

Are the Amyloid and Tau Alzheimer's Disease Literatures Equally Credible? A Variable-Power P-Curve Analysis of Bias Signatures

Michael Flanders¹, Kendra D. Sims², Audrey R. Murchland², Isabel Elaine Allen³, Michelle Caunca³, Carlos Morales Valentín⁴, Lon S. Schneider⁵, Sarah F. Ackley¹

¹ Department of Epidemiology, Brown University, Providence, RI ² Department of Epidemiology, Boston University, Boston, MA ³ University of California, San Francisco, CA ⁴ Dept. of Psychology, University of Puerto Rico Río Piedras, San Juan, PR ⁵ Keck School of Medicine of USC, Los Angeles, CA

AD Biomarker Research & Credibility

AD pathology: Amyloid- β plaques and tau tangles define Alzheimer's disease (AD) pathology.

Credibility: The amyloid field has disproportionately faced concerns about retractions, data fabrication, overinvestment, and groupthink.

Reporting bias: Beyond individual cases of misconduct, we examine field-wide patterns consistent with p-hacking and publication bias.

Aducanumab Approval: The 2021 FDA approval provides a cut point for comparing earlier vs. recent reporting patterns.

Is Amyloid More Biased than Tau?

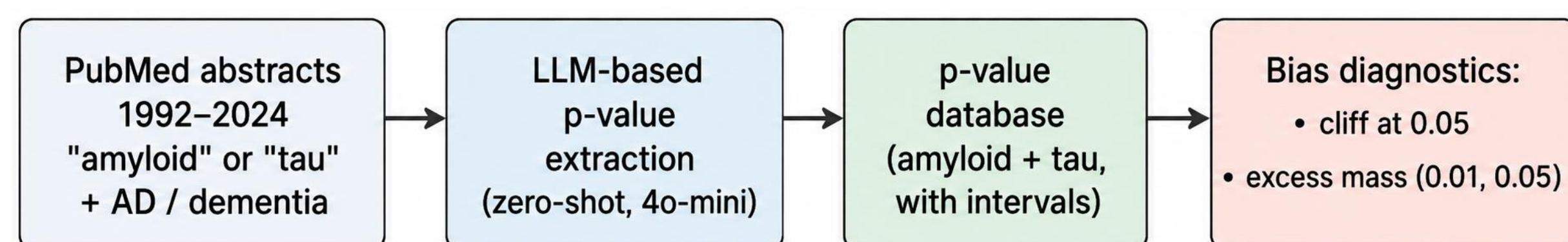
Do the amyloid and tau AD literatures show distinct signatures of p-hacking and publication bias?

Do differences narrow after the 2021 aducanumab approval?

PubMed Abstracts on Alzheimer's Disease or Dementia

Source. PubMed-indexed abstracts, 1992–present, referencing "amyloid" or "tau" + AD/dementia. **Comparator.** Tau acts as an internal field-matched control: same audience, journals, statistical norms.

LLM Extraction of P-Values from Journal Abstracts



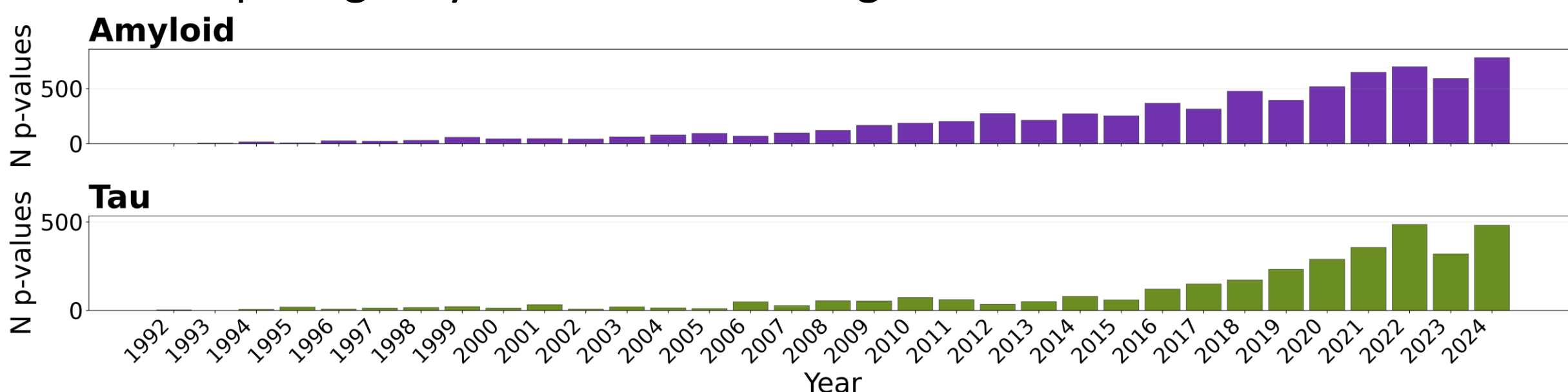
Pipeline. OpenAI API, ChatGPT 4o-mini model with zero-shot prompting; the LLM scans each abstract for p-values.

