# Randomised controlled trial designs in the setting of rare diseases: evaluation of a series of trials over a long-term research horizon

M-A Bayar[1,2], G Le Teuff[1,2], S Michiels[1,2], DJ Sargent[3], M-C Le Deley[1,2]

[1]Department of Biostatistics, Gustave Roussy, Villejuif, France,
[2]Paris-Sud University, CESP, INSERM, Villejuif, France
[3]Mayo Clinic, Rochester, MN, USA

# Background

- Randomised clinical trials: cornerstone of treatment evaluation, even in rare diseases
- Traditional clinical trial designs strive to definitively establish the superiority
  of an experimental treatment
  ⇨risk-adverse criteria and large sample sizes
- Clinical trials in rare diseases:
  A real conundrum…
- Several on-going research projects
  - Adaptive designs
  - Bayesian approach
  appealing approaches in this setting

# Study objectives

- To challenge the issue of the level of evidence requested in clinical trials
  when using a frequentist approach.
- To develop a simulation framework that
  offers a new perspective to evaluate trial designs

⇨ **How to achieve the greatest therapeutic gain from a clinical investigation of a limited number of patients?**
(Whitehead, Biometrics, 1985)

# In 2012: First step

- Extension of the work from Sposto and Stram
  A strategic view of randomized trial design
  in low-incidence paediatric cancer. Stat Med 1999

  Series of Phase III trials in paediatric cancers
  Cure rate survival model

⇨Generalisation to exponential survival models with various assumptions and disease scenarios

Taking the long view: how to design a series of Phase III trials to maximize cumulative therapeutic benefit.
Le Deley MC, Ballman K, Marandet, Sargent D.
Clinical Trials, 2012

# Moving from the first step

- First step: A new treatment was characterized by its relative efficacy to the current control treatment

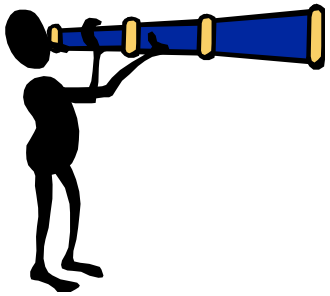
⇨ Possible accumulation of survival benefits trial after trial

?? Assumption arguable and likely overoptimistic

⇨ Current work:
Correction of the simulation framework
(Bayar, ISCB 2014, under review in Stat Med)

# Common features (1)

- Consider a trial as part of a series of trials over a long period rather than in isolation
- Assess benefits and risks on a longer research horizon



- Search for the best compromise between evidence criteria and sample size in terms of total survival benefit

# Common features (2)

## Simulation Model

- Succession of *K* two-arm trials over a 15-year horizon
- Experimental arm *E vs.* control arm *C*
- Treatment selected after each trial becomes the control of the next trial
- Primary endpoint: overall survival – logrank test

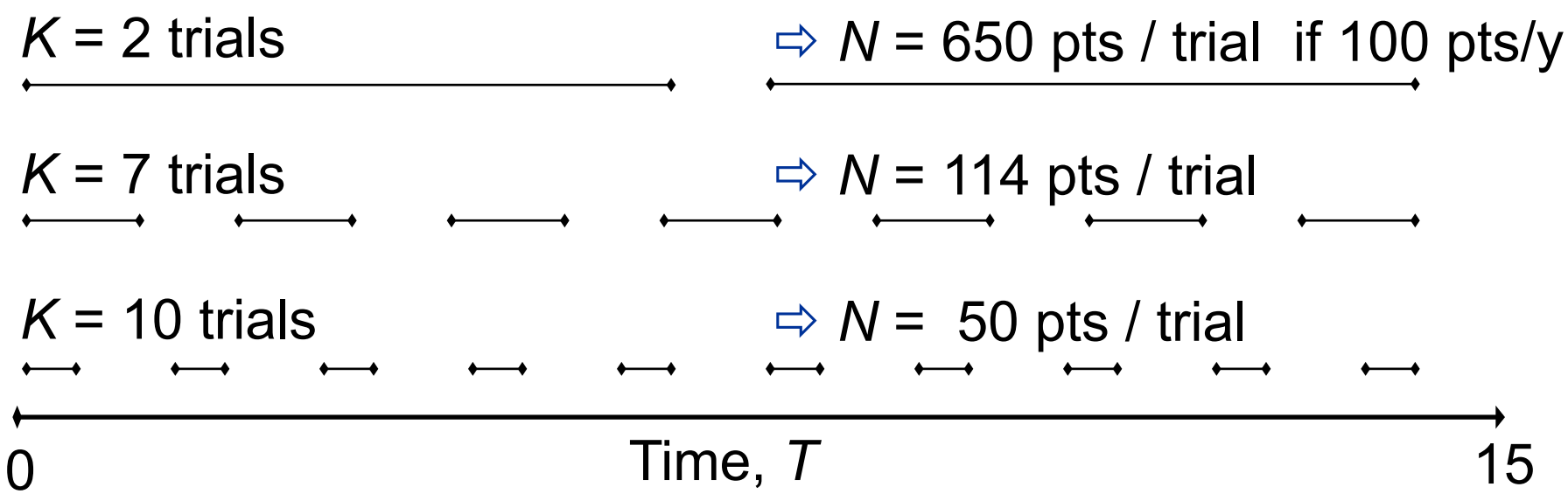
## Assumptions

- Exponential distribution of survival times for each trial *i*, $(\lambda_i^C, \lambda_i^E)$, $i \in [1,K]$
- No patient lost to follow-up
- Fixed follow-up time *FU*

# Common features (3)

## Design parameters to be evaluated

- **α-level for treatment selection**
  2.5%, 5%, 10%, 20%, 30%, 40%, 50%
- **Trial sample size, N,** derived from
  the number of trials K run over a 15-year period



# Common features (4)

## Performance metrics

**For a series of trials**

at the end of the 15-year research period

- Overall hazard ratio, $HR\downarrow o = \lambda\downarrow K+1\uparrow C \,/\lambda\downarrow 1\uparrow C$
- Total survival benefit, $Benefit = [1/HR\downarrow o \; -1]$

**Simulation of 5000 repetitions**

of the 15-year research period

⇨Expected total survival benefit, $E[Benefit]$

⇨Probability of a detrimental effect, $P[HR\downarrow o > 1]$

# Common features (6)

## Scenarios of the underlying disease

### Disease severity

Survival in the control arm at the beginning of the 15 years

| Survival | | *FU* |
|---|---|---|
| Median survival of 6 months | | 6 months |
| Median survival of 1 year | | 1 year |
| Median survival of 2 years | | 2 years |
| 2-year OS of 75% | | 2 years |

**Accrual rate**: 50, 100 and 200 patients per year

# Main difference: Treatment effect

## In the previous work (Le Deley, Clin Trials, 2012)

New treatment effect characterized by
its **relative efficacy** to the current control treatment

