

# *Skeptical review: Cognitive-Structural Decoupling in Long-Lived Bats: Quantifying Resilience Beyond Age and Global Brain Structure*

---

## Summary

The manuscript asks whether cognitive performance in long-lived Egyptian fruit bats is “decoupled” from biological age (DNAmAge) and global white-matter structure (DTI-derived global FA/MD). The authors derive a behavioral score, Cognitive Adaptation Efficiency (CAE), from perseverative error rates across short- and long-term memory phases of a multi-phase foraging task (Sec. 2.2.1–2.2.2; Sec. 3.2). They then define a Cognitive-Structural Decoupling Index (CSDI) as the residual from a multiple regression predicting CAE from DNAmAge, sex, and global FA/MD (Sec. 2.4.2; Sec. 3.4.1–3.4.2). The reported regression has low explanatory power and is not statistically significant, which is interpreted as evidence for decoupling (Sec. 3.5; Conclusion). The study is conceptually interesting and the residual-based framing could be useful, but the current version is undermined by (i) inconsistent cohort/sample-size reporting across Methods vs Results, (ii) DTI values that appear physiologically implausible and thus call into question the validity of key predictors, (iii) incomplete/ambiguous formal definition and edge-case handling for CAE, and (iv) over-interpretation of a null/low-power regression as “direct evidence” of decoupling. Addressing data accounting, behavioral metric specification/robustness, imaging QC, and more cautious statistical interpretation would substantially strengthen the manuscript’s claims and impact.

## Strengths

- Compelling biological motivation: bats are a valuable system for studying aging and potential cognitive maintenance in long-lived mammals (Introduction/Sec. 1).
- Multi-modal dataset integrating behavior, DNAmAge, and DTI, with an overall pipeline that (once clarified) could be a useful template for future work (Sec. 2.1–2.4).
- The CAE and CSDI constructs provide an intuitively appealing operationalization of “better-than-expected” cognition using a standard residualization framework (Sec. 2.2.2; Sec. 2.4.2).
- Manuscript organization is generally clear, with transparent acknowledgement of several limitations (Sec. 3.3; Sec. 3.5; Conclusion).
- Figures use clear multi-panel layouts and generally support reader orientation (though several need stronger quantitative/statistical annotation).

## Major issues

1. **Cohort definition and sample size reporting are internally inconsistent across the manuscript, making it unclear which animals contributed to which analyses and undermining reproducibility. Methods report a final**

cohort of  $N = 28$  with specific sex/colony counts (Sec. 2.1), while Abstract/Results repeatedly use  $N = 30$  and provide different sex/colony compositions (Sec. 3.1–3.4; Table 2; figures). DNAmAge descriptive statistics also differ between Methods and Results despite the same stated range (Sec. 2.1 vs Sec. 3.1).

*Recommendation:* Provide a single, unambiguous accounting of the dataset: a CONSORT-like flow from initial  $N$  (e.g.,  $N = 41$ ) through exclusions (with reasons) to the final  $N$  for each modality and each analysis (behavior, DNAmAge, DTI, regression). Add a table listing per-animal data availability (or at least counts of complete cases by modality). Then harmonize  $N$ , sex/colony counts, DNAmAge mean $\pm$ SD, figure captions, and Table 2 degrees of freedom to match the actually analyzed cohorts (Sec. 2.1; Sec. 3.1–3.4).

2. DTI predictors (Global\_FA, Global\_MD) appear physiologically implausible (e.g., global FA  $\approx 0.99$  with extremely low diffusivities), strongly suggesting acquisition/preprocessing/unit-scaling/masking problems. Because these variables are core predictors in the CAE regression and central to the structural-decoupling narrative, their questionable validity jeopardizes the main conclusions (Sec. 2.3; Sec. 3.3; Sec. 3.4.1; Sec. 3.5).

*Recommendation:* Add (and act on) a focused DTI QA/QC section. Concretely: (i) explicitly state units and scaling for MD/AD/RD ( $\text{mm}^2/\text{s}$  vs  $m^2/\text{s}$ ; whether  $1e^{-3}$  scaling is applied), (ii) describe brain/WM mask generation and exactly how “global mean” metrics were computed (whole-brain vs WM-only vs atlas ROIs vs skeletonized voxels), (iii) report preprocessing steps and settings (eddy/motion/susceptibility correction; outlier handling) and software/versions (Sec. 2.3), and (iv) provide basic QC visualizations (representative FA/MD maps + masks; FA/MD histograms across voxels; motion/SNR summaries). If an error is found, recompute global metrics and update all downstream analyses. If uncertainty remains, move structure-based conclusions to exploratory/sensitivity analyses and foreground results that do not depend on DTI integrity (e.g., CAE  $\sim$  DNAmAge; Sec. 3.4.1).

