

Directions for new developments on statistical design and analysis of small population group trials

Prof dr Kit Roes
Professor of Clinical Trial Methodology
Julius Center, UMC Utrecht

Principal Investigator of EU FP7 **Asterix** project

Outline

- Why attention for statistical methodology?
- IRDiRC recommendations
- Agenda for new developments
- Regulatory impact

Why attention to statistical methodology?

EMA Guideline on Small populations: “No methods exist that are relevant to small studies that are not also applicable to large studies. However.....less conventionalmethodological approaches may be acceptable if they help to improve the interpretability of the study results”.

Average 761 (median 538) patients enrolled in orphan drug trials.

Average 3,549 (median 1588) in non-orphan drug trials.

Why attention to statistical methodology?

- More often an area of high medical need (no treatment)
- Rare disease with large heterogeneity between patients in disease course.
- In (very) rare disease a relatively large fraction of the population to treat could be included in clinical trials (finite “patient horizon”).
- Challenge of appropriate (clinical) endpoints and biomarkers.
- Evidence synthesis more challenging (replication of trials, between study heterogeneity).

Example in Fabry's disease

Galafold (migalastat) in Fabry's disease, auth. May 26, 2016.

Long-term treatment of adults and adolescents aged 16 years and older with a confirmed diagnosis of Fabry disease (galactosidase A deficiency) and who have an amenable mutation.

4 open label Phase 2 studies in 27 patients (23 in longer term follow-up).

2 pivotal comparative studies:

- db placebo controlled. 67 patients randomized (34 Galafold, 33 placebo).

- open label, ERT controlled. 60 patients randomized (36 Galafold, 24 ERT).

