MRI-derived Face Age vs Brain Age from the Same T1 Scan: a Longitudinal Single-Subject Stress Test

Ekaterina Kondrateva[1] EKATERINA.KONDRATEVA@MAASTRICHTUNIVERSITY.NL
Ramil Khafizov[2]
Gleb Bobrovskikh[2]
[1] *Department of Radiation Oncology (Maastro), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht University, Maastricht, The Netherlands*
[2] *Applied AI Institute*

Editors: Under Review for MIDL 2026

Abstract

A non-defaced T1-weighted MRI contains two distinct aging signals in a single file: the brain parenchyma and the complete facial anatomy. We present an open pipeline that extracts both, running a brain-age model (SynthBA) and two face-age methods—multi-view surface renders scored by FaceAge, and morphometric features from BioFace3D-20 landmarks—in parallel on the same volume. The SIMON single-subject dataset (99 scans, 36 scanners, age 29–46 yr) reveals a critical failure mode: SynthBA predicts a near-constant age of 27.4±1.26 yr regardless of the subject's actual age (MAE 16.2 yr), demonstrating that high scanner reproducibility and severe out-of-distribution bias can coexist. Face morphometrics, by contrast, remain biased (+5.1 yr) but show a detectable positive trend with chronological age, while FaceAge renders show a larger positive bias (+8.5 yr) consistent with the MRI-to-photo domain gap. The main result is methodological: reproducibility alone is not evidence of validity, and same-scan facial geometry may retain longitudinal aging information missed by current brain-age models.

Keywords: brain age, face age, MRI, longitudinal, reproducibility, out-of-distribution, SIMON

1. Introduction

Standard MRI pre-processing pipelines routinely remove facial structure ("defacing") to protect participant identity before data sharing. However, non-defaced T1-weighted volumes still contain both a brain and a complete facial surface within the same acquisition. This makes it possible to estimate two age signals from one file and ask whether they behave similarly under repeated scanning.

Prior work on brain age (Peng et al., 2021; Lemaître et al., 2022) and face age (Bontempi et al., 2025) usually treats these as separate modalities acquired by different instruments. A same-scan comparison removes that acquisition mismatch, and a longitudinal single-subject dataset adds a stricter test: can a model follow aging within one person rather than merely produce scanner-stable predictions?

More broadly, biological-age estimation may become useful for earlier disease detection or risk stratification, not only for age prediction itself. For example, FaceAge suggested

that face-derived biological age carries clinically relevant signal beyond chronological age in oncology (Bontempi et al., 2025).

We use the SIMON dataset (Bhagwat et al., 2021), which contains 99 T1 scans of one healthy subject acquired across 36 scanners while he ages from 29 to 46 years, as a controlled stress test. Our goals are: (1) run face-age and brain-age estimation on the same T1 volume; and (2) measure whether each branch tracks longitudinal aging or collapses to a reproducible but incorrect age estimate.

2. Methods

Face branch. Given a T1 NIfTI volume, we applied the BioFace3D head extraction pipeline, including histogram matching to the IXI intensity template and N4 bias field correction. The head isosurface was extracted with VTK FlyingEdges3D using a scan-specific threshold estimated from image intensity statistics, with quality control for failed extractions.

We rendered the resulting head mesh with PyVista from nine viewpoints ($\pm 10^\circ$ yaw/pitch) under randomized lighting. We then evaluated two face-based age estimators: FaceAge on the rendered views, and BioFace3D-20 landmarks followed by GPA+EDMA morphometric features and Ridge regression. Figure 1A shows a synthetic example of the MRI-derived multi-view render grid used as input to the face branch.

Brain branch. Skull stripping is performed with SynthStrip (Hoopes et al., 2022). We then run SynthBA (Lemaître et al., 2022), a segmentation-based model designed to be robust to acquisition variability, on the stripped volume.

Dataset and evaluation setting. The **SIMON** dataset (Bhagwat et al., 2021) provides 99 T1 scans of a single healthy male collected across 73 sessions on 36 scanners (age range 29.6–46.4 years). Because identity is fixed, between-scan spread reflects scanner sensitivity rather than subject heterogeneity, while the slope of predictions over time provides a direct readout of longitudinal validity.

3. Results

Figure 1 plots chronological age against predicted age for all three methods. Three patterns stand out.

SynthBA (brain) produces a nearly flat response: mean predicted age 27.4 ± 1.26 years across all 99 scans despite the subject aging from approximately 30 to 46 years. Scanner-to-scanner reproducibility is excellent (SD 1.26 yr), but absolute accuracy is poor (MAE 16.2 yr, bias -16.2 yr). The model is effectively frozen near a training-distribution anchor and does not track longitudinal change in this subject.

