Statistical Challenges in Medical Device Investigations

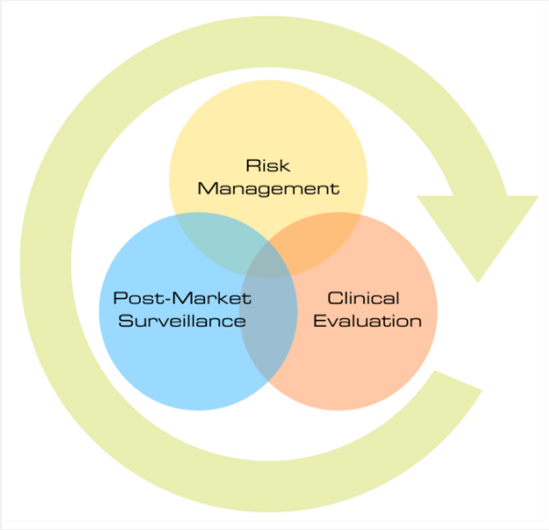
PSDM - 08 November 2022

Corine Baljé-Volkers (c.balje@clin-q.com)

Disclaimer

- Based on own experiences
- 20 minutes is way too short to get into depth, so high-level presentation
- Multiple publications/references available, advisable to read
- Statistical challenges in the phase of medical device *development* (e.g. quality control statistics, lot conformance, production consistency as per 21 CFR 820) is outside the scope of today's presentation.
- US and EU (notified bodies) may look at things (slightly) different

