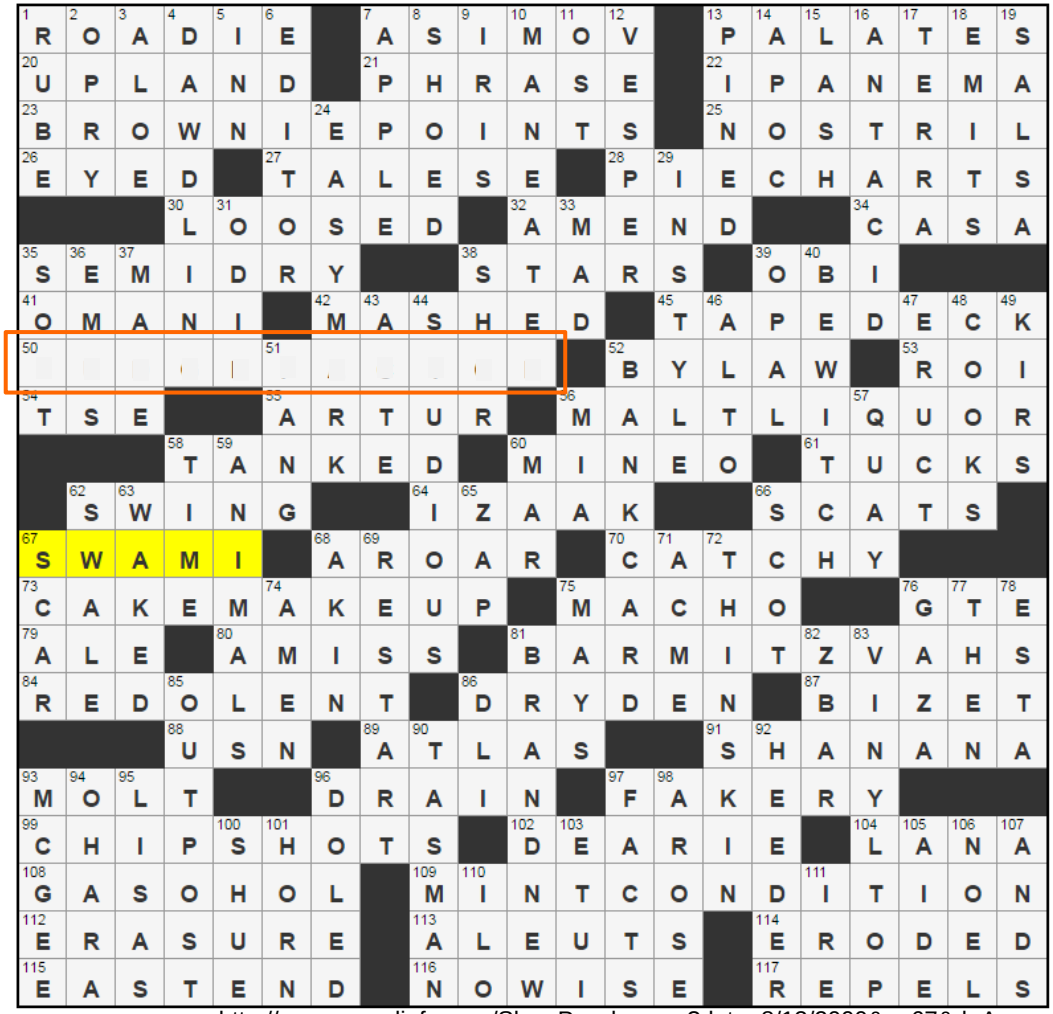
Comment on “ICH E9(R1): Estimands and Sensitivity Analyses in Clinical Trials” by Dr. Wang

Toshimitsu Hamasaki, PhD, Pstat®
National Cerebral and Cardiovascular Center

National Cerebral and Cardiovascular Center

What is statistician's margin of error?