Face morphometrics (GPA+EDMA features, Ridge regression) show a positive bias of $+5.1$ yr and a spread of SD 5.6 yr, but they also show a visible positive trend with chronological age. The geometric representation therefore appears to retain some longitudinal aging signal even without subject-specific retraining.

FaceAge multi-view renders show the largest bias (+8.5 yr, SD 5.7 yr) and the widest scatter. This is consistent with the domain gap between MRI surface renders and the photographic training distribution of the model.

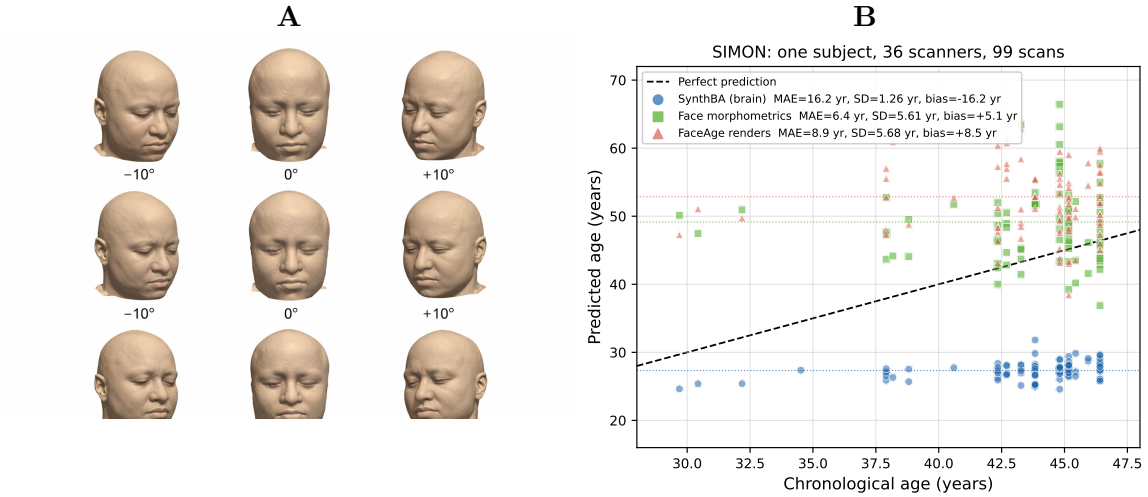

Figure 1: (A) Synthetic example of the MRI-derived multi-view face render grid used as input to the face branch. (B) Chronological vs predicted age on SIMON (one subject, 36 scanners, 99 scans). Dashed line: perfect prediction. Dotted horizontal lines: mean predicted age per method. SynthBA predicts a near-constant age regardless of the subject's actual age at scan, a clear out-of-distribution failure. Face morphometrics show a positive trend with chronological age despite positive bias.

4. Discussion and Conclusion

A non-defaced T1 MRI still contains two age-sensitive anatomical signals, but the SIMON stress test shows that extracting them reliably is harder than running pretrained models on the same file. The main lesson is that low scanner variance can mask severe biological error: SynthBA is reproducible across hardware yet fails to follow the subject's aging trajectory.

The face branches fail differently. FaceAge is strongly biased on MRI surface renders, consistent with the photo-to-MRI domain gap. The morphometric branch discards color and texture yet preserves a visible longitudinal trend, suggesting that facial geometry from T1 MRI may retain aging information that current brain-age predictions miss.

Limitations. This work is only a first step. We have not yet fully reproduced several of the most interesting recent papers in face age and brain age, so the current comparison is still incomplete. We also have not yet looked at other biomarkers or additional datasets that could test whether the observed face–brain mismatch generalises beyond SIMON. Finally, IXI brain-age results are intentionally excluded here because a potential train/test leakage concern makes them unsuitable as main evidence.

References

Nikhil Bhagwat et al. The SIMON dataset: A longitudinal single-subject MRI dataset acquired across many scanner platforms. *GigaScience*, 10(4), 2021.

Dennis Bontempi et al. FaceAge: Predicting biological age from face photographs with deep learning. *The Lancet Digital Health*, 2025. preprint / in press.

Andrew Hoopes, Jocelyn S Mora, Adrian V Dalca, Bruce Fischl, and Malte Hoffmann. SynthStrip: Skull-stripping for any brain image. https://surfer.nmr.mgh.harvard.edu/docs/synthstrip/, 2022.

Paul Lemaître et al. SynthBA: A robust brain age estimator leveraging synthetic data. *arXiv preprint arXiv:2210.xxxxx*, 2022.

Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. In *Medical Image Analysis*, volume 68, page 101871. Elsevier, 2021.