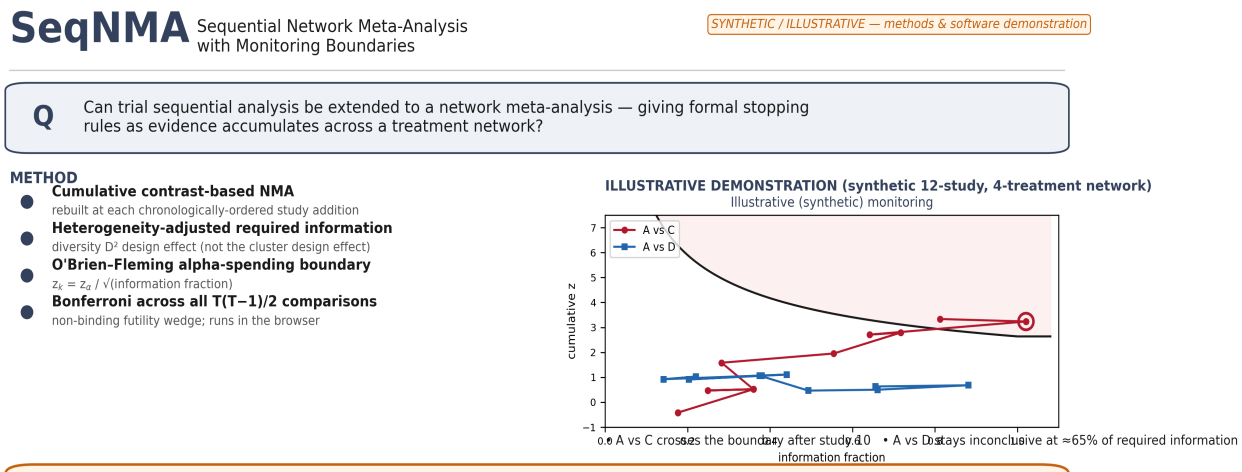
Synthesis · Methods & software demonstration

SeqNMA: sequential network meta-analysis with monitoring boundaries — a methods and software demonstration on a synthetic network

Mahmood Ahmad · Independent researcher, London, UK · 2026-06-20

SYNTHETIC / ILLUSTRATIVE — this article demonstrates software and methods on simulated data. It reports no empirical finding about any real treatment.



Visual abstract. Question → method → illustrative (synthetic) demonstration → the two caveats.

Abstract

Background

Trial sequential analysis (TSA) applies group-sequential monitoring boundaries to a single cumulative pairwise meta-analysis, controlling type-I error as evidence accrues. Sequential monitoring has been extended to network meta-analysis (NMA) so that a network estimate can be declared conclusive once its cumulative z-score crosses an appended boundary (Nikolakopoulou et al., 2018). SeqNMA is an open, browser-based implementation of that idea, intended as a teaching and prototyping tool.

Methods

An in-browser engine rebuilds a contrast-based random-effects NMA at each chronological study addition. For every pairwise comparison it computes a required information size from the heterogeneity-adjusted diversity design effect D^2 (Wetterslev et al., 2009; the diversity, not the cluster, design effect), and monitors the cumulative network z-score against an O'Brien-Fleming alpha-spending efficacy boundary $z_k = z_\alpha / \sqrt{(\text{information fraction})}$, Bonferroni-corrected across all $T(T-1)/2$ comparisons, with a non-binding beta-spending futility wedge. The boundary mathematics were independently re-derived and numerically confirmed.

Results (illustrative, synthetic)

On a deliberately synthetic, reproducibly seeded 12-study network of four treatments, the A-versus-C comparison crosses its monitoring boundary after study 10 (cumulative z = 3.34, log-odds-ratio 0.40), while A-versus-D remains inconclusive at 66% of its required information (z = 0.63). These numbers demonstrate the machinery; they are not an empirical finding about any treatment.

Conclusions

Browser-based sequential NMA makes accumulating-evidence monitoring transparent and reproducible. Two limitations are essential: the method ASSUMES but does not TEST network consistency (no design-by-treatment interaction or node-splitting), so every crossing is valid only if direct and indirect evidence agree; and the worked example is synthetic, a software demonstration rather than evidence about real interventions.