Example in Fabry's disease

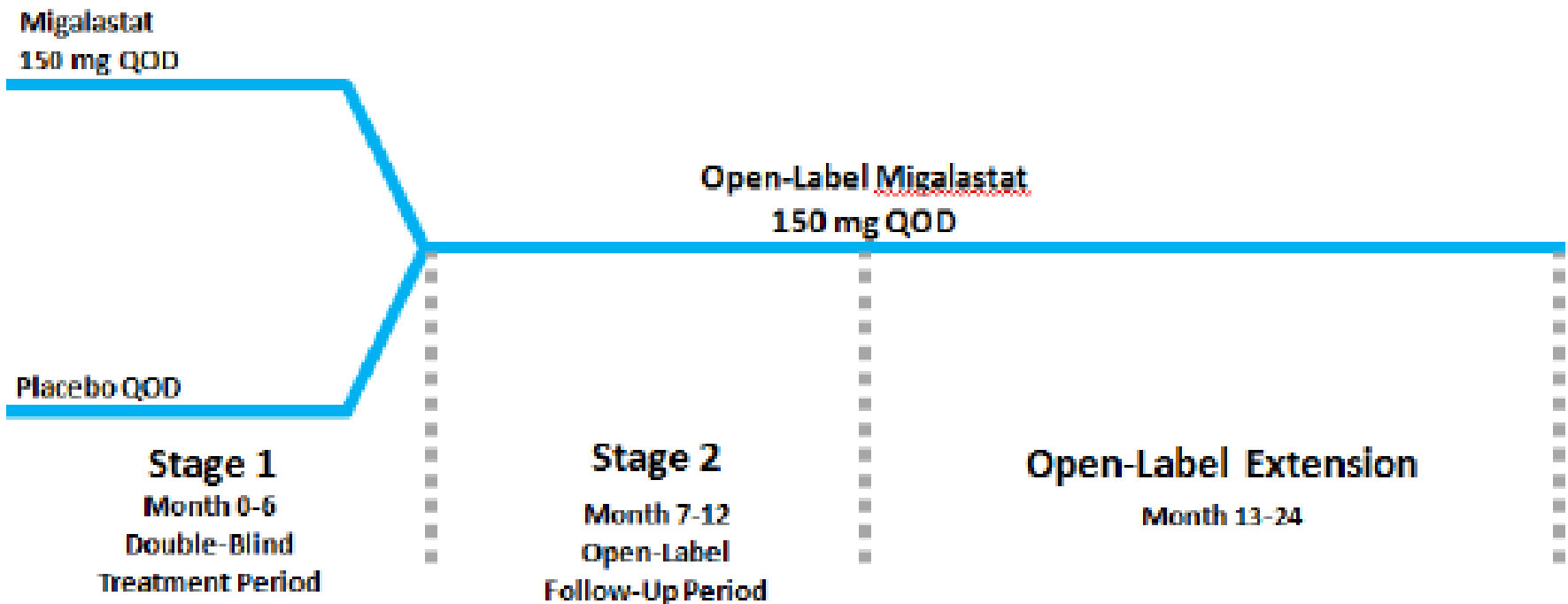


Figure 3: AT1001-011: Study Design

Example in Fabry's disease

- Primary endpoint analysis for Stage 1: The **proportion of successes** (i.e., percentage of patients with a $\geq 50\%$ reduction from Baseline in the average number of IC GL-3 inclusions)

Migalastat

13/34

Placebo

9/33

p-value

0.3

- Secondary (a.o): Mean percentage change in number of IC GL-3 inclusions.

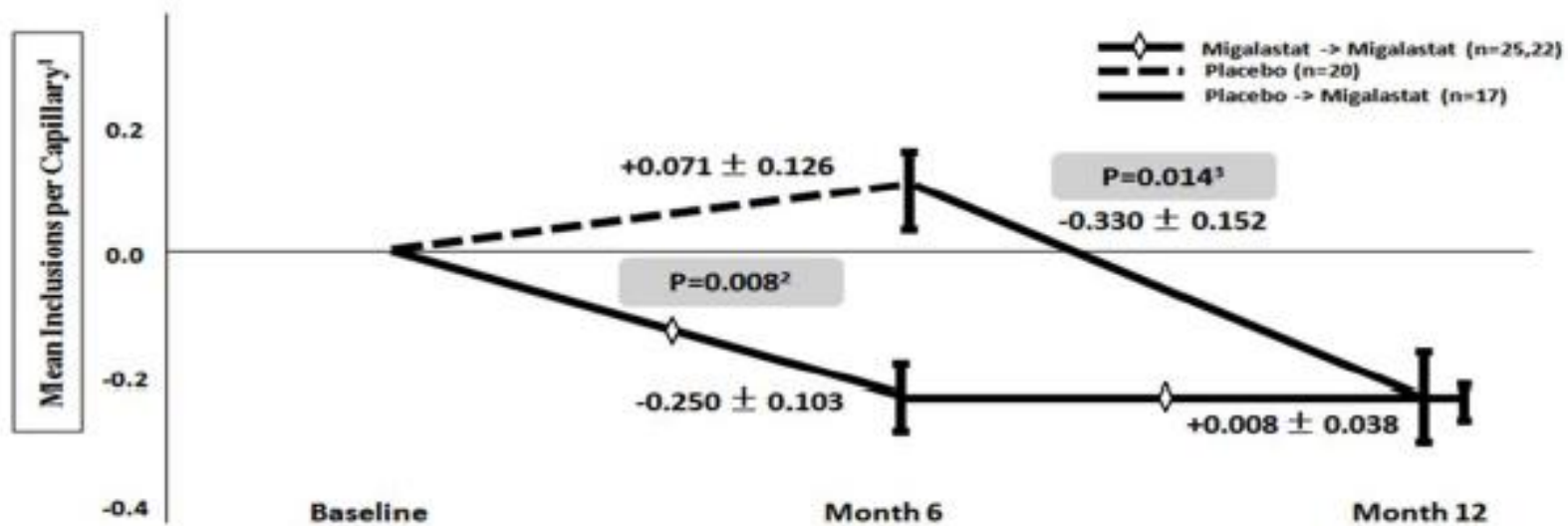
-8.0 +/- 105

13.0 +/- 90.5

0.097

Example in Fabry's disease

Figure 4: AT1001-011: Change from Baseline in the Mean Number of GL-3 Inclusions per Kidney Interstitial Capillary



Post-hoc analysis In 50/67 patients with amenable mutations.

