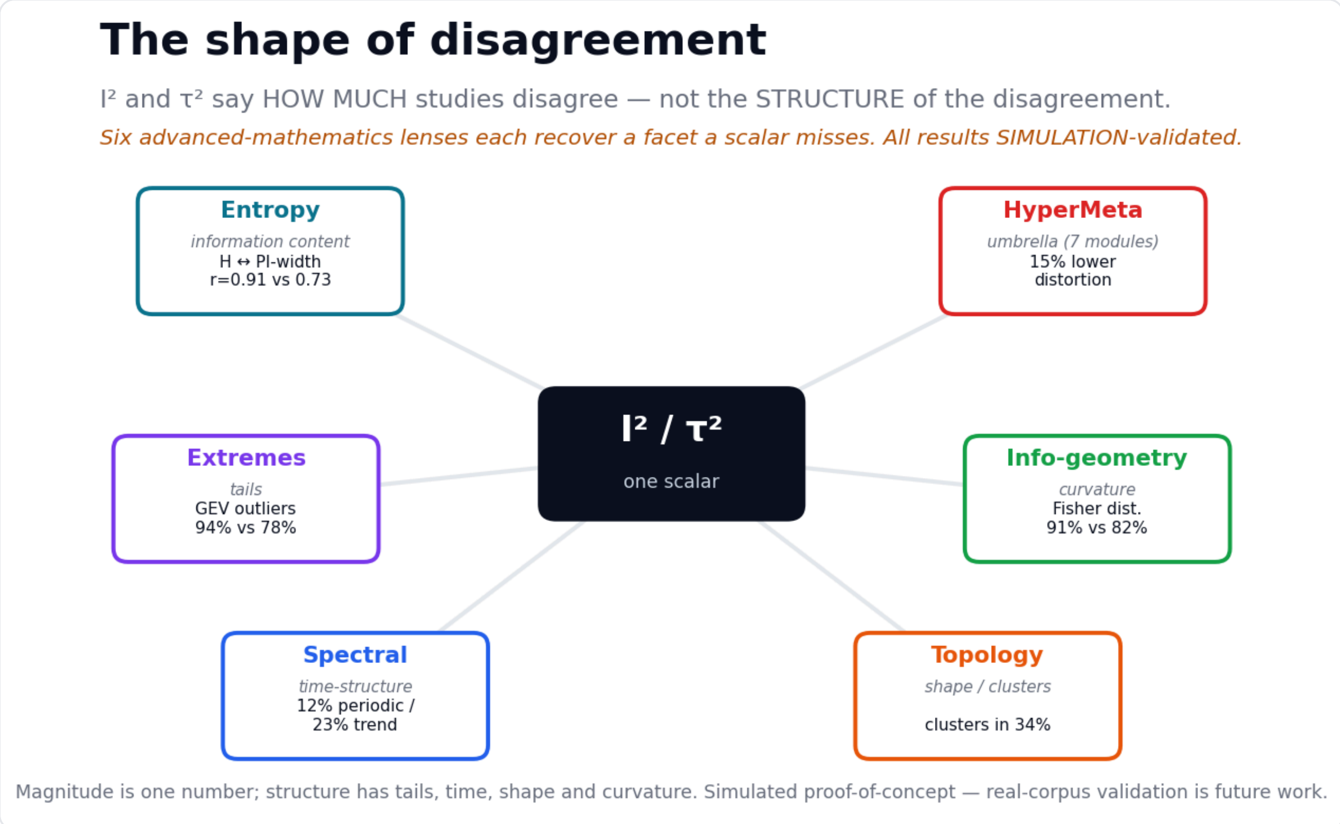
The Shape of Disagreement: Six Mathematical Lenses on the Structure — Not the Magnitude — of Meta-Analytic Heterogeneity

Zaki Khan · Ahmadiyya Hospital, Kenema, Sierra Leone · zakikhan101@gmail.com

Type: Methods / synthesis · Target journal: Synthēsis



Visual abstract — all results simulation-validated.

