Westfälische Wilhelms-Universität Münster

medizinische fakultät
Westfälische Wilhelms-Universität Münster

Can policy makers trust the outcomes of Bayesian analyses?
The case of clinical trials and rare diseases

Dr. rer. nat. Joachim Gerß, Dipl.-Stat.

joachim.gerss@ukmuenster.de

Institute of Biostatistics and Clinical Research

IBKF
Institut für Biometrie und Klinische Forschung

wissen.leben
WWU Münster

Regulators‘ view: EMA CHMP
Guideline on clinical trials in small populations

- There are no special methods for designing, carrying out or analysing clinical trials in small populations. There are, however approaches to increase the efficiency of clinical trials.

- The following […] approaches […] may be helpful in particular situations.

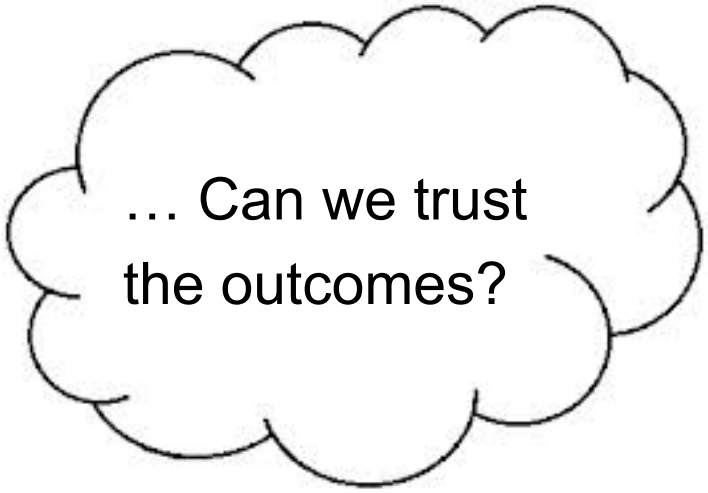
  – Bayesian methods […] are a way to formally combine knowledge from previous data or prior ‘beliefs’ with data from a study. Such methods may be advantageous when faced with small datasets, although introducing prior beliefs is often a concern in drug regulation.

Regulators‘ view: EMA CHMP
Draft Guideline on clinical trials in small populations

- There are no special methods for designing, carrying out or analysing clinical trials in small populations. There are, however approaches to increase the efficiency of clinical trials. Further, some methodological approaches, not acceptable in large trials, may be considered acceptable for trials in small and very small populations.
- The following […] approaches […] may be helpful in particular situations.
  - Relaxing the type I error boundary increases the risk of false positive trial results. There are however, situations where such an approach is acceptable.
  - Bayesian methods […] are a way to formally combine knowledge from previous data or prior ‘beliefs’ with data from a study. Such methods may be advantageous when faced with small datasets, although introducing prior beliefs is often a concern in drug regulation. […] being able to use knowledge of likely effects of drugs due to their chemical form, likeness to other existing compounds, mechanism of action, and so on, is a very valuable addition to sparse data.

Popular Bayesian Methods

- Combination of knowledge from previous data or prior ‘beliefs’ with data from a current study
- Dose finding: *Continual reassessment method*
- *Response-adaptive randomization*
- Prediction of the study result using *predictive probabilities*
- Borrowing of information across related subpopulations

