Assurance methods for designing a survival trial with a delayed treatment effect

James Salsbury[1], Jeremy Oakley[1], Steven Julious[2], Lisa Hampson[3]

[1]School of Mathematics and Statistics, The University of Sheffield
[2]School of Health and Related Research, The University of Sheffield
[3]Novartis Pharma AG, Basel, Switzerland

The University Of Sheffield.

NOVARTIS

Introduction

- Overview of assurance/probability of success
- Delayed treatment effects
- Assurance with DTEs

What is assurance?

- Power is

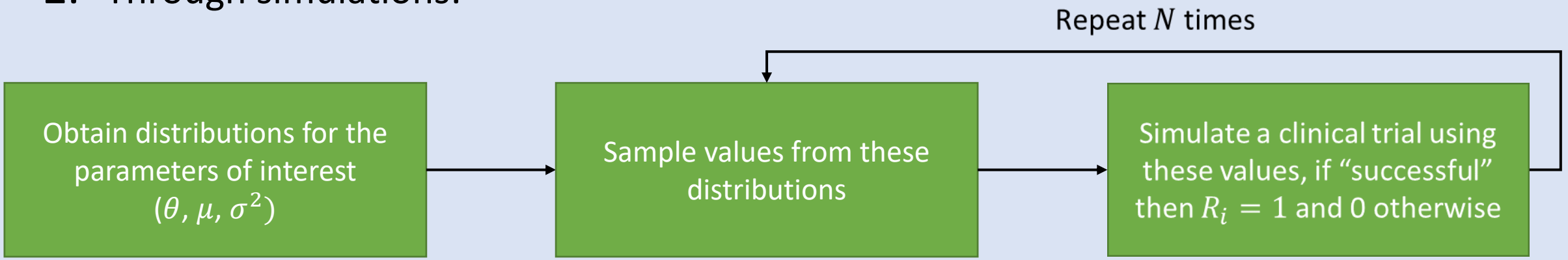
*"The probability of a trial being "successful" **given** a difference θ exists"*

- However, this is **conditional**. Can we do better?
- If we obtain a prior distribution $p(\theta)$ (instead of assuming θ takes a fixed value), and integrate over this prior distribution then this probability is now **unconditional**
- This is called an **assurance calculation**[1]
- **Assurance** is also known as expected power, average power, predictive power, probability of success etc..

[1]O'Hagan, Anthony & Stevens, John & Campbell, Michael. (2005). Assurance in clinical trial design. Pharmaceutical Statistics. 4. 187 - 201. 10.1002/pst.175.

How do you calculate assurance?

- Two ways:
  1. Analytically
  2. Through simulations:



Assurance is then estimated as $\hat{P}(R) = \frac{1}{N}\sum_{i=1}^{N} R_i$

Probability of Success (PoS) at Novartis

END OF PH2

Probability of "Success"

Within a Phase	60%	95%	70%
Probability of Approval	60% * 95% = **57%**	(potential gap)	
Probability of Success	60% * 95% * 70% = **40%**		

	Information added	How	PoS at end of Ph2
1	7 program characteristics	Predictive model fitted to industry data	Step 1 estimate
2	Pivotal / Phase 3 risks	Bayesian analysis: prob of meeting TPP	Step 2 estimate
3	Beyond Pivotal / Ph 3 risks	Adjustment algorithm calibrated by experts	Step 3 estimate
4	Unaccounted risks / data	Subjective adjustment (Exceptions only)	Final estimate

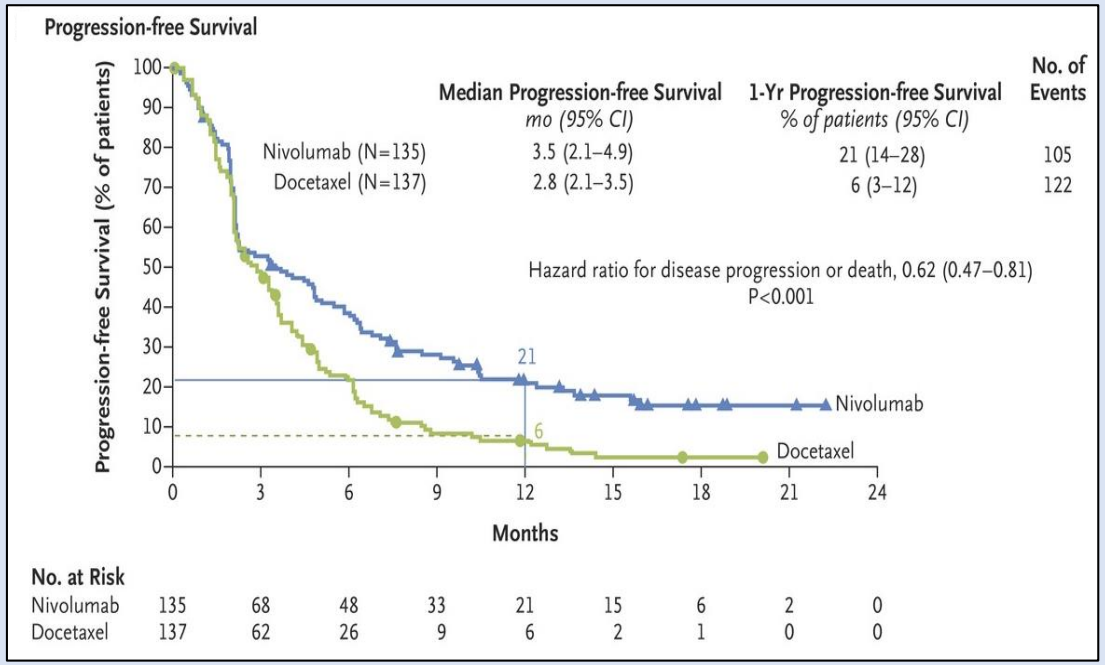
Incremental PoS adjustment

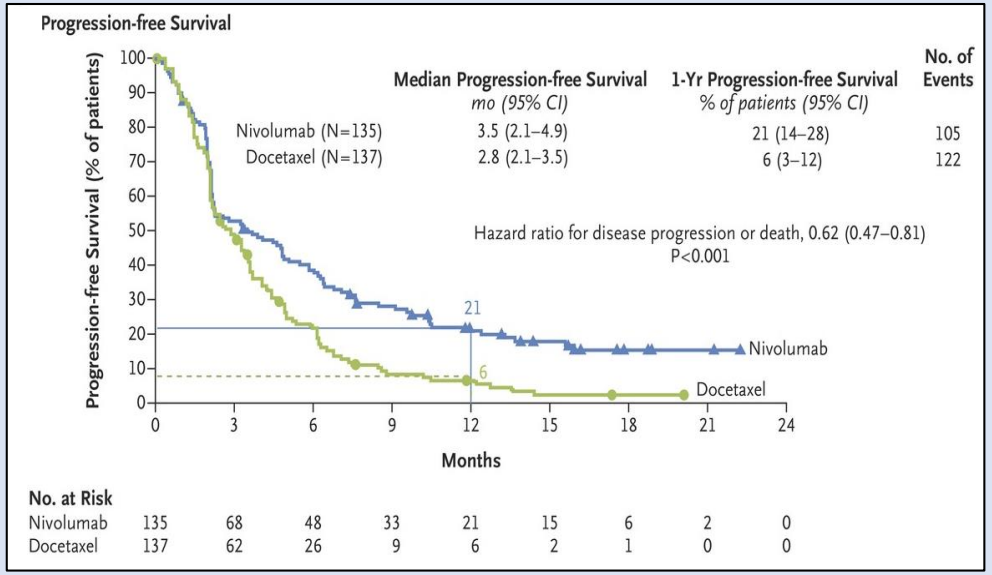
Hampson LV, Bornkamp B, Holzhauer B, et al. Improving the assessment of the probability of success in late stage drug development. Pharm Stats. 2022; 21(2): 439- 459. doi:10.1002/pst.2179

