Reproducing subgroup findings – worth the effort?

Marietta Kirchner, Heiko Götte

IMBI, Heidelberg **Merck, Darmstadt**

EFSPI Scientific Meeting: Reproducibility in Clinical Research Nov 2019

imbi Heidelberg Institute of Medical Biometry and Informatics

HEIDELBERG UNIVERSITY HOSPITAL

MERCK

Subgroup analysis fulfill different purposes

1. Confirmatory subgroup analyses
   - Prespecified, adjusted for multiple testing

2. Consistency check in confirmatory trials
   - Prespecified, definition of „consistent“? What to do if not „consistent“?

3. Exploratory post-hoc subgroup analyses
   - Goal: find subgroup with improved BR profile
   - One of several prespecified subgroups
   - Newly identified subgroup based on (high) number of covariates

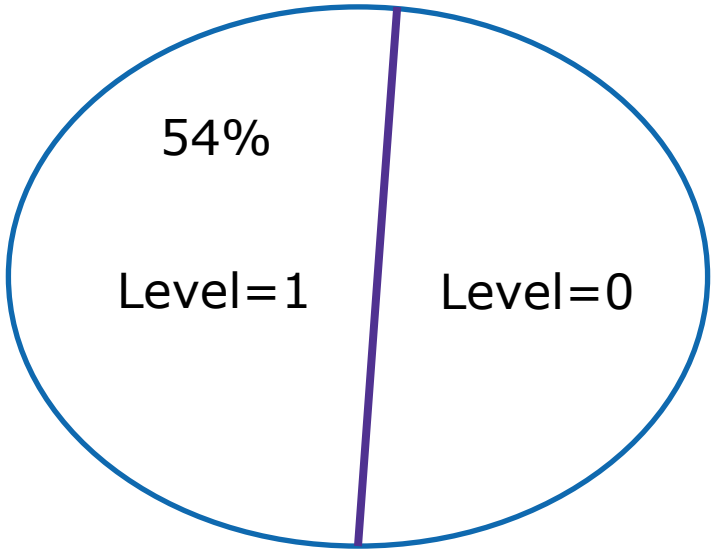
Simple case: one subgroup factor
Motivating example, binary endpoint

- Patients with ovarian cancer, endpoint = „objective response“ (RECIST criteria), n=120
- Full population: OR=1.52 (68% (41/60) in the control arm and 77% (46/60) in the experimental arm)

	Subgroup level=0		
	Control	Experimental	
	Response rate		Odds ratio
FIGO stage	0.73	0.64	0.65
	Subgroup level=1		
	Control	Experimental	
	Response rate		Odds ratio
FIGO stage	0.63	0.86	**3.47**

Subgroup factor “FIGO stage” at baseline

- Subgroup level = 0 → FIGO stage = ”IV”
- Subgroup level = 1 → FIGO stage = ”III or better”



Estimates of subgroup treatment effects in overall nonsignificant trials: To what extent should we believe in them?

Julien Tanniou[1,2] | Ingeborg van der Tweel[1] | Steven Teerenstra[2,3] | Kit C.B. Roes[1,2]

[1]Julius Center for Health Sciences and Primary Care, Department of Biostatistics, UMC Utrecht, Utrecht, Netherlands

[2]Medicines Evaluation Board, College ter Beoordeling van Geneesmiddelen, Utrecht, Netherlands

[3]Radboud Institute for Health Sciences, Department of Health Evidence, section Biostatistics, Radboud UMC, Nijmegen, Netherlands

Correspondence
Julien Tanniou, Julius Center for Health Sciences and Primary Care, Department of Biostatistics, UMC Utrecht, Utrecht,

In drug development, it sometimes occurs that a new drug does not demonstrate effectiveness for the full study population but appears to be beneficial in a relevant subgroup. In case the subgroup of interest was not part of a confirmatory testing strategy, the inflation of the overall type I error is substantial and therefore such a subgroup analysis finding can only be seen as exploratory at best. To support such exploratory findings, an appropriate replication of the subgroup finding should be undertaken in a new trial. We should, however, be reasonably confident in the observed treatment effect size to be able to use this estimate in a replication trial in the subpopulation of interest. We were therefore interested in evaluating the bias of the estimate of the subgroup treatment effect, after selection based on significance for the subgroup in an overall "failed" trial. Different scenarios, involving continu-

Pharmaceutical Statistics. 2017;16:280-295

Simulation study by Tanniou et al 2017

Overall result not significant – Subgroup is significant

TABLE 4 Probability of observing a significant subgroup test ($\alpha_S = 0.004$) under the null hypothesis

Proportion of the subgroup	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of $p < .004$	0.157	0.146	0.136	0.118	0.102	0.087	0.057	0.033	0.006

„This pragmatic strategy could therefore be of real interest to judge whether a subgroup finding should be considered relevant and hence be replicated."

Pharmaceutical Statistics. 2017;16:280-295

Simulation study by Tanniou et al 2017

Overall result not significant – Subgroup is significant

TABLE 4 Probability of observing a significant subgroup test ($\alpha_S = 0.004$) under the null hypothesis

Proportion of the subgroup	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of $p < .004$	0.157	0.146	0.136	0.118	0.102	0.087	0.057	0.033	0.006

- Here only one subgroup was considered
- Usually, there are several overlapping subgroups

→ Quantify the bias in case of overlapping subgroups

→ Propose bias adjustment method

Simulation study by Tanniou et al 2017

Overall result not significant – Subgroup is significant

TABLE 4 Probability of observing a significant subgroup test ($\alpha_S = 0.004$) under the null hypothesis

Proportion of the subgroup	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of $p < .004$	0.157	0.146	0.136	0.118	0.102	0.087	0.057	0.033	0.006

- Here only one subgroup was considered
- Usually, there are several overlapping subgroups

→ **Quantify the bias in case of overlapping subgroups** 1.

→ Propose bias adjustment method: applied to

a) binary endpoint, b) time-to-event 2.

Several overlapping subgroups
Motivating example, binary endpoint

- Full population: OR=1.52 (68% (41/60) in the control arm and 77% (46/60) in the experimental arm)

	Subgroup level=0		
	Control	Experimental	
	Response rate	Response rate	Odds ratio
FIGO stage			
ECOG score			
BL SOD			
Tu ascites			
Tu pelvic			
Diff grade			

Overlap with FIGO status

	ECOG score	BL SOD	Tu ascites	Tu pelvic	Diff grade
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 0)$	0.56	0.40	0.40	0.35	0.16
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 1)$	0.58	0.66	0.32	0.38	0.26

FIGO stage=FIGO disease stage at baseline (1=”III or better”, 0=”IV”), ECOG score=ECOG performance score (1=”0”,0=”≥1”), BL SOD=baseline tumor sum of diameters (1=”≤50mm”, 0=”>50mm”), Tu ascites= tumor location ascites at baseline (1=”yes”, 0=”no”), Tu pelvic=tumor location at baseline pelvic soft tissue (1=”yes”, 0=”no”), Diff grade=differentiation grade of tumor at baseline (1=”moderate or better”, 0=”poor or worse”).

Several overlapping subgroups
Motivating example, binary endpoint

- Full population: OR=1.52 (68% (41/60) in the control arm and 77% (46/60) in the experimental arm)
- Subset with the largest observed treatment effect: OR=3.47 (complement: OR=0.65)

	Subgroup level=0		
	Control	Experimental	
	Response rate	Response rate	Odds ratio
FIGO stage	0.73	0.64	0.65
ECOG score	0.57	0.80	3.00
BL SOD	0.70	0.64	0.76
Tu ascites	0.62	0.75	1.83
Tu pelvic	0.68	0.79	1.86
Diff grade	0.68	0.78	1.65

	Subgroup level=1		
	Control	Experimental	
	Response rate	Response rate	Odds ratio
FIGO stage	**0.63**	**0.86**	**3.47**
ECOG score	0.74	0.73	0.95
BL SOD	0.67	0.86	3.00
Tu ascites	0.78	0.80	1.11
Tu pelvic	0.70	0.71	1.09
Diff grade	0.69	0.70	1.06

Overlap with FIGO status

	ECOG score	BL SOD	Tu ascites	Tu pelvic	Diff grade
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 0)$	0.56	0.40	0.40	0.35	0.16
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 1)$	0.58	0.66	0.32	0.38	0.26

Replication of the subgroup finding in a new trial → Worth the effort?

→ Probability of success (PoS)?

→ Bias due to selection?

Subgroup setting: Illustrating bias for PoS estimate for a subset q

Power with fixed true effect: $\theta_q = \log(OR)$

$$\mathrm{P}\left(Reject\ H_0 | \theta_q\right)$$

True PoS

PoS for phase III is estimated based on phase II data $\hat{\theta}_q$

$$\int_{-\infty}^{\infty} \mathrm{P}\left(Reject\ H_0 | \theta_q\right) f(\theta_q | \hat{\theta}_q) d\theta_q$$

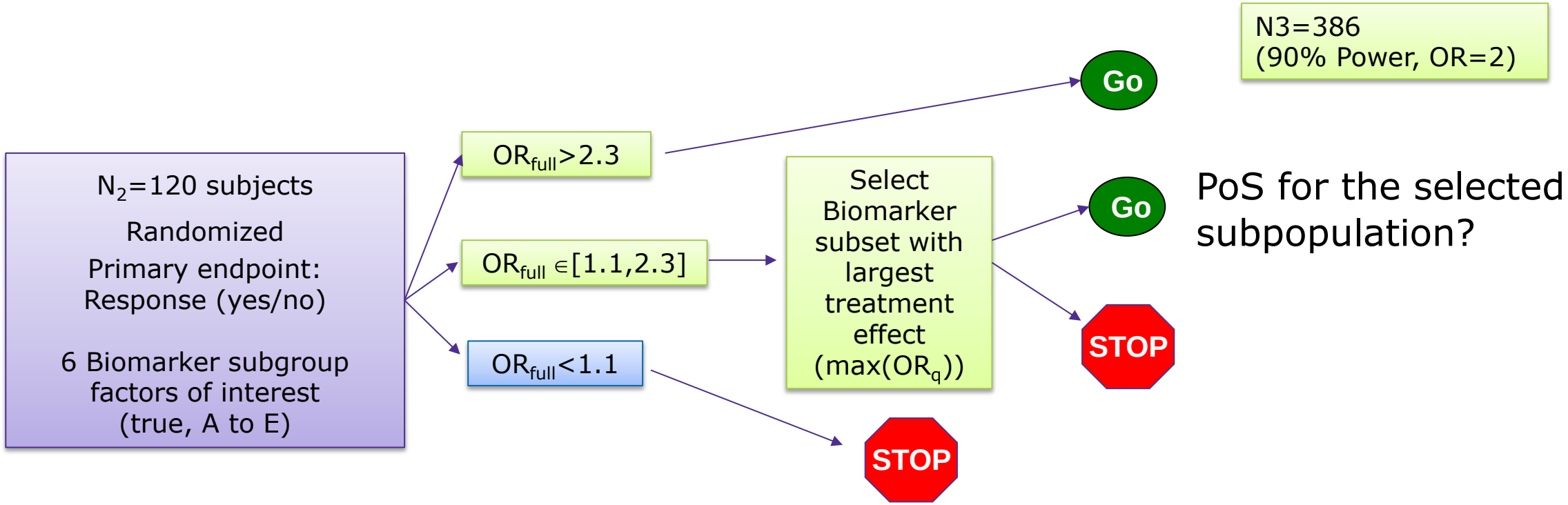
Estimated PoS

Quantify the bias in a simulation study!

Trial design used for the simulation study

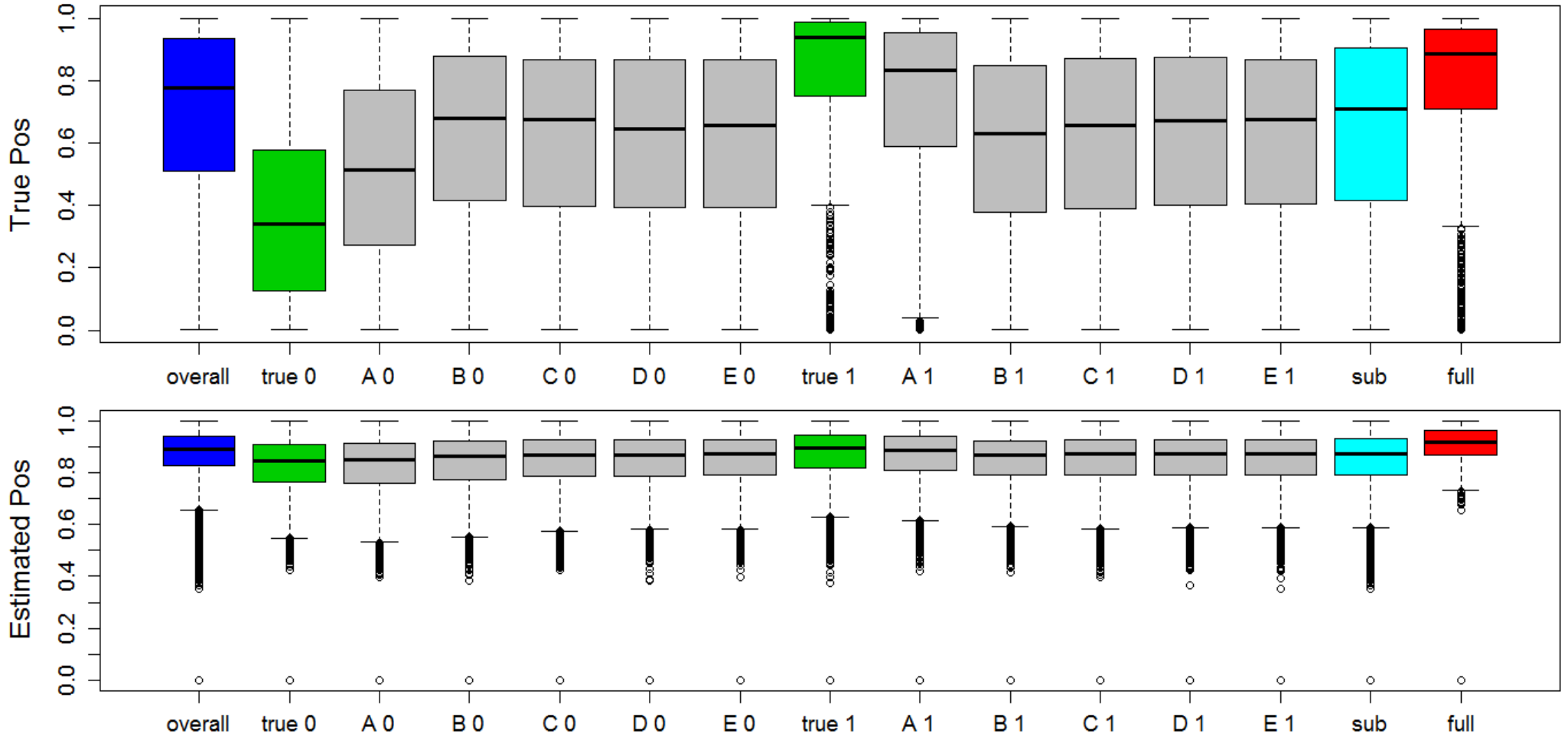
Early development

Late development



PoS for the selected subpopulation?

