Beyond RCTs – using historical data in pivotal clinical trials

DSBS meeting on
Extrapolation in paediatrics and use of external data in pivotal clinical trials
Copenhagen, November 14, 2018

Franz König
Medical University of Vienna
Center for Medical Statistics, Informatics and Intelligent Systems
www.meduniwien.ac.at/medstat
Franz.Koenig@meduniwien.ac.at

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013{602552 (IDEAL - http://www.ideal.rwth-aachen.de/).

MEDIZINISCHE UNIVERSITÄT WIEN

Presentation based on paper by Eichler et al. (2016) in CP & T

STATE OF THE ART

"Threshold-crossing": A Useful Way to Establish the Counterfactual in Clinical Trials?

H-G Eichler[1], B Bloechl-Daum[2], P Bauer[3], F Bretz[4], J Brown[5], LV Hampson[6], P Honig[7], M Krams[8], H Leufkens[9], R Lim[10], MM Lumpkin[11], MJ Murphy[12], F Pignatti[1], M Posch[3], S Schneeweiss[13], M Trusheim[14] and F Koenig[3]

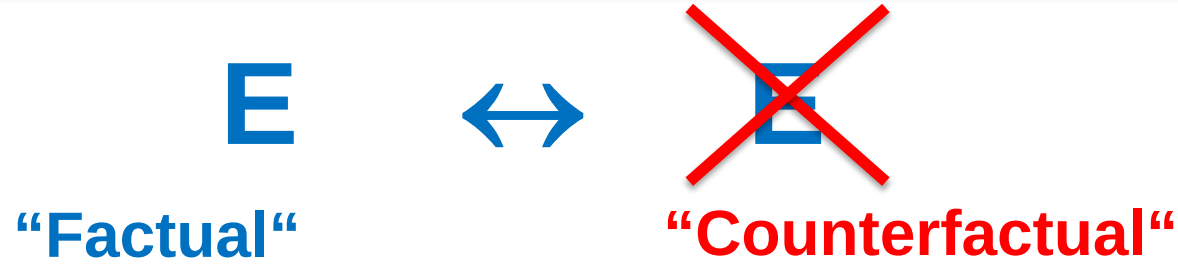
A central question in the assessment of benefit/harm of new treatments is: how does the average outcome on the new treatment (the factual) compare to the average outcome had patients received no treatment or a different treatment known to be effective (the counterfactual)? Randomized controlled trials (RCTs) are the standard for comparing the factual with the counterfactual. Recent developments necessitate and enable a new way of determining the counterfactual for some new medicines. For select situations, we propose a new framework for evidence generation, which we call "threshold-crossing." This framework leverages the wealth of information that is becoming available from completed RCTs and from real world data sources. Relying on formalized procedures, information gleaned from these data is used to estimate the counterfactual, enabling efficacy assessment of new drugs. We propose future (research) activities to enable "threshold-crossing" for carefully selected products and indications in which RCTs are not feasible.

DISCLAIMER

The views expressed in this article are the personal views of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the agencies or organizations with which the authors are affiliated.

http://dx.doi.org/10.1002/cpt.515 (open-access)

One central question when developing/prescribe a drug …



- How does the outcome of (experimental) treatment (**the factual**) compare to “what would have happened [if patients] had not received the test treatment or if they had received a different treatment known to be effective”[1] (**the counterfactual**)
- Asked by patients, clinicians treating individual patients, population-level decision-makers (including drug developers, regulators, HTA bodies, and payers of health care)

- However, the counterfactual of individual patients cannot be observed

[1] ICH E10: Choice of Control Groups in Clinical Trials

Randomized controlled trials (RCTs) …

- … became the **gold standard** for comparing the factual with the counterfactual

T – recognition that the counterfactual for individuals are not known, as opposed to average counterfactual for groups, leading to the comparison of group averages

C – average treatment effect comparing experimental with a control

R – randomisation to minimize confounding and bias at baseline

- … allow to establish *causal effects*

Which evidence is sufficient to proof efficacy of a drug?

- When is the counterfactual sufficiently clear to allow robust inferences about the causal effects of a new treatment (the factual) when an RCT is not feasible?

and/or

- How can we make the counterfactual sufficiently clear, not just for obvious *parachute cases*?

"Parachute cases": where the factual and counterfactual are sufficiently well understood and difference is likely to be sufficiently large to reasonable exclude chance or bias.

Demand for alternatives ...

Ethical concerns

- Epidemic and nonepidemic situations with high unmet need (natural history of the disease linked with high levels of morbidity and/or mortality).
- Promising results (animal studies, PK/PD, case studies, …)
  - loss of equipoise
  - ethical dilemmas for randomization
  - unwillingness of patients to participate in trials if not receiving experimental medication.

- Example: Ebola Virus Disease (EVD) epidemic demonstrated difficulty of implementing gold-standard RCTs … "*learn as much as possible, as quickly as possible, without compromising patient care …*" (WHO, 2014)

Demand for alternatives ...

Ethical concerns

The rise of one-time interventions with long-term outcomes

- New generation of theraphies (gene / cell theraphies, or tissue engineered products, …)
- Challenges:
  - some are administered only once in a lifetime,
    but effects can only be measured after prolonged periods
  - blinding not realistic - potential of high dropouts in a RCT
  - …
- Example: Holoclar, first cell-based therapy authorized in EU (see EMA EPAR)

Demand for alternatives ...

Ethical concerns

The rise of one-time interventions with long-term outcomes

Smaller treatment-eligible populations

- Growing number of drugs targeting small populations
- Precision/stratified medicine [Simon et al., 2015], molecular diagnostic profiling
- Limited number of patients available within a reasonable time frame
- Large fraction of approved drugs for rare conditions were not studied in RCT [Onakpoya et al. 2015]

Demand for alternatives ...

Ethical concerns

The rise of one-time interventions with long-term outcomes

Smaller treatment-elgible populations

Personalized treatment combinations

- Single drug interventions will not be sufficient in many pathologies
- Individual combination theraphies (based on clincial and biomarker predictors)
- Thousands of patients to be screened for reasonably powered RCT (e.g, Klauschen et al.)
- Alternative designs suggested: e.g, comparing selected patients to unselected patients
  - do not compare like-with-like;
  - what if predicitive BM are prognostic

Demand for alternatives ...

Ethical concerns

The rise of one-time interventions with long-term outcomes

Smaller treatment-elgible populations

Personalized treatment combinations

Interindividual variance: shift from noise to focus of interest

- Research question changed from

  "Is A better than B in a group of patients?"

- to

  "If A truly modulates target X, i.e. has pharmacodynamic activity,
  (how) can we identify patients who benefit from a combination that work?"

Demand for alternatives ...

Ethical concerns

The rise of one-time interventions with long-term outcomes

Smaller treatment-elgible populations

Personalized treatment combinations

Interindividual variance: shift from noise to focus of interest

What else has changed …

… six years ago (22/11/2012) at the EMA Workshop on clinical-trial data and transparency an avalanche was set off …

Guido Rasi, Excecutive Director of EMA:

"…we are not here to decide if we publish clinical-trial data, but how!"

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

24 June 2013
EMA/240810/2013
Executive Director

Publication and access to clinical-trial data

POLICY/0070
Status: Draft for public consultation
Effective date:
Review date:
Supersedes: N.A.

DRAFT

1. Introduction and purpose

The aim of the European Medicines Agency ('the Agency') is to protect and foster public health. Transparency is a key consideration for the Agency in delivering its service to patients and society.

There is growing demand from external stakeholders for full transparency, not only about the Agency's deliberations and actions, but also about the data and results from clinical trials (CTs) on which regulatory decisions are based. Following consultations with a broad range of external stakeholders and European bodies, including the European Ombudsman and the European Data Protection Supervisor, the Agency has drafted this policy, which complements the existing 'Policy on access to documents (related to medicinal products for human and veterinary use)' (POLICY/0042) (EMA/110196/2006), which came into effect in December 2010. To ensure consistency, the existing policy on access to documents and this policy on publication and access to clinical-trial data, once finalised, will be aligned.

Allowing external parties access to CT data held by the Agency will directly or indirectly affect different stakeholders' rights, interests and values. In addressing many competing objectives, the Agency takes the following views and positions, which inform the policy:

Enabling public scrutiny and secondary analysis of CTs: Access to CT data in an analysable format will benefit public health in future. It will make drug development more efficient by establishing a level playing field that allows all drug developers to learn from past successes and failures, and it will enable the wider scientific community to make use of detailed and high-quality CT data to develop new knowledge in the interest of public health. The Agency also takes the view that a high degree of transparency will take regulatory decision-making one step closer to EU citizens and patients, and promote better-informed use of medicines. Independent replication of CT data analysis is a legitimate

7 Westferry Circus • Canary Wharf • London E14 4HB • United Kingdom
Telephone +44 (0)20 7418 8400 Facsimile +44 (0)20 7418 8409
E-mail info@ema.europa.eu Website www.ema.europa.eu

Open access to Clinical Study Report (CSR): designates the entirety of elements submitted as study reports in CTD Module 5, following the format of the ICH E3 document

Controlled access to Raw CT data (meaning individual patient data sets, individual patient line-listings, individual Case Report Forms (CRFs), and documentation explaining the structure and content of data sets

Further Clinical Trial Data Transparency Initiatives

- **BMJ Open Data Campaign**
  "As of January 2013, the BMJ will no longer publish any trial of drugs or devices where the authors do not commit to making the relevant anonymised patient level data available, upon reasonable request."
- **FDA Transparency Initiative**
  Availability of Masked and De-identified Non-Summary Safety and Efficacy Data
- **All Trials Initiative**
  "All Trials Registered, All Results Reported"
- **Individual Pharmaceutical Industry Initiatives: ClinicalStudyDataRequest.com**
  GSK Data transparency initiative, Roche Global Policy on Sharing of clinical Trial Data, …
  Researchers may receive access to raw data after requests have been reviewed by an independent panel of experts
- **Yale University Open Data Access (YODA) Project**
  … a model to facilitate access to patient-level clinical research data to promote wider availability of clinical trial data and independent analysis by external investigators
- **Project Data Sphere (PDS): Sharing of comparator arm data from historic cancer clinical trials**
- **Cochrane Collaboration statement on access to clinical trial data**
  "All data from all randomised clinical trials, including raw anonymised individual participant data that do not allow identification of individual participants, and the corresponding trial protocols, to become publicly available free of charge and in easily accessible electronic formats"
- **Joint Statement of EFPIA and PHRMA**
  Principles for Responsible Clinical Trial Data Sharing
- **New EU regulation on clinical trials on medicinal products for human use**
- ….

