The influence of design elements on Bayesian outcome-adaptive randomization

The issue of considering differential hypothesis test criteria.

L. Garcia Barrado[1] T. Burzykowski[1,2]

[1]Interuniversity Institute for Biostatistics and Statistical Bioinformatics (I-BioStat)
[2]International Drug Development Institute (IDDI)

Correspondence: Leandro.garciabarrado@uhasselt.be

UHASSELT I-BioStat KU LEUVEN

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Outline

- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- Simulation Study
- Results
- Conclusion

Outline

- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- Simulation Study
- Results
- Conclusion

Introduction

- Idea to design a clinical trial
  - 2 treatments
    - Control (t=0)
    - Experimental (t=1)
  - 2 biomarker strata
    - Biomarker negative (s=0)
    - Biomarker positive (s=1)
  - Binary clinical outcome (response y)

Measured by perfect assay ($Se_A = Sp_A = 1$)

$$P(S = 0|B = 0) = P(S = 1|B = 1)$$

$$P(S = 1) = P(B = 1)$$

Goal is to identify stratum-treatment combinations which have an acceptable outcome probability.

Outline

- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- Simulation Study
- Results
- Conclusion

Methods

Bayesian biomarker-based outcome-adaptive phase-II study design

- Model
- Adaptive trial characteristics
- Decision rule(s)

(Zhou et al. 2008 [BATTLE]; Kim et al. 2011 [BATTLE]; Barry et al. 2015; Gu et al. 2016 [BATTLE-2])

Methods (Model)

- Consider the following hierarchical probit model

$$\pi_{ist} \equiv P(y_{ist} = 1) = P(z_{ist} > 0) = \Phi(\mu_{st})$$

$$\mu_{st} \sim N(\phi_t, \sigma^2)$$

$$\phi_t \sim N(\alpha, \tau^2)$$

- With
  - $z_{ist} \sim N(\mu_{st}, 1) =$ latent, normally distributed random variable
  - $y_{ist} =$ response of patient i in stratum s treated with treatment t $\quad (y_{ist} = 1$ if $z_{ist} > 0)$
  - $\pi_{ist} =$ probability of positive response

  - Hyperparameters α, σ^2, and τ^2

Methods (Adaptive trial characteristics)

- Biomarker-status established by perfect assay
- Adaptive randomization starts after n_0 1:1 randomized patients
- Randomization ratios adapted according to 'max-mapping' strategy

$$r_{st,n} = P\left(\prod_{\substack{t' \neq t \\ t' \in \Omega_{s;n}}} \mu_{st,n} > \mu_{st',n} | y_n\right)$$

Ratio $r_{st,n}$ is equal to the posterior probability that after n patients, treatment t is superior to all other treatments still under consideration in stratum s

Methods (Decision rule(s))

Goal is to identify stratum-treatment combinations which have an acceptable outcome probability.

Methods (Decision rule(s))

Goal is to identify stratum-treatment combinations which have an acceptable outcome probability.

- Define
  - $\pi_1 =$ target response probability
  - $\pi_0 =$ undesirable response probability

- Allow stopping for futility during the trial after each patient
- Test efficacy only at the end of the trial

Methods (Decision rule(s))

- Bayesian hypothesis test of futility

$$F_{st,n} = \begin{cases} 1 \; if \; P(\mu_{st} \geq \Phi^{-1}(\pi_1) | y_n) \leq \delta_l \\ 0 \; otherwise, \end{cases}$$

Issue!
Both could hold for certain posteriors

Stop treating patients in biomarker stratum s with treatment t when the posterior probabilty of a positive response exceeding target π_1 is smaller than small probability δ_l

- Bayesian hypothesis test of efficacy

$$S_{st,n} = \begin{cases} 1 \; if \; P(\mu_{st} \geq \Phi^{-1}(\pi_0) | y_n) > \delta_u \\ 0 \; otherwise, \end{cases}$$

Consider treatment t in biomarker stratum s efficacious when the posterior probabilty of a positive response exceeding undesirable rate π_0 is larger than large probability δ_u

Methods (Decision rule(s))

- Alternatively consider (see Berry et al. 2010)
  - $\pi_0 =$ undesirable response probability
  - $\beta =$ clinically meaningful improvement over π_0

$$F_{st,n} = \begin{cases} 1 \; if \; P(\mu_{st} \geq \Phi^{-1}(\pi_0 + \beta)|y_n) \leq \delta_l \\ 0 \; otherwise, \end{cases}$$

$$S_{st,n} = \begin{cases} 1 \; if \; P(\mu_{st} \geq \Phi^{-1}(\pi_0 + \beta)|y_n) > \delta_u \\ 0 \; otherwise, \end{cases}$$

