

Skeptical review: Aging and Cognition in Long-Lived Egyptian Fruit Bats: Behavioral Performance and the Unmet Promise of Microstructural Biomarkers

Summary

The manuscript investigates cognitive aging and putative resilience in **32** long-lived Egyptian fruit bats by relating skin-derived DNA methylation age (DNAmAge) to performance in a multi-phase spatial foraging task (Introduction; Sec. 2.1–2.2; Sec. 3.1–3.2). The authors extract multiple behavioral measures (learning, short-/long-term memory, perseveration, visit efficiency), z -score and sign-align them, and aggregate them into a composite Cognitive Performance Index (CPI), then model CPI as a function of DNAmAge with sex and colony of origin as covariates (Sec. 2.2.2–2.2.3; Sec. 2.4.2; Sec. 3.2–3.4). A “Cognitive Resilience Score” is defined as the residual from $CPI \sim DNAmAge$ to represent age-adjusted performance (Sec. 2.4.3; Sec. 3.5–3.6). The originally advertised diffusion-MRI contribution—a proposed Normalized Directional Diffusion Variance (NDDV) microstructural metric—could not be executed because the available diffusion files were **3D** rather than the required **4D** multi-direction acquisitions, and an atlas file was missing (Sec. 2.1.1; Sec. 2.3; Sec. 3.3; Sec. 4.1–4.2). Within the available behavioral/epigenetic data, the authors report no statistically significant association between DNAmAge and CPI over the sampled epigenetic age range (Sec. 3.4), with substantial individual variability and a trend-level colony effect. The dataset and transparency about the imaging failure are valuable, but the paper currently reads partly as a project report with an unrealized central aim; strengthening the positioning, quantitative reporting, task/method detail, CPI validation, and uncertainty-focused interpretation of the null finding would substantially improve the scientific contribution and credibility of the “resilience” framing.

Strengths

- Addresses an interesting and underrepresented question—cognitive aging in a long-lived, cognitively sophisticated non-traditional model organism—using DNAmAge and an ecologically relevant spatial foraging paradigm (Introduction; Sec. 2.1–2.2).
- Transparent reporting of the diffusion-MRI data limitation (**3D** vs **4D**; missing atlas) and an effort to extract a generalizable “lessons learned” message about data integrity and QC (Sec. 2.3; Sec. 3.3; Sec. 4.1–4.4).
- Clear high-level pipeline: behavioral metrics \rightarrow standardized composite CPI \rightarrow GLM with DNAmAge/sex/origin \rightarrow residual-based age-adjusted score (Sec. 2.2–2.4; Sec. 3.2–3.6).
- The CPI and residual-based score are, in principle, simple to reproduce and could be useful phenotypes for future longitudinal or multimodal studies if properly validated and reported (Sec. 2.4.1–2.4.3).

- Figures use familiar plot types and attempt to show distributions and model relationships; the intent to include diagnostics and descriptive views is good (Figs. 2–7; Sec. 3.2–3.5).

Major issues

1. **Mismatch between advertised aims and delivered results: the title/Abstract/Introduction/Conclusions foreground a diffusion-MRI microstructural biomarker (NDDV) linking brain integrity, DNAmAge, and cognition, but NDDV cannot be computed because diffusion data are unusable (3D not 4D) and an atlas is missing (Sec. 1; Sec. 2.1.1; Sec. 2.3; Sec. 3.3; Sec. 4.1–4.2).** This creates a substantial positioning problem: readers may expect an imaging biomarker study, while the empirical contribution is behavioral–epigenetic only plus a data-QC cautionary note.

Recommendation: Reframe the manuscript explicitly around what is actually supported by the data: (i) behavioral phenotype extraction and CPI construction; (ii) effect-size/uncertainty bounds on DNAmAge–performance association in this cohort; (iii) heterogeneity (e.g., colony) and limitations; and (iv) a concise, actionable “DTI data QC” case study. Revise the title and Abstract to de-emphasize NDDV/microstructural biomarkers (or clearly label NDDV as conceptual/unimplemented). Move most NDDV derivation/pipeline detail to an Appendix and keep only a brief, internally consistent conceptual description in the main text (Sec. 1; Sec. 2.3; Sec. 3.3).

