BBS spring seminar on synthetic controls: summary & follow-up

Simon Wandel, Director Statistics

4th EFSPI Workshop on Regulatory Statistics, Basel

September 23, 2019

NOVARTIS | Reimagining Medicine

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of Novartis. Novartis does not guarantee the accuracy or reliability of the information provided herein.

Agenda

Seminar highlights

Follow-up

Conclusion

Seminar highlights

Seminar on synthetic controls: overview

- Tom Brookland (Roche): *RWD/RWE Global Regulatory Overview*
- Kaspar Rufibach and Hans Ulrich Burger (Roche): *External controls in drug development*
- Somnath Sarkar (Flatiron): *Considerations for Developing External Control Arm from Real-World Data*
- Laurence Colin and Yue Li (Novartis): *Making better use of early phase safety data*
- Cornelia Dunger-Baldauf and Charis Papavassilis (Novartis): *For the sake of the patient – reducing placebo exposure by using historical controls*
- Gonzalo Duran-Pacheco (Roche): *Electronic Health Records used to derive Control Arms for Single-Arm oncology trials: Proof of concept using RCT's in lung cancer*
- Chris Harbron (Roche): *A Decision Making Framework For Utilising External Control Arms*
- Meinhard Kieser (University of Heidelberg): *Synthetic controls – what do we need and how far can we go? Rejoinder*
- Norbert Benda* (BfArM): *Synthetic controls – what do we need and how far can we go? Rejoinder*
- Kit Roes* (Radboud University): *Synthetic controls – what do we need and how far can we go? Rejoinder*
- Jan Müller-Berghaus* (Paul-Ehrlich-Institute): *Synthetic controls – what do we need and how far can we go? Rejoinder*
- Anja Schiel* (Norwegian Medicines Agency): *Synthetic controls – what do we need and how far can we go? Rejoinder*

Full program and slides: h*ttp://bbs.ceb-institute.org/?p=1248*

Terminology (Tom Brookland)

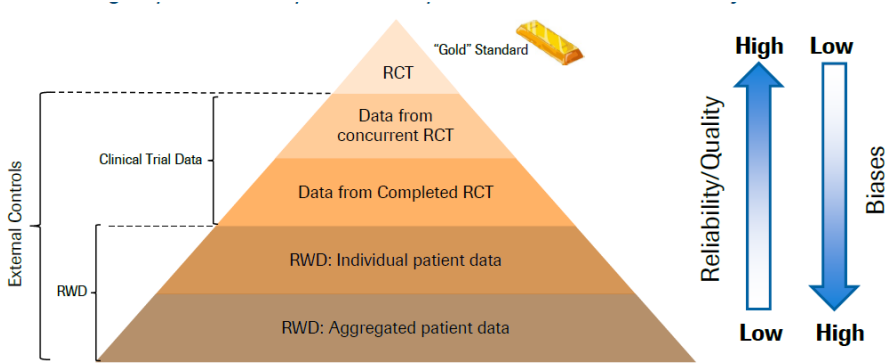
- Inconcistencies everywhere
  - definitions
  - terminology
  - use of terminology
- For example, some feedback about the seminar title
  - What are synthetic controls?
  - Why not using the term historical controls?
- Common ground
  - yes we use different terminologies
  - we (should) care about *features of the data* (Rufibach & Burger)

