



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

Franz König (Medical U Vienna)
Frank Bretz (Novartis)
Martin Posch (Medical U Vienna)



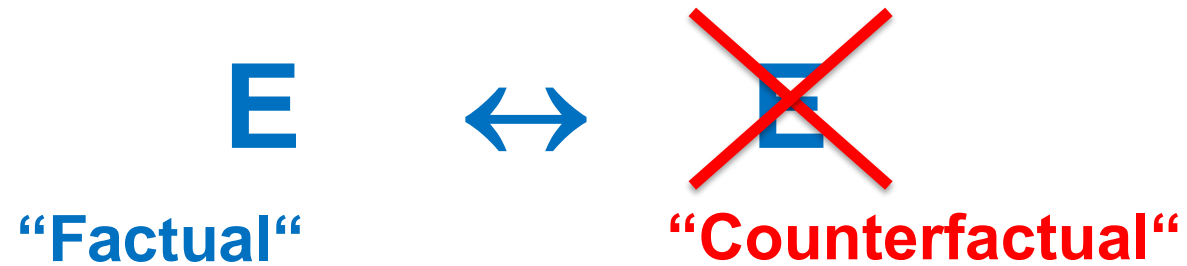
This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013{602552 (IDEAL - <http://www.ideal.rwth-aachen.de/>).

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

H-G Eichler¹, B Bloechl-Daum², P Bauer³, F Bretz⁴, J Brown⁵, LV Hampson⁶, P Honig⁷, M Krams⁸, H Leufkens⁹, R Lim¹⁰, MM Lumpkin¹¹, MJ Murphy¹², F Pignatti¹, M Posch³, S Schneeweiss¹³, M Trusheim¹⁴ and F Koenig³

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.

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*

Demand for alternatives ...

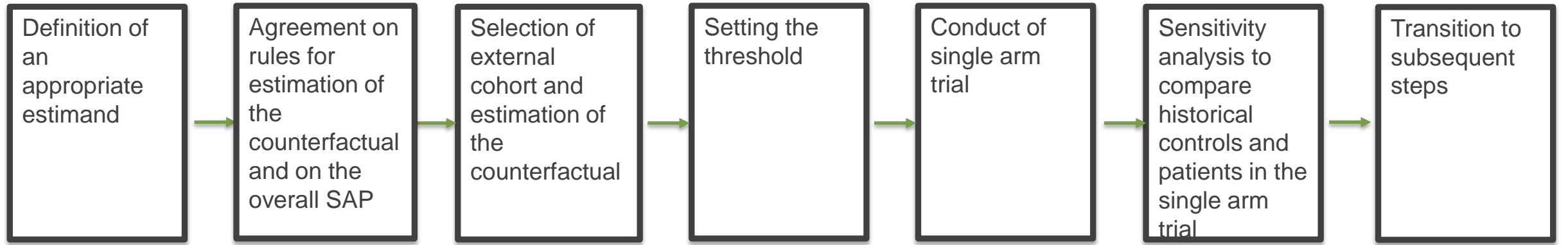
- ▶ Ethical concerns
 - Epidemic and nonepidemic situations with high unmet need (e.g. Ebola)
- ▶ The rise of one-time interventions with long-term outcomes
 - New generation of therapies (gene / cell therapies, tissue engineered products) that may be administered only once in a lifetime, but effects can only be measured after prolonged periods (e.g. Holoclar)
- ▶ Smaller treatment-eligible 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?*”

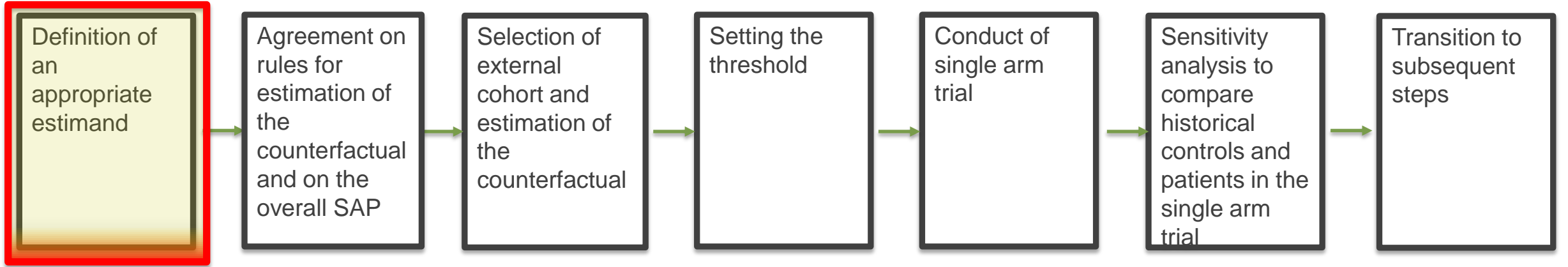
A new framework?