2. **Insufficient quantitative reporting and template/metadata problems undermine credibility and reproducibility: Table 1 is unfilled (Sec. 2.4.1/Sec. 3.1), Table 2 (GLM output) is referenced but missing, and at least one reported statistic appears as a placeholder/formatting error (e.g., adjusted R-squared “00.00” in Sec. 3.4).** The manuscript also contains clear template mismatches (astronomy-related keywords after the Abstract) and (per the unstructured report) non-scientific placeholder affiliation text; such issues can signal to readers that other parts may be similarly unvetted.

Recommendation: Fully populate Table 1 (N , mean, SD, min, max for DNAmAge and CPI; counts for Sex and Origin) and include Table 2 with complete regression output (coefficients, SEs, 95% CIs, test statistics, df, exact p -values, R^2 /adjusted R^2) (Sec. 2.4.1; Sec. 3.1; Sec. 3.4). Remove placeholder/incorrect values (e.g., “00.00”) and verify all reported statistics against the analysis output. Replace astronomy keywords with appropriate terms and ensure affiliations and other front-matter metadata are correct and publication-ready (Abstract/front matter).

3. **CPI construction is underspecified and insufficiently validated for its central role. CPI is an unweighted sum of z-scored components spanning heterogeneous types (continuous latencies, counts, and binary perseveration indicators), potentially across multiple phases, without demonstrating di-**

mensional coherence, reliability, or robustness (Sec. 2.2.2–2.2.3; Sec. 3.2). The exact CPI component set is not enumerated (e.g., whether Visit_Efficiency enters once or per phase), preventing replication and making the CPI scale hard to interpret (Sec. 2.4.1; Sec. 3.2).

Recommendation: In Sec. 2.2.3 and Sec. 3.2, explicitly list all CPI components (with phase indexing), indicate sign flips, and provide the CPI formula as a sum over a clearly defined metric set. Add a component-level table: distribution summaries (mean/SD/range), missingness, and directionality. Report a correlation matrix among components and at least one reliability/structure check (e.g., Cronbach’s α ; simple PCA/factor analysis) to justify aggregation. Provide sensitivity analyses showing whether the DNAmAge association remains similar when (i) excluding binary perseveration variables, (ii) using PCA-derived factor scores, and/or (iii) using domain sub-indices (learning/STM/LTM/efficiency) rather than a single CPI (Sec. 3.2–3.4).

4. **Behavioral task description is not sufficiently operational for readers outside the immediate project, limiting interpretability of CPI components and the cognitive domains claimed.** The multi-phase spatial foraging task needs more concrete details (arena/box layout, number of boxes, cues, timing, how phases are separated, what defines “correct,” and how/when it changes), and explicit handling of key edge cases (e.g., never visiting the correct box, censored times) (Sec. 2.1–2.2; Sec. 2.2.2). Without this, ceiling/floor effects and what the task truly measures (learning vs flexibility vs inhibition) are hard to assess.

Recommendation: Expand Sec. 2.1–2.2 with a step-by-step task specification: number and arrangement of boxes, cue availability, phase definitions and timing (within-day vs across days), trial/session duration, inter-phase interval, and rule for “correct” location across phases. In Sec. 2.2.2, define metric computation with explicit edge-case rules: censoring/maximum time if correct is never visited, treatment of missing timestamps or simultaneous events, and how perseveration is coded when entries are zero. Report how often such edge cases occurred and whether they affect CPI distribution (Sec. 3.2).

5. **Interpretation of the null DNAmAge–CPI result is overstated relative to the design and reporting. The manuscript sometimes implies “lack of age-related cognitive decline” or “remarkable resilience,” but the study is cross-sectional (not longitudinal), modest in size (N=32), and spans a limited mid-to-late adult epigenetic-age range (Sec. 2.4.2; Sec. 3.1; Sec. 3.4; Sec. 3.6; Sec. 4.3–4.4).** Absence of statistical significance is not evidence of absence; without effect sizes, CIs, and detectable-effect analyses, readers cannot judge what magnitude of decline is ruled out.