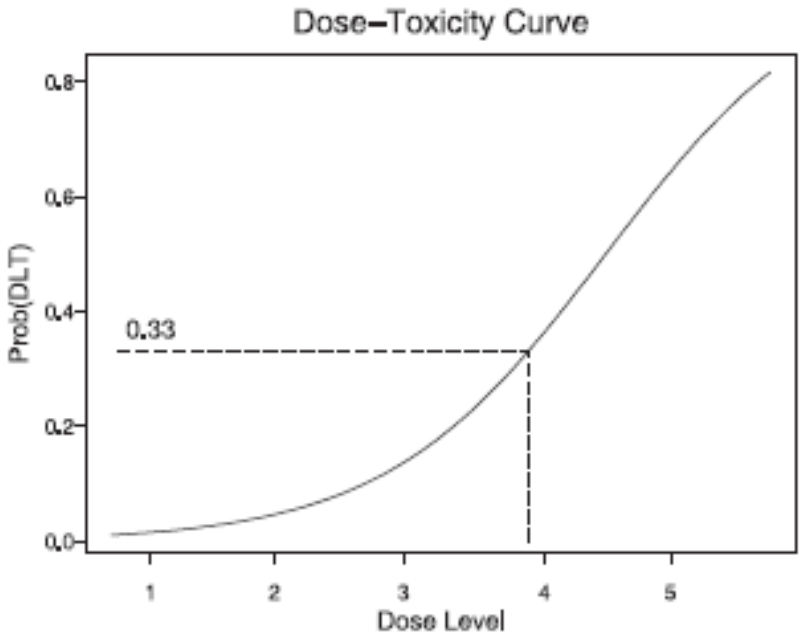


Contents

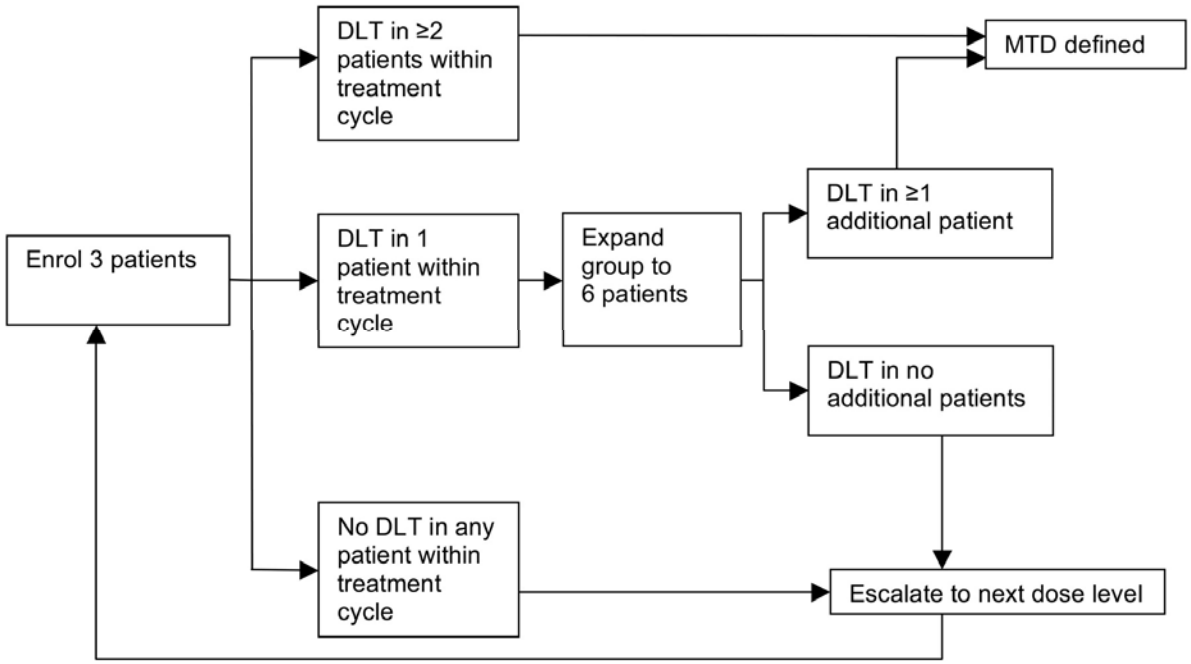
Contents

1. Dose finding in Phase I trials

- Set a starting dose and further increasing dose levels
- Define a *dose-limiting toxicity* (DLT)
- Define the *target toxicity level* (TTL), typically TTL = 20% - 33%
- Identify the *maximum tolerated dose* (MTD), i.e. the dose with Prob(DLT) = TTL



1. Dose finding in Phase I trials
Traditional 3+3 Design



Example

	Dose Level				
Cohort	1	2	3	4	5
1	0/3				
2		0/3			
3			1/3		
4			0/3		
5				2/3	
MTD			***		

1. Dose finding in Phase I trials
The Bayesian Continual Reassessment Method (CRM)

1. Parametric model of the dose-toxicity relationship $p(d) = exp\{3+ad\} / (1+exp\{3+ad\})$
2. Set target toxicity level TTL=30%.
3. Treat the first patient at the dose level closest to the current estimate of the MTD.
4. Observe the toxicity outcome.
5. Obtain the *posterior distribution* of the probability of toxicity $p(d)$ at each dose level d.
6. Treat the next patient at the dose level closest to the updated estimate of MTD.
7. Repeat Steps 4-6.

Example

	Dose Level	Tox	Dose Level 1	Dose Level 2	Dose Level 3	Dose Level 4	Dose Level 5	Dose Level 6
			p=0.05	p=0.10	p=0.20	**p=0.30**	p=0.50	p=0.70
Patient 1	4	0	p=0.00	p=0.00	p=0.02	p=0.04	p=0.14	**p=0.38**
Patient 2	6	1	p=0.05	p=0.10	p=0.20	**p=0.31**	p=0.51	p=0.70
Patient 3	4	1	**p=0.34**	p=0.45	p=0.57	p=0.65	p=0.76	p=0.84
Patient 4	1	0	p=0.16	**p=0.25**	p=0.39	p=0.49	p=0.65	p=0.79

1. Dose finding in Phase I trials
Bayesian CRM versus Traditional 3+3 Design

Operating characteristics (Simulation studies)

Target toxicity level (TTL)

average sample size

Pct. of patients developing DLT

		Dose 1	2	3	4	5	Ave N	% DLT
Scenario 1	P(DLT):	0.05	0.15	0.30	0.45	0.60		
3+3	% patients	26.0	32.5	27.2	12.1	2.3	15.2	21.1
	% MTD	20.5	42.7	27.5	5.7	0		
CRM with cohort size 1	% patients	15.6	24.1	34.7	19.0	6.7	18.5	27.0
	% MTD	1.0	21.4	52.4	23.0	2.2		
CRM with cohort size 3	% patients	21.3	31.4	29.1	15.8	2.5	19.0	23.3
	% MTD	1.5	22.6	49.8	23.7	2.4		

% of patients treated at target toxicity level (TTL)