3. CAE is central to the paper but is not specified with a fully explicit, unambiguous mathematical definition, and current edge-case handling can bias CAE upward. The displayed formula is ambiguous without parentheses and may contain a sign/precedence interpretation that would incorrectly increase CAE with higher LTM perseveration. Additionally, perseverative error rate is set to 0 when total entries in a phase are 0 (division-by-zero handling), which treats non-participation as perfect performance and can inflate CAE; small denominators also make rates unstable (Sec. 2.2.1–2.2.2; Sec. 3.2).

*Recommendation:* In Sec. 2.2.2, provide the exact CAE formula with explicit parentheses, signs, weighting (STM vs LTM), and theoretical range; include a worked example. Verify the typeset equation matches the implemented code (and fix if not). Replace the division-by-zero convention: treat phases with 0 entries as missing/undefined (or enforce a minimum-entry threshold), and report how many bats/phases are affected. Add sensitivity analyses for CAE (e.g., excluding low-engagement phases; alternative thresholds; treating STM/LTM separately; or modeling perseverative counts using an appropriate binomial framework with exposure/denominator) and report whether key conclusions persist (Sec. 3.2; Sec. 3.4).

4. **The manuscript over-interprets a non-significant, low- $R^2$  regression as “direct evidence” of decoupling. With  $N \approx 30$ , multiple predictors, potential measurement noise (behavior and especially DTI), bounded/skewed outcome, and possible nonlinearities, failure to reject the null is not strong evidence of absence of association (Sec. 3.4.1; Sec. 3.5; Abstract; Conclusion).**

*Recommendation:* Reframe claims in the Abstract, Sec. 3.5, and Conclusion to emphasize that the study did not detect strong linear associations under the current design and power, rather than asserting direct evidence of decoupling. Report effect sizes with confidence intervals (and standardized coefficients) for all predictors, including the borderline DNAmAge effect noted in Sec. 3.4.1. Add a power/sensitivity analysis (minimum detectable effect size given  $N$ ) and/or Bayesian estimation to quantify evidence for near-zero effects. Include simpler supporting analyses (univariate CAE  $\sim$  DNAmAge; CAE  $\sim$  FA; CAE  $\sim$  MD; partial correlations) and clearly distinguish “absence of evidence” from “evidence of absence.”

5. **CSDI is presented as an individualized “resilience/decoupling” index, but its validity and robustness are not established and it likely inherits instability from the weak/possibly misspecified model and questionable DTI predictors. Residuals can largely reflect noise, collinearity, outliers, or model choice rather than a stable trait-like resilience measure (Sec. 2.4.2; Sec. 3.4.2).**

*Recommendation:* Treat CSDI explicitly as exploratory unless you can demonstrate robustness. Add stability checks: bootstrap the regression and report uncertainty for each CSDI (or at least rank stability), and compare CSDI values across alternative specifications (e.g., CAE  $\sim$  DNAmAge + Sex only; with/without FA/MD; robust regression). Report whether high/low CSDI individuals remain consistent across models. Also test associations between CSDI and potential nuisance variables (total entries/engagement, colony, session/batch) to ensure CSDI is not primarily capturing these factors (Sec. 3.4.2; Sec. 3.1–3.2).

6. **Potential confounding/hierarchical structure is insufficiently addressed.** Colony (Aseret vs Herzeliya) is reported but not modeled; it may proxy environmental differences, handling, scanning sessions/batch effects, or task exposure differences. Engagement (total entries) may also confound CAE if activity levels vary substantially across individuals (Sec. 2.1; Sec. 3.1–3.2; Sec. 2.2.2).

*Recommendation:* At minimum, include colony as a covariate in the CAE model (or justify exclusion given power). Report descriptive comparisons by colony/sex for DNAmAge, CAE, and (if valid) DTI metrics (Sec. 3.1–3.2). Evaluate whether CAE correlates with total entries (and consider including total entries as a covariate or switching to a count-based model of perseveration with appropriate exposure). If methylation age or imaging were processed in batches, report and (if possible) adjust for batch/scanning session effects (Sec. 2.1; Sec. 2.3; Sec. 2.4.2).