Terminology

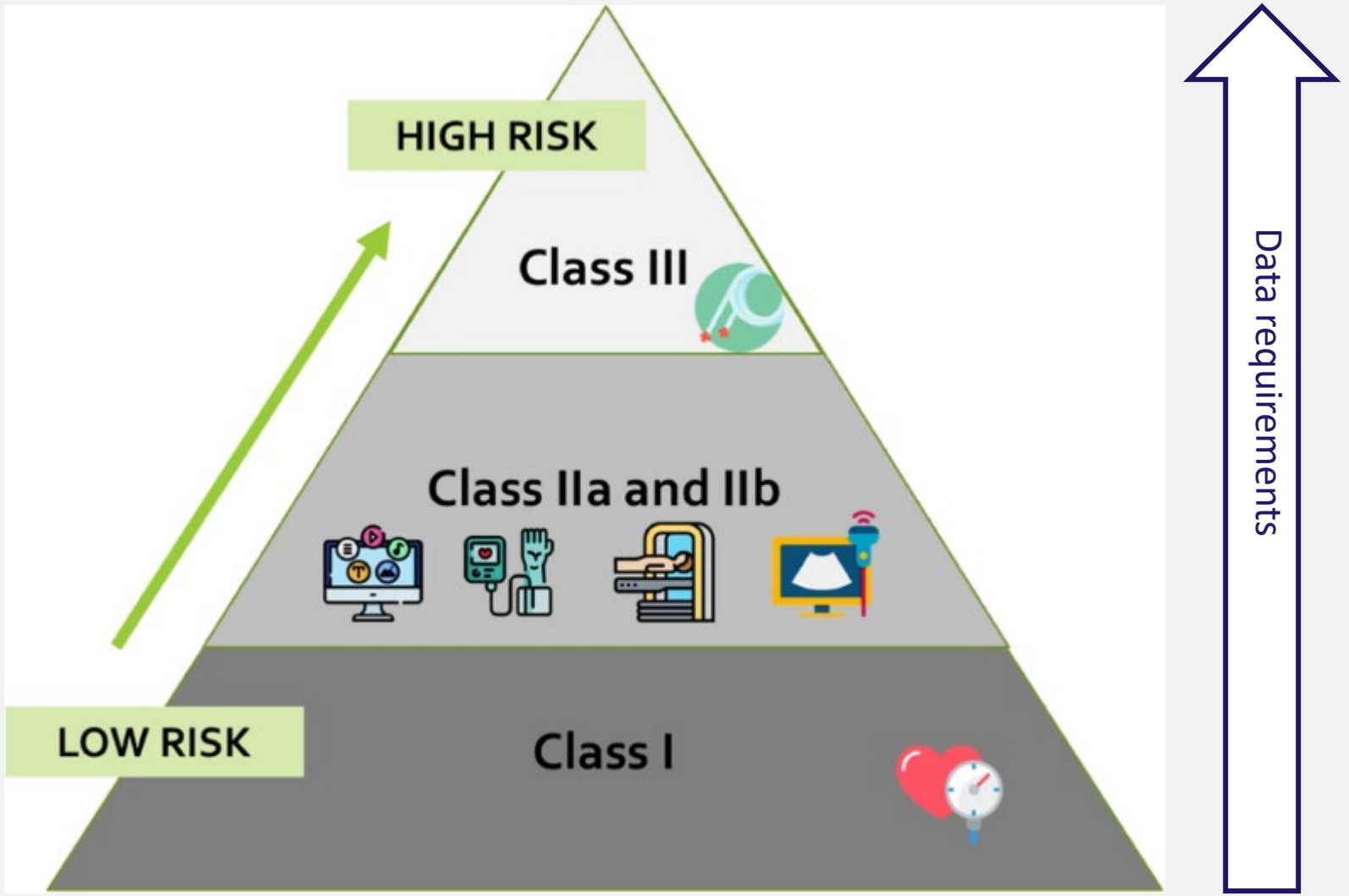


- Clinical investigation ≡ Clinical study/trial
- Clinical investigation plan (CIP) ≡ Clinical study/trial protocol (CSP)
- Drug/treatment ≡ Medical Device/Investigational Device
- Safety and Performance ≡ Efficacy and Safety → Device performs according to claim and is safe in use
- US: FDA (Center for Devices and Radiological Health (CDRH)), EU: Notified bodies (EU member state level)
- Clinical study report (CSR) ≡ Clinical evaluation report (CER) - **living** document throughout expected lifetime of device
- EU: *CE Marking* device meets EU safety, health and environmental requirements. US: "federal/state-enforced USA product compliance" → approval for market release
- Important documents: MDR (Medical Device Regulation, **REGULATION (EU) 2017/745**) and **ISO 14155:2020** "*Clinical investigation of medical devices for human subjects – Good clinical practice*"

Terminology

Classification based on intended use, invasiveness, duration of use, and the risks/potential harms associated with their use.

Class I: low risk → don't require a clinical investigation.



Differences with pharmaceuticals

Pharmaceutical			Medical Device	
Phase			Stage	
"0"	Pilot	Not common. N≈10	Pilot/FIH	Preliminary safety and performance, possible device modifications
I	Safety/toxicity	FIH, small group of healthy volunteers	Feasibility	Safety and performance of (near-) final device, small group of patients
II	Safety & efficacy	Assess dosing, small group of patients		
III	Clinical efficacy	Large group of patients, efficacy and safety	Pivotal	Large group of patients, safety and performance/effectiveness
IV	Postmarket	Longterm safety and efficacy	Postmarket	Longterm usage, safety and performance/effectiveness

Performance = the ability of a device to achieve its intended purpose as stated by the manufacture (could possibly be a clinical effect/benefit)

Types of devices

- Very diverse → interesting & challenging, especially when it comes to study design and statistics.
- Statistical methods: generally the same, underlying design issues are the challenges.
- Nondiagnostic: therapeutic and aesthetic devices (including e.g. implants)
  - **Therapeutic**: for *treating* a condition or disease
- **Diagnostic**:
  - Intended for *detecting* a condition or disease
  - *In vitro* diagnostic tests (applied to human samples such as blood, spit, urine, or tissue) and diagnostic imaging systems
  - Molecular diagnostic tests (precision medicine)