% of patients treated with doses above TTL

% of dose level selected as the MTD

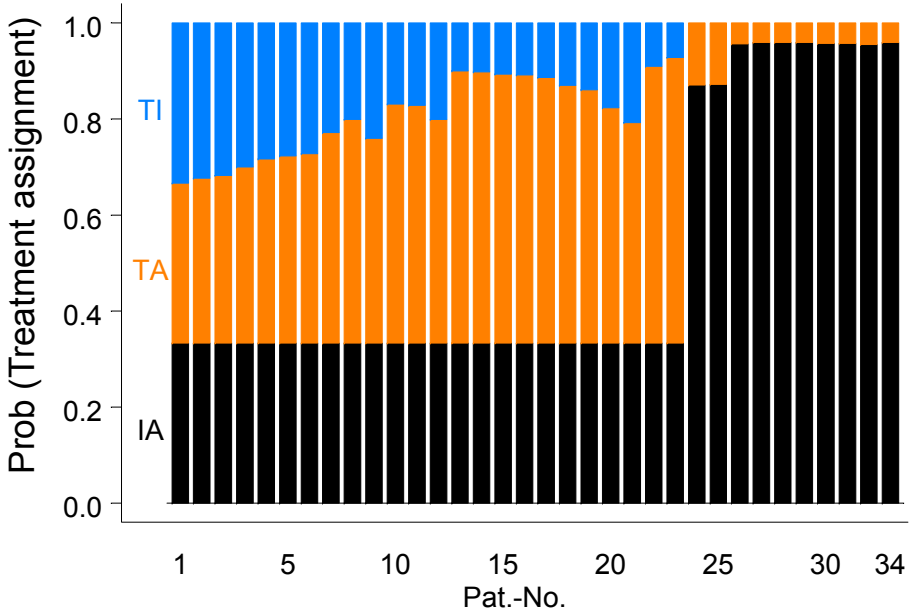
Results

- CRM more accurate identifies the MTD compared to 3+3 design.
- On the other hand, CRM treats more patients at the MTD level and above. CRM with cohort size of 3 instead of 1 offers some protection.

Contents

2. Response-adaptive randomization

- Consider a randomized two- or multi-arm clinical trial
- *Response-adaptive randomization*:
  Randomized Treatment Assignment not with equal and fixed probabilities, but increased assignment of patients to more promising treatments
-



2. Response-adaptive randomization
Example: Phase IIB design with binary response

- 2 treatment arms k=1,2
- Denote θ_k the response probability in arm $k \in \{1,2\}$
- <u>Goal:</u> Find the treatment arm k corresponding to the largest response probability θ_k
- <u>Algorithm</u>

1. Randomize the first 14 patients to treatment arm 1 and 2 with equal probability 1/2.
2. After each observed outcome, compute the (posterior) probability of each arm k to be the best arm, using all currently available data („Prob(arm k is best)“)

any time during the trial

3. „*Early loser*“: If Prob(arm k is best) < p_L= 0.025, then arm k is declared the loser and is suspended.
4. „*Early winner*“: If Prob(arm k is best) > p_U=0.975, then arm k is declared the winner and the trial is stopped early.
5. „*Futility*“: If Prob(θ_k>0.5|data)<0.05, then arm k is declared futile and is suspended.
6. Assign patients to treatment groups with probability proportional to Prob(arm k is best)^c (with tuning parameter c=1), but never lower than 0.1.
7. „*Final winner*“ at max. no. pat.=60: If Prob(arm k is best)>p^*_U=0.9, declare arm k winner

2. Response-adaptive randomization
Example: Simulated operating characteristics

1. „Null" scenario
2. „Most likely" scenario
3. „Optimistic" scenario

Scenario 1
Average Trial Length: 22.5 months

Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%, 97.5%)
Arm1	0.55	0.01	0	0.11	19.6 (5, 38)
Arm2	0.55	0.16	0.11	0	35.6 (8, 53)

Type I error = 0.17

Scenario 2
Average Trial Length: 16.4 months

Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%, 97.5%)
Arm1	0.55	0	0	0.55	10.1 (4, 22)
Arm2	0.7	0.74	0.55	0	30.8 (4, 51)

Scenario 3
Average Trial Length: 10.8 months

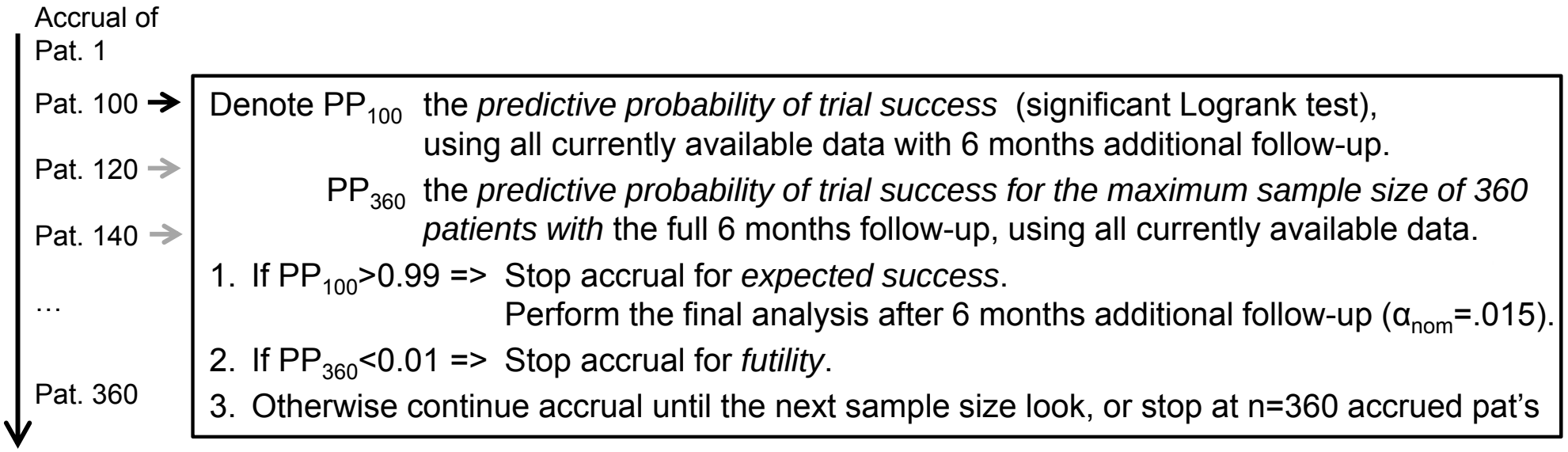
Arm	True Pr (success)	Pr (select)	Pr(select early)	Pr(stop early)	# Patients (2.5%, 97.5%)
Arm1	0.55	0	0	0.89	7.01 (4, 16)
Arm2	0.8	0.96	0.89	0	20.1 (4, 51)

