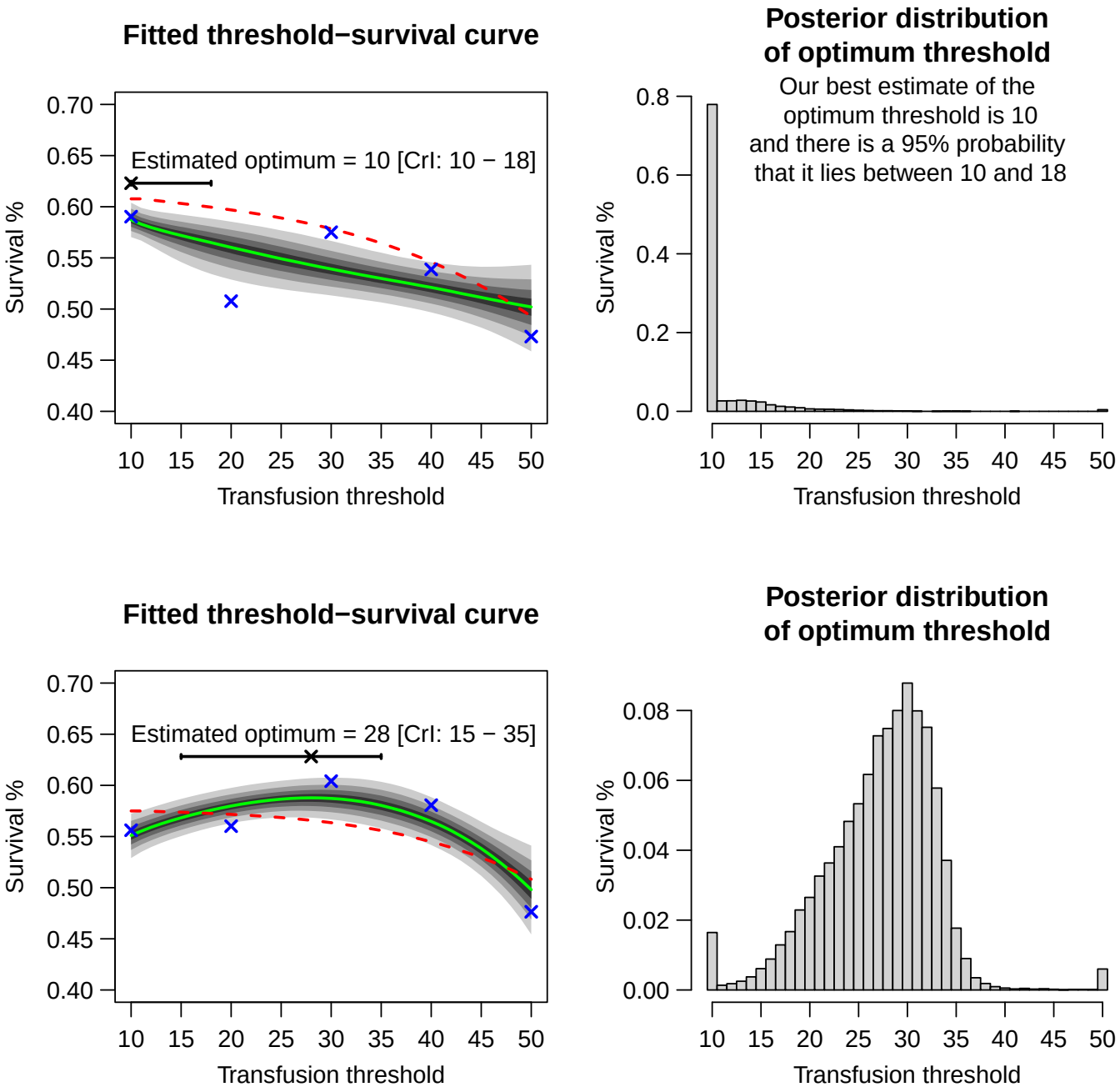
[3] the mean width of the 95% HDI for the recommended threshold (across realisations)

[4] probability that the 95% HDI excludes the true maximum

Table 3: Mean patients randomised across 200 realisations for Scenarios 1-7

	Threshold					Mean Survival %
	10	20	30	40	50	
S1: ER	479	483	480	478	480	56.3
S1: RAR	926	671	317	277	209	58.5
S2: ER	477	482	483	478	479	51.1
S2: RAR	235	299	332	400	1133	53.2
S3: ER	478	484	481	478	480	53.5
S3: RAR	519	521	456	429	475	53.5
S4: ER	481	480	482	477	479	54.3
S4: RAR	541	601	500	389	369	54.5
S5: ER	481	479	480	479	481	52.9
S5: RAR	371	445	482	474	629	53.1
S6: ER	479	482	481	479	479	55.3
S6: RAR	768	680	388	313	252	56.3
S7: ER	483	477	480	480	479	53.6
S7: RAR	480	511	475	448	486	53.7

Figure 2: Examples of presentation of trial results using the adaptive design with a fractional polynomial model (Top: scenario 1, realisation 1; Bottom: scenario 6, realisation 96)



Green line = posterior median of the threshold-survival curve; grey shading = uncertainty in 10% bands; blue crosses = simulated trial data and the red dashed line = true threshold-survival curve [available for simulations only]. The estimated optimum is the posterior median [black cross] and the CrI is the 95% HDI [black line above threshold-survival curve].

Neal, R. M. (1999). Regression and classification using gaussian process priors. *Bayesian Statistics 6*.

Shah, A., Gould, D. W., Doidge, J., Mouncey, P. R., Harrison, D. A., Young, J. D., Stanworth, S. J., and Watkinson, P. J. (2020). A UK national survey of prophylactic platelet transfusion thresholds in non-bleeding, critically ill adults. *Transfusion Medicine*, **30**, 515—-517.