7. **Aims and narrative emphasize regional “Structural Preservation Hotspots” and region-wise neural signatures, but the Results state regional analyses could not be performed due to technical issues. This creates a mismatch between promised contributions and delivered results (Introduction/Sec. 1; Sec. 2.4.1–2.4.3; Sec. 3.3).**

*Recommendation:* Refocus the manuscript around what was actually executed (global metrics + behavioral/DNAmAge analyses) and clearly label regional hotspot mapping as planned/future work. Move unexecuted regional methods (Sec. 2.4.1; Sec. 2.4.3) to an Appendix or a dedicated “Planned analyses” subsection, and update the Abstract/Introduction/Conclusion so they do not imply regional neural signatures were identified in this study (Sec. 3.3; Sec. 3.5).

8. **Modeling and reporting choices are not well matched to the data characteristics and currently hinder interpretability: CAE is bounded in  $[0, 1]$  and appears skewed/near-ceiling, yet OLS is used without adequate diagnostics; the regression output/table appears corrupted and/or inconsistent with described software; and numerical stability/collinearity concerns are suggested (e.g., extremely large condition number mentioned in the unstructured report) but not addressed (Sec. 3.4.1; Figure 4; Figure 6; Sec. 2.4.2).**

*Recommendation:* Improve statistical reporting and align the model to the outcome. Provide a clean regression table (coefficients, SEs,  $t$ , exact  $p$ , CIs,  $R^2$ /adj- $R^2$ ,  $F$  and  $p$ , AIC/BIC) and clarify the software used (R vs Python/statsmodels) with package versions (Sec. 3.4.1; Sec. 2.4). Add diagnostics: residual plots, influence, and multicollinearity checks (VIF/condition number), and standardize continuous predictors for interpretability. Consider a model appropriate for bounded outcomes (beta regression; quasi-binomial on error rates; or robust/Spearman-based analyses) and report whether conclusions are consistent across approaches (Figure 4; Sec. 3.4).

## Minor issues

1. Figure 1 asserts no selection bias between initial and final cohorts, but the figure/caption as described do not provide direct side-by-side evidence or statistical comparisons supporting this claim (Figure 1; Sec. 3.1).

*Recommendation:* Add explicit comparisons between initial vs final cohorts (side-by-side distributions/densities for DNAmAge; proportions for sex/colony) and include statistical tests or standardized differences in the caption. Alternatively, remove/soften the “no selection bias” statement if not supported.

2. Behavioral task description lacks key parameters needed to judge difficulty, engagement, and reproducibility (number of boxes, session structure, phase timing/definition of STM vs LTM separation, reward randomization, training/habituation, exclusion criteria for problematic logs) (Sec. 2.2.1–2.2.2).

*Recommendation:* Expand Sec. 2.2.1–2.2.2 with essential task design details and explicit criteria for excluding/imputing sessions/phases. Report distributions of total entries per phase and discuss whether CAE shows ceiling/floor effects given task difficulty (Sec. 3.2).

3. Figures 4 and 6 do not provide sufficient quantitative/statistical annotations to support the narrative; Figure 6 caption contains unclear/stray numbers and does not state whether predictions are in-sample or cross-validated (Figure 4; Figure 6; Sec. 3.4.1).

*Recommendation:* Annotate plots with slopes/effect sizes, CIs,  $p$ -values (or correlation coefficients), and  $N$ . For Figure 6, explicitly state what is on each axis, whether predictions are in-sample or cross-validated, and add calibration/fit metrics ( $R^2$ , RMSE/MAE). Clean up the caption to remove stray artifacts.

4. Predictor choice and encoding are under-justified: AD/RD are computed but not used; sex coding is not specified; interactions ( $\text{Age} \times \text{Sex}$ ;  $\text{Age} \times \text{DTI}$ ) and nonlinearities are not discussed (Sec. 2.3; Sec. 2.4.2; Sec. 3.4.1).

*Recommendation:* In Sec. 2.4.2, justify the specific predictor set, specify sex encoding (e.g., 0/1), and briefly discuss whether alternative specifications (dropping one of FA/MD; using AD/RD; adding nonlinear terms/interactions) were considered and why they were/weren’t included given power.

5. Ethics/animal welfare information is missing despite live-animal behavioral testing and MRI scanning (Sec. 2.1–2.3).

*Recommendation:* Add an ethics statement (e.g., end of Sec. 2.1 or a new section) with approval identifiers, housing/enrichment, handling, and MRI anesthesia/restraint protocols and welfare measures.

