

# *Skeptical review: Brain Structural Preservation in Long-Lived Bats: An Epigenetic Investigation of *Rousettus aegyptiacus**

---

## Summary

This manuscript tests whether DNA methylation (DNAm) epigenetic age is associated with global brain structure in the long-lived Egyptian fruit bat (*Rousettus aegyptiacus*). Using an existing species-specific DNAm clock derived from skin (Sec. 2.2.1) and DTI-derived total brain volume (TBV) computed from skull-stripped B0 images (Sec. 2.2.3), the authors analyze **33** bats with complete epigenetic and imaging data (from an initial cohort of **41**; Sec. 2.1, 3.1). The primary analysis is an OLS regression of standardized TBV on standardized DNAm age with sex and origin colony as covariates (Sec. 2.4–2.4.1, 3.2). The reported association is null ( $\beta \approx 0.0073$ ,  $p \approx 0.968$ ; adjusted  $R^2 \approx 0.017$ ; Sec. 3.2), and residual diagnostics indicate strong deviations from normality and likely influential points (Sec. 3.3, Fig. 5). The paper’s broader motivating aim—to relate epigenetic age, brain structure, and spatial-memory performance—cannot be addressed because behavioral outcomes could not be extracted due to data-parsing failures (Sec. 2.2.2, 3.1). The question and dataset are promising for comparative biogerontology, but the current version reads as a preliminary feasibility/progress report: the imaging phenotype (TBV) requires stronger validation, inference should be more effect-size/precision-centered and robust to assumption violations, and the framing should be tightened to match what is actually tested given missing behavioral results and under-specified cohort/design details.

## Strengths

- Addresses a timely comparative aging question in a non-traditional, exceptionally long-lived mammalian model (*Rousettus aegyptiacus*), which is underrepresented in the brain-aging literature (Sec. 1).
- Leverages a species-specific DNAm clock (Sec. 2.2.1), which is conceptually appropriate for studying aging biology beyond chronological age alone.
- Assembles a valuable multimodal dataset (epigenetics + neuroimaging + intended behavior) in a challenging species, and provides a clear high-level description of the TBV computation approach (Sec. 2.2.3).
- Reports the core regression outputs (coefficients, p-values, model F-test, adjusted  $R^2$ ) and includes diagnostic plots that appropriately raise concerns about OLS assumptions (Sec. 3.2–3.3; Figs. 3–5).
- The Discussion acknowledges several key limitations (e.g., missing behavioral outcomes; coarseness of TBV) and points toward more granular imaging phenotypes and improved methods in future work (Sec. 3.3, 4.4).

## Major issues

1. **Framing and interpretation overreach: the manuscript treats a null association** ( $\beta \approx 0.0073$ ,  $p \approx 0.968$ ; adjusted  $R^2 \approx 0.017$ ; Sec. 3.2) as evidence for “remarkable preservation”/slower atrophy with age (Sec. 1, 3.3, 4.3–4.4). With  $n = 33$ , cross-sectional data, an epigenetic-age range that is not contextualized relative to lifespan (Sec. 3.1), and assumption violations (Sec. 3.3), “no detectable association” is not equivalent to “preservation,” nor does it quantify an atrophy rate or support cross-species comparisons.

*Recommendation:* Reframe throughout (Abstract, Sec. 1, Sec. 3.3, Sec. 4.3–4.4) to state the main result as an absence of detectable association between DNAm age and TBV in this sample. Replace or qualify strong claims (e.g., “remarkable preservation,” “supports our hypothesis”) with effect-size/precision language. Report the unstandardized DNAm-age slope ( $\text{mm}^3$  per DNAm-year) and a 95% CI alongside the standardized  $\beta$  (Sec. 3.2), and interpret what effect sizes remain compatible with the CI. Add a short “precision/power” paragraph (Sec. 3.3 or 4.3) describing minimum detectable effects given the observed TBV variance and age range. If you want to make a preservation-style claim, consider an equivalence/interval-null framing (define a smallest effect size of interest such as  $\pm X\%$  TBV per year, and discuss whether the CI excludes it). Explicitly note that absence of evidence is not evidence of absence of atrophy and that longitudinal data would be required to estimate within-individual change.

