Statistical Design and Analysis of a Non-Inferiority Clinical Trial

Áine Glynn[1], Filip Kłosowski[1]

[1] School of Mathematical and Statistical Sciences, University of Galway

Background

Non-Inferiority (NI) trials are increasingly used in clinical research, especially when a placebo is unethical and when new treatments aim for similar efficacy with other advantages[1,2]. However, they are frequently poorly designed and interpreted. Confusion arises around the specification of NI margins, selection of appropriate active controls and interpretation of statistical conclusions. This may lead to adverse consequences for manufacturers, clinicians and the wider public.

What is a Non-Inferiority Trial?

- **Non-Inferiority Trials:** Clinical studies designed to demonstrate a new treatment is not clinically worse than an active control by more than a pre-specified margin.
- **Non-Inferiority Margin (Δ):** Pre-specified and approved threshold the new treatment must meet to prove it preserves a clinically meaningful portion of the active control's effect [3].
- NI Trials typically run like randomised control trials but compare the new treatment with an **active control**; an established standard of care used.
- NI must show the new treatment's estimated effect, along with its Confidence Interval (CI), lies within the pre-specified NI margin.

Aims & Objectives

1. Justify the Δ using historical evidence.
2. Design and assess appropriate NI trial design parameters.
3. Develop a Shiny application for exploring sample size requirements.
4. Analyse simulated data and evaluate NI conclusions.
5. Examine the assumptions and limitations of NI designs.

Biomimics 3D Stent

It's 2008 and we have just been hired as statisticians for BioMimics 3D Vascular Stent's pivotal NI trial:

The BioMimics 3D stent is a peripheral vascular stent implanted in the leg to improve blood flow in patients with peripheral vascular disease (narrowing of the peripheral blood vessels). Unlike conventional straight stents, it features a 3D helical design, intended to improve vascular performance and blood flow to affected vessels.

From a statistical perspective, this study is a single arm trial evaluated against a fixed Performance Goal (PG). Such trials are commonplace for demonstrating NI in medical devices due to cost, extended timelines and challenges with recruitment and blinding. Strictly speaking they are no different than a single arm trial against a fixed endpoint. The similarity to an NI trial is that the endpoint *(i.e. the PG)* is chosen to represent a 'worst case' for safety (or efficacy) to claim NI. As such, many NI trials reported are actually single arm trials with a PG.

We derived the PG and the corresponding NI margins by setting safety and efficacy endpoints using a targeted literature review and previous meta-analysis[4,5,6]. This was done using IPA statistics and applying a random-effects model.

For safety, the margin is set at the upper bound of the 95% CI, representing the maximum acceptable level of harm. For efficacy, the margin is set at the lower bound of the 95% CI, representing the minimum clinical benefit that must be preserved. Crossing either bound results in failure to demonstrate NI.

Our project involves working with BioMimics' lead scientist, to design, assess and interpret this trial from a statistician's perspective. This includes:

- Defining and justifying appropriate safety and efficacy endpoints.
- Ensuring the trial is statistically powered and ethically justified.
- Selecting valid analysis methods.
- Correctly interpreting and communicating NI conclusions.

Meta-Analysis

Prior to being available on the market all medical products must obtain approval from the relevant regulatory authority. This is done through a series of clinical studies where the medical product is monitored for safety and effectiveness.

Initially, a feasibility study is conducted. If successful, a pivotal study is proposed. Our pivotal study is a single-arm PG trial. The PG and NI margin must be statistically justified, adequately powered and grounded in clinical evidence.