Results for the setting with differential prior, 5000 simulation runs
Estimated PoS is nearly always high (>80%) - True PoS is not



Götte H, Kirchner M, Sailer MO. Probability of success for phase III after exploratory biomarker analysis in phase II. Pharm Stat. 2017;16:178–191.

Available methods to correct for overoptimism due to subgroup selection

1. Ignoring selection procedure

a) Methods for multiple testing

- Focus on the number of subgroups, independent decision bounds
- Comparing subgroup effects with each other (e.g., maximum effect) is ignored

b) Hierarchical models

- Bayesian approach for independent (non-overlapping) subgroups
- True subgroup effects are realizations from joined distribution

$\Rightarrow$ Subgroup effects are shrunken towards the overall mean

c) Model averaging

Available methods to correct for overoptimism due to subgroup selection

2. Taking selection procedure into account

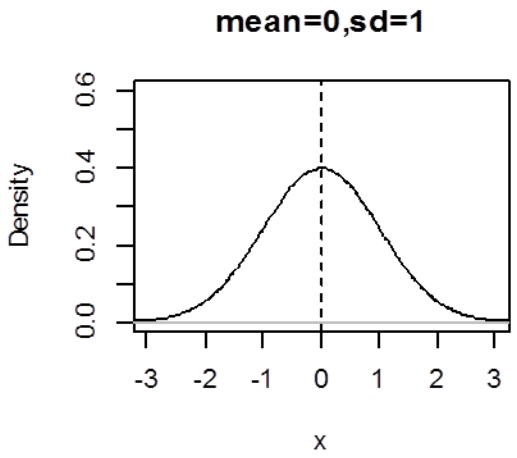
a) Resampling

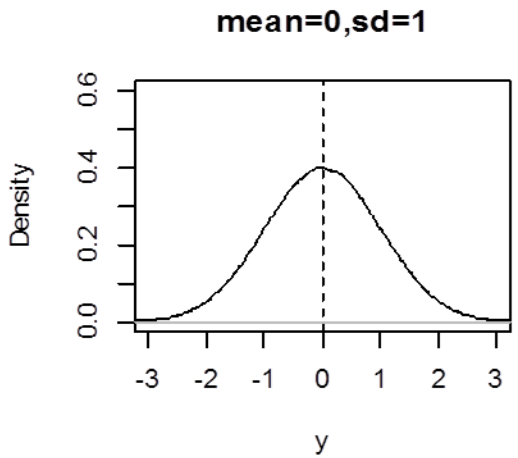
- Bootstrap, permutation
- Identify the magnitude of selection bias and subtract it from the unadjusted estimate
- Quality of adjustment depends on how often the most influential data points are included in the bootstrap samples

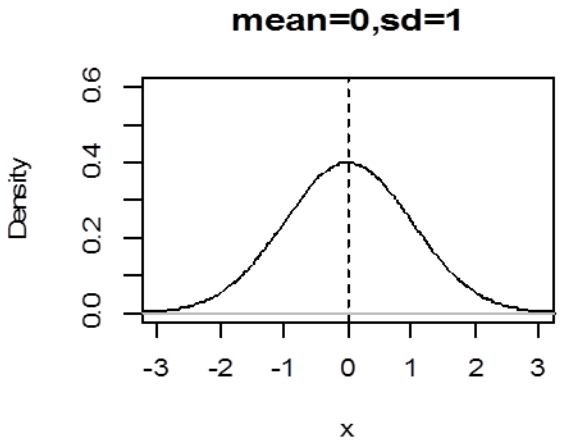
→ level of adjustment reflects whether the observed effects are based on a few or on many observations.

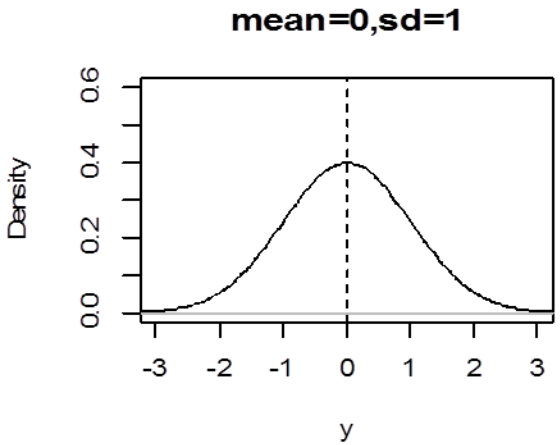
Best approach? → Step back and see where the problem is!

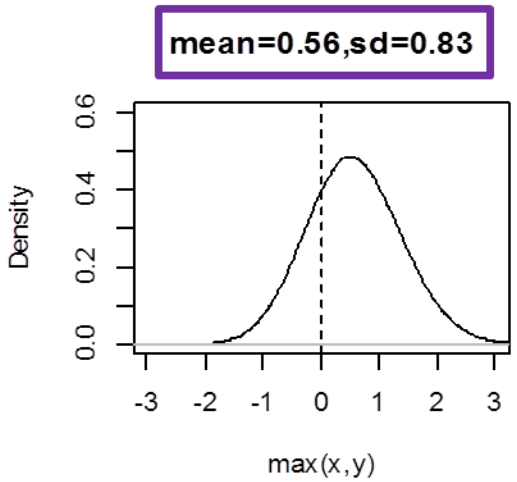
More options ⇒ more bias

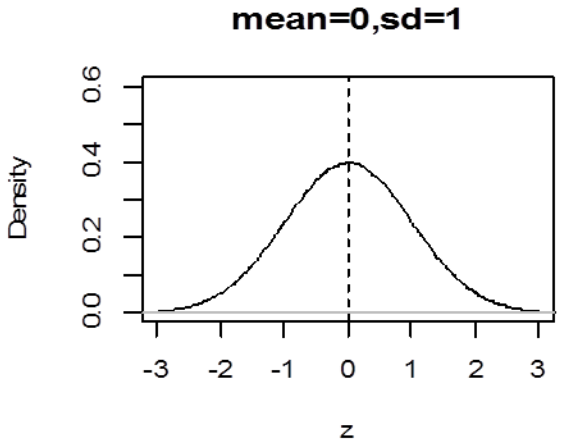


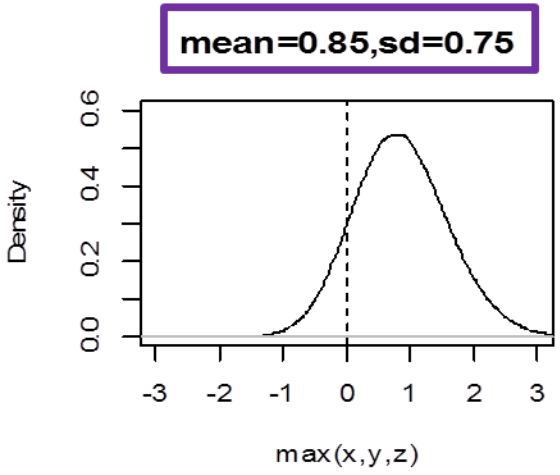




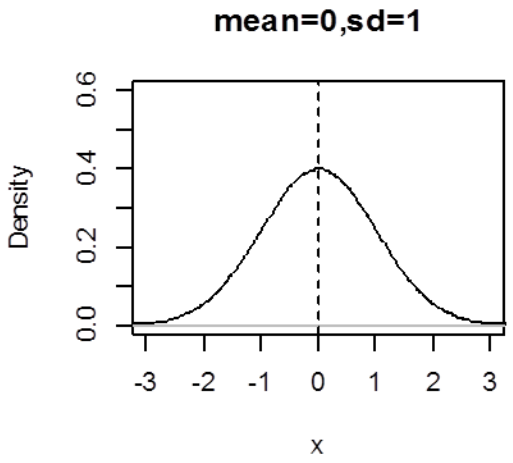


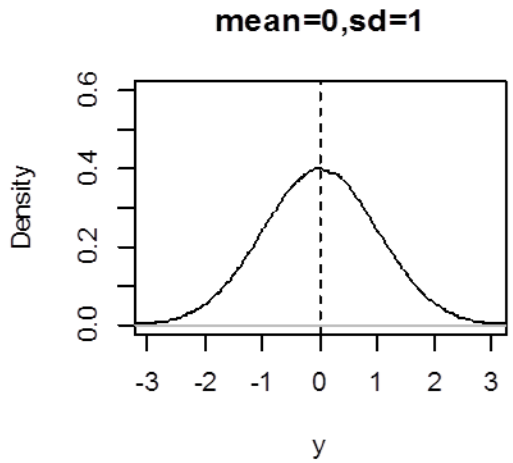


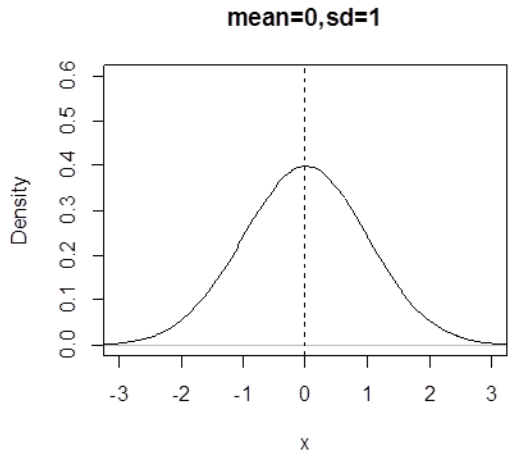


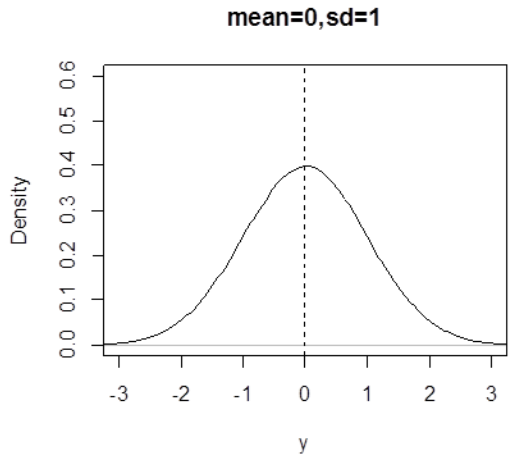


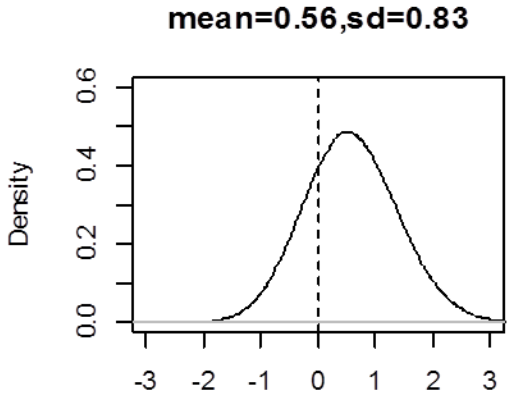
Level of bias depends on selection rule





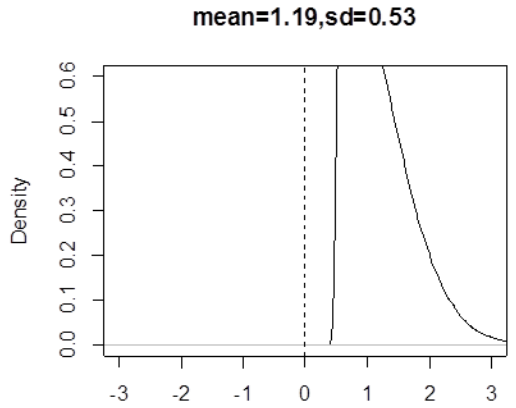






max(x,y)

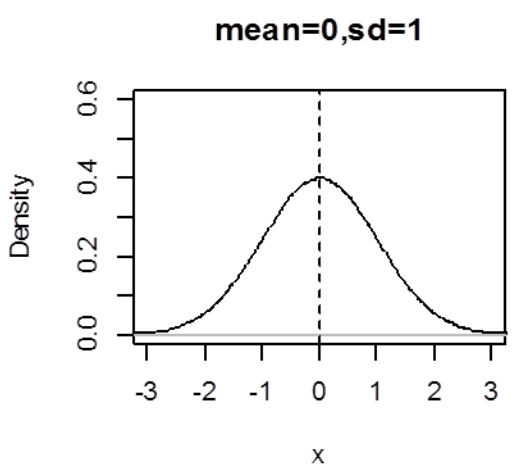
	x	y	Either
P(select)	50%	50%	100%
Bias	0.56	0.56	0.56