Outline

- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- **Simulation Study**
- Results
- Conclusion

Simulation study (Example from Barry et al 2015)

- $P(B = 1) = P(S = 1) = 0.5$
- Underlying true response probabilities P_{st}

		Treatment	
		t=0	t=1
Biomarker status	s=0	$P_{00} = 0.25$	$P_{01} = 0.25$
	s=1	$P_{10} = 0.25$	$P_{11} = 0.5$

- $\pi_0 = 0.25$
- $\pi_1 = 0.5$
- $\beta = 0.1$
- $\delta_l = 0.025$
- $\delta_u = 0.9$

Simulation study (Example from Barry et al 2015)

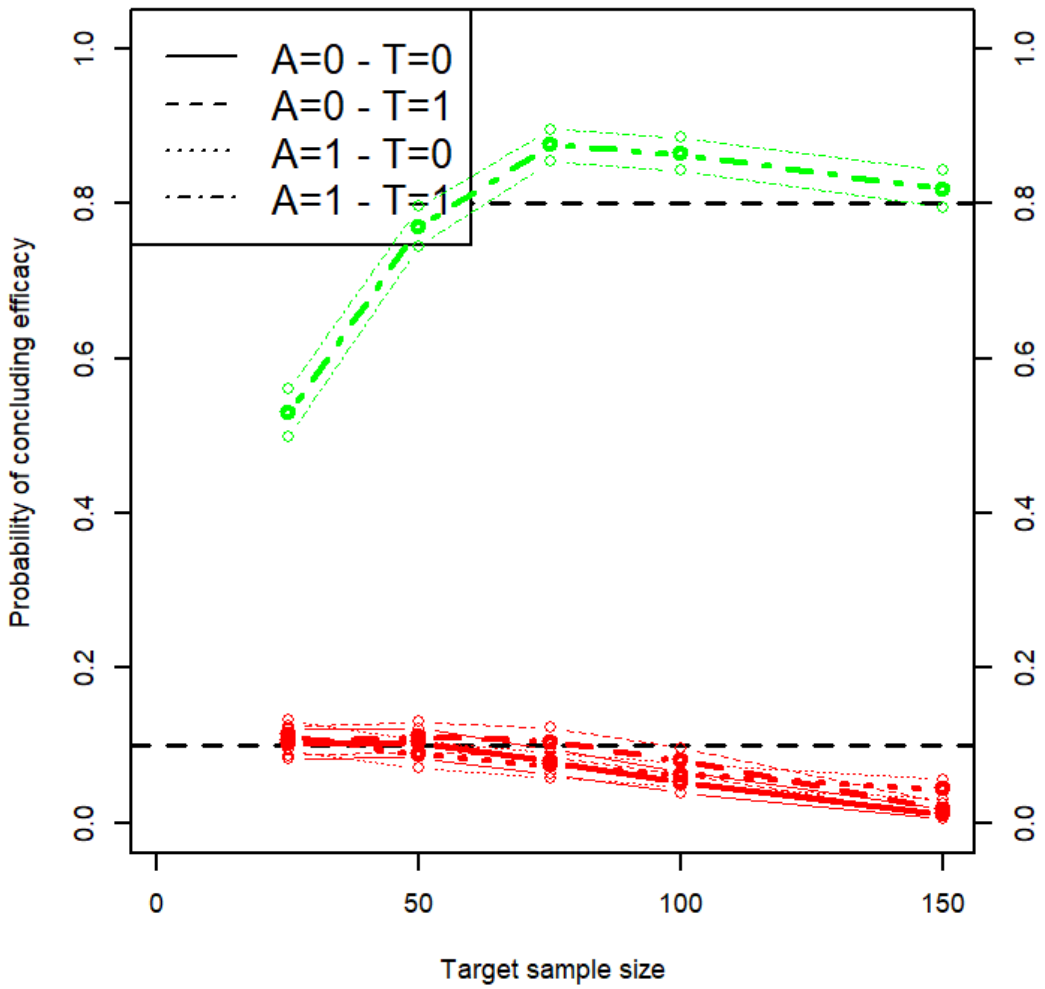
- Priors
  - Conservative prior for test for futility $(\sigma^2 = 0.1; \tau^2 = 0.1; \alpha = \Phi(\pi_1))$
  - Non-liberal prior randomisation ratio update $(\sigma^2 = 1; \tau^2 = 0.01; \alpha = \Phi(\pi_0) + \Phi(\pi_1)/2)$
  - Flat prior for test for efficacy $(\sigma^2 = 1; \tau^2 = 0.01; \alpha = \Phi(0.5))$