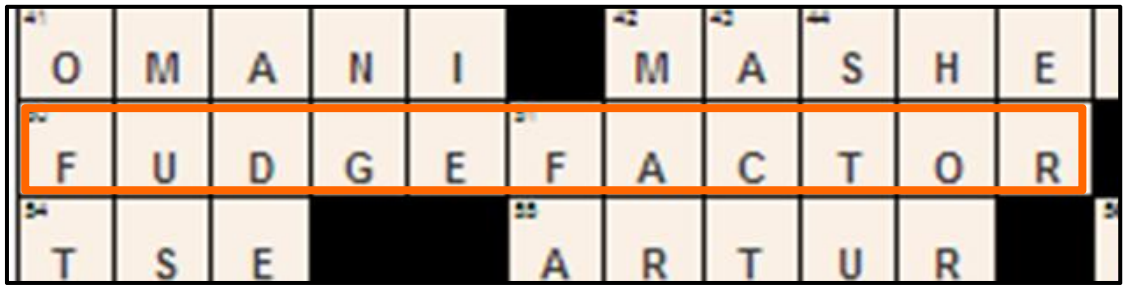
The New York Times

Crossword puzzle
New York Times, August 13, 2000

11-letter answer
"Statistician's margin of error"

- **PLUS-OR-MINUS?**
- **INFERIORITY?**
- **SUPERIORITY?**

“Fudgefactor” is statistician’s margin of error



Example: Einstein’s field equation

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

cosmological constant Λ

Fudge

1. to cheat or welsh (often followed by on): to fudge on an exam; to fudge on one's campaign promises.
2. to avoid coming to grips with something: to fudge on an issue.
3. to exaggerate a cost, estimate, etc., in order to allow leeway for error

http://www.dictionary.com/browse/fudge

Wikipedia In cosmology, the cosmological constant is the value of the energy density of the vacuum of space. It was originally introduced by Albert Einstein in 1917, as an addition to his theory of general relativity to “hold back gravity” and achieve a static universe, which was the accepted view at the time. Einstein abandoned the concept after Hubble's 1929 discovery that all galaxies outside the Local Group are moving away from each other, implying an overall expanding universe. From 1929 until the early 1990s, most cosmology researchers assumed the cosmological constant to be zero.

https://en.wikipedia.org/wiki/Cosmological_constant

Fudgefactors for estimands in clinical trials?

A **quantity of interest** whose true value "**we**" want to know

A **metho**d for estimating the **estimand**.

A **numerical estimate** of the estimand that results from the use of a particular **estimator**

estimand
target populations
primary variables
methods for estimating the estimand
study designs
quality of clinical trial conduct

What is Bernard's estimand?

IX. The Use of Calculation in Study of Living Beings; Averages and Statistics

I will cite still another example borrowed from surgery. A great surgeon performs operations for stone by a single method; later he makes a statistical summary of deaths and recoveries, and he concludes from **these statistics that the mortality law for this operation is two out of five**. Well, I say that **this ratio means literally nothing scientifically and gives us no certainty in performing the next operation**; for we do not know whether the next case will be among the recoveries or the deaths. What really should be done, instead of gathering facts empirically, is to study them more accurately, each in its special determinism. We must study cases of death with great care and try to discover in them the cause of mortal accidents, so as to master the cause and avoid the accidents. Thus, if we accurately know the cause of recovery and the cause of death, we shall always have a recovery in a definite case. We cannot, indeed, admit…

Claude Bernard (1813-1878) was a French physiologist, known as one of the greatest of all men of science. He was one of the first to suggest to the use of blind experiments to ensure the objectivity of scientific observations.

Bernard, C. (1865). *Introduction à l'étude de la médecine expérimentale.* Paris: J.B. Baillière. ("An Introduction to the Study of Experimental Medicine", translated by H.C. Greene)