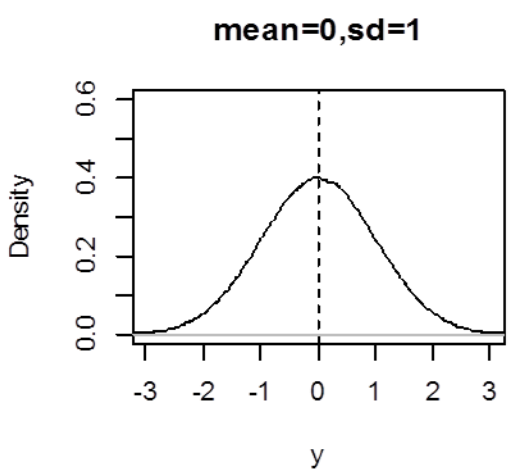


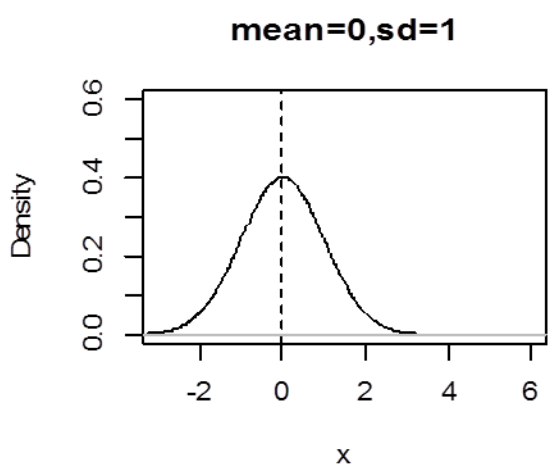
max(x,y) if greater than 0.5

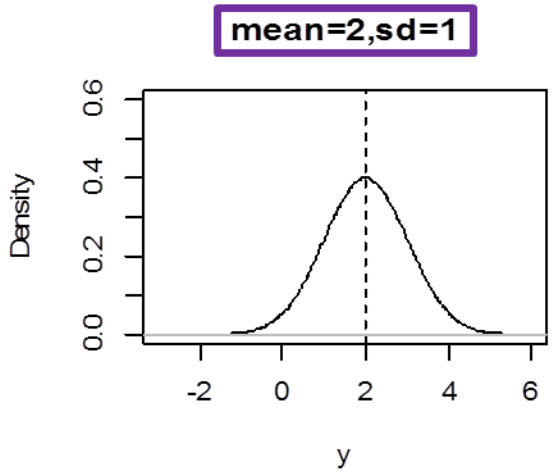
	x	y	Either
P(select)	26%	26%	52%
Bias	1.19	1.19	1.19

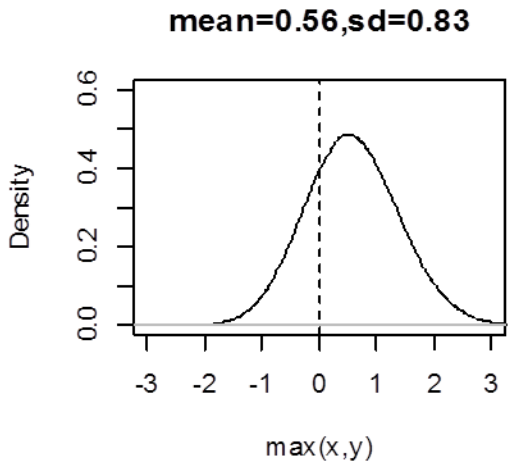
Level of bias depend on underlying distributions



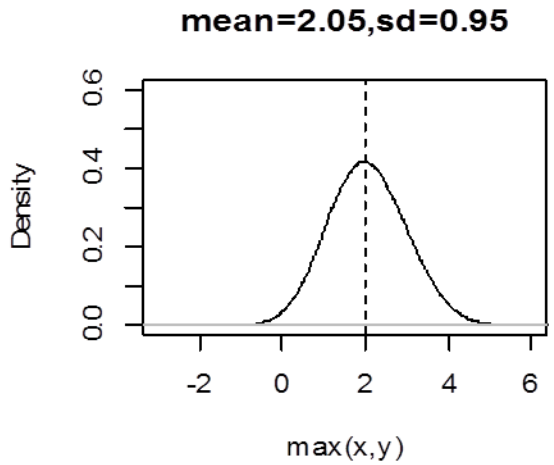








	x	y	Either
P(select)	50%	50%	100%
Bias	0.56	0.56	0.56



	x	y	Either
P(select)	8%	92%	100%
Bias	1.30	0.11	0.21

Summary

With the correct model, bias is only generated by selection, however, magnitude depends on

- Number of options for selection

  → The more options the greater is the bias

- Type of selection rule

  → More restriction increases bias but reduces selection probability

- Underlying true distribution

  → True values are unknown ⇒ uncertainty in underlying distribution

Summary

With the correct model, bias is only generated by selection, however, magnitude depends on

- Number of options for selection

  → The more options the greater is the bias

- Type of selection rule

  → More restriction increases bias but reduces selection probability

- Underlying true distribution

  → True values are unknown ⇒ uncertainty in underlying distribution

⇒ „One fits all" correction does not work

⇒ None of the presented methods fulfill all criteria

⇒ Ideal: Bayesian approach that takes selection procedure into account

Classical Bayesian approach difficult to implement

Bayesian approach allows to make statements about the true effect given the data

$$\pi(\theta_{select} \mid y_{obs}) \text{ with } \; y_{obs} = (resp_i, treat_i, x_{i1}, \ldots x_{iK}),\ x_{ik} \in \{0,1\}$$

→Complex structure of Prior and Likelihood due to overlapping subsets and selection procedure

→Difficulty to obtain an analytical expression for the Likelihood

Approximate **B**ayesian **C**omputation
Approximate posterior by ABC approach

Determine: $\pi(\theta_{select} \mid y_{obs})$ with $y_{obs} = (resp_i, treat_i, x_{i1}, \dots x_{iK})$, $x_{ik} \in \{0,1\}$

→Complex structure of Prior and Likelihood due to overlapping subsets and selection procedure

→Difficulty to obtain an analytical expression for the Likelihood

<u>Solution:</u> use simulations to derive posterior distribution → ABC approach

- ABC method bypass the evaluation of the likelihood function
- Approximate posterior by $\pi_{ABC}(\theta_{select} \mid y_{obs})$
- Here: rejection sampling

Approximate Bayesian Computation
Steps of the ABC approach

One data set: observed data are summarized →

<u>ABC steps:</u>

1. Draw n_{sim} values θ_i from the prior distribution
2. Simulate a data set based on each θ_i
3. Calculate the summary statistic
4. Compare with observed data
5. Accept simulation if "similar" to observed data

→ Posterior distribution of model parameter θ is approximated by the distribution of θ_i from accepted simulations

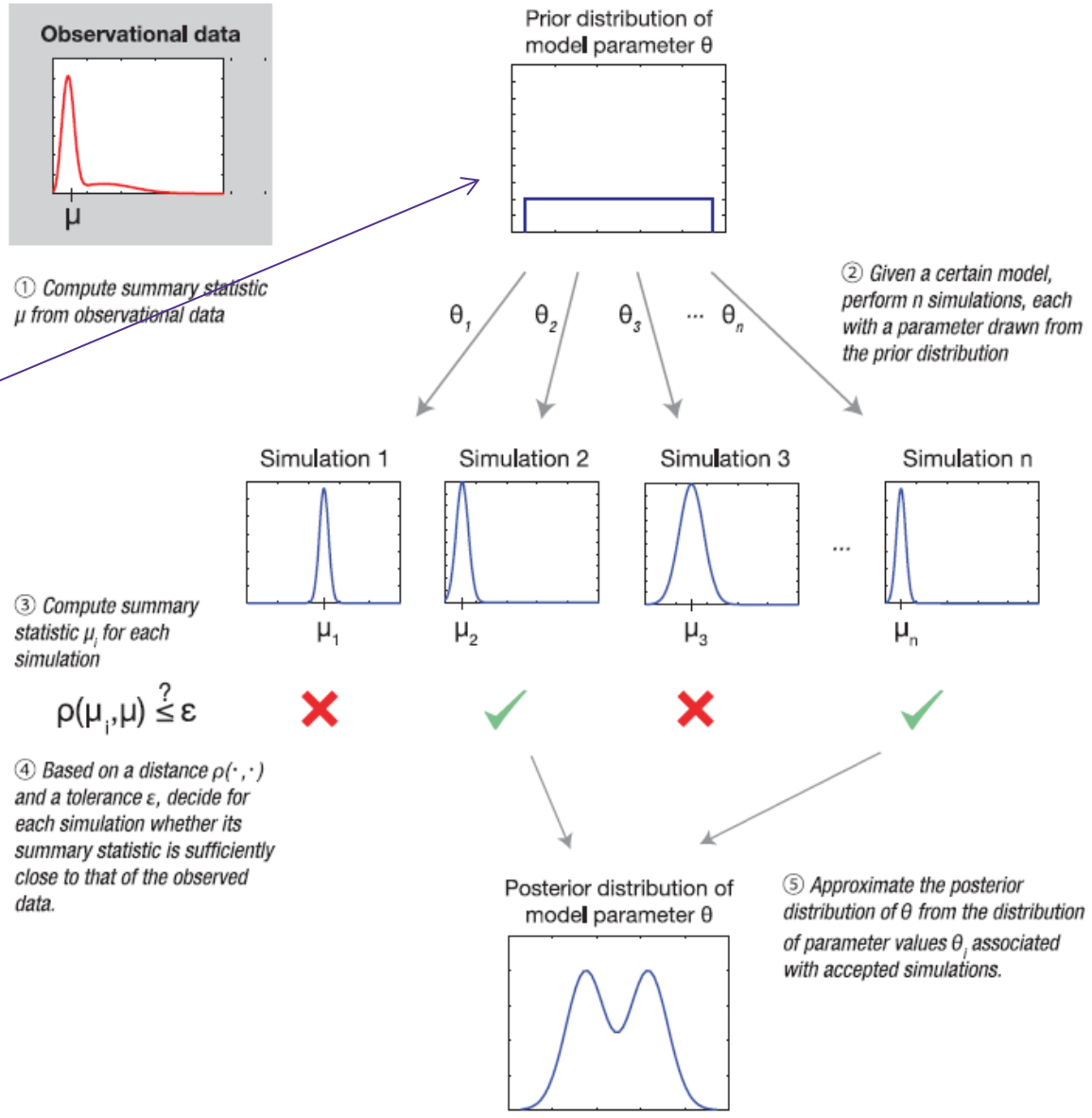


Figure 1. Parameter estimation by Approximate Bayesian Computation: a conceptual overview.
doi:10.1371/journal.pcbi.1002803.g001

Citation: Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, et al. (2013) Approximate Bayesian Computation. PLoS Comput Biol 9(1): e1002803. doi:10.1371/journal.pcbi.1002803

Simple example of ABC approach

- Binary endpoint: response
- Intervention (treat=1) vs control (treat=0) group
- No subgroup → example just for illustration purpose

Binary endpoint, Logistic regression model
Treatment defines P(response)

$$P(Y_i = 1|treat_i\ ) = \frac{e^{(\vartheta+\theta*treat_i)}}{1+e^{(\vartheta+\theta*treat_i)}}$$

Control group: $P(Y_i=1 \mid treat_i=0) = \frac{e^{(\vartheta)}}{1+e^{(\vartheta)}}$,

Treatment group: $P(Y_i=1 \mid treat_i=1) = \frac{e^{(\vartheta+\theta)}}{1+e^{(\vartheta+\theta)}}$

Prior for ABC:

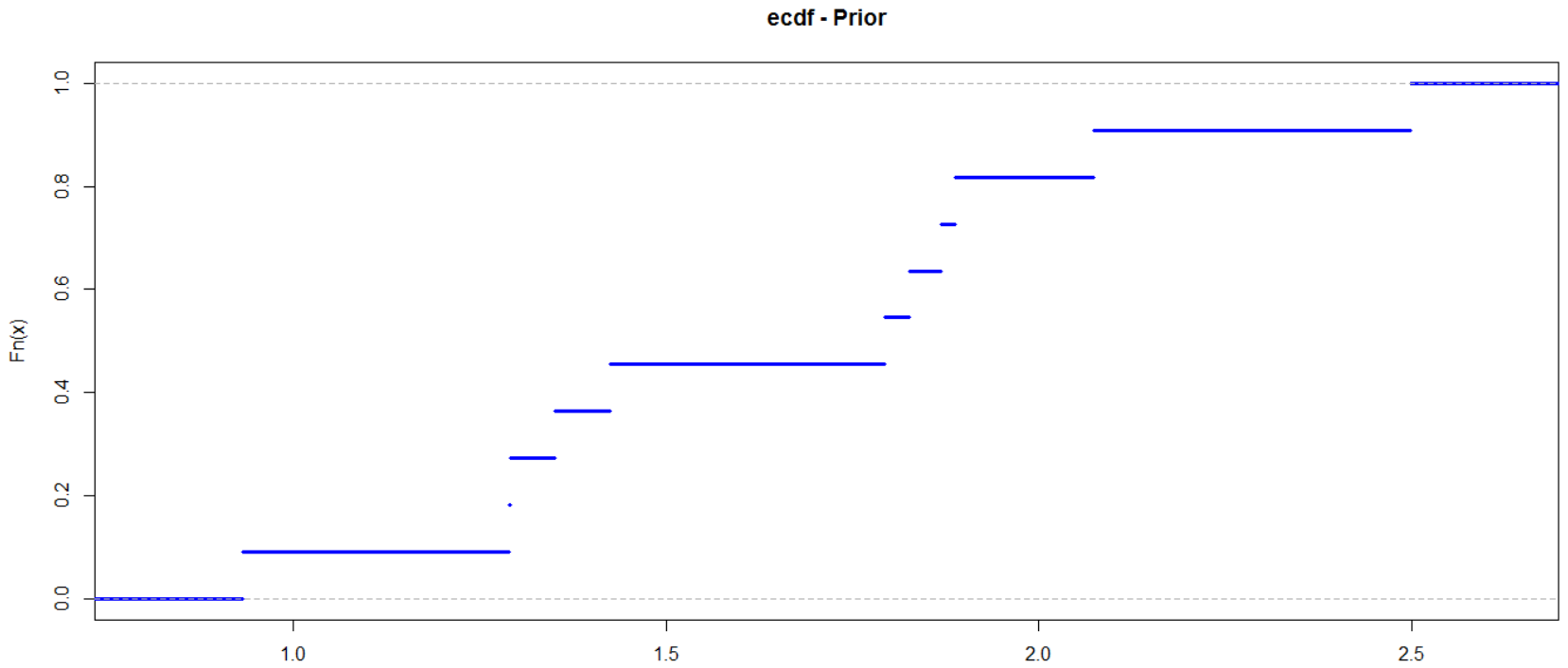
$$(\vartheta,\theta)\sim N\left(\begin{pmatrix}\log\left(\frac{0.3}{0.7}\right)\\ \log(1.5)\end{pmatrix},\begin{pmatrix}0 & 0\\ 0 & 1/25\end{pmatrix}\right)$$ → response rate control group = 30%

Step 1 – Draw from a prior distribution
Principle of ABC approach

Simulation

$\frac{e^{(\vartheta)}}{1+e^{(\vartheta)}}$	**True OR = exp(θ)**
0.3	1.89
0.3	2.07
0.3	0.93
0.3	1.42
0.3	1.79
0.3	1.29
0.3	1.29
0.3	1.83
0.3	1.87
0.3	2.50
0.3	1.35

