Email contact: yevgen.ryeznik@math.uu.se

Treatment allocation adaptive randomization methods in clinical trials with few individuals may influence model parameter estimation.

Yevgen Ryeznik[1,2], Oleksandr Sverdlov[3] and Andrew C. Hooker[2]

UPPSALA UNIVERSITET

[1]Department of Mathematics, Uppsala University, [2]Department of Pharmaceutical Biosciences, Uppsala University, [3]Novartis Pharmaceutical Corporation.

Background and Objectives

In dose-response studies with censored time-to-event outcomes, D-optimal designs depend on the true model parameters and the number of censored outcomes. In order to implement such a design in practice, an adaptive design that incorporates updated knowledge about the dose-response curve at interim analyses can be used [1]. Further, treatment allocation should involve randomization, which is essential to mitigate various experimental biases and perform valid statistical inference at the end of the trial. Here, we compare several randomization procedures and their impact on model estimation.

Methods and Materials

Model: An accelerated failure time model [2] assuming a quadratic dose-response shape for log-transformed time-to-event outcomes with a Weibull distribution that are subject to right censoring with a fixed censoring time is assumed:

$$\log T = \beta_0 + \beta_1 x + \beta_2 x^2 + b\varepsilon,$$

$t = \min(T, \tau)$ — observed time,

where τ is censoring time, x is dose (without loss of generality assume that $x \sim [0,1]$), ε has a p.d.f. $f(w) = \exp(w - \exp(w))$ with $\mathbf{E}(\varepsilon) = -\gamma$ and $\mathbf{Var}(\varepsilon) = \pi^2/6$

Designs: For the above model, the *D*-optimal design for the most precise estimation of the dose-response curve (assuming no censoring) is supported at doses of 0, ½, and 1, with equal allocation proportions. If the outcomes are censored, then the design still has 3 support points, but the design depends on model parameters $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, b)$ [1], and may be found numerically with a first order (exchange) algorithm [3]. In practice, $\boldsymbol{\theta}$ is unknown, and a two-stage adaptive design can be used. At Stage 1, a cohort of $n^{(1)}$ patients is randomized according to the uniform design. Then, before Stage 2, the design is updated based on outcome data from the first cohort, and the second cohort of size $n^{(2)}$ is randomized according to the updated design. In this work, one experimental scenario with 50% censoring is explored. We consider:

- A single-stage design with uniform balanced dose allocation
- A single-stage design with the true (theoretical) D-optimal design
- A two-stage adaptive design ($n^{(1)} + n^{(2)} = 60$),
- A multi-stage adaptive design with a stopping rule based on the D-criterion [1]. Patients randomized in cohorts of size 15 each. After each cohort is randomized, we check if the determinant of the observed Fisher information matrix exceeded some predefined threshold. If so, the randomization stops.

Randomization: The cohort sizes can be small in practice, and an experimenter must ensure that actual allocation numbers are as close as possible to the targeted ones. At the same time, the allocation must involve a random element to minimize the potential for selection bias [4]. The following randomization procedures were investigated :

- *Completely Randomized Design (CRD)*
- *Permuted Block Design (PBD)*

Chronological bias: we assume that there is an effect due to a time trend as in [5]

Selection bias: We adopt the approach described in [6] for a 3-arm randomization setting. Based on a current disbalance in treatment assignments, we assign "sicker" patients to a "better" doses and vise versa.

Evaluation: For single-stage and two-stage adaptive designs, D-efficiency:

$$\mathrm{D}_{\mathrm{eff}}(n) = (|M(\xi_n, \boldsymbol{\theta})|/|M(\xi^*, \boldsymbol{\theta})|)^{1/4},$$

where ξ^* is the true D-optimal design, and ξ_n is the design with the same support points as ξ^* and allocation proportions obtained after randomization. For multi-stage designs, the procedures are compared in terms of distribution of a sample size upon termination.

Results

Table 1 shows average $\mathrm{D}_{\mathrm{eff}}(n)$ for a single-stage trial with locally D-optimal (or uniform) design with different randomization methods.

- CRD is less efficient than PBD for small sample sizes.
- Uniform PBD is much less efficient than optimized CRD and PBD.

	CRD D-optimal	PBD D-optimal	PBD Uniform
n= 15	0.93	1.00	0.74
n= 30	0.97	1.00	0.74
n= 45	0.98	1.00	0.74
n= 60	0.99	1.00	0.74

Table 2 shows, for two-stage adaptive designs, the percentage of simulation runs for which the MLE of $\boldsymbol{\theta}$ could not be obtained from the first stage data (and, therefore, an estimate of the D-optimal design for stage 2)

- the choice of the randomization procedure is important.
- CRD results in a much larger percentage of fails compared to PBD.

Procedure in Stage 1	Size of Stage 1 ($n^{(1)}$)		
	15	30	45
CRD	23.00%	4.00%	1.70%
PBD	12.30%	1.40%	0.30%

Table 3 shows average $\mathrm{D}_{\mathrm{eff}}(n)$ of 2-stage adaptive design strategies.

- CRD→CRD (i.e. CRD used in both Stage 1 and Stage 2) results in 1–2% loss compared to the PBD→PBD strategy.
- When the total sample size is split equally ($\mathrm{n}^{(1)} = \mathrm{n}^{(2)} = 30$), the 2-stage adaptive designs have highest $\mathrm{D}_{\mathrm{eff}}(n)$.
- Uniform designs (non-adaptive) have worse performance.