- Interest in effect on 'frequentist' power and type-I error probability
  - $N_{max} \in (25,50,75,100,150)$
  - $N_0 = 25$
  - 1000 simulated trials for each N_{max}

Outline

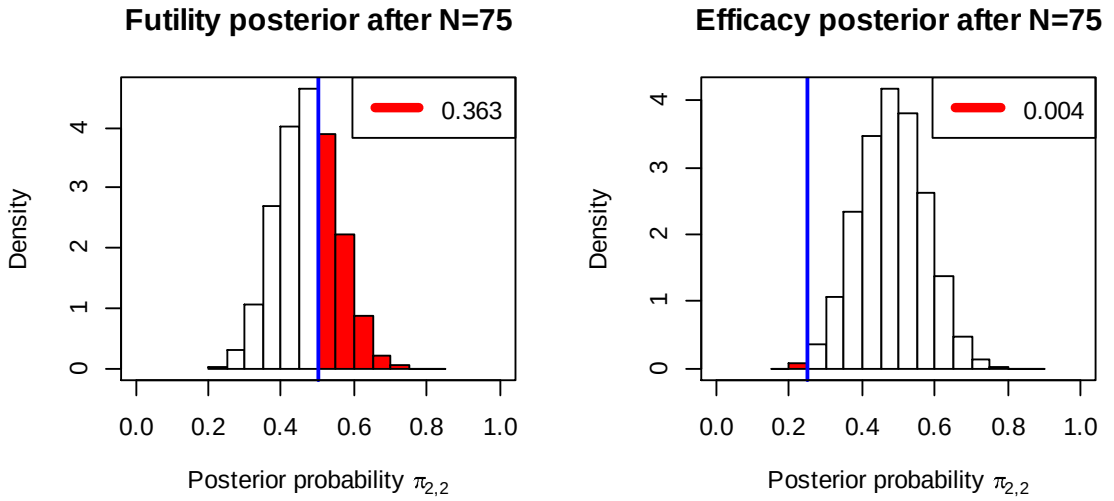
- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- Simulation Study
- Results
- Conclusion

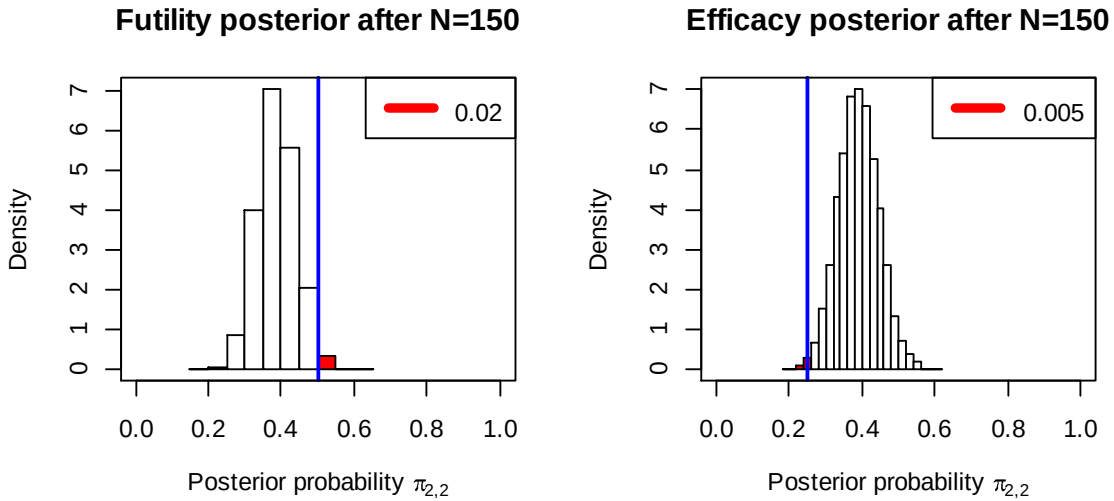
Results (Two-criteria decisions)



- At $N_{max} = 75$
  - about 87% of trials end in declaring the underlying true efficacious combination as efficacious
  - About 10% of trials denote the inefficacious combinations as efficacious
- Rather surprising drop in power (green curve) with increasing N_{max}!