Validation. Precision $\approx 90\%$; recall $\approx 90\%$ on human-checked samples. Larger and few-shot variants did not outperform.

Why LLMs. LLMs flexibly capture the natural-language variants regex pipelines miss.

P-values by year show different publication timing for amyloid and tau.

This timing imbalance makes calendar year an important confounder when comparing amyloid and tau bias signatures.



P-Curve Methods, Interval-Censored Values, & Fitting

Gamma-Mixture P-Curve Distributions. Not every study has the same chance of finding a small p-value. We model p-values as a mix of studies with different statistical power & estimate that mix.

Interval-Censored Values. Models use the probability of each reported interval, rather than fitting exact values.

Examples: "p = 0.04" $\rightarrow$ [0.035, 0.045]; "p < 0.05" $\rightarrow$ (0, 0.05)

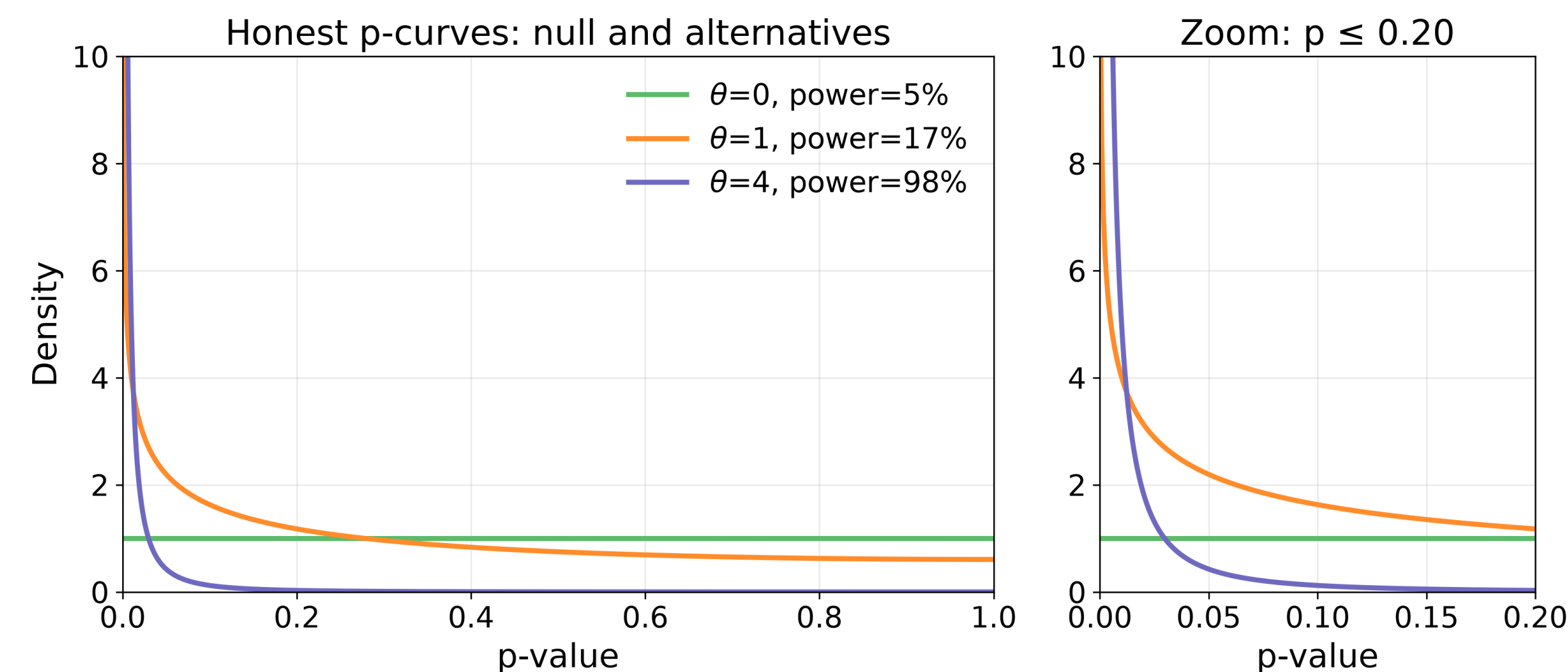
Bias Signatures:

- Model-free cliff statistic.** $\Pr(p \in [0.04, 0.05]) - \Pr(p \in [0.05, 0.06])$, a simple discontinuity measure at 0.05.
- Excess mass in (0.01, 0.05).** Fit the models excluding this window; compare predicted vs. observed mass inside it.