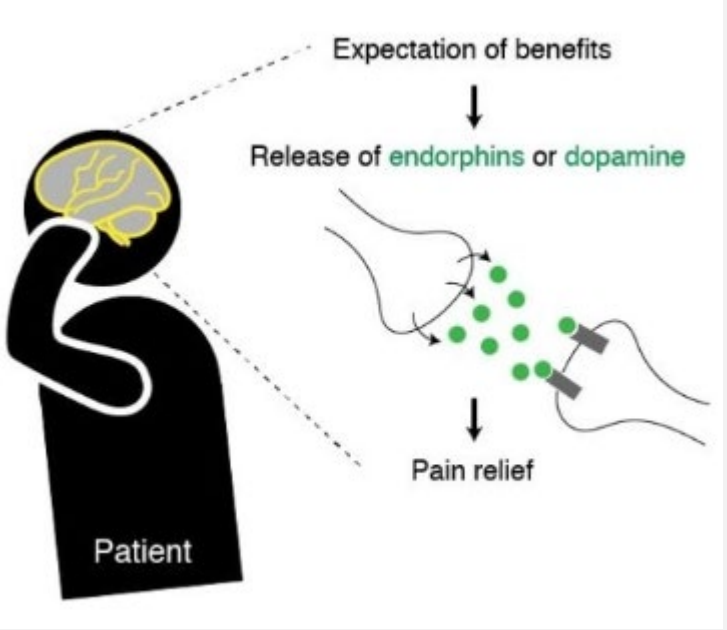
Important unique considerations

- Postmarketing surveillance cycle of reporting
- Diagnostic devices versus therapeutic devices: very different designs and statistical analyses
- Therapeutic medical devices: short commercial life cycle/fast evolution cycle → updated devices possible during clinical investigation.
- Statistical considerations for therapeutic devices: placebo effect, (sham) controls, blinding, missing data, non-inferiority, survival analysis, repeated measures, and historical controls.
- Statistical considerations for diagnostic devices: molecular diagnostic is mainly adaptive design and Bayesian approaches.

- Will discuss a few challenges today, with examples

Use of placebo/control

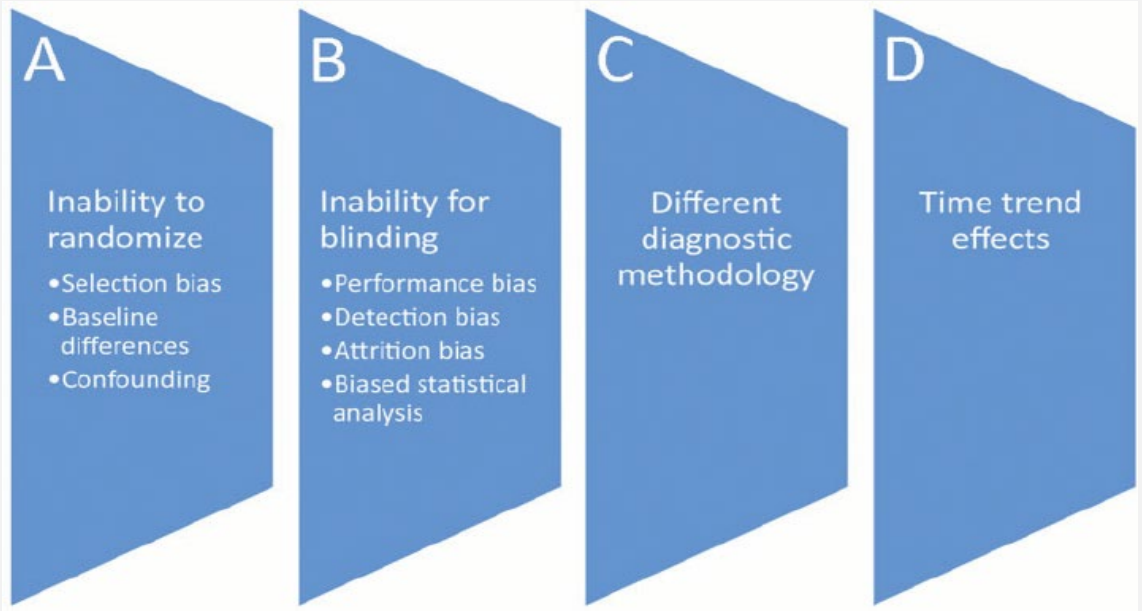
- RCT/double-blind not always possible, may not even recruit patients for this.
- Primary endpoint a subjective patient-reported measure (e.g. pain scores), large placebo effect
  - Change primary endpoint to objective measure
  - Include placebo arm (if ethical/practical, e.g. sham surgery for implanted devices?)
- Active control:
  - Availability
  - Same placebo effect?
  - Blinding
  - Non-inferiority design common
- <Historical/external control: next slide>