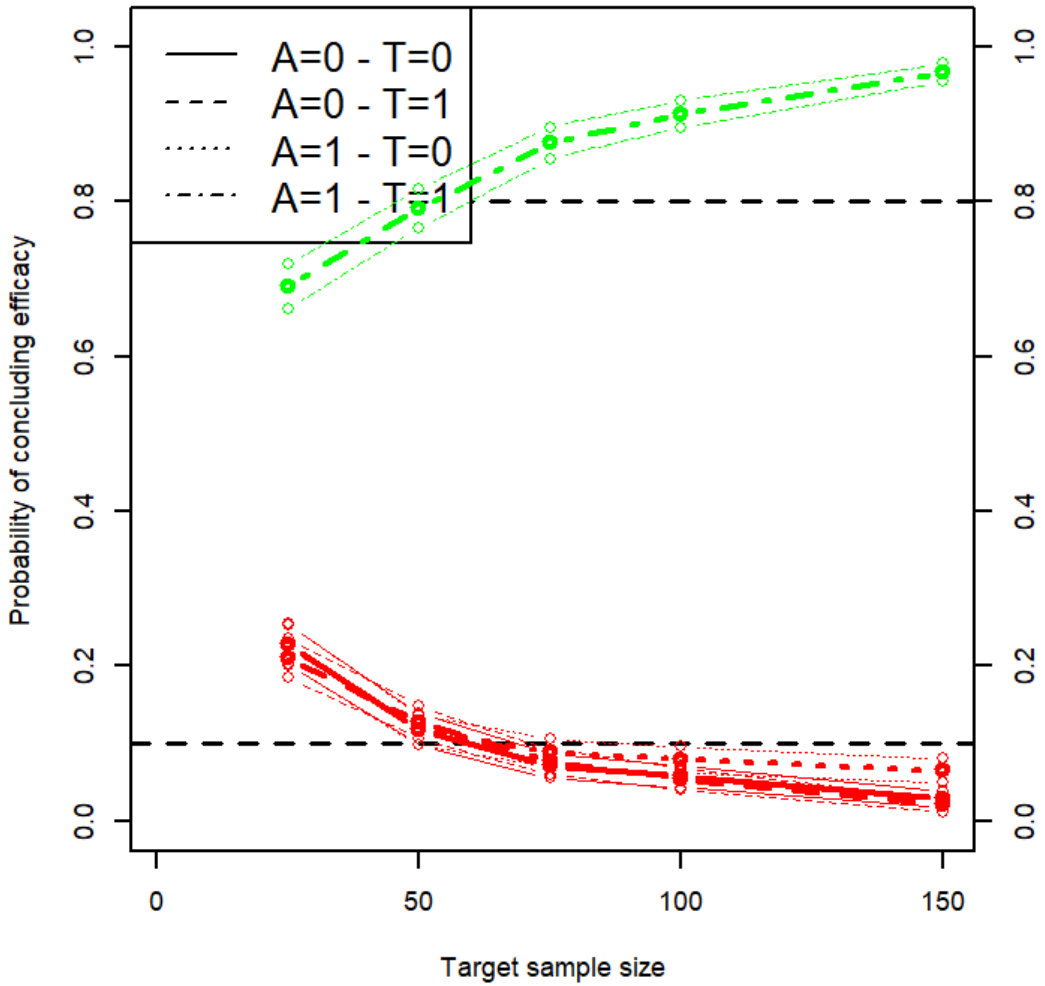
Results (Two-criteria decisions)





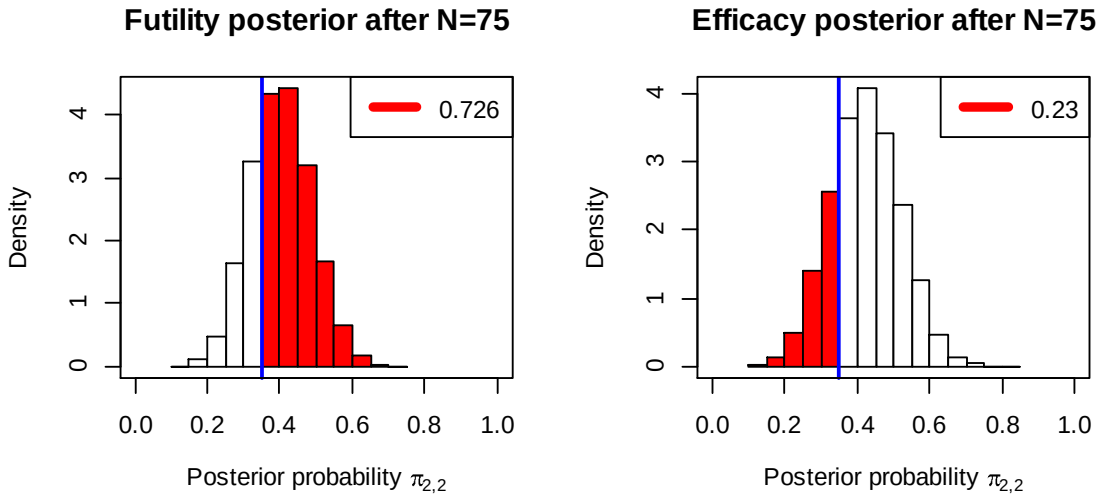
- For about 10% of trials
  - After $n = 75$ significant efficacy test
  - After $n = 150$ both efficacy **and** futility test are significant
- Random variability => Posterior centered around $\pi_{2,2} < 0.5$
- Increased sample size => Reduced posterior variability