P-Value Distributions: Theory Imposes Shape Restrictions

Bumps or cliffs at $p = 0.05$ are inconsistent with an unbiased literature.

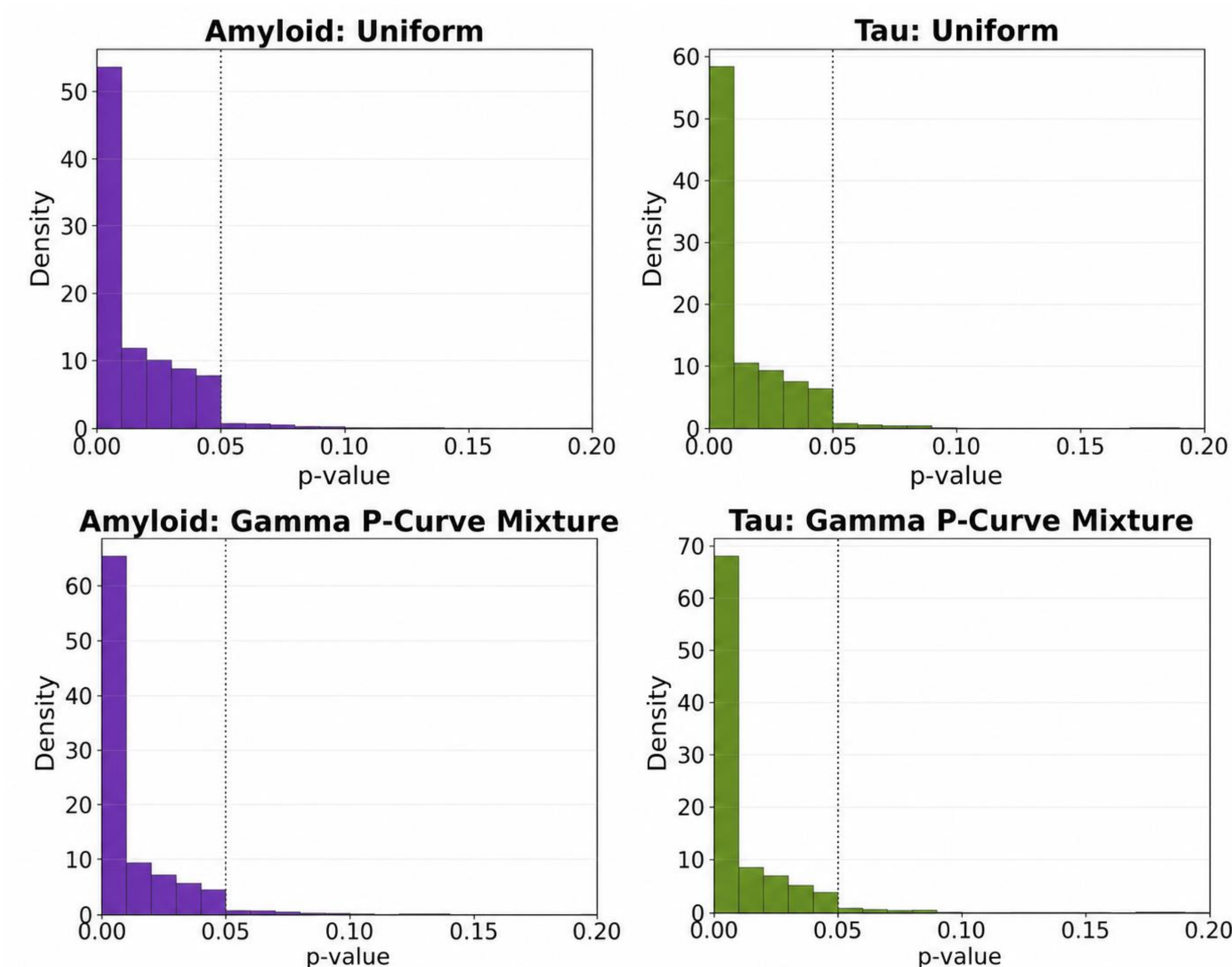
Under an honest normal model, p-curve densities are smooth, non-increasing, and convex on (0,1). Any nonnegative mixture, such as a gamma mixture, preserves convexity.