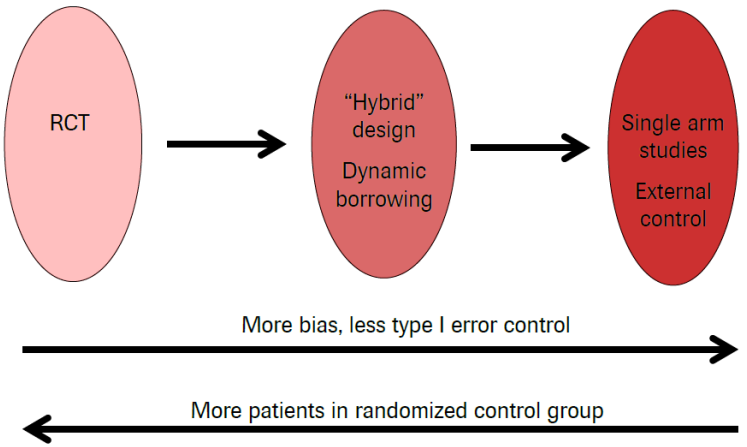
Features of the data

- Focus on **features of data**, such as:
  - randomized or not,
  - concurrecy,
  - systematically collected or not (e.g. tumor assessments),
  - robustness of endpoints,
  - relevant data available, e.g. to identify population of interest at baseline,
  - IPD vs. summary statistics only,
  - …

Rufibach & Burger

The «evidence» pyramide



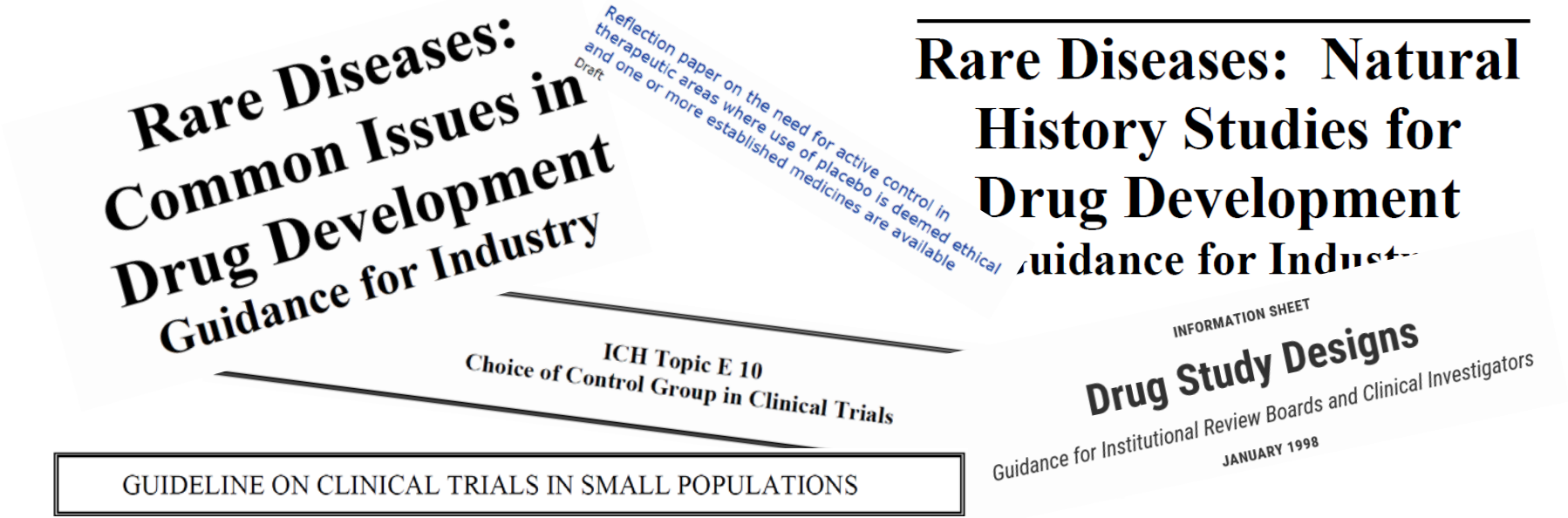


Brookland

Rufibach & Burger

When/where to use synthetic controls?