6. Reproducibility details are incomplete (software packages/versions, key preprocessing parameters, availability of scripts/code) (Sec. 2.3; Sec. 2.4).

*Recommendation:* Add a reproducibility paragraph listing software and versions (DTI preprocessing, registration, statistical modeling), critical parameters (mask thresholds, correction settings), and whether code is public or available upon request.

7. Section 2.4.3 proposes regressing region-wise measures on DNAmAge and Sex after defining CSDI as residuals from a model already containing DNAmAge and Sex; in standard OLS with identical design matrices, residuals are orthogonal to those covariates, making this appear redundant/confusing (Sec. 2.4.3).

*Recommendation:* Clarify the intended rationale (e.g., different sample/encoding, robustness) or remove DNAmAge/Sex from the secondary model and explain the orthogonality property explicitly.

8. The manuscript states CSDI is “expected” to be approximately normally distributed because it is a residual; mean-zero is guaranteed (with intercept), but normality is not (Sec. 3.4.2).

*Recommendation:* Rephrase to separate guaranteed residual properties (mean  $\approx 0$ ; orthogonality to regressors) from distributional assumptions, and only claim normality if supported empirically (e.g., Q–Q plot) and if modeling assumptions warrant it.

9. Keywords are generic and not well aligned with the scientific content (Abstract).

*Recommendation:* Use domain-relevant keywords (e.g., cognitive aging, Egyptian fruit bat, epigenetic clock, DNAmAge, diffusion tensor imaging, white matter, cognitive flexibility, resilience).

## Very minor issues

1. Formatting/typographical inconsistencies appear throughout (e.g., stray ‘#’ in headings; inconsistent variable naming like “Global\_FA” vs “Global FA”; inconsistent unit formatting; caption artifacts) (Sec. 1–4; Figure captions; Table 2).

*Recommendation:* Perform a careful copy-editing pass to standardize headings, variable notation, math/unit formatting, and remove stray characters/numbers from captions/tables.

2. Regression table internal consistency/rounding issues are present (e.g., coefficient/SE not matching the printed  $t$ -statistic for DNAmAge) (Table 2; Sec. 3.4.1).

*Recommendation:* Regenerate Table 2 directly from the analysis output with consistent rounding/precision so that coef/SE/ $t/p$  values match exactly (within rounding).

3. Some figure presentation details reduce accessibility/clarity (missing panel labels, axis units, overplotting, colorblind-safe palette not specified) (Figures 1–7).

*Recommendation:* Add consistent panel labels (A/B/C...), axis units (e.g., DNAmAge in years), transparency/jitter for overplotting, and a colorblind-safe palette; include key summary stats ( $N$ , mean/SD) where helpful.

4. Neutral scientific tone could be improved by reducing rhetorical/subjective wording in a few places (Introduction/Sec. 1; Conclusion/Sec. 4).

*Recommendation:* Edit for concision and measured claims, especially where results are null/uncertain or dependent on exploratory indices.

5. Metadata/presentation placeholders noted in the unstructured report (e.g., author/affiliation formatting) would be inappropriate for submission if present in the manuscript.

*Recommendation:* Replace any placeholder author/affiliation/title formatting with the final journal-compliant information before resubmission.

## Key statements and references

- • Bats are an extraordinary mammalian order that exhibit exceptional longevity relative to body size and show resistance to many age-related diseases, with accumulating evidence indicating that they maintain robust physiological function throughout their extended lifespans, yet the specific neural and cognitive mechanisms underlying their resistance to age-related cognitive decline remain largely unexplored [11].
- *Reference(s):* [11]
- • DNA methylation age (DNAmAge) is used in this study as a robust and precise biomarker of biological aging, consistent with prior work demonstrating that epigenetic clocks based on DNA methylation patterns can accurately estimate biological age across individuals [11].
- *Reference(s):* [11]
- • Diffusion Tensor Imaging (DTI) metrics such as Fractional Anisotropy (FA) and Mean Diffusivity (MD) are employed here as global measures of white matter structural integrity, building on prior evidence that these diffusion-based indices provide crucial insights into white matter organization and microstructural health in the mammalian brain [11].
- *Reference(s):* [11]
- • The conventional paradigm in aging research, supported by extensive prior work, posits a tight, near-inevitable link between biological aging, brain structural integrity (including white matter degradation and neuronal loss), and cognitive performance, such that age-related cognitive decline is typically accompanied by observable structural brain changes across diverse mammalian species [11].