**Hazard ratio** drawn from a hypothetical distribution

## In the current work (Bayar, under review)

New treatment effect characterized by
its associated hazard rate

**Hazard rate** drawn from a hypothetical distribution

# Previous work: e.g., 5 successive simulated trials

Truth

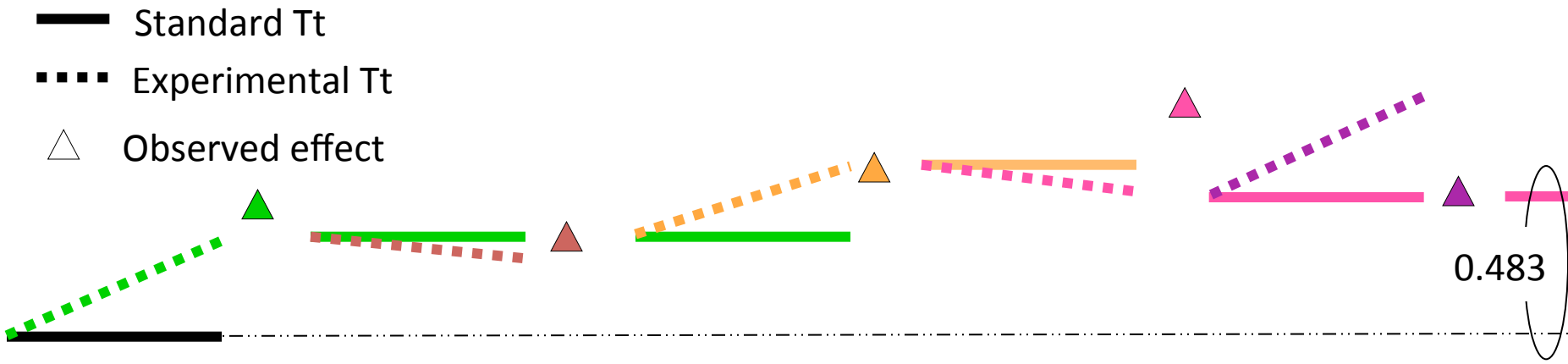
$HR_1 = 0.6$ $HR_2 = 1.1$ $HR_3 = 0.7$ $HR_4 = 1.15$ $HR_5 = 0.5$

Observed data

$\hat{HR}_1 = 0.5$ $\hat{HR}_2 = 1.01$ $\hat{HR}_3 = 0.75$ $\hat{HR}_4 = 0.8$ $\hat{HR}_5 = 0.98$

Conclusion of the trial

+ (True+) Neg (True -) + (True +) + (False +) Neg (False -)



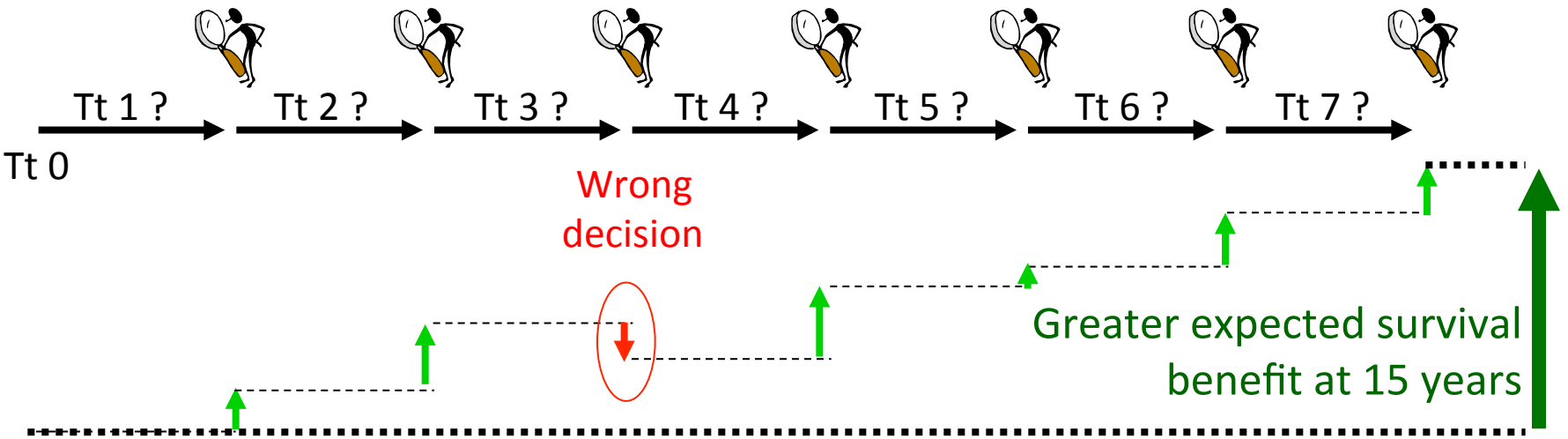

Overall Hazard Ratio measuring the Gain at the end of the five trials

*HRo* = 0.6 x 1 x 0.7 x 1.15 x 1

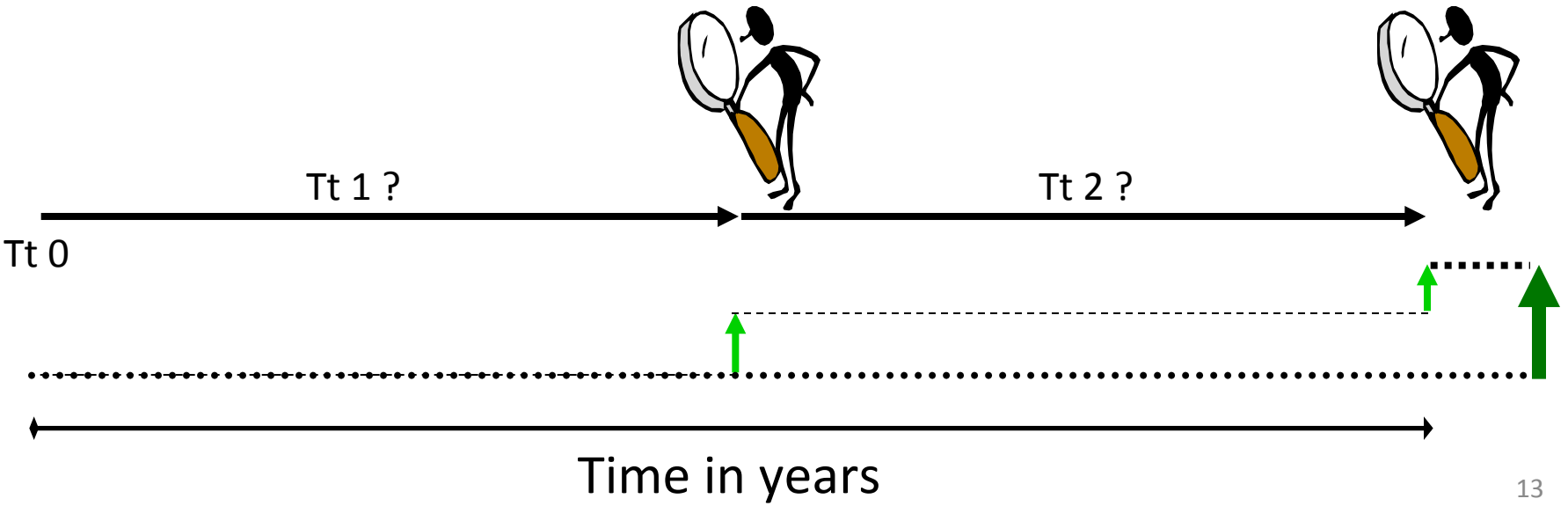
= 0.483

# ⇨Possible accumulation of survival benefits trial after trial

Optimal: 7 smaller trials (114 pts) and relaxed evidence criteria (alpha = 20%)




Traditional: 2 large trials (650 pts) and stringent evidence criteria (alpha = 2.5%)



# Hypothetical treatment effect (1)

## In the current work

$\lambda\downarrow T\downarrow k\ \uparrow E$ **hazard rate** of the experimental arm in the trial k, initiated at time $T\downarrow k$ , $T\downarrow k \in [0,15[$

derived from the expected distribution of $\mathrm{HR} \sim logN$

$\Rightarrow \ln[\lambda\downarrow T\downarrow k\ \uparrow E \downarrow \uparrow] \sim \mathcal{N}(\mu(T\downarrow k), \sigma^2)$

With $\mu(T\downarrow k) = a \times T\downarrow k + b$

and assuming a fixed scale parameter $\sigma^2$

Parameters a, b et $\sigma^2$ characterize the expected treatment distribution

# Hypothetical treatment effect (2)

## In the current work: four hypothetical distributions

**D1: "Historical" distribution**

derived from the meta-analysis of 698 RCT, >200 000 pts, performed by Djulbegovic et al. (Cochrane 2012)