Delayed Treatment Effects

- A survival trial in which the control and treatment survival arms follow the same trajectory for some time, T, at which time they separate
- This phenomenon is known as **delayed treatment effects** (DTEs) and is a form of **non-proportional hazards** (NPHs)

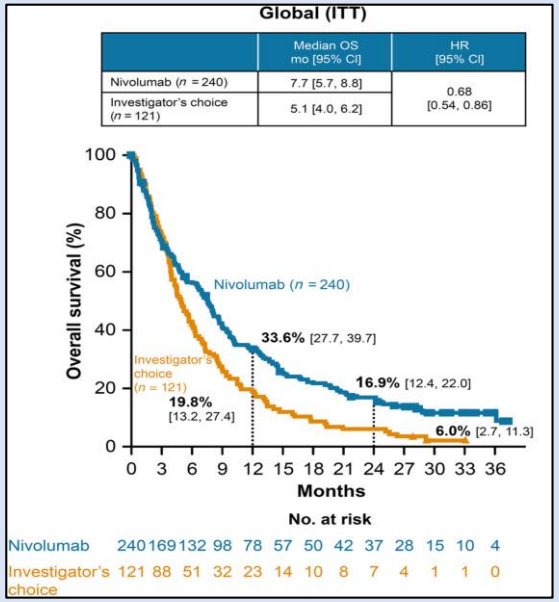
Checkmate 017



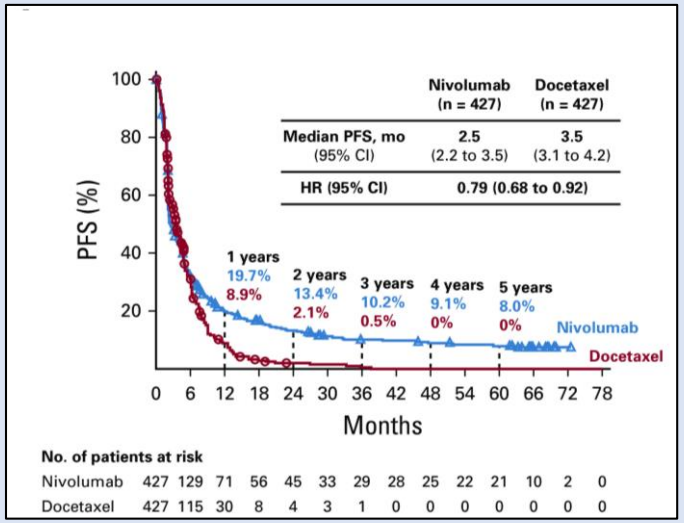


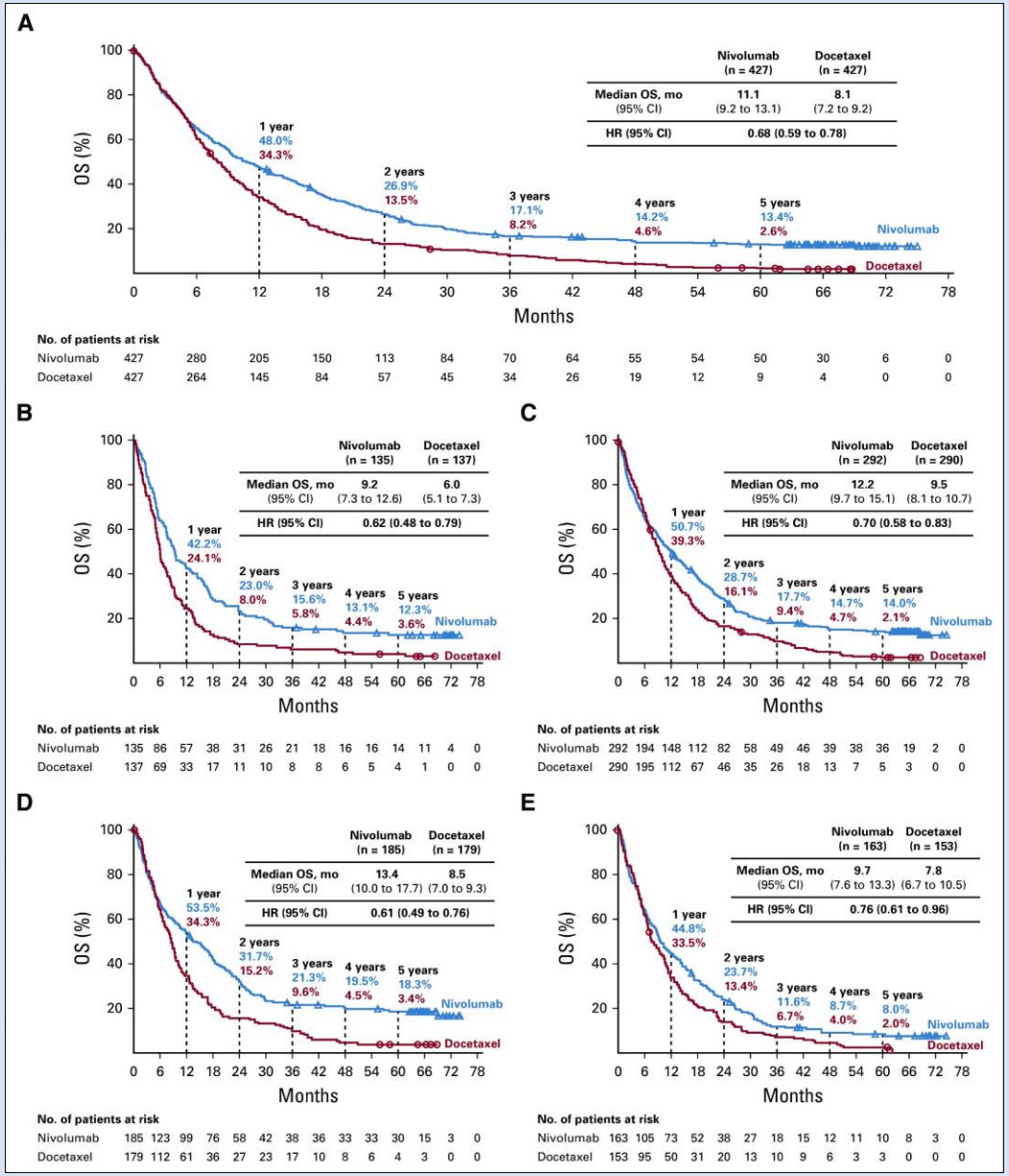
Checkmate 017

Checkmate 141



Checkmate 017 & Checkmate 057 (combined)





