

Skeptical review: Cognitive Resilience and the Neuroepigenetic Landscape of Spatial Memory in Aging Egyptian Fruit Bats

Summary

This manuscript examines cognitive aging and putative resilience in Egyptian fruit bats (*Rousettus aegyptiacus*) using a multi-modal design combining (i) behavior from an ecologically motivated three-phase foraging task, (ii) diffusion MRI microstructure (DTI mean diffusivity; MD) extracted from 24 atlas-defined ROIs plus a global MD measure, and (iii) an epigenetic age estimate (DNAmAge) derived from skin DNA methylation (Sec. 1, Sec. 2.1–2.4). The authors introduce two novel behavioral metrics computed for phases 2 and 3: Spatial Memory Adaptation Efficiency (unique incorrect box entries before the first correct entry) and Prior Memory Interference Index (proportion of entries to the previously correct box) (Sec. 2.3, Sec. 3.2). In a complete-case analytic cohort of $n = 30$ bats with behavioral, DTI, and DNAmAge data (Sec. 3.1), they report no significant Spearman associations between DNAmAge and the behavioral metrics (Sec. 3.3) and no significant ROI-wise associations between MD and behavior after FDR correction across 25 MD measures \times 4 behavioral outcomes (Sec. 3.4). A pre-planned moderation analysis (DNAmAge moderating MD–behavior relationships) is not carried out because no primary MD–behavior associations survive correction (Sec. 3.4). The paper interprets these null findings as consistent with preserved performance across the sampled age range and discusses possible explanations (power limitations, non-linear trajectories, distributed substrates, limited sensitivity of MD) (Sec. 3.5, Conclusions/Sec. 4). The question and model system are compelling, and reporting null results is valuable, but the current draft requires substantial strengthening of methodological reporting, behavioral metric validation/robustness, alignment of statistical models with bounded/zero-inflated outcomes, explicit sensitivity/power characterization, and cleanup of template/placeholder artifacts before strong claims about “cognitive resilience” can be supported.

Strengths

- Clear conceptual focus on dynamic cognitive processes (adaptation and interference resolution) rather than only global decline, well motivated in Sec. 1.
- Ecologically relevant, multi-phase foraging task in a long-lived, spatially sophisticated species, extending cognitive aging work beyond standard laboratory organisms (Sec. 2.3).
- Novel, explicitly stated behavioral metrics (Adaptation Efficiency; Interference Index) that are simple to compute and interpret at face value (Sec. 2.3, Sec. 3.2).
- Ambitious multi-modal integration of behavior, diffusion MRI microstructure (MD), and epigenetic aging (DNAmAge), a combination that is rarely available in non-traditional model species (Sec. 2.1–2.4).

- Appropriate use of nonparametric Spearman correlations for DNAmAge–behavior and multiple-comparisons correction (FDR) for the ROI-wise brain–behavior screen, with transparent reporting of null results (Sec. 2.5–2.6, Sec. 3.3–3.4).
- Discussion acknowledges several plausible reasons for null findings (limited power, nonlinearity, distributed networks, measurement sensitivity), which is the right direction for framing negative results (Sec. 3.5, Sec. 4).

Major issues

1. **Reproducibility-critical methodological details are missing or too vague across all three modalities (behavior, DTI, DNAmAge), preventing rigorous evaluation of confounds and limiting reuse/replication (Sec. 2.1–2.4).**

For the foraging task, key parameters are not fully specified (number of boxes, spatial layout, arena geometry, phase durations/termination, inter-phase timing, reward schedule, habituation/training, counterbalancing/randomization of correct locations, and criteria for valid trials/phases) (Sec. 2.1, Sec. 2.3, Sec. 3.2). For DTI, acquisition and preprocessing are under-described (scanner/field strength, sequence, b-values, directions, resolution, TR/TE, anesthesia and motion mitigation, distortion/eddy correction, tensor fitting, skull stripping, atlas origin, registration strategy, ROI extraction, and QC thresholds) (Sec. 2.4). For DNAmAge, the methylation platform, preprocessing/normalization, clock model and its calibration/validation for this species/tissue/age range, and QC are not provided (Sec. 2.1–2.2).

