Early reporting of efficacy endpoints and its potential impact Biostatistical part

Jan Bogaerts

EORTC

Conflict of interest

- None

Interim analysis

- For futility: decide that continuing enrollment is of no further use, because a positive finding is unlikely
  - (Wisely) giving up
- For superiority: decide that the statistical comparison process can end here, because we consider the result significant at this time
  - Stop accrual, possibly change treatment for some
  - Submit data for registration and/or publication
  - This one is the topic of today

Three major concerns

- Maturity
- Information volume
- Correct estimation

Maturity

- Early stopping for superiority will mean some results will never be available:
  - OS full comparison (if decision on PFS)
  - Full accrual may not be reached
  - HTA assessment will suffer
- The interim database is typically ‘in flux’, so despite best efforts data will still evolve
  - Build in a buffer (not part of the statistical approach!)

Information volume

- The full trial is designed to have a well-powered dataset for the question
- Stopping early means there is a trade-off between the effect size observed (so far) and the volume of data available
- Typically the bulk of the comparative data at interim is during the early experience (since randomization) of patients
  - Risk of overlooking late effects
  - This flows into the next topic

Estimation

- Hazard ratio (logrank, Cox) is the average ratio in risk of experiencing the event:
  - Over time since randomization
  - Weighed by the available data
  - -> at interim, weighed towards early experience
- Hazard ratio at each point in time is conditioning on not having reached the event up to that time
  - Important to understand

Some math

- Assuming constant hazard ratio over time
- On an individual trial basis, there is no mathematical trick to stop at an unbiased point: all estimates are unbiased
- However, if we consider the x% trials that are stopped at interim: those estimates are (collectively) biased and will regress to (a) mean if such trials are allowed to continue

??? ->> Let's give you a way to think about this.

