

Foundation artificial intelligence models for health recognition by use of face photographs (FAHR-Face): a model development and validation study

Fridolin Haugg*, Grace Lee*, John He, Leonard Nürnberg, Dennis Bontempi, Danielle S Bitterman, Paul Catalano, Vasco Prudente, Dmitrii Glubokov, Andrew Warrington, Suraj Pai, Dirk De Ruyscher, Christian Guthier, Benjamin H Kann, Vadim N Gladyshev, Hugo JWL Aerts, Raymond H Mak



Summary

Background Facial appearance offers a non-invasive window into health. We built Foundation Artificial Intelligence for Health Recognition-Face (FAHR-Face), a foundation model trained on more than 40 million facial photographs and fine-tuned it for two distinct tasks: biological age estimation (FAHR-FaceAge) and survival prediction (FAHR-FaceSurvival).

Methods FAHR-FaceAge underwent two-stage, age-balanced fine-tuning on ten public image datasets; FAHR-FaceSurvival was fine-tuned on 34 363 photographs of patients with cancer. Model robustness (cosmetic surgery, make-up, pose, and lighting) and independence (attention mapping) were tested. Both models were clinically tested in two independent cancer patient datasets with survival analysed by multivariate Cox models and adjusted for clinical prognostic factors.

Findings For age estimation, FAHR-FaceAge had the lowest mean absolute error of 5.1 years on public datasets, outperforming benchmark models and maintaining accuracy across the full human lifespan. In patients with cancer, FAHR-FaceAge outperformed a previous age estimation model in survival prognostication. FAHR-FaceSurvival showed robust prediction of mortality, and the highest-risk quartile had more than triple the mortality of the lowest (adjusted hazard ratio 3.23; $p < 0.0001$). These findings were validated in the independent cohort and both models showed generalisability across age, sex, race, and cancer subgroups. The two algorithms provided distinct, complementary prognostic information; attention mapping revealed that each model relied on distinct facial regions. The combination of FAHR-FaceAge and FAHR-FaceSurvival improved prognostic accuracy.

Interpretation A single foundation model can generate inexpensive, scalable facial biomarkers that capture both biological ageing and disease-related mortality risk. The foundation model enabled effective training by use of relatively small clinical datasets. Routine quantitative facial analysis could enhance risk stratification when traditional biomarkers are unavailable.

Funding EU European Research Council.

Copyright © 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Facial images have emerged as a promising biomarker for health assessment, offering a non-invasive and readily accessible means of capturing physiological changes associated with ageing and overall health.^{1,2} Influenced by various physiological factors, disease processes, and lifestyle choices, the human face might provide valuable insights into an individual's health status, reflecting the cumulative effects of genetic predisposition, environmental exposures, existing medical conditions, and lifestyle habits.^{3,4} These facial changes can mirror the underlying biological ageing process, connected with the development of various diseases, particularly cancer, through the time-dependent accumulation of cellular damage.^{5,6} Despite this potential, effectively quantifying facial appearance for health assessment has been challenging, because facial features are complex and difficult to measure objectively.

Studies have shown the potential of artificial intelligence (AI)-based approaches to extract health-related information from facial images, including chronological age estimation,^{7,8} biological age estimation,^{9,10} cardiovascular risk assessment,⁹ and identification of individuals with rare genetic syndromes.² Biological age here refers to the predicted age from a facial image that reflects ageing physiology rather than time lived. Across imaging modalities and anatomical regions, AI-based biological age predictions and their deviation from chronological age are associated with survival and other health outcomes.¹¹ Consistent with previous studies¹² and our previous study,¹⁰ we interpret higher biological age, in our case defined as appearing older than chronological age, as a marker of poorer health and greater clinical risk of adverse health outcomes, including mortality. In that study, we developed FaceAge, a deep-learning model that estimates biological age from facial photographs, and

Lancet Digit Health 2026

Published Online
<https://doi.org/10.1016/j.landig.2026.101056>

*Joint first authors

Department of Radiation Oncology, Mass General Brigham Cancer Institute, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA (F Haugg MS, G Lee MD, J He BS, L Nürnberg MS, D Bontempi PhD, D S Bitterman MD, V Prudente MS, A Warrington BS, S Pai PhD, C Guthier PhD, B H Kann MD, Prof H J W L Aerts PhD, Prof R H Mak MD); Artificial Intelligence in Medicine Program, Mass General Brigham, Harvard Medical School, Boston, MA, USA (F Haugg, G Lee, L Nürnberg, D Bontempi, D S Bitterman, V Prudente, S Pai, C Guthier, B H Kann, Prof H J W L Aerts, Prof R H Mak); Department of Radiology and Nuclear Medicine, CARIM & GROW, Maastricht University, Maastricht, Netherlands (L Nürnberg, D Bontempi, V Prudente, S Pai, Prof H J W L Aerts); Department of Radiation Oncology (Maastricht), Maastricht University, Maastricht, Netherlands (L Nürnberg, D Bontempi); Department of Biostatistics, Harvard T H Chan School of Public Health, Boston, MA, USA (P Catalano ScD); Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA (P Catalano); Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA (D Glubokov MS, Prof V N Gladyshev PhD); Department of Radiation Oncology (Maastricht), Maastricht University Medical Center, GROW School, Maastricht, Netherlands (Prof D De Ruyscher MD); Department of Radiology,