Recommendation: Substantially expand Methods. (i) In Sec. 2.1 and Sec. 2.3, fully specify apparatus and protocol (arena dimensions; number/layout of boxes; cues/lighting; phase durations/stop rules; reward type/amount/schedule; habituation/training; whether phases are same day; and how correct-box locations are randomized/counterbalanced across bats and phases). State inclusion/exclusion rules (e.g., minimum activity/entries) and whether scorers were blinded. (ii) In Sec. 2.4, report full DTI acquisition parameters and a step-by-step preprocessing/QC pipeline (software + versions; motion/eddy/distortion correction; denoising; skull stripping; tensor fitting; atlas/template provenance; registration direction and method; ROI extraction; partial-volume mitigation; and QC thresholds plus number excluded for imaging QC). (iii) In Sec. 2.1–2.2 (or a dedicated subsection), describe biopsy processing, methylation assay, normalization, the epigenetic clock model (training set/species/tissue), expected error, QC filters, and how DNAmAge was computed. If space is limited, move full details to Supplementary Material but keep a complete summary in the main text.

2. **The behavioral metrics are the paper’s central contribution, but their construct validity, robustness, and full mathematical definition are not yet established (Sec. 2.3, Sec. 3.2–3.5).**

(a) Adaptation Efficiency (unique incorrect boxes before first correct) ignores repeated errors/perseveration and sequence/time structure; two bats with very different perseverative profiles can score identically. (b)

Interference Index (old-correct entries / total entries) can change due to denominator variation (overall exploration/activity), making interpretation ambiguous without also analyzing numerator/denominator and activity covariates. (c) Edge cases are not defined (e.g., if a bat never finds the new correct box; if total entries are zero) (Sec. 2.3.1–2.3.2). (d) Reliability/consistency across phases is not assessed, yet the discussion sometimes reads as if these are stable cognitive traits (Sec. 3.2–3.5).

Recommendation: In Sec. 2.3.1–2.3.2, provide fully specified computation rules (what counts as an “entry”; whether counts stop at first correct; handling of rapid re-entries; whether repeated visits to the same wrong box are counted; and explicit definitions for ‘never correct’ and zero-denominator phases—e.g., NA with exclusion rules, or capped values). Report task parameters needed to interpret ranges (e.g., total number of boxes; whether the stated 0–5 range is intrinsic or a cap) (Sec. 2.3; see also Sec. 3.2). Add a short ‘metric validation/psychometrics’ Results subsection: relate each metric to simpler behavioral quantities (total incorrect entries, latency to first correct, total entries/activity), and report cross-phase consistency (e.g., correlation of phase-2 vs phase-3 versions). Consider reporting complementary metrics that capture perseveration and timing (e.g., total incorrect entries before first correct; number of revisits; latency) alongside the proposed metrics to support construct validity.

- 3. The analytic cohort is complete-case ($n = 30$ from an initial $n = 41$), but missingness and potential selection bias are not characterized (Sec. 3.1).** If missing DTI/DNA_mAge/behavior is related to age, sex, colony, or performance (e.g., motion-prone animals, low-activity animals, scan failures), effect estimates and null conclusions can be biased, and generalizability is unclear.

Recommendation: In Sec. 3.1 (and/or a Supplementary table), provide a CONSORT-style flow: starting N , numbers missing behavior/DTI/DNA_mAge, and reasons (motion/QC failure, incomplete task, assay failure, etc.). Add an included-vs-excluded comparison table on available variables (sex, origin colony, DNA_mAge, any behavioral summaries, and chronological age if known) and test whether key variables differ between groups. If differences exist, acknowledge potential bias and consider sensitivity analyses (e.g., inverse probability weighting, or at minimum qualitative discussion).

- 4. The statistical modeling framework is incompletely described and, in places, poorly matched to the bounded/zero-inflated outcomes, limiting power and interpretability (Sec. 2.5–2.6, Sec. 3.3–3.4).** The paper uses Spearman for DNA_mAge–behavior but ROI-wise ‘mass-univariate’ analyses via linear regression with covariates; however, Adaptation Efficiency is a bounded count (reported 0–5) and Interference Index is a bounded proportion with many zeros—settings where OLS assumptions can be violated and effect estimates can be inefficient. In addition, the moderation-analysis Methods text contains formatting/corruption artifacts and lacks a clear executable specification (Sec. 2.6.2), and covariate handling is inconsistent between DNA_mAge–behavior correlations and MD–behavior models.