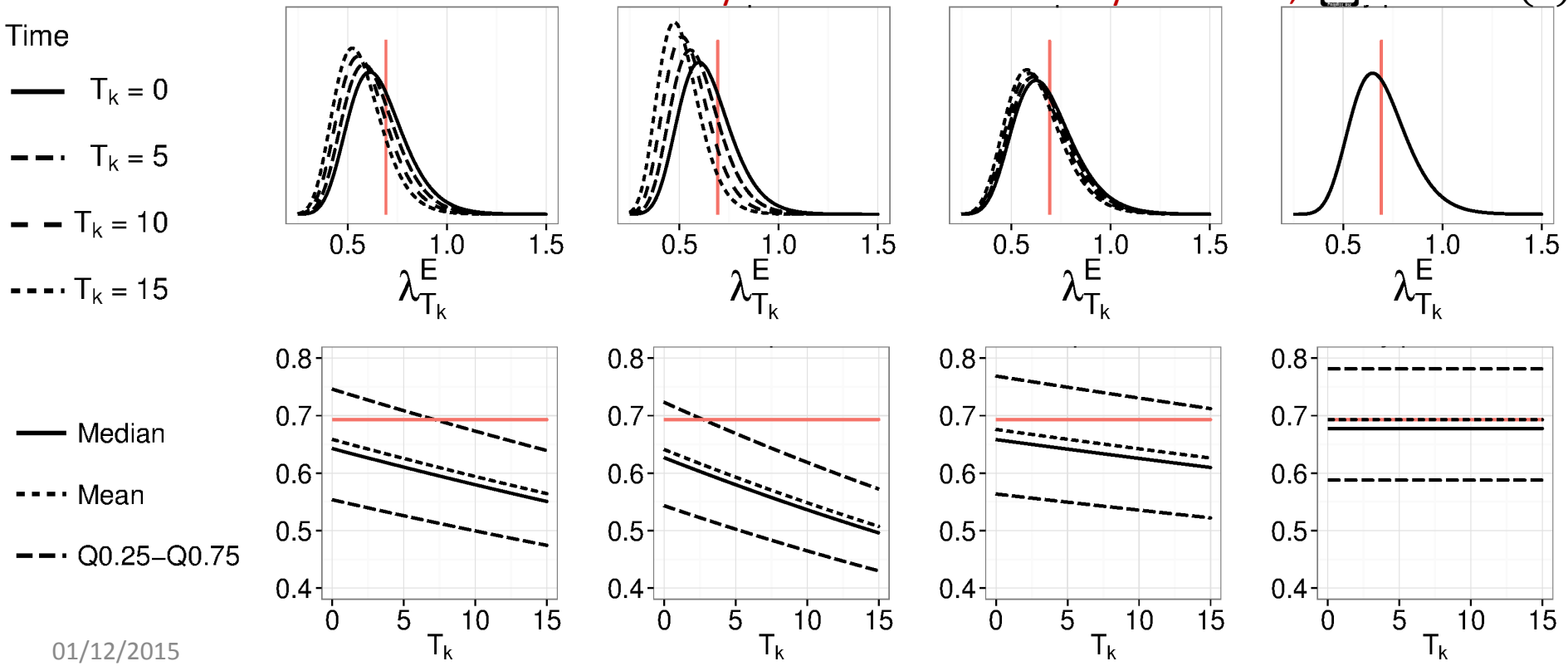
- $\mathrm{HR} \sim logN$
- $\mathrm{E}(\mathrm{HR}) = 0.95$
- Probability of a breakthrough $\mathrm{P}[HR < 0.5] = 0.02$

**Other distributions +/- optimistic or pessimistic**

# Hypothetical treatment effect (3)

| Distribution | D1: historical distribution | D2: more optimistic | | D4: very pessimistic |
|---|---|---|---|---|
| | 0.95 | 0.925 | 0.975 | 1 |
| Probability of breakthrough | 0.02 | 0.02 | 0.01 | 0.01 |

For a disease scenario characterised by a median survival of 1 year at T=0, $\lambda_1 = \ln(2)$