- *Reference(s)*: [11]

## Mathematical consistency audit

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

**Maths relevance:** substantial

The paper’s analytical content centers on (i) defining a new behavioral metric (CAE) from normalized perseverative error rates, (ii) fitting a multiple linear regression model predicting CAE from DNAmAge, sex, and global DTI metrics, and (iii) defining CSDI as the regression residual to quantify ‘better-than-expected’ cognition. There are no long derivations, but internal consistency hinges on correct formula specification (especially CAE), consistent cohort size ( $N$ ), and consistent use of residual properties.

### Checked items

1. ✓ **Perseverative error rates (STM/LTM) normalization** (Sec. 2.2.2, p.3)
  - **Claim:** Perseverative error rates are defined as error counts divided by total entries in the corresponding phase (Phase 2 for STM; Phase 3 for LTM).
  - **Checks:** definition consistency, range/bounds sanity check
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:** Total entries in a phase is a nonnegative integer., Perseverative error count is  $\leq$  total entries in that phase.
  - **Notes:** If error counts are a subset of entries, rates lie in  $[0, 1]$ , which is consistent with later interpretation of CAE as near 1 for good performance.
2. ⚠ **Division-by-zero convention for error rates** (Sec. 2.2.2, p.3)
  - **Claim:** If a bat made no entries in a phase, the corresponding error rate is assigned 0 (rather than undefined).
  - **Checks:** definition consistency, edge-case sanity check
  - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
  - **Assumptions/inputs:** A phase can have total entries = 0.
  - **Notes:** The convention is mathematically well-defined but conceptually ambiguous: it treats ‘no opportunity to err’ as ‘perfect performance,’ which can inflate CAE. The paper does not justify why 0 is the correct limit/definition rather than missingness.
3. ✗ **CAE aggregation formula (signs/parentheses)** (Sec. 2.2.2, p.3)
  - **Claim:** CAE combines STM and LTM perseverative error rates into a single score where higher CAE indicates fewer errors, with values closer to 1 indicating better performance.

- **Checks:** algebra/precedence check, monotonicity sanity check, range/bounds sanity check, notation consistency
  - **Verdict:** FAIL; confidence: medium; impact: critical
  - **Assumptions/inputs:**  $\text{Perseverative}_{\text{ErrorRate}_{\text{STM}}}$  and  $LTM$  are in  $[0, 1]$ .
  - **Notes:** The displayed equation is ambiguous and, under standard operator precedence, reads as  $\text{CAE} = 1 - \text{Rate}_{\text{STM}} + \text{Rate}_{\text{LTM}}/2$ . This increases as  $\text{Rate}_{\text{LTM}}$  increases, contradicting the verbal claim that fewer perseverative errors yield higher CAE. If the intended formula is  $\text{CAE} = 1 - (\text{Rate}_{\text{STM}} + \text{Rate}_{\text{LTM}})/2$ , the manuscript must add parentheses; otherwise the written equation is inconsistent with the interpretation and boundedness.
4. **⚠ CAE boundedness vs reported range** (Sec. 3.2, p.5–6 (CAE described as closer to 1 is better; range reported 0.46 to 1.00))
- **Claim:** CAE is a score closer to 1 for better performance and empirically falls within approximately  $[0, 1]$ .
  - **Checks:** range/bounds sanity check, consistency between definition and narrative
  - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
  - **Assumptions/inputs:** Error rates are between 0 and 1.
  - **Notes:** Boundedness in  $[0, 1]$  holds if  $\text{CAE} = 1 - (\text{Rate}_{\text{STM}} + \text{Rate}_{\text{LTM}})/2$ , but not necessarily if CAE is interpreted as  $1 - \text{Rate}_{\text{STM}} + \text{Rate}_{\text{LTM}}/2$  (which can exceed 1). Because the equation formatting is unclear, consistency with the reported  $[0.46, 1.00]$  range cannot be verified from the text alone.
5. **✓ CAE regression model specification** (Sec. 2.4.2, p.4; Sec. 3.4.1, p.7)
- **Claim:** A multiple linear regression predicts CAE from DNAmAge, Sex, Global\_FA, and Global\_MD.
  - **Checks:** symbol/definition consistency
  - **Verdict:** PASS; confidence: high; impact: moderate
  - **Assumptions/inputs:** Linear additive model with an intercept is used., Sex is encoded numerically ( $\text{Sex}_{\text{numeric}}$ ).
  - **Notes:** The model is consistently stated in Methods and Results with the same set of predictors.
6. **✓ Regression degrees of freedom consistency** (Table 2 (OLS Regression Results), Sec. 3.4.1, p.7)
- **Claim:** With 30 observations and 4 predictors, the residual degrees of freedom is 25 and model degrees of freedom is 4.
  - **Checks:** algebraic consistency (DoF)
  - **Verdict:** PASS; confidence: high; impact: minor