13

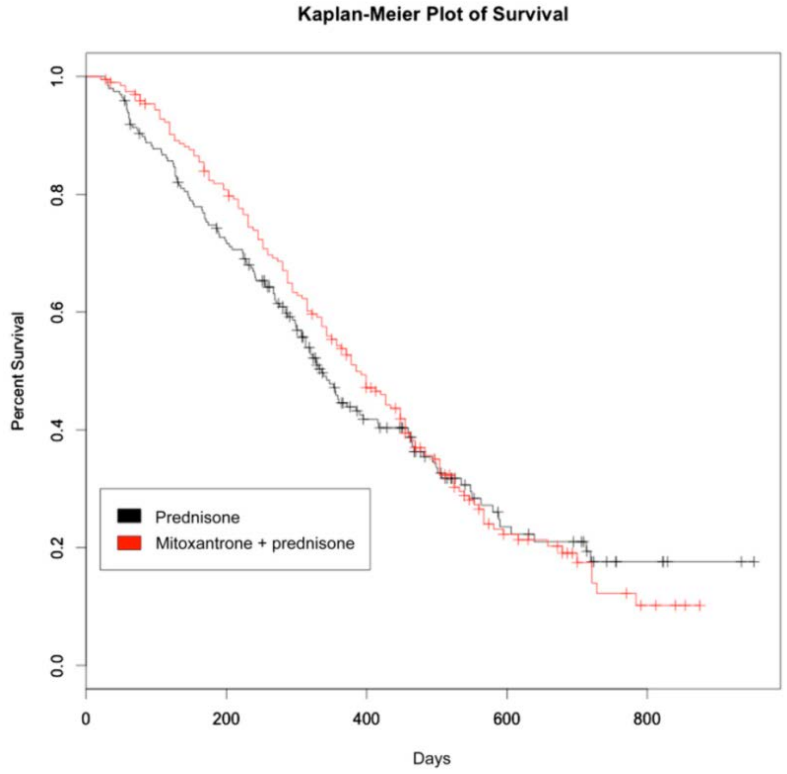
The Oncologist®

Genitourinary Cancer

Comparative Effectiveness of Mitoxantrone Plus Prednisone Versus Prednisone Alone in Metastatic Castrate-Resistant Prostate Cancer After Docetaxel Failure

Angela K. Green,[a,*] Robert W. Corty,[b,*] William A. Wood,[a] Mathew Meeneghan,[a] Katherine E. Reeder-Hayes,[a] Ethan Basch,[a] Matthew I. Milowsky,[a] Stacie B. Dusetzina[c,d]

[a]UNC Lineberger Comprehensive Cancer Center, [b]Division of Hematology and Oncology, School of Medicine, [c]Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, and [d]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA



- n=562;
- Not “real-world data”
- Not a single patient was enrolled for this study
- What is it?

Background. Mitoxantrone was approved for use in metastatic castrate-resistant prostate cancer (mCRPC) based on pain palliation without observed survival benefit in a small phase III trial in 1996. To re-evaluate for possible survival benefits in a larger contemporary sample and to demonstrate analytic uses of the newly available Project Data Sphere online resource, we used data from control arms of completed clinical trials to compare survival and toxicity among patients with postdocetaxel mCRPC treated with mitoxantrone and prednisone.

Green, et al. The Oncologist 2015;20:516-522

141,124
Patients

183
Datasets

15,893
Downloads

2,098
Users

30
Data Providers

9,547
Logins

What if we could share, integrate, and analyze our collective historical cancer research data in a single location?

Cancer researchers are working tirelessly and remarkable discoveries have been made, yet every year, 8.2 million lives are lost to cancer around the world.

Sadly, we're losing nearly the same number of people today as we were 40 years ago. With researchers working independently and with fewer resources, we're simply not finding solutions quickly enough.

The true power of this platform will come from an ever increasing volume of data and the continuing engagement of a diverse global community focused on finding solutions for cancer patients. Imagine what will happen when the entire cancer community joins efforts.

Availability of …

… of patient-level RCT data

- Offer unprecedented opportunities to learn about the counterfactual
- High quality data
- Often large (allowing precise estimates)
- Good Standardisation
- Include information on substantial number of covariates
  (i.e., provides in-depth understanding of patient population)

… real world data (RWD)

- High external validity
- Multisourced data from different healthcare environments allows to assess reproducibiltiy
- Speed and relevance
- RWD messier than RCT
- Limited data standardization
  (differently defined variables, time points for measurements, exposure and event definition, different coding systems), missing data and lack of information (e.g, patient-reported outcomes).

Data transparency initiatives will increase quality and interoperability of data

The framework of

Threshold-Crossing

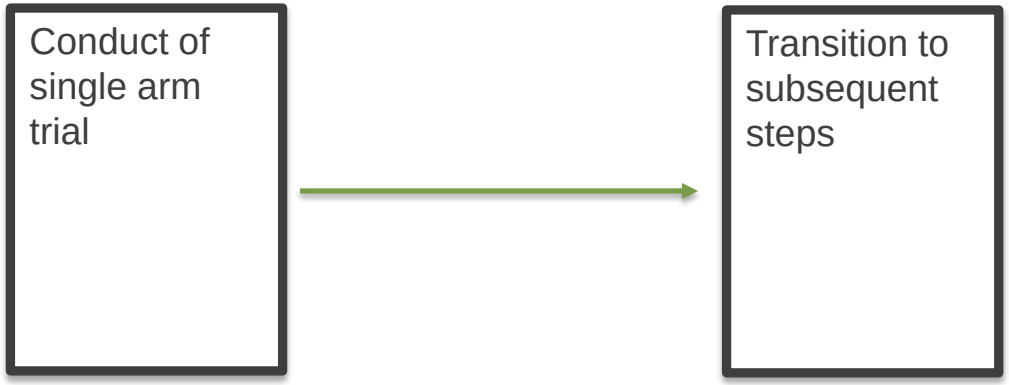
- Addresses the demands for alternatives trial designs
- Pre-specified incorporation of existing data (RCT and/or RWD)
- New trial with experimental treatment only
- Upfront pre-specification is key to avoid post-hoc cherry picking

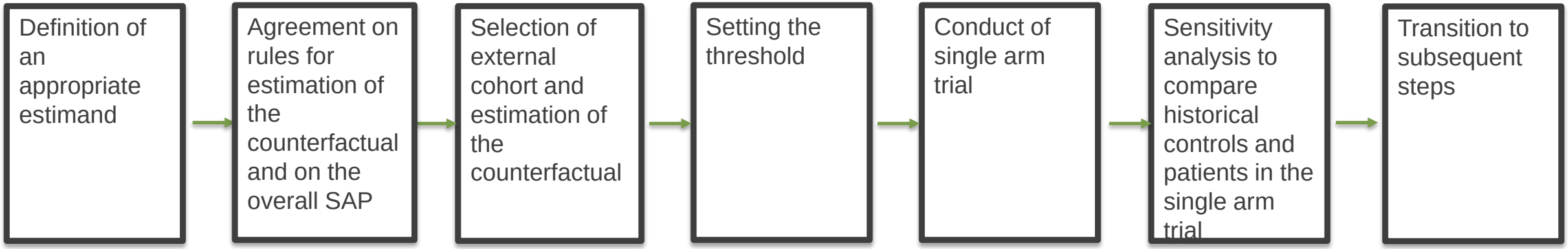
Threshold-crossing?

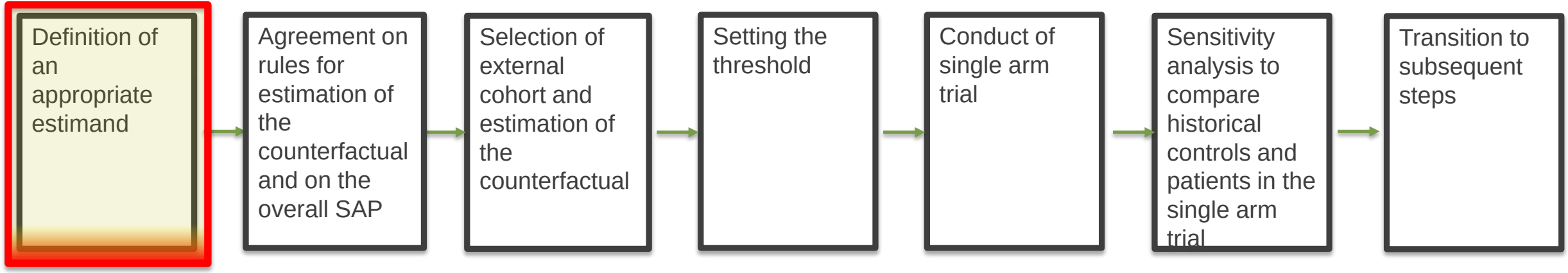
- “It may be tempting in exceptional cases to initiate an externally controlled trial, hoping for a convincingly dramatic effect, with a prompt switch to randomized trials if this does not materialize”

  [ICH E10 guideline]

- Can we operationalise the concept?

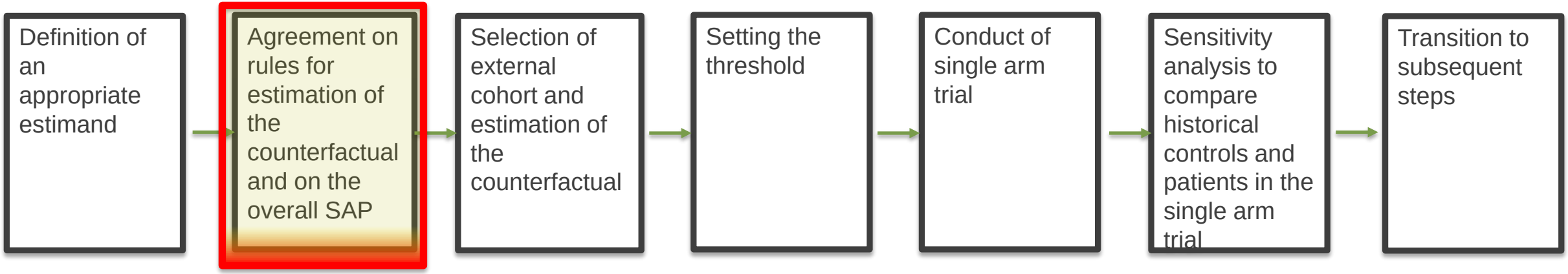

Conduct of single arm trial
Transition to subsequent steps


Definition of an appropriate estimand
Agreement on rules for estimation of the counterfactual and on the overall SAP
Selection of external cohort and estimation of the counterfactual
Setting the threshold
Conduct of single arm trial
Sensitivity analysis to compare historical controls and patients in the single arm trial
Transition to subsequent steps

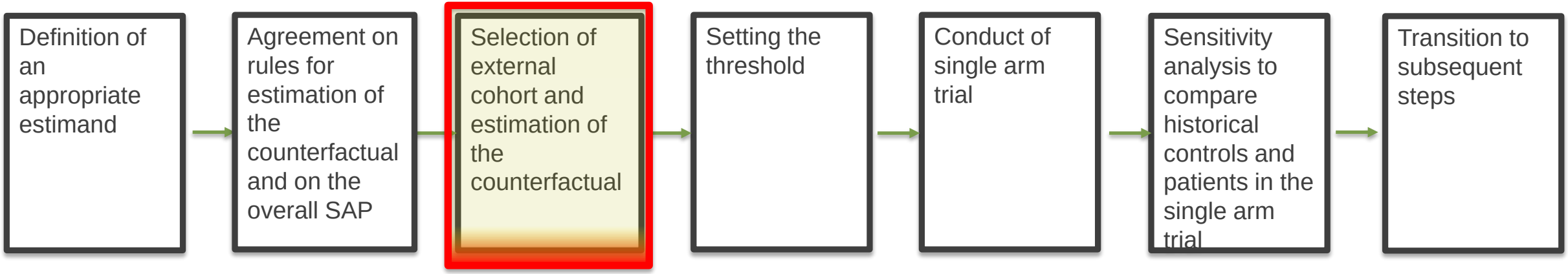


- Precise definition of the estimand (what needs to be estimated to address scientific question)
  - Including treatment-eligible population
    - described by phenotypic and genotypic criteria
    - precise selection criteria to allow for unequivocal definition of historical controls and contemporary intervention cohort
  - Variable(s) of interest (what, when and how it is measured)
  - The measure for intervention effect (quantifying the treatment benefit in terms of the variable(s) of interest)