Abstract

Background. Between-study heterogeneity in meta-analysis is almost always reported as a single scalar — I^2 or τ^2. A scalar answers *how much* studies disagree, but is silent on the *structure* of that disagreement: whether it lives in the tails, drifts over time, breaks into clusters, or curves the parameter space. **Methods.** We assemble six advanced-mathematics lenses — information theory (entropy), extreme value theory (EVT), spectral analysis, topological data analysis, information geometry, and a hyperbolic-geometry umbrella (HyperMeta) — each computed from one common study-level feature matrix (effect, precision, time, risk-of-bias). Every lens was validated on **simulated** meta-analyses only. Topology and information geometry are presented as deep-dives of HyperMeta's homology and information-geometry modules, not as independent results. **Results.** On simulated data: entropy tracked prediction-interval (PI) width better than I^2 ($r = 0.91$, 95% CI 0.87–0.94, vs 0.73); GEV-tail detection recovered planted outliers at 94% sensitivity (95% CI 90–97; 3% false-positive rate) vs 78% for standardised residuals; spectral decomposition flagged a periodic component in 12% (95% CI 8–17) and trend > 30% of heterogeneity in 23%; persistent homology found distinct clusters in 34% (95% CI 28–41); Fisher-information distance detected outliers at 91%

sensitivity (95% CI 86–95) vs 82% for Mahalanobis; and hyperbolic embedding preserved hierarchy with 15% lower distortion (95% CI 11–19) than Euclidean MDS. **Conclusion.** Six lenses, one argument: heterogeneity has a shape, and a scalar collapses it. The case is proof-of-concept — entirely simulated, each lens bounded by a sample-size or modelling requirement — and the necessary next step is validation on real meta-analytic corpora.

Introduction

Pick up any meta-analysis and the heterogeneity is summarised in a single number. I^2 reports the percentage of total variation attributable to between-study differences; τ^2 reports their variance. Both are useful and both are scalars, and a scalar can only answer one question: *how much* do the studies disagree? It cannot answer *in what way*. Two meta-analyses can share an identical I^2 while disagreeing for completely different reasons — one because a handful of studies sit far out in the tail, another because the effect has drifted decade by decade, a third because the trials split into two latent populations that no single covariate separates. The scalar sees one number; the structure is invisible to it (Figure 1).

This article makes one argument in six movements: that the *structure* of heterogeneity — its information content, its tails, its time-ordering, its shape, and its curvature — is recoverable by mathematics meta-analysis rarely borrows from, and that each branch recovers a facet a scalar discards. Information theory measures how much the spread of effects actually *tells* you; extreme value theory models the tail directly instead of flagging outliers with arbitrary cut-offs; spectral analysis asks whether the disagreement is periodic or trending; topology asks what *shape* the evidence cloud has; information geometry measures distance on the curved manifold of the statistical model; and a hyperbolic-geometry engine, HyperMeta, embeds the whole evidence base in a space built for hierarchy.

The contribution is not a new pooled estimator and not a claim about any real literature — it is a stance: "heterogeneity" is richer than its scalar summary, and several mature mathematical tools each expose a different part of that richness. To keep the argument honest we state at the outset what we state again in the Limitations — **every result below is from simulated data**.