Power

Contents

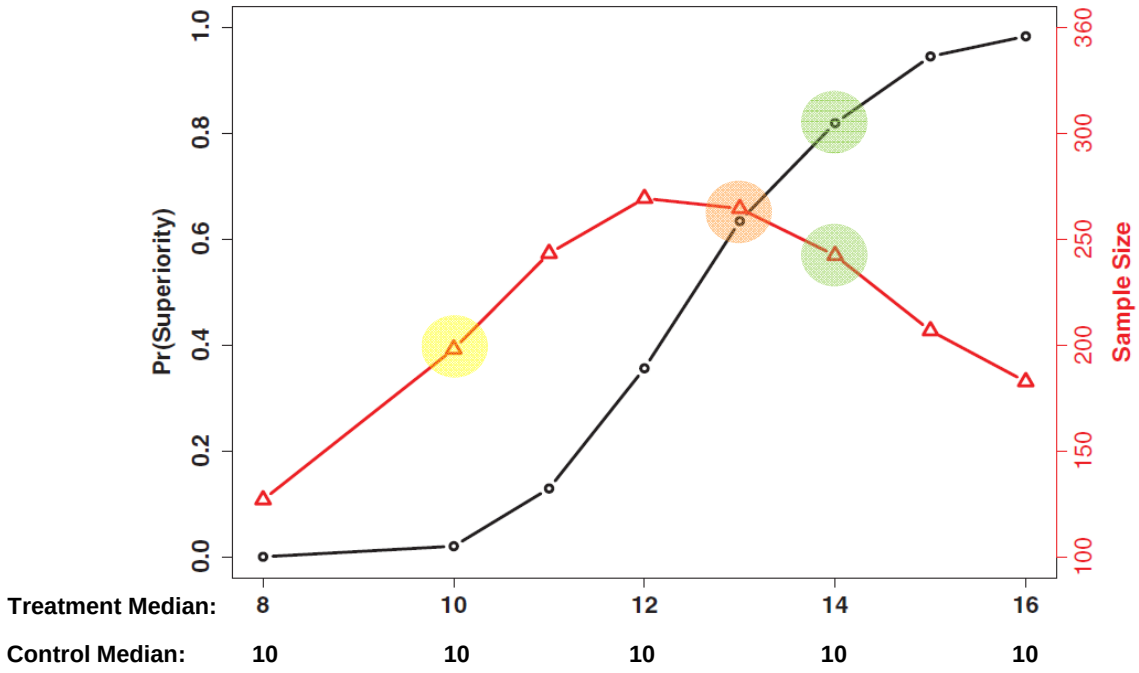
3. Interim analyses using predictive probabilities
Example: Randomized survival trial

- Standard frequentist design:
  - 360 subjects
  - 1:1 randomization
  - 6-month follow-up
  - Final analysis: Logrank test with two-sided significance level α=0.05
  - In case median survival times = 13 versus 10 weeks => power=70%
    In case median survival times = 14 versus 10 weeks => power=89%
- Alternative Bayesian adaptive design:
  - As above, but with additional *interim analyses*:



3. Interim analyses using predictive probabilities Example: Simulated Operating characteristics

- Standard frequentist design:
  - In case median survival times = 13 versus 10 weeks => power=70%
    In case median survival times = 14 versus 10 weeks => power=89%
- Alternative Bayesian adaptive design:



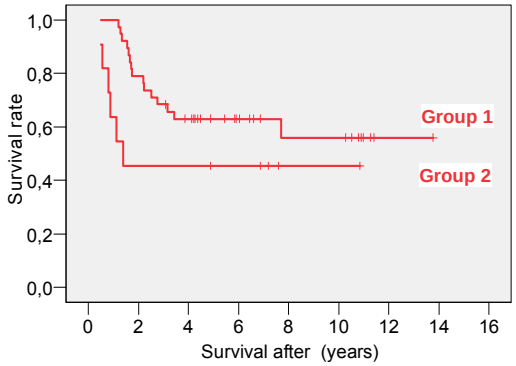
- If the treatment is ineffective the design determines this very effciently (mean sample size = 202).
- Median survival times = 13 vs. 10 weeks => power = 63.4%, with a mean sample size of only 271.
- Median survival times = 14 vs. 10 weeks => power = 81.9%, with a mean sample size of only 248.