2. **Statistical inference is fragile given clear OLS assumption violations and outliers** (Sec. 3.3; Fig. 5). The manuscript documents non-normal residuals and apparent influential points but still relies on standard OLS p-values/CIs as the primary basis for inference (Sec. 3.2–3.3), without robust SEs, robust regression, bootstrapping/permutation, or pre-specified sensitivity analyses.

*Recommendation:* Strengthen Sec. 2.4 (analysis plan) and Sec. 3.2–3.3 (results) with robustness checks: (i) report heteroskedasticity-robust SEs (e.g., HC3/HC4) for OLS; (ii) fit a robust regression (Huber or Tukey) and report the DNAm-age coefficient and CI; (iii) provide a nonparametric association (Spearman) and/or a permutation test for the DNAm-age slope; (iv) quantify influence (Cook’s distance/leverage/DFBETAs), define objective outlier criteria, and run sensitivity analyses with/without influential observations. Also check functional form (e.g., add a spline term or at least overlay LOESS in Fig. 4) to ensure a linear model is appropriate. Base the Discussion (Sec. 4.3–4.4) on the ensemble of robust/primary analyses rather than a single OLS p-value.

3. **Outcome definition/validity: TBV computed as (number of non-zero voxels in skull-stripped averaged B0)  $\times$  voxel volume** (Sec. 2.2.3) is unconventional as a volumetric phenotype and may be sensitive to preprocessing artifacts (skull-stripping behavior, implicit thresholding, intensity scaling),

**EPI distortions, within-brain zeros, resampling/geometry differences, and scan/session effects.** As written, it is unclear whether “non-zero voxels” is exactly equivalent to a binary brain mask, and whether all scans share identical voxel size and image dimensions.

*Recommendation:* In Sec. 2.2.3, explicitly define the binarization rule (e.g.,  $> 0$  vs  $> \epsilon$ ) and/or (preferably) compute TBV from an explicit binary brain mask rather than intensity non-zeros. Provide validation/QC: confirm (on all subjects or a representative subset) that non-zero counting matches mask-based volume; show example QC images (skull-stripping results) and report how failures were handled. Add acquisition and preprocessing details needed to judge comparability (scanner/field strength, sequence, voxel size, b-values, number of directions, resampling, distortion/motion correction, skull-stripping tool and parameters; Sec. 2.2.3). If scans differ by protocol/session, incorporate scan/batch covariates (or justify why not) and discuss potential bias. Consider adding a brief reliability/sensitivity check (e.g., TBV from first B0 vs averaged B0; sensitivity to threshold  $\epsilon$ ) to show the phenotype is stable.

- 4. Scope mismatch due to missing behavioral outcomes: the paper’s stated integrative aim (epigenetic age  $\leftrightarrow$  brain structure  $\leftrightarrow$  spatial memory) is central in the Abstract/Introduction (Sec. 1) but cannot be evaluated because behavioral metrics could not be extracted (Sec. 2.2.2, 3.1).** As a result, the current manuscript is primarily a single association test (DNAm age vs TBV), and the cognitive narrative currently overpromises relative to delivered analyses.

*Recommendation:* Align framing with actual content. Revise title/Abstract (and early Sec. 1) to position this as a preliminary/feasibility analysis of DNAm age vs global TBV, with behavioral analyses explicitly out of scope due to parsing failure. In Sec. 2.2.2 and 3.1, quantify the behavioral-data failure (how many files/animals affected, what exactly broke, what was attempted) and move detailed behavioral metric definitions to an Appendix to reduce distraction. In Sec. 4.1–4.4, clearly separate conclusions supported by current data (structural association test) from planned future work (cognition linkage), and outline a concrete recovery plan (schema validation, header harmonization rules, semi-manual extraction, unit tests for parsers) or justify why recovery is not feasible.

- 5. Epigenetic age measure is under-described and not calibrated/benchmarked within this cohort (Sec. 2.2.1, 3.1).** The DNAm clock is described as “previously established and validated,” but key details are missing (training set, tissue(s), CpG count, accuracy/MAE, age range, normalization/QC, implementation). Additionally, if chronological age is known, the manuscript does not report DNAm–chronological correspondence or consider age-acceleration formulations; if chronological age is unknown, this limitation should be explicit because it affects interpretability of “advanced epigenetic age.”