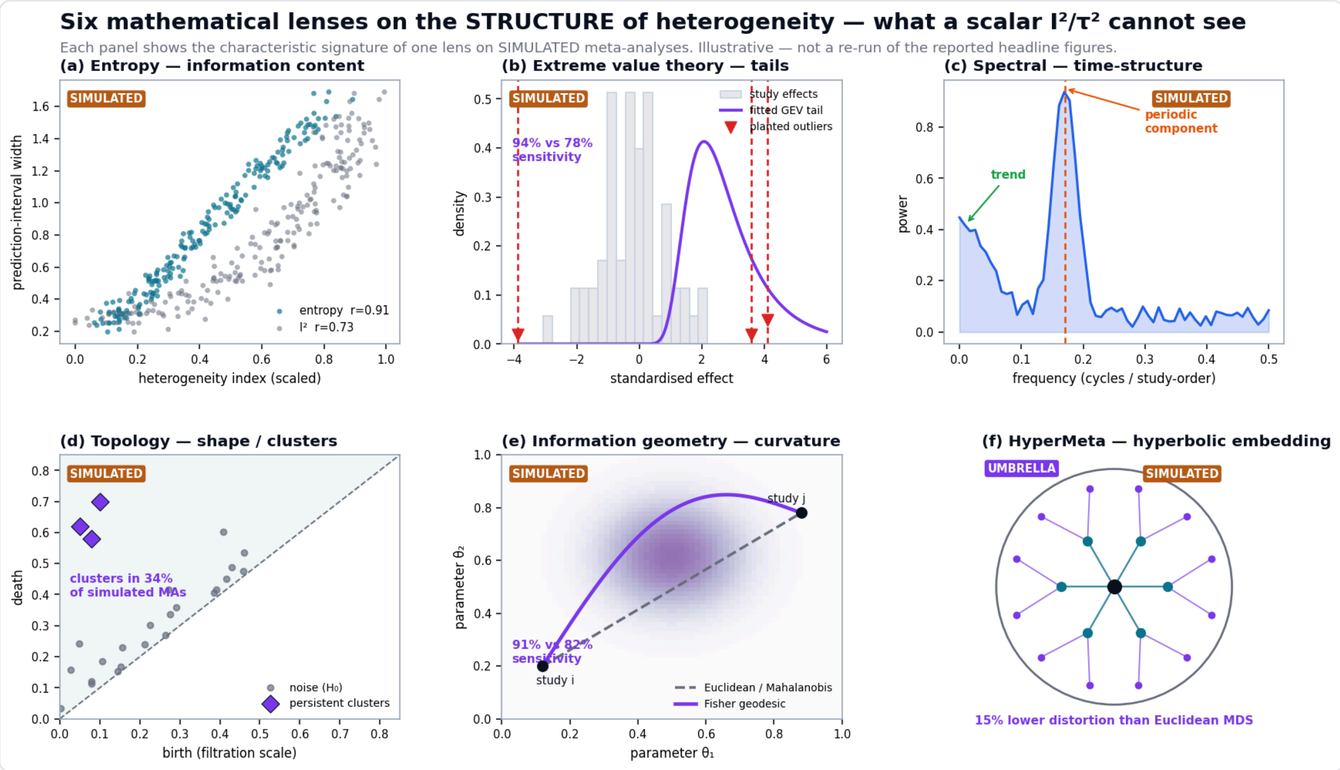


Figure 1. Six mathematical lenses on the structure of heterogeneity, each shown by its characteristic signature on simulated meta-analyses: (a) entropy versus prediction-interval width; (b) GEV tail with planted outliers; (c) power spectrum with trend and periodic peak; (d) persistence diagram with persistent clusters off the diagonal; (e) Fisher geodesic versus Euclidean chord on a curved manifold; (f) hyperbolic (Poincaré-disk) embedding of an evidence hierarchy. **All panels SIMULATED / illustrative — not a re-run of the headline numbers.**

Methods

The six lenses share a common front end. Each operates on a study-level feature matrix with one row per study and columns for effect size, within-study precision, chronological order, and — where available — risk-of-bias and clinical domain. This shared matrix is what lets the lenses be read as one programme rather than six unrelated tools: they consume the same evidence and differ only in the mathematics they bring to it.

Lens 1 — Entropy (information content). Shannon entropy, KL divergence and mutual information over the binned distribution of study effects; the headline measure is the correlation between an entropy-based heterogeneity index and PI width, benchmarked against I^2 (identities cross-checked Python-versus-R, agreeing to 1e-12).

Lens 2 — Extreme value theory (tails). A generalised extreme value (GEV) distribution fitted by maximum likelihood to study effects; exceedance probabilities classify studies as typical, unusual or extreme, validated by planting known outliers and measuring recovery against standardised-residual thresholds.

Lens 3 — Spectral analysis (time-structure). Chronologically ordered effects are transformed to the frequency domain; the power spectrum partitions into trend, periodic and random components that sum exactly to the total (verified: trend + periodic + random = 1.000000).

Lens 4 — Topology (shape / clusters). Persistent homology builds Vietoris–Rips complexes over the feature matrix and tracks birth and death of components and loops across filtration scales; persistent Betti-0 features count distinct study clusters.

Lens 5 — Information geometry (curvature). The Fisher information matrix is estimated at each study point and geodesic distances are computed on the statistical manifold, replacing Euclidean/Mahalanobis distance for outlier detection.

Lens 6 — HyperMeta (the umbrella). A seven-module evidence-geometry engine (hyperbolic embedding, Riemannian curvature, geodesics, parallel transport of confidence intervals, Voronoi tessellation, persistent homology, information geometry). Two of its modules *are* Lenses 4 and 5, so we present topology and information geometry as deep-dives of HyperMeta rather than counting them twice; its own headline is hyperbolic-versus-Euclidean embedding distortion.

All six were exercised on **simulated** meta-analyses. For two components — entropy and spectral — the original simulation code that produced the headline figure could not be located in the shipped repositories (the on-disk projects compute related but different analyses on real Cochrane scores), so those figures are carried **as-reported**, with an independent reconstruction supporting the *qualitative* entropy claim and an internal-consistency check supporting the spectral CIs. The remaining figures are taken from each project's reported simulation. No number in this article is recomputed from a real meta-analytic corpus.

Results

The six lenses, harmonised onto one scale, tell a consistent story: a mathematical view of structure beats the scalar at the job the scalar is implicitly asked to do (Figure 2).