Contents

4. Combination of prior ‘beliefs’ with data from a study
Randomized survival trial

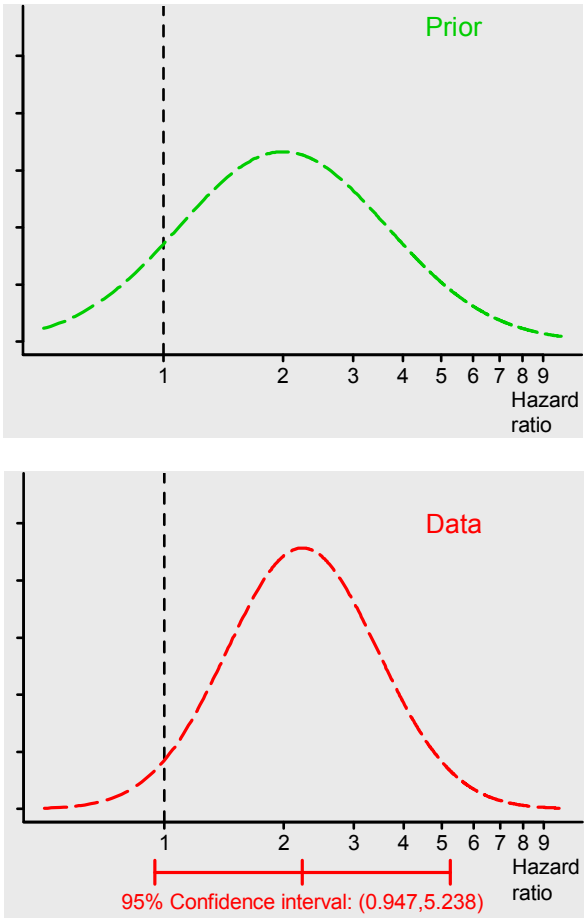
Classical „frequentist“ statistical analysis