Example in Fabry's disease

- Classical type I error (5%), acknowledging limited power.
- An inefficient choice of primary endpoint (dichotomized).
 - ▶ That *as endpoint* does not really play a role in the actual overall interpretation and assessment of the results.
- Although by design there was Type I error control, the actual Type I error of the conclusion cannot be ascertained.
- Multiple (secondary) endpoints.

IRDiRC recommendations

Identified area's in the workshop (a.o.)

- Different methods/designs versus types of conditions
- Stimulate use of existing methods to increase efficiency
- Decision analytic approaches and rational approaches to adjusting levels of evidence
- Extrapolation problems and opportunities
- Patients' engagement in study design



Agenda for new developments

POSITION STATEMENT

Open Access



Directions for new developments on statistical design and analysis of small population group trials

Ralf-Dieter Hilgers^{1*}, Kit Roes^{2†}, Nigel Stallard^{3†} and for the IDeAI, Asterix and InSPIRe project groups

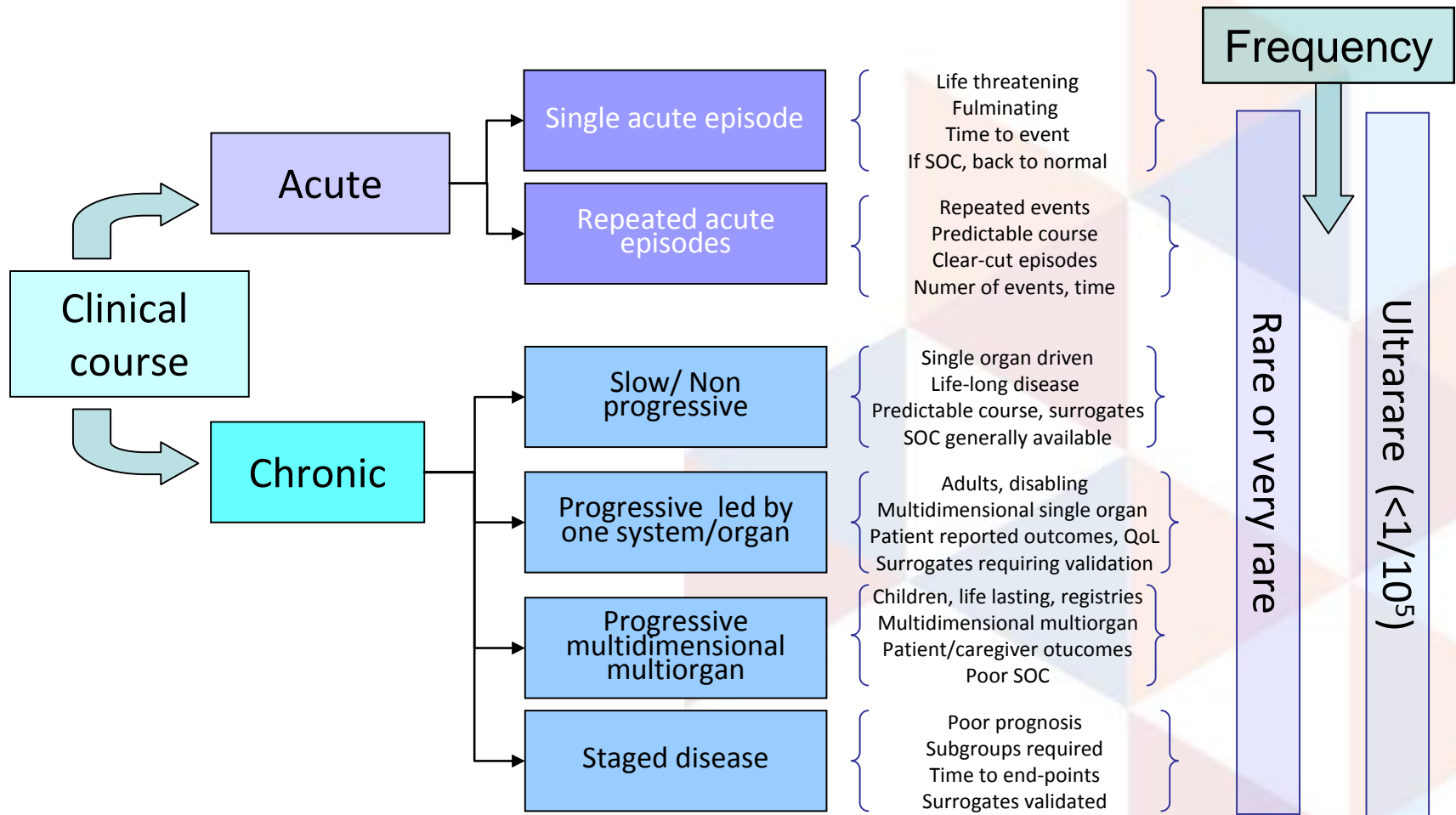


- Level of evidence
- Pharmacological considerations
- Methodological and statistical considerations
- Extrapolation
- Patient involvement and ethical considerations

Methods vs types of conditions

- Guidance on design at disease level no longer practical (over 8000 rare diseases).
- One general document (at present) may not provide sufficient guidance.
- Framework with intermediate approach, driven by key characteristics of disease and treatment.
- Developed with clinical and statistical considerations, based on about 100 EMA dossiers.

Proposed framework (asterix)



Level of evidence: Patient horizon

Design by Type I and Type II errors aims to balance favorable and unfavorable decisions for future patients.

N: Total number of patients up to certain time horizon.
(e.g. next 10 years, or up to next treatment option)

n: The number of patients to include in the clinical trial(s)