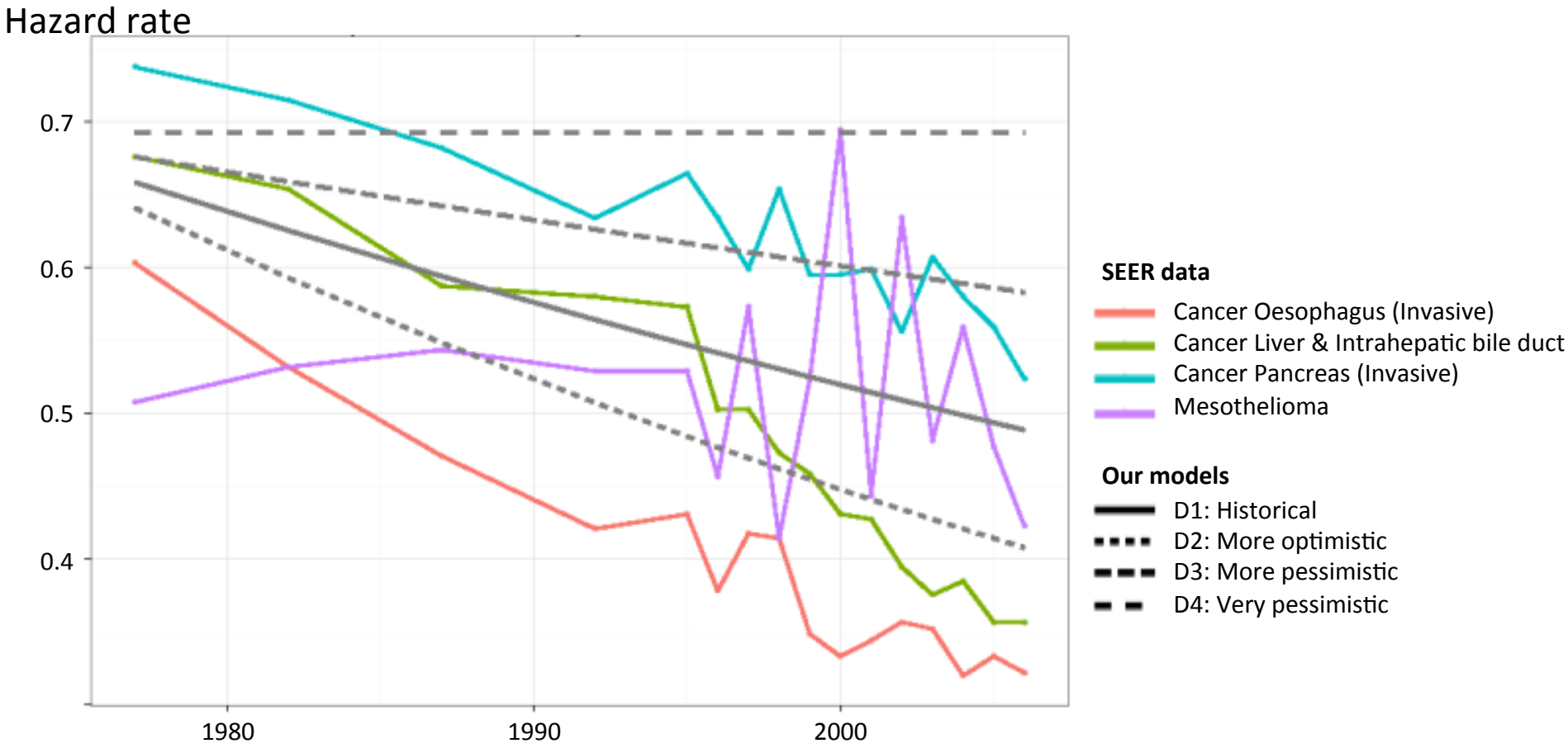


# Hypothetical treatment effect (4)

## Empirical validation of our assumption

SEER data plotted against our modeling

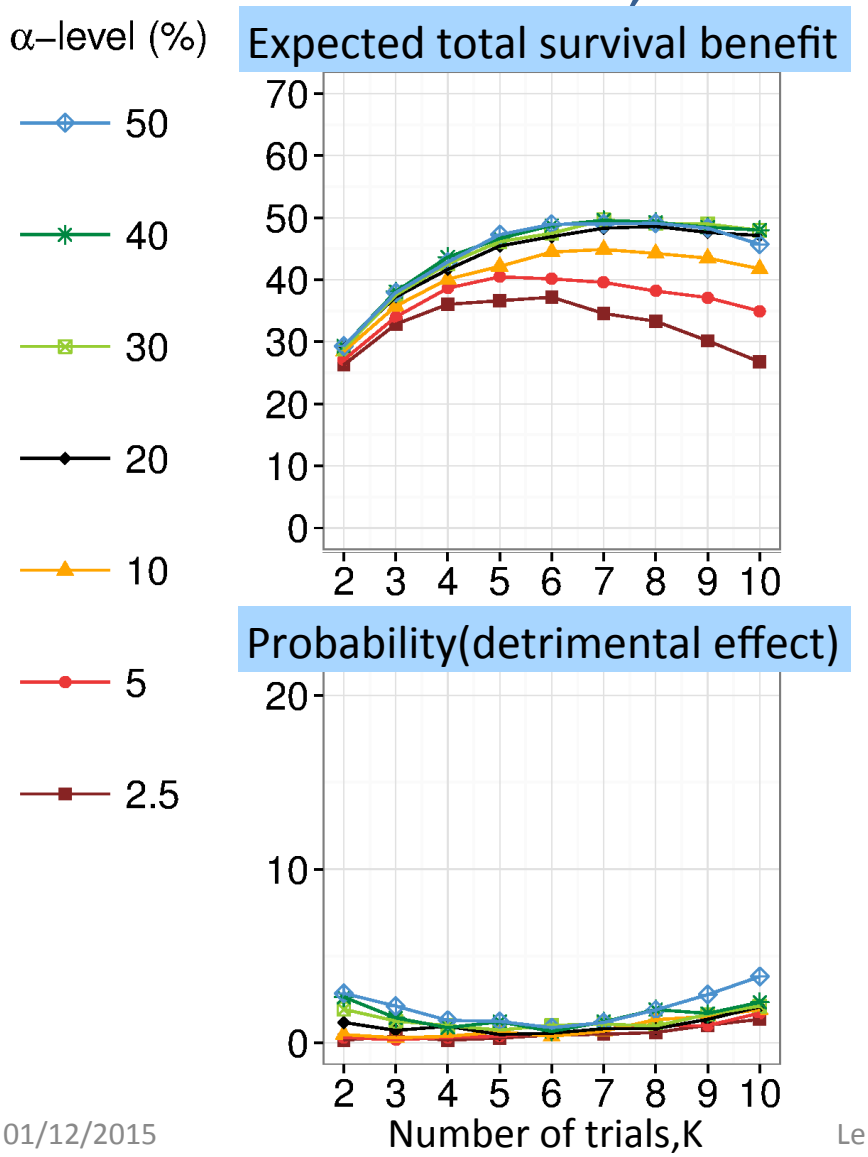
E.g. for a disease characterized by a 1-year median survival



# Results – From one situation

## Disease scenario: 100 patients/year, median survival = 1 year

## Historical distribution, D1




**Expected total survival benefit**

- Increases with increasing α-level
- Little additional gain for α>20%
- Large impact of α-level when the number of trials increases (small sample size)
- Non monotonous increase with K
- Interplay between K and α-level