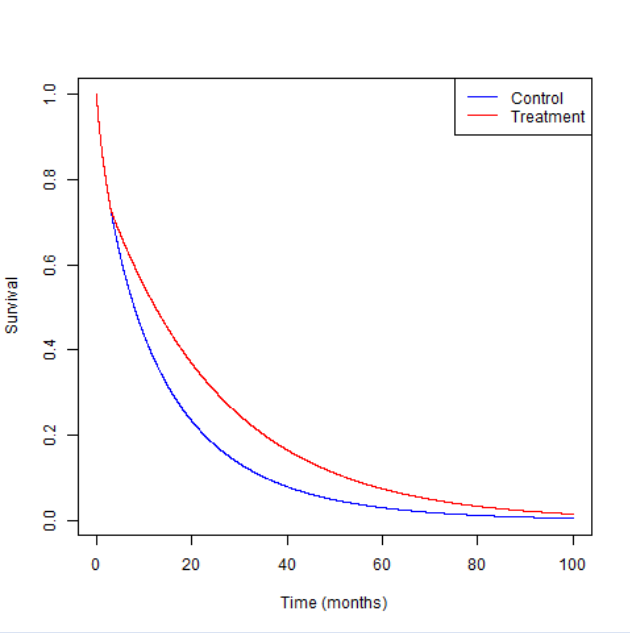
What makes DTE hard?

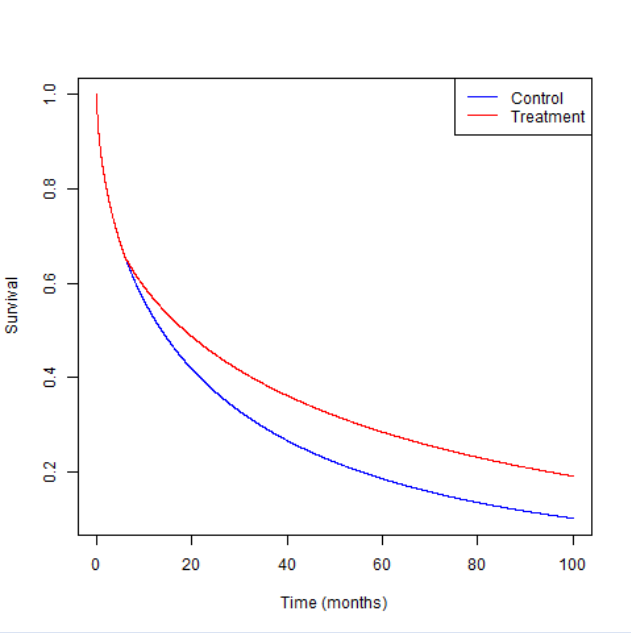
- In design:
  - When are the curves going to separate?
  - If we don't account for the delay, we lose power
  - When to plan for interim analyses?

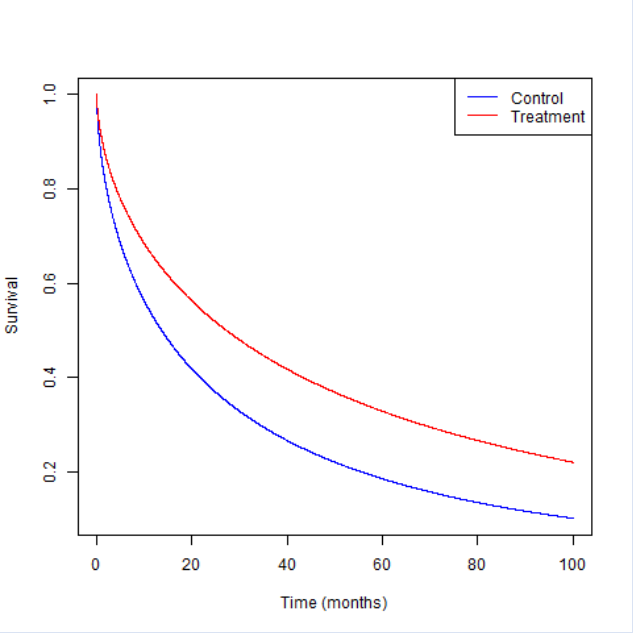
- In analysis:
  - Proportional hazards are violated, how to account for this?
  - Weighted log-rank test, RMST etc..

Parameterisation

- $S_c(t) = \exp\{-(\lambda_c t)^{\gamma_c}\}$
- $S_t(t) = \begin{cases} \exp\{-(\lambda_c t)^{\gamma_c}\}, & 0 \le t \le T \\ \exp\left\{-(\lambda_c T)^{\gamma_c} - \lambda_t^{\gamma_t}(t^{\gamma_t} - T^{\gamma_t})\right\}, & t > T \end{cases}$







Elicitation

- We have five unknown parameters in our parameterisation: $\lambda_c, \gamma_c, \lambda_t, \gamma_t$ and T
- How do we elicit beliefs about these parameters?
- For λ_c and γ_c:
  - We can use historical data (`RBesT`[1])
  - Ren and Oakley (2014)[2] consider eliciting Weibull parameters
- For T:
  - We can ask questions directly about the length of delay
- For λ_t and γ_t:
  - We can ask questions such as:
    - Median survival time on experimental treatment
    - Survival probability at time t
    - Hazard ratio at time t

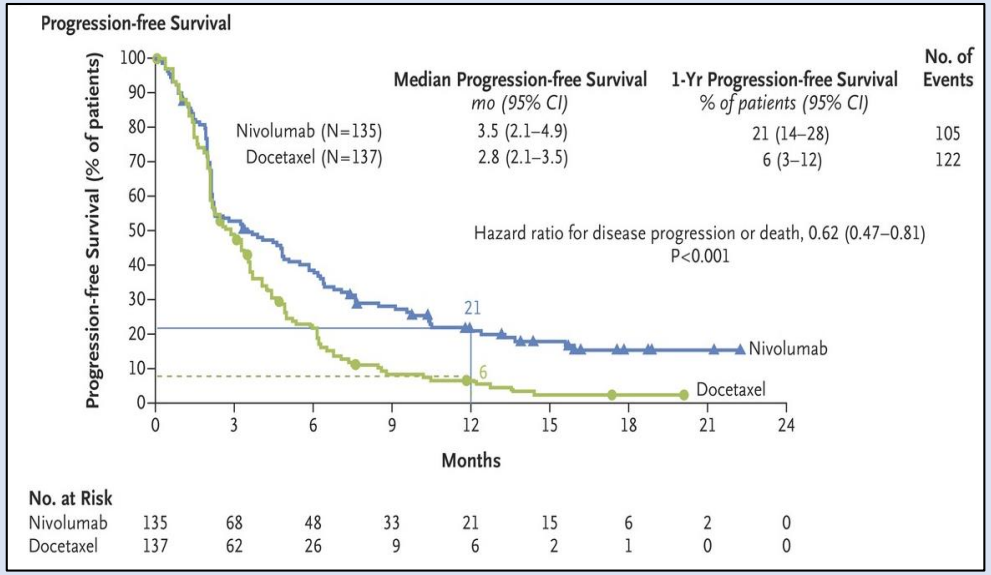
[1]Weber S, Li Y, Seaman JW, Kakizume T, Schmidli H (2021). "Applying Meta-Analytic-Predictive Priors with the R Bayesian Evidence Synthesis Tools." _Journal of Statistical Software_, *100*(19), 1-32. doi: 10.18637/jss.v100.i19
[2]Ren S, Oakley JE. Assurance calculations for planning clinical trials with time-to-event outcomes. Stat Med. 2014;33(1):31-45. doi:10.1002/sim.5916.