($n = n_1 + n_2$, on new and comparator respectively)

At the end of the trial a decision is taken –

to apply treatment i to the remaining $N - n_1 - n_2$ patients.

Level of evidence: Patient horizon

Determine $n = n_1 + n_2$: *Expected benefit* over all N patients “maximized”.

Preliminary (first and old!) results: *Optimal* $O(\sqrt{N})$

Biometrical Journal 00 (2016) 00, 1–17 DOI: 10.1002/bimj.201500228

1

Determination of the optimal sample size for a clinical trial accounting for the population size

Nigel Stallard^{*1}, Frank Miller², Simon Day³, Siew Wan Hee¹, Jason Madan⁴, Sarah Zohar⁵, and Martin Posch⁶

Further research: Can it provide rational basis for Type I / Type II errors?

Evidence synthesis

- Prospective meta-analysis as part of evidence synthesis.
 - ▶ Similar studies, or even different designs.
 - ▶ Methodological challenges: between study heterogeneity, rare events.
 - ▶ Robust estimation and confidence intervals, valid under wide range of heterogeneity (hybrid bayesian – frequentist).
- Incorporate prior (control) data to reduce sample size of new trial.
 - ▶ Prospectively planned, Type I error controlled

Incorporate prior data to reduce sample size

- Include first study as “prior information” into analysis of the second (Phase III).
- Weight of study decreases with increasing heterogeneity.
- Assess and control type 1 error properties.
- *Hybrid Bayesian-Frequentist approach, already being proposed in Scientific Advices*

Regulatory impact

- Assessment & increasingly advanced methodology needs more expert involvement throughout EU. Cannot be fully captured in guidance.
- New guidance for small populations along the proposed framework to make it more applied and useful.
- Level of evidence & rational standards: beyond type I error and power; impact not only in rare diseases.
- Prospective (robust) meta-analysis should be more often considered, including different data sources (registries).