- Various documents (guidances, reflection papers, etc.) touch base on the topic



Excurse

- Guideline on Missing Data in Confirmatory Clinical Trials

> Other simple approaches for single imputation of missing data are to replace the unobserved measurements by values derived from other sources. Possible sources include information from the same subject collected before withdrawal, from other subjects with similar baseline characteristics, a predicted value from an empirically developed model or historical data. Examples of empirically

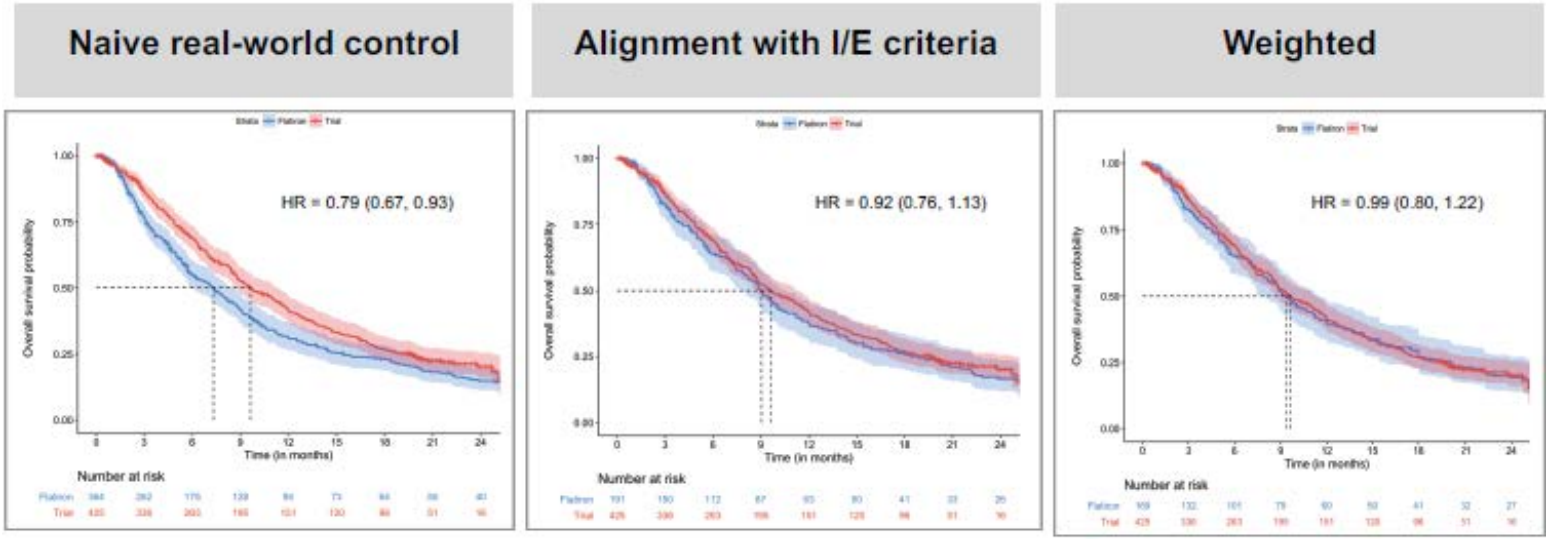
- Guideline on the investigation of subgroups in confirmatory clinical trials

> ***4.6. 'Credibility' of a subgroup finding***
>
> As indicated in 4.3, biological plausibility and the ability to find replication are key elements to evaluate the credibility of a subgroup finding. Whilst some evidence will be available when planning the trial on which factors are likely to be prognostic for patient outcome or predictive of therapeutic response the credibility of findings in a subgroup of interest needs to be re-evaluated based on the data generated in the clinical trials supporting the MAA, and other external data or knowledge that has emerged during

Examples (Sarkar)

- Focus on (semi-)structured data: EHR
- There's a lot of work to be done before these data can be used!