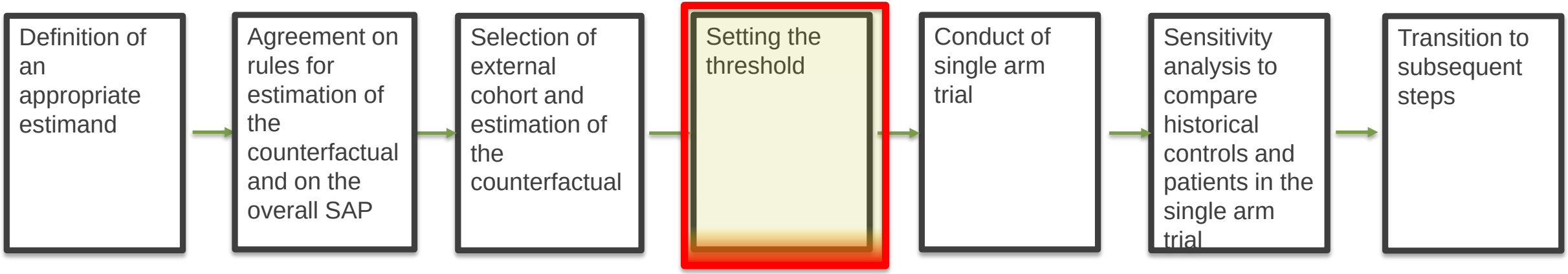
- See ICH E9 guidance, forthcoming addendum



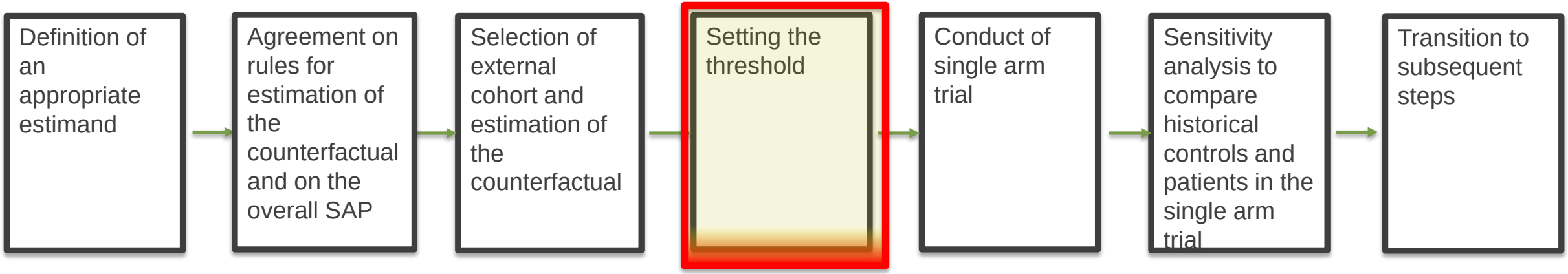
- Rules for estimation of counterfactual for the chosen estimand have to be established before selecting historical cohort
- Develop Statistical Analysis Plan (SAP) including sensitivity analyses



- Based on selection criteria (step 1), select one ore more suitable control cohorts from RWD, RCT or combination of both
  - Normally, patients in the control cohorts will have received standard of care, best supportive care, etc.
- Bias: How to avoid risk of cherry picking of a favourable historical control (e.g, selection of controls where the outcome/effect of comparator treatment is artificially poor).
  - Historical controls identified from systematic, transparent, and reproducible review of existing evidence following a pre-specified protocol
  - If possible, more than one control cohort from different sources
  - Controls identified before patients are enrolled in the prospective, single arm trial
- After establishing the control cohort, estimate the counterfactual by quantifying the historical/external information (according to step 2)

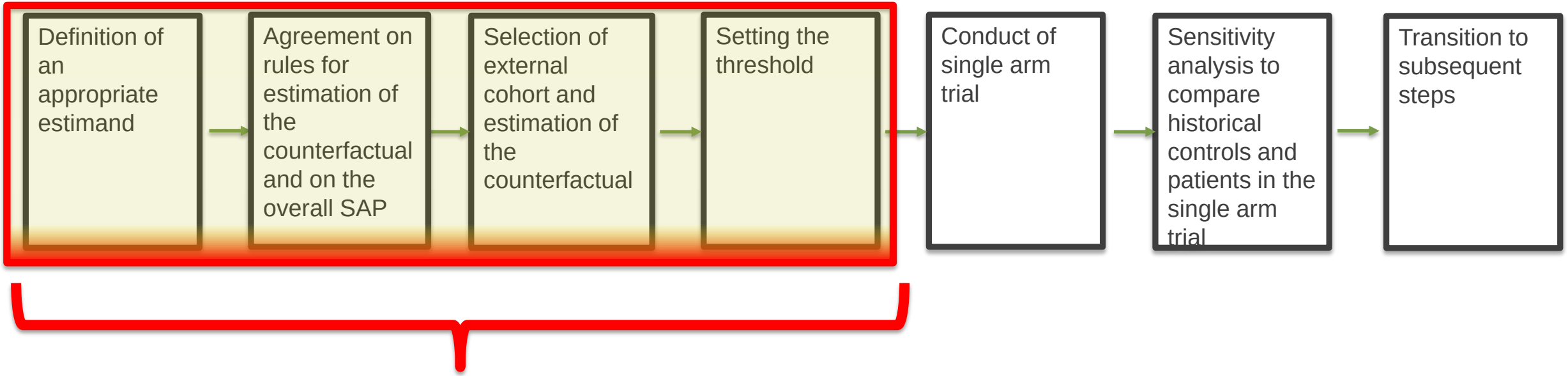


- Set efficacy threshold based on historical data
  - Serves as benchmark for primary analysis
  - Needs to be pre-specified to avoid cherry-picking
  - New data (e.g. from the ongoing trial) can be used for sensitivity analyses
  - Sponsors may wish to additionally define a futility threshold
- Setting the threshold high or low?
  - Large distance between estimate of counterfactual and threshold (high hurdle): small risk of false-positive, but high risk of false-negative conclusion
  - Small distance (low hurdle): vice-versa

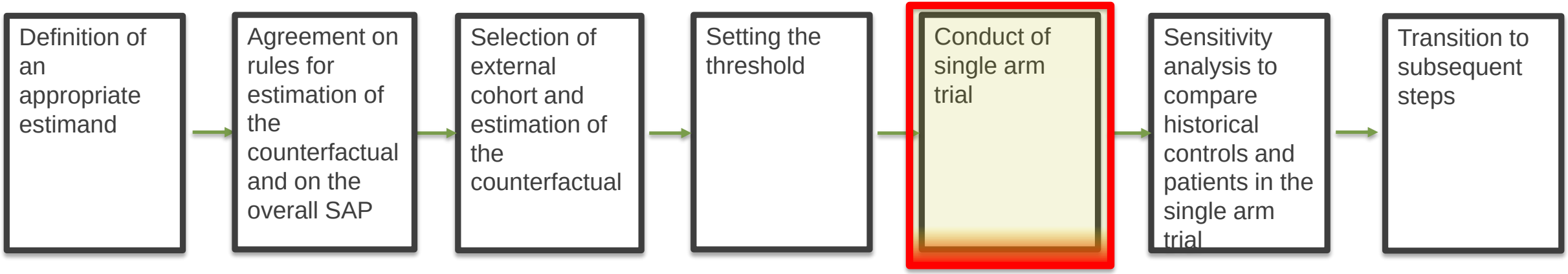


Threshold should be determined by ...

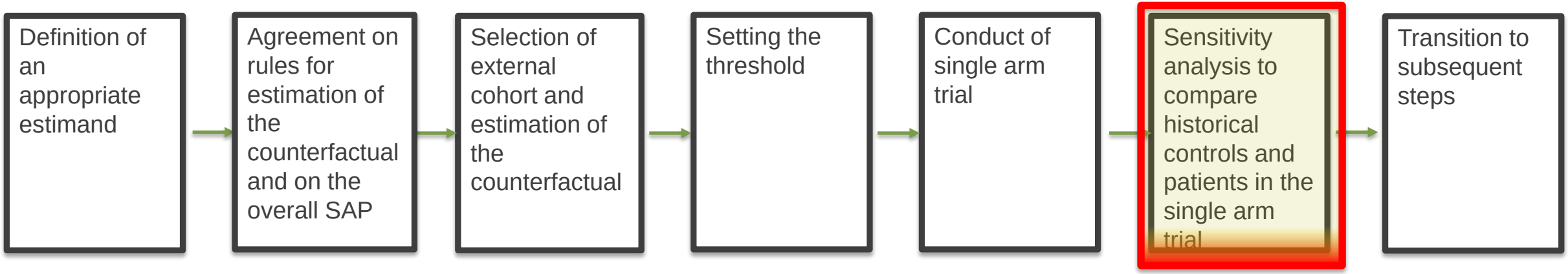
- **Methodological considerations**
  - Accuracy and precision of counterfactual
  - Quality and completeness of data-set(s)
  - Total number of patients
  - Number of sources
  - Degree of agreement between different sources
- **Ethical considerations**
  - Severity of disease
  - Unmet need of target population
  - Availability of alternative treatments
  - Patients' input on what is clinically relevant
  - Social burden of disease
  - Expected frequency of serious adverse events