Recommendation: Rewrite Sec. 2.5–2.6 as a coherent, final analysis plan with explicit model equations and consistent covariates. For DNAmAge–behavior, either (i) report covariate-adjusted associations via regression (e.g., $\text{Behavior} \sim \text{DNAmAge} + \text{Sex} + \text{Colony}$) or (ii) justify unadjusted Spearman and state it clearly. For ROI analyses, consider outcome-appropriate GLMs: (a) Adaptation Efficiency as count/ordinal (Poisson/negative binomial or ordinal regression; include overdispersion checks), and (b) Interference as binomial using counts (old-correct entries out of total entries) rather than a raw ratio, optionally with random effects if repeated measures are used. If you retain OLS for comparability, add residual diagnostics/robust regression and explicitly justify. In Sec. 2.6.2, remove numbering/table artifacts and provide the exact moderation model (e.g., $\text{Behavior} \sim \text{MD} + \text{DNAmAge} + \text{MD} \times \text{DNAmAge} + \text{Sex} + \text{Colony}$), plus criteria for running it (pre-specified even under null main effects vs only if main effects pass FDR).

5. **Null findings are interpreted relatively strongly despite small n and heavy multiplicity (25 MD measures \times 4 outcomes = 100 tests), without quantitative sensitivity/power characterization (Sec. 3.3–3.5, Conclusions/Sec. 4).** Under these conditions, ‘no significant associations’ is ambiguous: it may reflect truly small effects, model mismatch, nonlinearity, or limited power after correction. The manuscript would be much more informative if it quantified what effect sizes are (and are not) compatible with the data.

Recommendation: Add a dedicated sensitivity/power subsection (end of Sec. 2.6 or start of Sec. 3.5). Report minimum detectable effect sizes for: (i) DNAmAge–behavior correlations (with chosen α), and (ii) ROI-wise MD–behavior tests under the applied FDR regime. In Results, report effect sizes with uncertainty (e.g., correlation/regression estimates with confidence intervals) rather than only p -values, and summarize the distribution of observed effects across ROIs (e.g., median and maximum $|\text{effect}|$). Consider equivalence tests or Bayesian analyses for DNAmAge–behavior to quantify evidence against moderate-to-large effects. In the Discussion/Conclusions, revise language to ‘consistent with preserved performance across this sample/age range under the tested models’ and avoid implying strong evidence of absence without these sensitivity bounds.

6. **DNAmAge is treated as a key aging variable, but its biological meaning in this cohort is insufficiently supported without reporting calibration to chronological age, tissue limitations, and potential measurement error (Sec. 2.1–2.2, Sec. 3.1–3.3).** It is currently unclear (i) whether chronological ages are known and how DNAmAge tracks them in this sample, (ii) whether the clock is validated for *R.aegyptiacus* skin, and (iii) whether ‘age acceleration’ (DNAmAge residualized on chronological age) could be a more interpretable predictor if chronological age is available.

Recommendation: In Sec. 3.1–3.3, report chronological age availability, its distribution, and DNAmAge–chronological age concordance (correlation and scatterplot). If chronological age is known, consider adding age-acceleration analyses (DNAmAge residuals) and/or include chronological age alongside DNAmAge to clarify whether DNAmAge adds information beyond age. In Methods (Sec. 2.1–2.2), provide clock provenance/validation and expected error; discuss tissue specificity and attenuation due to measurement error in Sec. 3.5. If chronological age is unknown/estimated, state estimation method and limitations explicitly.

7. **The manuscript appears not fully finalized: key tables contain placeholders/corrupted entries and some section formatting is broken, which undermines confidence in the results presentation and makes review/actionability difficult (Sec. 2.5, Sec. 3.1, Sec. 2.6.2).** Examples include Table 1 containing ‘TBD’ entries, Table 2 showing corrupted headers and implausible values (e.g., Global MD Min/Max both 0.00), and the malformed moderation-analysis section header/numbering artifacts.

Recommendation: Before resubmission, fully clean and validate the manuscript outputs. Replace ‘TBD’ with actual descriptive statistics or remove redundant tables; correct Table 2 headers and values and ensure MD values are in plausible ranges with units. Fix all section-numbering artifacts (e.g., stray ‘#’, malformed table-style headings), and cross-check consistency of cohort sizes across all tables/figures/text. Re-run scripts to regenerate tables/figures directly from the analysis pipeline to avoid manual transcription errors; note software versions and provide reproducibility details where possible.

Minor issues

1. Abstract contains irrelevant/template keywords (e.g., “Astronomy data analysis/data reduction”) and interpretive phrasing that can overstate conclusions given null results and limited power; key design details (ROIs/tests; moderation not performed) are not clearly stated (Abstract).