Fit-for-purpose RWE



NOVARTIS | Reimagining Medicine

Examples (Colin & Li)

- Safety is a major cause of failed drug approvals
  - Use existing data (pool) & compare like with like

Novartis healthy volunteer studies with placebo

N = 1775 subjects (99 studies)		n	%
Gender: Male		1471	82.9
Ethnicity: White		1203	67.8
Asian		287	16.2
Black		215	12.1
Native American		12	0.68
Other		58	3.27
	median	IQR	range
Age (years)	34	26-44	18-78
Height (cm)	175	168.5-181	143.8-199
Weight (kg)	77.9	69-86.4	47.7-116.1

Risk prediction for each subject (had they been taking placebo)

Subject	Baseline ALT (U/L)	ULN (U/L)	Number of post-baseline samples taken	Age (years)	Weight (kg)	Predicted probability ALT>ULN under placebo
1	16	55	5	22	75	1.4%
2	21	55	5	32	78	2.9%
3	28	55	5	47	70	6.2%
4	35	55	5	25	80	15.7%
5	40	55	5	52	76	14.4%
6	50	55	5	35	77	33.9%

Examples (Dunger-Baldauf & Papavassilis)

Motivation for considering the use of historical controls

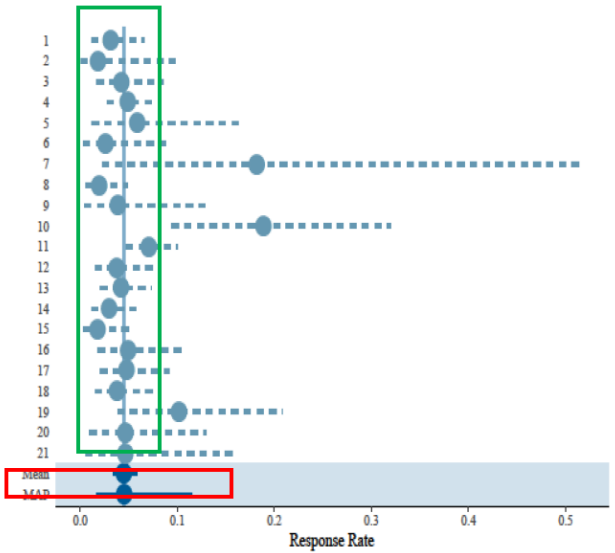
- During drug development, frequently a part of the study population is exposed to (ineffective) placebo, a burden for the patient and the Sponsor

Summary and Discussion

- More experience is needed with applications of the approach discussed here (MAP based on a meta-analysis and functional uniform priors for Bayesian dose response estimation) – planning ongoing

- **21 historical randomized controlled clinical studies in psoriasis, with 3,071 pbo subjects**
- Note that most studies show similar response rates

Prediction of response rate in new study (mean =0.05 95% Credible Interval 0.017,0.115)

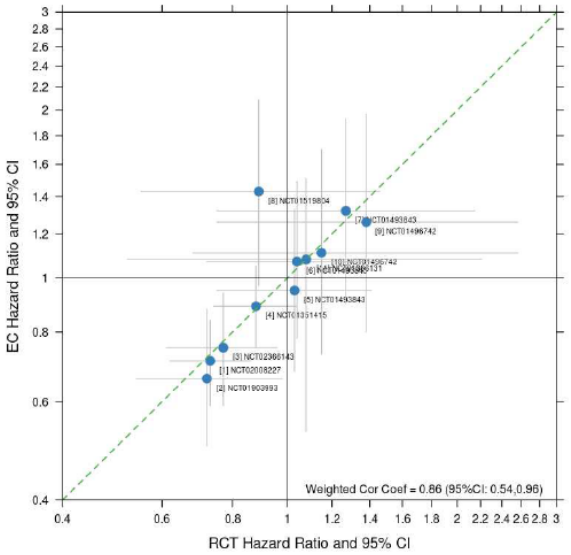


Examples (Duran-Pacheco)