- **Assumptions/inputs:** An intercept term is included.
  - **Notes:**  $n = 30$ ; parameters = 1 (intercept) +4 predictors = 5; residual DoF =  $30 - 5 = 25$ , matching the table.
7. ✓ **CSDI definition as residual** (Sec. 2.4.2, p.4; Sec. 3.4.2, p.8)
- **Claim:** CSDI is defined as  $\text{CSDI} = \text{CAE}_{\text{observed}} - \text{CAE}_{\text{predicted}}$  (residual from the CAE regression model).
  - **Checks:** definition consistency, sign convention check
  - **Verdict:** PASS; confidence: high; impact: critical
  - **Assumptions/inputs:**  $\text{CAE}_{\text{predicted}}$  denotes the fitted value from the regression model (including intercept).
  - **Notes:** The sign convention aligns with the stated interpretation: positive residual means better-than-predicted CAE.
8. ✓ **Mean of CSDI residuals approximately zero** (Sec. 3.4.2, p.8)
- **Claim:** CSDI distribution is centered around zero (mean reported  $\sim 4.67e - 15$ ), as expected for model residuals.
  - **Checks:** property check (OLS residual mean)
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** OLS includes an intercept and residuals are computed on the same data used for fitting.
  - **Notes:** With an intercept, OLS residuals sum to zero (up to floating-point rounding), consistent with the reported mean near 0.
9. △ **Normality expectation for residuals** (Sec. 3.4.2, p.8)
- **Claim:** Residuals (CSDI) are ‘expected’ to be approximately normally distributed.
  - **Checks:** logical/statistical assumption check
  - **Verdict:** UNCERTAIN; confidence: high; impact: minor
  - **Assumptions/inputs:** Normality is not an OLS identity; it requires additional assumptions about the error term.
  - **Notes:** Mean-zero is expected; normality is not guaranteed from the regression definition alone. The manuscript does not state assumptions that would justify ‘expected normality’.
10. ✘ **Cohort size consistency across manuscript** (Sec. 2.1, p.2 ( $N = 28$ ); Sec. 3.1, p.5 ( $N = 30$ ); Table 2, p.7 (No. Observations: 30); Abstract, p.1 (30 individuals))
- **Claim:** The same final analysis cohort size is used throughout for the core models/metrics.
  - **Checks:** definition consistency, cross-section consistency
  - **Verdict:** FAIL; confidence: high; impact: critical

- **Assumptions/inputs:** CAE regression and CSDI are computed on the stated final cohort.
  - **Notes:** Methods states final cohort  $N = 28$ ; later Results and the regression table use  $N = 30$ . This is an internal inconsistency that affects model definition and interpretability.
11. **△ Regional modeling plan vs stated technical limitation** (Sec. 2.4.1–2.4.3, p.4–5 (regional analyses described); Sec. 3.3, p.7 (regional extraction could not be performed))
- **Claim:** The manuscript’s mathematical/statistical plan is consistent with what is later stated as actually executed.
  - **Checks:** internal workflow consistency
  - **Verdict:** UNCERTAIN; confidence: high; impact: moderate
  - **Assumptions/inputs:** If regional metrics are unavailable, region-wise models cannot be fit.
  - **Notes:** Results state regional extraction/analysis could not be performed, but Methods still present full region-wise regression frameworks (Structural Preservation Hotspots;  $\text{CSDI} \sim \text{regional metrics}$ ). This is not an algebraic error, but it is an internal analytic inconsistency unless clearly marked as ‘planned but not executed’.

### Limitations

- The PDF provides few explicitly typeset equations and no equation numbering; the CAE formula appears to suffer from typesetting/precedence ambiguity that cannot be resolved without a clearly parenthesized expression.
- This audit does not assess numerical plausibility of reported parameter values, DTI metric magnitudes, or statistical significance computations, per scope.