Endpoint	Pooled Estimate	95% CI	Recommended Performance Goal	Justifiable Range
30-Day Amputation	0	[0.0000, 1.0000]	1%	0–1%
30-Day Death	0	[0.0000, 1.0000]	1%	0–1%
30-Day Target Vessel Revascularisation (TVR)	0.0517	[0.0234, 0.1104]	11%	2–11%
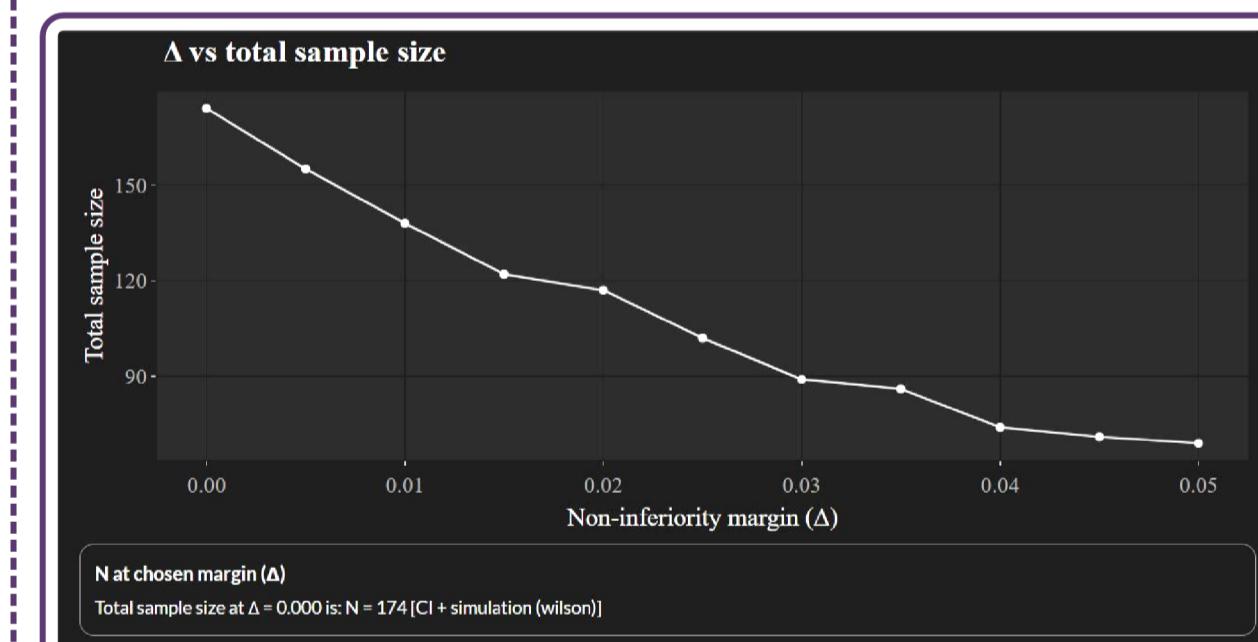
Endpoint	Pooled Estimate	95% CI	Recommended Performance Goal	Justifiable Range
Rutherford Classification Change (12 months): Improved or No Change	0.9583	[0.8786, 0.9865]	87%	87–99%
Rutherford Classification Change (12 months): Increase by One Class	0.0441	[0.0143, 0.1280]	1%	1–13%

Sample Size

The sample size was calculated for the safety outcome. Based on an estimated safety proportion of 0.95 a sample size of **n = 174** had 80% power based on a one-sided test for a binomial at the $\alpha = 0.025$ significance level to declare NI against a PG for safety of 0.89. The sample size was determined via Monte Carlo simulation using a Wilson score CI-based decision rule.

Shiny Application

As part of the sample size calculation we developed an interactive Shiny application to support statistical planning for this trial and future trials of the same nature. The application allows users to explore how sample size requirements vary under different design assumptions, including power, significance level and effect size.

Δ vs total sample size

N at chosen margin (Δ)
Total sample size at Δ = 0.000 is: N = 174 (CI + simulation [wilson])

The application also supports multiple CI–based decision rules. Exploratory analyses within the app were used to compare alternative methods and assess their impact on NI conclusions and required sample size. The Wilson score CI approach was ultimately selected based on these analyses.

Next Steps

- Carry out an extensive simulation study to compare the different approaches available to generate a CI for a population proportion in single arm medical device trials.
- Calculate the sample size needed for efficacy assuming a gate-keeping approach for joint outcomes.
- Calculate the sample size needed for safety and efficacy if an interim analysis is required.
- Finalise a Statistical Analysis Plan (SAP) specifying endpoints, estimators, CI methods and NI decision rules prior to data unblinding.
- Conduct sensitivity and tipping-point analyses to evaluate the robustness of conclusions to small changes in assumptions or observed event counts.

GitHub

Meta-Analysis:

github.com/FilipMKgit/MDX_Meta-Analysis

Shiny:

github.com/FilipMKgit/Margin-Jinn

Poster:

github.com/FilipMKgit/MDX_Project_Poster

References

1. Cuzick & Sasieni, 2022. , doi: 10.1038/s41416-022-01937-w↵
2. Sandie et al., 2022, doi: 10.1186/s13063-022-06118-x↵
3. FDA, 2022. , 2016 https://www.fda.gov/media/78504/download↵
4. FDA, 2022. , 2016 https://www.fda.gov/media/78504/download↵
5. Werk et al., 2008. , doi: 10.1161/CIRCULATIONAHA.107.735985↵
6. Tepe et al., 2008. , doi: 10.1056/NEJMoa0706356↵