Use of historical control

- Historical/external control:
  - meta-analysis of published data
  - possibly the use of (full study) data of a previous similar clinical investigation
- Regulators: “unusual circumstances”:
  - unmet need
  - highly predictable disease course (objectively measured)
  - expectation of a large effect

Potential limitations associated with use of historical control:



Use of historical control

- Observational study comparing with external/historical control (HC):
  - Need to have HC and investigational device arm comparable with regards to characteristics/covariates relevant to the outcome of the study

  → Matching or stratification (*propensity scores*) to mitigate imbalances

  → Two-stage design (without access to/knowledge of primary outcome data):
    - Stage 1 = selection of control group, preliminary estimation sample size, specifications covariates, identify independent statistician
    - Stage 2 (all pts enrolled and all baseline covariate information is available) = estimate propensity score, match patients in investigational arm with patients in control group, check balance in covariate distributions, final assessment on control group formulation and sample size estimation, SAP for future analysis.

Blinding

- Mostly impossible to blind the patient and/or the investigator to device receipt
  - Keep other people involved blinded (e.g. data manager, programmer, statistician, client, ..)
  - Independent blinded endpoint evaluator
  - Use undisputable objective outcome measures

- In case of single/double-blinding: use questionnaire on patient perception/beliefs of treatment assignment to assess bias and correct for it in the statistical assessments (*patient's treatmentality*).

Missing data

"Missing data will not be imputed"

"LOCF will be used for missing data imputation"

"The analysis of the primary endpoint will be based on available data only."

"Subjects who die prior to reaching Day xx will be excluded from analysis."

- More missing data as compared to pharmaceutical trials, variety of reasons:
  - Patients are unblinded, possibly resulting in increased dropout
  - Failed implantation (plus removal) or device failure – rescue treatment may be necessary
  - Lower patient compliance
  - Device studies tend to be more difficult to perform

Missing data

- Completers may differ from people who drop out/have missing data
- May also result in different missing patterns across the treatment arms

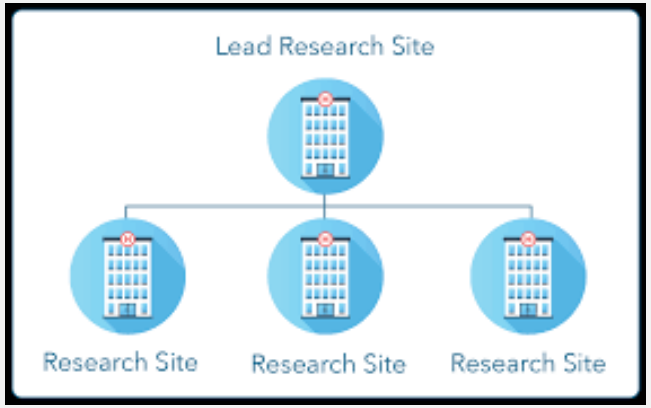
➔ Possibly jeopardizing the validity of the study and interpretation of the results

- Therefore:
  - Sufficient attention at design stage to reduce frequency of missing data and minimize the impact of missing data.
  - Investigate robustness and sensitivity for the study results to the missingness: e.g. interpolation, multiple imputation, **tipping-point analysis**. Worst-case analysis is too extreme to be realistic, LOCF may be too simplistic.

Other possible considerations

- Changes to protocol or device during investigation/study
  - Per MDR: In general, a change to the investigational medical device is a substantial modification and may require a new application for the clinical investigation.
  - Sponsor need to perform a risk analysis to assess the impact of this change on the safety and effectiveness of the device.