Should be agreed with regulators and other relevant decision makers

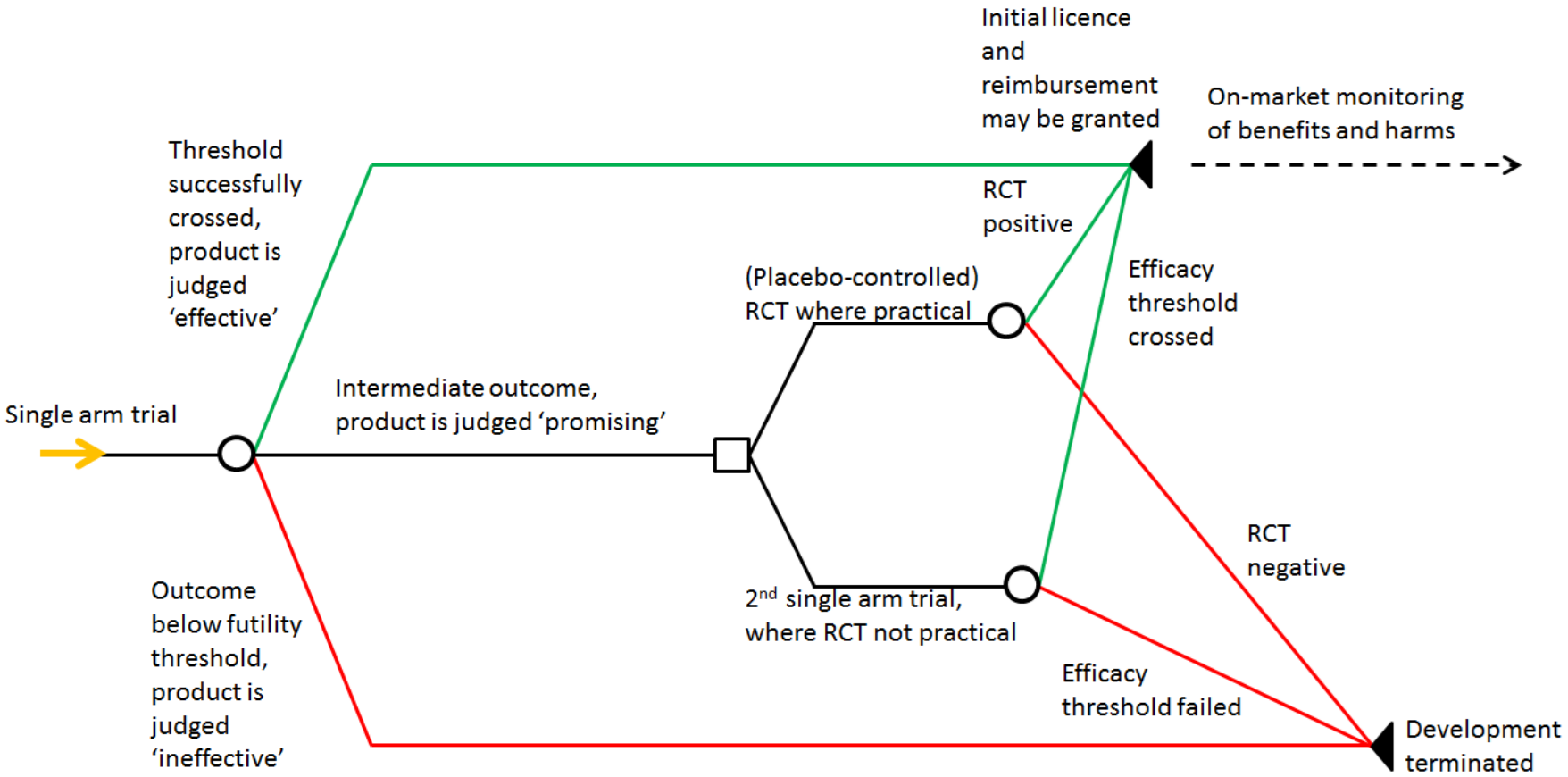


- Interventional phase
- Single-arm trial where all patients receive experimental treatment
- Trial participants (experimental group) have to be selected according to same criteria as historical control group(s)
- Same caveats apply as for any other single-arm trial
  - Several sources of bias (no concealed allocation)
  - Blinding assessors to endpoint
  - …



- Compare historical controls and patients from the single-arm trial via pre-defined threshold
- Conduct further sensitivity analyses
  - Comparability of patient populations
  - Sensitivity analyses to verify the robustness of conclusions
  - Methods of causal inference to control for confounding (e.g, multivariabel regression model adjusting for confounding, weighting or stratifying analyses by propensity scores derived from high dimensional covariate analysis, …)
  - Acknowledge impact of (untestable) assumptions on the validity of the final results as well as the impact of unknown and/or unmeasured confounders

Definition of an appropriate estimand → Agreement on rules for estimation of the counterfactual and on the overall SAP → Selection of external cohort and estimation of the counterfactual → Setting the threshold → Conduct of single arm trial → Sensitivity analysis to compare historical controls and patients in the single arm trial → Transition to subsequent steps



Comparable to mutistage approach developed by Cooper et al. (2015) for Ebola treatments

Thresholding can learn from …

- Single arm trials (e.g., oncology)
- Non-Inferiority trials (and specification of NI margin)
- Meta-Analysis (methods and guidelines for conducting and reporting systematic reviews)
- Adaptive designs (sequential conduct of stages)
- Statistical methods beyond RCTs (causal inference, Bayesian methods, …)

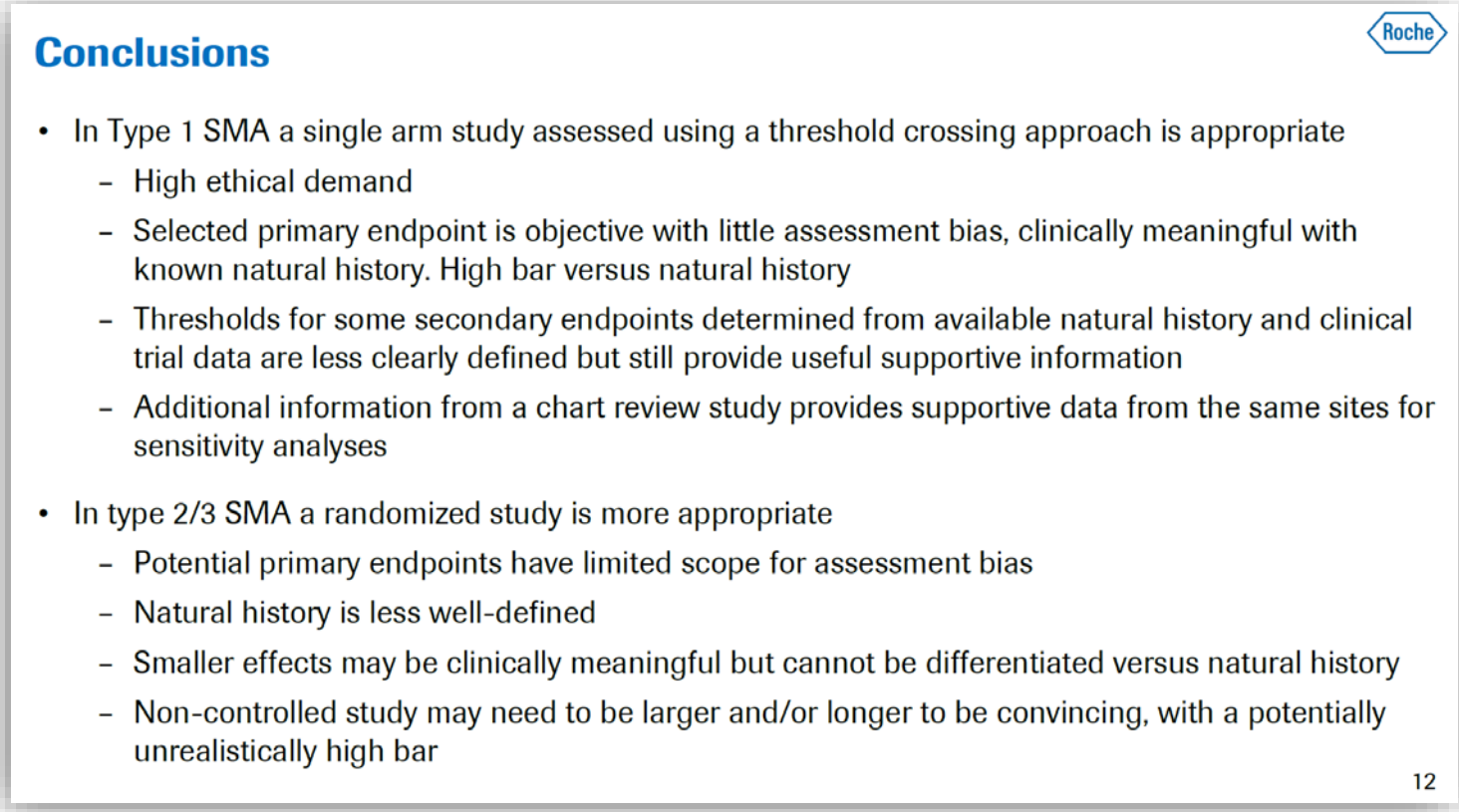
Remarks

- Applicable where RCTs are not feasible or ethical
- Full transparency of all steps
  (as opposed to an uncontrolled study and "hope for the best")
- Reuse of existing data makes drug development faster and economical
- Bias in favour of products that are either highly effective or (near-)ineffective
- Opportunity to steer pharmaceutical research and development to areas of greater unmet need
- Note the focus on an effciacy threshold; in practice, the approach will have top be implemented with a view to demonstrate an acceptable benefit-risk profile

- Methodological risks: No randomisation and blinding – increased risk of bias
- Threshold determined for primary efficacy endpoint, what about safety?
- Expectation risks: Setting (un)realistic thresholds?

Case study in Spinal Muscular Atrophy (SMA)

- Slide from Carol Reid and Uli Burger. “When a Threshold Crossing approach may and may not be appropriate: A Case Study in SMA”. EFSPI Regulatory Statistics Workshop, 24-25th September 2018



Conclusions

Roche

- In Type 1 SMA a single arm study assessed using a threshold crossing approach is appropriate
  - High ethical demand
  - Selected primary endpoint is objective with little assessment bias, clinically meaningful with known natural history. High bar versus natural history
  - Thresholds for some secondary endpoints determined from available natural history and clinical trial data are less clearly defined but still provide useful supportive information
  - Additional information from a chart review study provides supportive data from the same sites for sensitivity analyses
- In type 2/3 SMA a randomized study is more appropriate
  - Potential primary endpoints have limited scope for assessment bias
  - Natural history is less well-defined
  - Smaller effects may be clinically meaningful but cannot be differentiated versus natural history
  - Non-controlled study may need to be larger and/or longer to be convincing, with a potentially unrealistically high bar

12

See https://www.efspi.org/EFSPI/Events/Regulatory_Meetings/3rd_efspi_workshop_on_regulatory_statistics.aspx?hkey=4e080028-7086-44f2-a892-473190ef7324

The evolution of “non-RCT evidence”

- We now have resources that were not available to the RCT pioneers in the mid-20th century: Rich data on past and current patients from RW and RCTs
- We are now starting to develop methodologies and skill sets to make use of these resources – to overcome the stigma of “non-randomization”?
- Evidence can be based on a diverse family of data sources and methodologies complementing (not: replacing) RCTs.

Counterfactuals are everywhere …

- Socrates was once asked by a young man whether he should get married …

- Socartes´ reply

Do as you wish, you will likely regret, no matter what you choose.

MANY THANKS FOR YOUR PATIENCE

Backup Slides - Simulations

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design

RCT:

- H_0: $\mu_N \leq \mu_C$ vs H_1: $\mu_N > \mu_C$
  at α=2.5% (one-sided)
  with two-sample t-test

Threshold:

- H_0^t: $\mu_N \leq t$ vs H_1^t: $\mu_N > t$
  using one-sample t-test at α=2.5%

  t …a-priori fixed threshold from historical controls.