Prior distribution



Step 2 – Simulate data sets
Principle of ABC approach

Simulation

$\frac{e^{(\vartheta)}}{1+e^{(\vartheta)}}$	**True OR = exp(θ)**	**n_{CN}**	**n_{CR}**	**n_{EN}**	**n_{ER}**
0.3	1.89	3	2	5	0
0.3	2.07	4	1	2	3
0.3	0.93	3	2	4	1
0.3	1.42	4	1	2	3
0.3	1.79	5	0	4	1
0.3	1.29	4	1	4	1
0.3	1.29	4	1	4	1
0.3	1.83	2	3	2	3
0.3	1.87	4	1	1	4
0.3	2.50	5	0	1	4
0.3	1.35	4	1	2	3

Step 3 and 4 – Compare summary statistic with observed data
Principle of ABC approach

Simulation

$\frac{e^{(\vartheta)}}{1+e^{(\vartheta)}}$	True OR = exp(θ)	n_{CN}	n_{CR}	n_{EN}	n_{ER}
0.3	1.89	3	2	5	0
0.3	2.07	4	1	2	3
0.3	0.93	3	2	4	1
0.3	1.42	4	1	2	3
0.3	1.79	5	0	4	1
0.3	1.29	4	1	4	1
0.3	1.29	4	1	4	1
0.3	1.83	2	3	2	3
0.3	1.87	4	1	1	4
0.3	2.50	5	0	1	4
0.3	1.35	4	1	2	3

Observed data

n_{CN}	n_{CR}	n_{EN}	n_{ER}
4	1	2	3

Step 5 – Accept simulations, get posterior distribution
Principle of ABC approach

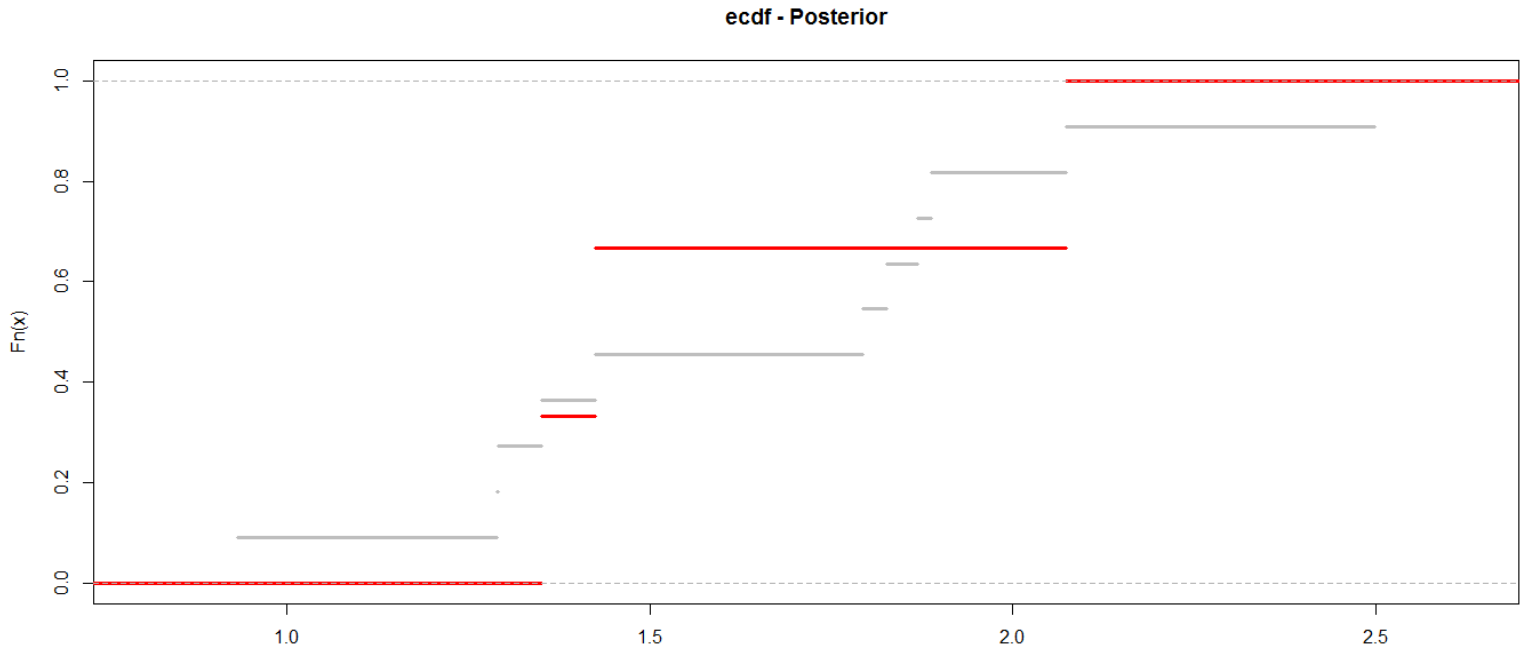
Simulation

$\frac{e^{(\vartheta)}}{1+e^{(\vartheta)}}$	True OR = exp(θ)	n_{CN}	n_{CR}	n_{EN}	n_{ER}
0.3	2.07	4	1	2	3
0.3	1.42	4	1	2	3
0.3	1.35	4	1	2	3

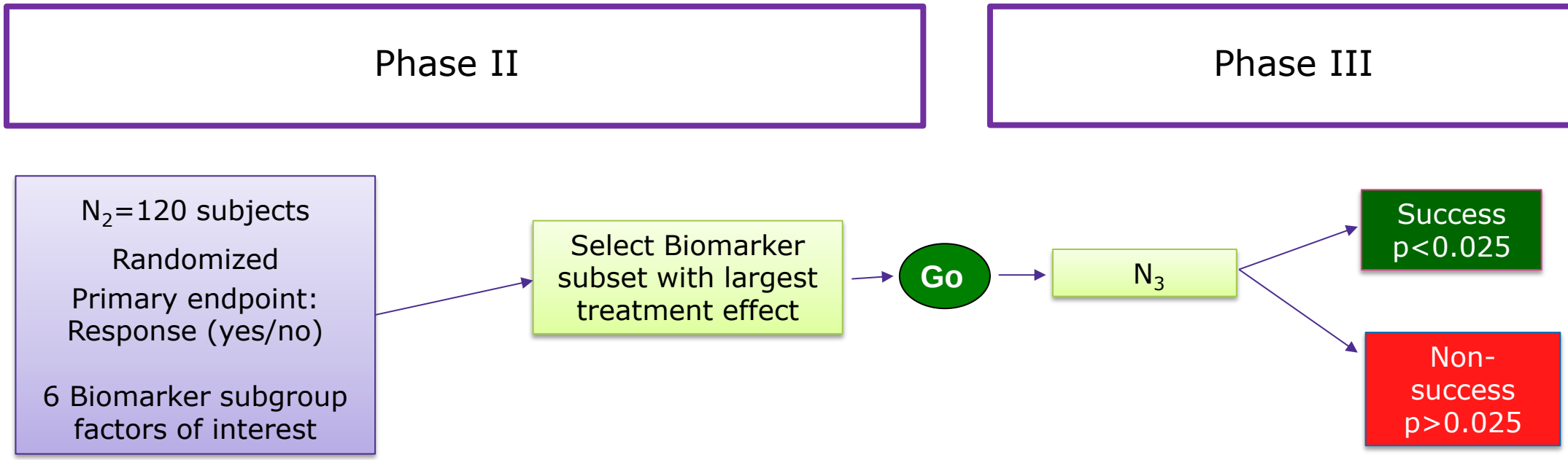
Posterior distribution

Observed data

n_{CN}	n_{CR}	n_{EN}	n_{ER}
4	1	2	3



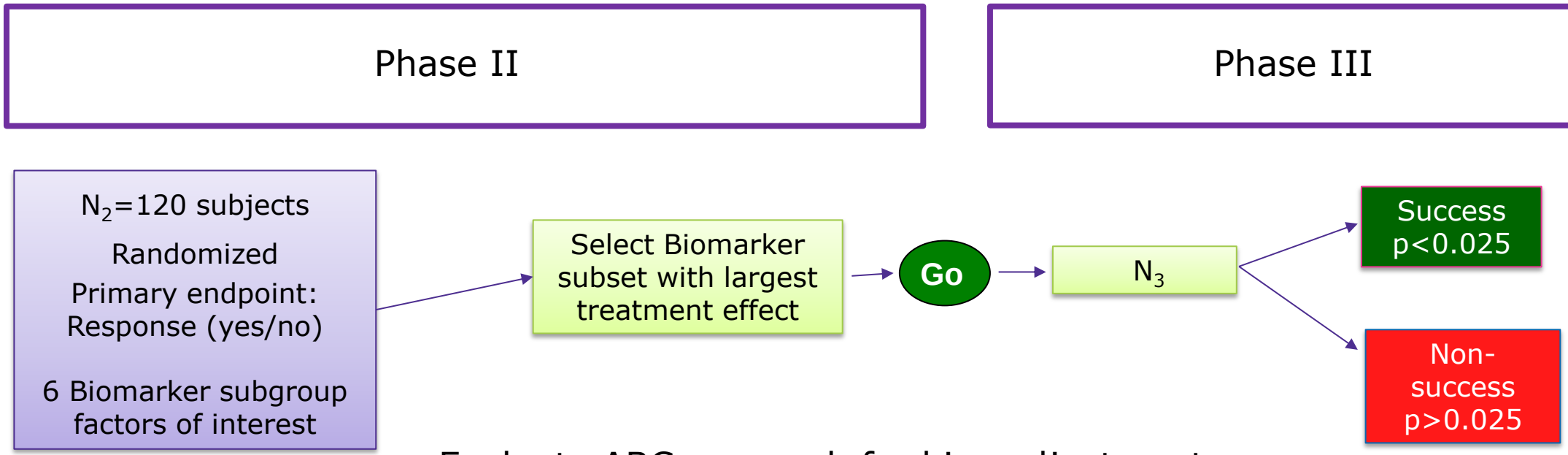
Bias adjustment based on ABC approach
Trial design for simulation



Go decision influenced by

- **Treatment effect estimate log(OR): $\hat{\theta}_{select}$**
- Posterior probability $P(\theta_{select} > \theta_{rel}|data)$
- PoS for phase III: $\int_{-\infty}^{\infty} \mathrm{P}(p < 0.025|\theta_{select})\, f(\theta_{select}|data) d\theta_{select}$

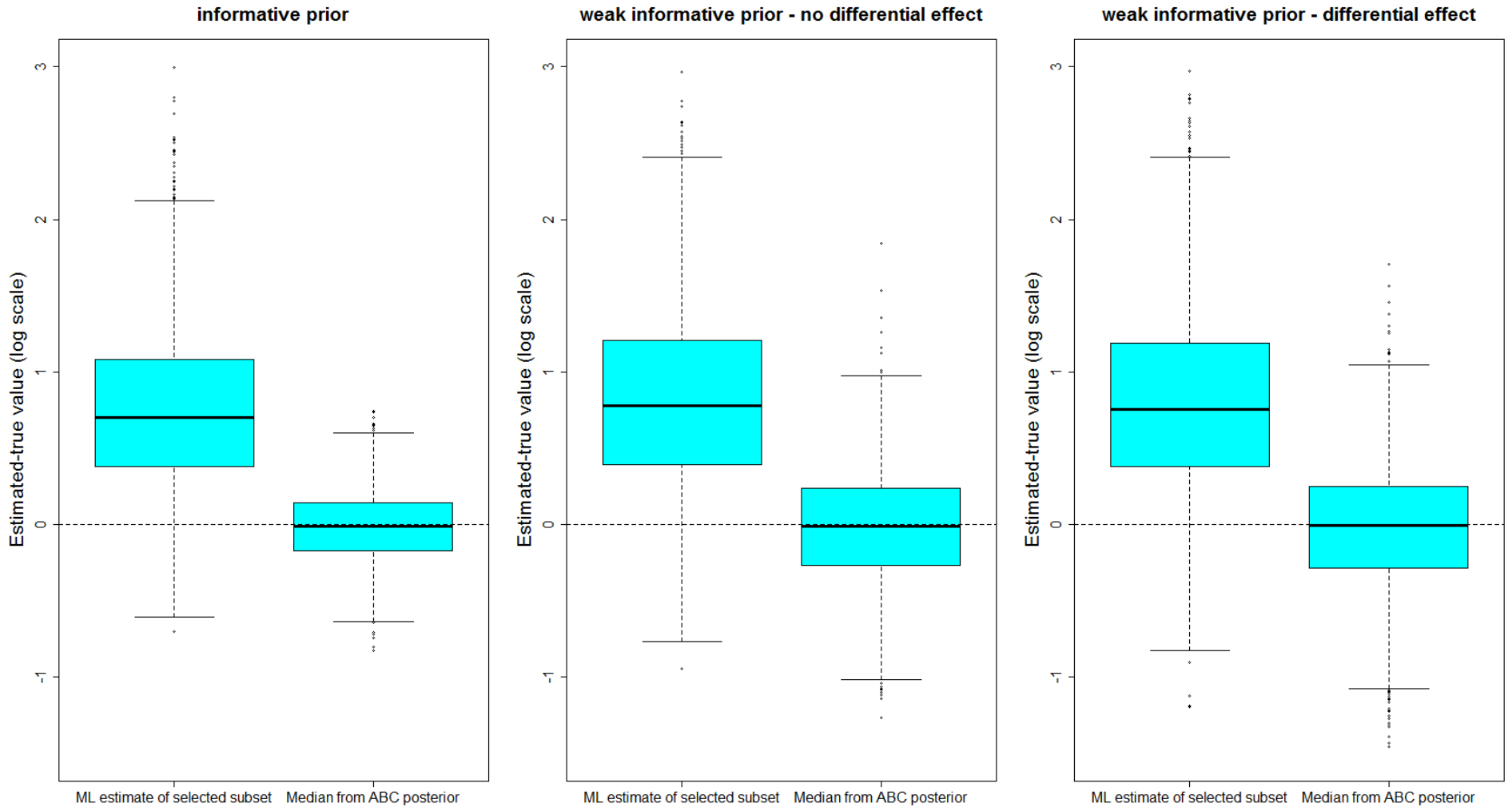
Bias adjustment based on ABC approach
Trial design for simulation