Recommendation: Recast conclusions in the Abstract, Sec. 3.6, and Sec. 4.3–4.4 to: “no detectable association in this sample/age range,” rather than broad claims of preserved cognition across aging. Report standardized effect sizes (e.g., CPI SD change

per DNAmAge year) with 95% CIs and interpret the CI as an upper bound on plausible decline. Add sensitivity checks for nonlinearity (quadratic DNAmAge term; or spline if feasible) and robustness (Spearman correlation; robust regression or permutation test) (Sec. 2.4.2; Sec. 3.4). Optionally include a detectable-effect-size or post-hoc power analysis to clarify what effects this design could reasonably detect.

6. **The “Cognitive Resilience Score” (residual from $\text{CPI} \sim \text{DNAmAge}$) is presented as a substantive phenotype, but when the DNAmAge slope is near zero the residual will be almost identical to CPI, and the term “resilience” may mislead readers into inferring demonstrated age-related decline that individuals are resilient to (Sec. 2.4.3; Sec. 3.5–3.6).** The score’s distribution, outlier sensitivity, and dependence on model choice are not characterized.

Recommendation: If retained, explicitly label it as “age-adjusted performance (residual)” and clarify its interpretation when no age effect is detected. Report its mean/SD/range, identify influential points (Cook’s distance/leverage), and report the correlation between CPI and the residual score to show what is gained by the transformation (Sec. 3.5–3.6). Include sensitivity to alternative specifications (e.g., adding Origin, Sex, or nonlinearity in the baseline model) and note that residualization is model-dependent (Sec. 2.4.3).

7. **Potential confounding and heterogeneity (especially colony of origin) are not analyzed deeply enough. Origin shows a trend-level association with CPI (Sec. 3.4) that could reflect housing/rearing differences, genetic structure/relatedness, prior task exposure, sensory/health status, or different DNAmAge distributions across colonies, potentially masking or mimicking aging patterns. Interactions ($\text{DNAmAge} \times \text{Origin}$, $\text{DNAmAge} \times \text{Sex}$) are not discussed.**

Recommendation: In Sec. 3.1 and Sec. 3.4–3.6, describe what is known to differ between colonies (environmental history, husbandry, capture/rearing, testing context). Provide descriptive comparisons: DNAmAge by Origin and CPI by Origin (means/SDs, plots) and check for DNAmAge–Origin confounding. If power permits, test $\text{DNAmAge} \times \text{Origin}$ (and $\text{DNAmAge} \times \text{Sex}$) interactions or clearly justify why they are omitted. Regardless, treat the Origin effect as exploratory (report estimate and 95% CI; avoid “trend toward significance” framing) and discuss plausible mechanisms and future controls (Sec. 3.6; Sec. 4.4).

8. **Reproducibility and data provenance details are incomplete. The Methods lack software/package versions, explicit inclusion/exclusion criteria, handling of missing/anomalous events, categorical coding choices, and model diagnostics (Sec. 2.1–2.2; Sec. 2.4; Sec. 3.4).** Additionally, inclusion of absolute internal filesystem paths (as noted in the unstructured report) is not appropriate for a publication and may expose private infrastructure while not helping others reproduce the work.

Recommendation: Add a reproducibility subsection (Sec. 2.4 or end of Sec. 2) specifying: languages and package versions; data cleaning rules; subject/trial exclusion criteria; handling of missing/censored metrics; and exact coding of Sex and Origin (reference levels). Summarize model diagnostics (residual plots, influence checks) supporting GLM assumptions (Sec. 3.4). Remove any absolute internal file paths and replace with portable descriptions; state code/data availability (repository/DOI) or, if not shareable, provide a detailed analysis workflow and synthetic example inputs (Sec. 2; Data availability statement).

9. **Figures and terminology currently limit standalone interpretability and may mislead: Figure 1’s overlap depiction is hard to read and may imply near-total overlap without numeric labels; Figures 2–7 often lack units, sample sizes, and key numeric/statistical annotations; and there is at least one documented inconsistency in the color mapping for the resilience score (Sec. 3.5; Fig. 7).** Some terminology is inconsistent (e.g., CPI vs. CAI) and CPI scale is hard to interpret without the number of components.