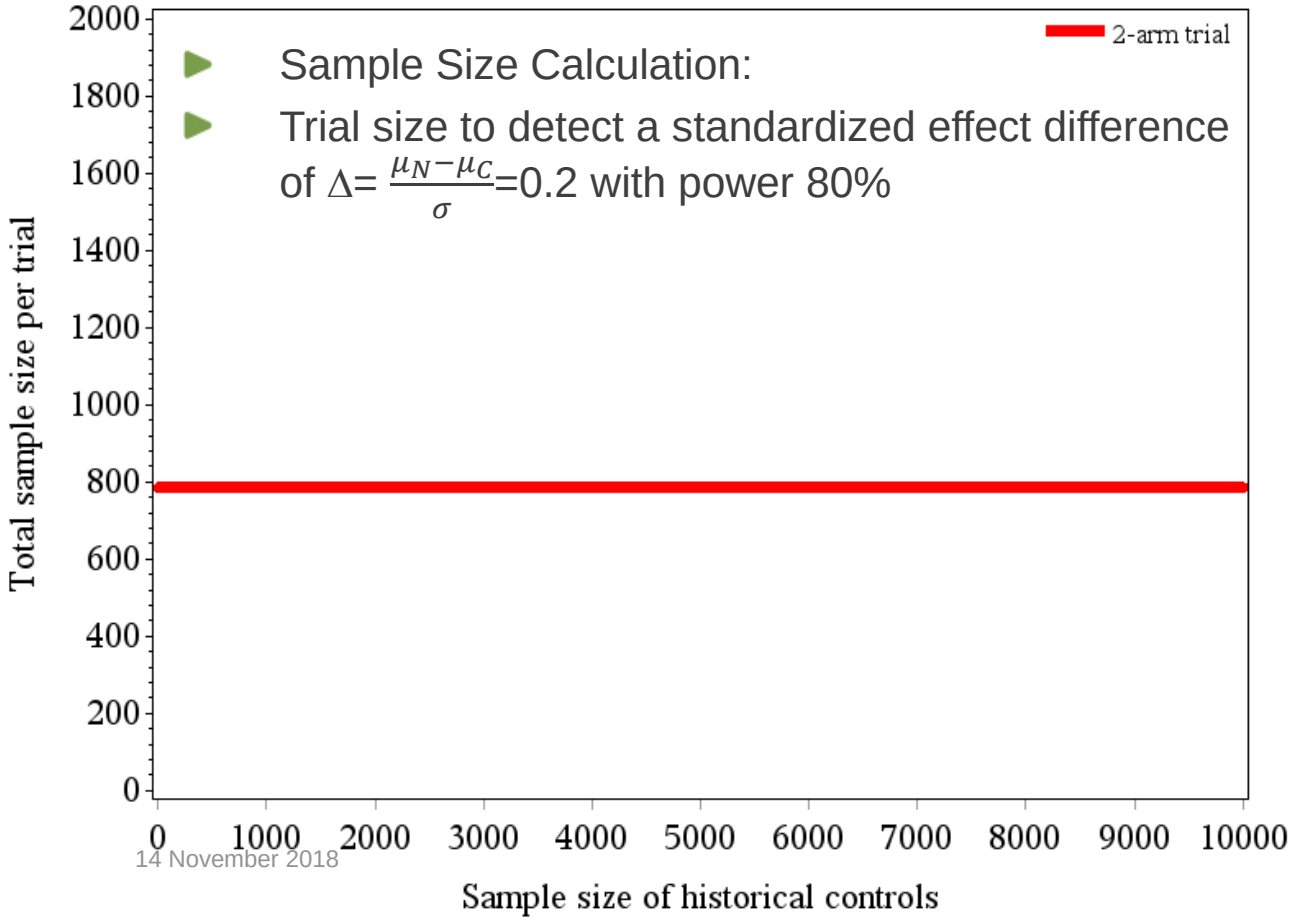
- Can we take a rejection of H_0^t: $\mu_N \leq t$ naively as a rejection for H_0: $\mu_N \leq \mu_C$?

Comparison of sample sizes, power, risk of false positives and impact of historical data base size

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design

RCT:

- H_0: $\mu_N \leq \mu_C$ vs H_1: $\mu_N > \mu_C$
  at α=2.5% (one-sided)

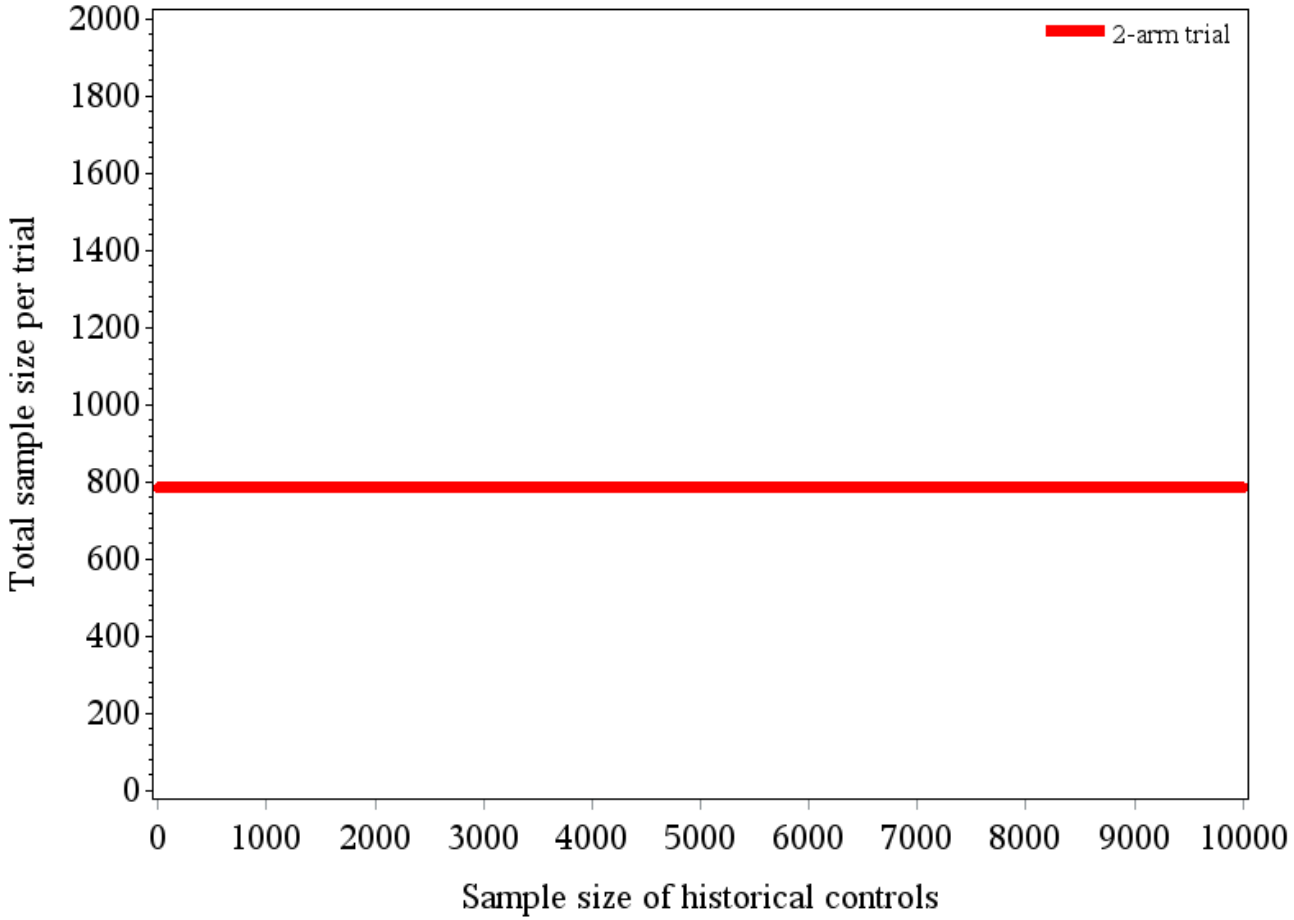


- Sample Size Calculation:
- Trial size to detect a standardized effect difference of $\Delta = \frac{\mu_N - \mu_C}{\sigma}$=0.2 with power 80%

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design

RCT:

- H_0: $\mu_N \leq \mu_C$ vs H_1: $\mu_N > \mu_C$



Threshold:

- H_0^t: $\mu_N \leq t$ vs H_1^t: $\mu_N > t$

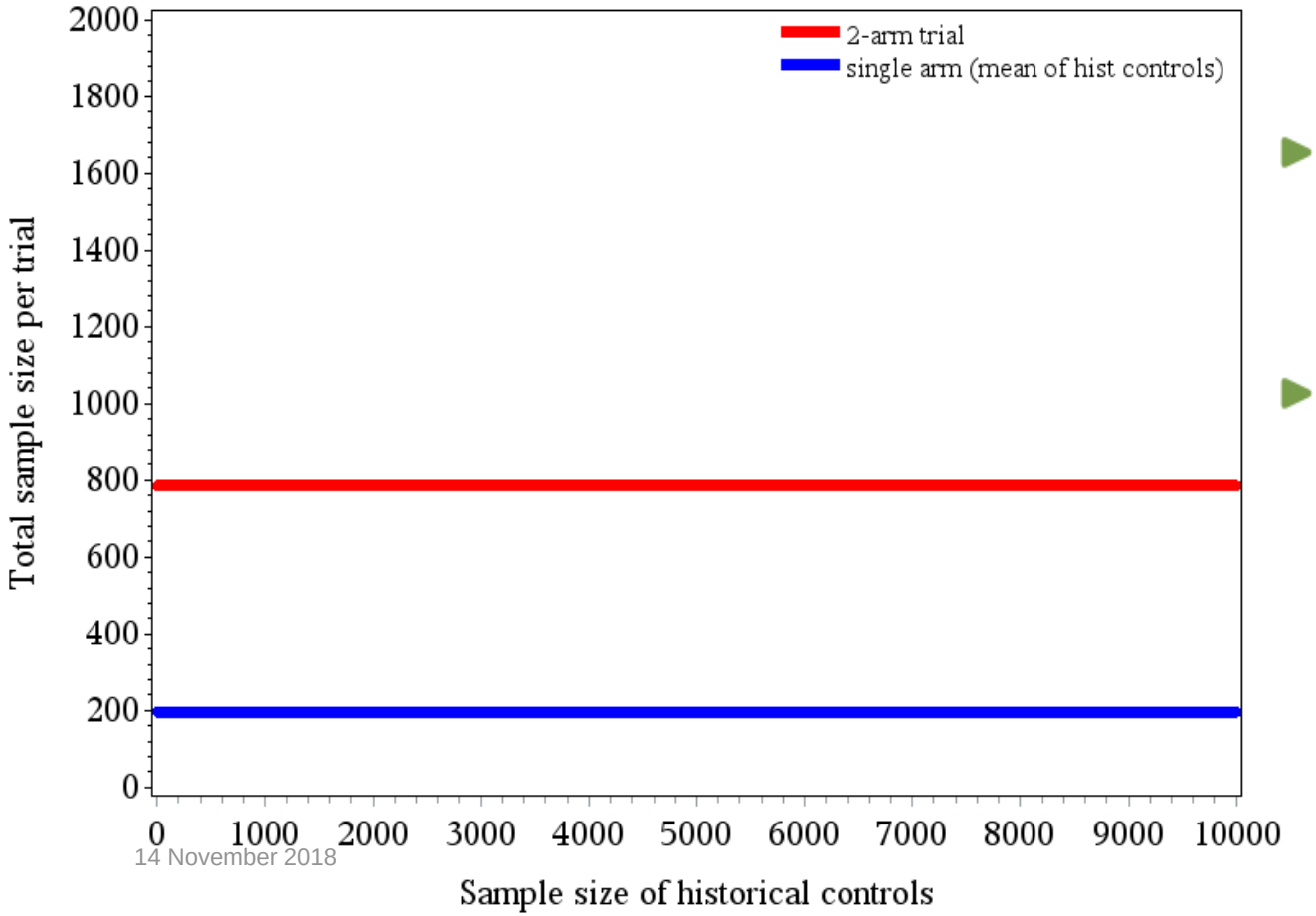
  t …a-priori fixed threshold from historical controls.

- Apply sample size calculation for single arm
  - What if t is observed mean from historical control used directly as threshold t?
  - More cautious strategies?

- Impact on error rates,
  taking rejection of H_0^t: $\mu_N \leq t$ naively as a rejection for H_0: $\mu_N \leq \mu_C$?