*Recommendation:* Expand Sec. 2.2.1 with full clock provenance: cite the clock paper/tool, describe training data (tissue, age range,  $n$ ), number of CpGs, reported performance (MAE/correlation), and your pipeline (normalization/QC/software versions). Clarify whether DNAm ages were computed anew or imported. If chronological ages exist, report DNAm age vs chronological age correlation and consider (at least in supplement) models using chronological age and/or DNAm age acceleration (DNAm residuals controlling for chronological age). If chronological ages do not exist, explicitly state this and temper interpretations of DNAm age as “biological age,” noting the additional uncertainty (including that the clock is trained on skin, not brain, which may limit inferences about brain aging).

6. **Cohort/design and missingness are under-specified, limiting interpretability and raising potential bias concerns (Sec. 2.1, 3.1, 3.3).** Eight of 41 bats are excluded due to missing imaging/TBV, but reasons for missingness and whether it relates to age/sex/colony are not documented. The age distribution is not contextualized relative to species lifespan, and potential confounders strongly tied to brain volume (body mass/head size, health status, scan/session effects) are not addressed beyond sex and origin colony (Sec. 2.4, 3.2).

*Recommendation:* In Sec. 2.1 and 3.1, document sampling/inclusion criteria, captive vs wild-derived status, and why 8/41 lacked usable imaging (scan failure, QC exclusion, preprocessing failure). Provide a missingness check: compare DNAm age/sex/colony between included vs excluded animals. Contextualize the DNAm-age range (6.62–15.07 years; Sec. 3.1) against typical/maximum lifespan and life stage in *R. aegyptiacus*. If available, add morphometrics (mass, forearm length, head size) and relevant health indices as covariates or at minimum report them and discuss as limitations; likewise, clarify whether all DTI data were acquired under a single protocol/session and, if not, incorporate batch/session/scanner covariates.

7. **Reporting/reproducibility gaps: key descriptive statistics, full model outputs, and methodological details needed to evaluate and reproduce results are incomplete (Sec. 2.2.1, 2.2.3, 2.3.2, 3.1–3.2).** Current reporting emphasizes standardized  $\beta$  and a p-value, but does not present a full regression table (coefficients/SEs/CIs), complete descriptives for the  $n = 33$  subset, or clear statements about standardization conventions and reference categories.

*Recommendation:* Add (main text or supplement): (i) a descriptive table for the  $n = 33$  analysis subset (mean/SD/median/min/max for DNAm age and TBV; counts by sex/colony); (ii) the full regression table including unstandardized and standardized coefficients, SEs, 95% CIs, p-values, and model fit metrics; (iii) a clear definition of z-scoring (computed over which sample; sample SD vs population SD; Sec. 2.4), and reference categories for  $C(\text{Sex})$  and  $C(\text{Origin\_colony})$  (Sec. 2.4.1). Include a brief code/data availability statement and QC criteria summary (even if data are restricted).

## Minor issues

1. Figure set around behavioral failure is redundant and potentially misleading: Figures 1 and 2 both visualize the all-zero behavioral outcome and include inferential-looking elements despite degenerate data (Sec. 3.1). Several figures also suffer from low resolution/legibility and inconsistent terminology/units (Figs. 1–5).

*Recommendation:* Merge Figures 1 and 2 into one high-resolution diagnostic (or a concise table) that documents the parsing failure, sample sizes, and completeness—without regression lines. For Figures 3–5, improve legibility (font sizes/export quality), standardize axis labels/units and variable names, add  $n$  and key summaries in captions, and soften caption language to descriptive statements rather than claims of “confirmation” (Sec. 3.2–3.3). Consider annotating influential points and adding robust-fit overlays if you implement robustness analyses.

2. Model specification/notation inconsistencies and typos create ambiguity about the fitted model: mismatched formulas across Methods vs Results (Sec. 2.4.1 vs 3.2), stray symbols (e.g., “ $\hat{+}$ ”), and confusion about standardizing the response vs predictors (Sec. 2.4). Coefficient symbol inconsistency (“ff” vs  $\beta$ ) appears in the Abstract/Results.

*Recommendation:* Provide one definitive model formula in Sec. 2.4/2.4.1 (explicitly stating whether TBV and/or DNAm age are standardized) and use the same notation in Sec. 3.2. Fix the “ff” symbol in the Abstract to  $\beta$  (and define  $\beta$ ). Clarify standardization wording to distinguish predictors from response (e.g., DNAmAge\$*text{std}* predictor; TBV*text{std}*\$ response, if that is what was done).