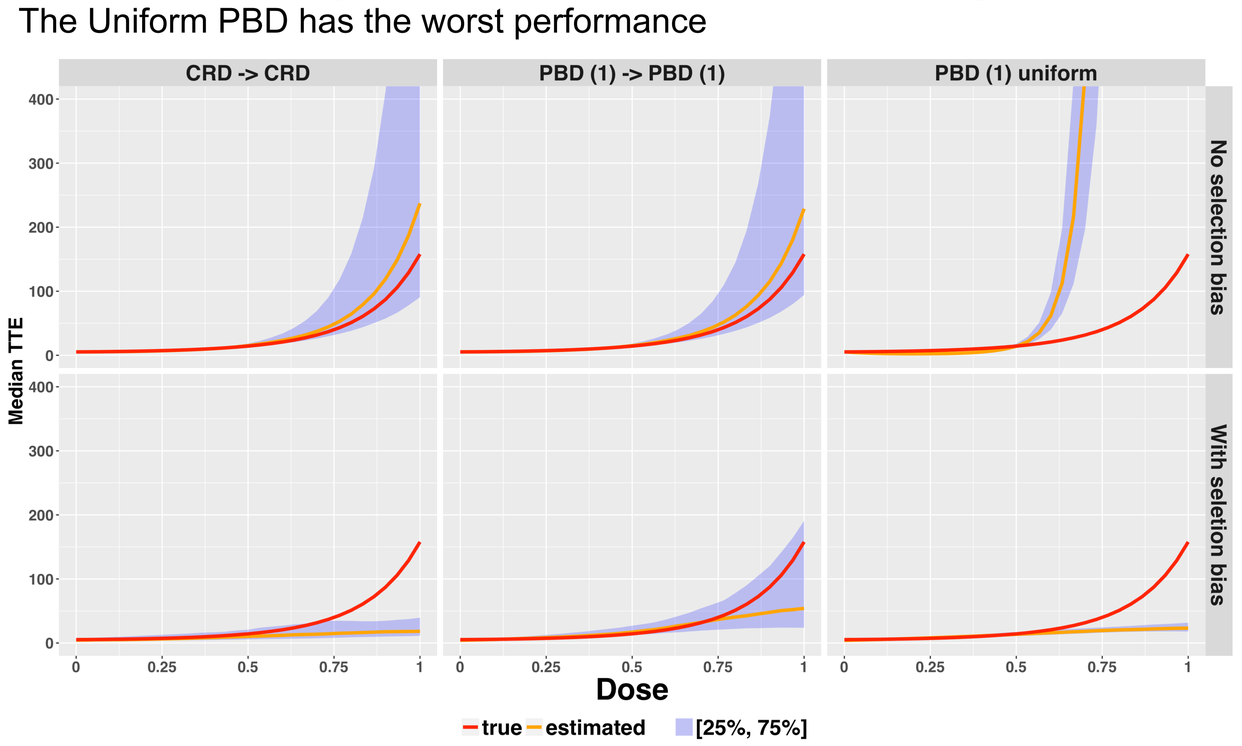
	CRD→ CRD	PBD→ PBD	PBD Uniform
$n^{(1)} = 15$ $n^{(2)} = 45$	0.79	0.81	0.74
$n^{(1)} = 30$ $n^{(2)} = 30$	0.84	0.85	0.74
$n^{(1)} = 45$ $n^{(2)} = 15$	0.82	0.83	0.74

Table 4 shows distributions of the sample size at study termination for two adaptive design strategies and the uniform allocation design.

- Designs have the same median sample size, but CRD→CRD has Larger maximum sample size.
- Non-optimal PBD uniform requires several more cohorts to achieve the same level of estimation accuracy as for the adaptive designs

Multi-stage randomization procedure			
Sample size statistics	CRD → CRD →	PBD → PBD →	PBD Uniform
min	15	15	15
25%	30	30	90
median	45	45	210
75%	60	45	270
max	120	105	345

Figure 1 shows that the presence of selection bias has a negative impact on quality of estimation: the designs tend to systematically underestimate the dose–response curve at higher dose levels. The "least affected" design is PBD → PBD. The Uniform PBD has the worst performance



We find that ***chronological bias*** had no impact on D-efficiency for any of the considered design strategies—the average values of $\mathrm{D}_{\mathrm{eff}}(n)$ were identical in the no-trend case and in the cases when the trend was present.

Conclusions

Both the choice of an allocation design and a randomization procedure to implement the target allocation impacts the quality of dose–response estimation, especially for small samples. The choice of randomization procedure can be crucial, especially for small populations.

References

[1] Ryeznik Y., Sverdlov O., Hooker A. C. *Adaptive optimal designs for dose-finding studies with time-to-event outcomes*, The AAPS Journal, 20:24, 2018

[2] Lawless JF. *Statistical models and methods for lifetime data*, 2nd ed., New York: Wiley, 2003

[3] Fedorov V.V., Hackl P. *Model-oriented designs of experiments,* Berlin: Springer, 1997

[4] Berger V.V. Selection bias and covariate imbalances in randomized clinical trials, West Sussex: Wiley, 2005

[5] Tamm M, Hilgers R.D. *Chronological bias in randomized clinical trials arising from different types of unobserved time trends*. Methods of Information in Medicine 53, 501-510, 2014

[6] Rückbeil MV, Hilgers RD, Heussen N *Assessing the impact of selection bias on test decisions in trials with a time-to-event outcome*. Statistics in Medicine **36**, 2656-2668, 2017