Objective

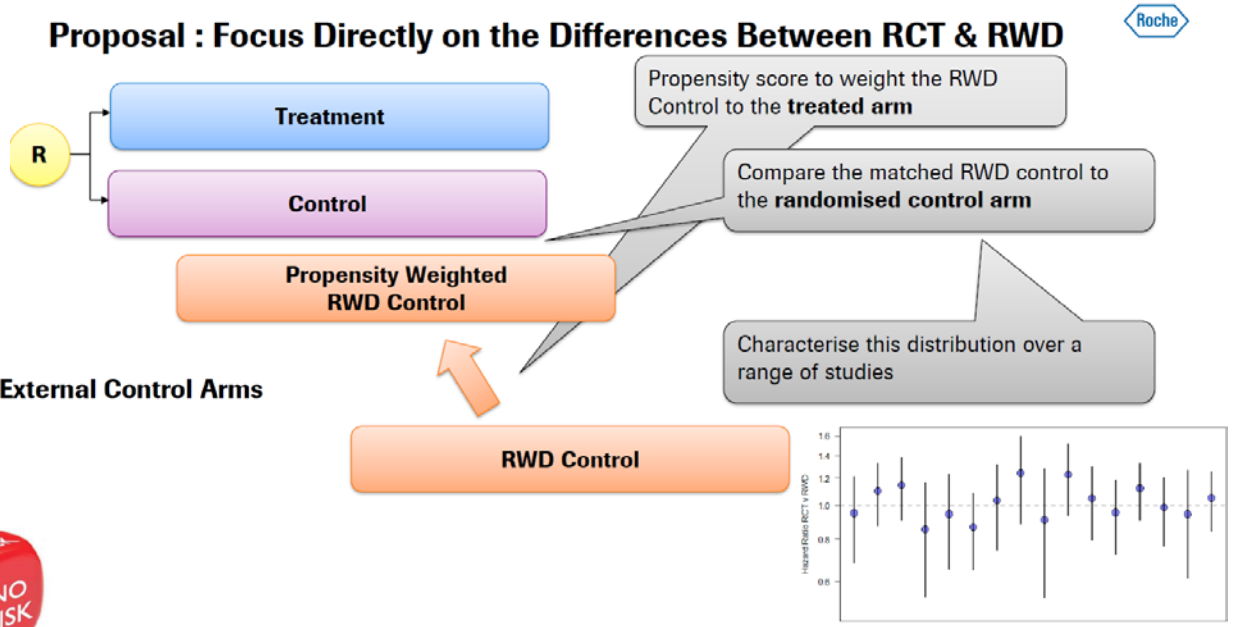
- To assess how closely results from RCTs on aNSCLC could be replicated by substituting EHR-based EC groups as the comparator

- Trials results replicated:
  - Treatment effect estimates, except for one trial
  - Conclusions from statistical tests (H0: logHR = 0)



NOVARTIS | Reimagining Medicine

Examples (Harbron)



A Decision Making Framework For Utilising External Control Arms

Basel Biometric Section – May 2019
Chris Harbron, MCO, Roche

NOVARTIS | Reimagining Medicine

Rejoinder (Kieser)

- Remain realistic
- Keep up with standards
- A call for statisticians to see it as an opportunity, also for innovative designs

Some recent headlines – facts, prophecies, fictions?

"…[the Flatiron acquisition is] ***a showcase in the future: trial designs where the currently used ones look like stagecoaches versus ICE trains*****".**

(Gross, 14.05.2018)

Standards

- don't undervalue **the merits of standards**!
- for **common clinical trials**, established standards exist for…
  - data source
  - design
  - conduct
  - analysis
  - reporting
- which **standards** are available **for clinical trials with synthetic control**, which have to be created?