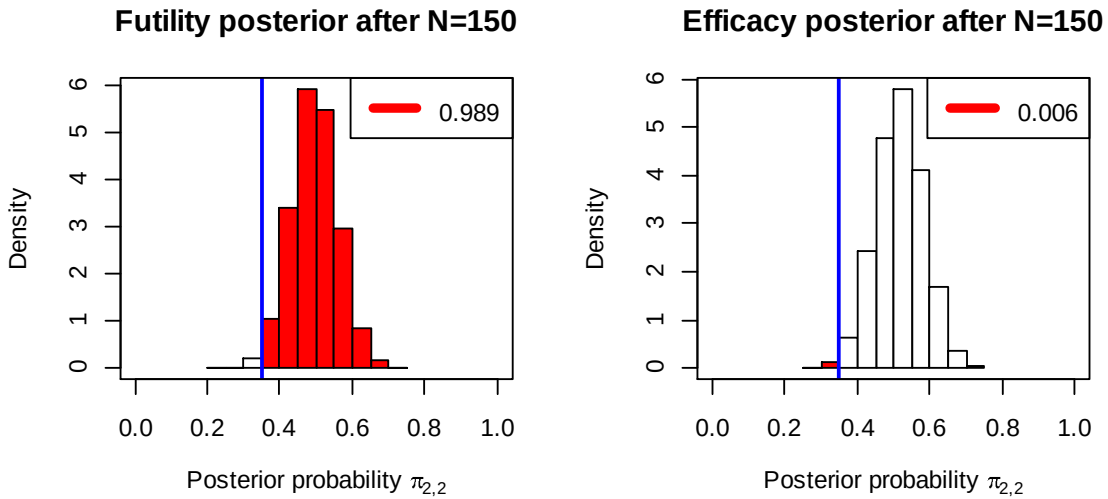
Results (One-criterion decisions)



- Artificially matched the power at $N_{max} = 75$ with two-criteria decisions by considering $\delta_u = 0.62$
- Power increases with N_{max} as theoretically expected!

Results (One-criterion decisions)





- Increased sample size => Reduced posterior variability

Leads only to lower probability of concluding futility and higher probability of concluding efficacy

Outline

- Introduction
- Methods
  - Model
  - Adaptive Trial Characteristics
  - Decision Rule(s)
- Simulation Study
- Results
- Conclusion

Conclusion

- Assuming 2 different criteria for stopping and concluding efficacy leads to contradictory results
  - Reduced posterior variability when increasing sample size

- Even in the (highly unlikely) setting where $P_{1,1} = P_{1,2} = P_{2,1} \equiv \pi_0$ and $P_{2,2} \equiv \pi_1$

- Overestimating the response probability $(P_{2,2} < \pi_1)$ increases the probability of contradictory results
  - Posterior distributions will be centred somewhere between π_0 and π_1

References

- Barry, W.T., Perou, C.M., Marcom, P.K., Carey, L.A., and Ibrahim, J.G. (2015). The use of Bayesian hierarchical models for adataptive randomization in biomarker-driven phase II studies. *Journal of Biopharmaceutical statistics.* 25, 66-88.
- Berry, S. M., Carlin, B. P., Lee, J. J., & Muller, P. (2010). *Bayesian adaptive methods for clinical trials*. CRC press.
- Gu, X., Chen, N., Wei, C., Liu, S., Papadimitrakopoulou, V.A., Herbst,R.S., and Lee, J.J. (2016). Bayesian tow-stage biomarker-based adaptive design for targeted therapy development. *Statistics in biosciences.* 8, 99-128.
- Kim, E.S., Herbst, R.S., Wistuba, I.I., Lee, J.J., Blumenschein Jr., G.R., et al. (2011). The BATTLE trial: Personalizing Therapy for Lung Cancer. *Cancer Discovery*. 1, 44-53.
- Zhou, X., Liu, S., Kim, E.S., Herbst, R.S., and Lee, J.J. (2008). Bayesian adaptive design for targeted therapy development in lung cancer – a step toward personalized medicine. *Clinical Trials.* 5, 181-193.

Thank you for your attention!