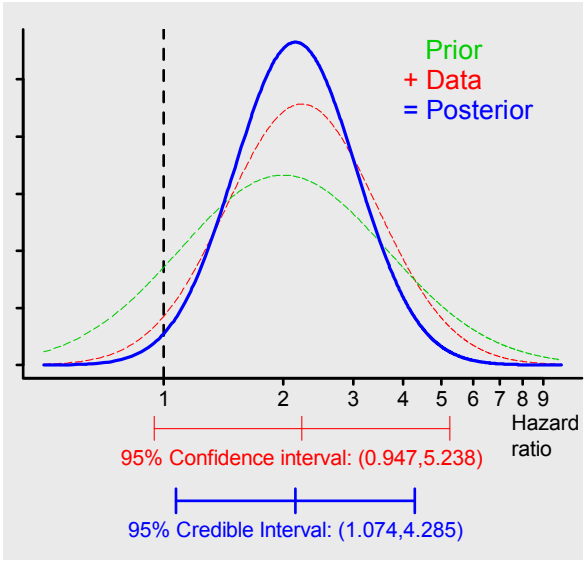
Example 1



HR=2.227
95% CI 0.947-5.238
p=0.0990

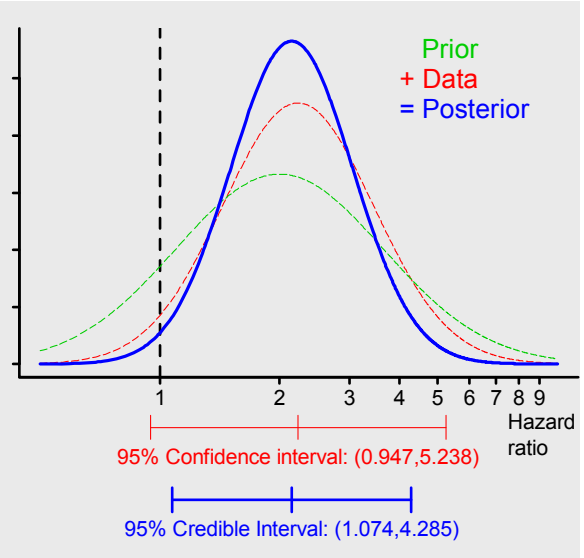
Bayesian analysis



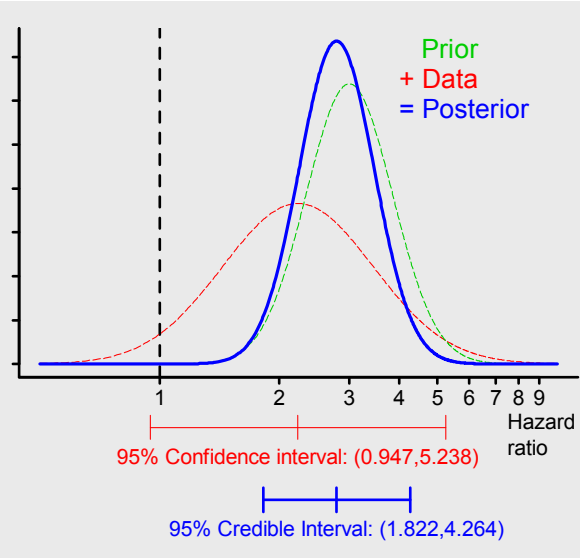


4. Combination of prior ‘beliefs’ with data from a study
Randomized survival trial

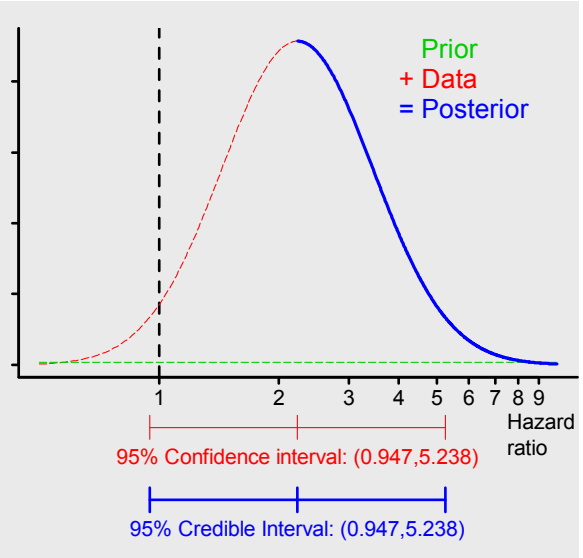
Example 1



Example 2

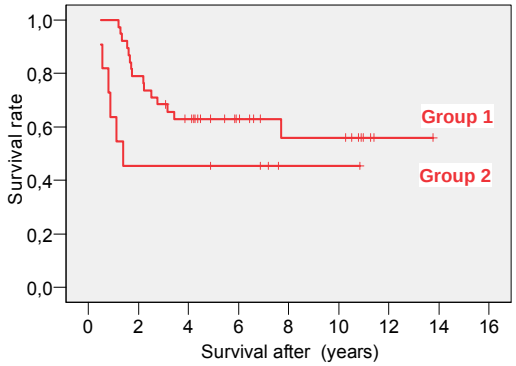


Example 3
„Noninformative“ prior



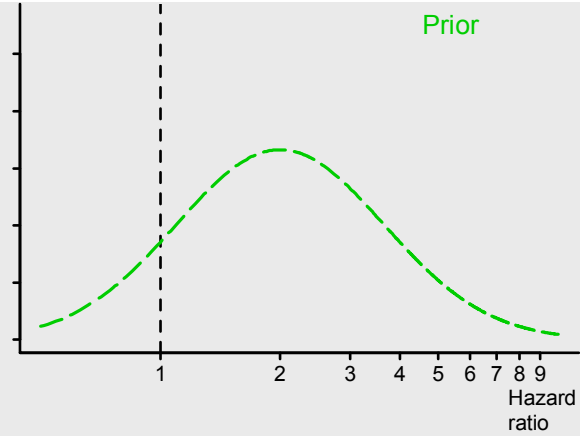
4. Combination of prior ‘beliefs’ with data from a study
Randomized survival trial: Frequentist & Bayesian analysis

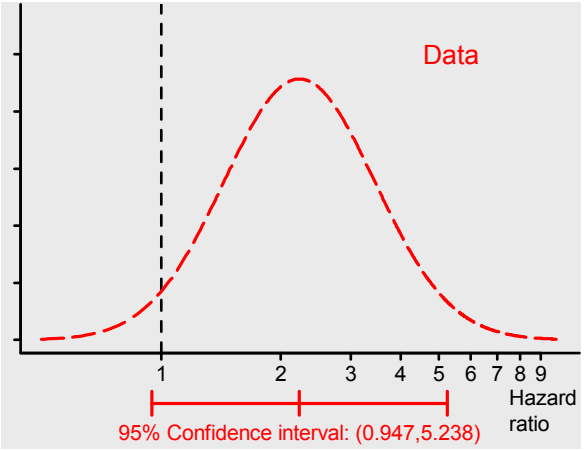
Classical „frequentist“ statistical analysis

