EORTC
European Organisation for Research and Treatment of Cancer

Innovation in late-phase trials

Jan Bogaerts
Scientific Director
EORTC

Overview

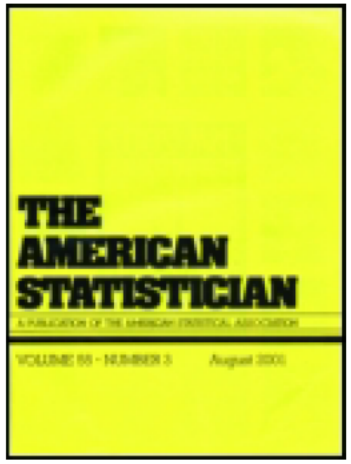
- Brief intro
- p-value crisis continues / how can we better represent collected proof or information?
  - 2 axes
  - endpoints
  - graphical methods
  - estimands
- PRO/QOL as an opportunity

About me

- Mathematician / Statistician
- 10 years in industry
- 20 years at EORTC
- Long time in RECIST
- Statistician on the EMA SAG-O
  - I represent my own opinions
- ESMO - Magnitude of Benefit Scale

- I chose some topics, but of course this is a selection

The p-value crisis is still active



Taylor & Francis
Taylor & Francis Group

The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

The ASA's statement on p-values: context, process, and purpose

Ronald L. Wasserstein & Nicole A. Lazar

2006 …

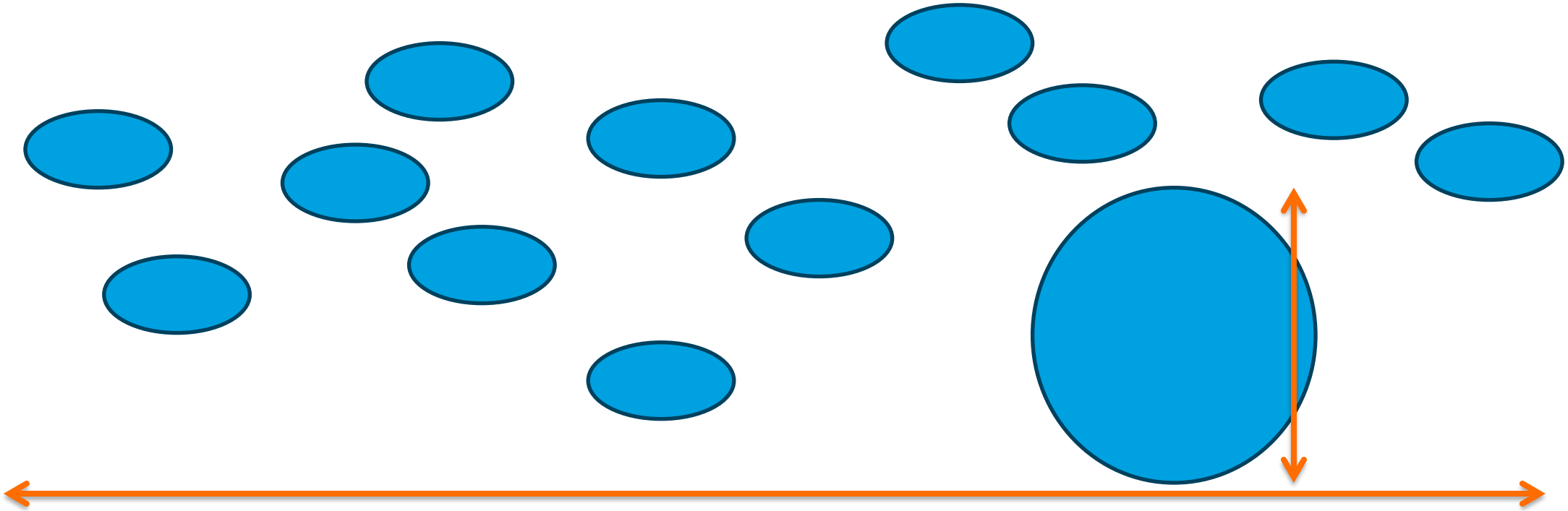
The movement of statistics from being informative to being decisional

Our problem is still very much the same

- In his influential book Statistical Methods for Research Workers (1925), Fisher proposed the level p = 0.05
  - Corneel Coens pointed this out to me