<u>Evaluate ABC approach for bias adjustment:</u>

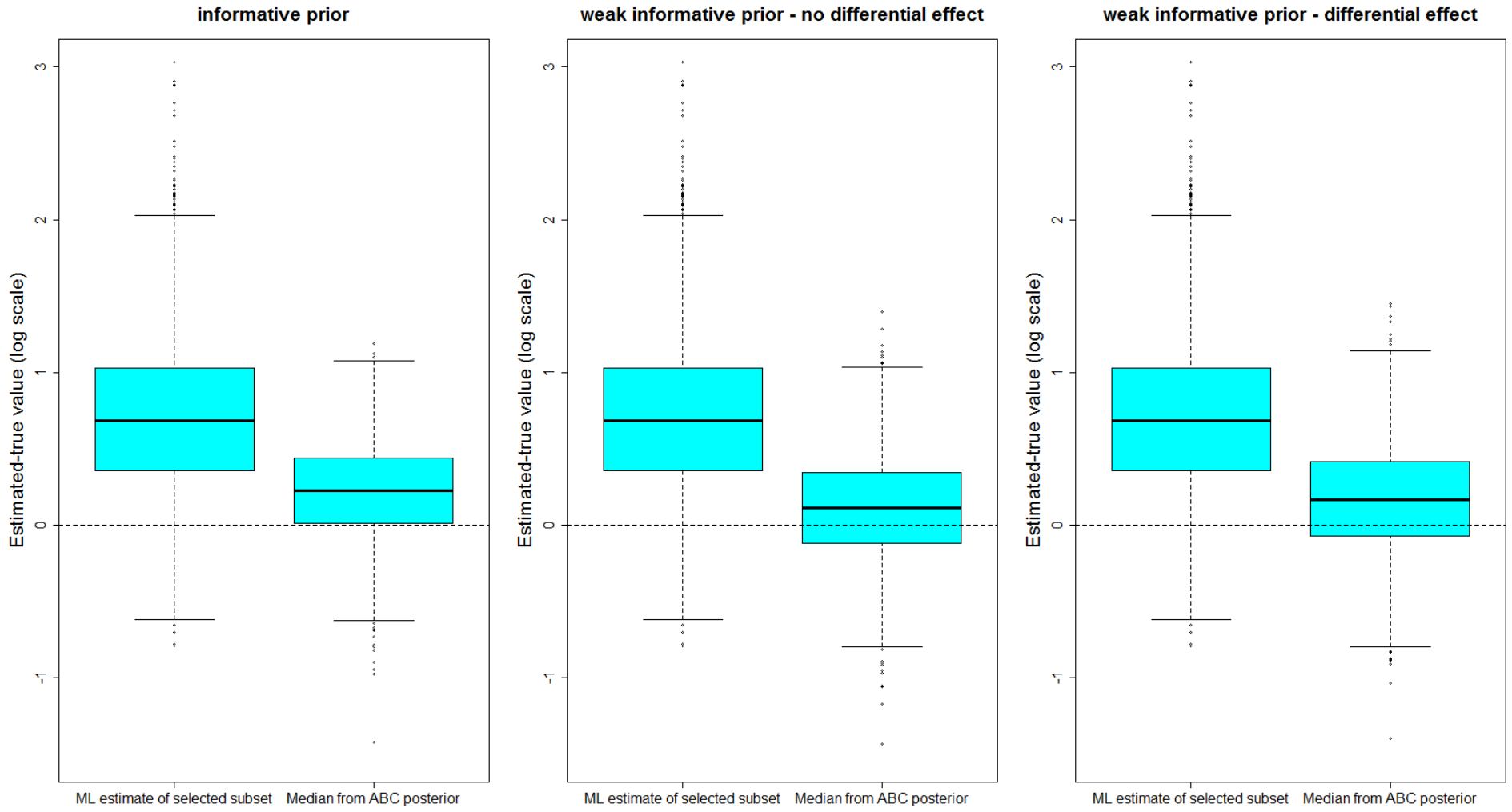
- Generate 2000 „observed" data sets
- For each data set: apply ABC as described before
- Determine $\widehat{\boldsymbol{\theta}}_{\boldsymbol{select}}$:
  - a) Maximum-likelihood estimator
  - b) ABC-adjusted: Median of $\pi_{ABC}(\theta_{select} \mid S_{obs})$

Using same prior for ABC simulation and for generating observations
Simulation: Bias for treatment effect estimate log(OR)



Using different prior for ABC simulation and for generating observations

Simulation: Bias for treatment effect estimate log(OR)



Coming back to our motivating example

- Full population: OR=1.52 (68% (41/60) in the control arm and 77% (46/60) in the experimental arm)
- Subset with the largest observed treatment effect: OR=3.47 (complement: OR=0.65)

	Subgroup level=0		
	Control	Experimental	
	Response rate	Response rate	Odds ratio
FIGO stage	0.73	0.64	0.65
ECOG score	0.57	0.80	3.00
BL SOD	0.70	0.64	0.76
Tu ascites	0.62	0.75	1.83
Tu pelvic	0.68	0.79	1.86
Diff grade	0.68	0.78	1.65

	Subgroup level=1		
	Control	Experimental	
	Response rate	Response rate	Odds ratio
FIGO stage	**0.63**	**0.86**	**3.47**
ECOG score	0.74	0.73	0.95
BL SOD	0.67	0.86	3.00
Tu ascites	0.78	0.80	1.11
Tu pelvic	0.70	0.71	1.09
Diff grade	0.69	0.70	1.06

Overlap with FIGO status

	ECOG score	BL SOD	Tu ascites	Tu pelvic	Diff grade
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 0)$	0.56	0.40	0.40	0.35	0.16
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 1)$	0.58	0.66	0.32	0.38	0.26

Replication of the subgroup finding in a new trial → Worth the effort?

→ Probability of success?

→ Bias of the observed OR after selection?

Results of ABC-adjustment
Motivating example

- Full population: OR=1.52
- Subset with the largest observed treatment effect: OR=3.47 (complement: OR=0.65)

	Full population	Selected subset (ignoring selection)	Selected subset (adjusted for selection)
OR	1.52	3.47	1.76
Posterior probability P(OR>2 given data)	25%	76%	37%
PoS for a future trial with 510 subjects	52%	89%	61%

Götte H, Kirchner M, Sailer MO, Kieser M. Simulation-based adjustment after exploratory biomarker subgroup selection in phase II. Stat Med. 2017; 36:2378–2390.

Conclusion
Motivating example

- Full population: OR=1.52
- Subset with the largest observed treatment effect: OR=3.47 (complement: OR=0.65)

OR
Posterior pro P(OR>2 give
PoS for a futu with 510 sub

- Overall, there is no evidence that there are differential effects among the subsets.
- If the sponsor would decide to continue with the development at all, then phase III should be conducted with the full population.

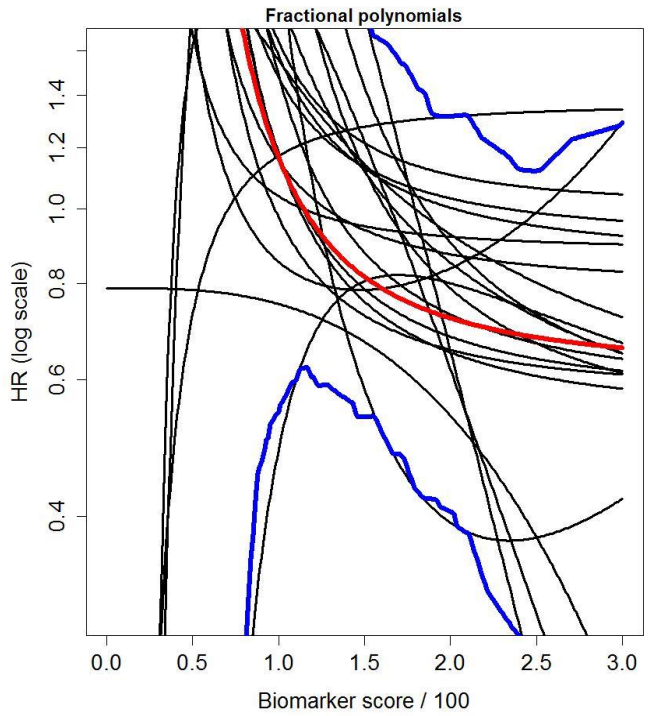
Götte H, Kirchner M, Sailer MO, Kieser M. Simulation-based adjustment after exploratory biomarker subgroup selection in phase II. Stat Med. 2017; 36:2378–2390.

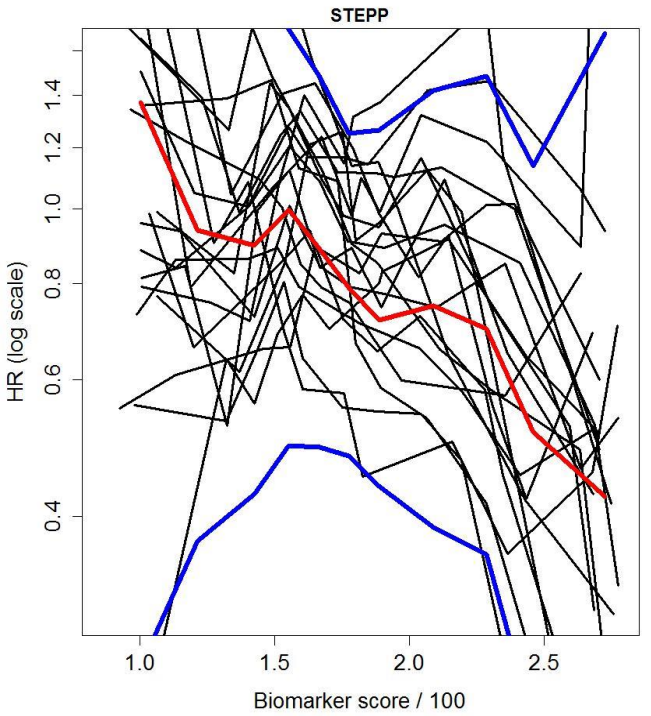
Further example: survival endpoint – cut-off selection

	Full population (130 subjects with 82 events)
HR	0.791
PoS for a future trial with 373 events	55%
Posterior probability $P(HR \leq 0.75 \mid \text{data})$	41%

Further example: survival endpoint – cut-off selection

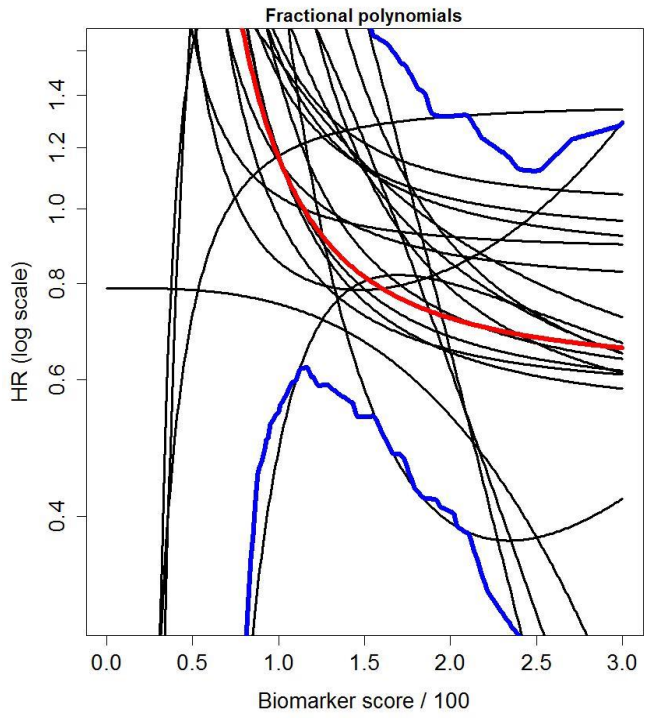
	Full population (130 subjects with 82 events)
HR	0.791
PoS for a future trial with 373 events	55%
Posterior probability $P(HR \leq 0.75 \mid \text{data})$	41%

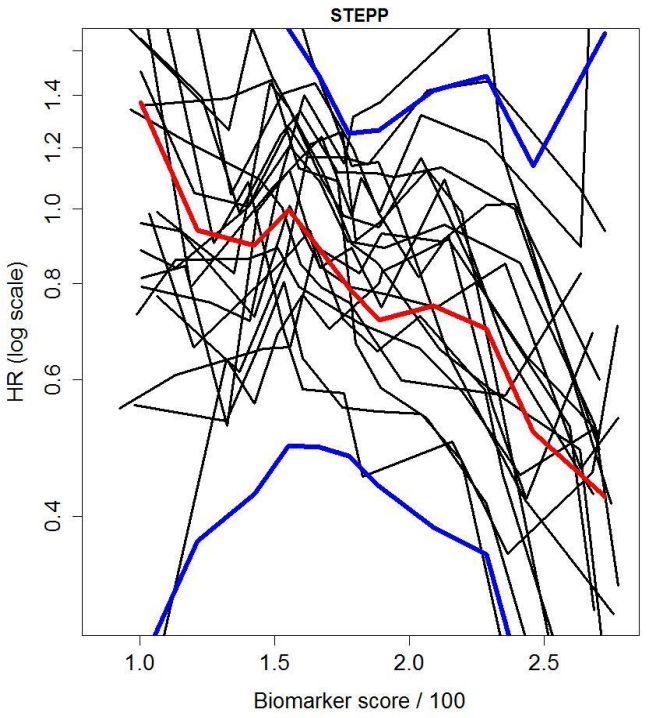




Further example: survival endpoint – cut-off selection

Quantile	Cut-off on original scale		HR (Cox model)
30%	140	<	1.089
		≥	0.697
40%	165	<	0.902
		≥	0.751
50%	184	<	0.895
		≥	0.698
60%	196	<	0.966
		≥	0.597
70%	213	<	1.030
		≥	0.449

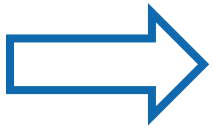




Selection procedure: subset with largest treatment effect (minimum HR) with observed HR for the selected subgroup is smaller than 0.80 and larger than 0.85 for the complement.