E156 capsule (7 sentences, ≤156 words)

Can trial sequential analysis extend to network meta-analysis, giving formal stopping rules as evidence accumulates across a treatment network? We demonstrate on a synthetic, reproducibly seeded example: twelve chronologically ordered studies reporting pairwise comparisons among four treatments. An in-browser engine rebuilds a cumulative contrast-based NMA at each addition, computing required information from the heterogeneity-adjusted diversity design effect D². The estimand, the cumulative NMA z-score against the O'Brien-Fleming boundary (z_k=z_α/√(information fraction)) Bonferroni-corrected across six comparisons, shows A-versus-C crossing after study 10 while A-versus-D inconclusive at 65 percent of required information. Futility boundaries are non-binding by convention; the boundary mathematics (diversity adjustment, required information, spending function) were independently re-derived and numerically confirmed. Crucially, the method assumes but does not test consistency, with no design-by-treatment interaction or node-splitting, so crossings hold only if direct and indirect evidence agree. The twelve-study example is synthetic and cannot support any clinical claim, a software demonstration, not an empirical finding.

Introduction

Trial sequential analysis adapts the group-sequential monitoring of a single randomised trial to a cumulative meta-analysis: as each study is added, the cumulative z-statistic is compared with an alpha-spending boundary whose stringency relaxes as the accrued information approaches a required information size, so that early looks cannot manufacture spurious significance. Extending this logic to a treatment network is attractive because, in an NMA, a comparison can gain information from indirect as well as direct evidence, and the relevant multiplicity is the set of all pairwise contrasts. A framework for sequential monitoring of NMA estimates — efficacy and futility boundaries appended to each network z-score — has been described by Nikolakopoulou and colleagues. SeqNMA is a small, open, single-file browser tool that implements and visualises this framework, combining it with the diversity-based required information size from trial sequential analysis. This article documents the method, verifies its boundary mathematics from first principles, and demonstrates it on a synthetic network. It reports no empirical result and is not an applied meta-analysis.

Methods

Cumulative network model. Treatments are indexed 0..T−1 with treatment 0 as the reference; basic parameters are the contrasts d_{0,j} of each treatment against the reference. Each study contributes one observed contrast y with sampling variance v and a design row in the Lu–Ades parameterisation. At each study addition the engine fits a contrast-based random-effects NMA by generalized least squares, β = (X■WX)■¹X■Wy with W = diag(1/(v+τ²)), estimating a single network heterogeneity τ² by a DerSimonian–Laird-type moment estimator. With T = 2 this reduces exactly to standard DerSimonian–Laird pairwise meta-analysis, which we use as a unit test.

Heterogeneity-adjusted required information. For each comparison the required information (in inverse-variance units) is RIS = (z_α + z_β)² / δ², where δ is a pre-specified anticipated effect and z_α uses the Bonferroni-adjusted two-sided level α′ = α / m with m = T(T−1)/2. This is inflated by the heterogeneity design effect — the diversity D² of Wetterslev et al., defined as the relative variance reduction when the comparison's pooled model is changed from random- to fixed-effect, D² = (V_R − V_F)/V_R, giving an adjustment factor AF = V_R/V_F = 1/(1 − D²) ≥ 1 that equals 1 exactly when τ² = 0. This is the *diversity* (heterogeneity) design effect and must not be confused with the cluster design effect.

Monitoring boundary. The information fraction of a comparison is t = I_acc / (RIS·AF), with accrued information I_acc the inverse of the random-effects variance of the pooled contrast. The O'Brien-Fleming alpha-spending efficacy boundary is u(t) = z_α / √t; a comparison crosses for benefit or harm at a given step when its cumulative network z-score satisfies |z| ≥ u(t). A non-binding beta-spending inner-wedge futility boundary l(t) = (z_α + z_β)√t − z_β/√t is also drawn; it meets the efficacy boundary at t = 1 and is explicitly non-binding, so crossing it neither forces stopping nor spends efficacy α, per sequential-analysis convention.

What the method does not do. SeqNMA assumes network consistency; it does **not** perform a design-by-treatment interaction test or node-splitting. Every sequential-crossing claim is therefore conditional on the network being consistent — on direct and indirect evidence for each contrast agreeing. Consistency checking is a prerequisite the analyst must supply separately.