Recommendation: Revise Figure 1 to an UpSet plot or an area-proportional Euler diagram with explicit counts for every region/overlap; increase resolution and use a color-blind-safe palette. For Figs. 2–7, add axis units (DNAMAge in years; clarify CPI units as “sum of k z -scores”), annotate N per panel/group, and include key statistics (slope/CI/ R^2) on the regression plot (Sec. 3.2–3.5). Fix the Fig. 7 color convention so text/caption/colorbar all match (positive residual = higher age-adjusted performance) and standardize CPI terminology throughout.

Minor issues

1. NDDV description is internally inconsistent across sections, including conflicting statements about whether $b = 0$ images are required, and ambiguous equation typesetting (placement of normalization relative to the square root) (Sec. 1; Sec. 2.3; Sec. 3.3). The name “Normalized Directional Diffusion Variance” also appears inconsistent with a square-root-of-variance (i.e., SD) formula.

Recommendation: Harmonize NDDV inputs and definition across Sec. 1/2.3/3.3: explicitly state the required 4D diffusion series and $b = 0$ images (as currently defined), add parentheses to disambiguate the equation, and either rename NDDV to reflect SD/dispersion or change the computation to true variance. Since NDDV is not empirically computed here, keep the main-text description minimal and consistent and move detailed derivations to an Appendix.

2. Handling of non-normal behavioral metrics is not clearly connected to analysis choices. z -scoring rescales but does not correct skew; whether CPI and GLM residuals satisfy assumptions is not fully reported (Sec. 2.4.1; Sec. 3.2–3.4).

Recommendation: Clarify in Sec. 2.4.1 that z -scoring was used for comparability across metrics, not to enforce normality. Provide distribution diagnostics for CPI (e.g., QQ plot/Shapiro–Wilk) and residual diagnostics for the GLM. Consider reporting a rank-based (Spearman) DNAmAge–CPI association alongside OLS as a robustness check (Sec. 3.4).

3. The paper sometimes refers to “age” or “across the lifespan” when the predictor is skin-derived DNAmAge; this can overgeneralize beyond epigenetic age and beyond the sampled adult range (Abstract; Sec. 1; Sec. 3.6; Sec. 4.2–4.4).

Recommendation: Systematically use “DNAmAge/epigenetic age” when discussing the predictor, and explicitly delimit the sampled range (6.62–13.84 DNAmAge years). Briefly note what is known (or unknown) about correspondence between skin DNAmAge, chronological age, and brain aging in this species.

4. Ethics/animal welfare approval information is missing or not explicit despite live-animal behavioral experiments (Sec. 2).

Recommendation: Add an ethics statement at the end of Sec. 2.1 specifying the approving body, protocol/permit number, and relevant guidelines for bat housing and experimentation; if covered elsewhere, cite it but still summarize approvals here.

5. The DTI limitation is described mainly as “3D not 4D,” but it is unclear whether the issue is irrecoverable (e.g., header/concatenation error) and whether bvec/bval files or original DICOMs exist (Sec. 2.3; Sec. 3.3). This matters for the “lessons learned” value and for future remediation.

Recommendation: In Sec. 3.3 (and/or Sec. 4.4), state what exactly was available (NIfTI only vs DICOMs; presence/absence of bvec/bval), whether reconversion or concatenation was attempted, and what checks would have detected this earlier (e.g., automated dimensionality verification).

6. The “lessons learned” QC discussion is useful but somewhat generic (Sec. 4.4).

Recommendation: Make Sec. 4.4 more actionable: list concrete pre-acquisition and post-acquisition QC steps (verify 4D dimensionality, number of directions and b_0 s, presence of atlas, standardized naming, automated scripts, and documentation/data-sharing practices).

7. Colony effect is discussed as a “trend” without full uncertainty reporting and could be overemphasized (Sec. 3.4; Sec. 3.6).

Recommendation: Report the Origin coefficient with 95% CI and exact p -value in Sec. 3.4, treat it as exploratory, and avoid near-significance language. If multiple model variants were explored, briefly acknowledge multiplicity.