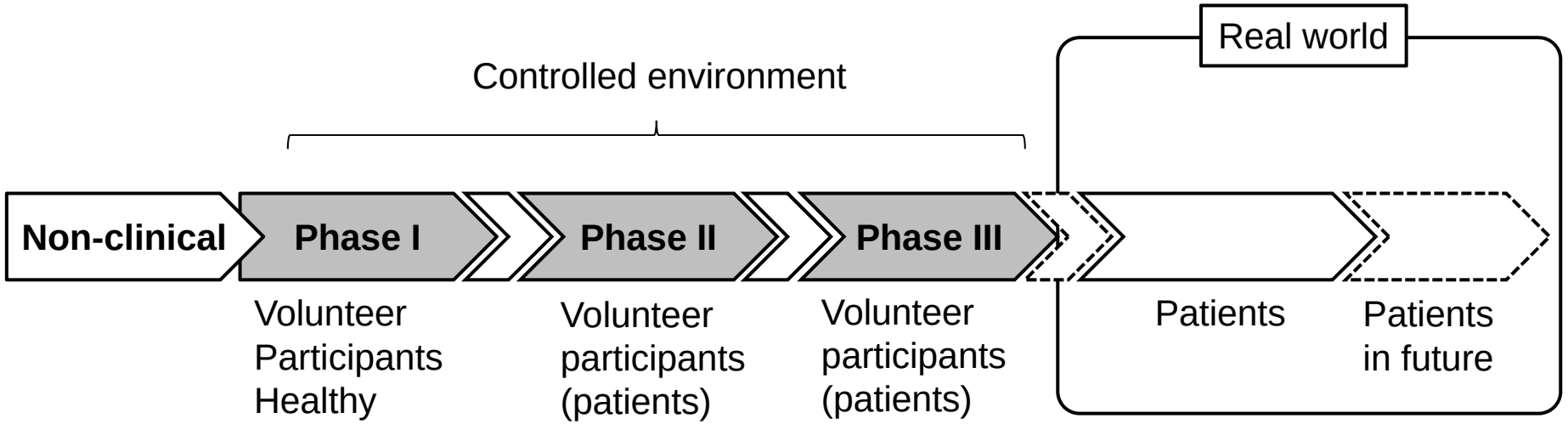
Who are "we" in medical product development?

Estimand

A **quantity of interest** whose true value "we" want to know

- People who may have disease in the future
- Patients' families
- Patients
- Physician investigators
- Volunteer participants in clinical trials
- Regulatory agency
  - Statisticians
  - Clinical Scientists
  - …
- IRB/IDMC
- Industry
  - Statisticians
  - Clinical Scientists
  - ….

“Estimand” may differ from phase to phase



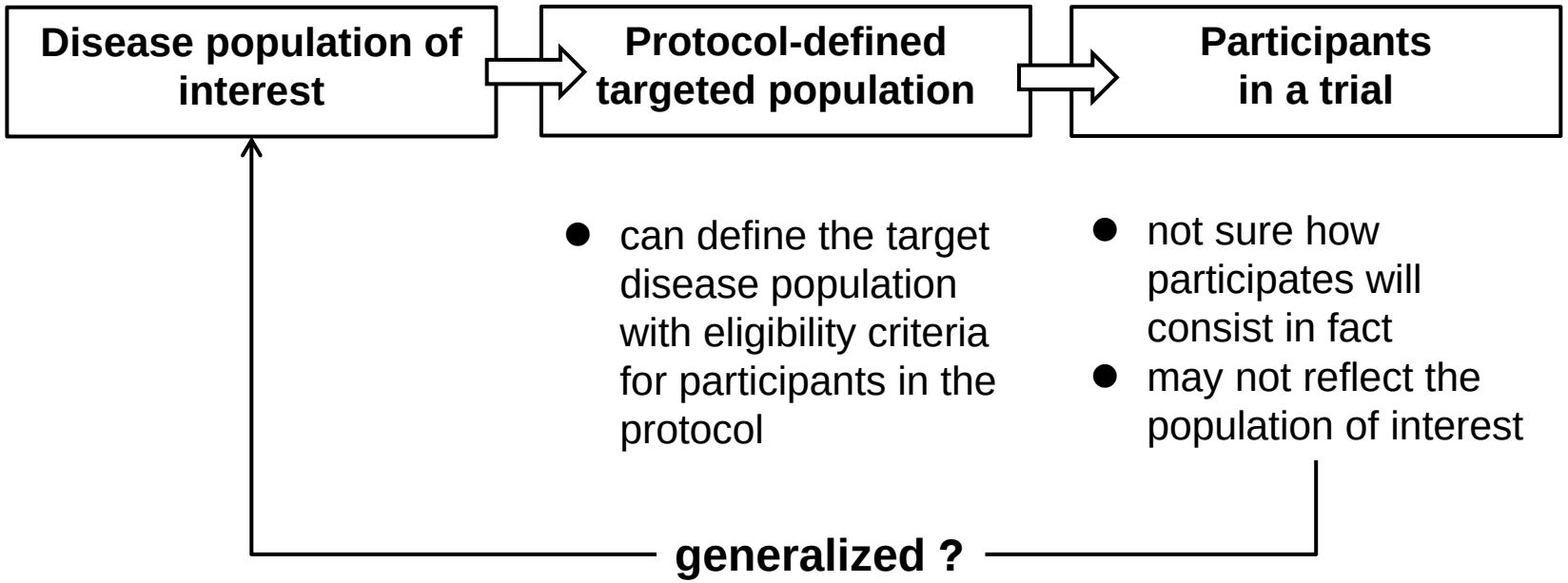
Relatively smaller, restricted population
Internal validity
Explanatory
Experimental
Well-designed