3. Covariate rationale is not well motivated (Sec. 2.4). Sex and origin colony are included, but there is no explanation for why these were chosen over (or in addition to) other plausible confounders, and no discussion of potential scan/batch effects if applicable.

*Recommendation:* Add a brief justification in Sec. 2.4 for covariate selection, and explicitly list other candidate covariates that were unavailable or incomplete. If scan protocol/batch varies, state how this was handled (or acknowledge as a limitation) and consider adding batch covariates where feasible.

4. Related-work context is thin for readers outside the immediate subfield (Sec. 1). The manuscript would benefit from a clearer synthesis of (i) typical mammalian brain atrophy patterns, (ii) what is known about bat longevity/brain aging, (iii) prior neuroimaging in bats, and (iv) epigenetic clocks in long-lived mammals.

*Recommendation:* Add a compact related-work paragraph/section in Sec. 1 that situates this study relative to existing evidence and clarifies novelty (e.g., first DNAm-age vs global TBV test in this species; limitations due to cross-sectional design and coarse phenotype).

5. Editorial/content hygiene issues reduce professionalism and discoverability: the keyword list contains unrelated astronomy terms, and the affiliation/authorship line contains placeholder text (unstructured report notes: “Anthropic, Gemini & OpenAI servers. Planet Earth.”).

*Recommendation:* Replace keywords with relevant terms (epigenetic aging, DNA methylation clock, DTI, brain volume, bat longevity, comparative aging). Replace placeholder affiliations with correct institutional information per journal requirements.

6. Ethics/animal welfare approvals and permits are not explicitly documented (Sec. 2.1–2.2), which is typically required for vertebrate animal research.

*Recommendation:* Add a short ethics statement subsection with IACUC (or equivalent) approvals/permit numbers and relevant guidelines for capture/housing/imaging/tissue sampling, consistent with journal policy.

## Very minor issues

1. Terminology and naming are inconsistent across text/figures (DNAm\_Age vs DNAmAge; TBV vs TotalBrainVolume vs TotalBrain\_Volume; standardized forms inconsistently named), and units/p-value formatting vary (Secs. 2.2–3.3).

*Recommendation:* Adopt canonical variable names (e.g., TotalBrainVolume\_mm<sup>3</sup>; TotalBrainVolume\_std; DNAmAge\_years; DNAmAge\_std) and use them consistently throughout. Standardize unit formatting (mm<sup>3</sup>) and p-value style to the journal’s conventions.

2. TBV definition depends on an implicit assumption that non-brain voxels are exactly zero and brain voxels non-zero; the thresholding/masking rule is not explicitly stated (Sec. 2.2.3).

*Recommendation:* State the exact criterion used to classify voxels as brain (mask-based preferred; otherwise specify threshold  $\epsilon$  and how it was chosen) and confirm that image preprocessing does not introduce within-brain zeros that would bias volume.

3. Minor formatting/ $\LaTeX$ /heading issues: malformed  $\LaTeX\{\}$  for species formatting in the title, stray line breaks, inconsistent heading styles, and occasional overly interpretive figure captions (Secs. 1, 3.2–3.3, 4).

*Recommendation:* Proofread for  $\LaTeX\{\}$  correctness (italicize species names consistently), remove stray breaks, normalize heading formatting, and keep captions primarily descriptive (reserve interpretation for the main text).

## Key statements and references

- • Epigenetic age for each bat was determined using a previously established and validated DNA methylation clock specific to *Rousettus aegyptiacus*, based on high-throughput methylation profiling of skin tissue CpG sites, with the resulting ‘DNAmAgeBat.Rousettus.aegyptiacus\_Skin’ values taken directly from existing metadata.
- *Reference(s)*: (none)
- • The study’s central hypothesis and interpretation rely on prior evidence that *Rousettus aegyptiacus* is an exceptionally long-lived mammalian species whose maximum lifespan substantially exceeds that predicted by its body size and that it shows resistance to various age-related pathologies, including those affecting the brain.
- *Reference(s)*: (none)
- • The use of DNA methylation-based epigenetic clocks as accurate estimators of biological age that can deviate from chronological age, thereby reflecting an individual’s true biological aging rate, is grounded in prior work establishing such clocks from specific CpG methylation patterns.
- *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 limited formal mathematics; the main analytic content is an OLS regression model relating epigenetic age to total brain volume with categorical covariates, and a procedural definition of TBV from voxel counts and voxel dimensions. The primary consistency issues are around notation/model specification alignment and clarity about which variables are standardized, plus an implicit assumption in the TBV voxel-count definition.

