Boehringer Ingelheim

Can NI-margin calibration safeguard against invalid results in an evolving population?

“A silent killer of head-to-head trials”

Nuala Peter — Boehringer Ingelheim, Global Biostatistics & Data Sciences

15th June 2026 | PSI Conference 2026, Belfast

Life forward

Based on:

Peter N, et al. Calibrated non-inferiority margin: a new pragmatic method to account for population shift in stroke trials. European Stroke Journal 2026;11(1)

What I'll cover

1. **The problem: why a fixed margin can quietly become invalid**
2. How margins work today — and where existing fixes fall short
3. **Our concept: subgroup, reweight, calibrate**
4. A worked example: TASTE study (tenecteplase vs alteplase)
5. **What it means — and a free calculator you can download**

Boehringer Ingelheim

Quick poll -
How do you set a non-inferiority (NI) margin in a head-to-head trial?

A. Expert / clinical consensus

B. Fixed 95–50-95 rule on historical data

C. A method that adjusts for population change

D. I've never had to set one

Boehringer Ingelheim

The problem: populations don't stand still

- Head-to-head trials test a new treatment against an active standard
- The NI margin is justified using historical data — when the standard beat placebo
- **But clinical practice evolves: who we treat, and how fast, changes over time**
  - **Even if each subgroup's effect is unchanged, a shift in the MIX changes the overall effect**
- **Result: a margin imported from the past may no longer be valid for today's patients**

Boehringer Ingelheim

When does a margin become invalid?

- M1 = the full benefit of the standard vs control in the past (conservative estimate)
- M2 = the fraction of that benefit we insist on preserving (often 50%) → the value of the margin (the actual margin is '– M2')
- **Both rest on the constancy assumption: the historical effect still holds today**
- **If the population shifted on an effect-modifying covariate, constancy can break**
- **Then M1 is wrong → the margin is wrong → the trial's conclusion can be invalid**

Boehringer Ingelheim

Existing fixes have limitations

- Most methods need individual patient data from the historical RCTs
  - Data sharing is blocked by legal and administrative barriers
- Others rest on strong statistical assumptions
  - …that are often impossible to verify
- **Trialists are repeatedly told to adjust margins - with no practical way to do it**

Our concept: subgroup, reweight, calibrate

- **Split both datasets into homogeneous subgroups by effect-modifying covariates**
- Estimate the standard-vs-control effect within each subgroup (historical data)
- **Reweight those effects to the NEW trial's patient mix → a calibrated M1**
- Derive the M2 from the reweighted effect with simple maths
- NIM = -M2
- **Only subgroup estimates are needed — not full patient-level outcome data**

Boehringer Ingelheim

A fully pre-specifiable, four-step procedure

1

Create subgroups

by combining effect-modifying covariates (≥ 15 obs & ≥ 1 event per arm)

2

Estimate effects

risk difference + SE within each historical subgroup

3

Determine weights

from the NEW trial's subgroup proportions (ignores treatment arm)