Very minor issues

1. Front-matter and formatting contain clear template artifacts and inconsistencies: astronomy-related keywords (Abstract), typographic issues (e.g., adjusted R-squared formatted as “00.00”), inconsistent variable-name quoting/backticks, occasional broken words/line breaks, and inconsistent section label formatting (Sec. 2.3–2.4).

Recommendation: Proofread and remove template artifacts; replace keywords with appropriate terms; standardize variable-name formatting and section labeling; fix broken line breaks and numeric formatting; ensure all referenced tables/figures exist and are correctly numbered.

2. Figure/caption inconsistencies and minor numerical presentation issues: Fig. 7 resilience color mapping is inconsistent between text and caption (Sec. 3.5); CPI mean reported as “−0.00” may confuse; some axes/legends use ambiguous labels or inconsistent abbreviations (e.g., CPI vs. CAI).

Recommendation: Make color conventions consistent across text/caption/colorbar (Fig. 7; Sec. 3.5), format signed-zero values as 0.00, and standardize abbreviations and axis labels across all figures.

Mathematical consistency audit

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

Maths relevance: light

The paper contains a small number of central mathematical definitions: (i) a proposed NDDV diffusion-signal dispersion metric normalized by mean $\mathbf{b} = \mathbf{0}$ signal, (ii) a Cognitive Performance Index (CPI) formed by summing sign-adjusted z -scores of behavioral metrics, (iii) linear-model specifications for CPI vs DNAmAge with covariates, and (iv) a Cognitive Resilience Score defined as residuals from $\text{CPI} \sim \text{DNAmAge}$. Most of the manuscript is methodological/interpretive rather than derivational; the key internal-consistency risks are definitional ambiguity and contradictory sign/color conventions.

Checked items

1. ✓ **NDDV definition (dispersion normalized by $\mathbf{b} = \mathbf{0}$)** (Sec. 1 (definition paragraph), p.2)
 - **Claim:** Defines NDDV using the sample variance across diffusion directions ($k = 1..N$) and normalizes by mean $\mathbf{b} = \mathbf{0}$ signal.
 - **Checks:** symbol/definition consistency, dimensional consistency
 - **Verdict:** PASS; confidence: medium; impact: moderate

- **Assumptions/inputs:** S_k are DWI signal intensities for N directions ($N = 30$), \bar{S} is the mean of the N DWI signals, $\bar{S}_{b=0}$ is the mean $b = 0$ signal intensity.
 - **Notes:** As a normalized dispersion metric, (std dev of DWIs)/(mean $b = 0$) is dimensionless and internally coherent. However, later items note ambiguity in the equation's grouping and a naming mismatch (variance vs std dev).
2. **△ Ambiguity of NDDV equation grouping (division inside vs outside sqrt)** (Sec. 1, p.2 (NDDV displayed formula))
- **Claim:** The formula indicates a square root of a sample-variance-like term divided by $\bar{S}_{b=0}$.
 - **Checks:** algebra/notation clarity
 - **Verdict:** UNCERTAIN; confidence: medium; impact: moderate
 - **Assumptions/inputs:** Displayed equation formatting may be read as either $\sqrt{A}/\bar{S}_{b=0}$ or $\sqrt{A/\bar{S}_{b=0}}$.
 - **Notes:** In the parsed rendering, the placement of $\bar{S}_{b=0}$ *relative to the radical is unclear. These two interpretations differ by a factor of $\sqrt{\bar{S}_{b=0}}$.* Parentheses are needed to make the intended definition unambiguous.
3. **✘ NDDV name vs computed quantity (variance vs standard deviation)** (Sec. 1, p.2 (metric name and formula); also Sec. 2.3, p.4)
- **Claim:** Metric is called a variance but uses a square root of a variance term.
 - **Checks:** definition consistency
 - **Verdict:** FAIL; confidence: high; impact: moderate
 - **Assumptions/inputs:** The equation uses $\sqrt{\frac{1}{N-1} \sum (S_k - \bar{S})^2}$, i.e., sample standard deviation.
 - **Notes:** As written, the metric is a normalized sample standard deviation, not a variance. This is a definitional mismatch (terminology vs formula).
4. **✘ Consistency of needing $b = 0$ images for NDDV** (Sec. 2.3, p.4 vs Sec. 3.3, p.6)
- **Claim:** NDDV is described as not requiring $b = 0$ images, yet the definition uses $\bar{S}_{b=0}$.
 - **Checks:** symbol/definition consistency, logical consistency
 - **Verdict:** FAIL; confidence: high; impact: moderate
 - **Assumptions/inputs:** Sec. 2.3 says NDDV is normalized by mean $b = 0$ map; Sec. 1 uses $\bar{S}_{b=0}$ in denominator.
 - **Notes:** Sec. 3.3 wording conflicts with the earlier definition and planned computation, both of which require $b = 0$ images to obtain $\bar{S}_{b=0}$.
5. **✓ CPI construction via z-scores and sign inversion** (Sec. 2.2.3, p.3–4)