- ▶ “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?

Threshold-Crossing

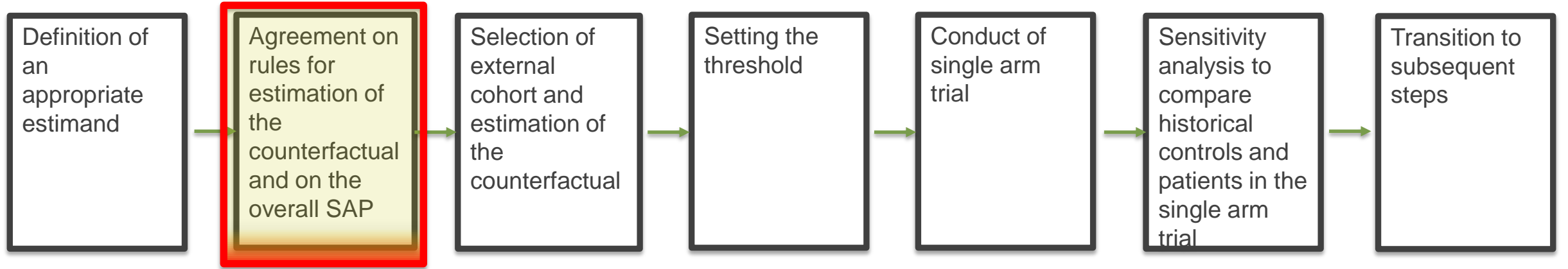
- ▶ Addresses the demands for alternatives trial designs
- ▶ Pre-specified incorporation of existing data (RCT and/or RWD)
- ▶ New trial with experimental treatment only
 - If threshold is crossed, the product is deemed effective and – in the absence of prohibitive risks – may be granted initial license/reimbursement
 - Otherwise, either the product development terminates or a conventional RCT is started
- ▶ Upfront pre-specification is key to avoid post-hoc cherry picking



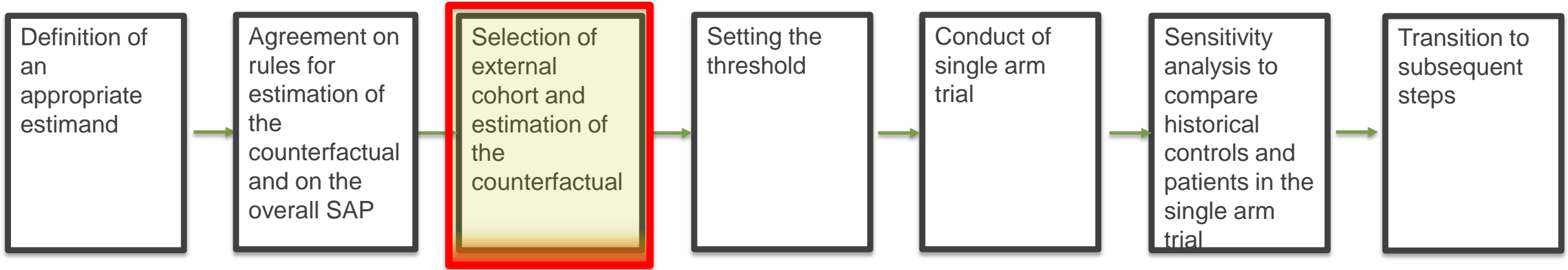


- ▶ Precise definition of the estimand (what needs to be estimated to address scientific question)
 - Including treatment-eligible population
 - Variable(s) of interest (what, when and how it is measured)
 - The measure for intervention effect (quantifying the treatment benefit in terms of the variable(s) of interest)