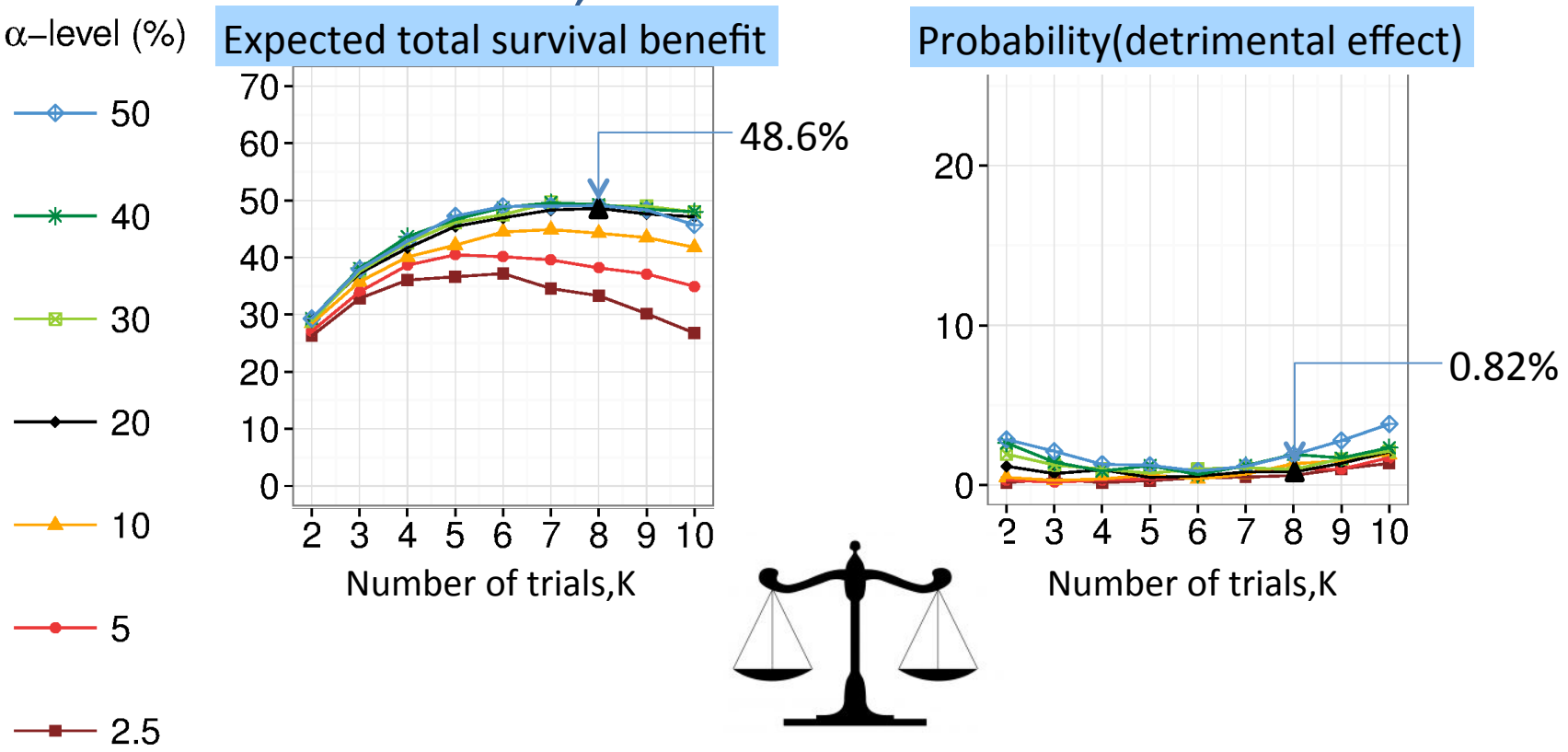
**Probability(detrimental effect)**

- Increases with increasing α-level
- Increases with K for α-level 2.5-20%

# Results – From one situation

## Disease scenario: 100 patients/year, median survival = 1 year

## Historical distribution, D1




## Optimal design parameters ?

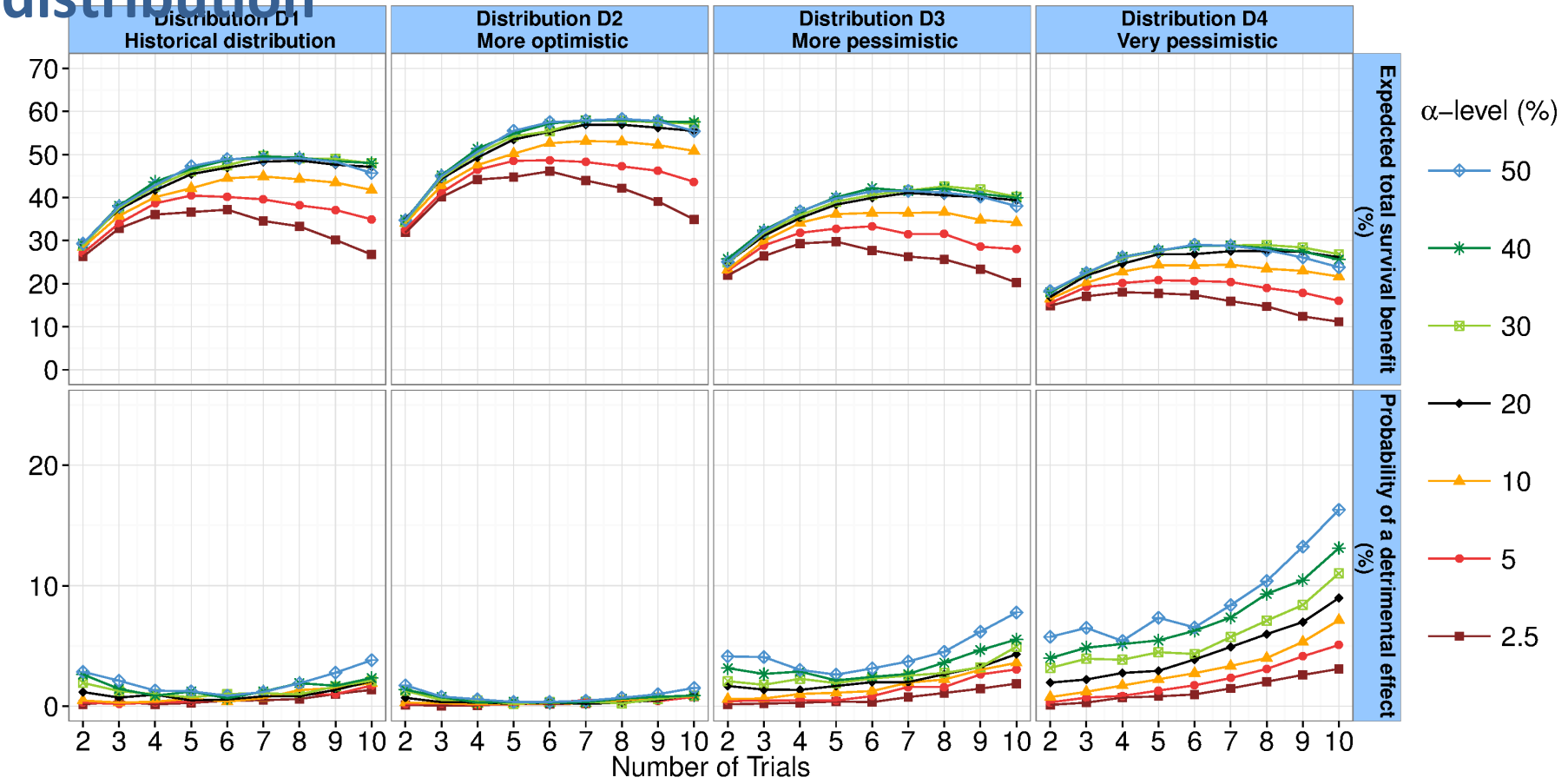
α-values & trial sample size maximizing expected total survival benefit provided that the probability of a detrimental effect remains <1%

For this scenario: α-values=20% and K=8 (⇨N=88)

# Results – Generalisation

Disease scenario: 100 patients/year, median survival = 1 year

**Results according to the hypothetical treatment distribution**



# Results – Comparison to traditional design

Disease scenario: 100 patients/year, median survival = 1 year

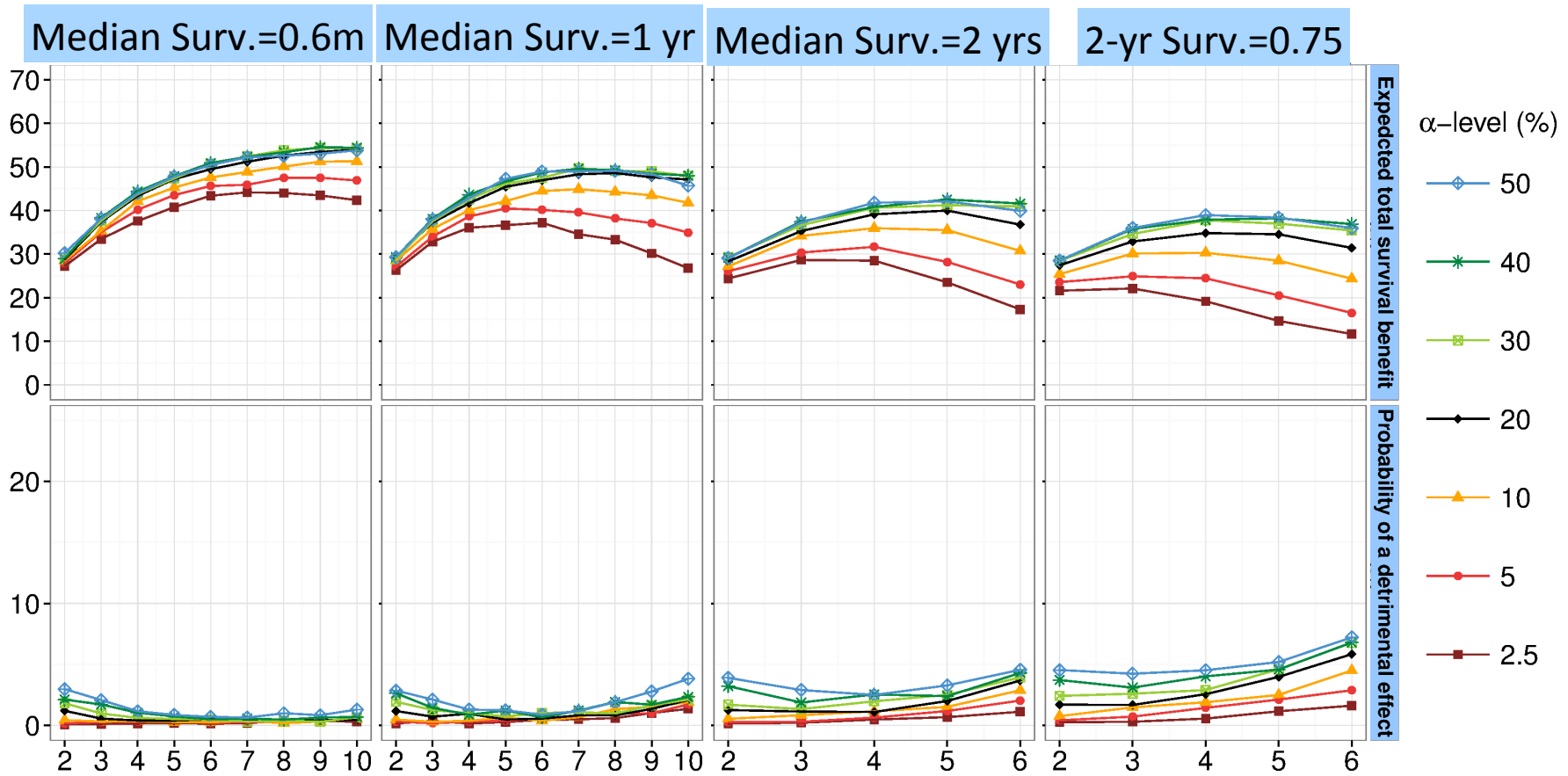
**Results according to the hypothetical treatment distribution**

| Treatment effect distribution | Optimal design parameters | | | | | Traditional design α=2.5%, K=2, N=650 | |
|---|---|---|---|---|---|---|---|
| | α | K | N | E(Benefit) % | P(HRo>1) % | E(Benefit) % | P(HRo>1) % |
| D1: historical | 20% | 8 | 88 | 48.6 | 0.82 | 26.3 | 0.16 |
| D2: more optimistic | 20% | 7 | 114 | 56.9 | 0.20 | 31.9 | 0.12 |
| D3: more pessimistic | 5% | 6 | 150 | 33.3 | 0.84 | 21.9 | 0.16 |
| D4: very pessimistic | 5% | 4 | 275 | 20.1 | 0.86 | 14.9 | 0.12 |