Empirical P-Value Distributions: Bias Present in Both Literatures

Interval-Smoothed Histograms

- Reported p-values are imprecise intervals, not exact points.
- Instead of forcing each p-value to a single number, we spread each observation's mass across its reported interval.



Top: Uniform Interval Smoothing.

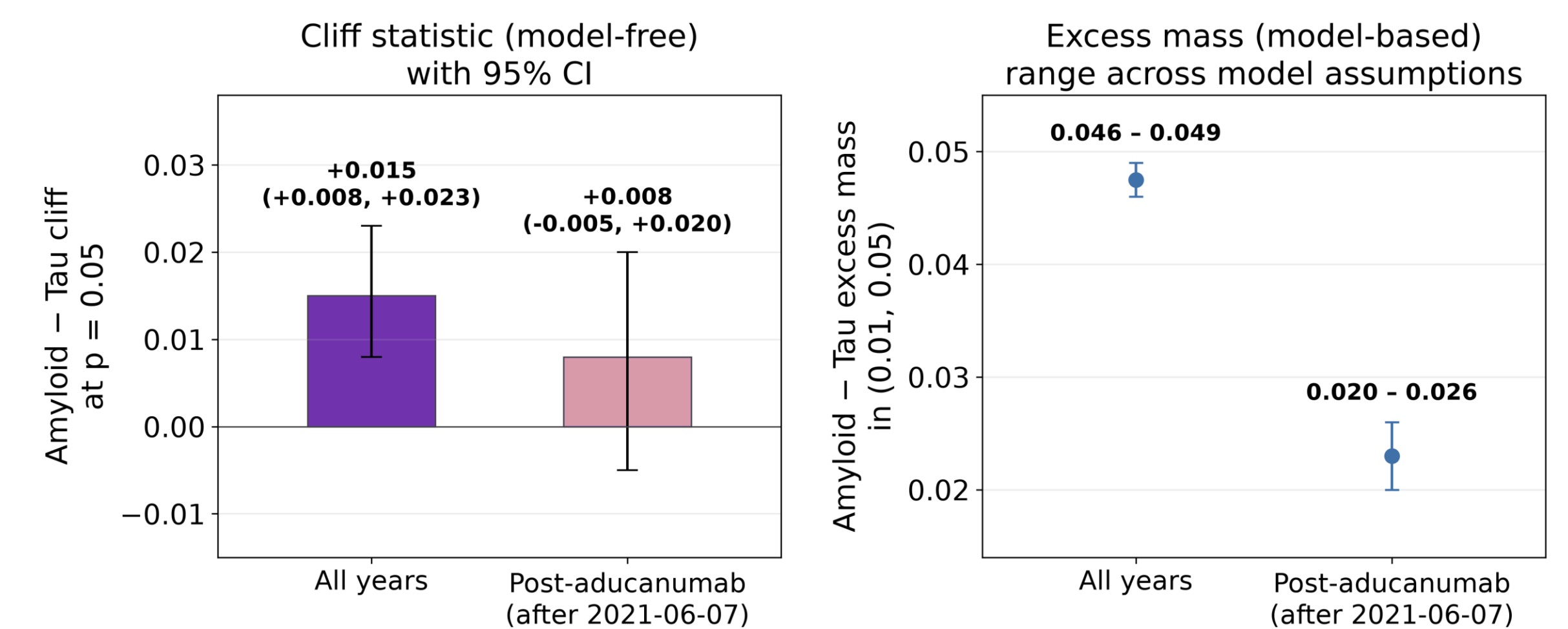
Amyloid (n = 7,223) and tau (n = 3,384) p-values from 1992–present were spread uniformly across their reported precision intervals before binning.

Bottom: P-Curve Gamma-Mixture Interval Smoothing.

The same data are reweighted using a Gamma p-curve mixture reference, representing a moderate-to-high-power honest literature.

The p = 0.05 cliff and excess mass below 0.05 remain, indicating that bias signatures are not an artifact of interval smoothing.

Bias Signatures: 1992–Present & 2021–Present



Across the full study period, amyloid shows stronger bias signatures than tau.