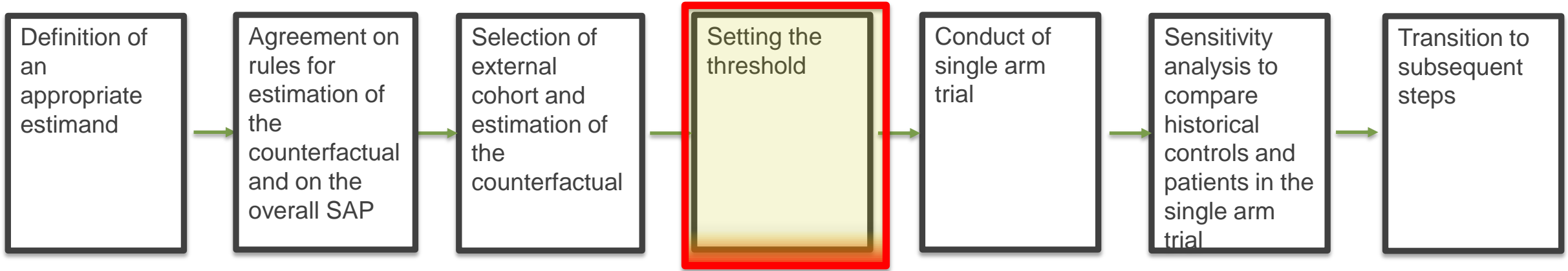
- ▶ See ICH E9 guidance, forthcoming addendum



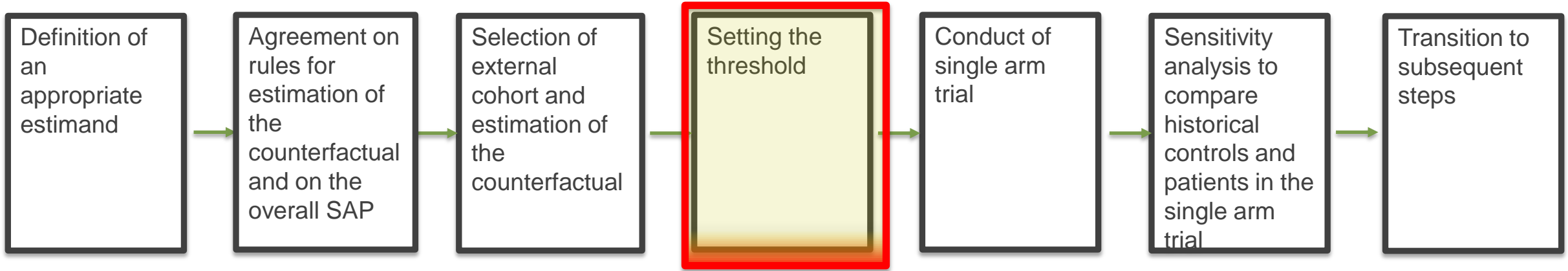
- ▶ Rules for estimation targeting the chosen estimand have to be established before selecting historical cohort
- ▶ Similar to developing a Statistical Analysis Plan (SAP), including sensitivity analyses



- ▶ Based on selection criteria (step 1), select one or 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



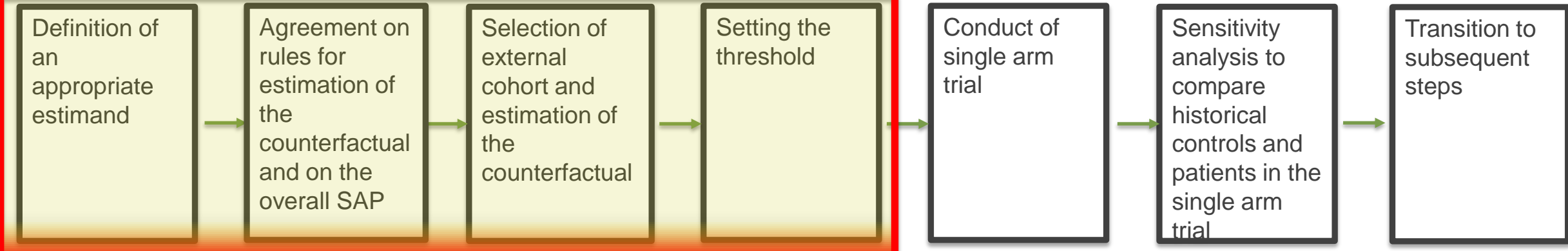
Threshold should be determined by ...