- What is the impact of data base size of historical controls

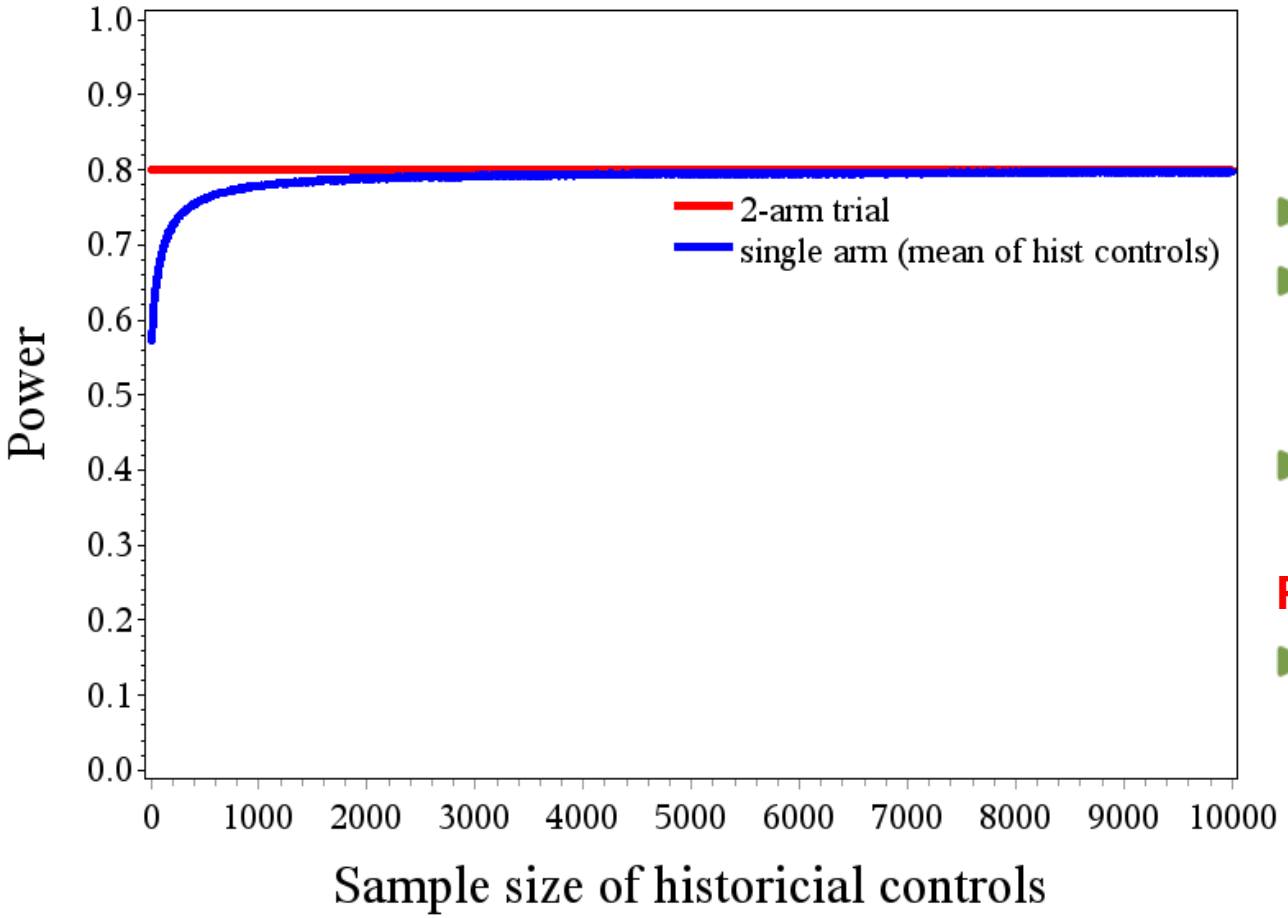
Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design



Threshold Design (**blue line**)

- Best-case scenario: (with no uncertainty on effect size of control, no shift, …) sample size can be reduced to a quarter relative to a parallel RCT
- Due to sampling variability, the observed mean in controls typically does not coincide with true population mean μ_C (even assuming μ_C would be identical for historical and concurrent controls)
- Impact on power and type I error rate?

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design



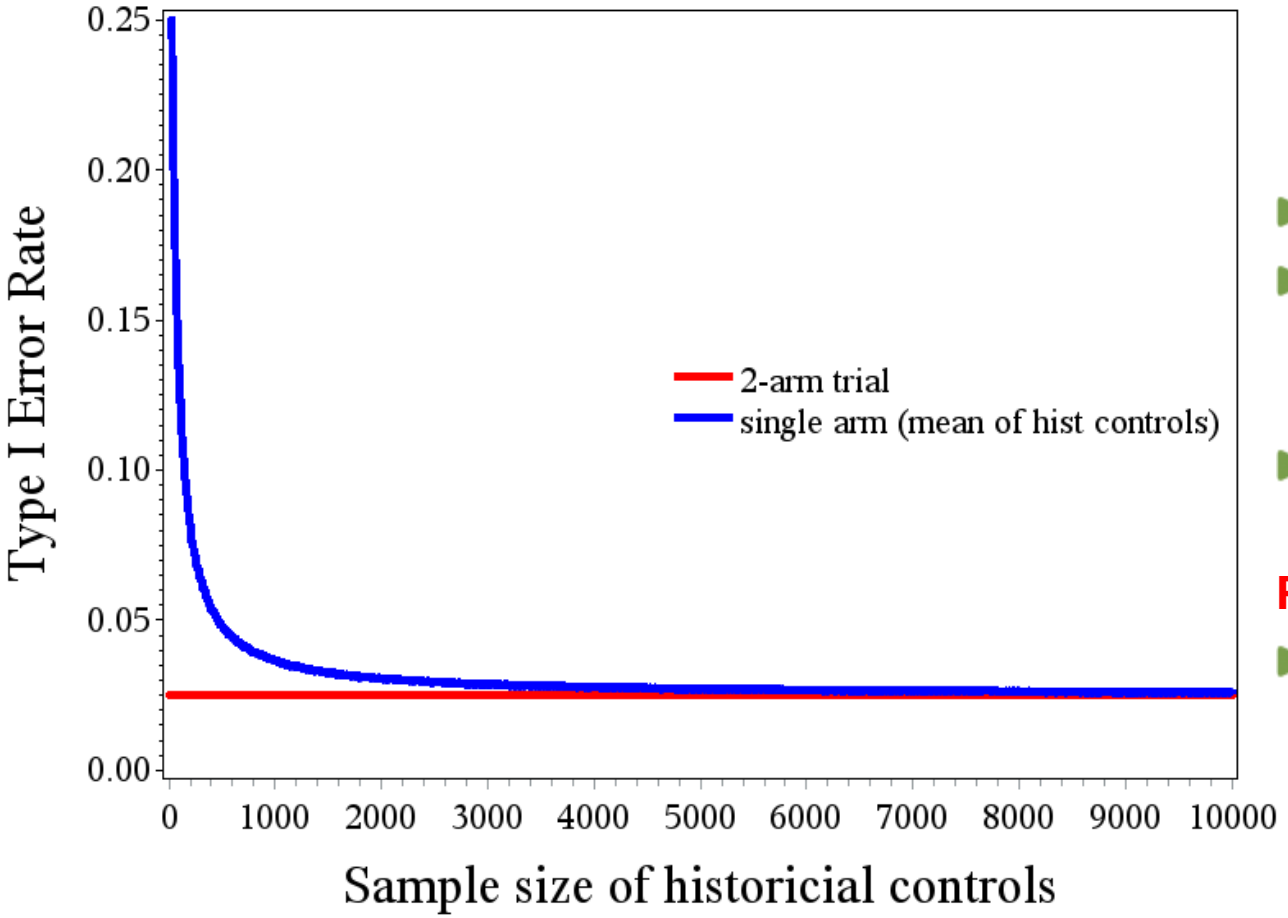
Threshold Design (**blue line**)

- Due to sampling variability, the observed mean in controls typically does not coincide with true population mean μ_C (even assuming μ_C would be identical for historical and concurrent controls)
- Impact on power and type I error rate?
- Power decreases with decreasing sample size in the historical controls due to increasing variability of the historical estimate
- The type I error rate substantially inflated for small sample sizes of historical controls (blue line)

RCT (red line)

- Both type I error rate and power do not depend on the historical data

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design



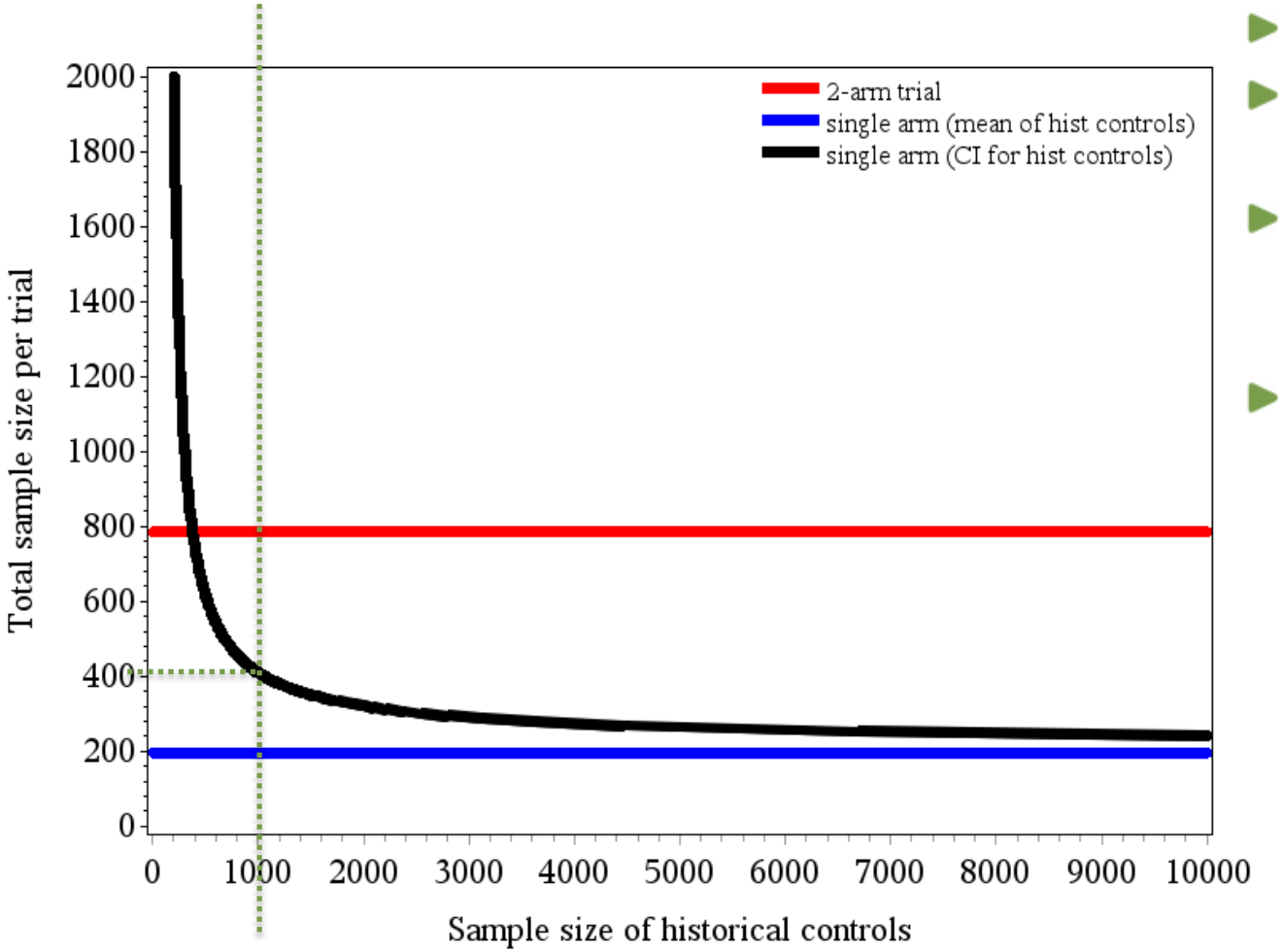
Threshold Design (**blue line**)