Recommendation: Replace keywords with study-relevant terms (spatial memory, cognitive aging, DNAmAge, diffusion MRI/DTI, fruit bat). In the Abstract, explicitly note the 24 ROIs + global MD and that planned moderation was not run because no MD–behavior associations survived FDR. Soften claims to ‘consistent with’ rather than implying strong evidence of resilience.

2. Introduction could more clearly separate conceptual motivation from testable hypotheses and pre-specified analyses; some repetition dilutes the main contribution (Sec. 1).

Recommendation: Streamline Sec. 1 and add a short end-of-Introduction paragraph listing explicit hypotheses (DNAmAge–behavior; MD–behavior; DNAmAge moderation) and the planned analysis approach (including correction strategy). Add brief re-

lated-work context on bat cognition, mammalian cognitive aging, and epigenetic clocks to position novelty.

3. ROI set and anatomical interpretation are difficult to evaluate because the 24 ROIs are not listed/categorized and near-significant effects are described without naming regions (Sec. 2.4.2, Sec. 3.4–3.5).

Recommendation: In Sec. 2.4.2, list all ROIs (or provide a categorized list) and specify atlas provenance. Add a supplementary table reporting (per ROI and behavioral outcome) effect size, SE/CI, uncorrected p , and FDR q . When mentioning nominal trends in Sec. 3.4, name the ROI(s).

4. Choice to analyze only mean diffusivity (MD) is not justified in Methods; this limits interpretability of null findings because other diffusion metrics may capture different microstructural processes (Sec. 2.4, Sec. 3.5).

Recommendation: In Sec. 2.4, justify MD as the primary metric (e.g., gray-matter sensitivity, prior bat work, data quality). Clarify whether FA/AD/RD were computed and, if feasible, provide exploratory analyses or at least a concrete plan for future work in Sec. 3.5.

5. Results text leans heavily on qualitative descriptions and figure references; effect sizes and uncertainty are not consistently summarized in-text (Sec. 3.1–3.4).

Recommendation: Add concise numerical summaries: distributions/ranges of behavioral metrics and MD (consistent with corrected tables), and representative effect sizes (e.g., maximum $|\rho|$, median $|\beta|$ across ROIs) with confidence intervals. Ensure each figure is introduced before citation and captions contain n , units, and key takeaway.

6. Ethics/welfare reporting is too brief for animal imaging and skin biopsy procedures (Sec. 2.1).

Recommendation: Expand Sec. 2.1 with approving bodies, protocol/permit numbers, anesthesia/analgesia and monitoring details for MRI and biopsy, and post-procedure care, consistent with journal standards.

7. Conclusions partially blur empirical findings (null under tested models) with speculative interpretations (distributed networks; non-linear trajectories; evolutionary resilience) (Conclusions/Sec. 4).

Recommendation: Restructure Sec. 4 to (i) state primary tested models and null outcomes, (ii) clearly label speculative explanations as hypotheses, and (iii) emphasize concrete next steps (larger n , longitudinal sampling, richer diffusion metrics, outcome-appropriate GLMs, preregistered ROIs).

8. Terminology mismatch: Results state no significant “linear or monotonic” **DNAmAge**–behavior correlations, but Methods specify Spearman (monotonic) rather than Pearson (linear) (Sec. 2.6.1, Sec. 3.3).

Recommendation: Revise wording to ‘no significant monotonic association (Spearman)’ or additionally report an explicit linear model if you intend to claim lack of linear association.

9. Several figures need improved standalone interpretability (units, n , legibility, and definitions), particularly for MD units and ROI labeling (e.g., Fig. 6–8) and for task schematic clarity (Fig. 1–3).

Recommendation: Increase font sizes and export as vector/high-resolution. Add units (MD), sample sizes, clear legends, and consistent variable naming. For ROI-heavy plots, consider grouping ROIs by system and/or moving full ROI labels to supplementary figures.

Very minor issues

1. Numerous typographical/LaTeX and formatting artifacts reduce professionalism and readability (e.g., broken words across lines; malformed italics for species name; inconsistent ‘p-values’ formatting) (multiple sections).

Recommendation: Proofread and fix LaTeX formatting (e.g., `\textit{Rousettus aegyptiacus}`), remove stray line breaks, and standardize statistical notation/rounding throughout.

2. Inconsistent variable naming for epigenetic age (`DNAmAge` vs `\text{DNAm_Age}`) across text/figures/tables (Sec. 2.2, Sec. 3.x; figures/tables).