Elicitation – hazard ratio

- The **hazard ratio** is:

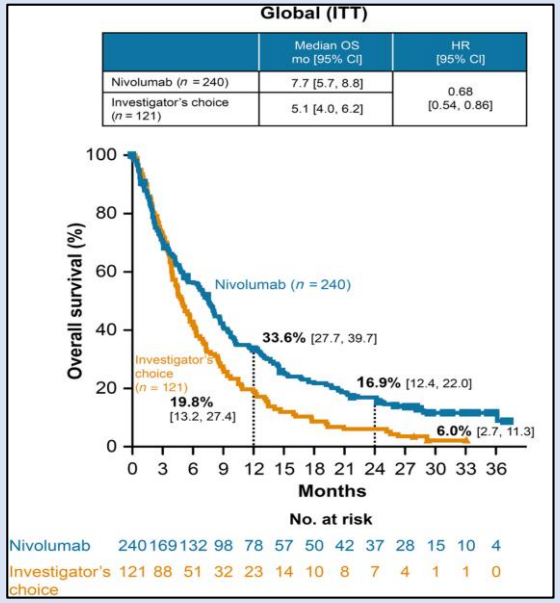
$$\mathrm{HR}(t)\begin{cases} 1, & 0 \le t \le T \\ \dfrac{\gamma_t \lambda_t^{\gamma_t} t^{\gamma_t - 1}}{\gamma_c \lambda_c^{\gamma_c} t^{\gamma_c - 1}}, & t > T \end{cases}$$

- We observe that after the delay, the hazard ratio seems to be **constant**

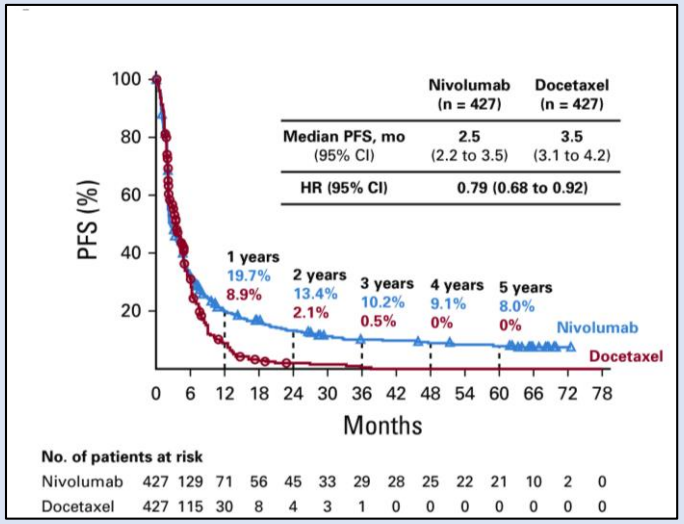


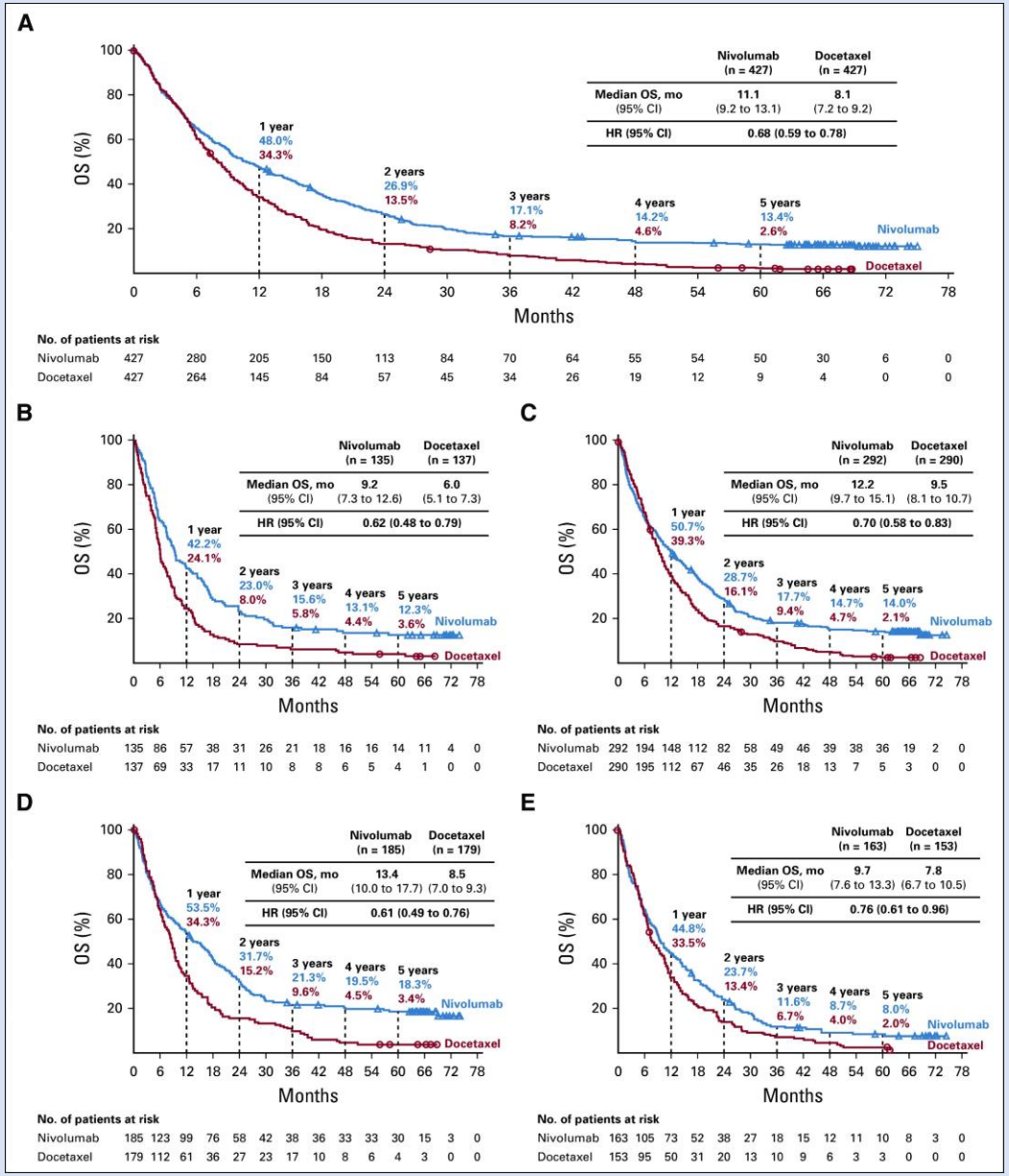
Checkmate 017

Checkmate 141



Checkmate 017 & Checkmate 057 (combined)