- **Claim:** Standardize each metric across the cohort via z -scores; invert metrics where lower is better by multiplying by -1 ; sum adjusted z -scores to obtain CPI.
 - **Checks:** definition consistency, algebraic correctness
 - **Verdict:** PASS; confidence: high; impact: moderate
 - **Assumptions/inputs:** z -scoring uses cohort mean and standard deviation per metric., Sign inversion applied to listed metrics.
 - **Notes:** The described transformation is algebraically correct and yields a composite where higher values mean better performance. No internal contradictions in the sign-flip procedure itself.
6. \triangle **CPI component set is fully specified** (Sec. 2.2.2–2.2.3, p.3–4; Sec. 3.2, p.5–6)
- **Claim:** The CPI equals the sum of all adjusted z -scored metrics derived from the task.
 - **Checks:** definition completeness
 - **Verdict:** UNCERTAIN; confidence: medium; impact: minor
 - **Assumptions/inputs:** Visit_Efficiency is computed for each phase, but CPI summation does not explicitly state whether all phase-specific versions enter the sum.
 - **Notes:** The paper does not explicitly enumerate the exact CPI summands (including whether Visit_Efficiency contributes once or three times by phase). Without the precise component list, later interpretive statements about CPI scaling cannot be strictly verified from the text alone.
7. \checkmark **GLM model specification for CPI** (Sec. 2.4.2, p.4; reiterated Sec. 3.4, p.6)
- **Claim:** Fits $\text{CPI} \sim \text{DNAmAge} + \text{Sex} + \text{Origin}$ as a general linear model.
 - **Checks:** symbol/definition consistency
 - **Verdict:** PASS; confidence: high; impact: minor
 - **Assumptions/inputs:** CPI is treated as continuous; Sex/Origin are categorical covariates.
 - **Notes:** Model formula is consistently stated in Methods and Results. Encoding details (reference levels) are not given, but that is not an internal mathematical inconsistency.
8. \checkmark **Cognitive Resilience Score as residuals** (Sec. 2.4.3, p.4; Eq-like line in Sec. 3.5, p.7)
- **Claim:** Defines resilience as $\text{\text{CPI}}_{\text{\textit{observed}}} - \text{\text{CPI}}_{\text{\textit{predicted}}}$, *where* $\text{\text{CPI}}_{\text{\textit{predicted}}} \text{\textit{comes from the regression}} \text{\text{CPI}} \sim \text{\text{DNAmAge}}$.
 - **Checks:** algebraic correctness, definition consistency
 - **Verdict:** PASS; confidence: high; impact: moderate

- **Assumptions/inputs:** $CPI_{\text{predicted}}$ denotes fitted values from the simple linear model.
 - **Notes:** Residual definition is standard and consistent with the stated interpretation: positive residual = better-than-expected performance for age.
9. ✘ **Resilience color/sign convention across text and Figure 7 caption** (Sec. 3.5 text, p.7; Figure 7 caption, p.8)
- **Claim:** Colors map consistently to higher/lower resilience.
 - **Checks:** definition consistency
 - **Verdict:** FAIL; confidence: high; impact: minor
 - **Assumptions/inputs:** Sec. 3.5 text says blue = higher resilience, red = lower., Figure 7 caption says red = higher resilience, blue = lower.
 - **Notes:** Direct contradiction between text and caption. This can invert interpretation of the visualization.