4

Pool & derive margin

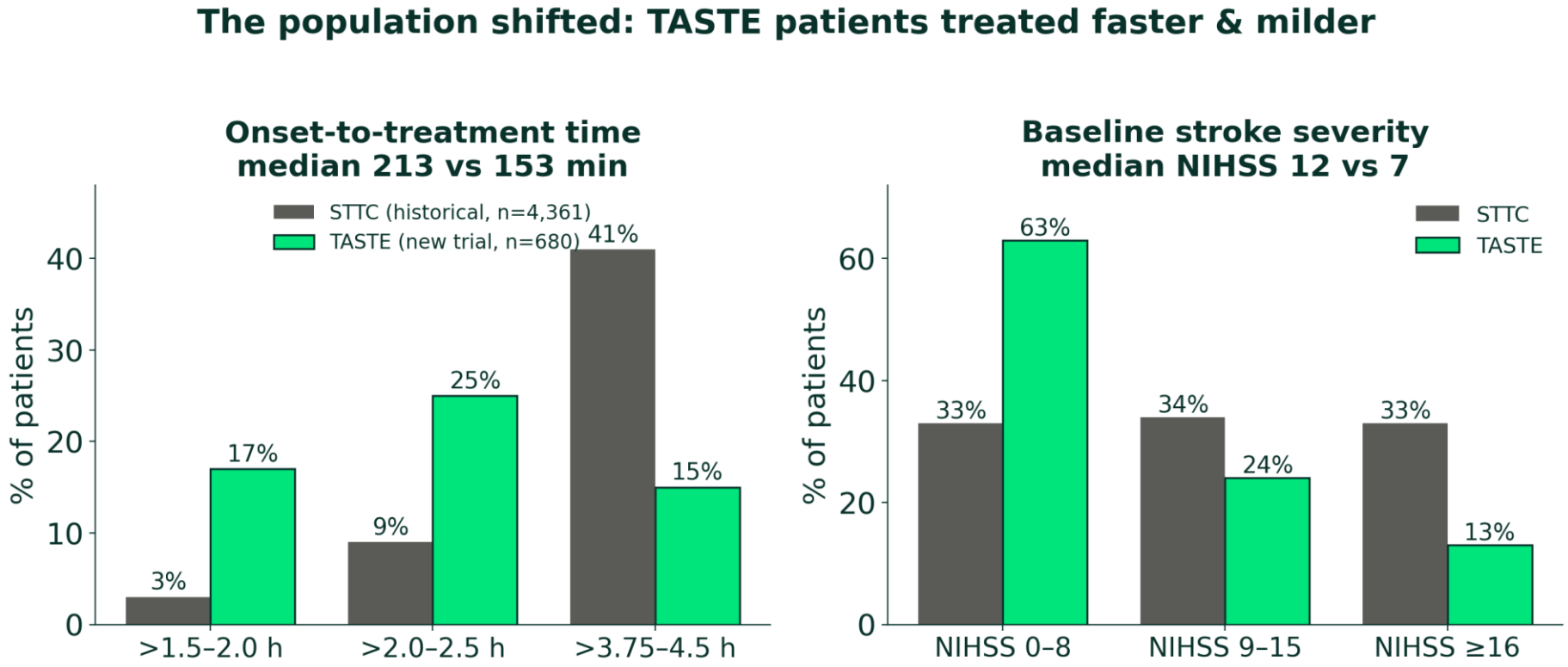
reweighted effect → 95–50–95 fixed-margin rule → NI margin

Pre-specified, blinded, and independent of outcomes → no impact on type-I error

Boehringer Ingelheim

Worked example: TASTE (2024) vs historical STTC (2014)

TASTE (tenecteplase vs alteplase, n=680) treated patients faster & milder than STTC (alteplase vs control, n=4,361)