Elicitation – hazard ratio

- The **hazard ratio** is:

$$\mathrm{HR}(t)\begin{cases} 1, & 0 \le t \le T \\ \dfrac{\gamma_t \lambda_t^{\gamma_t} t^{\gamma_t - 1}}{\gamma_c \lambda_c^{\gamma_c} t^{\gamma_c - 1}}, & t > T \end{cases}$$

- We observe that after the delay, the hazard ratio seems to be **constant**
- We incorporate this into the parameterisation by setting $\gamma_c = \gamma_t$
- The post-delay HR (HR^*) is now

$$\mathrm{HR}^* = \left(\frac{\lambda_t}{\lambda_c}\right)^{\gamma_c}$$

- We are able to indirectly elicit beliefs about λ_t by asking questions about HR^*

Elicitation methods - SHELF

- How do you actually elicit beliefs from experts? Non-trivial..
- **SH**effield **EL**icitation **F**ramework (SHELF)[1] is a package of documents, templates and software to carry out elicitation of probability distributions for uncertain quantities from a group of experts
- Two (most) common ways of elicitation are:
  1. Trial roulette method
  2. Quantile method
- Both then involve a least squares fit to a standard parametric distribution (usually)

[1]Oakley JE, O'Hagan A. SHELF: the Sheffield elicitation framework (version 2.0), School of Mathematics and Statistics, University of Sheffield, 2010 (http://www.jeremy-oakley.staff.shef.ac.uk/shelf/).

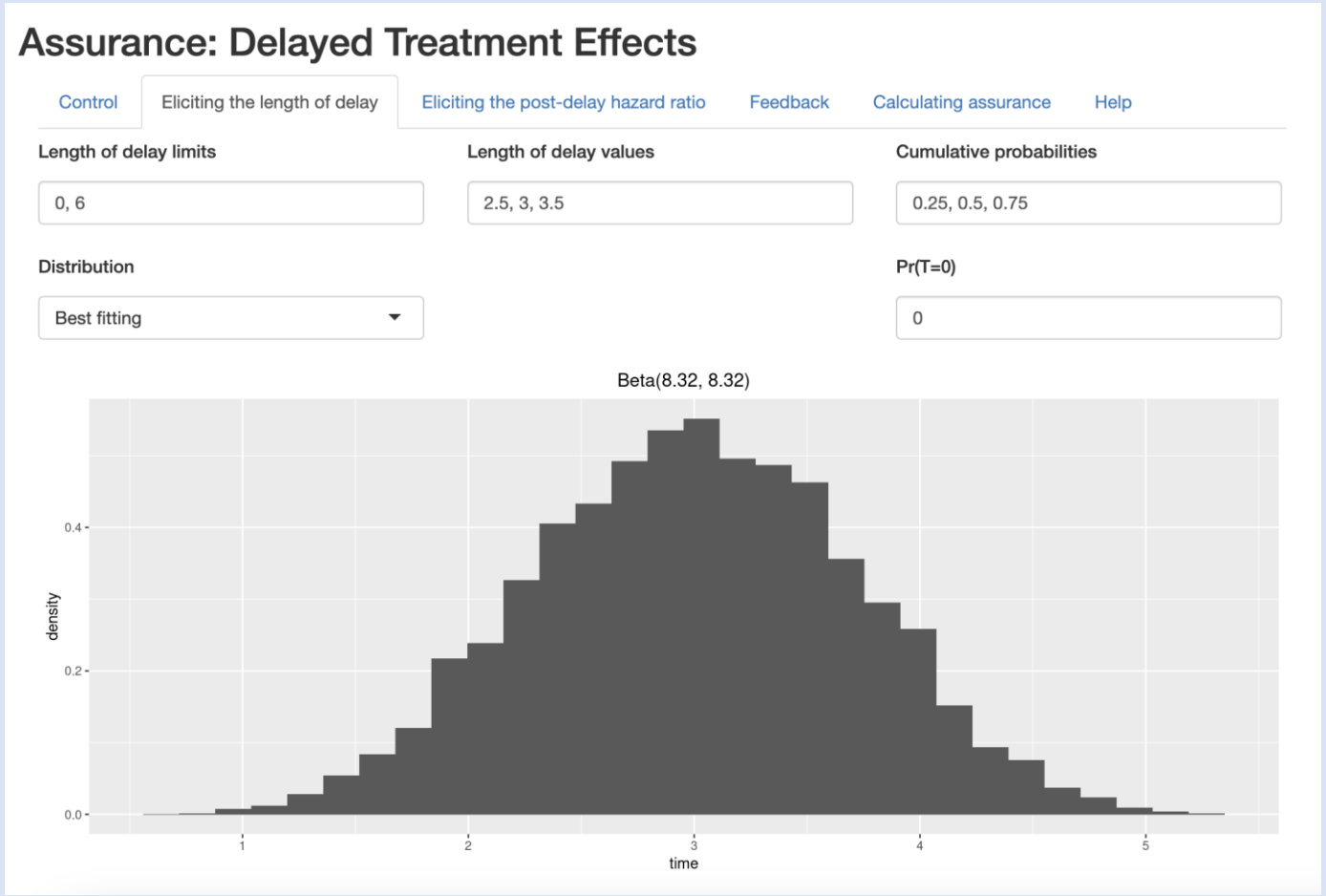
Calculating assurance

- Once we have distributions for control, T and HR^* we can use these to calculate assurance
- For $i = 1, \ldots, M$:
  - Simulate control data from $p(\mathrm{control})$
  - Sample T_i, HR_i^* from $p(T), p(\mathrm{HR}^*)$
  - Simulate treatment data from T_i, HR_i^*
  - Simulate a clinical trial using control and treatment data
  - If trial successful then $R_i = 1, 0$ otherwise
- End for
- Assurance is