- Multi-center trials: treatment-by-center interactions
  - Possible more protocol deviations or more training needed in use of device
  - Learning curve/one center very experienced



Frequently applied designs/statistical methods

- Non-inferiority – issues with device creep (~ biocreep for pharmaceuticals)
  - Active control can be quickly outdated
  - Take into account placebo effect with active control arm, but may not be information available (no sham arm) – variable margins
  - Possibilities to switch to superiority

- Devices requiring a long-term FU: survival analysis.
  - Usually a smaller sample size.
  - Matched-pair designs (e.g. bilateral knees, nostrils, eyes, skin locations..) require non-standard survival analysis methods
  - Applied to repeated events instead of first event (e.g. restenosis after stent implant, or infections after cochlear implants)

Frequently applied designs/statistical methods

- Repeated measures
  - challenge when device is updated during clinical investigation
- Performance of device/diagnostic accuracy when there is a golden standard:
  - sensitivity/specificity, predictive values, ROC analysis
- Bayesian statistics for Molecular diagnostic tests (precision medicine):
  - E.g. Bayesian latent class model (LCM) analysis if there is no golden standard

Postmarket Surveillance

- Surveillance of medical devices after marketing permission: mainly finding trends/signals in a database of medical device adverse event reports
  - E.g. nonparametric regression or change-point analysis: series of time ordered data in order to detect whether any changes have occurred (CUSUM plots)
- Also further/long-term conformance of clinical performance/effectiveness
  - e.g. when there is uncertainty about the device in use,
  - for high-risk medical devices,
  - unexpected serious/rare adverse events,
  - in case of expedited market access,
  - assessment of long-term effects,
  - new intended use

Publications/references

- MDR (Medical Device Regulation, **REGULATION (EU) 2017/745**)
- **ISO 14155:2020** "*Clinical investigation of medical devices for human subjects – Good clinical practice*"

- Campbell, G e.a. (2011) Missing Data in the Regulation of Medical Devices *J Biopharm Stat.*, 21:2, 180-195
- Campbell, G. and Yue, L.Q. (2016) Statistical innovations in the medical device world sparked by FDA. *J. Biopharm. Stat.*, 26 (1), 3–16.
- Heng Li & Lilly Q. Yue (2007) Statistical and Regulatory Issues in Nonrandomized Medical Device Clinical Studies, *J Biopharm Stat.*, 18:1,20-30.
- Lao, C.S. Statistical Issues Involved in Medical Device Postmarketing Surveillance. *Ther Innov Regul Sci* 34, 483–493 (2000).
- Levenson M. and Yue L. Regulatory Issues of Propensity ScoreMethodology Application to Drug and Device Safety Studies. *J Biopharm Stat*. Volume 23, 2013, Issue 1:
- Redberg, RF. "Sham Controls in Medical Device Trials". *The New England Journal of Medicine* 2014;371(10):892-3:
- Sandie, A.B., Molinari, N., Wanjoya, A. *et al.* Non-inferiority test for a continuous variable with a flexible margin in an active controlled trial: an application to the "Stratall ANRS 12110 / ESTHER" trial. *Trials* 23, 202 (2022)
- Wiens, B.L., Lystig, T.C., and Berry, S.M. (2013) Recent statistical contributions to medical device development. *Ther. Innov. Regul. Sci.*, 48 (1), 90–97.
- Xu, Z e.a. Signal detection using change point analysis in postmarket surveillance. *Pharmacoepidemiol Drug Saf.* 2015 Jun; 24(6): 663–668.
- Zhang, Z., Kotz, R.M.,Wang, C., *et al.* (2013) A causal model for joint evaluation of placebo and treatment-specific effects in clinical trials. *Biometrics*, 69, 318–327.
- Yuan X e.a. Missing data handling methods in medical device clinical trials. *J Biopharm Stat*. 2009 Nov;19(6):1085-98.
- Yue L, Statistical and regulatory issues with the application of propensity score analysis to nonrandomized medical device clinical studies. *J Biopharm Stat*. 2007;17(1):1-13; discussion 15-7, 19-21, 23-7 passim.