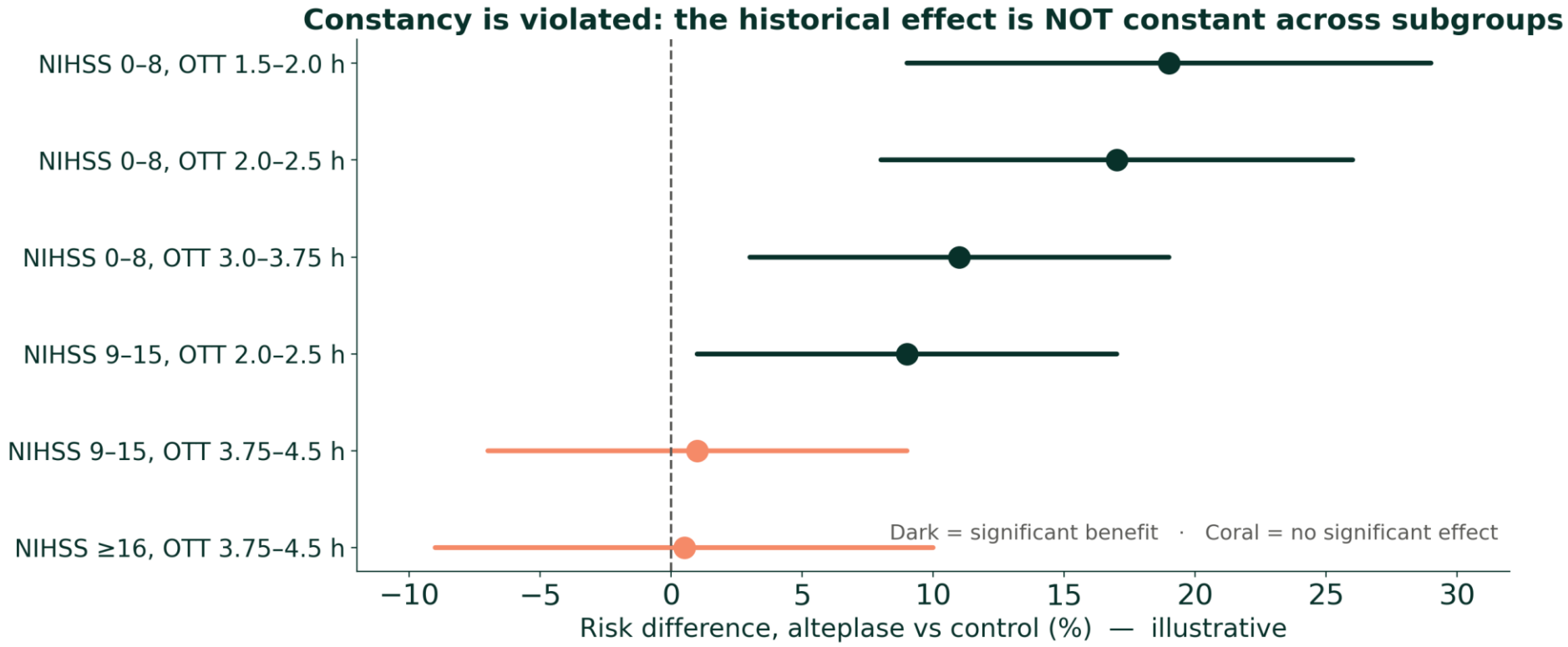


Source: Peter et al., Eur Stroke J 2026 (Table 1).

Boehringer Ingelheim

Constancy is violated

The historical alteplase effect is large in mild/fast patients, absent in severe/late — so the mix matters

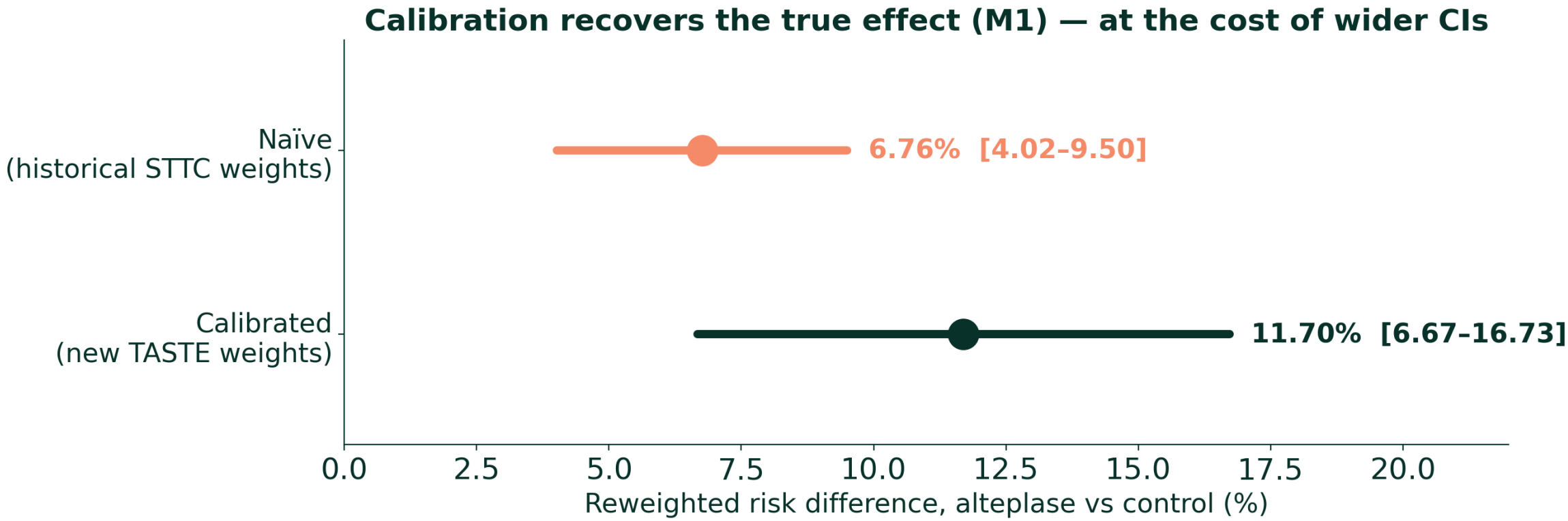


Illustrative, based on the subgroup pattern reported in Peter et al. 2026 (Fig 4).