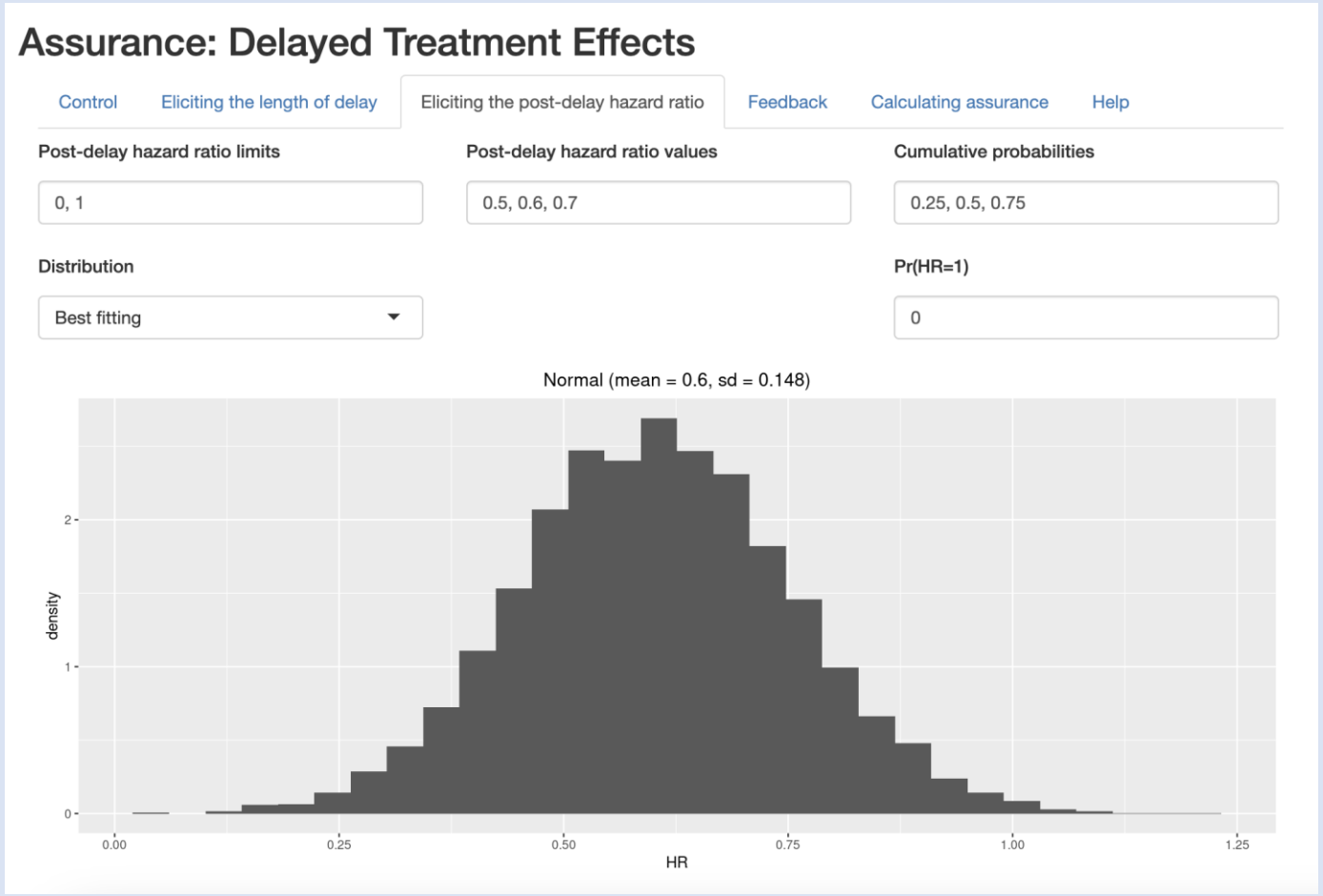
$$\frac{1}{M}\sum_{i=1}^{M} R_i$$

Interim analyses (futility, efficacy..), choice of analysis (weighted log-rank test, RMST..) can be changed here

DTE Shiny App (1/3)



DTE Shiny App (2/3)



DTE Shiny App (3/3)

Assurance: Delayed Treatment Effects

Control | Eliciting the length of delay | Eliciting the post-delay hazard ratio | Feedback | Calculating assurance | Help

Maximum number of patients in the trial

1000

Recruitment length

12

Ratio control **Ratio treatment**

1 1

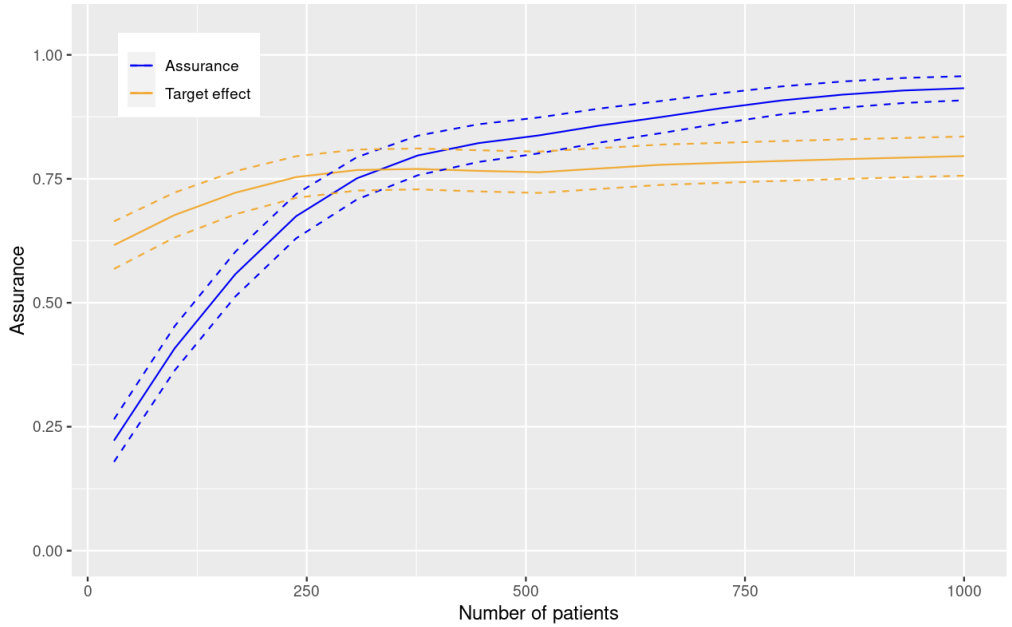
Maximum trial duration (including recruitment time)

60

Target effect (average hazard ratio)

0.8

Produce plot



The **blue** line is the proportion of trials that give rise to a 'successful' outcome.
The **orange** line is the proportion of trials in which the estimated average hazard ratio is less than the target effect - 0.8.
On average, 910 events are seen when 1000 patients are enroled for 60 months.

DTE assurance paper

Statistics > Applications

[Submitted on 10 Oct 2023]

Assurance Methods for designing a clinical trial with a delayed treatment effect

James Salsbury, Jeremy Oakley, Steven Julious, Lisa Hampson

An assurance calculation is a Bayesian alternative to a power calculation. One may be performed to aid the planning of a clinical trial, specifically setting the sample size or to support decisions about whether or not to perform a study. Immuno-oncology (IO) is a rapidly evolving area in the development of anticancer drugs. A common phenomenon that arises from IO trials is one of delayed treatment effects, that is, there is a delay in the separation of the survival curves. To calculate assurance for a trial in which a delayed treatment effect is likely to be present, uncertainty about key parameters needs to be considered. If uncertainty is not considered, then the number of patients recruited may not be enough to ensure we have adequate statistical power to detect a clinically relevant treatment effect. We present a new elicitation technique for when a delayed treatment effect is likely to be present and show how to compute assurance using these elicited prior distributions. We provide an example to illustrate how this could be used in practice. Open-source software is provided for implementing our methods. Our methodology makes the benefits of assurance methods available for the planning of IO trials (and others where a delayed treatment expect is likely to occur).