Rejoinder (Benda)

Summary

- "real-world" data describe what happens but not which treatment decision is best – and so do synthetic controls
  - data only vs experiments
- use of synthetic controls
  - may be used as additional evidence to (self-standing) RCTs
  - if you know that placebo response would be 0
    - when do you know ?
    - is this the relevant endpoint ? – what about survival ?
  - if discussed in case of limited options to generate randomized data
    - small RCTs may still be possible
    - validity may still be difficult to justify
    - results to be qualified, may require stricter requirements w.r.t. the null hypothesis
  - usually based on assumptions difficult to verify
  - difficult to control type-1 error / false positive decisions

Federal Institute for Drugs and Medical Devices

Rejoinder (Roes)

When: Necessary – but not sufficient

*"In fact, **most** orphan drugs and paediatric indications submitted for regulatory approval are based on randomised controlled trials that follow generally accepted rules and guidance. Deviation from such standards is, therefore, uncommon and should only be considered when completely unavoidable and would need to be justified."*

- Rare condition
- High unmet need
- No satisfactory treatment
- *Randomisation not possible/ethical*
- No obvious control arm

- *Single arm effect "Dramatic", "Unprecedented",....*

How: Experimental design and modeling

Q -> Design -> *Data* <-> *Analysis* -> Conclusion

Design:

- The external data may be richer than "just" to distill a control group. Could we move to a more DoE approach (incl modeling) to leverage the richness?

Data <-> *Analysis in this setting:*

- Agreeing a priori on a plausible model - and sticking to it because of T1E - may lead to larger errors than making sure that statistical inference is based on a model that is adequately supported by the data.

- We need a broader approach to quantifying "error" (characteristics of the decision procedure), to include the model building step.

- This is not unique to "exceptional circumstances": estimands, new high volume data, new treatment modalities,... will require the same.

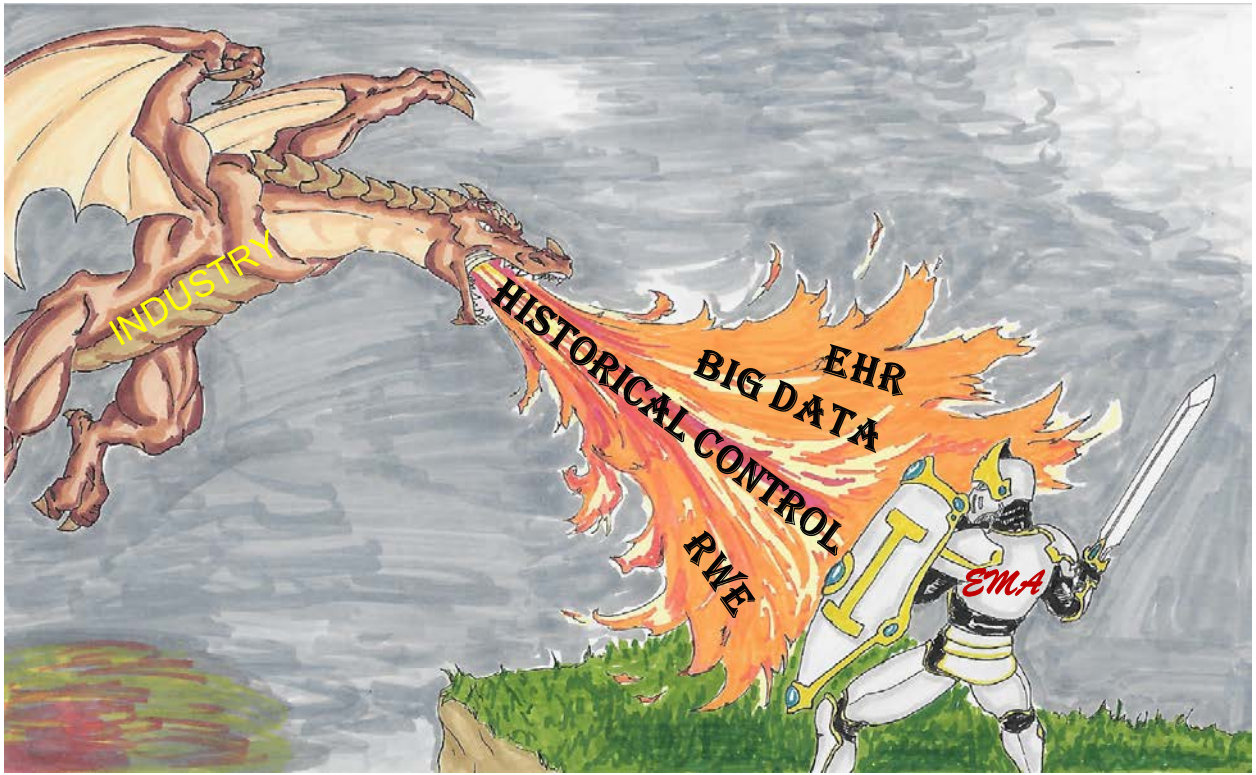
Rejoinder (Müller-Berghaus, Schiel)