▶ **Methodological considerations**

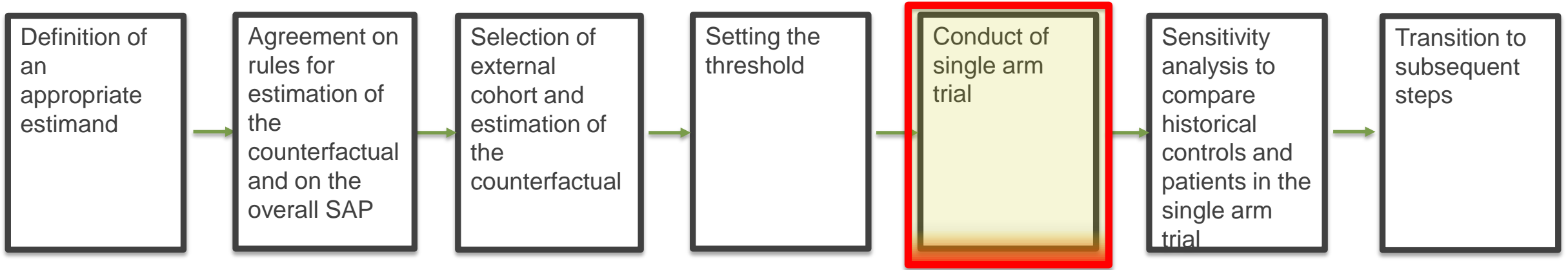
- Accuracy and precision of counterfactual
- Quality and completeness of data-set(s)
- Total number of patients
- Number of sources
- Degree of agreement between different sources

▶ **Ethical considerations**

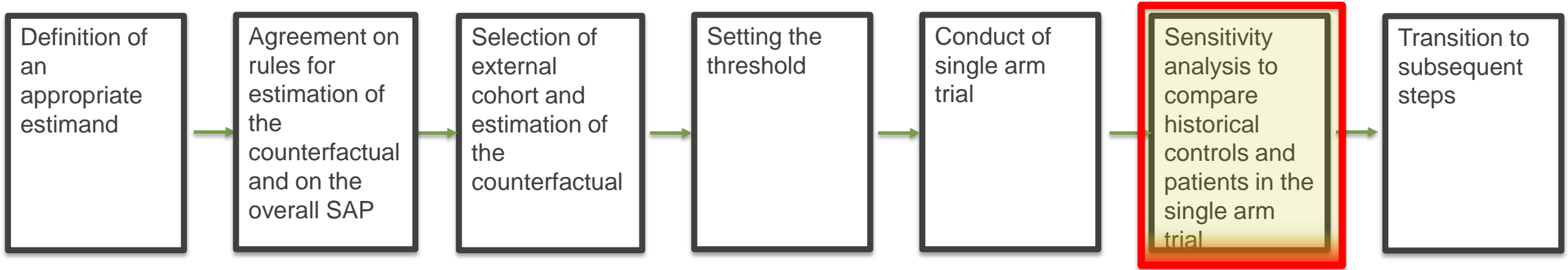
- Severity of disease
- Unmet need of target population
- Availability of alternative treatments
- Patients' input on what is clinically relevant
- Social burden of disease
- Expected frequency of serious adverse events