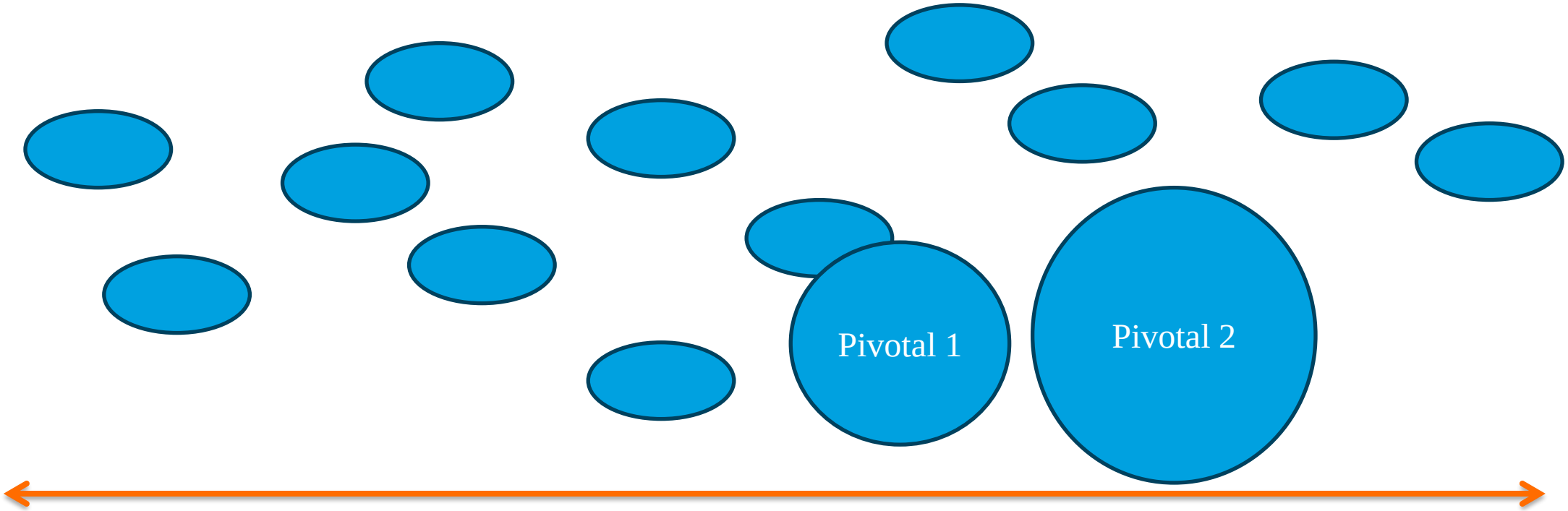
- The way we use p-values today remains very problematic
- It was certainly not intended this way