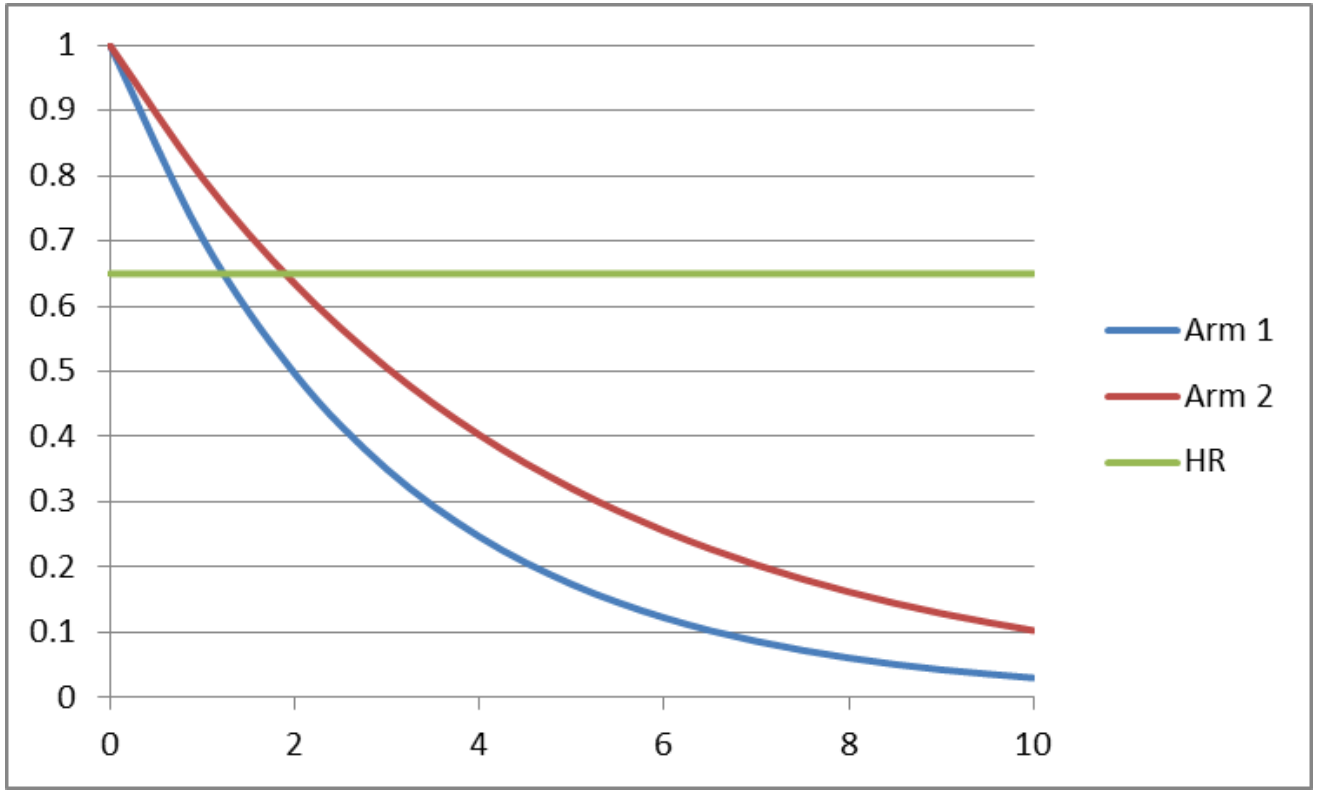
Some math – a parallel

- We play 10 rounds of toss-a-coin, one Euro per game (trial)
  - Final outcome is -10 Euro to +10 Euro, and everything in between
  - There is no strategy to "win" this game (on average)
- I stop after 5 rounds if winning (interim analysis)
  - Win +1 to +5 Euro
  - If I only look at the games I am stopping halfway, those estimates are too optimistic (it is a zero-sum game)

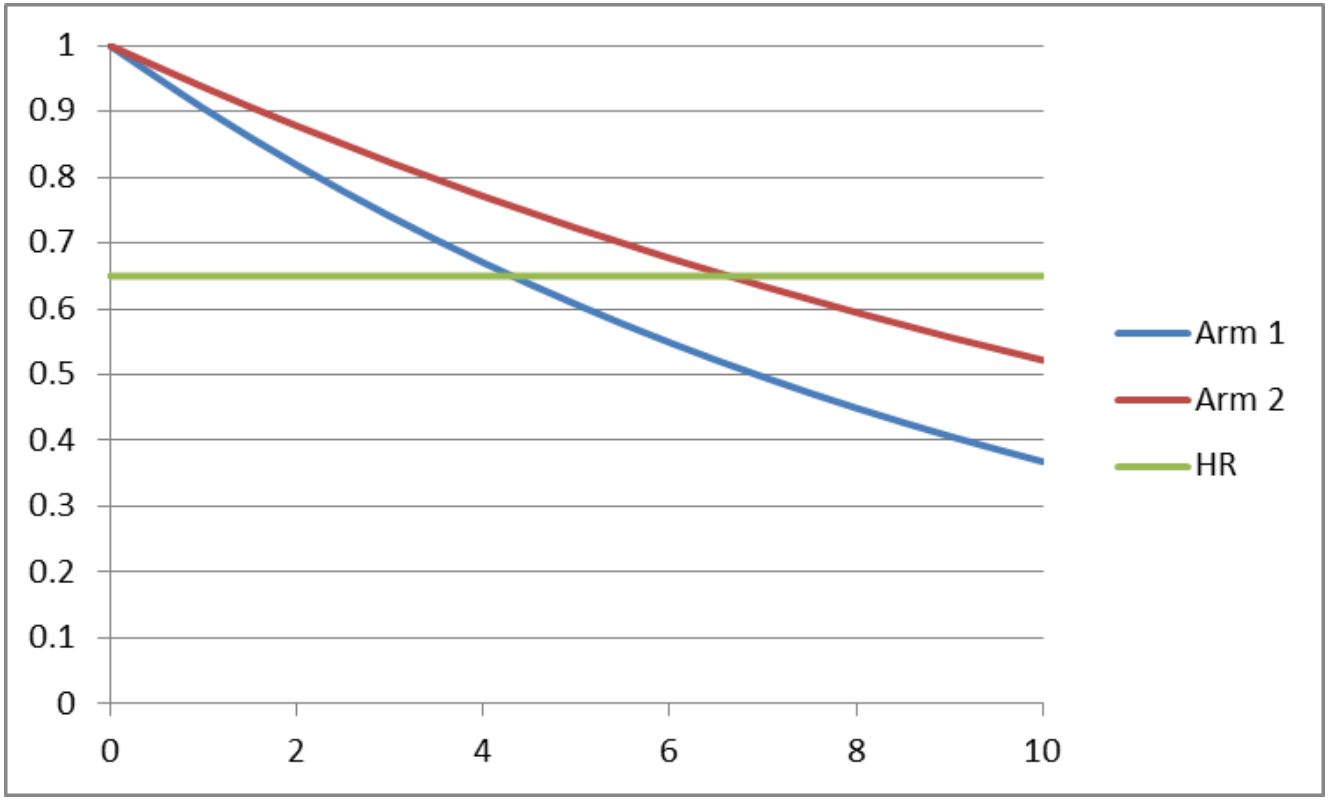
Following slides

- Estimates at interim will be
  - Unbiased for constant hazard ratios
  - Optimistically biased for converging hazard ratios
  - Pessimistically biased for diverging hazard ratios
- Immunotherapy -> diverging hazard ratios?