Boehringer Ingelheim

Calibration recovers the true effect

Reweighting to the new population ~1.5× the effect (M1: 6.76% → 11.70%) — at the cost of wider CIs

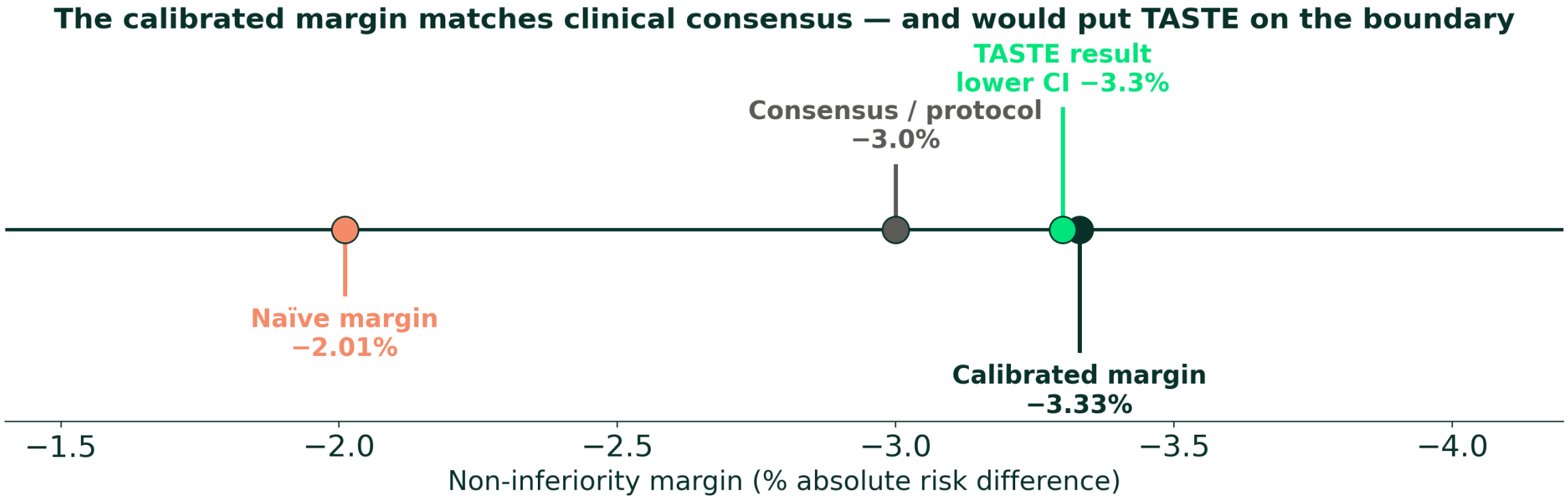


Source: Peter et al., Eur Stroke J 2026 (reweighted risk difference, alteplase vs control).

Boehringer Ingelheim

A credible, calibrated margin

Calibrated −3.33% matches clinical consensus (−3%); TASTE's result lands on the boundary



Source: Peter et al., Eur Stroke J 2026; consensus margin from ESO (Alamowitch et al. 2023).

Boehringer Ingelheim

Why this is practical, not just theoretical

- **Fully pre-specifiable in the protocol — no impact on type-I error**
- Needs only subgroup estimates — no historical patient-level outcomes
- **Comes with a free, downloadable Excel calculator**
- Robust across methods in sensitivity analyses (see back-up)
- **Already used to re-analyse AcT → EU approval of tenecteplase 25 mg**

Honest limitations

- Sample-size planning is harder — the margin isn't known until the trial starts
  - Use a plausible range, or pair with adaptive re-estimation
- Care with non-collapsible measures (odds / hazard ratios) — work on the log scale
- Developed for 1:1 allocation; uneven ratios need further work
- **Only needed when an effect-modifying covariate has actually shifted**

Take-home messages

- **Population shift can quietly invalidate a fixed NI margin**
- **Subgroup + reweight + calibrate fixes it — simply and transparently**
- Pre-specifiable, assumption-light, and already proven in regulatory practice
- *Try the calculator on your own trial*

Boehringer Ingelheim

Thank you — questions welcome

Nuala Peter · Boehringer Ingelheim, Global Biostatistics & Data Sciences

(nuala.peter@boehringer-ingelheim.com)

Read the paper & download the free calculator:

- Peter et al., European Stroke Journal 2026;11(1):aakaf022 — open access
- doi:10.1093/esj/aakaf022