### Checked items

1. ✓ **TBV units from voxel volume  $\times$  voxel count** (Sec. 2.2.3, p.3)
  - **Claim:** Total brain volume (TBV) is computed as the number of non-zero voxels in an averaged 3D B0 image multiplied by voxel volume derived from header zooms, yielding TBV in  $\text{mm}^3$ .
  - **Checks:** units/dimensions, definition consistency
  - **Verdict:** PASS; confidence: high; impact: moderate

- **Assumptions/inputs:** Header zooms provide voxel side lengths in mm for the three spatial axes., Non-brain voxels are exactly zero after skull stripping; brain voxels are non-zero., Counting is applied to the 3D averaged B0 image.
  - **Notes:** Dimensionally consistent:  $(\text{count}) \times (\text{mm} \times \text{mm} \times \text{mm}) = \text{mm}^3$ . The arithmetic definition is coherent given the stated assumptions.
2. **△ Averaging B0 volumes preserves zero background** (Sec. 2.2.3, Step 2–4, p.3)
- **Claim:** Averaging the first three B0 volumes and then counting non-zero voxels correctly captures brain volume because non-brain tissue voxels have intensity zero.
  - **Checks:** logical implication, definition consistency
  - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
  - **Assumptions/inputs:** Background voxels are exactly zero in each of the three B0 volumes., Averaging does not introduce non-zero values outside the brain.
  - **Notes:** If the background is exactly zero in each volume, the average remains zero; however, the paper does not specify a mask/threshold or prove that skull stripping enforces exact zeros everywhere outside brain, which is required for the “non-zero voxel count” definition to be robust.
3. **✘ Standardization role confusion (predictor vs response)** (Sec. 2.4, p.4)
- **Claim:** “Continuous predictor variables (DNAm\_Age, TotalBrain\_Volume) were standardized (z-scored) before entering them into the statistical models.”
  - **Checks:** symbol/role consistency, definition consistency
  - **Verdict:** FAIL; confidence: high; impact: moderate
  - **Assumptions/inputs:** Primary model uses TBV as outcome and DNAm-Age as predictor.
  - **Notes:** TotalBrain\_Volume is not a predictor in the stated primary model; it is the response. This is an internal role/definition inconsistency that affects coefficient interpretation clarity.
4. **✘ Model formula consistency: Methods vs Results** (Methods: Sec. 2.4.1, p.4; Results: Sec. 3.2, p.5)
- **Claim:** The primary OLS model is specified as  $\text{TotalBrainVolume} \sim \text{DNAm-Age} + \text{Sex} + \text{Origin\_colony}$ , but Results describe fitting  $\text{TotalBrainVolume\_std} \sim \text{DNAmAge\_std} + \text{C}(\text{Sex}) + \text{C}(\text{Origin\_colony})$ .
  - **Checks:** notation consistency, model specification consistency
  - **Verdict:** FAIL; confidence: high; impact: moderate
  - **Assumptions/inputs:** Standardization is applied as stated., Sex and Origin\_colony are treated as categorical factors with dummy coding.