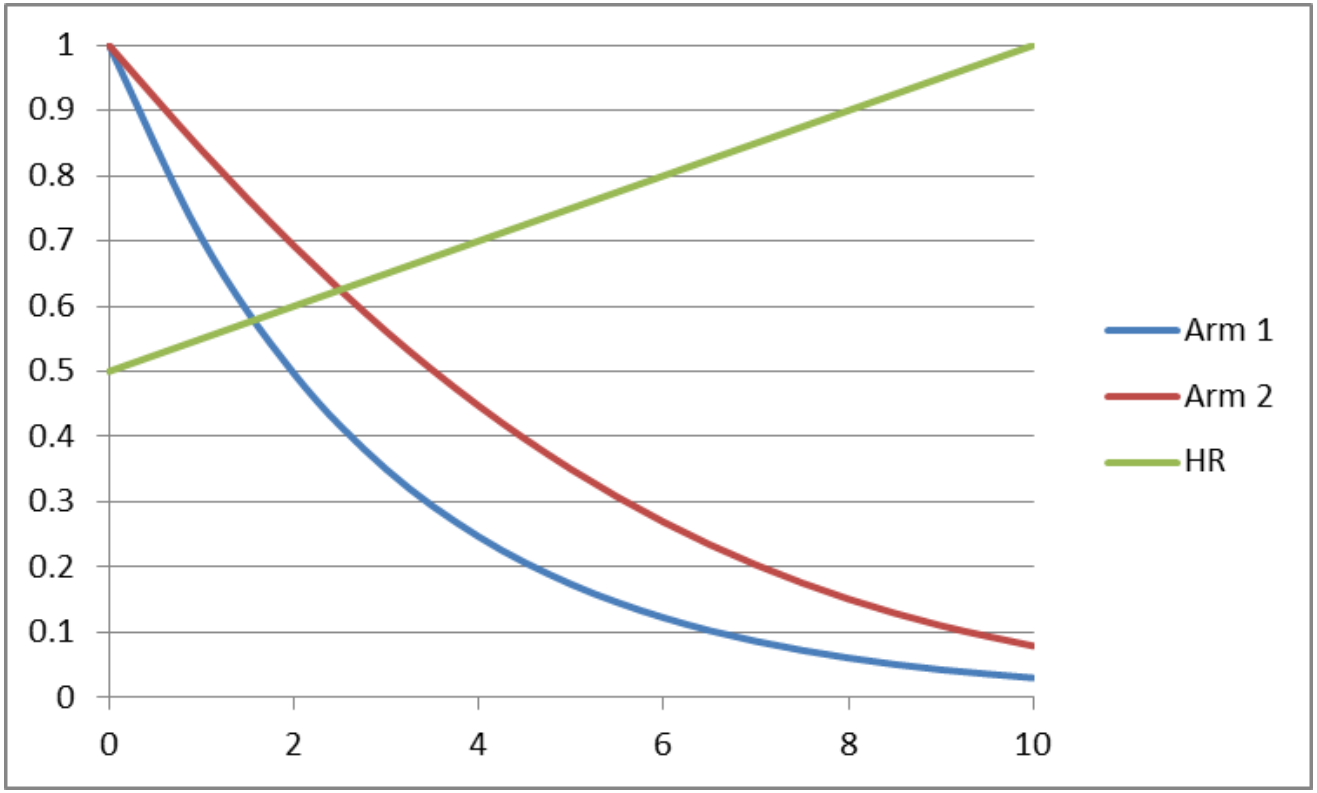
Constant hazard ratios



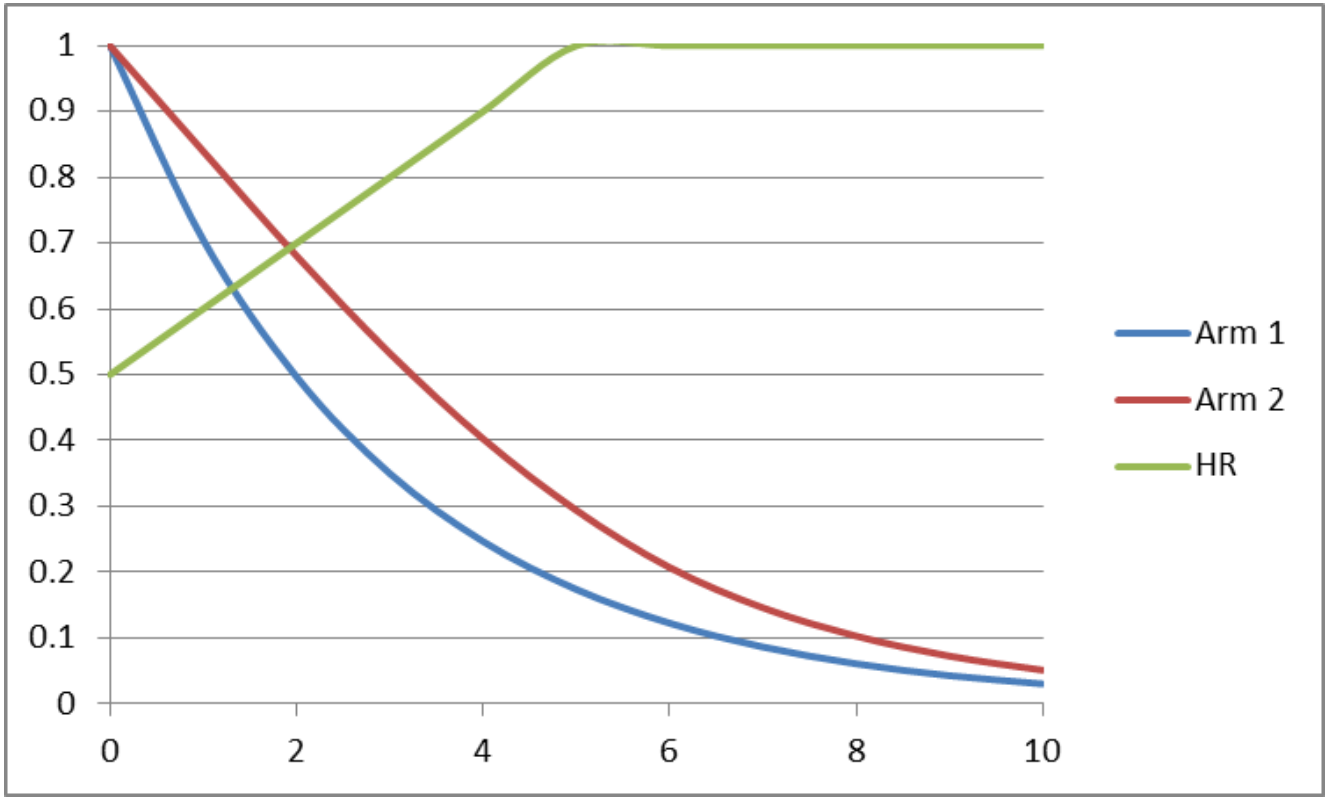
Constant hazard ratios



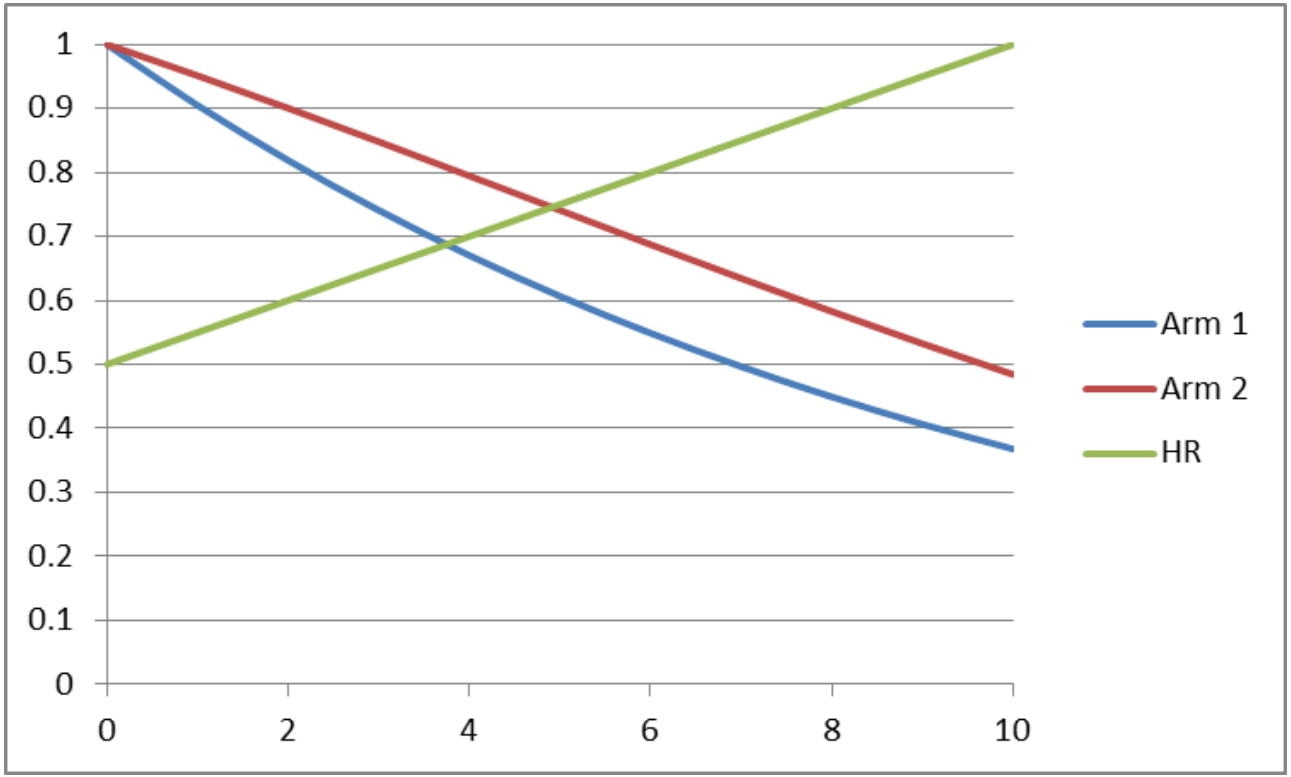
Converging hazard ratios



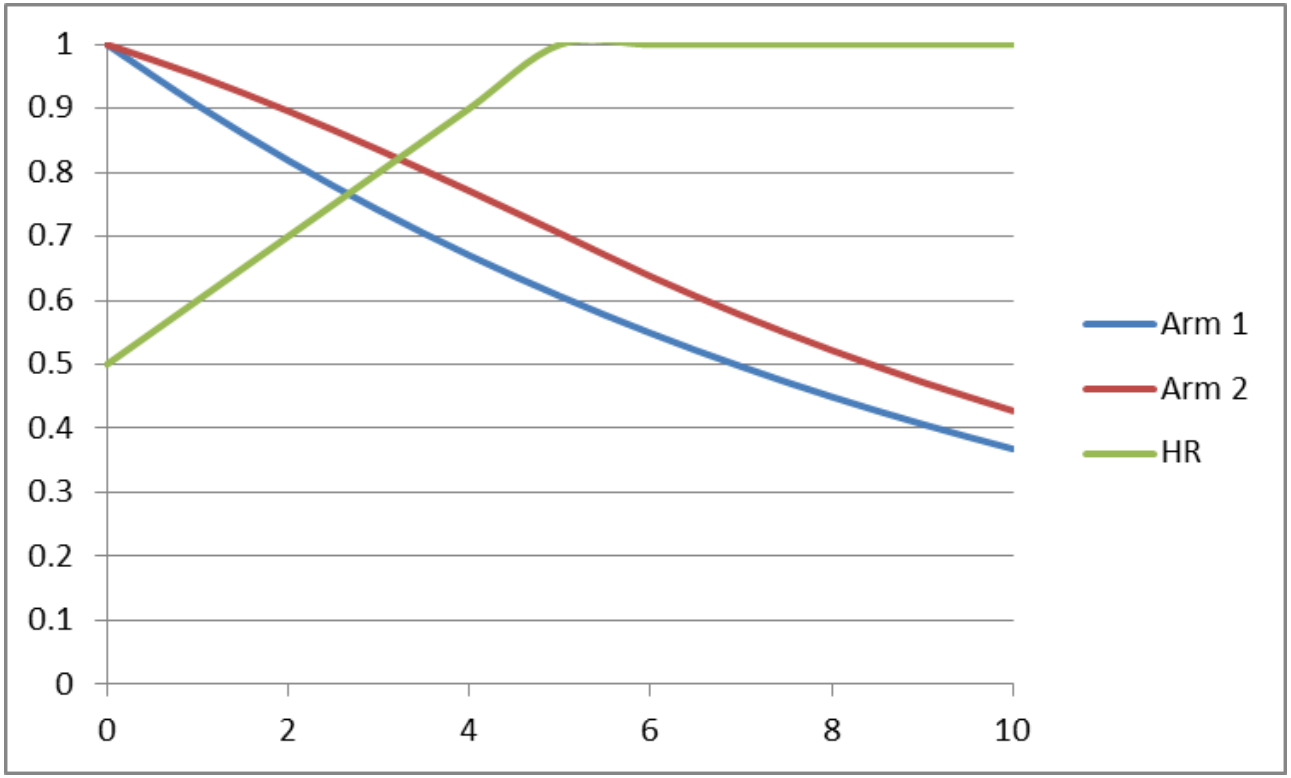
Converging hazard ratios



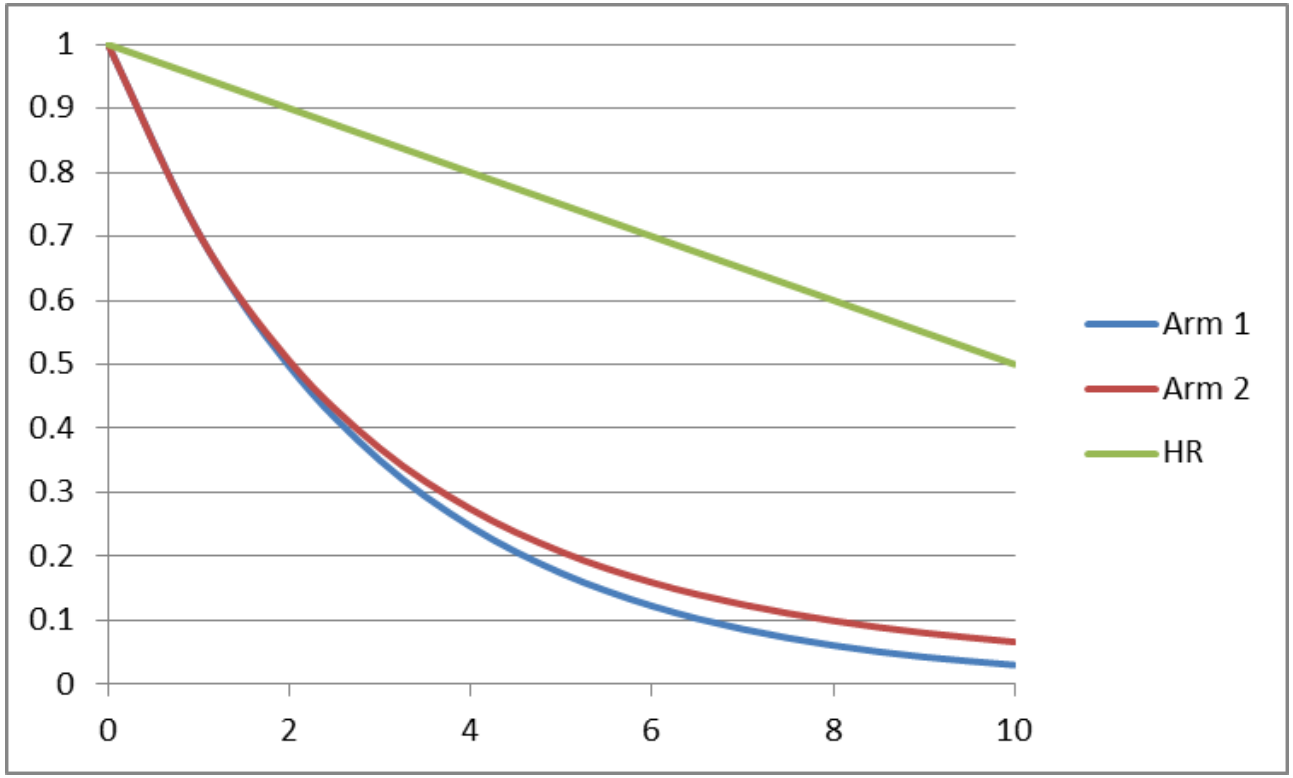