## Numerical results audit

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

Out of 18 automated numeric consistency checks, 15 passed and 3 failed. Passes include exact cohort accounting within the Results section (41 initial minus 11 excluded equals  $N = 30$ ) and multiple internal OLS table consistency checks (df,  $R^2$ -to-percent, F-test  $p$ -value, most  $t$ -stat calculations, and adjusted  $R^2$ ). Failures are concentrated in cross-section consistency between Methods and Results (cohort  $N$ /sex/colony counts and DNAmAge summary) plus one coefficient  $t$ -stat mismatch for DNAmAge in Table 2.

### Checked items

1. ✓ **C1\_cohort\_exclusions\_count** (p.5, Results §3.1)

- **Claim:** Initial pool of 41 individuals with 11 excluded due to missing DTI, forming final cohort  $N = 30$ .
  - **Checks:** parts\_vs\_total
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{final\_N} == \text{initial\_N} - \text{excluded\_missing\_DTI}$ .
2. ✓ **C2\_sex\_counts\_sum\_to\_N\_results** (p.5, Results §3.1; Fig.2 caption text)
- **Claim:** Final cohort comprised 20 males and 10 females ( $N = 30$ ).
  - **Checks:** parts\_vs\_total
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{males} + \text{females} == \text{total\_N}$ .
3. ✓ **C3\_colony\_counts\_sum\_to\_N\_results** (p.5, Results §3.1; Fig.2 caption text)
- **Claim:** Equal distribution of 15 bats from Aseret and 15 from Herzeliya ( $N = 30$ ).
  - **Checks:** parts\_vs\_total
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{aseret} + \text{herzeliya} == \text{total\_N}$ .
4. ✓ **C4\_age\_range\_ordering\_and\_inclusion\_results** (p.5, Results §3.1; Fig.2 caption text)
- **Claim:** DNAmAge range reported as 6.62 to 13.84 years, with mean  $9.43 \pm 1.62$  years.
  - **Checks:** range\_sanity\_check
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{age\_min} < \text{age\_mean} < \text{age\_max}$ .
5. ✓ **C5\_CAE\_range\_ordering\_and\_bounds** (p.5-6, Results §3.2)
- **Claim:** CAE mean  $0.77 \pm 0.15$  with scores ranging from 0.46 to 1.00; CAE closer to 1 indicates greater efficiency.
  - **Checks:** range\_sanity\_check
  - **Verdict:** PASS
  - **Notes:** Checked  $0 \leq \{\text{rm CAE}\}_{\min} \leq \{\text{rm CAE}\} \leq 1$  (with  $\text{abs\_tol}$ ).  $\leq \{\text{rm CAE}\}_{\max}$
6. ✓ **C6\_global\_FA\_within\_theoretical\_bounds** (p.6, Results §3.3)
- **Claim:** Mean Global FA reported as  $0.991 \pm 0.003$ ; FA theoretical maximum is 1.0 and ranges 0 to 1.
  - **Checks:** unit\_consistent\_numeric\_comparison
  - **Verdict:** PASS

- **Notes:** Checked  $FA_{\min}^{\text{theoretical}} \leq \text{Global}_F A_{\text{mean}} \leq FA_{\max}^{\text{theoretical}}$ .
7. ✓ **C7\_regression\_df\_residuals\_check** (p.7, Table 2 (OLS Regression Results))
- **Claim:** No. Observations: 30, Df Model: 4, Df Residuals: 25.
  - **Checks:** df\_consistency
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{df\_resid\_reported} == n\_obs - \text{df\_model} - 1$  (intercept included).
8. ✓ **C8\_regression\_R2\_percentage\_conversion** (p.7, Table 2 and paragraph below)
- **Claim:** Model accounted for only 16.4% of the variance in CAE ( $R^2 = 0.164$ ).
  - **Checks:** percentage\_conversion
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{percent\_reported} = 100 \times R^2$  (rounding allowed).
9. ✓ **C9\_regression\_F\_pvalue\_internal\_consistency** (p.7, Table 2 (OLS Regression Results))
- **Claim:** F-statistic: 1.229; Prob (F-statistic): 0.324 with df\_model= 4 and df\_resid= 25.
  - **Checks:** recompute\_p\_value\_from\_F
  - **Verdict:** PASS
  - **Notes:** Computed  $p = \text{sf}(F_{\text{stat}}; \text{df}_1, \text{df}_2)$  and compared to reported.
10. ✗ **C10\_coefficient\_t\_stat\_consistency\_DNAge** (p.7, Table 2)
- **Claim:** DNAge coef 0.0309, std err 0.018,  $t = 1.745$ .
  - **Checks:** t\_stat\_equals\_coef\_over\_se
  - **Verdict:** FAIL
  - **Notes:** Reported  $t$  does not match coef/std\_err within tolerance.
11. ✓ **C11\_coefficient\_t\_stat\_consistency\_Sex** (p.7, Table 2)
- **Claim:** Sex<sub>numeric</sub> coef 0.0548, std err 0.060,  $t = 0.914$ .
  - **Checks:** t\_stat\_equals\_coef\_over\_se
  - **Verdict:** PASS
  - **Notes:** Checked  $t_{\text{reported}} \approx \text{coef}/\text{std}_{\text{err}}$ .
12. ✓ **C12\_coefficient\_t\_stat\_consistency\_Global\_FA** (p.7, Table 2)
- **Claim:** Global<sub>F</sub>A coef  $-8.0668$ , std err 9.988,  $t = -0.808$ .
  - **Checks:** t\_stat\_equals\_coef\_over\_se
  - **Verdict:** PASS