Boehringer Ingelheim

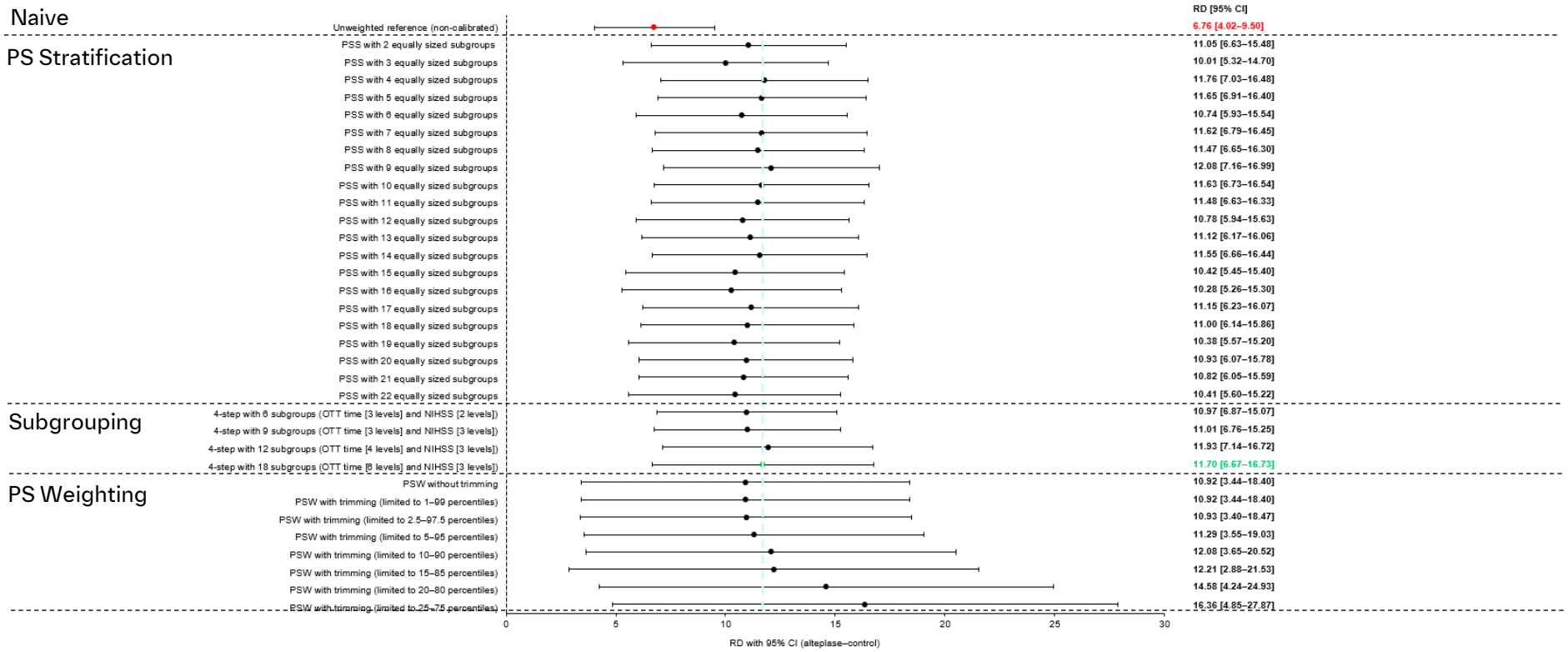
Back-up

Sensitivity analyses, definitions & references

Boehringer Ingelheim

Back-up: robustness across methods

Calibrated effect is consistent across PSS, PSW and the 4-step method; PSW inflates variance most



Source: Response to reviewers, Fig 1 (STTC calibrated for TASTE). Red = non-calibrated, green = calibrated 4-step.

Boehringer Ingelheim

Back-up: M1, M2 and the 95–50–95 rule

- M1 — conservative estimate of the standard's benefit (lower 95% bound of reweighted RD)
  - **Calibrated: M1 = 6.67% (vs naïve 4.02%)**
- M2 — acceptable reduction; here 50% of M1
  - **Calibrated: M2 = 3.33% → margin −3.33% (vs naïve −2.01%)**
- *100% preserved = superiority test in head-to-head trial; 0% = superiority over control*

Boehringer Ingelheim

Back-up: key references

- Peter N, et al. Calibrated non-inferiority margin… Eur Stroke J 2026;11(1):aakaf022
- Parsons MW, et al. TASTE trial. Lancet Neurol 2024;23:775–786
- Emberson J, et al. STTC IPD meta-analysis. Lancet 2014;384:1929–1935
- Schumi J, Wittes JT. Understanding non-inferiority. Trials 2011;12:106
- Alamowitch S, et al. ESO tenecteplase recommendation. Eur Stroke J 2023;8:8–54
- Peter N, et al. AcT regulatory re-analysis. SVIN 2025;5:e001705