Recommendation: Standardize to one notation (preferably `DNAmAge`) everywhere; if `\text{DNAm_Age}` is a raw column name, note the mapping once in Methods.

3. Title may overpromise breadth (e.g., “neuroepigenetic landscape”) relative to analyzing one epigenetic summary (skin `DNAmAge`) and one diffusion metric (MD) (Title; Sec. 1–2).

Recommendation: Either refine the title to match scope (`DNAmAge` + MD + spatial memory) or explicitly define early that ‘landscape’ is operationalized via these proxies.

4. Metric definitions omit explicit handling of rare but possible edge cases (e.g., zero total entries for Interference Index; never entering the new correct box for Adaptation Efficiency) (Sec. 2.3.1–2.3.2).

Recommendation: Add one-line conventions for each edge case (e.g., set to NA and exclude with stated rule; or cap to maximum) and report how many instances occurred.

Key statements and references

- Spearman rank correlations between DNAm_Age and each of the four behavioral metrics in the 30-bat cohort showed no statistically significant associations, with correlation coefficients and p -values of $\rho = -0.09$, $p = 0.619$ (Adaptation Efficiency P2), $\rho = -0.25$, $p = 0.192$ (Interference Index P2), $\rho = 0.10$, $p = 0.601$ (Adaptation Efficiency P3), and $\rho = -0.06$, $p = 0.751$ (Interference Index P3).
- *Reference(s)*: Benjamini-Hochberg
- A mass-univariate analysis relating each of the four behavioral metrics to Mean Diffusivity (MD) in 24 brain regions of interest plus a global MD measure (100 tests total), using linear regression models with Sex and Origin colony as covariates and Benjamini–Hochberg False Discovery Rate correction, found that none of the brain–behavior associations were statistically significant (all FDR-corrected p -values > 0.05).
- *Reference(s)*: Benjamini-Hochberg
- Because no regional or global MD measures showed significant associations with the behavioral metrics after multiple-comparisons correction, the planned moderation analysis testing the interaction term $\text{Regional_MD} \times \text{DNAmAge}$ in linear models of behavioral performance could not be carried out, as there were no primary significant brain–behavior relationships to moderate.
- *Reference(s)*: (none)
- Within the studied DNAmAge range of 6.62 to 13.84 years in Egyptian fruit bats, the absence of significant linear or monotonic correlations between DNAmAge and spatial memory adaptation or interference metrics indicates that individual differences in these cognitive functions are not predicted by epigenetic age, suggesting preservation of spatial memory adaptation and interference resolution across this segment of the lifespan.
- *Reference(s)*: (none)

Mathematical consistency audit

This section audits **symbolic/analytic** mathematical consistency (algebra, derivations, dimensional/unit checks, definition consistency).

Maths relevance: light

The paper contains light formal mathematics: two custom behavioral metric definitions expressed verbally (a count of unique incorrect entries before first correct entry; a ratio of entries to the previously-correct location), a linear moderation model formula with an interaction

term, and analytic bookkeeping (number of tests in mass-univariate analysis). There are no extended derivations; the main audit focus is internal definitional completeness, symbol consistency, and basic algebra/range checks for the metrics and model specification.

Checked items

1. \triangle **Adaptation Efficiency metric definition** (Sec. 2.3.1, p.3)

- **Claim:** $\text{\text{Adaptation_Efficiency}}$ for a phase equals the number of unique incorrect box entries made before the first successful entry into that phase's newly correct box (entry = action 'E' or 'F'); lower is better.
- **Checks:** definition well-posedness, edge-case analysis, symbol/term consistency
- **Verdict:** UNCERTAIN; confidence: high; impact: critical
- **Assumptions/inputs:** A phase has a well-defined 'newly correct box' location., Behavioral logs contain a sequence of actions with box identifiers., At least one successful entry into the newly correct box occurs, or a rule exists if it does not.
- **Notes:** As written, the metric requires a 'first successful entry' into the newly correct box. The paper does not define what happens if that event never occurs within the phase (metric undefined), nor does it state any truncation window or exclusion/censoring rule. This is a core construct of the paper, so the missing definition prevents full internal verification.