- Due to sampling variability, the observed mean in controls typically does not coincide with true population mean μ_C (even assuming μ_C would be identical for historical and concurrent controls)
- Impact on power and type I error rate?
- Power decreases with decreasing sample size in the historical controls due to increasing variability of the historical estimate
- The type I error rate substantially inflated for small sample sizes of historical controls (blue line)

RCT (red line)

- both type I error rate and power do not depend on the historical data

Example: Demonstrating the efficacy of a new drug (Continuous Endpoint) RCT vs threshold single arm design



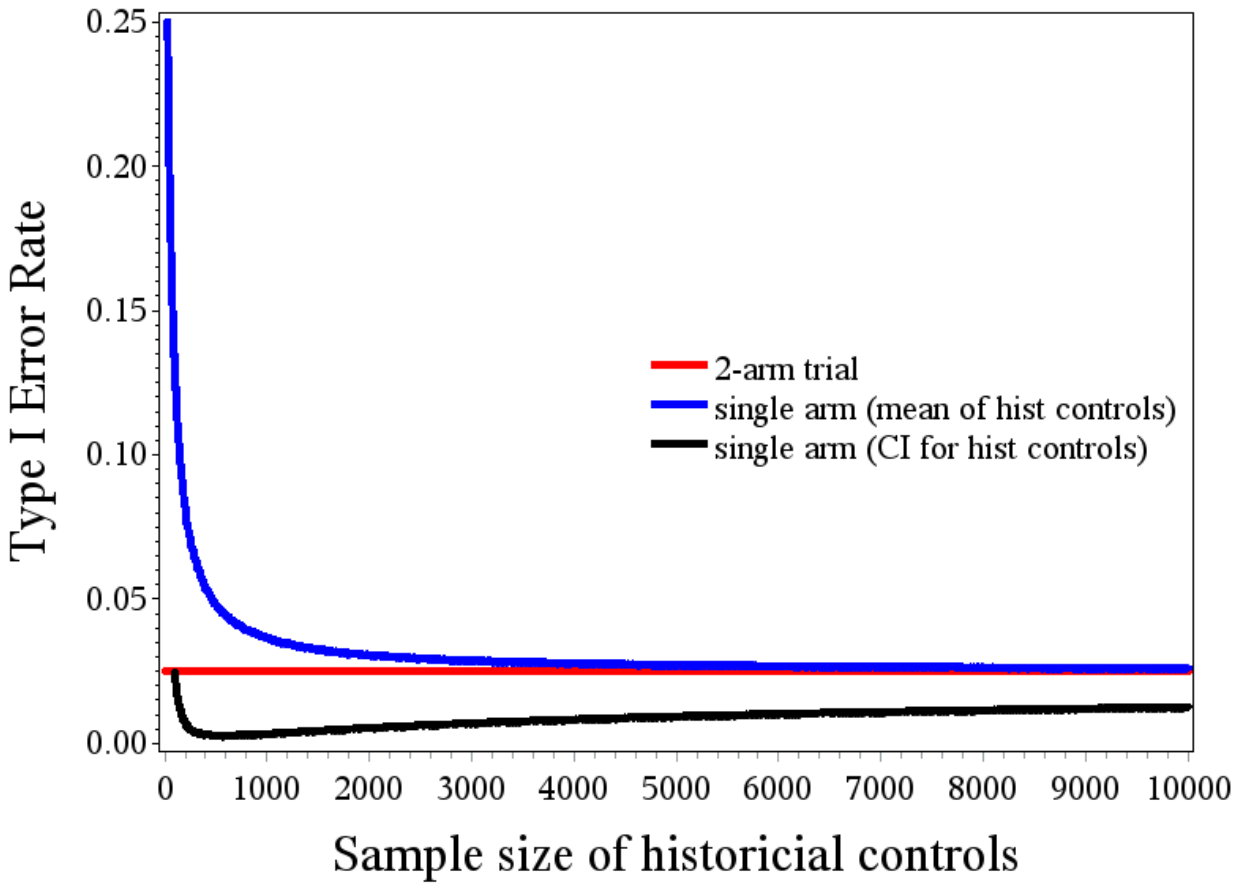
Threshold Design (**blue line**)

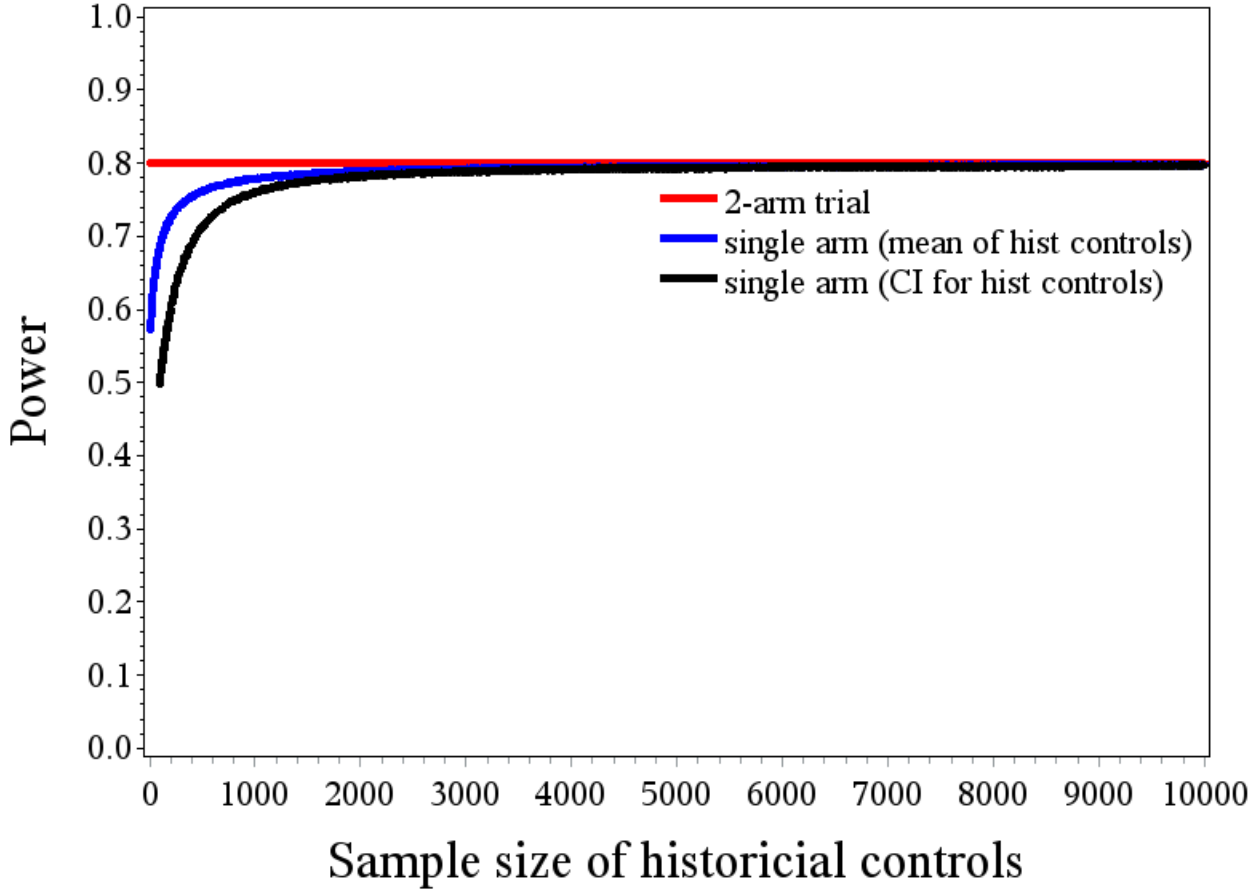
- More cautious choice of the threshold
- E.g, take upper bound of 95%-confidence interval for μ_C computed from historical controls
- Results in sample size of about 400 (=half of that for the parallel group design), if about 1000 historical controls were available.
- The more historical data available, the lower N for the threshold-crossing trial

Impact on power and type I error rate?

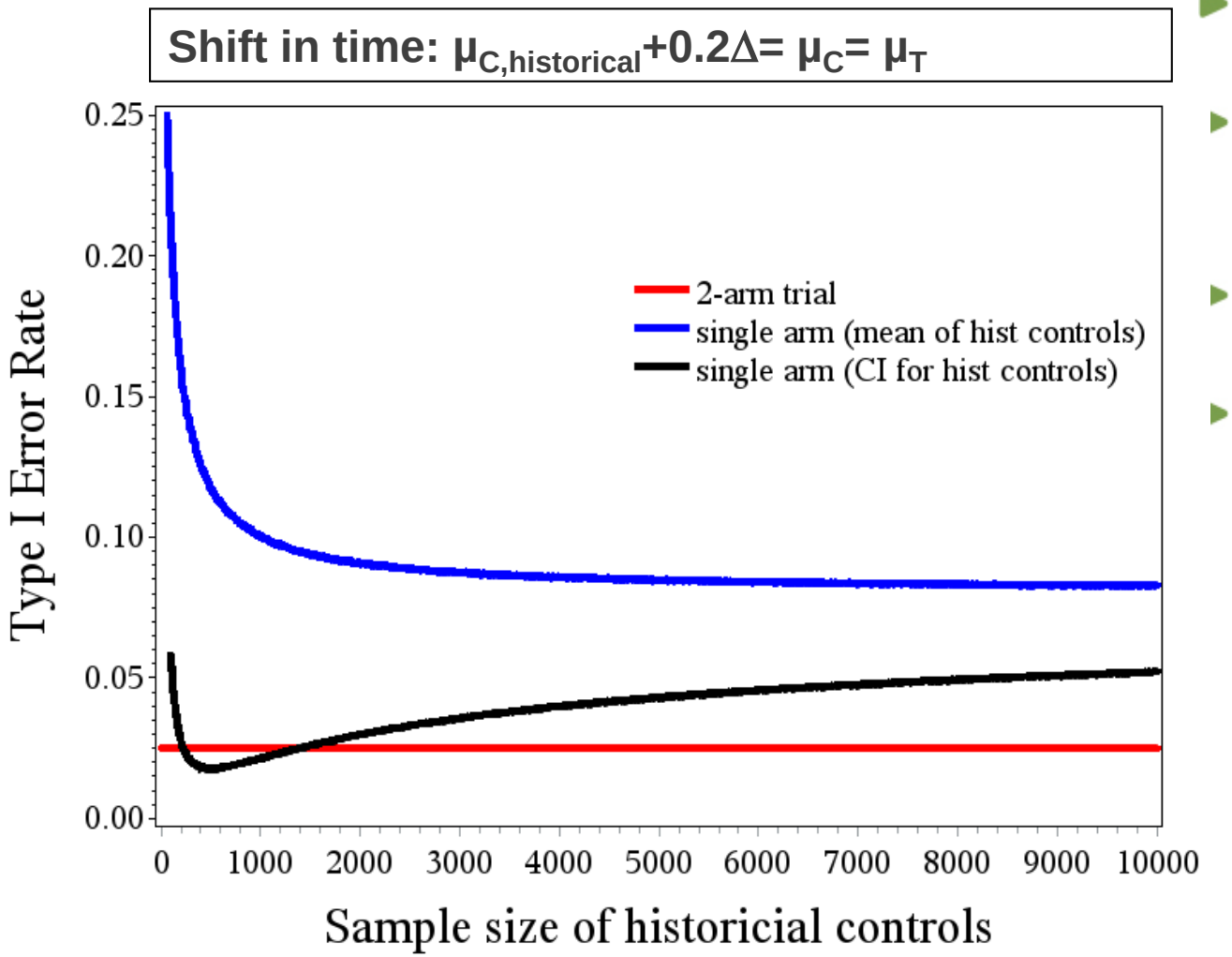
Assuming μ_C is identical for historical and concurrent controls,
the type I error rate is controlled (back line),

however a further loss of power is observed if the historical control data base is small