Intervention versus Disease
Population-level response
ATE

Relatively larger population
External validity
Pragmatic
Observational
Strategic

Patients versus Disease
Patient-level response
ATT

“Population of interest” and “Participants in a trial”



Definition of estimand in cardiovascular trials

STATISTICAL PRINCIPLES FOR CLINICAL TRIALS

ICH Harmonised Tripartite Guideline

Having reached *Step 4* of the ICH Process at the ICH Steering Committee meeting on 5 February 1998, this guideline is recommended for adoption to the three regulatory parties to ICH

2.2.2 Primary and Secondary Variables

In many cases, the approach to assessing subject outcome may not be straightforward and should be carefully defined. For example, it is inadequate to specify mortality as a primary variable without further clarification; mortality may be assessed **by comparing proportions alive at fixed points in time**, or **by comparing overall distributions of survival times over a specified interval**. Another common example is a recurring event; the measure of treatment effect may again be a simple dichotomous variable (any occurrence during a specified interval), time to first occurrence, rate of occurrence (events per time units of observation), etc. … (**ICH E9 guideline**)

How to measure and estimate intervention effects

When comparing two response rates P_E (experimental) and P_C (control)

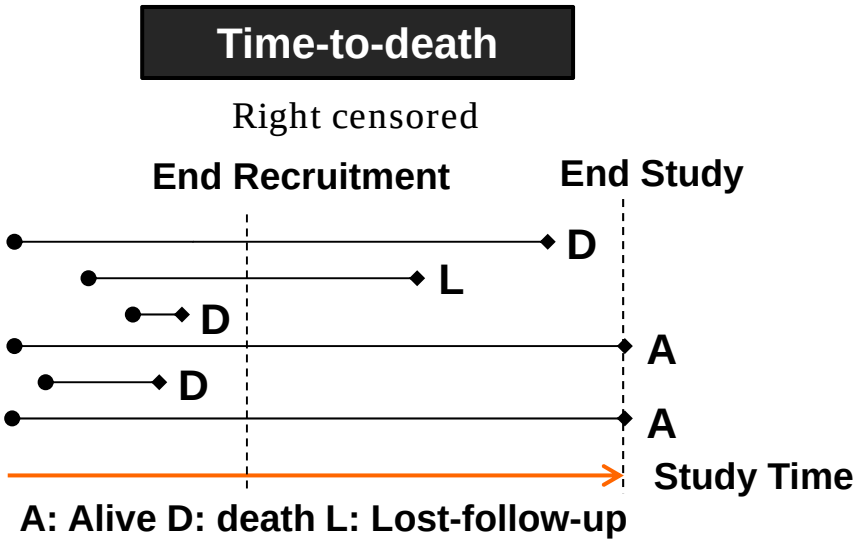
Measure of intervention effects

- Ratio $P_{\mathrm{E}}/P_{\mathrm{C}}$ (Relative Risk or Risk Ratio)
- Relative Risk Reduction $1 - P_{\mathrm{E}}/P_{\mathrm{C}}$
- Odds Ratio $\{P_{\mathrm{E}}/(1-P_{\mathrm{E}})\}/\{P_{\mathrm{C}}/(1-P_{\mathrm{C}})\}$
- Risk difference $P_{\mathrm{E}} - P_{\mathrm{C}}$
- # of needed to treat (NNT) $1/(P_{\mathrm{E}} - P_{\mathrm{C}})$