2. \checkmark **Interference Index metric definition and bounds** (Sec. 2.3.2, p.3)

- **Claim:** $\text{\text{Interference_Index}}$ for a phase equals (number of entries into the previous phase's correct box during the current phase) divided by (total number of all box entries during the current phase).
- **Checks:** range/normalization, definition consistency, edge-case analysis
- **Verdict:** PASS; confidence: medium; impact: moderate
- **Assumptions/inputs:** Numerator counts a subset of the denominator events (entries)., Denominator is positive (at least one entry in the phase)., Entry is consistently defined across numerator and denominator (implicit from Sec. 2.3.1).
- **Notes:** If numerator counts a subset of the same 'entry' events counted in the denominator, the ratio is guaranteed to lie in $[0, 1]$, matching the Results narrative. However, the denominator-zero case is not specified (likely rare). Also, the paper should explicitly confirm numerator/denominator use the same entry definition ('E'/'F').

3. \checkmark **Moderation model specification** (Sec. 2.6.2, p.4 (also shown again in the combined page image at end))

- **Claim:** A moderation (interaction) model is $\text{Behavioral_Metric} \sim 1 + \text{Regional_MD} + \text{DNAmAge} + \text{Regional_MD} \times \text{DNAmAge}$; a significant interaction indicates age modulation of the MD–behavior relationship.
 - **Checks:** model form/algebra, notation consistency
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** Behavioral_Metric is treated as a continuous dependent variable in a linear model., Regional_MD and DNAmAge are numeric predictors., Interaction term is the product of the two predictors.
 - **Notes:** The stated formula correctly includes intercept, both main effects, and their interaction for moderation testing. Notational variants (\times vs $*$) appear, but they denote the same product interaction.
4. ✓ **Mass-univariate test count arithmetic** (Sec. 3.4, p.7)
- **Claim:** Total tests = 100 = (25 MD measures: 24 ROIs + global) \times (4 behavioral metrics).
 - **Checks:** algebra/arithmetic consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** Exactly 24 ROI MD values are tested plus one global MD value., Exactly four behavioral metrics are tested.
 - **Notes:** The multiplication $25 \times 4 = 100$ is correct and matches the described set of predictors/outcomes.
5. ✓ **Use of 'correlation' vs regression with covariates** (Sec. 2.6.1, p.4 and Sec. 3.4, p.7)
- **Claim:** The paper describes a mass-univariate 'correlation analysis' but also states that Sex and Origin colony were included as covariates in a linear regression model.
 - **Checks:** definition/terminology consistency
 - **Verdict:** PASS; confidence: medium; impact: minor
 - **Assumptions/inputs:** The intended analysis is a linear regression/partial association test rather than a simple bivariate correlation.
 - **Notes:** Analytically, a regression with covariates is a coherent way to test adjusted associations, but calling it 'correlation' can confuse what statistic is being tested (bivariate vs adjusted/partial). This is mainly clarity rather than a mathematical contradiction.
6. ✗ **Claim of no 'linear or monotonic' correlation vs stated Spearman method** (Sec. 2.5, p.4 and Sec. 3.3, p.6)

- **Claim:** Methods: Spearman rank correlation used. Results: 'no statistically significant linear or monotonic correlations' between `\text{DNAm_Age}` and behavioral metrics.
 - **Checks:** method-claim consistency
 - **Verdict:** FAIL; confidence: high; impact: moderate
 - **Assumptions/inputs:** Only Spearman correlations were computed as described.
 - **Notes:** Spearman tests monotonic association, not specifically linear association. The Results' inclusion of 'linear' overstates what was tested unless an additional linear (e.g., Pearson or regression slope) test is also performed and reported, which is not stated.
7. ✘ **Epigenetic age symbol consistency (DNAmAge vs `\text{DNAm_Age}`)** (Sec. 2.2, p.3; Table 1 p.4; Table 2 p.5; Sec. 3.3 p.6)
- **Claim:** The epigenetic age variable is consistently referenced across Methods and Results.
 - **Checks:** notation consistency
 - **Verdict:** FAIL; confidence: high; impact: minor
 - **Assumptions/inputs:** DNAmAge and `\text{DNAm_Age}` refer to the same underlying quantity.
 - **Notes:** The paper switches between 'DNAmAge' and '`\text{DNAm_Age}`' without an explicit equivalence statement. This is likely a naming/formatting inconsistency but should be standardized to avoid ambiguity.
8. △ **Table 1 vs Table 2 descriptive statistics coherence** (Table 1, p.4; Table 2, p.5; Sec. 3.1, p.5)
- **Claim:** Descriptive statistics reported in Table 1 and Table 2 are compatible and refer to clearly defined cohorts.
 - **Checks:** definition/cohort consistency
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
 - **Assumptions/inputs:** Table 1 may refer to the initial dataset and Table 2 to the complete-case final cohort.
 - **Notes:** Table 1 reports DNAmAge Max = 15.07, while the final cohort in Sec. 3.1/Table 2 has Max = 13.84. This could be consistent if Table 1 summarizes the pre-filtered dataset, but Table 1 does not explicitly state its cohort/sample, leaving ambiguity.
9. △ **Reported metric ranges vs definitions** (Sec. 3.2, p.6; Table 2, p.5)
- **Claim:** Interference Index ranges 0 to 1; Adaptation Efficiency ranges 0 to 5
 - **Checks:** range/normalization, dependency on unspecified task parameters
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor

- **Assumptions/inputs:** Interference numerator is a subset of total entries., The task has a finite number of distinct boxes; metric may be implicitly capped by that number minus one.
- **Notes:** Interference Index in $[0,1]$ is consistent with the ratio definition. Adaptation Efficiency's stated upper bound (5) cannot be derived without knowing the total number of available boxes (or an explicit truncation rule).

Limitations

- The document contains very few explicit equations and no step-by-step derivations; most 'math' is in verbal metric definitions and model descriptions, limiting the depth of algebraic auditing.
- No explicit notation section or formal definitions (with domains) are provided for several quantities (e.g., number of boxes, phase duration), so some range/edge-case checks cannot be fully completed.
- Figures are referenced for statistical outputs, but the underlying analytic expressions (e.g., exact regression equations, test statistics) are not shown, preventing verification beyond the stated model forms.

Numerical results audit

This section audits **numerical/empirical** consistency: reported metrics, experimental design, baseline comparisons, statistical evidence, leakage risks, and reproducibility.

Of 14 numerical checks, 13 passed and 1 failed. The only flagged inconsistency is a cross-table mismatch in DNAm age mean/SD between Table 1 and Table 2; other internal arithmetic, cohort-size consistency ($n = 30$), reported test-count arithmetic ($25 \times 4 = 100$), and multiple sanity/range checks passed.

Checked items

1. ✓ **C1** (p.5, Results §3.1 (Cohort characteristics))
 - **Claim:** Initial dataset comprised 41 individual bat records; final analytical cohort of 30 bats.
 - **Checks:** difference_check
 - **Verdict:** PASS
 - **Notes:** Computed excluded = $41 - 30 = 11$; matches expected excluded count and is non-negative.
2. ✓ **C2** (p.7, Results §3.4 (Mass-univariate analysis description))
 - **Claim:** Total tests conducted: 100 in total: 25 MD measures \times 4 behavioral metrics.
 - **Checks:** product_check
 - **Verdict:** PASS
 - **Notes:** Verified $25 \times 4 = 100$ equals reported total tests.

3. ✓ **C3** (p.7, Results §3.4 and p.9 Conclusions (ROIs count))
 - **Claim:** MD extracted from 24 predefined brain ROIs, plus a global measure (implied 25 MD measures).
 - **Checks:** sum_check
 - **Verdict:** PASS
 - **Notes:** Verified 24 ROIs + 1 global measure = 25 MD measures.
4. ✗ **C4** (p.4 Table 1 vs p.5 Table 2 (DNAm age descriptive stats))
 - **Claim:** DNAm age summary differs between Table 1 and Table 2 (e.g., mean 9.68 vs 9.45; max 15.07 vs 13.84).
 - **Checks:** cross_table_consistency_check
 - **Verdict:** FAIL
 - **Notes:** Mins match exactly (6.62 vs 6.62) and $\max(\text{Table 2}) = 13.84$ is $\leq \max(\text{Table 1}) = 15.07$, but mean and SD differences exceed rounding tolerance (mean diff 0.23; SD diff 0.17).
5. ✓ **C5** (p.5 Results §3.1 narrative vs Table 2 (`\text{DNAm_Age}` max))
 - **Claim:** Narrative: `\text{DNAm_Age}` ranged from 6.62 to 13.84 years; Table 2 lists max 13.84 (consistent).
 - **Checks:** within_section_consistency_check
 - **Verdict:** PASS
 - **Notes:** Narrative min/max exactly match Table 2 min/max (6.62 to 13.84).
6. ✓ **C6** (p.5 Table 2 (`\text{DNAm_Age}` mean relative to min/max))
 - **Claim:** `\text{DNAm_Age}` mean 9.45 should lie within [6.62, 13.84].
 - **Checks:** range_check
 - **Verdict:** PASS
 - **Notes:** Verified $6.62 \leq 9.45 \leq 13.84$.
7. ✓ **C7** (p.5 Table 2 (Adaptation Efficiency P2))
 - **Claim:** Adaptation Efficiency P2: mean 1.97, SD 1.56, median 2.00, min 0.00, max 5.00.
 - **Checks:** range_sanity_check
 - **Verdict:** PASS
 - **Notes:** Sanity constraints satisfied: $\min \leq \text{median} \leq \max$, $\min \leq \text{mean} \leq \max$, $\text{SD} \geq 0$.
8. ✓ **C8** (p.5 Table 2 (Adaptation Efficiency P3))
 - **Claim:** Adaptation Efficiency P3: mean 2.50, SD 1.83, median 2.00, min 0.00, max 5.00.
 - **Checks:** range_sanity_check