- Berghaus
  - in response to Dunger-Baldauf & Papavassilis «burden to patient»
    - disease specific!
    - what might be true for dermatology does not need to be true for other diseases

- Schiel
  - nobody prevents Pharma companies from using these methods in ph I, II, IV
  - for pivotal (registration) studies: we don't want it
    - very strong preference for «classical», randomized, type-I-error controlled trial
    - at least speak to us (regulators) early when considering such approaches

In summary

In summary



Excurse (2)

Adaptive Designs for
Clinical Trials of Drugs
and Biologics
Guidance for Industry

DRAFT GUIDANCE

Special considerations apply to Type I error probability estimation when a sponsor and FDA have agreed that a trial can explicitly borrow external information via informative prior distributions. Type I error probability simulations need to assume that the prior data were generated under the null hypothesis. This is usually not a sensible assumption, as the prior data are typically being used specifically because they are not compatible with the null hypothesis. Furthermore, controlling Type I error probability at a conventional level in cases where formal borrowing is being used generally limits or completely eliminates the benefits of borrowing. It may still be useful to perform simulations in these cases, but it should be understood that estimated Type I error probabilities represent a worst-case scenario in the event that the prior data (which are typically fixed at the time of trial design) were generated under the null hypothesis. A comprehensive discussion of Bayesian approaches is beyond the scope of this

Follow-up

What happened since then?

- Not unexpected: some publications (I will name just two)
- Increased awareness of the value and importance of causal inference
  - BBS satellite seminar
  - BBS/BES jointly organized course & seminar on causal inference
- And many other things...

Publications (1)

Biom J. 2019 Jul 16. doi: 10.1002/bimj.201800250. [Epub ahead of print]

A multistate model for early decision-making in oncology.

Beyer U[1], Dejardin D[1], Meller M[1], Rufibach K[1], Burger HU[1].

Author information

1 Department of Biostatistics, MDBB 663, F. Hoffmann-La Roche Ltd., Basel, Switzerland.

Rufibach, Burger

Abstract

The development of oncology drugs progresses through multiple phases, where after each phase, a decision is made about whether to move a molecule forward. Early phase efficacy decisions are often made on the basis of single-arm studies based on a set of rules to define whether the tumor improves ("responds"), remains stable, or progresses (response evaluation criteria in solid tumors [RECIST]). These decision rules are implicitly assuming some form of surrogacy between tumor response and long-term endpoints like progression-free survival (PFS) or overall survival (OS). With the emergence of new therapies, for which the link between RECIST tumor response and long-term endpoints is either not accessible yet, or the link is weaker than with classical chemotherapies, tumor response-based rules may not be optimal. In this paper, we explore the use of a multistate model for decision-making based on single-arm early phase trials. The multistate model allows to account for more information than the simple RECIST response status, namely, the time to get to response, the duration of response, the PFS time, and time to death. We propose to base the decision on efficacy on the OS hazard ratio (HR) comparing historical control to data from the experimental treatment, with the latter predicted from a multistate model based on early phase data with limited survival follow-up. Using two case studies, we illustrate feasibility of the estimation of such an OS HR. We argue that, in the presence of limited follow-up and small sample size, and making realistic assumptions within the multistate model, the OS prediction is acceptable and may lead to better early decisions within the development of a drug.