If parameters are defined under D1 whereas the true treatment effect is D4
E(Benefit) = 27.6%  and $P(HR_o>1) = 5.98\%$

# Results – According to disease severity

Disease scenario: 100 patients/year,
**Under the "historical" treatment effect distribution D1**



# Main conclusion

Under reasonably optimistic assumptions regarding the future treatment effects

**optimal designs** with

- reduced sample size and
- relaxed $\alpha$-level

**outperform traditional ones** when

- disease is severe (baseline median survival ≤ 1 yr)
- accrual is ≥ 100 patients/year

No major improvement is observed in diseases with a better prognosis and/or very low accrual

# Discussion

## What is new compared to the previous work?

(Le Deley, Clinical Trials, 2012)

- Similar pattern and conclusions
- Expected survival benefit smaller,
  due to the correction of a possible bias in the previous work
- Current framework now allows the evaluation of more complex designs addressing the issue of making the best use of a limited number of patients over time.

# Discussion

## Sensitivity analyses

- Impact of reducing the variance parameter of the treatment effect distribution
  - Under D1, D2: smaller benefit but recommendations relatively stable
  - More conservative recommendations under D3, D4
- Impact of changing the length of the research horizon
  - General pattern is enhanced with longer research horizon

# Discussion

## What does this simulation study illustrate?

- The traditional sample size calculation using standard evidence criteria can be challenged in rare diseases.
- Treatment effect is not limited to the punctual null and alternative hypotheses.
- Evaluation of operating characteristics of RCT based on the whole hypothetical distribution of future treatment effects.
- Our perspective may be helpful to define a better trade-off between false-positive and false-negative risks specified at the trial level.

# Discussion

Reducing the trial sample size is not an objective per se but the consequence of performing more trials; the reduced performance of each trial considered in isolation is then corrected by the increased number of new evaluated treatments.

Our recommendation is only valid when considering a series of trials run over a relatively long research horizon. and when the supply of new treatments is large.

# Discussion

## Perspectives

- In the current simulation framework, fixed sample sizes, no interim analysis identical through the series of trials

⇨On-going work to extend the study to evaluate more flexible designs, incorporating interim analyses

⇨How can we make the best “use” of limited number of patients?

⇨Evaluation of adaptive designs

# Discussion

## Perspectives

- Reducing the level of evidence at the level of a single trial may be acceptable if reliable external data are available for a Bayesian design.
- In the current framework, trials are independent and we ignore external data.

⇨This simulation framework should also allow us to evaluate Bayesian designs were information from the previous trial is borrowed to better estimate the relative treatment effect of the new experimental treatment (information on the control arm).