- **Notes:** The paper presents two different model specifications (raw vs standardized response/predictor; explicit categorical encoding appears only in Results). This does not prove the analysis is wrong, but it is an internal inconsistency in the mathematical/statistical description.
5. ✓ **Degrees of freedom of the overall F-test** (Sec. 3.2, p.5)
- **Claim:** Overall model test reported as  $F(3, 29)$  for  $n = 33$ .
  - **Checks:** df/parameter count, internal consistency
  - **Verdict:** PASS; confidence: high; impact: minor
  - **Assumptions/inputs:** Predictors are DNAmAge (1 df) + Sex (2 levels → 1 df) + Origin\_colony (2 levels → 1 df)., An intercept is included.
  - **Notes:** With intercept + 3 regressors,  $df_{model} = 3$  and  $df_{resid} = 33 - 4 = 29$ , consistent with  $F(3, 29)$ . If either categorical had  $> 2$  levels, this would break; the text indicates exactly two colonies and binary sex, so it matches.
6. ✓ **Interpretation of standardized coefficient  $\beta$**  (Sec. 3.2, p.5)
- **Claim:**  $\beta = 0.0073$  is interpreted as the change in standardized TBV per one standard deviation increase in epigenetic age, holding covariates constant.
  - **Checks:** interpretation consistency, definition consistency
  - **Verdict:** PASS; confidence: medium; impact: minor
  - **Assumptions/inputs:** Both TBV and DNAmAge were z-scored (standardized) in the fitted model., OLS includes covariates Sex and Origin\_colony.
  - **Notes:** Given the stated standardized model, this interpretation is consistent. The remaining ambiguity is whether the model indeed used standardized TBV (conflicts with Methods), so the pass is conditional on the Results model being the actual fitted one.
7. ✗ **Unstandardized interpretation statement in Methods** (Sec. 2.4.1, p.4)
- **Claim:** The DNAm\_Age coefficient “quantifies the average change in total brain volume per unit increase in epigenetic age”.
  - **Checks:** interpretation consistency
  - **Verdict:** FAIL; confidence: high; impact: minor
  - **Assumptions/inputs:** Earlier text states variables were standardized (z-scored) before modeling.
  - **Notes:** If TBV and/or DNAmAge are standardized, the coefficient is not “per unit year” and not in  $\text{mm}^3$ ; it is in SD units. This clashes with the later standardized-model description and could mislead readers about units of the effect.
8. ✗ **Coefficient symbol in abstract ( $\text{ff}$ ) vs  $\beta$  in Results** (Abstract, p.1; Sec. 3.2, p.5)

- **Claim:** Abstract reports “ $\beta = 0.0073$ ” while Results report “ $\beta = 0.0073$ ”.
- **Checks:** notation consistency
- **Verdict:** FAIL; confidence: high; impact: minor
- **Assumptions/inputs:** Only one primary slope is being discussed.
- **Notes:** “ $\beta$ ” is undefined and inconsistent with “ $\beta$ ” used elsewhere; likely a typographical/notation error that should be corrected for mathematical clarity.

9. ✓ **Significance criterion statement** (Sec. 2.4.1, p.4)

- **Claim:** Two-sided p-value  $< 0.05$  is used as the significance threshold.
- **Checks:** definition consistency
- **Verdict:** PASS; confidence: high; impact: minor
- **Notes:** No internal contradictions observed; later p-values are reported in the same two-sided convention (not contradicted elsewhere).

10. ✓ **Categorical covariate encoding clarity** (Sec. 3.2, p.5)

- **Claim:** Model uses  $C(\text{Sex})$  and  $C(\text{Origin\_colony})$ , implying categorical encoding.
- **Checks:** model specification consistency
- **Verdict:** PASS; confidence: medium; impact: minor
- **Assumptions/inputs:** Sex and Origin\_colony each have exactly two categories as stated in Sec. 2.1.
- **Notes:** The categorical encoding is consistent with the stated data structure (two colonies, binary sex). The paper does not specify reference categories; that affects interpretation of those coefficients but not internal consistency.

## Limitations

- The PDF contains very few explicit equations; most mathematical content is described verbally, limiting step-by-step derivation checks.
- No explicit z-score formula is provided, so verification of standardization conventions is limited to consistency checks of interpretation rather than formal derivation.
- The TBV computation relies on implicit image-mask properties (exact zeros outside brain) that are asserted but not formally defined; this prevents full analytic verification of the TBV definition without additional procedural detail.

## Numerical results audit

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

10 numeric checks were executed: 9 PASS and 1 FAIL. Passes include cohort composition sums (colony and sex), repeated subset size (33 of 41) consistency, OLS df consistency for  $F(3, 29)$  with  $n = 33$ , range checks for epigenetic age and TBV means within min/max, recomputed

endpoint differences (age range and TBV range), repeated DNAmAge coefficient/p-value consistency, and significance-threshold comparisons for normality diagnostic p-values. One check failed due to an inconsistency in how F-statistic components were grouped for repeated-constant consistency (mixing F value and degrees of freedom).