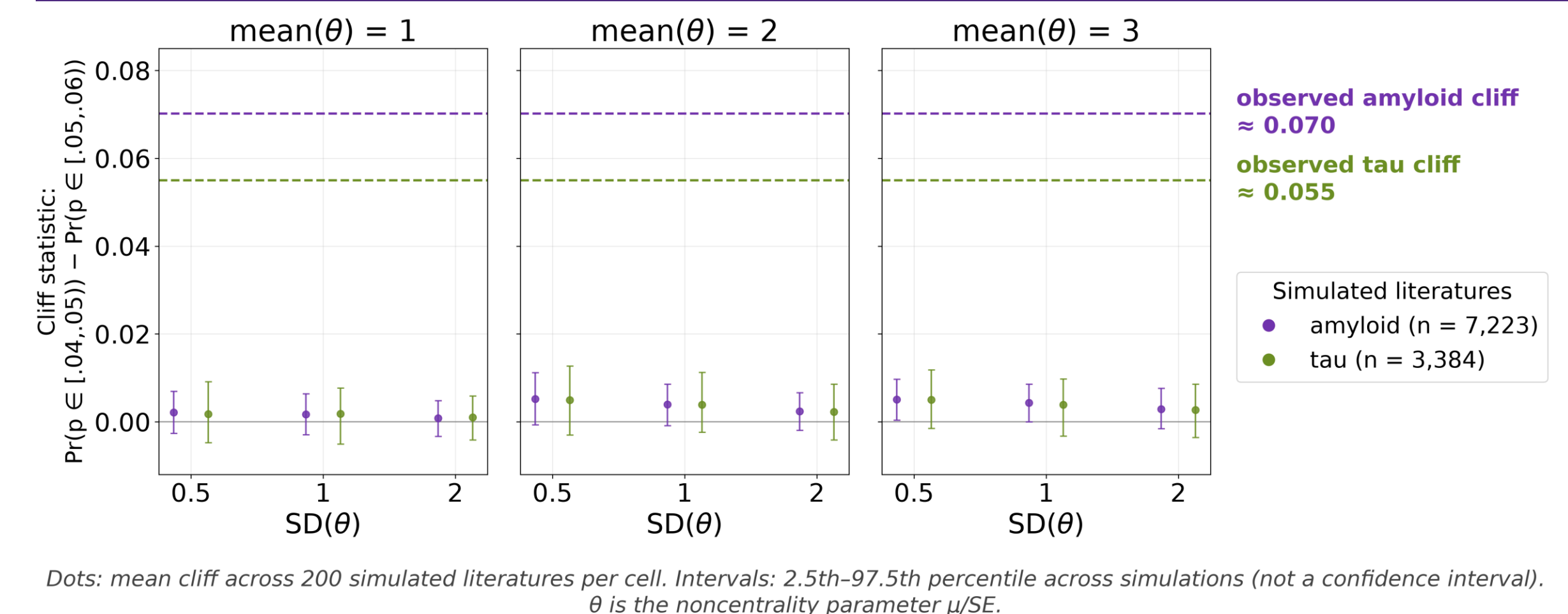
Amyloid has a larger cliff at p = 0.05, indicating a sharper discontinuity at the conventional significance threshold.

Amyloid also shows more excess mass below p = 0.05, suggesting more just-significant p-values than expected.

In the recent window, the amyloid-tau gap narrows.

The cliff gap is no longer clearly distinguishable from zero, and the excess-mass gap is substantially smaller.

Simulation: Honest Distributions Do Not Produce Large Cliffs



Dots: mean cliff across 200 simulated literatures per cell. Intervals: 2.5th–97.5th percentile across simulations (not a confidence interval). θ is the noncentrality parameter μ/SE .

Bias Signatures Converge After Accounting for Calendar Time

Both amyloid and tau show strong bias signatures. Sharp cliffs at $p = 0.05$ and substantial excess mass in (0.01, 0.05).

Calendar time is a confounder. Tau expanded later in the AD literature, so full-period comparisons mix an earlier-growing amyloid literature with a later-growing tau literature.

When restricted to the same recent period, amyloid and tau look similar. Bias signatures remain large in both literatures, but the amyloid-tau gap is much smaller.

Interpretation. Substantial p-hacking and/or publication bias is present in both literatures, with apparent amyloid-tau differences largely reflecting publication era and evolving reporting norms.



Funding: NIH NIA (R00AG073454, P01AG082653, R35AG071916), NIH NINDS (UE5NS070680), Alzheimer's Association (P0590624), and the Dana Foundation.