Further example: survival endpoint – cut-off selection

	Full population	Selected subset (ignoring selection)
HR	0.791	0.449
PoS for a future trial with 373 events	55%	91%
Posterior probability $P(HR \leq 0.75 \mid \text{data})$	41%	88%

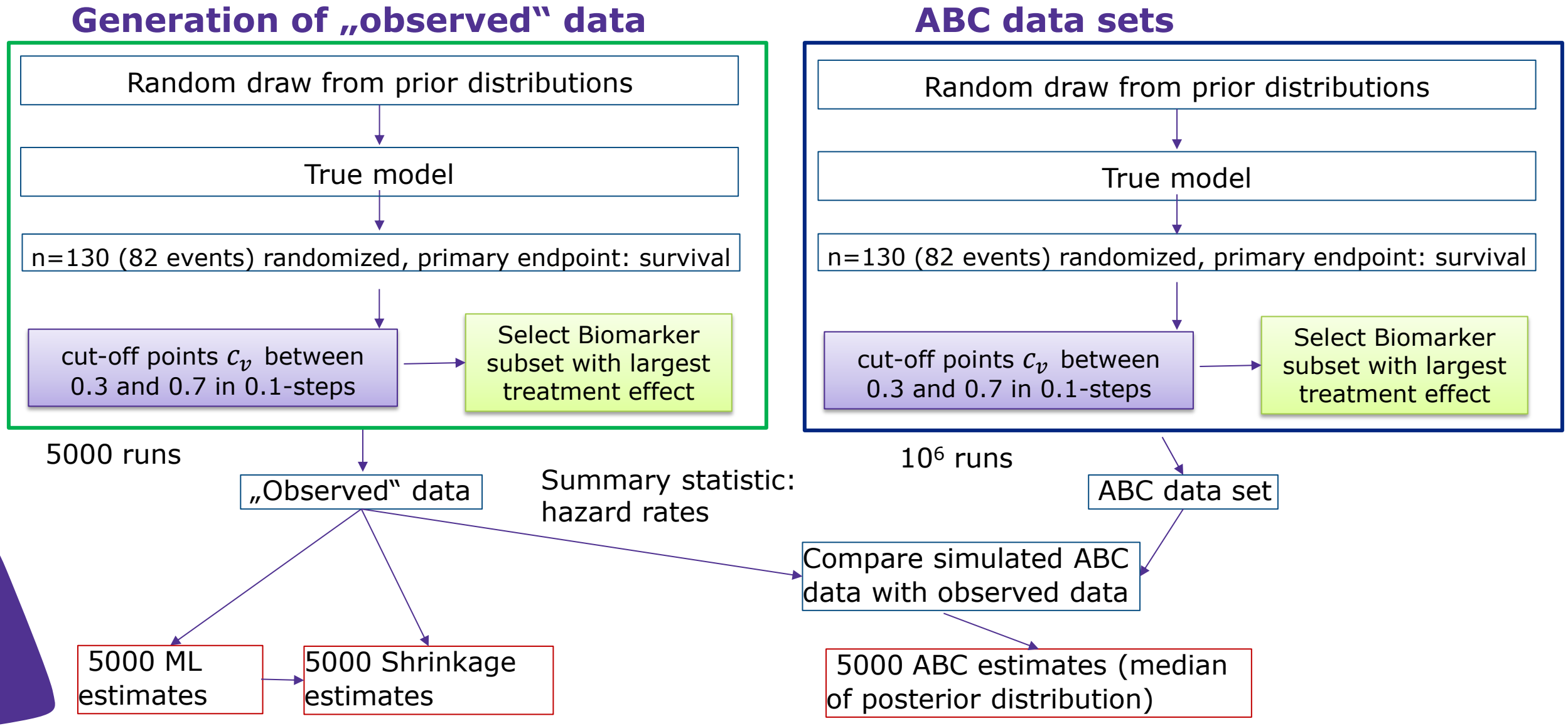


Replication of the subgroup finding in a new trial → Worth the effort?

→ Probability of success (PoS)?

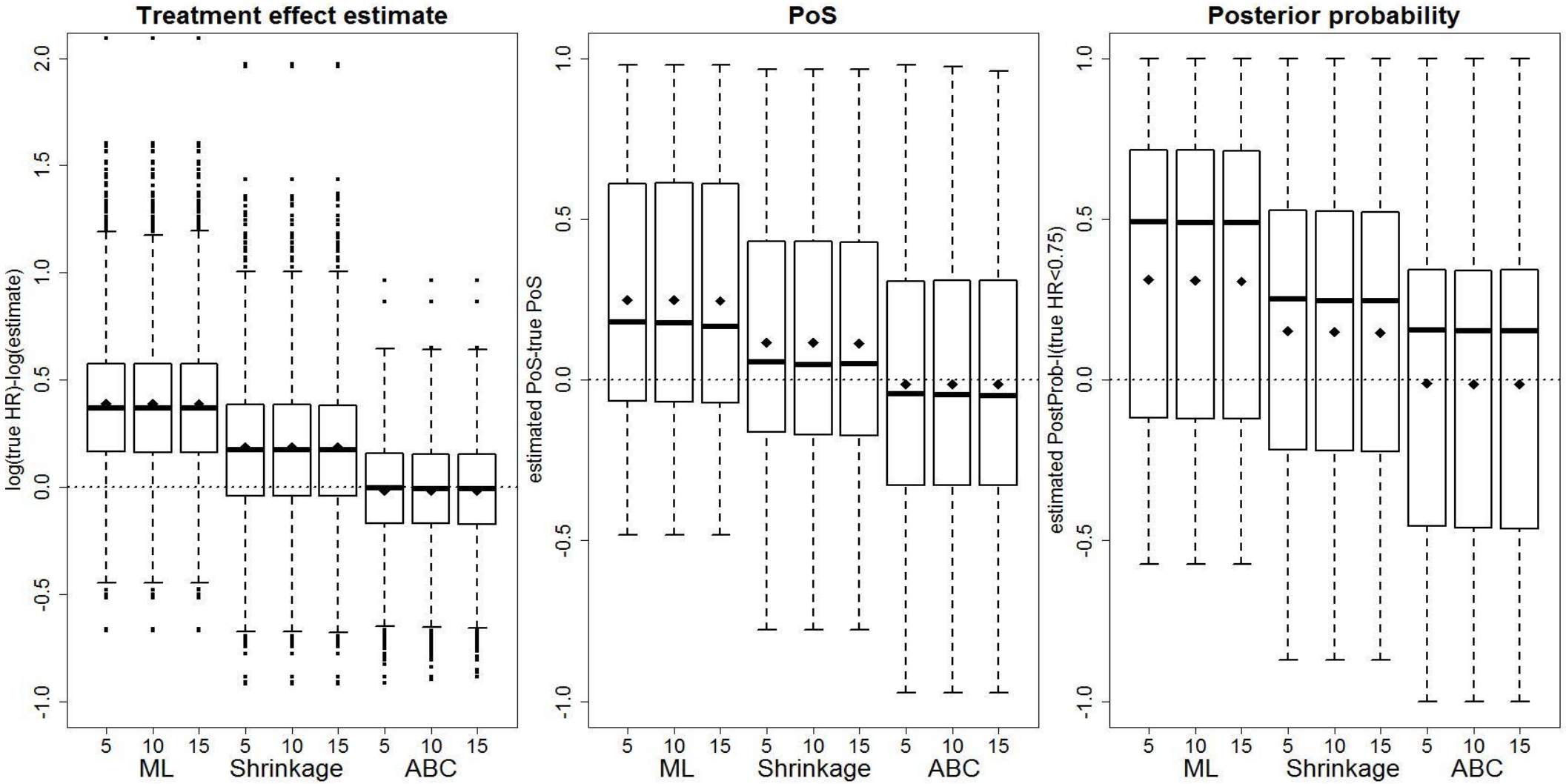
→ Bias-adjusted estimates?

Simulation study: survival endpoint, cut-off selection

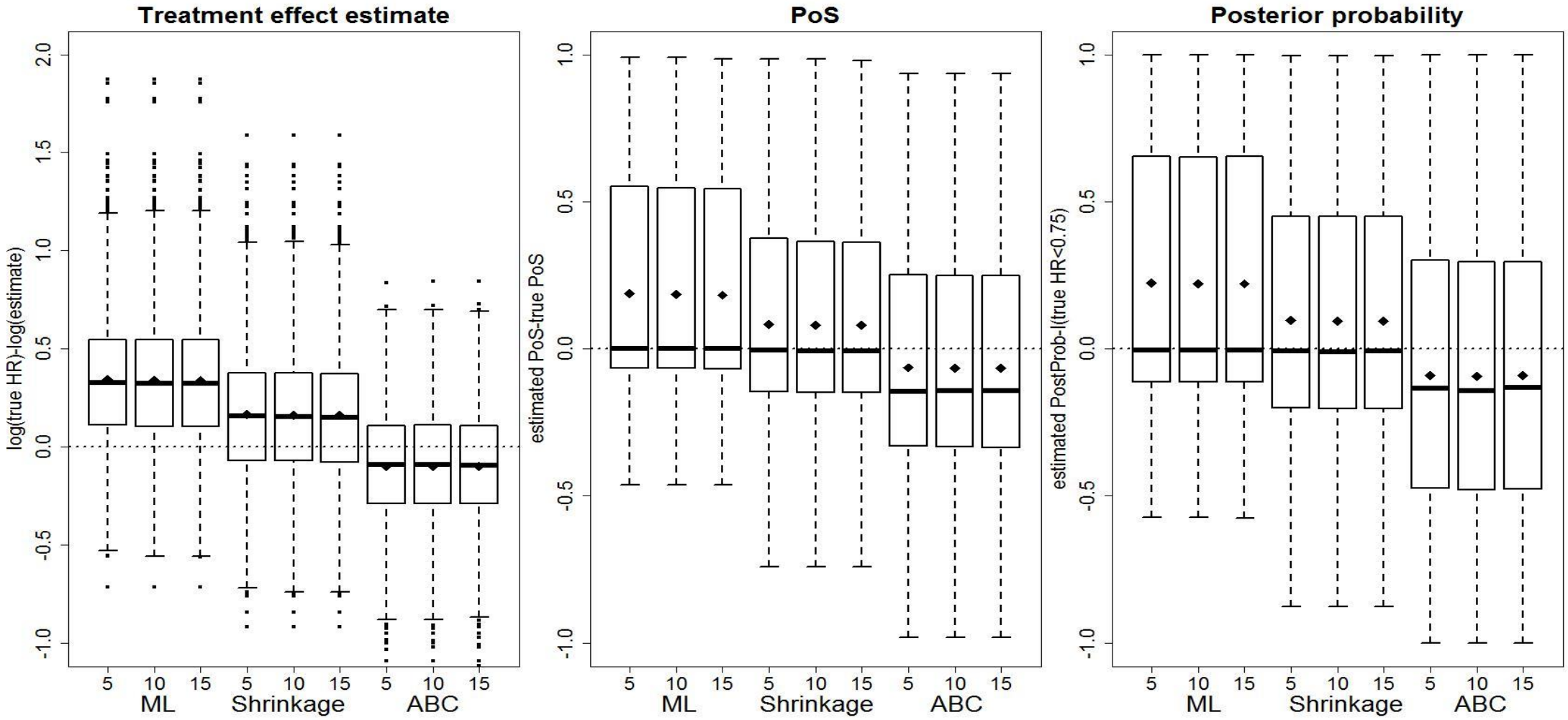


Using same prior for ABC simulation and for generating observations

Simulation: Bias for treatment effect estimate, PoS, posterior probability



Using different prior for ABC simulation and for generating observations
Simulation: Bias for treatment effect estimate, PoS, posterior probability



Results
Coming back to our example

Received: 24 September 2018 | Revised: 4 September 2019 | Accepted: 6 September 2019
DOI: 10.1002/bimj.201800302
Biometrical Journal
RESEARCH PAPER

Adjustment for exploratory cut-off selection in randomized clinical trials with survival endpoint

Heiko Götte[1] | Marietta Kirchner[2] | Meinhard Kieser[2]

	Full population	Selected subset (ignoring selection)	Selected subset (shrinkage)	Selected subset (ABC-adjusted for any selection)	Selected subset (ABC-adjusted for selection of "$>70\%$-quantile")
HR	0.791	0.449	0.571	0.703	0.705
PoS for a future trial with 373 events	55%	91%	83%	73%	70%
Posterior probability $P(HR \leq 0.75 \mid \text{data})$	41%	88%	77%	64%	63%

Conclusion

Coming back to our example

	Full population	Selected subset (ignoring	Selected subset (shrinkage)	Selected subset (ABC-	Selected subset (ABC-adjusted for selection of
PoS fo trial w event					
Poster proba ($HR \leq$ data)					

- Recommendation: continue development with the "Biomarker ≥ 213"-subgroup
  → Treatment effect around 0.7 can be expected in future trials
- Adaptive enrichment designs with the possibility to exclude the complement after the interim analysis may be attractive option for further development

Summary

ABC approach can help to assess whether it is worth reproducing subgroup finding:

- Requires simulation with
  - Fixed selection procedure
  - Choice of prior distributions
  - True effects need to be calculated based on a true model
- Provides bias adjustment that is superior to shrinkage estimation
  - High number of potential subsets increase the chance of observing strongly differential effects
  - Selected subgroup and its complement show strong differences $\Rightarrow$ level of shrinkage is rather low $\Rightarrow$ no adequate bias correction

Outlook

- The problem of selection bias is universal and does not only apply to subgroup selection in randomized phase II trials

  All „new“ designs like platform, basket, umbrella,... trials:

  - Observed effects are compared with each other and focus is on the „best“ results
  - ABC adjustment can be useful as well $\Rightarrow$ further research

- ABC approach can be used where a Bayesian simulation is possible but analytical derivation of Likelihood/ Prior is difficult

Thank you

References

Tanniou J, van der Tweel I, Teerenstra S, Roes KCB. Estimates of subgrou treatment effects in overall nonsignificant trials: To what extent should we believe in them. Pharm Stats. 2017;16:280-295.

Götte H, Kirchner M, Sailer MO. Probability of success for phase III after exploratory biomarker analysis in phase II. Pharm Stat. 2017;16:178–191.

Sunnaker M, Busetto AG, Numminen E, Corander J, Foll M, et al. Approximate Bayesian Computation. PLoS Comput Biol. 2013;9:e1002803.

Götte H, Kirchner M, Sailer MO, Kieser M. Simulation-based adjustment after exploratory biomarker subgroup selection in phase II. Stat Med. 2017;36:2378–2390.

Götte H, Kirchner M, Kieser M. Adjustment for exploratory cut-off selection in randomized clinical trials with survival endpoint. Biometrical Journal. 2019

Back up

ASCO 2017
Subgroup results in phase II should be further investigated

J Clin Oncol 35, 2017 (suppl 4S; abstract 252)

Conclusions:

Although the primary endpoint was not reached in this Phase II all-comer HCC study, the results of the subgroup analysis suggest a population-specific effect for the combination therapy, especially in one which is HBV+. This warrants the further development of this combination as first-line therapy in a well-defined subset of pts with advanced HCC. Clinical trial information: NCT02400788

J Clin Oncol 35, 2017 (suppl; abstr 4123)

Conclusions:

Pts with poor IN status may have treatment benefit of induction CT followed by CRT for LAPC. Clinical trial information: UMIN000006811.