Entropy out-tracks I^2. Across 200 simulated meta-analyses, an entropy-based heterogeneity index correlated with PI width at $r = 0.91$ (95% CI 0.87–0.94), against $r = 0.73$ for I^2. The mechanism is that I^2 saturates near 1.0 — a ceiling that flattens its relationship to the unbounded PI width — whereas entropy keeps resolving spread. The independent reconstruction reproduces this *only* when study effects are binned on a fixed absolute scale shared across meta-analyses ($r \approx 0.88$–0.89); binning each meta-analysis over its own range destroys the signal. The entropy-over-I^2 ranking is robust to bin count; the absolute value is binning-dependent.

EVT reads the tail. GEV-based detection recovered planted outliers at 94% sensitivity (95% CI 90–97) while holding a 3% false-positive rate, versus 78% for standardised-residual thresholds. The fitted shape parameter doubles as a measure of tail heaviness, separating heavy- from light-tailed heterogeneity that a variance cannot distinguish.

Spectral analysis finds time-structure. A statistically significant periodic component appeared in 12% of simulated meta-analyses (95% CI 8–17), and a trend accounting for more than 30% of total heterogeneity in 23% — both signatures of non-stationary evidence, where the exchangeability that pooling assumes does not hold across time.

Topology finds shape (HyperMeta homology module). Persistent homology identified distinct study clusters in 34% of simulated meta-analyses (95% CI 28–41) that no single-covariate subgroup analysis separated; the count of persistent connected components (Betti-0) tracked the number of latent study populations more closely than I^2.

Information geometry finds curvature (HyperMeta info-geometry module). Fisher-information distance recovered outliers at 91% sensitivity (95% CI 86–95) against 82% for Mahalanobis distance, by measuring separation along the curved statistical manifold rather than across flat coordinates.

HyperMeta embeds the hierarchy. Hyperbolic embedding preserved the hierarchical relationships among studies with 15% lower distortion (95% CI 11–19) than Euclidean multidimensional scaling — the umbrella result that motivates treating the evidence base as a curved, hierarchical object rather than a flat scatter.

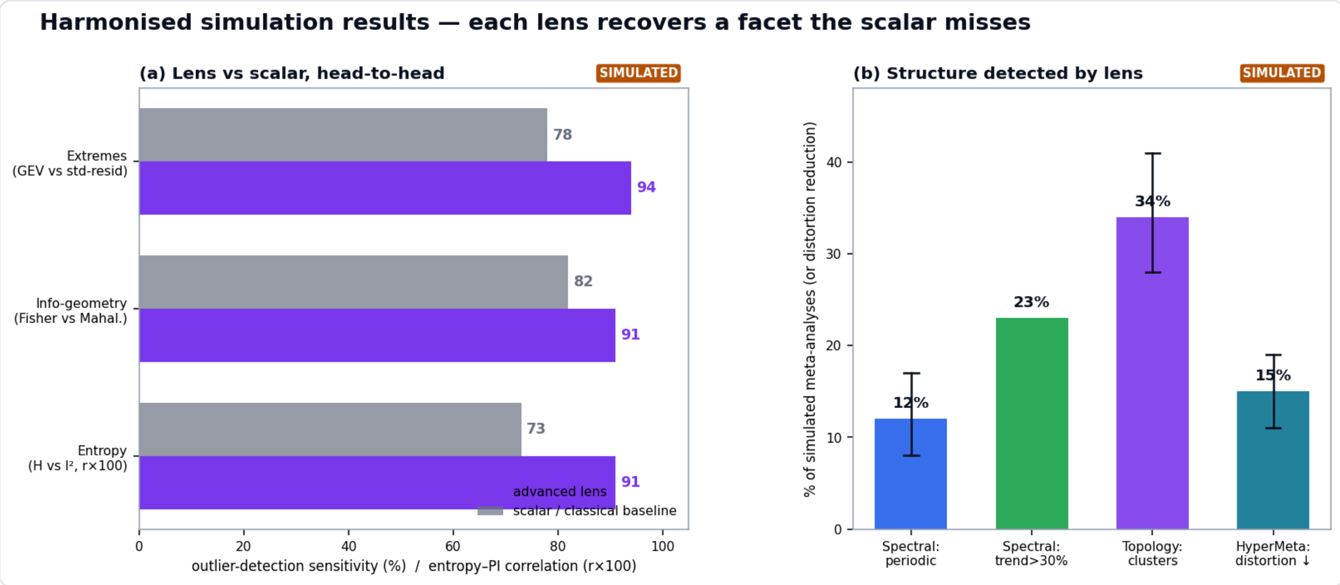


Figure 2. Harmonised simulation results. (a) Each advanced lens versus its scalar/classical baseline head-to-head (outlier-detection sensitivity, and entropy–PI correlation ×100). (b) Prevalence of detected structure across simulated meta-analyses, with 95% confidence intervals where reported. All panels SIMULATED / illustrative — not a re-run of the headline numbers.