# Thank you for your attention!

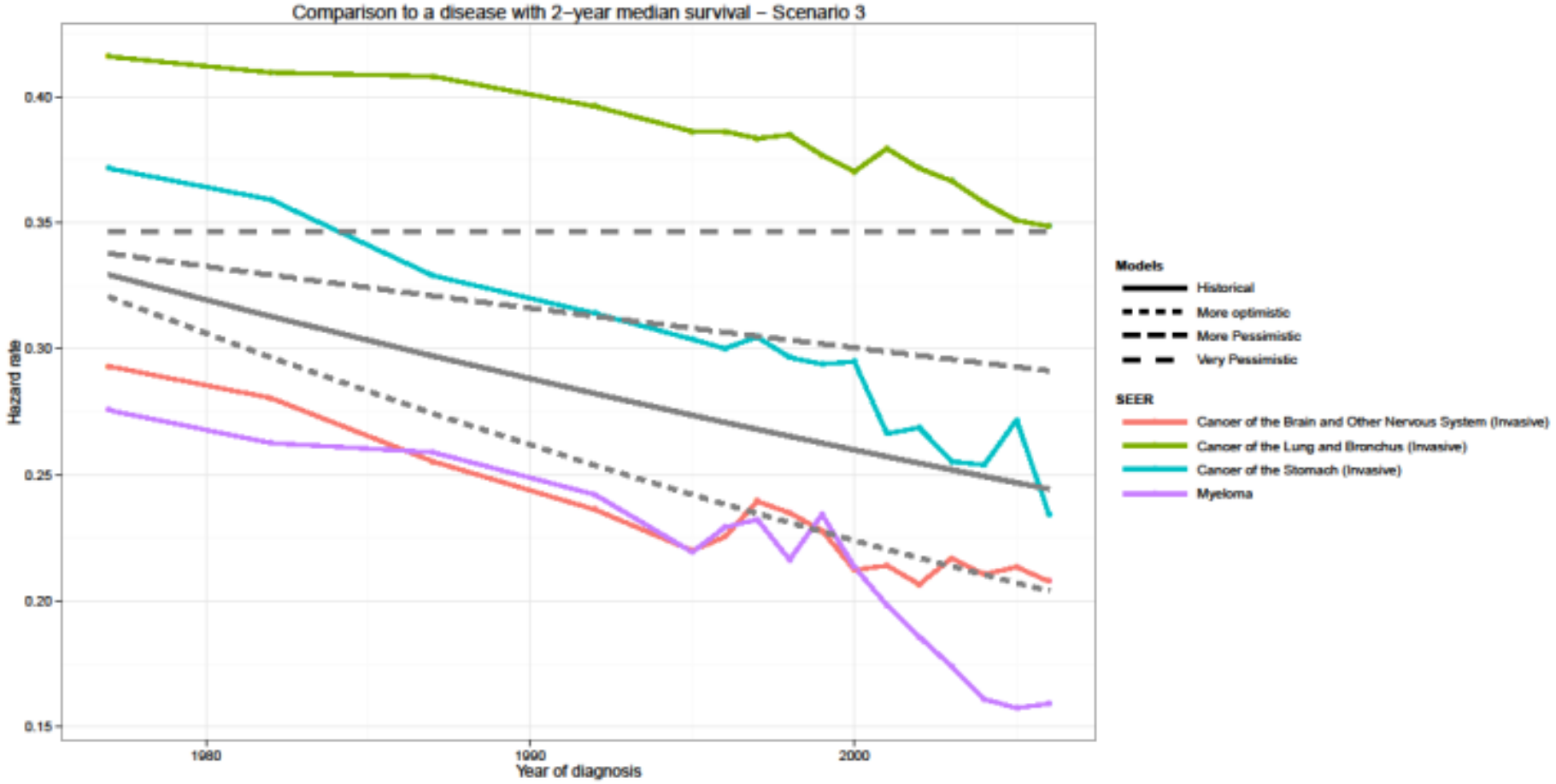
# BACK-UP SLIDES

# Hypothetical treatment effect (5)

## Empirical validation of our assumption

SEER data plotted against our modeling

E.g. for a disease characterized by a 2-year median survival

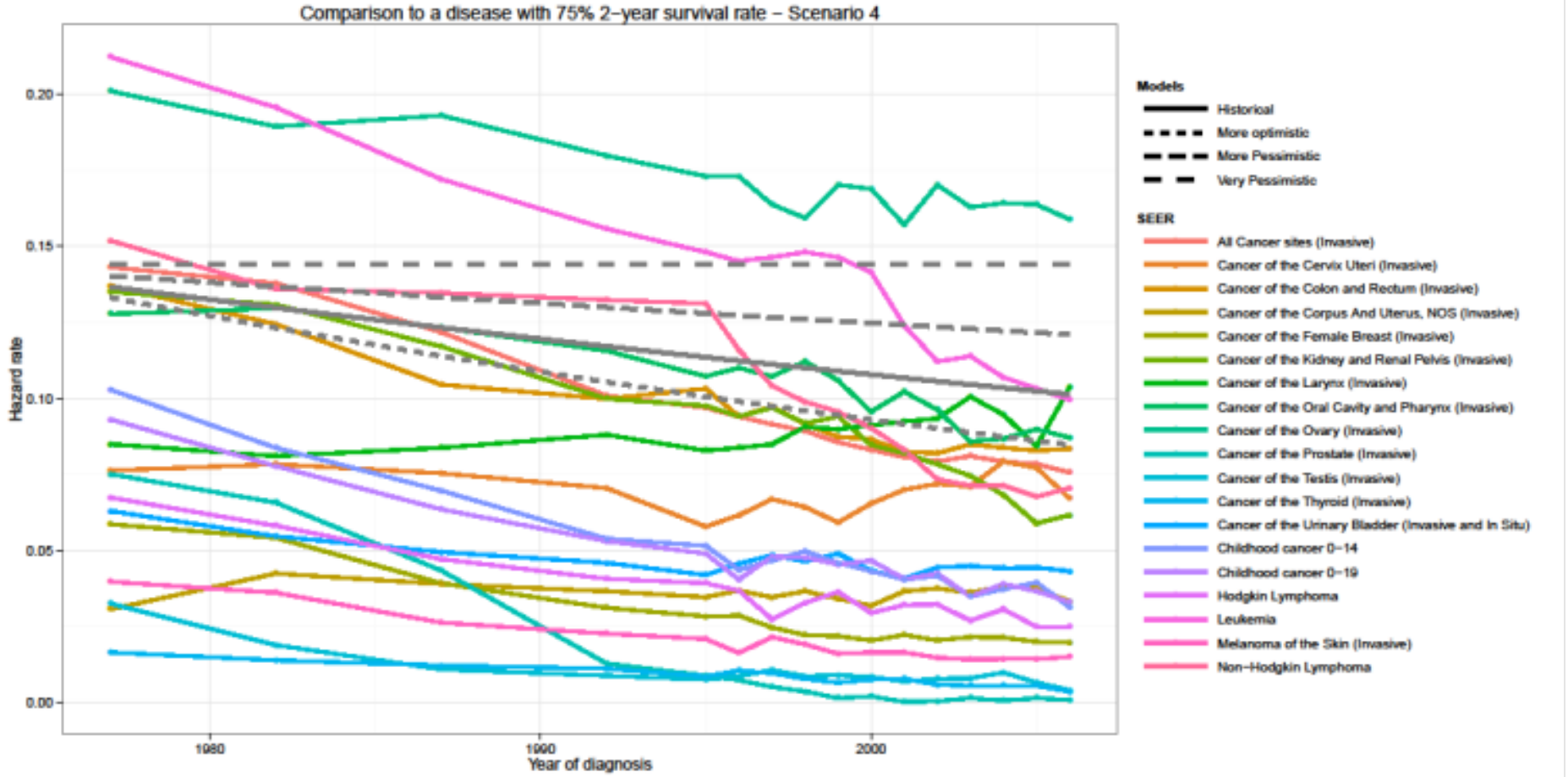


# Hypothetical treatment effect (6)

## Empirical validation of our assumption

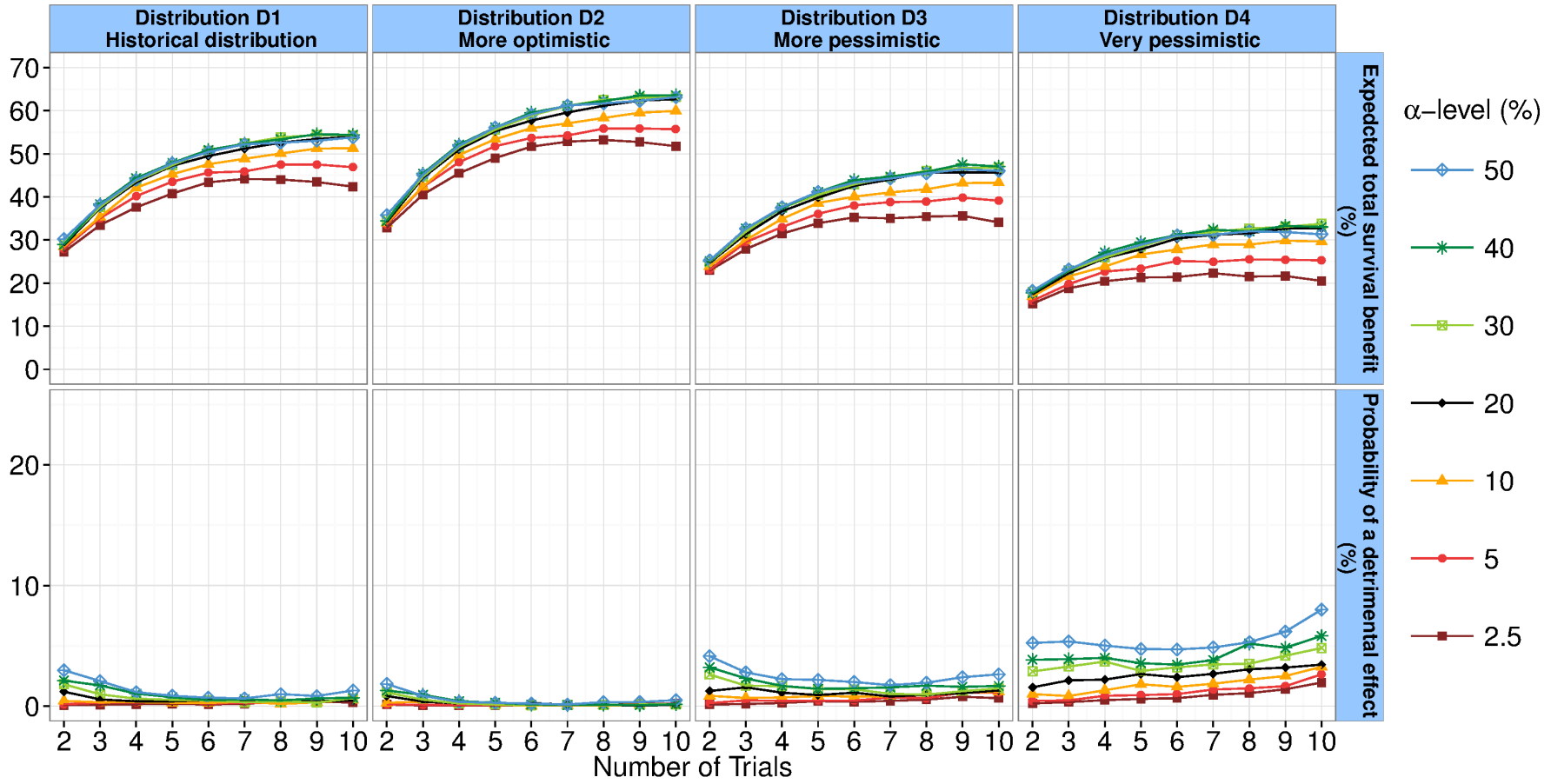
SEER data plotted against our modeling

E.g. for a disease characterized by a 2-year of 75%



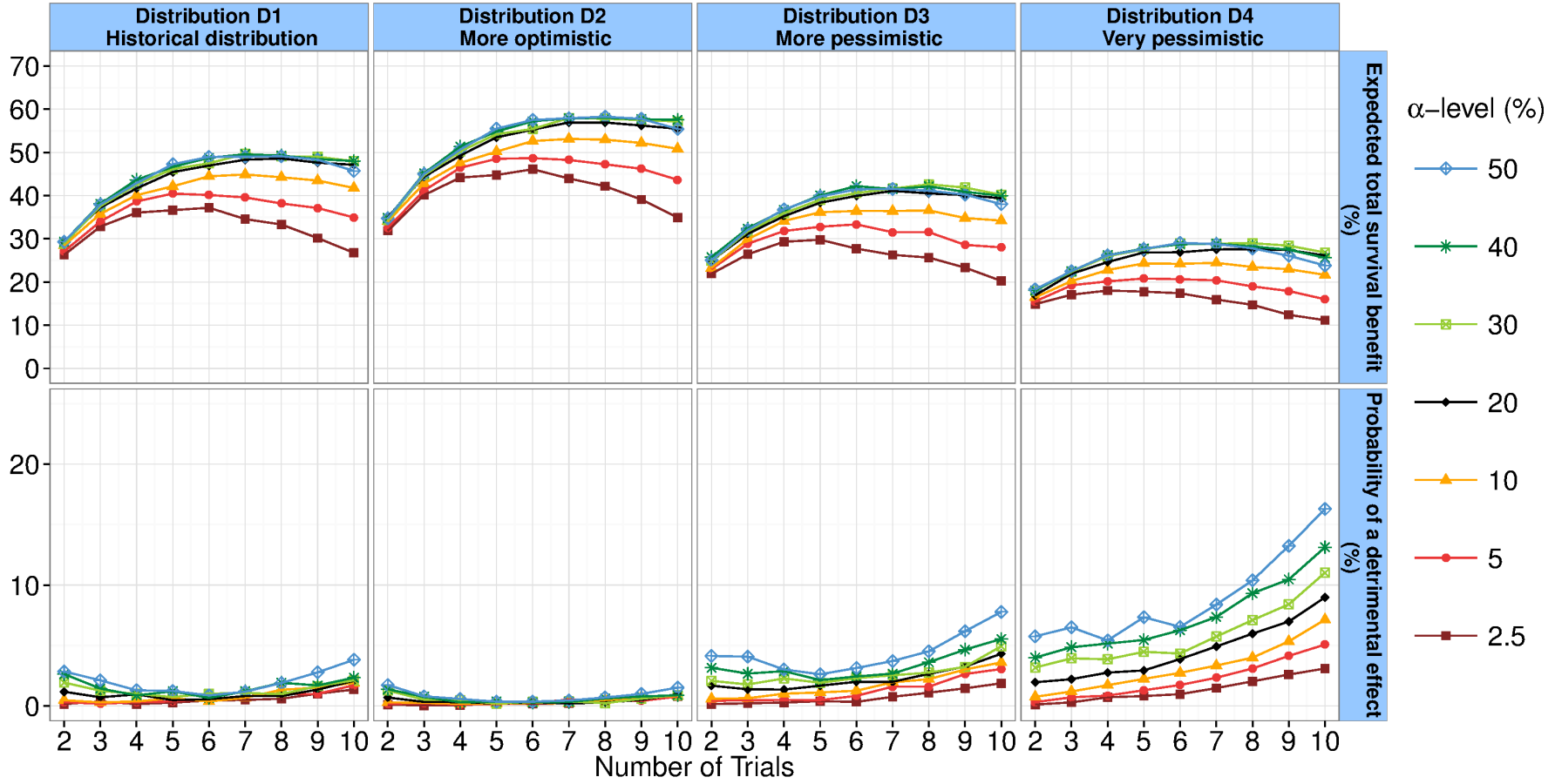