J Clin Oncol 35, 2017 (suppl; abstr 1089)

Conclusions:

In this study, the mOS of pts with Dx+ TNBC who received 0-1 prior lines of therapy appears longer than that of unselected historic controls. ENZA may represent a therapeutic option in pts with AR+ TNBC who would otherwise receive cytotoxic chemotherapy and is currently being evaluated in ENDEAR, a phase 3 study in pts with Dx+ advanced TNBC and 0-1 prior lines of therapy. Clinical trial information: NCT01889238

Simulation study by Tanniou et al 2017

Overall result not significant – Subgroup is significant

To investigate the potential bias involved in the subgroup treatment effect estimate, simulation studies were performed with several design parameters: the 1-sided significance levels for the overall and the subgroup test ($\alpha_O = 0.025$ and $\alpha_S = 0.025$, respectively), the power of the overall analysis ($1 - \beta = 0.80$), and the proportion r (0.1, 0.2, 0.3, …, 0.9) of the subgroup of interest in the full study population. The significance levels α_O and α_S for the overall test and the subgroup test, respectively, constitute the selection part of the subgroup finding hence are the source of any potential bias.

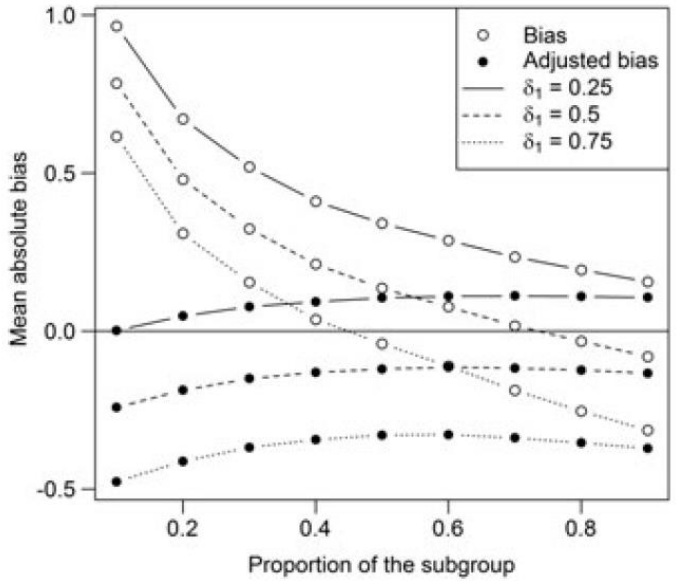


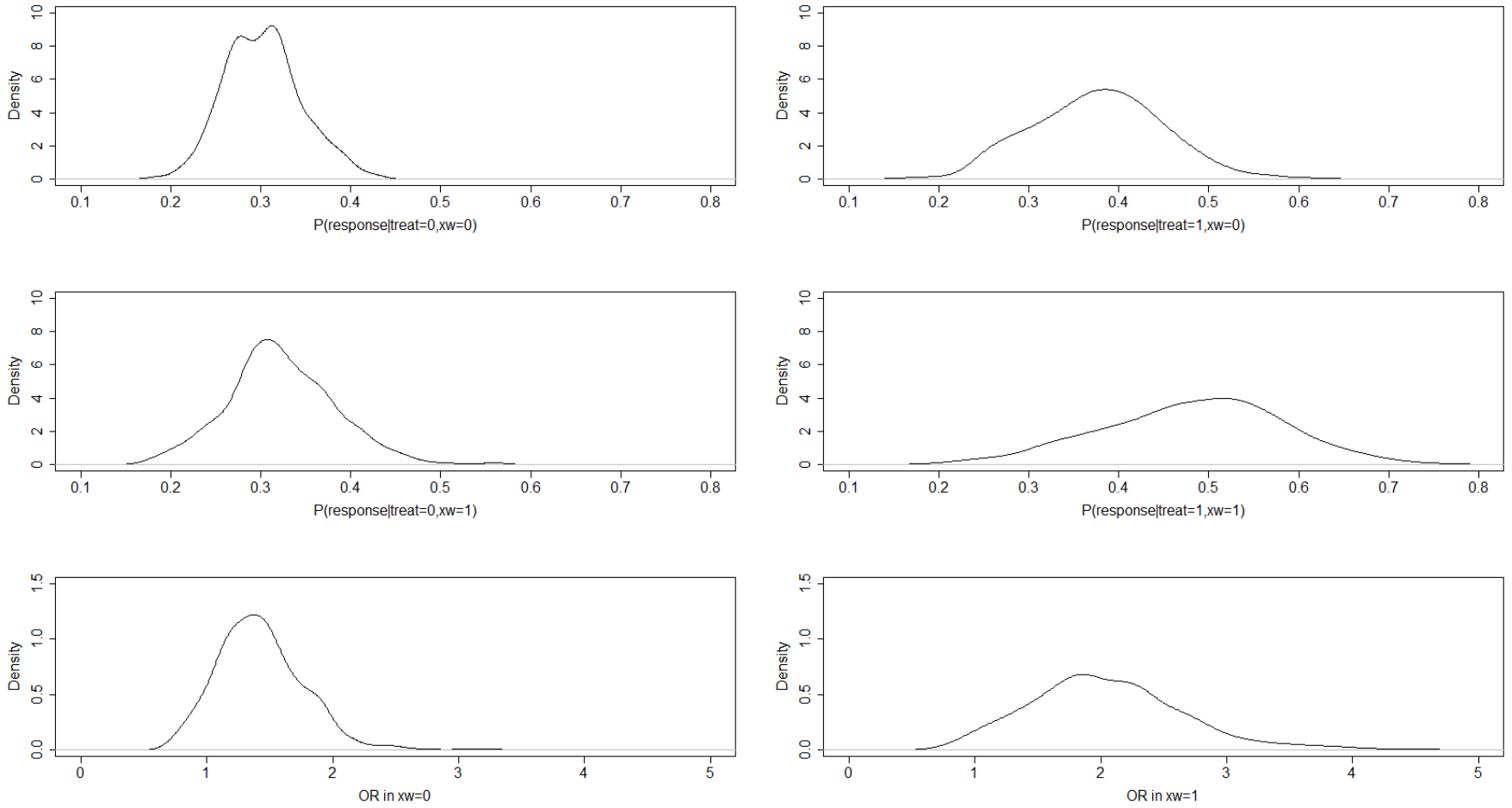
FIGURE 7 Absolute bias and shrinkage correction for continuous outcomes as a function of the proportion of the subgroup for $\Delta = 0.5$ (scenarios 4-6)

Pharmaceutical Statistics. 2017;16:280-295

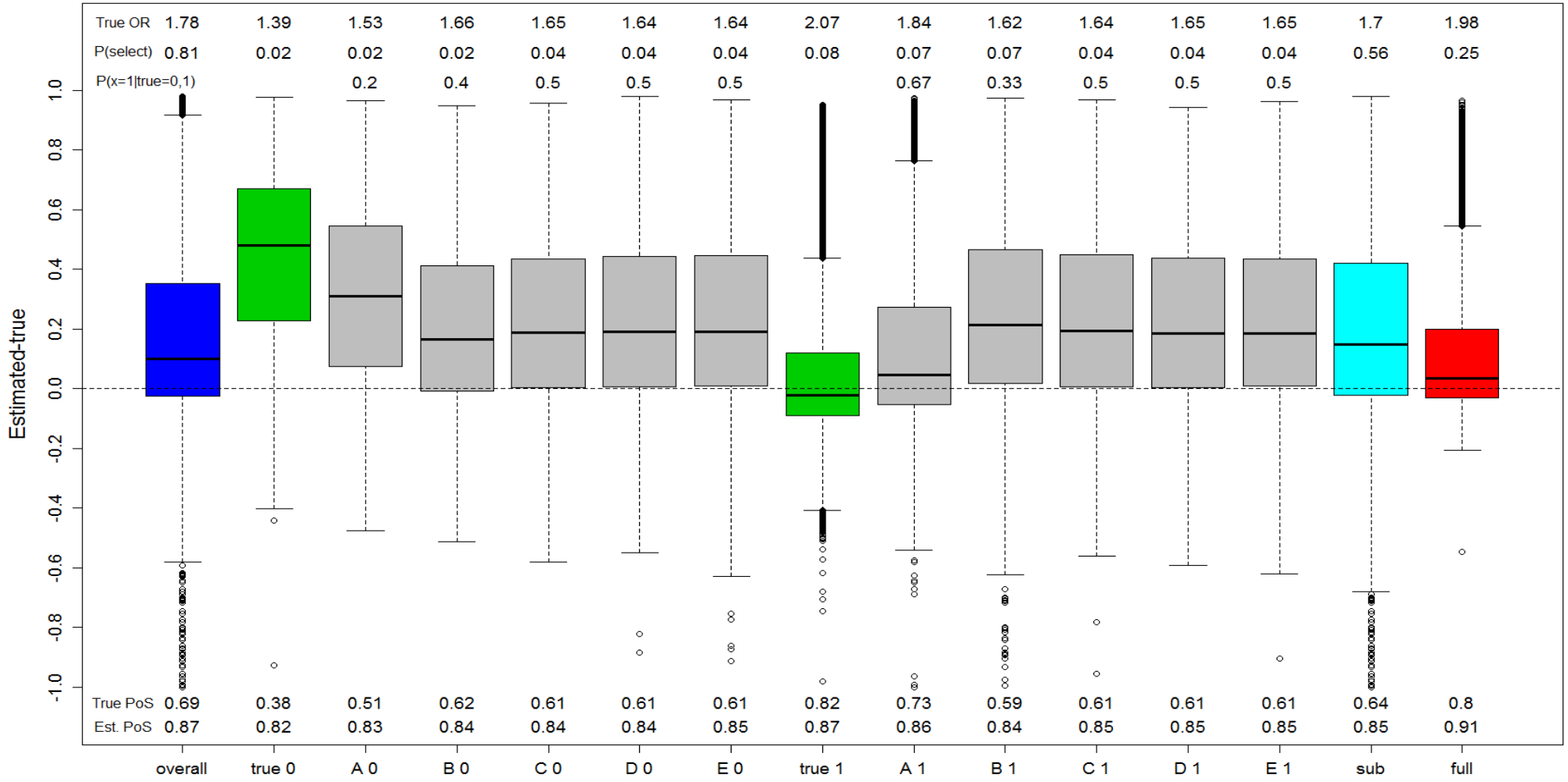
Generation of true effects
Differential prior

$$P(Y_i = 1|treat_i, x_{i\omega}) = \frac{e^{(\vartheta+\theta*treat_i+\lambda*x_{i\omega}+\tau*treat_i*x_{i\omega})}}{1+e^{(\vartheta+\theta*treat_i+\lambda*x_{i\omega}+\tau*treat_i*x_{i\omega})}}$$

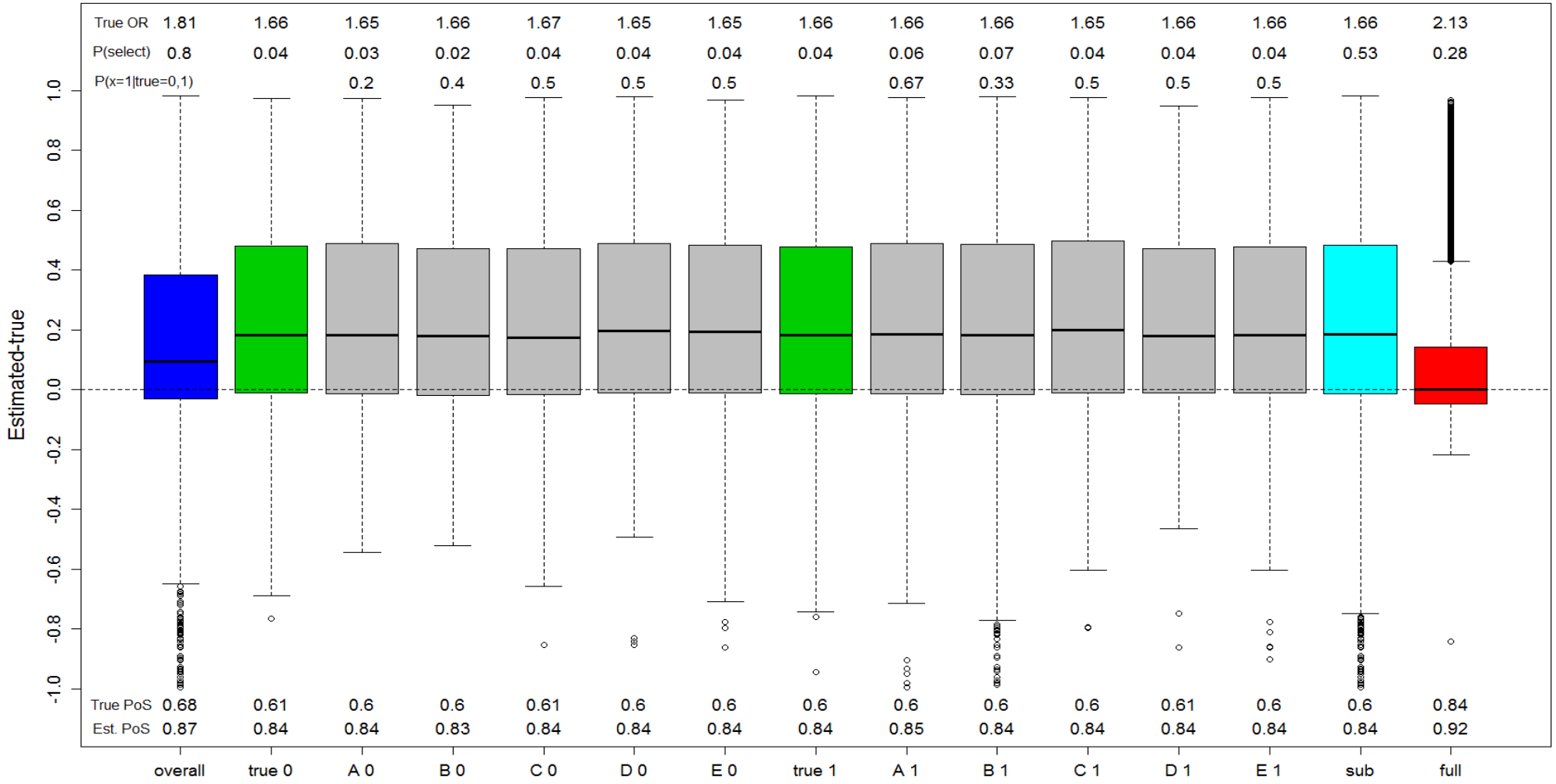
$$\vartheta_\omega \sim N\left(\log\left(\frac{0.3}{0.7}\right), \left(\frac{1}{5}\right)^2\right),$$
$$\theta_\omega \sim N\left(\log(1.4), \left(\frac{1}{4}\right)^2\right),$$
$$\lambda_\omega \sim N\left(\log(1.1), \left(\frac{1}{5}\right)^2\right),$$
$$\tau_\omega \sim N\left(\log(1.4), \left(\frac{1}{5}\right)^2\right).$$