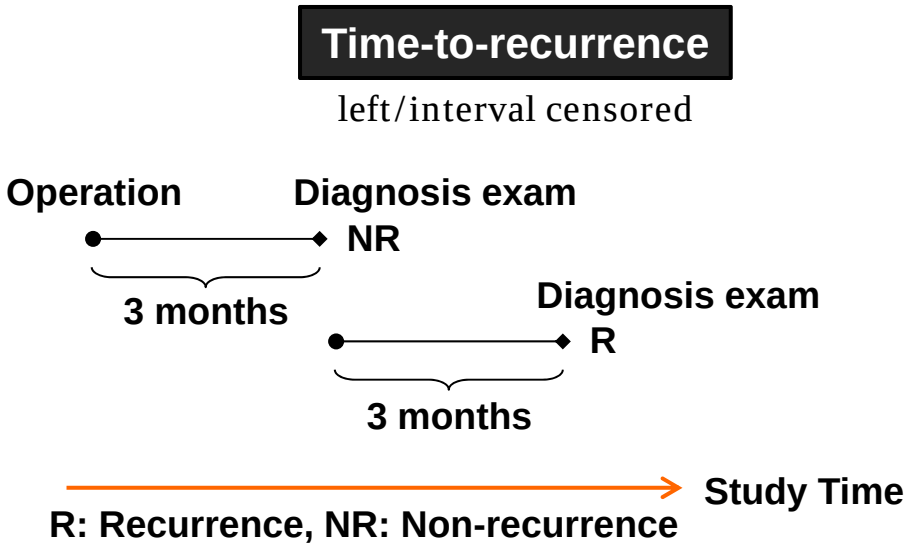
Estimator

- Wald-type estimator
- Maximum likelihood estimator
- Score-type estimator
- More “direct” type estimator
- …

Time-to-event outcomes and censoring scheme



- Estimated time to death may be **shorter** than the true time to death, but the true time to death is unknown



- Estimated time to recurrence may be **longer** than the true time to recurrence, but the true time to recurrence is unknown
  - Depends how often the diagnosis exam scheduled
  - Frequently-scheduled diagnosis test would lead to an ethical issue if it is Invasive

Estimand in noninferiority clinical trials

5.2.3 Roles of the Different Analysis Sets

…

The full analysis set and the per protocol set play different roles in superiority trials (which seek to show the investigational product to be superior), and in equivalence or non-inferiority trials (which seek to show the investigational product to be comparable, see section 3.3.2). In superiority trials the full analysis set is used in the primary analysis (apart from exceptional circumstances) because it tends to avoid over-optimistic estimates of efficacy resulting from a per protocol analysis, since the non-compliers included in the full analysis set will generally diminish the estimated treatment effect. **However, in an equivalence or non-inferiority trial use of the full analysis set is generally not conservative and its role should be considered very carefully.**

ICH-E9 guidance on “Statistical Principle for Clinical Trials”

Choice of analysis population in noninferiority trials

F. Study Quality and Choice of Analysis Population

Traditionally, the primary analysis of a randomized clinical superiority trial follows the intention-to-treat (ITT) principle, namely, all randomized patients are analyzed according to the treatment to which they were randomized, including patients who leave the study prematurely... **Adhering to the ITT principle in superiority trials is generally considered conservative, in that poor study quality resulting in a large number of protocol violations will tend to bias the results towards the null hypothesis of no difference between treatments. The opposite is true for NI trials. Quality issues could result in treatment groups appearing similar (i.e., biasing the results towards the alternative hypothesis for NI trials), when, in fact, the test drug may be inferior, as mentioned in section III.D.3**. Many problems that may cause a superiority trial to fail, such as non-adherence, misclassification of the primary endpoint, or attrition, can bias the results toward no treatment difference (success) and undermine the validity of the trial, creating apparent non-inferiority when the test drug is in fact inferior. **Imputation of missing data under the inferiority null hypothesis is one possible approach to countering the bias due to attrition.**