Subjects: **Applications (stat.AP)**; Methodology (stat.ME)

Cite as: arXiv:2310.06673 [stat.AP]

(or arXiv:2310.06673v1 [stat.AP] for this version)

https://doi.org/10.48550/arXiv.2310.06673

Submission history

From: James Salsbury [view email]

[v1] Tue, 10 Oct 2023 14:48:26 UTC (3,259 KB)

References

- O'Hagan, Anthony & Stevens, John & Campbell, Michael. (2005). Assurance in clinical trial design. Pharmaceutical Statistics. 4. 187 - 201. 10.1002/pst.175.
- Hampson LV, Bornkamp B, Holzhauer B, et al. Improving the assessment of the probability of success in late stage drug development. Pharm Stats. 2022; 21(2): 439- 459. doi:10.1002/pst.2179
- Weber S, Li Y, Seaman JW, Kakizume T, Schmidli H (2021). “Applying Meta-Analytic-Predictive Priors with the R Bayesian Evidence Synthesis Tools.” _Journal of Statistical Software_, *100*(19), 1-32. doi: 10.18637/jss.v100.i19
- Ren, Shijie & Oakley, Jeremy. (2014). Assurance calculations for planning clinical trials with time-to-event outcomes. Statistics in medicine. 33. 10.1002/sim.5916.
- **Checkmate 017** - Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. N Engl J Med. 2015;373(2):123-135. doi:10.1056/NEJMoa1504627
- **Checkmate 141** - Yen CJ, Kiyota N, Hanai N, et al. Two-year follow-up of a randomized phase III clinical trial of nivolumab vs. the investigator’s choice of therapy in the Asian population for recurrent or metastatic squamous cell carcinoma of the head and neck (CheckMate 141). Head Neck. 2020;42(10):2852-2862. doi:10.1002/hed.26331
- **Checkmate 017 & Checkmate 057 (combined)** - Borghaei H, Gettinger S, Vokes EE, et al. Five-Year Outcomes From the Randomized, Phase III Trials CheckMate 017 and 057: Nivolumab Versus Docetaxel in Previously Treated Non-Small-Cell Lung Cancer [published correction appears in J Clin Oncol. 2021 Apr 1;39(10):1190]. J Clin Oncol. 2021;39(7):723-733. doi:10.1200/JCO.20.01605
- Oakley JE, O'Hagan A. SHELF: the Sheffield elicitation framework (version 2.0), School of Mathematics and Statistics, University of Sheffield, 2010 (http://www.jeremy-oakley.staff.shef.ac.uk/shelf/).

Thank you! Any questions?

Arxiv paper

jsalsbury1@sheffield.ac.uk

jamesalsbury.github.io

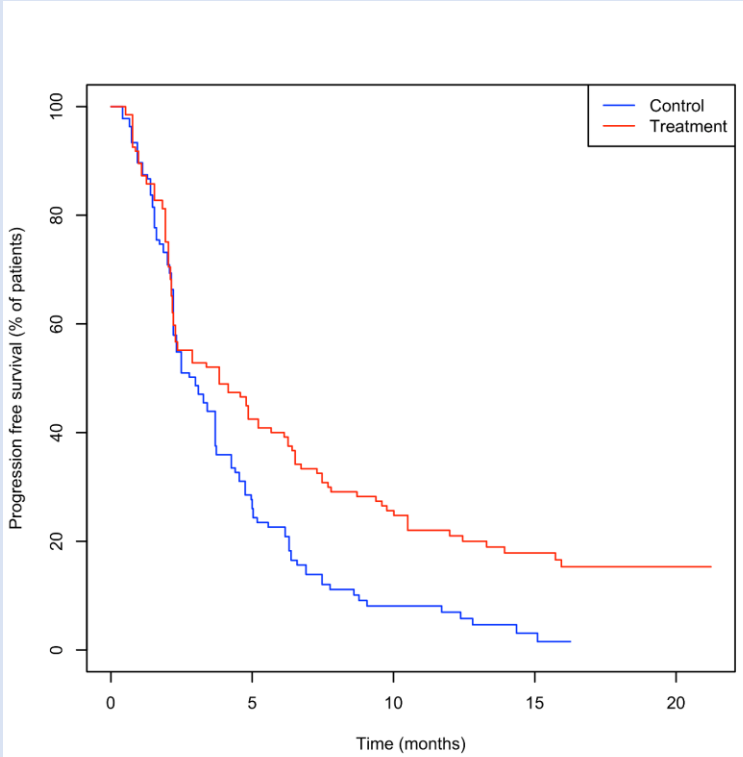
james-salsbury

DTE Shiny App

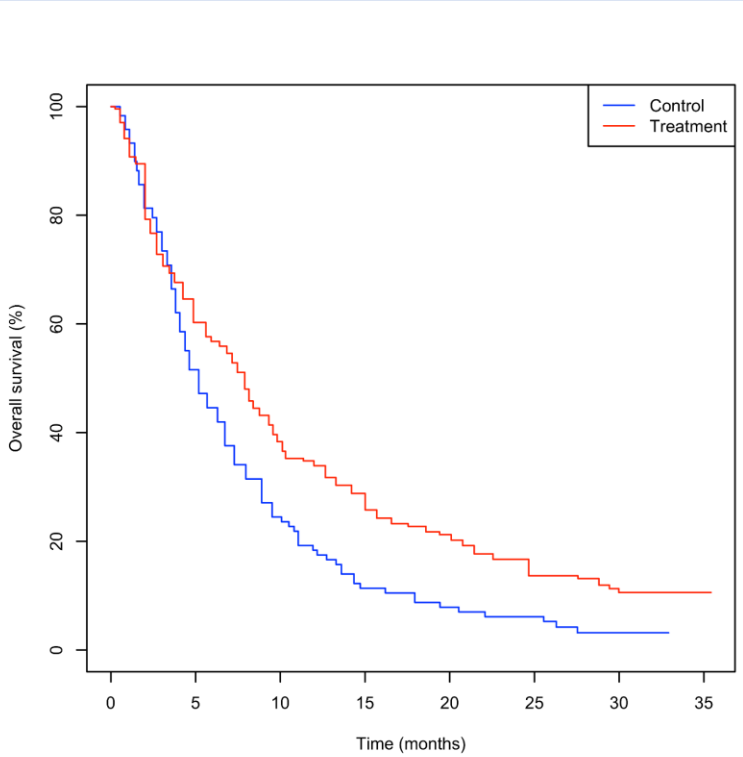
The University Of Sheffield.