Limitations

- Audit is based only on the provided PDF text/images; tables labeled “not shown” (Table 1, Table 2) are unavailable for symbolic cross-checking.
- Equation rendering in the parsed text may lose LaTeX grouping/parentheses; ambiguity findings reflect what is visible in the PDF rendering and extracted text rather than underlying source.
- No attempt was made to validate any numerical results, p -values, or model diagnostics; only symbolic/definitional consistency was assessed.

Numerical results audit

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

All 13 programmed numeric consistency checks passed. Cohort counts (intersection, sex, origin) are internally consistent with $N = 32$. DNAmAge and CPI ranges are ordered and yield coherent widths, and DNAmAge mean/SD satisfy basic sanity checks. The GLM summary statistics are mutually consistent: residual df matches N and number of predictors, and the reported F statistic aligns with R^2 within tolerance. One item (adjusted R^2) is effectively at a reporting/rounding boundary (computed ≈ 0.0069 vs reported 0.00).

Checked items

1. ✓ **C1_cohort_intersection_counts** (Results §3.1 (page 5): 'From an initial pool of 41 bats... and 33 bats with DTI data, the intersection yielded... 32')
 - **Claim:** From 41 bats with metadata+behavioral and 33 bats with DTI data, the intersection yielded a final cohort of 32 bats.
 - **Checks:** set_intersection_bounds
 - **Verdict:** PASS

- **Notes:** Checked $0 \leq \text{intersection} \leq \min(\text{sets})$. Computed implied non-overlap counts were nonnegative ($41 - 32 = 9$; $33 - 32 = 1$).
- 2. ✓ **C2_sex_counts_sum_to_total** (Results §3.1 (page 5): 'The cohort comprised 21 males and 11 females.')
 - **Claim:** Sex breakdown is 21 males and 11 females in the 32-bat cohort.
 - **Checks:** parts_to_total
 - **Verdict:** PASS
 - **Notes:** $21 + 11 = 32$.
- 3. ✓ **C3_origin_counts_sum_to_total** (Results §3.1 (page 5): 'Aseret (n=17) and Herzeliya (n=15)')
 - **Claim:** Origin (colony) breakdown is Aseret $n = 17$ and Herzeliya $n = 15$ in the 32-bat cohort.
 - **Checks:** parts_to_total
 - **Verdict:** PASS
 - **Notes:** $17 + 15 = 32$.
- 4. ✓ **C4_dnamage_range_order_and_width** (Results §3.1 (page 5): 'DNAmAge ranged from 6.62 to 13.84 years')
 - **Claim:** DNAmAge range is 6.62 to 13.84 years (min to max).
 - **Checks:** range_consistency
 - **Verdict:** PASS
 - **Notes:** Order check passed ($6.62 < 13.84$). Computed width = $13.84 - 6.62 = 7.22$.
- 5. ✓ **C5_dnamage_mean_within_range** (Results §3.1 (page 5): 'mean of 9.46 years and a standard deviation of 1.60 years')
 - **Claim:** DNAmAge mean=9.46 years should lie within stated min=6.62 and max=13.84.
 - **Checks:** mean_within_min_max
 - **Verdict:** PASS
 - **Notes:** Checked $6.62 \leq 9.46 \leq 13.84$.
- 6. ✓ **C6_dnamage_sd_positive_and_plausible** (Results §3.1 (page 5): 'standard deviation of 1.60 years')
 - **Claim:** DNAmAge standard deviation is 1.60 years; should be non-negative and not exceed half the full range width.
 - **Checks:** sd_sanity_vs_range
 - **Verdict:** PASS
 - **Notes:** Computed width= 7.22 and half-width= 3.61; sd= 1.60 is nonnegative and ≤ 3.61 .