«We propose... comparing historical control data to data from the experimental treatment,...»

Publications (2)

J Biopharm Stat. 2019 Aug 28:1-15. doi: 10.1080/10543406.2019.1657132. [Epub ahead of print]

Utilizing shared internal control arms and historical information in small-sized platform clinical trials.

Jiao F[1], Tu W[2], Jimenez S[3], Crentsil V[3], Chen YF[3].

Author information

1 Center for Devices and Radiological Health, U.S. Food and Drug Administration , Silver Spring , MD , USA.
2 Department of Statistics & Actuarial Science, University of Iowa , Iowa City , IA , USA.
3 Center for Drug Evaluation and Research, U.S. Food and Drug Administration , Silver Spring , MD , USA.

Furthermore, historical borrowing, ..., may further enhance a clinical trial's efficiency.

Abstract

Recruitment of patients in concurrent control arms can be very challenging for clinical trials for pediatric and rare diseases. Innovative approaches, such as platform trial designs, including shared internal control arm(s), can potentially reduce the needed sample size, improving the efficiency and speed of the drug development program. Furthermore, historical borrowing, which involves leveraging information from control arms in previous relevant clinical trials, may further enhance a clinical trial's efficiency. In this paper, we discuss platform trials highlighting their advantages and limitations. We then compare various strategies that borrow historical data or information, such as pooling data from different studies, analyzing data from studies separately, test-then-pool, dynamic pooling, and Bayesian hierarchical modeling, which focuses on the meta-analytic-predictive (MAP) prior. We further propose a procedure to illustrate the feasibility of utilizing historical controls under a platform setting and describe the statistical performance of our method via simulations.

Raising causal inference awareness

- 3 events on the above topic occurred in Basel after May
  - July 5: BES organized a Satellite Seminar on Causal Inference
  - Aug 19/20: course on causal inference taught by Prof. Miguel Hernán, jointly organized by BBS/BES
  - Aug 21: seminar on causal inference jointly organized by BBS/BES
- Why is this important?
  - many critics of synthetic controls around a lack of the ability to establish causality
  - better understanding methods for causal inference can help to «fill the gap»

Other things... Publications (3)

- Naci H, Davis C, Savović et al. *Design characteristics, risk of bias, and reporting of randomized controlled trials supporting approvals of cancer drugs by European Medicines Agency, 2014-16: cross sectional analysis.* BMJ 2019 (epub ahead of print)

- *"In this study, we evaluated the evidence base underpinning the EMA's recent cancer drug approvals. Between 2014 and 2016, a quarter of pivotal studies supporting cancer drug approvals were not randomised designs. Of the 39 randomized controlled trials that formed the basis of new cancer drug approvals, almost three quarters did not measure overall survival or quality of life outcomes as primary endpoints. Using the revised Cochrane tool, we judged 49% of randomised controlled trials to be at high risk of bias."*

Conclusion

Key points

- Synthetic controls
  - increasingly used in studies
  - somewhat different views on usability by industry and regulators

- No need to oversell approaches using synthetic controls
  - requires thorough, time-consuming preparations
  - do not solve all problems & pose their own challenges
  - applicability depends on disease area

- «Black or white» perspective not meaningful
  - Type I error control: not the only thing that counts ($p < 0.05$ not the only metric for decision making)
  - dialogue should continue

“The truth is always more heroic than the hype.”

- Jessica Lynch, Opening Statement Before House Oversight & Govt. Reform Committee, delivered 24 April 2007, Washington, D.C.

Thank you

NOVARTIS | Reimagining Medicine