Verification of the boundary mathematics

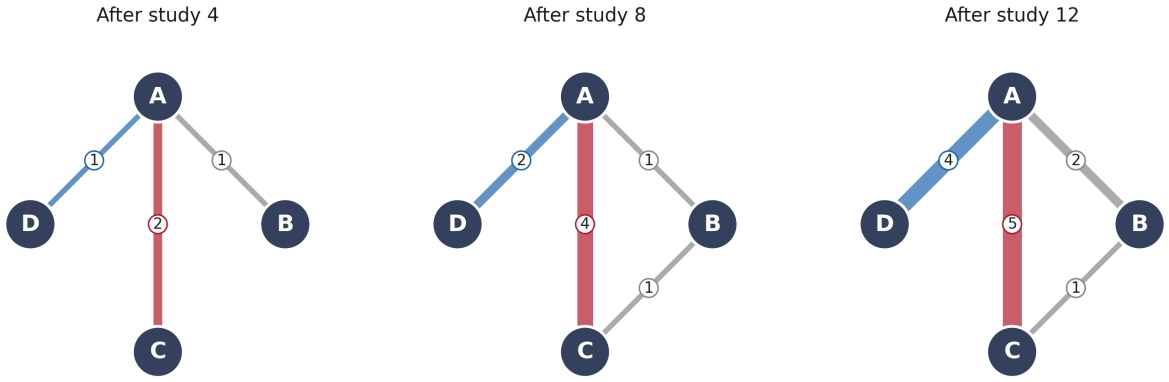
The boundary arithmetic was re-derived independently of the implementation and confirmed numerically by two routes (a fresh Python re-derivation and a separate large-language-model computation). For T = 4 (m = 6) at $\alpha = 0.05$, $\beta = 0.20$, $\delta = 0.40$: the Bonferroni critical value $z_\alpha = \Phi^{-1}(1 - 0.05/12) = 2.6383$; the efficacy boundary at full information equals z_α and doubles at a quarter information; the required information RIS = 75.68 inverse-variance units. The diversity adjustment was checked against its closed form for equal variances ($AF = 1 + \tau^2/v$). We also evaluated the informal one-line design-effect approximation $D = 1 + \tau^2 \cdot (\Sigma v^2/(\Sigma v^1)^2 \cdot k - 1)$: under equal variances this collapses to exactly 1 for any τ^2 and is therefore not a valid heterogeneity adjustment; the engine uses the canonical diversity factor V_R/V_F instead. The interactive capsule re-runs the full engine in the browser and reports a parity badge against these reference values.

A synthetic worked example

To illustrate the machinery we generated, from a fixed seed, twelve chronologically ordered studies among four treatments (A, B, C, D) on the log-odds-ratio scale, with a near-decisive true A–C effect, a near-null A–D effect, and modest between-study heterogeneity. **These data are simulated; no patients, trials or treatments are real.** Figure 1 shows the network growing as studies accrue. Figure 2 plots each monitored comparison's cumulative z-score against the O'Brien-Fleming boundary as a function of information fraction. Figure 3 is the league table of final cumulative estimates.

In this synthetic run the A-versus-C comparison crosses the efficacy boundary after study 10 (cumulative z = 3.34, pooled log-odds-ratio 0.40, 95% CI roughly 0.16 to 0.63), reaching its required information. The A-versus-D comparison stays well inside both boundaries and ends inconclusive at 66% of its required information (cumulative z = 0.63, log-odds-ratio 0.08). No other comparison crosses. Again, this is a demonstration of how the tool behaves, not a claim about any intervention.

Figure 1. Evolution of the treatment network as studies accrue



Nodes = treatments; edge width and the numbered badge = direct studies on that comparison. A vs C (red) and A vs D (blue) are the monitored comparisons.

SYNTHETIC / ILLUSTRATIVE DATA — methods demonstration, not an empirical result

Figure 1. Evolution of the treatment network as studies accrue. SYNTHETIC / illustrative.