Estimation of PoS – differential effect



Estimation of PoS – overall effect



Details ABC approach, binary endpoint

ABC approach with subgroup selection

- Sample from prior distributions for (a_1, a_2, a_3, a_4), $P(x_{i\omega} = 1)$, $P(x_{ik} = 1 \mid x_{i\omega} = 0)$ and $P(x_{ik} = 1 \mid x_{i\omega} = 1)$ for all k
- Generate data: $y = (resp_i, treat_i, x_{i1}, \dots x_{iK})$
  - Generate $treat_i, x_{i\omega}$
  - Generate $resp_i$ based on $P(Y_i=1 \mid treat_i, x_{i\omega}) = \frac{e^{(a_1+a_2*treat_i+a_3*x_{i\omega}+a_4*treat_i*x_{i\omega})}}{1+e^{(a_1+a_2*treat_i+a_3*x_{i\omega}+a_4*treat_i*x_{i\omega})}}$
  - Generate $x_{i1}, \dots, x_{i\omega-1}, x_{i\omega+1}, \dots, x_{iK}$ based on $P(x_{ik} = 1 \mid x_{i\omega} = 0)$ and $P(x_{ik} = 1 \mid x_{i\omega} = 1)$
- Select subset ι for which the observed effect $\hat{\theta}_\iota$ is the maximum among all components of $(\hat{\theta}_1, \dots, \hat{\theta}_{2K})$
- If $S_{obs,select} \in [s_{select} - \varepsilon, s_{select} + \varepsilon]$ and $S_{obs,full} \in [s_{full} - \varepsilon, s_{full} + \varepsilon]$ then $w_j = 1$ otherwise $w_j = 0$

$$s_{select} = \left(r_{selectC} = \frac{n_{selectCR}}{n_{selectCN} + n_{selectCR}}, r_{selectE} = \frac{n_{selectER}}{n_{selectEN} + n_{selectER}} \right)$$

$$s_{full} = \left(r_{fullC} = \frac{n_{fullCR}}{n_{fullCN} + n_{fulllCR}}, r_{fulllE} = \frac{n_{fullER}}{n_{fullEN} + n_{fullER}} \right)$$

ε=0.02 N_{sim}=10^6 simulations

ABC approach with subgroup selection

- Sample from prior distributions for (a_1, a_2, a_3, a_4), $\mathrm{P}(\mathrm{x_{i\omega}} = 1)$, $P(\mathrm{x_{ik}} = 1 \mid \mathrm{x_{i\omega}} = 0)$ and $P(\mathrm{x_{ik}} = 1 \mid \mathrm{x_{i\omega}} = 1)$ for all k
- Generate data: $y = (resp_i, treat_i, x_{i1}, \dots\ x_{iK})$
  - Generate $\mathrm{treat_i}$,$\mathrm{x_{i\omega}}$
  - Generate $\mathrm{resp_i}$ based on $\mathrm{P(Y_i{=}1 | treat_i, x_{i\omega})} = \frac{e^{(a_1 + a_2 * treat_i + a_3 * x_{i\omega} + a_4 * treat_i * x_{i\omega})}}{1 + e^{(a_1 + a_2 * treat_i + a_3 * x_{i\omega} + a_4 * treat_i * x_{i\omega})}}$
  - Generate $x_{i1}, \dots, x_{i\omega-1}, x_{i\omega+1}, \dots, x_{iK}$ based on $P(\mathrm{x_{ik}} = 1 \mid \mathrm{x_{i\omega}} = 0)$ and $P(\mathrm{x_{ik}} = 1 \mid \mathrm{x_{i\omega}} = 1)$
- Select subset ι for which the observed effect $\hat{\theta}_\iota$ is the maximum among all components of $(\hat{\theta}_1, \dots, \hat{\theta}_{2K})$
- If $\mathrm{S_{obs,select}} \in [\mathrm{s_{select}} - \varepsilon, \mathrm{s_{select}} + \varepsilon]$ and $\mathrm{S_{obs,full}} \in [\mathrm{s_{full}} - \varepsilon, \mathrm{s_{full}} + \varepsilon]$ then $w_j = 1$ otherwise $w_j = 0$
- The posterior $\pi_{ABC}(\theta_{Select} | \mathrm{S_{obs}})$ is then determined by $\hat{F}(t) = \frac{1}{\sum_v^{N_{sim}} w_v} \sum_{j=1}^{N_{sim}} w_j \, \mathbb{I}_{\theta_{\iota j} \le t}$

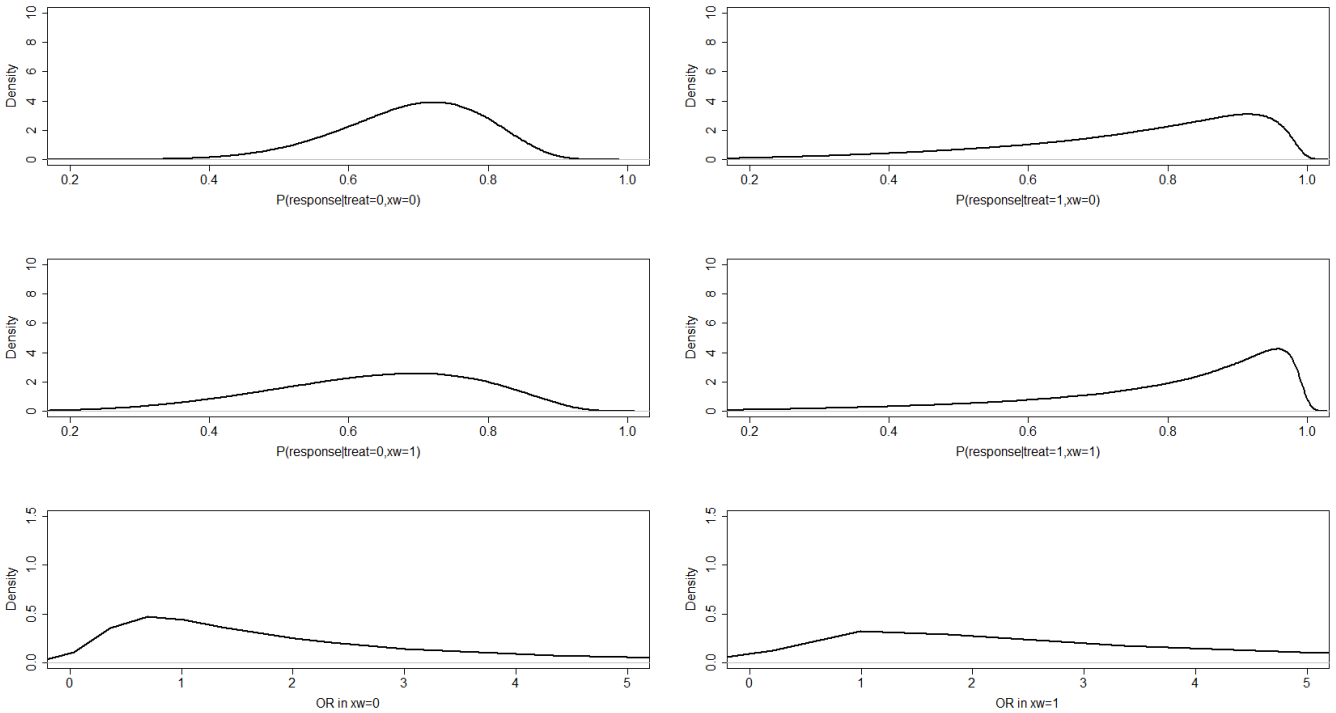
$PoS = \int_{-\infty}^{\infty} \mathrm{P}(p < 0.025 | \theta_{select})\, f_{ABC}(\theta_{select} | \mathrm{S_{obs}}) d\theta_{select}$ is determined by $\frac{1}{\sum_v^{N_{sim}} w_v} \sum_{j=1}^{N_{sim}} \mathrm{P}(p < 0.025 | \theta_{\iota j})\, w_j$

Real data example, binary endpoint

Real data example – Prior for ABC Simulation

- Full population: OR=1.52 (68% (41/60) in the control arm and 77% (46/60) in the experimental arm)
- Subset with the largest observed treatment effect: OR=3.47 (complement: OR=0.65)

$$(a_1, a_2, a_3, a_4) \sim N\left(\begin{pmatrix} \log\left(\frac{0.7}{0.3}\right) \\ \log(1.7) \\ \log(0.8) \\ \log(1.8) \end{pmatrix}, \begin{pmatrix} \frac{1}{4} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{9} \end{pmatrix}\right)$$



Overlap with FIGO status

	ECOG score	BL SOD	Tu ascites	Tu pelvic	Diff grade
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 0)$	0.56	0.40	0.40	0.35	0.16
$P(\text{subgroup level} = 1 \mid \text{FIGO} = 1)$	0.58	0.66	0.32	0.38	0.26

The observed prevalence for "FIGO disease stage=1" is 54%

$$P(x_{i\omega} = 1) \sim Beta(p_\omega * 50, (1 - p_\omega) * 50) \text{ with } p_\omega = 0.5$$

$$P(x_{ik} = 1 \mid x_{i\omega} = 0) \sim Beta(p_{k0} * 50, (1 - p_{k0}) * 50) \text{ and } P(x_{ik} = 1 \mid x_{i\omega} = 1) \sim Beta(p_{k1} * 50, (1 - p_{k1}) * 50)$$

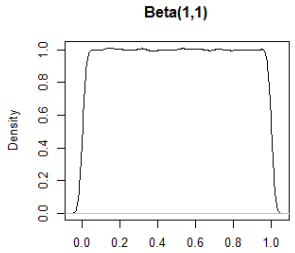
Survival example

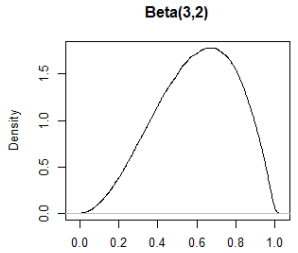
Less complex approach: Shrinkage estimation

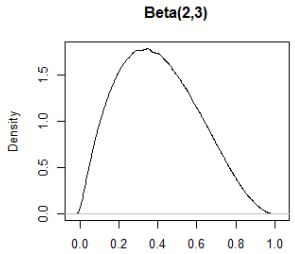
- Bayesian hierarchical model: $\hat{\theta}_{k_{ML}} \sim N\left(\theta_k, \frac{4}{d_k}\right)$ and $\theta_k \sim N(\theta, \tau^2)$ for all subsets k
  - θ and τ are estimated based on observed data:
    - $\hat{\theta}_{select_{shrink}} = \hat{\theta}_{select_{ML}} \cdot (1 - B) + \hat{\theta}_{Full_{ML}} \cdot B$ with $B = \frac{\frac{4}{d_\iota}}{\frac{4}{d_\iota} + \hat{\tau}^2}$
    - Estimate heterogeneity $\hat{\tau}$ based on data from selected subset and complement
      - "SJ"-estimator ("model error variance estimator")

Sidik, K., Jonkman, J.N. (2005). Simple heterogeneity variance estimation for meta-analysis. *Journal of the Royal Statistical Society* C54: 367–384.
Sidik, K., Jonkman, J.N. (2007). A comparison of heterogeneity variance estimators in combining results of studies. *Statistics in Medicine* 26:1964–1981.

Simulation: prior distribution for data and ABC identical







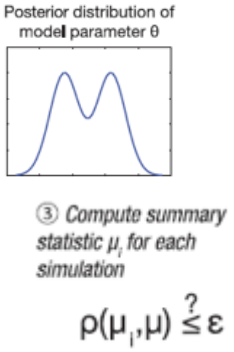
Multiplied with 300
$\rightarrow$ [0;300]

$c_\omega = 150$

$$h(t|treat_i, z_i) = a_1 e^{\left(a_2 \cdot treat_i + a_3 \cdot \mathbb{I}_{(z_i > c_\omega)} + a_4 \cdot treat_i \cdot \mathbb{I}_{(z_i > c_\omega)}\right)}$$

$$a_1 \sim Gamma(shape = -\frac{\log(0.5)}{10} \cdot 150, \text{scale} = \frac{1}{150}),\ a_2 \sim N\left(\log(0.9), \frac{9}{40}\right),\ a_3 \sim N\left(\log(2), \frac{1}{6}\right),\ a_4 \sim N(\log(0.75), 0.38).$$

Adjustment based on ABC approach



- What is needed for ABC approach?
  - Define true model and calculate true HR for selected subset
  - Summary statistic and boundary ε to identify elements of posterior
  - Define number of simulations N_{sim}

- True model: $h(t|treat_i, z_i) = a_1 e^{\left(a_2 \cdot treat_i + a_3 \cdot \mathbb{I}_{(z_i > c_\omega)} + a_4 \cdot treat_i \cdot \mathbb{I}_{(z_i > c_\omega)}\right)}$
  - c_ω defines a "true" subset (1) and its complement (2)
  - Other subsets, not defined by c_ω, lead to time dependent hazard (ratios)
  - $h(t) = p_1(t)h_1(t) + p_2(t)h_2(t)$ with time-dependent weights $p_1(t) = p_1 \frac{S_1(t)}{p_1 S_1(t) + p_2 S_2(t)}$ and $p_2(t) = p_2 \frac{S_2(t)}{p_1 S_1(t) + p_2 S_2(t)}$

- To identify elements of posterior we assume constant hazard and investigate impact of time dependency
  - Hazard rates in full/select and arm C/E within $\pm\ \varepsilon$
  - With ε=0.01 and N_{sim}=10^6 simulations

Time dependent hazard ratios depend on prognostic effects