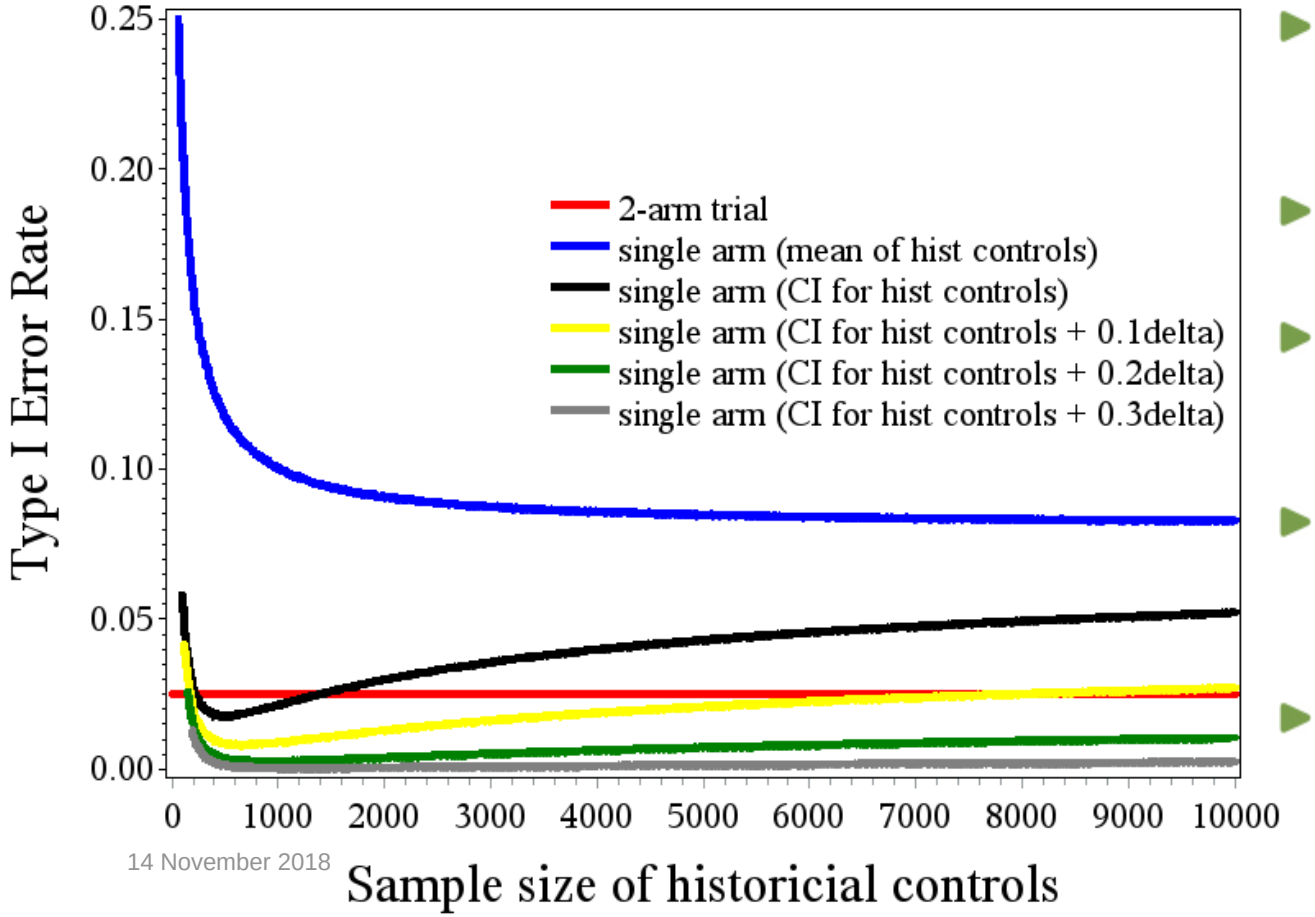


What if there is a shift, e.g. over time?

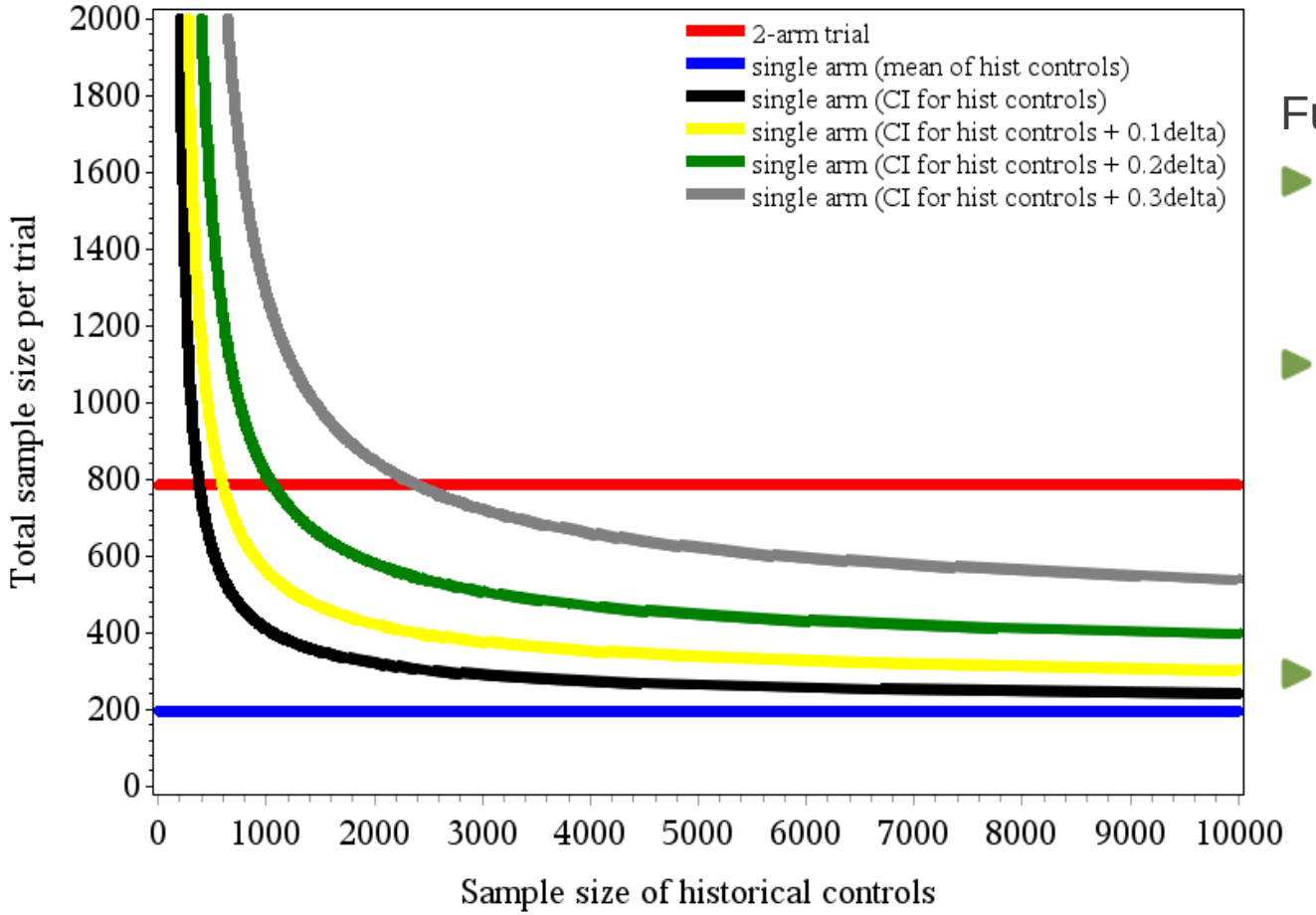


- What if μ_C is NOT identical for historical and concurrent controls,
- E.g., the mean response under control treatment is increasing over time
  →INFALTION OF TYPE I ERROR RATE
- To address such biases, apply more conservative (larger) thresholds t,
- Add a "buffer" to the upper bound of historical 95% confidence interval (e.g., 0.1Δ, 0.2Δ, and 0.3Δ for yellow, green and grey lines)

Shift in time: $\mu_{C,historical}+0.2\Delta=\mu_C=\mu_T$



- What if μ_C is NOT identical for historical and concurrent controls,
- E.g., the mean response under control treatment is increasing over time
  →INFALTION OF TYPE I ERROR RATE
- To address such biases, apply more conservative (larger) thresholds t,
- Add a "buffer" to the upper bound of historical 95% confidence interval (e.g., 0.1Δ, 0.2Δ, and 0.3Δ for yellow, green and grey lines)
- If "buffer" is sufficiently large, type I error inflation can be avoid
- However, …



However, …

- larger thresholds, larger sample sizes

Furthermore

- For simplicity we assumed that all historical controls come from one data source, e.g., a single clinical trial or a registry
- For several sources: account for between trial variability (e.g., meta-analytic estimate of μ_C obtained from a fixed or random effects meta-analysis of historical controls)
- … will increase required sample sizes further

JUMP

Backup

Borrowing strength from hybrid designs

- Existing patient-level RCT data: augment information on the control arm (i.e. the counterfactual) of an RCT
- Could allow for more efficient allocation of trial resources to the test treatment, fewer patients need to be randomised to the control group.
- Used by companies to incorporate historical data into phase II studies to inform internal go/no go decisions but not in pivotal trials (?)
- Hybrid designs may gain more traction, as data from past clinical trials are shared more widely

Rosmalen et al. 2017, Neuenschander et al., Viele et al., Gsteiger et al. 2013, Hobbs et al. 2013, Schmidli et al, 2014, …

Criteria for comparability

E.g., Pocock´s criteria for comparability of historical and concurrent controls

1. Such a group must have received a precisely defined standard treatment which must be the same as the treatment for the randomized controls.
2. The group must have been part of a recent clinical study which contained the same requirements for patient eligibility.
3. The methods of treatment evaluation must be the same.
4. The distributions of important patient characteristics in the group should be comparable with those in the new trial.
5. The previous study must have been performed in the same organization with largely the same clinical investigators.
6. There must be no other indications leading one to expect differing results between the randomized and historical controls.

Pocock SJ. The combination of randomized and historical controls in clinical trials. Journal of Chronic Diseases. 1976; 29:175–188.

Demand for alternatives ...

- Ethical concerns
  - Epidemic and nonepidemic situations with high unmet need (e.g. Ebola)
- The rise of one-time interventions with long-term outcomes
  - New generation of therapies (gene / cell therapies, tissue engineered products) that may be administered only once in a lifetime, but effects can only be measured after prolonged periods (e.g. Holoclar)
- Smaller treatment-elgible populations
  - Growing number of drugs targeting small populations (e.g. rare diseases)
- Personalized treatment combinations
  - Single drug interventions may not suffice in many pathologies and individual combination therapies (based on clinical and biomarker predictors) may be needed
- Interindividual variance: Shift from noise to focus of interest
  - Research question changed from "*Is A better than B in a group of patients?*" to "*If A truly modulates target X, how can we identify patients who benefit from A, rather than B?*"