### Checked items

1. ✓ **C1** (p.2 (Methods 2.1 Subject Cohort))
  - **Claim:** The study utilized **41** bats sourced from two colonies: Aseret ( $n = 24$ ) and Herzeliya ( $n = 17$ ).
  - **Checks:** parts\_vs\_total
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{sum}(\text{parts})$  vs total using total\_bats.
2. ✓ **C2** (p.2 (Methods 2.1 Subject Cohort))
  - **Claim:** The cohort comprised **18** females and **23** males (total **41**).
  - **Checks:** parts\_vs\_total
  - **Verdict:** PASS
  - **Notes:** Checked  $\text{sum}(\text{parts})$  vs total using total\_bats.
3. ✓ **C3** (p.1 Abstract; p.2 Methods 2.1; p.4 Results 3.1; p.7 Conclusions 4.2)
  - **Claim:** Multiple sections state imaging+epigenetic complete-data subset size is **33** out of **41**.
  - **Checks:** repeated\_constant\_consistency
  - **Verdict:** PASS
  - **Notes:** All repeated numeric constants consistent within tolerance. Also checked  $\text{subset} \leq \text{total}$ : OK.
4. ✓ **C4** (p.5 Results 3.2)
  - **Claim:** Overall OLS model:  $F(3, 29) = 1.185$  for  $n = 33$  implies residual  $df = n - (k \text{ predictors} + \text{intercept})$ .
  - **Checks:** df\_consistency
  - **Verdict:** PASS
  - **Notes:** Checked  $df_{resid} = n - (df_{model} + 1)$  assuming intercept.
5. ✓ **C5** (p.4 Results 3.1)
  - **Claim:** Epigenetic age range **6.62** to **15.07** years with mean **9.60** (SD **1.74**) is numerically consistent with being within the stated min/max.
  - **Checks:** range\_check
  - **Verdict:** PASS
  - **Notes:** Mean lies within stated min/max (with abs tolerance on bounds).
6. ✓ **C6** (p.4 Results 3.1)

- **Claim:** TBV range 4398.7 to 5201.4 mm<sup>3</sup> with mean 4851.3 mm<sup>3</sup> is numerically consistent with being within the stated min/max.
  - **Checks:** range\_check
  - **Verdict:** PASS
  - **Notes:** Mean lies within stated min/max (with abs tolerance on bounds).
7. ✓ **C7** (p.4 Results 3.1)
- **Claim:** Age range width is 15.07 – 6.62 years; TBV range width is 5201.4 – 4398.7 mm<sup>3</sup> (cheap recomputation from explicit numbers).
  - **Checks:** difference\_recomputation
  - **Verdict:** PASS
  - **Notes:** Recomputed differences from provided endpoints (no separate reported result to validate).
8. ✓ **C8** (p.5 Results 3.2; p.7 Conclusions 4.3; p.1 Abstract)
- **Claim:** Standardized coefficient for DNAmAge is reported as  $\beta = 0.0073$  and  $p = 0.968$  consistently across sections.
  - **Checks:** repeated\_constant\_consistency
  - **Verdict:** PASS
  - **Notes:** All repeated numeric constants consistent within tolerance.
9. ✗ **C9** (p.5 Results 3.2; p.7 Conclusions 4.3)
- **Claim:** Overall model statistics repeated:  $F(3, 29) = 1.185$ ,  $p = 0.333$ , adjusted  $R^2 = 0.017$ .
  - **Checks:** repeated\_constant\_consistency
  - **Verdict:** FAIL
  - **Notes:** Inconsistency in repeated constants for group(s): F.
10. ✓ **C10** (p.6 Results 3.3)
- **Claim:** Normality diagnostics: Omnibus test  $p = 0.000$  and Jarque-Bera  $p < 0.001$  are both consistent with a significance threshold of 0.05.
  - **Checks:** threshold\_comparison
  - **Verdict:** PASS
  - **Notes:** Checked p-values (JB treated as upper bound) against alpha.

### Limitations

- Only parsed text from the provided PDF pages was available; no numeric tables or full model output (e.g., coefficient SEs, t-stats,  $R^2$ ) were present to recompute many statistical quantities.
- No raw subject-level data are included in the PDF, preventing verification of descriptive statistics (means/SDs) and regression results beyond algebraic/df/range/repetition checks.

- Figures are present but numeric verification from plots is excluded (no pixel/value extraction).