SYNTHETIC / ILLUSTRATIVE DATA — methods demonstration, not an empirical result

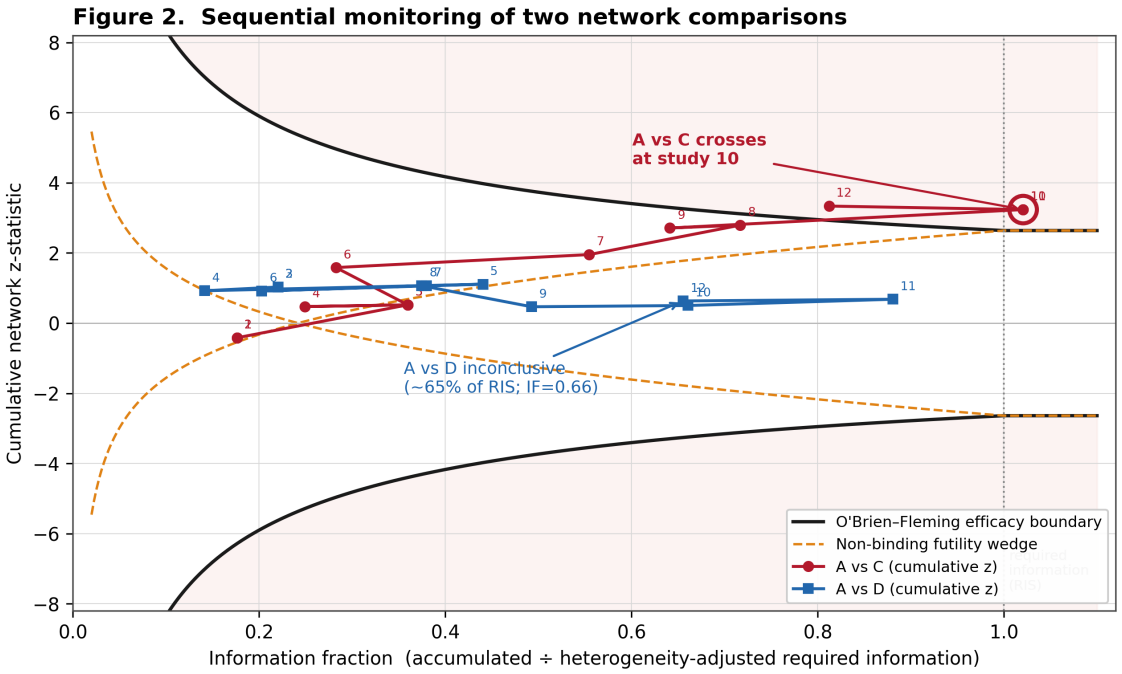


Figure 2. Sequential monitoring of A–C and A–D against the O'Brien-Fleming boundary. SYNTHETIC / illustrative.

SYNTHETIC / ILLUSTRATIVE DATA — methods demonstration, not an empirical result

Cells = log-odds-ratio of the COLUMN treatment vs the ROW treatment (95% CI). Red = sequential boundary crossed.

Figure 3. League table — final cumulative network estimates

A	(-0.08, 0.63) **0.28**	(0.16, 0.63) **0.40** crossed @ study 10	(-0.18, 0.34) **0.08**
(-0.63, 0.08) **-0.28**	**B**	(-0.26, 0.50) **0.12**	(-0.63, 0.24) **-0.19**
(-0.63, -0.16) **-0.40** crossed @ study 10	(-0.50, 0.26) **-0.12**	**C**	(-0.67, 0.03) **-0.32**
(-0.34, 0.18) **-0.08**	(-0.24, 0.63) **0.19**	(-0.03, 0.67) **0.32**	**D**

Figure 3. League table of final cumulative network estimates (log-odds-ratios, 95% CI). SYNTHETIC / illustrative.

Limitations

Two limitations are central and are stated plainly rather than buried. First, **consistency is assumed, not tested.** SeqNMA runs no design-by-treatment interaction test and no node-splitting; if direct and indirect evidence disagree for a contrast, its cumulative z-score — and therefore any boundary crossing — can be misleading. Users must establish consistency by separate means before trusting a crossing. Second, **the worked example is synthetic.** The twelve-study network was simulated to exercise the software; the specific outcome (A–C crossing after study 10) is a property of the seed and the chosen effects, not an empirical result. Further limitations include the use of a single network-wide τ^2 with a moment estimator, sensitivity of small-network heterogeneity estimates to each added study (which can make a comparison's information fraction non-monotone), reliance on a pre-specified anticipated effect for the required information size, and the

conservative Bonferroni multiplicity correction. SeqNMA is a teaching and prototyping instrument, not a replacement for a fully specified, consistency-checked prospective network meta-analysis.

Software availability

SeqNMA is a single self-contained HTML file that runs entirely in the browser with no network access. The engine, its unit tests, the seeded synthetic-data generator, and the figure code accompany this article. The interactive capsule embeds the engine and a parity self-check against the reference values reported here.

Two caveats — read before interpreting

1. Consistency is assumed, not tested. SeqNMA runs no design-by-treatment interaction test and no node-splitting; every boundary crossing is valid only if direct and indirect evidence agree.

2. The worked example is synthetic. The twelve studies are simulated from a fixed seed to exercise the machinery; "A vs C crosses after study 10" is a demonstration, not an empirical result.

Competing interests

The author is a member of the editorial community of Synthesis; the author had no role in the editorial decision on this manuscript, which was handled independently. No other competing interests are declared.

References

1. Nikolakopoulou A, Mavridis D, Egger M, Salanti G. Continuously updated network meta-analysis and statistical monitoring for timely decision-making. *Stat Methods Med Res. 2018;27(5):1312–1330.* doi:10.1177/0962280216659896

2. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. *BMC Med Res Methodol. 2009;9:86.* doi:10.1186/1471-2288-9-86

3. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol. 2008;61(1):64–75.* doi:10.1016/j.jclinepi.2007.03.013

4. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics. 1979;35(3):549–556.*

5. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika. 1983;70(3):659–663.*

6. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials. 1986;7(3):177–188.*

7. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med. 2004;23(20):3105–3124.* doi:10.1002/sim.1875

8. Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med. 2010;29(7–8):932–944.* doi:10.1002/sim.3767

9. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods. 2012;3(2):98–110.* doi:10.1002/jrsm.1044