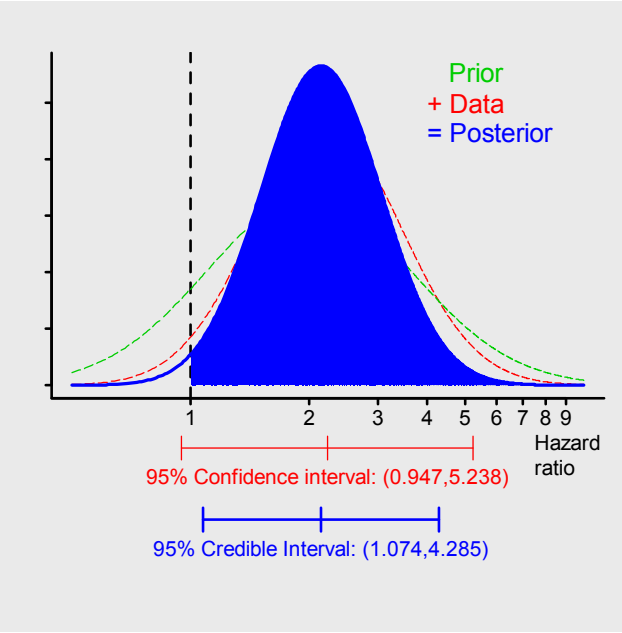


HR=2.227
95% CI 0.947-5.238
p=0.0990

Bayesian analysis



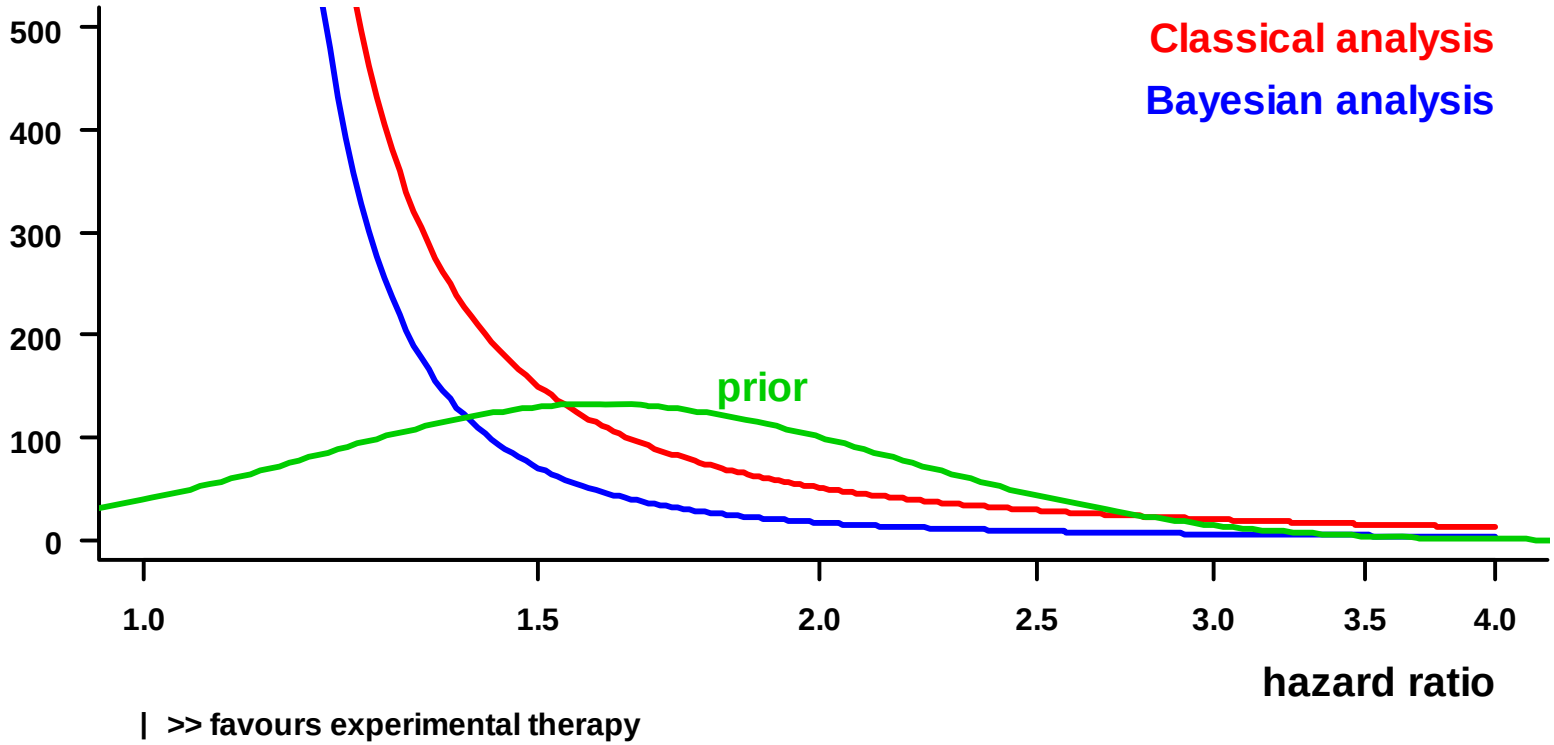




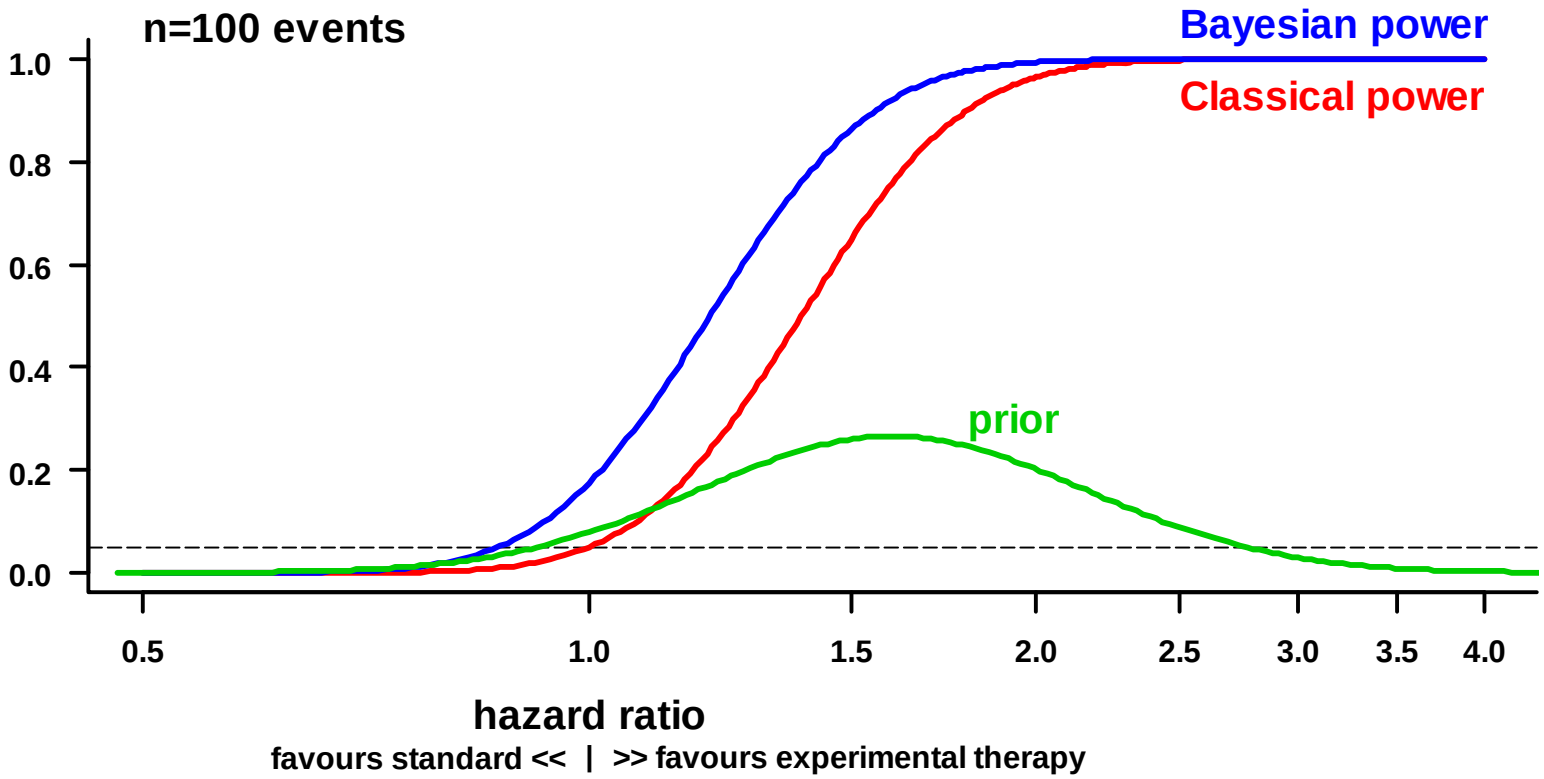
If p≤0.05 (<=> 1∉Confidence Interval)
=> „significant“