- **Verdict:** PASS
 - **Notes:** Sanity constraints satisfied: $\min \leq \text{median} \leq \max$, $\min \leq \text{mean} \leq \max$, $\text{SD} \geq 0$.
9. ✓ **C9** (p.5 Table 2 (Interference Index P2))
- **Claim:** Interference Index P2: mean 0.28, SD 0.29, median 0.22, min 0.00, max 1.00.
 - **Checks:** proportion_range_check
 - **Verdict:** PASS
 - **Notes:** All reported values lie within $[0, 1]$ where applicable; $\text{SD} \geq 0$.
10. ✓ **C10** (p.5 Table 2 (Interference Index P3))
- **Claim:** Interference Index P3: mean 0.16, SD 0.13, median 0.15, min 0.00, max 0.47.
 - **Checks:** proportion_range_check
 - **Verdict:** PASS
 - **Notes:** All reported values lie within $[0, 1]$ where applicable; $\text{SD} \geq 0$.
11. ✓ **C11** (p.5 Table 2 (Global MD summary))
- **Claim:** Global Mean Diffusivity (MD): mean 0.00073, SD 0.00004, median 0.00073, min 0.00065, max 0.00080.
 - **Checks:** range_sanity_check
 - **Verdict:** PASS
 - **Notes:** Sanity constraints satisfied: $\min \leq \text{median} \leq \max$, $\min \leq \text{mean} \leq \max$, $\text{SD} \geq 0$.
12. ✓ **C12** (p.5 Table 2 ($\text{\text{DNAm_Age}}$ summary))
- **Claim:** $\text{\text{DNAm_Age}}$ (years): mean 9.45, SD 1.62, median 9.22, min 6.62, max 13.84.
 - **Checks:** range_sanity_check
 - **Verdict:** PASS
 - **Notes:** Sanity constraints satisfied: $\min \leq \text{median} \leq \max$, $\min \leq \text{mean} \leq \max$, $\text{SD} \geq 0$.
13. ✓ **C13** (p.6 Results §3.3 (Spearman correlations reported))
- **Claim:** Spearman correlations: (ρ, p) pairs are: $(-0.09, 0.619)$, $(-0.25, 0.192)$, $(0.10, 0.601)$, $(-0.06, 0.751)$; all are non-significant at 0.05.
 - **Checks:** p-value_threshold_check
 - **Verdict:** PASS
 - **Notes:** All reported p -values exceed 0.05 (minimum p reported: 0.192).
14. ✓ **C14** (p.6 Results §3.3 and p.7 Fig.4 caption (n))

- **Claim:** Correlations are reported for cohort size $n = 30$.
- **Checks:** cross_section_consistency_check
- **Verdict:** PASS
- **Notes:** $n = 30$ matches across the referenced sections.

Limitations

- Audit uses only the numeric values explicitly present in the provided PDF text; underlying datasets (CSV/XLSX/NIfTI) are not available for recomputation of statistics.
- Figures are image-based; no numeric values were extracted from plot visuals or pixels, so figure-derived counts or distributions cannot be validated.
- Many claims are inferential/statistical (e.g., FDR-corrected null results) and require full test result tables to verify; only limited summary numbers are provided.
- Cannot recompute Spearman correlation coefficients and p -values without the underlying paired data.
- Cannot verify FDR-adjusted p -values (or the count of significant findings) without the full list of raw p -values/test results across all 100 tests.
- Figure-based cohort-flow details beyond the explicitly stated counts cannot be validated without extractable numeric figure values or an underlying missingness table.
- Behavioral metric computations (Adaptation Efficiency, Interference Index) cannot be validated without event-level behavioral data.