- **Notes:** Checked  $t_{reported} \approx \text{coef}/\text{std}_{err}$ .
13. ✓ **C13\_coefficient\_t\_stat\_consistency\_Global\_MD** (p.7, Table 2)
- **Claim:** Global<sub>MD</sub> coef  $-1518.5224$ , std err  $2936.843$ ,  $t = -0.517$ .
  - **Checks:**  $t\_stat\_equals\_coef\_over\_se$
  - **Verdict:** PASS
  - **Notes:** Checked  $t_{reported} \approx \text{coef}/\text{std}_{err}$ .
14. ✓ **C14\_adjusted\_R2\_consistency** (p.7, Table 2)
- **Claim:**  $R^2 : 0.164$ ; Adj.  $R^2 : 0.031$ ; No. Observations: **30**; Df Model: **4**.
  - **Checks:**  $recompute\_adjusted\_R2$
  - **Verdict:** PASS
  - **Notes:** Recomputed adjusted  $R^2 = 1 - (1 - R^2) \frac{n-1}{n-p-1}$ .
15. ✓ **C15\_CSDI\_mean\_near\_zero** (p.8, Results §3.4.2)
- **Claim:** CSDI scores centered at mean of zero (Mean =  $4.67e - 15$ ).
  - **Checks:**  $near\_zero\_check$
  - **Verdict:** PASS
  - **Notes:** Checked  $|CSDI_{mean}| \leq abs_{tol}$ .
16. ✓ **C16\_CSDI\_range\_ordering\_and\_inclusion\_of\_zero** (p.8, Results §3.4.2)
- **Claim:** CSDI range from  $-0.36$  to  $0.21$ ; distribution around zero.
  - **Checks:**  $range\_sanity\_check$
  - **Verdict:** PASS
  - **Notes:** Checked  $CSDI_{min} \leq CSDI_{mean} \leq CSDI_{max}$  and  $CSDI_{min} < 0 < CSDI_{max}$ .
17. ✗ **C17\_methods\_vs\_results\_N\_inconsistency\_flag** (p.2, Methods §2.1 vs p.5, Results §3.1)
- **Claim:** Methods report final cohort  $N = 28$  (17 males, 11 females; 16 Aseret, 12 Herzeliya) but Results report final cohort  $N = 30$  (20 males, 10 females; 15 Aseret, 15 Herzeliya).
  - **Checks:**  $cross\_section\_numeric\_consistency$
  - **Verdict:** FAIL
  - **Notes:** Strict equality check found mismatches across  $N/\text{sex}/\text{colony}$  counts (maximum absolute subgroup/count difference = 3).
18. ✗ **C18\_methods\_age\_summary\_vs\_results\_age\_summary** (p.2, Methods §2.1 vs p.5, Results §3.1)
- **Claim:** Methods: DNAMAge Mean  $\pm$  SD  $9.77 \pm 1.68$  years (Range 6.62–13.84); Results: mean  $9.43 \pm 1.62$  years (Range 6.62–13.84).

- **Checks:** cross\_section\_numeric\_consistency
- **Verdict:** FAIL
- **Notes:** Methods vs Results mean differs by **0.34** years and SD differs by **0.06** years, exceeding tolerance for cross-section consistency.

### **Limitations**

- Only parsed text from the provided PDF pages was used; no external data or assumptions were introduced.
- No candidate check relies on extracting numeric values from plot graphics; figures are treated as illustrative only.
- Many desirable validations (e.g., recomputing CAE, global DTI summaries, regression outputs, CSDI distribution) are not feasible without subject-level data tables, which are not present in the PDF.