Prob(HR>1|Data) ≥ 97.5%
=> „significant“

4. Combination of prior ‘beliefs’ with data from a study Sample size



4. Combination of prior ‘beliefs’ with data from a study Type I error and power



Contents

Summary and Conclusion

Examples:

2. Bayesian response-adaptive randomization →
3. Interim analyses using predictive probabilities →
4. Combination of prior ‘beliefs’ with data from a study →

Supplementary task	Final analysis
Bayesian response-adaptive randomization	Bayes
Bayesian predictive probabilities	Frequentist
----------	Bayes

Summary and Conclusion

Examples:

2. Bayesian response-adaptive randomization →
3. Interim analyses using predictive probabilities →
4. Combination of prior 'beliefs' with data from a study →

Supplementary task	Final analysis
Bayesian response-adaptive randomization	Bayes
Bayesian response-adaptive randomization	Frequentist
Bayesian predictive probabilities	Bayes
Bayesian predictive probabilities	Frequentist
----------	Bayes
----------	Frequentist

All Bayesian approaches: … using either an informative or a noninformative prior

Summary and Conclusion: Bayesian methods in clinical trials – Validity of analysis outcomes

	Trials in large populations	Trials in small populations, Medical device trials
Phase I trials (toxicity)	✓	✓
Phase II trials (efficacy)	✓	✓
Phase III trials (supplementary tasks)	✓	✓
Phase III trials (final analysis)		✓ **Cave!**

Summary and Conclusion: Bayesian methods in clinical trials – Validity of analysis outcomes and Regulators‘ view

	Trials in large populations		Trials in small populations, Medical device trials	
Phase I trials (toxicity)	✓	✓	✓	✓
Phase II trials (efficacy)	✓	✓	✓	✓
Phase III trials (supplementary tasks)	✓	✓	✓	✓
Phase III trials (final analysis)			✓ Cave!	(✓)

Summary and Conclusion:
Can policy makers trust outcomes of Bayesian analyses?

- **Early phase clinical trials („in-house studies“ w/o strict regulatory control):**

  Yes! And many Bayesian approaches are more powerful than classical approaches.

- **Late phase trials**

  - Yes! Bayesian methods can be constructed so that (frequentist long-run) operating characteristics are maintained (type I error, power)
  - … But Bayesian methods in a strict corset of frequentist quality criteria are usually not much more powerful than classical frequentist methods („Bayesians cannot perform magic.“)
  - Particularly flexible and powerful Bayesian methods are those that
    - apply some kind of paradigm shift w.r.t. quality criteria and/or
    - follow less stringent requirements w.r.t. e.g., type I error control

Summary and Conclusion:
Can policy makers trust outcomes of Bayesian analyses? – The case of clinical trials and rare diseases

- **Early phase clinical trials:** Apply powerful Bayesian approaches!
- **Late phase trials**
  - Large scale confirmatory trials are often not feasible
  - Some kind of paradigm shift w.r.t. quality criteria and/or less stringent requirements w.r.t. e.g., type I error control may be accepted.
  - Apply powerful Bayesian approaches, paying attention to
    - choose the appropriate model carefully,
    - choose the inputted (prior) information carefully and
    - check (classical) operating characteristics (type I error, power).

Literature

- Berry SM, Carlin BP, Lee JJ , Müller P (2010): Bayesian Adaptive Methods for Clinical Trials. Chapman & Hall/CRC Biostatistics.
- Spiegelhalter DJ, Abrams KR, Myles JP (2004): Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Wiley Series in Statistics in Practice.
- European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP) (2006): Guideline on Clinical Trials in Small Populations (CHMP/EWP/83561/2005). http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003615.pdf
- U. S. Food and Drug Administration, Center for Devices and Radiological Health (2010): Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials. http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm
- Tan SB, Dear KBG, Bruzzi P, Machin D (2003): Strategy for randomised clinical trials in rare cancers. British Medical Journal 327;47-49.