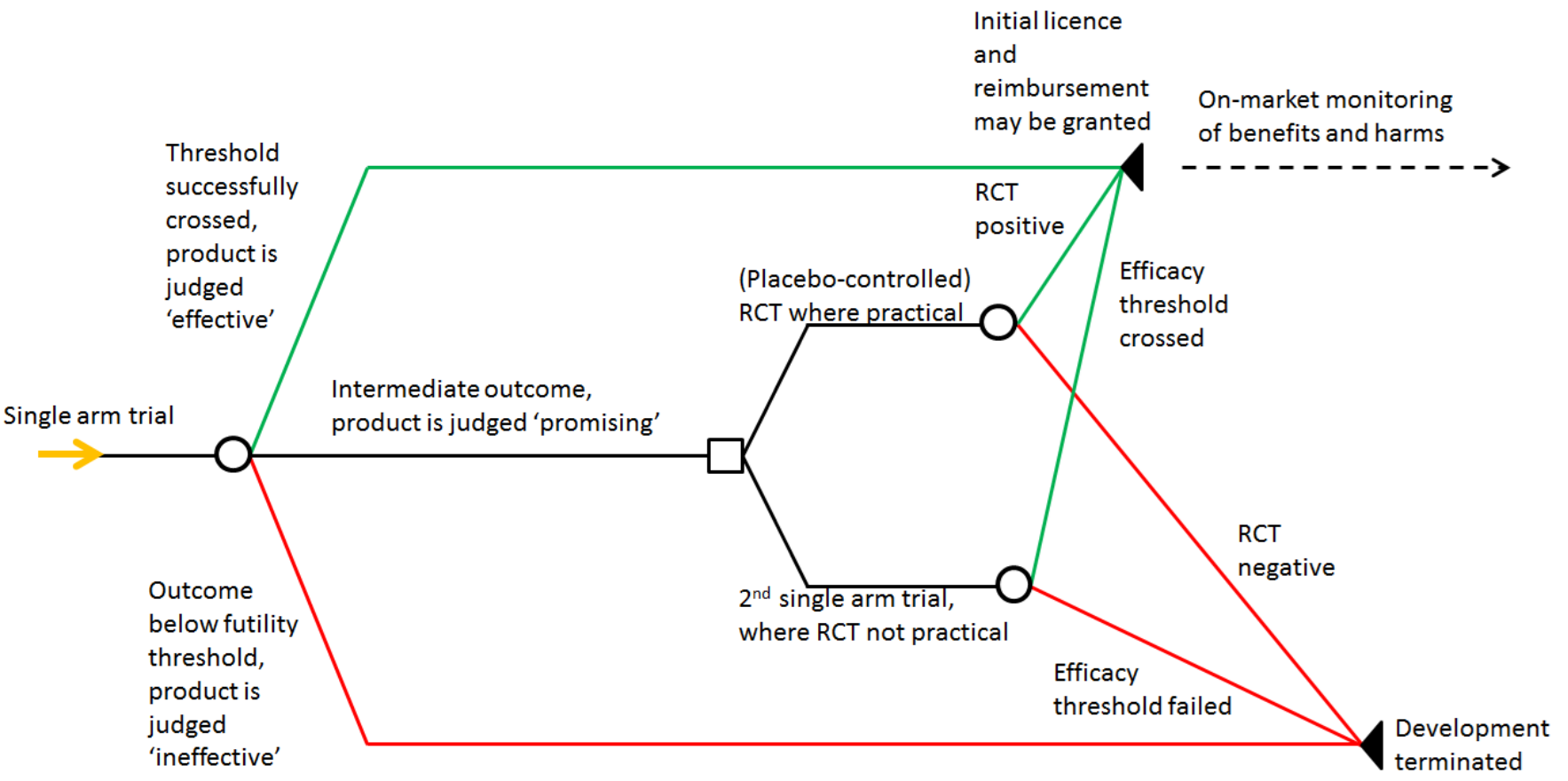
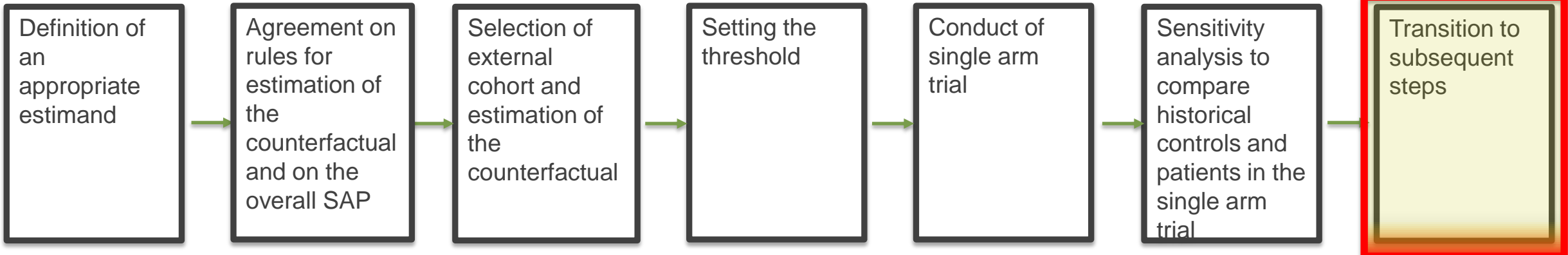
Should be agreed with regulators and other relevant decision makers



- ▶ 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



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

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 efficacy threshold; in practice, the approach will have to be implemented with a view to demonstrate an acceptable benefit-risk profile
- ▶ Methodological risks: No randomisation and blinding – increased risk of bias
- ▶ Expectation risks: Setting (un)realistic thresholds?

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.