The developing body of information on a drug*

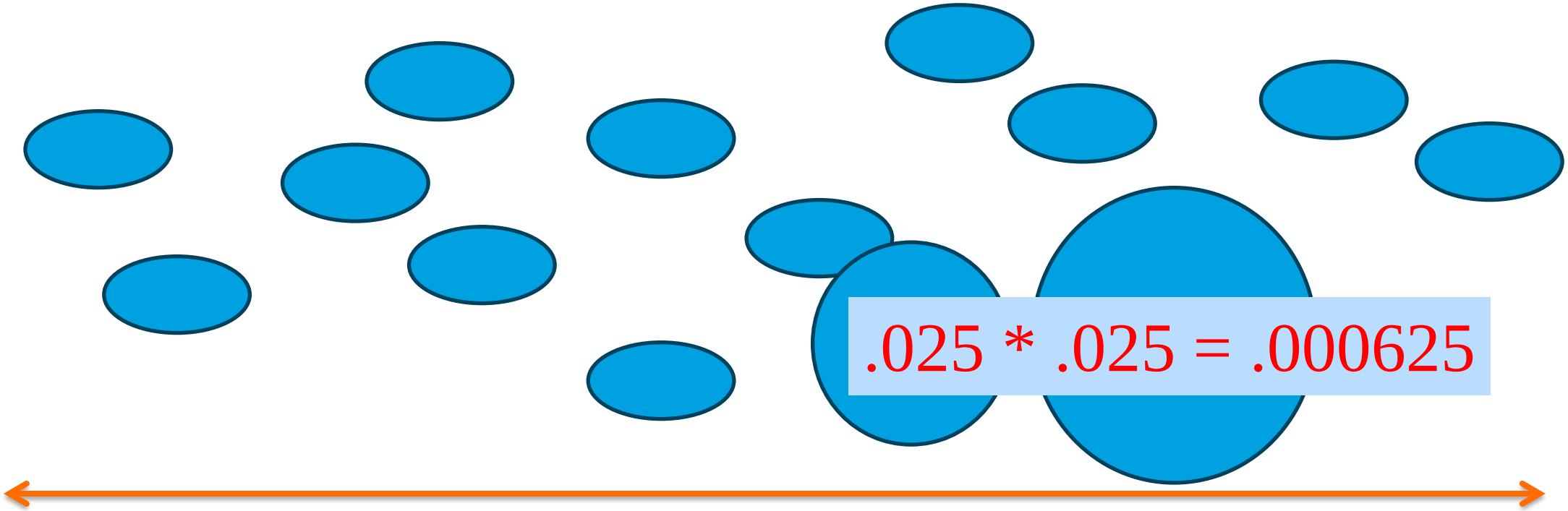


* For the purpose of this discussion. It could be much wider.