Converging hazard ratios



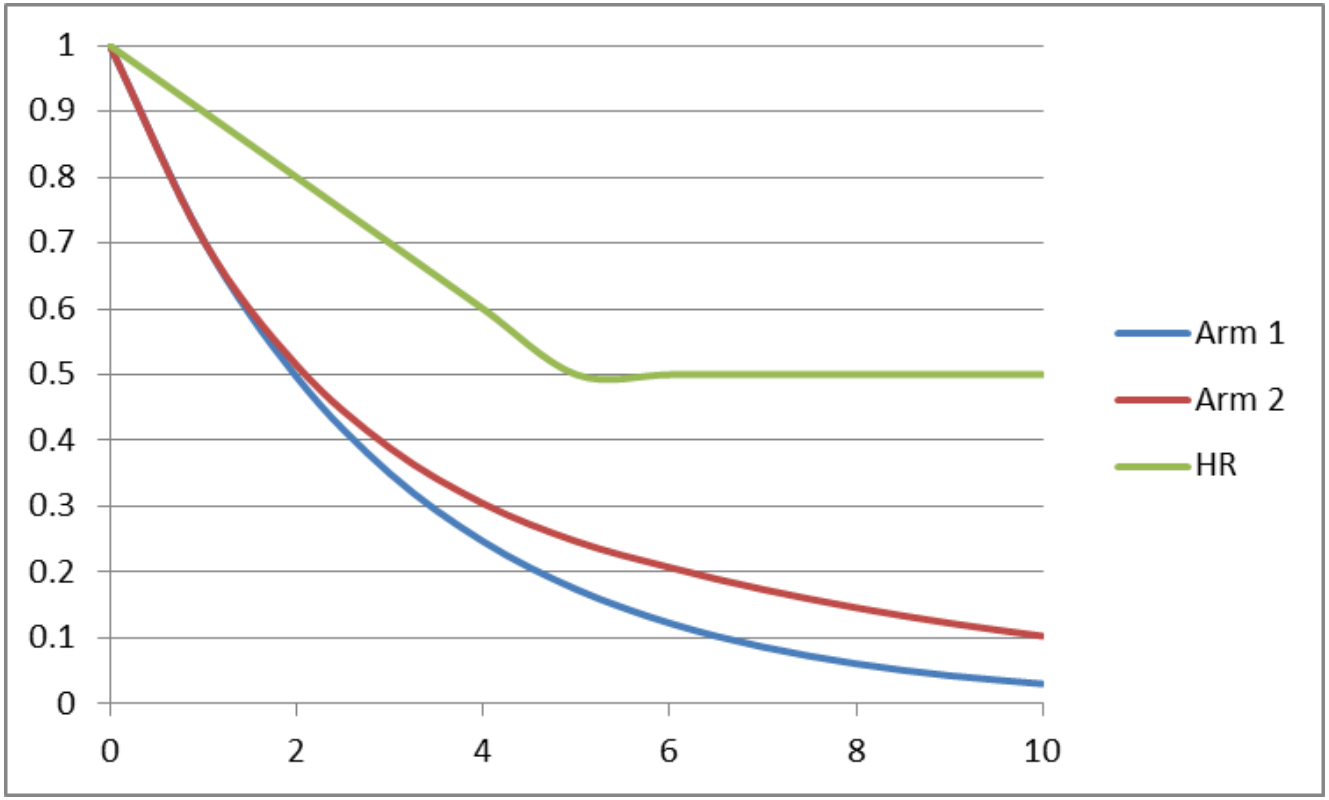
Converging hazard ratios



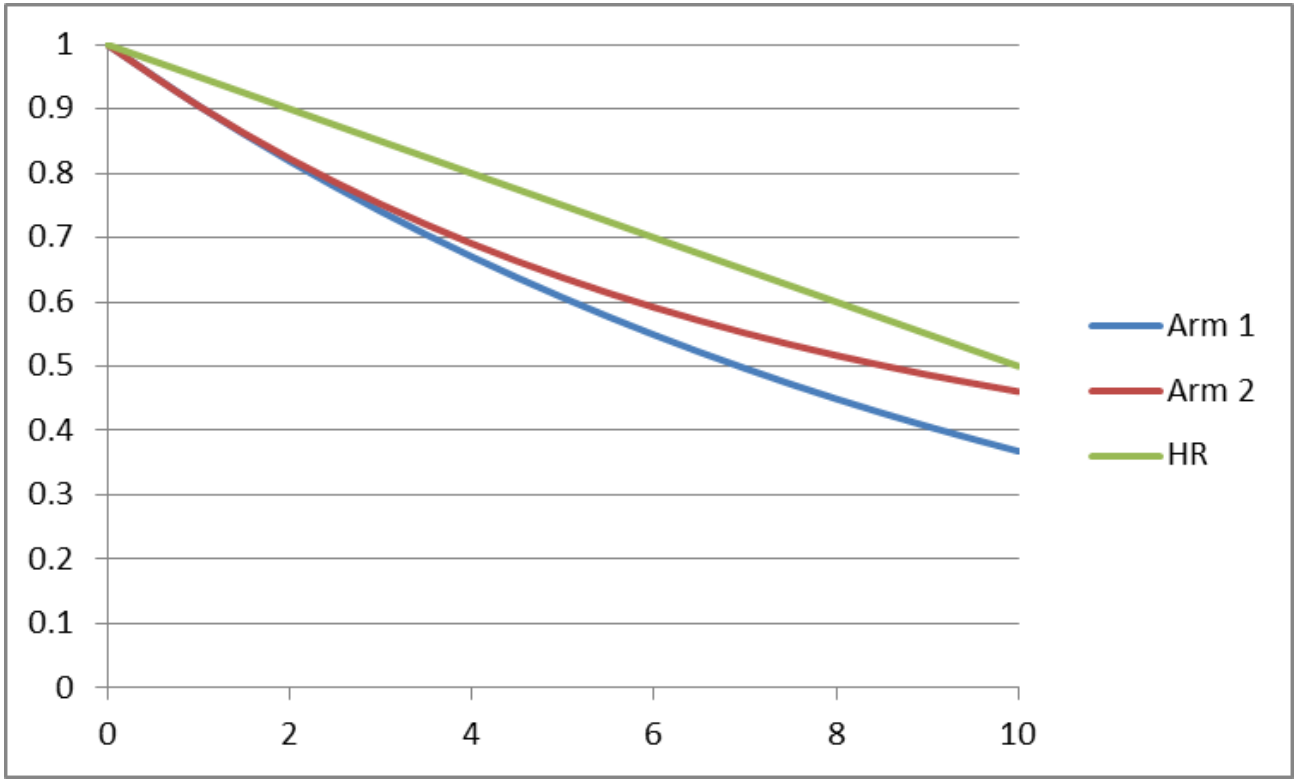
Diverging hazard ratios



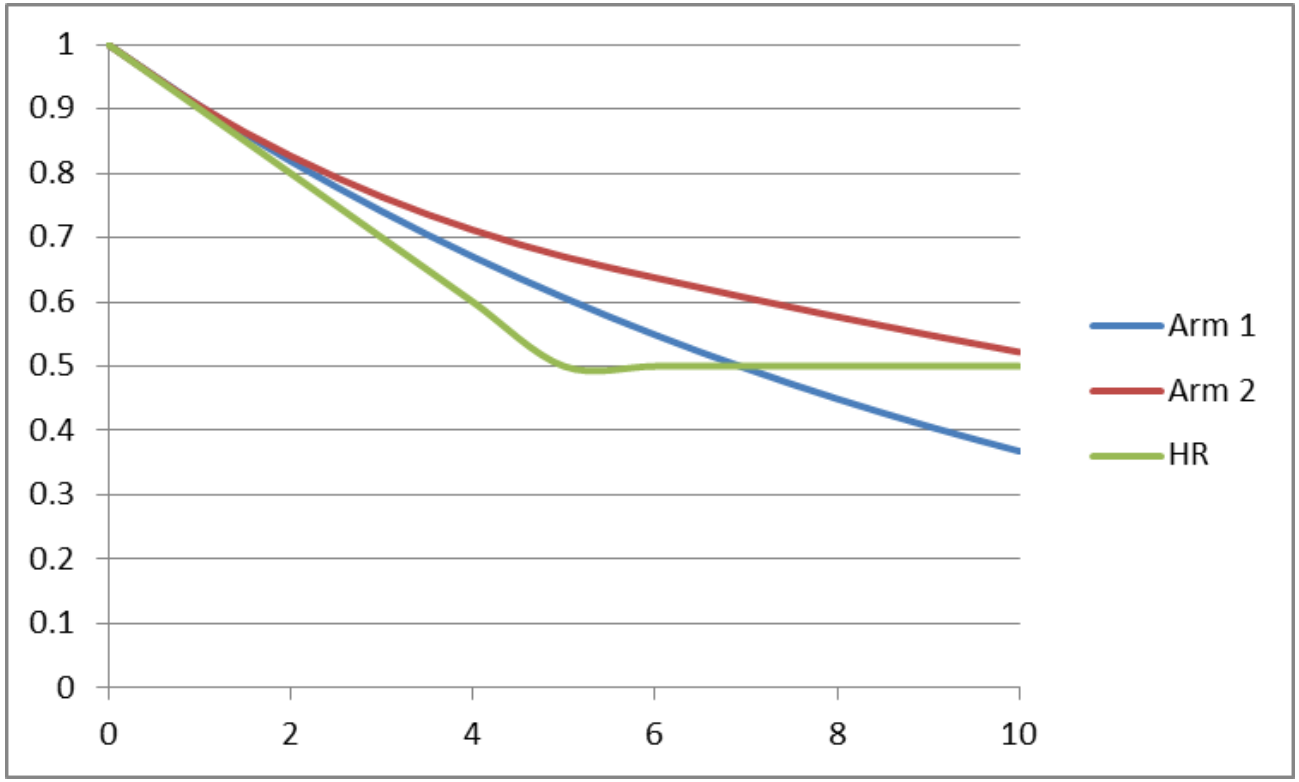
Diverging hazard ratios



Diverging hazard ratios



Diverging hazard ratios



Overall concern for any IDMC member

- Will these data sway everyone who will look at this in the future?
  - Impossible to answer
  - It is very dangerous to be led by a p-value (alone)
- Is the argument “this is how it was designed at the time the protocol was written” enough?
  - Should build in extra conditions on data volume and maturity

Future patients' interests / risks

On trial, may get

- Inefficient treatment which is current standard
- Are "few": those still to be enrolled on the control arm, or those not offered opportunity to switch

After application of findings

- May never get the treatment if data not compelling enough
- Are many more
- May get wrong treatment (if error at interim)

Three major concerns: again

- Maturity
- Information volume
- Correct estimation

Should the IDMC have a hotline to regulators, decision makers, the whole community to know if “this is going to be enough”?

It is a very high responsibility

Conclusions

- In most cases, early stopping for superiority should be avoided, unless:
  - This is the second randomized Phase III trial
  - There is truly overwhelming evidence (…)
  - Full accrual is reached and large majority of patients are off treatment
- Any balancing of interests between patients on the trial and future patients should likely be very one-sided