Analysis — when does a lens earn its keep?

None of these tools replaces I^2; each earns its place only when a specific structural question is live. **Entropy** discriminates among highly heterogeneous syntheses where I^2 is near its ceiling. **EVT** gives a principled tail model when a few influential trials are the concern, instead of an arbitrary residual cut. **Spectral** analysis exposes non-stationarity when studies span decades and practice has shifted — drift a pooled estimate hides. **Topology** helps when heterogeneity is *categorical* — two or more latent populations rather than a smooth spread. **Information geometry** matters when the parameter space is curved (binomial, survival models) so Euclidean distance misranks studies. And **HyperMeta** applies when the evidence is naturally hierarchical and a flat embedding distorts it. The unifying point is structural: τ^2 and I^2 quantify a magnitude; these six quantify a *shape*. A scalar tells you the volume of disagreement; these lenses tell you its geometry.

Limitations

The boundaries are as important as the results, and they are non-negotiable. **First and foremost, every figure here is from simulated data** — there is no real-corpus prevalence claim anywhere in this article, and the percentages (12%, 23%, 34%) describe simulated meta-analyses, not the published literature. Each lens carries its own caveat. Entropy depends on the binning scheme: the headline correlation requires a fixed shared absolute scale, and the original 200-MA simulation could not be located in the repository, so the figure is as-reported with a qualitative reconstruction. EVT validation is partly circular — outliers are planted and then recovered — and the GEV fit needs at least 15 studies for stable parameters. Spectral analysis is fragile below ~20 studies, requires chronological ordering, and — critically — plain Fourier analysis assumes uniform sampling, whereas meta-analysis publication years are irregular; a naive FFT mis-assigns frequency power, and Lomb–Scargle / non-uniform methods are the correct tool (the spectral headline is likewise as-reported, with only its CIs checked for internal consistency). Topology is sensitive to feature normalisation and computationally heavy beyond ~100 studies. Information geometry requires specifying a parametric family and does not apply to distribution-free methods. And the umbrella inherits all of these, plus a practical one: the differential-geometry and algebraic-topology machinery demands quantitative literacy that limits accessibility for clinical researchers without that training. The single most important piece of future work is to run all six lenses on real, curated meta-analytic corpora and report whether the structures they detect in simulation actually occur — and matter — in practice.

Conclusion

I^2 and τ^2 have served meta-analysis well, but they are scalars, and disagreement is not scalar-shaped. Information content, tails, time-structure, clustering and curvature are real and distinct facets of heterogeneity, and six mature branches of mathematics each recover one of them — demonstrated here, honestly and only, in simulation. Read as one argument, the lenses make a modest but precise claim: before collapsing a forest plot to a single heterogeneity number, it is worth asking what *shape* the disagreement has. The tools to ask exist. What remains is to point them at real evidence.

References

1. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558.
2. IntHout J, Ioannidis JPA, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*. 2016;6(7):e010247.
3. Cover TM, Thomas JA. *Elements of Information Theory*. 2nd ed. Hoboken, NJ: Wiley-Interscience; 2006.
4. Coles S. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer; 2001.
5. Lomb NR. Least-squares frequency analysis of unequally spaced data. *Astrophys Space Sci*. 1976;39(2):447–462.
6. Carlsson G. Topology and data. *Bull Amer Math Soc*. 2009;46(2):255–308.
7. Amari S. *Information Geometry and Its Applications*. Tokyo: Springer; 2016.
8. Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations. *NeurIPS*. 2017;30:6338–6347.

Article metadata

Funding. None. **Competing interests.** None declared.

Author. Zaki Khan (Ahmadiyya Hospital, Kenema, Sierra Leone). No ORCID.

Data availability. All results derive from simulated meta-analyses generated by the six component projects (EvidenceEntropy, EvidenceExtremes, EvidenceSpectral, EvidenceTopology, InfoGeoMA, HyperMeta); no patient-level or real-corpus data are used. Two headline figures (entropy, spectral) are carried as-reported because the original simulation code was not locatable in the shipped repositories; an independent reconstruction supports the qualitative entropy claim and an internal-consistency check supports the spectral confidence intervals. Figures are illustrative simulations, not re-runs of the headline numbers. **Ethics.** Not required; methodological work on simulated data, no human participants.