# Results

## Disease scenario: 100 patients/year, median survival = 6 mo
## Results according to the hypothetical treatment distribution



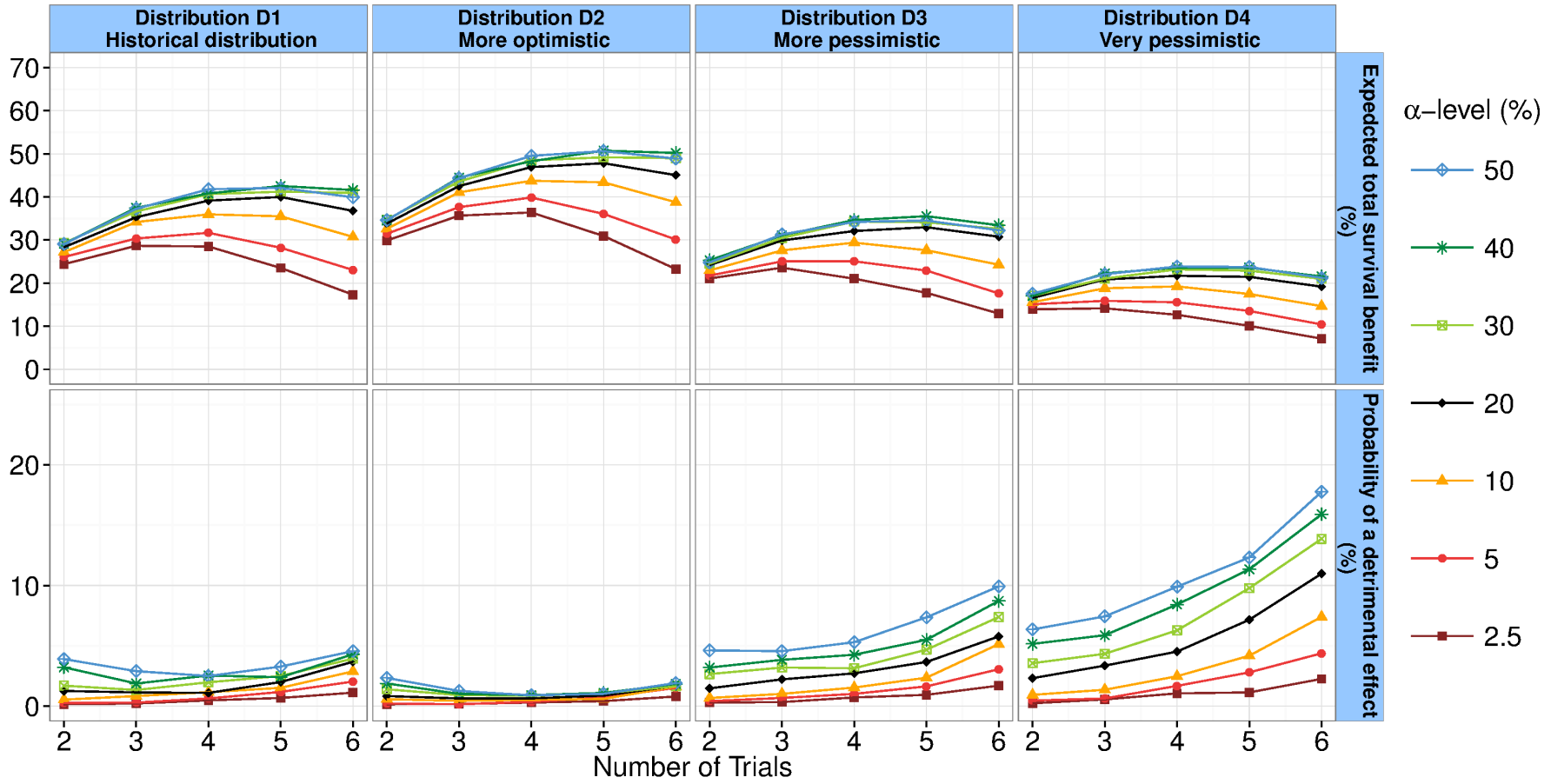
# Results

## Disease scenario: 100 patients/year, median survival = 1 year
## Results according to the hypothetical treatment distribution



# Results

## Disease scenario: 100 patients/year, median survival = 2 years
## Results according to the hypothetical treatment distribution



# Results

## Disease scenario: 100 patients/year, 2-year survival = 75%
## Results according to the hypothetical treatment distribution