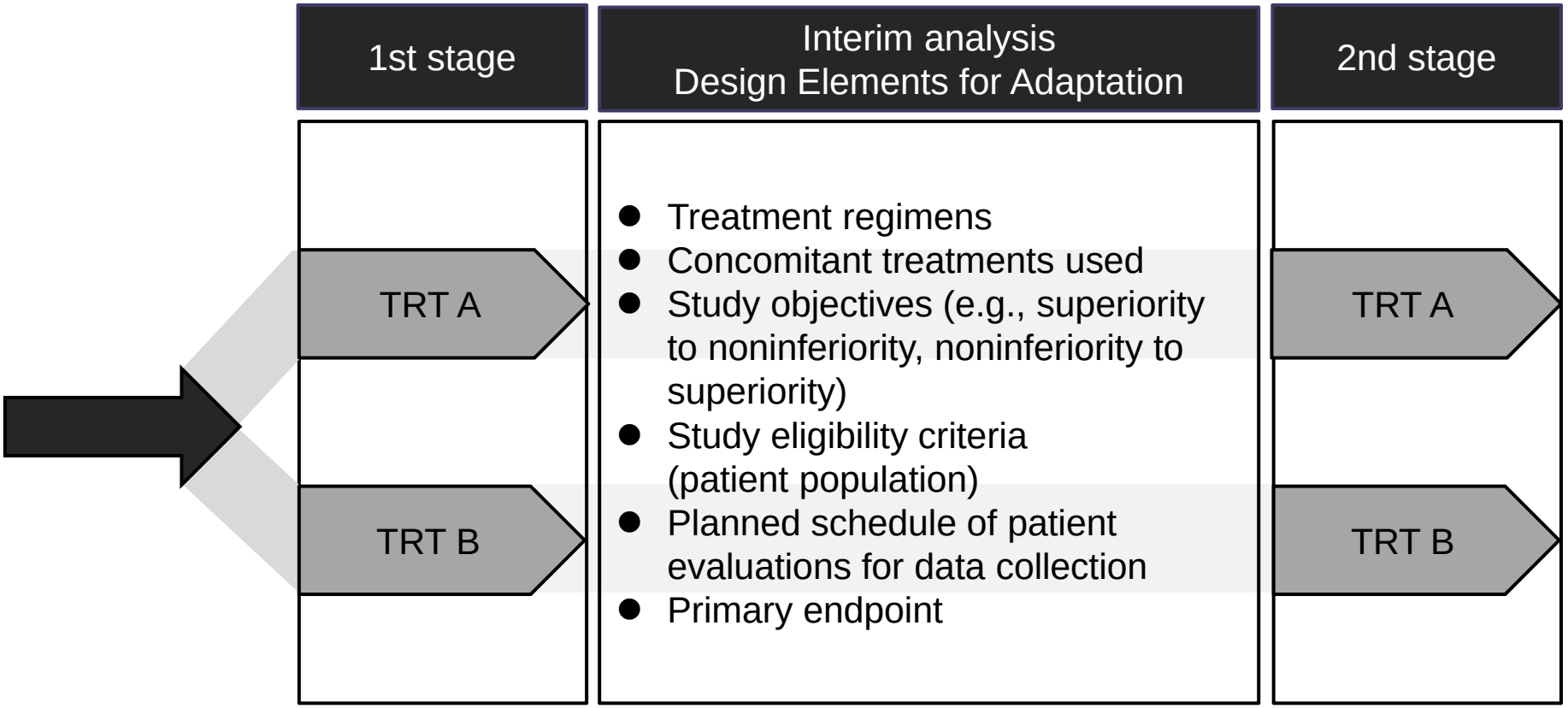
FDA (2016) Guidance for Industry: Non-Inferiority Clinical Trials to Establish Effectiveness

Sensitivity analysis

Sensitivity analysis is to evaluate the robustness of the findings or conclusions based on primary (major) analyses of data in clinical trials

- **Internal versus External- Sample-average versus Population-average**

A more complicated issue: conversation with Frank



Estimands may change!

Well-designed studies (including a good plan for trial monitoring) may increase an opportunity of providing a more reliable and robust estimate of the estimand that we want to know, using a simple estimator