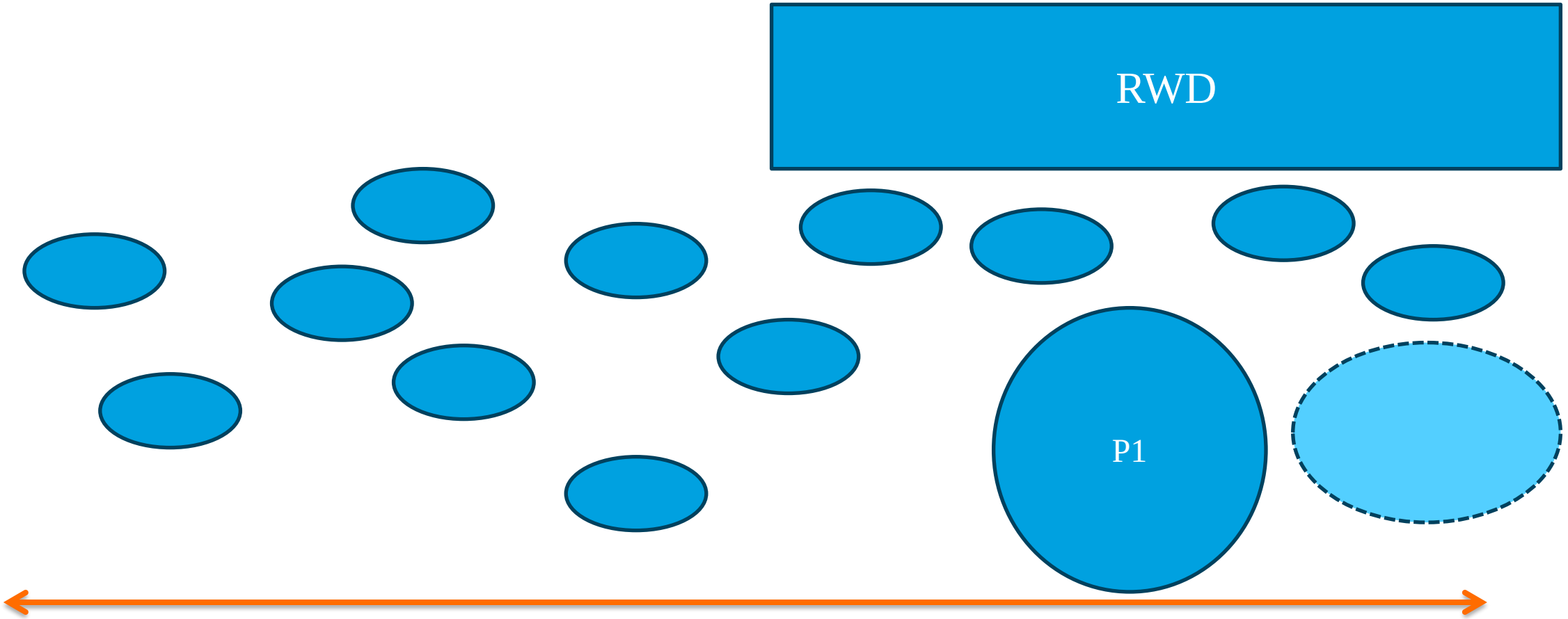
In the past …



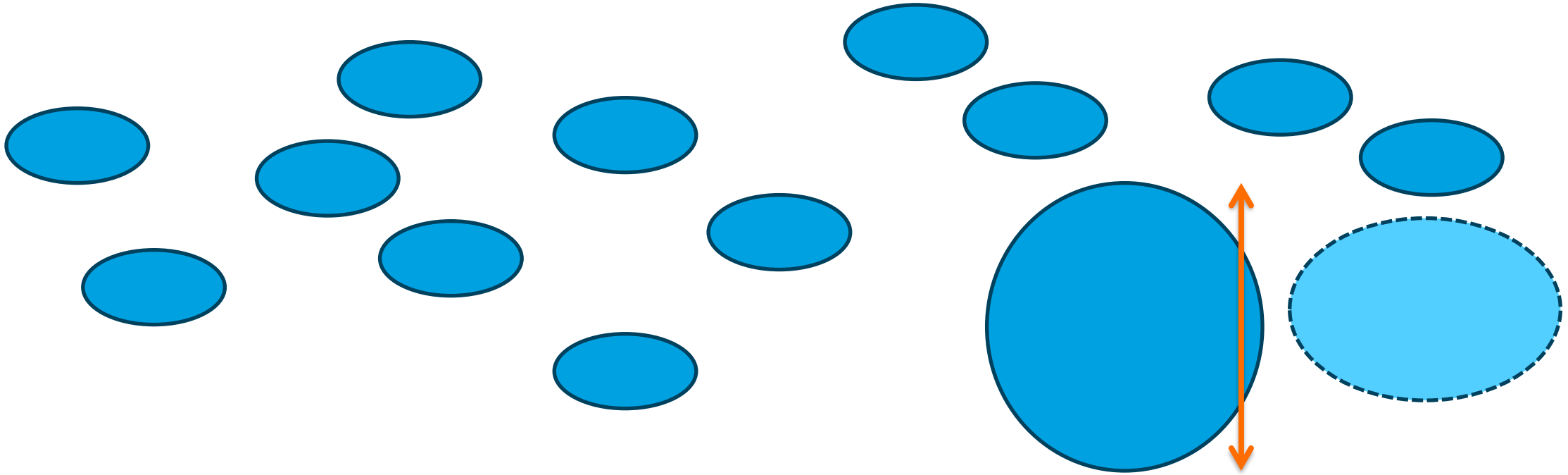
In the past …



The developing body of information on a drug*



The developing body of information on a drug*



Some scenarios for 3 endpoints

- OS is significant, but PFS and RR are equal
- PFS and RR are significant, but OS is equal (or even a bit worse)
- OS, PFS and RR are significant

- Let’s say these 3 are reasonably powered, and the statistical significance has a true clinical meaning

Some scenarios for 3 endpoints

- OS is significant, but PFS and RR are equal
- PFS and RR are significant, but OS is equal (or even a bit worse)
- OS, PFS and RR are significant

- Does it really matter which one was the ‘primary endpoint’?
  - As a statistician, I am supposed to say ‘yes’, and defend tooth and nail that the primary stands out
    - Mathematically, to make the p-values ‘function’ this is a necessity

Some scenarios for 3 endpoints

- OS is significant, but PFS and RR are equal
- PFS and RR are significant, but OS is equal (or even a bit worse)
- OS, PFS and RR are significant

- Does it really matter which one was the ‘primary endpoint’?
  - As a statistician, I am supposed to say ‘yes’, and defend tooth and nail that the primary stands out
    - Mathematically, to make the p-values ‘function’ this is a necessity

- High benefit of methods like the graphical methods (Bretz, Posch et al) to ‘structure’ the testing, paying attention to ‘what is needed’

A word about PFS

- We should pay more attention to our censoring rules
  - Missing observations should only matter as a censoring reason if they span a significant amount of time compared to the overall endpoint

- Here we have real opportunities to reap the benefits of estimands
  - Why are you censoring?
  - What is the underlying concern around the censoring rules you apply?

Last point

- I believe PRO / QOL are still a promising field to span the gap between registrational trials and HTA
- We are working very hard on this area at EORTC

Things I did not talk about, but would be happy to discuss with anyone

- RECIST current developments
  - Hyperprogression
  - Growth curve models
  - Radiomics
  - Minimal Residual disease / ctDNA
- The increasing difficulty of running trials in the post-approval space to optimize treatment
  - Pragmatic solutions
  - TWiCs

EORTC
European Organisation for Research and Treatment of Cancer

Thank you

I will be happy to have further discussions on any of this