7. ✓ **C7_cpi_mean_near_zero_reported_as_negative_zero** (Results §3.2 (page 6): 'mean of -0.00 and a standard deviation of 1.94')
- **Claim:** CPI mean is reported as -0.00 ; verify it is numerically equal to 0.00 under typical rounding and can be treated as zero.
 - **Checks:** `rounding_sign_zero`
 - **Verdict:** PASS
 - **Notes:** Verified -0.00 equals 0.00 numerically (signed zero formatting only).
8. ✓ **C8_cpi_range_order_and_width** (Results §3.2 (page 6): 'CPI values ranged from -5.87 to 3.33')
- **Claim:** CPI range is -5.87 to 3.33 (min to max).
 - **Checks:** `range_consistency`
 - **Verdict:** PASS
 - **Notes:** Order check passed ($-5.87 < 3.33$). Computed width = $3.33 - (-5.87) = 9.20$.
9. ✓ **C9_glm_df_consistency_with_sample_size** (Results §3.4 (page 6): 'F(3, 28) = 1.074' with $N = 32$ and predictors DNAmAge+Sex+Origin)
- **Claim:** In a linear model with $N = 32$ and 3 predictors (excluding intercept), the F -test degrees of freedom should be $(3, 28)$.
 - **Checks:** `regression_df_arithmetic`
 - **Verdict:** PASS
 - **Notes:** Residual df computed as $N - p - 1 = 32 - 3 - 1 = 28$, matching reported.
10. ✓ **C10_r2_vs_adj_r2_consistency_check** (Results §3.4 (page 6): 'adjusted R-squared = 00.00, R-squared = 0.103' with $N = 32$ and $p = 3$ predictors)
- **Claim:** Given $R^2 = 0.103$, $N = 32$, and $p = 3$ predictors, adjusted R^2 should be $1 - (1 - R^2) \frac{N-1}{N-p-1}$.
 - **Checks:** `adjusted_r2_recompute`
 - **Verdict:** PASS
 - **Notes:** Computed adjusted $R^2 \approx 0.00689$ (rounds to 0.01). Reported adjusted R^2 is 0.00 ; treated as pass at $abs_{tol} = 0.01$ but is at the tolerance boundary.
11. ✓ **C11_f_stat_vs_r2_identity_for_overall_model** (Results §3.4 (page 6): 'F(3, 28) = 1.074' and 'R-squared = 0.103')
- **Claim:** Overall F statistic should satisfy $F = (R^2/df_1)/((1 - R^2)/df_2)$ with $df_1 = 3$, $df_2 = 28$.
 - **Checks:** `f_stat_from_r2`

- **Verdict:** PASS
 - **Notes:** Computed $F_{\text{expected}} \approx 1.0717$ from R^2 and dfs; matches reported $F = 1.074$ within tolerance.
12. ✓ **C12_dti_shape_dimensionality_check** (Results §3.3 (page 6): 'skipped due to their (80, 80, 34) shape (representing 3D data)')
- **Claim:** DTI file shape reported as (80, 80, 34) indicates 3D rather than expected 4D.
 - **Checks:** dimensionality_from_shape_tuple
 - **Verdict:** PASS
 - **Notes:** Parsed shape to 3 entries ($ndim = 3$), confirming it is not 4D as expected for DTI.
13. ✓ **C13_nddv_formula_internal_N_value_consistency** (Introduction/Methods (page 2): NDDV definition with 'k = 1,...,30' and 'where N = 30')
- **Claim:** NDDV formula uses $k = 1..30$ and states $N = 30$; verify N matches count of diffusion-weighted images described.
 - **Checks:** repeated_constant_consistency
 - **Verdict:** PASS
 - **Notes:** N equals k_{max} (30), consistent within the formula text.

Limitations

- Only numeric statements explicitly present in the provided PDF text were used; referenced but missing tables (Table 1, Table 2) and any underlying datasets/files cannot be checked.
- No value extraction from plot pixels or image-only figures was performed; figure captions without explicit numbers were not used for numeric verification.
- Some checks (e.g., adjusted R^2 , F -from- R^2) assume standard OLS with an intercept and no missingness; if modeling differs, those relationships may not hold exactly.
- Several numeric items were left unverified due to missing supporting statistics (e.g., full regression coefficient tables/test statistics needed to recompute p -values; CPI SD expectations depend on component count and correlations).