NOVARTIS

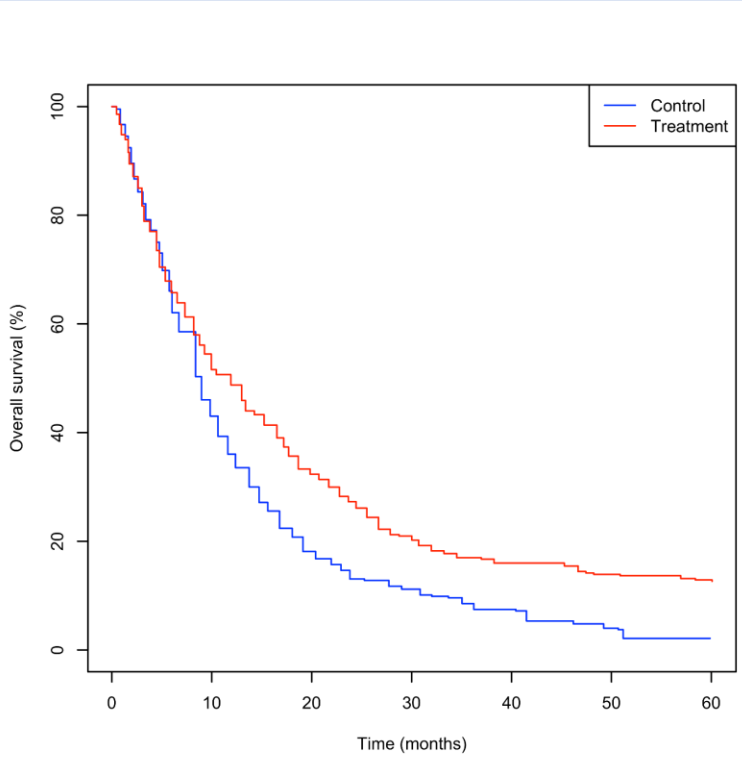
Back-up slides



Checkmate 017

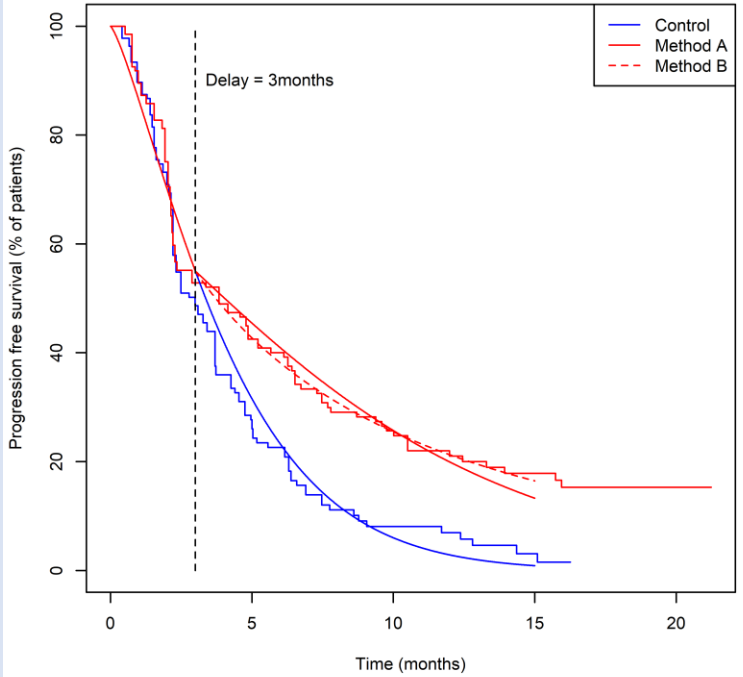


Checkmate 141

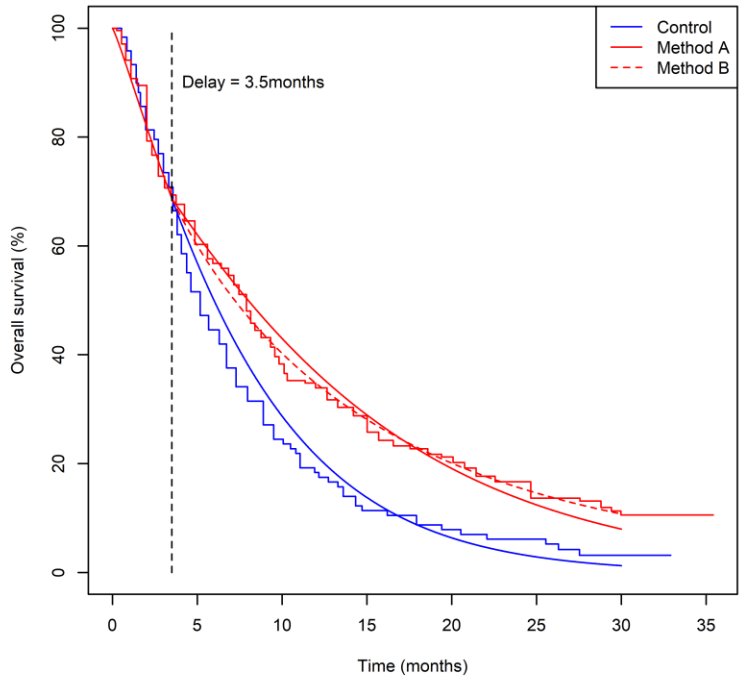


Checkmate 017 & Checkmate 057

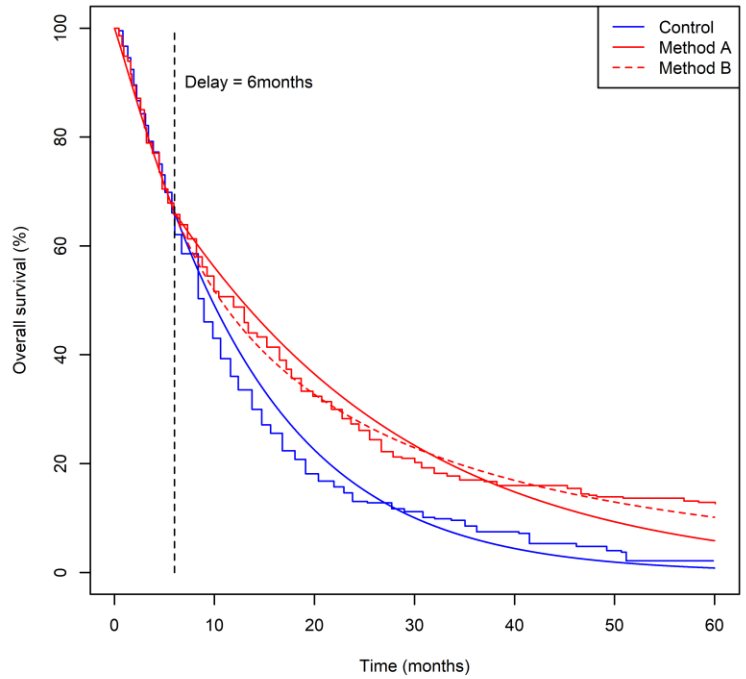
Back-up slides



Checkmate 017



Checkmate 141



Checkmate 017 & Checkmate 057