Brigham and Women's Hospital,
Harvard Medical School, Boston,
MA, USA (Prof H J W L Aerts)

Correspondence to:
Prof Raymond H Mak,
Department of Radiation
Oncology, Mass General Brigham
Cancer Institute, Dana-Farber
Cancer Institute, Harvard
Medical School, Boston,
MA 02115, USA
rmak@partners.org

or

Prof Hugo JWL Aerts, Artificial
Intelligence in Medicine
Program, Mass General Brigham,
Harvard Medical School, Boston,
MA 02115, USA
haerts@bwh.harvard.edu

See Online for appendix

Research in context

Evidence before this study

We systematically searched PubMed and Google Scholar for English language studies published from database inception until May 1, 2025, using terms such as “facial image analysis”, “biological age estimation”, “FaceAge,” “survival prediction”, and “foundation model”. Previous research showed the utility of facial images for chronological and biological age estimation, as well as associations with health outcomes. However, existing methods rely heavily on supervised training using incomplete, often biased datasets and show inconsistent generalisability outside their training distribution.

Added value of this study

This study introduces Foundation Artificial Intelligence for Health Recognition-Face (FAHR-Face), the first foundation artificial intelligence model pretrained via masked autoencoders on over

40 million diverse facial images, covering the entire human lifespan. FAHR-Face substantially surpasses previous methods in accuracy, robustness, and generalisability, reliably fine-tuning specialised clinical biomarkers, including biological age and survival risk, from small clinical datasets using only consumer-grade facial photographs.

Implications of all the available evidence

Clinicians already intuitively assess patient health based on appearance. FAHR-Face and its derived biomarkers (FAHR-FaceAge and FAHR-FaceSurvival) provide tools to standardise and quantify this intuitive practice. Integration of this model into clinical workflows could enhance prognostic accuracy and clinical decision making, particularly in settings where traditional biomarkers or comprehensive assessments are scarce.

found that appearing older than chronological age was associated with poorer survival in patients with cancer.

Building on these AI advancements, foundation models have revolutionised natural language processing and computer vision, and are now making substantial strides in medical applications. These models, trained on vast amounts of data, have shown capabilities in capturing complex patterns and transferring knowledge across diverse tasks.¹³ Unlike task-specific models, foundation models learn reusable representations that transfer across tasks, so one pretrained backbone can be adapted to many applications with minimal labels. In medical applications, foundation models have shown promise in clinical natural language processing, medical image analysis,¹⁴ and pathology.¹⁵

Among the techniques used to develop such foundation models, masked autoencoders have emerged as a particularly effective approach for image-based tasks.¹⁶ Masked autoencoders learn by randomly masking and reconstructing patches in images, creating a challenging self-supervisory task that requires a holistic understanding of the input.¹⁷ This approach is especially well suited for facial image analysis, as the face contains many correlated structures (eg, symmetry and typical spatial relationships between features), which the model can leverage to capture higher-level attributes potentially related to health.

Here, we introduce Foundation Artificial Intelligence for Health Recognition-Face (FAHR-Face), a deep-learning framework designed to assess health-related attributes and predict survival outcomes by use of consumer-grade digital photographs of the face. Our novel approach uses a masked autoencoder pretrained on over 40 million facial images to develop a foundation model capable of transforming subtle facial feature variations into quantifiable imaging biomarkers, and this foundation model can then be fine-tuned for multiple health-related tasks.

Our study shows the application of this foundation model in two key areas: FAHR-FaceAge for biological age

estimation and FAHR-FaceSurvival for predicting survival outcomes.

Methods

Model design and parameters

A foundation model (FAHR-Face) was trained by use of self-supervised learning on over 40 million unlabelled facial images from the WebFace42M dataset (see appendix p 1).¹⁸ Subsequently, two distinct supervised-learning models were developed by fine-tuning this foundation model: FAHR-FaceAge for biological age estimation and FAHR-FaceSurvival for predicting survival from facial photographs. The study design and dataset flow are illustrated in figure 1.

This study was done in accordance with the ethical principles outlined in the Declaration of Helsinki and was approved by the independent ethics review boards at Mass General Brigham (2020P002969) and the Dana-Farber Harvard Cancer Center (13-055 and 17-669) in Boston, USA, and Department of Radiation Oncology (Maastricht University (08-06-23/01) in the Netherlands. All clinical data were handled in compliance with institutional research policies. For both the Harvard and Maastricht patients, facial images and clinical variables were collected retrospectively from data acquired during routine clinical care. For the Harvard datasets the institutional review board (IRB) approved a waiver of consent (Mass General Brigham IRB protocol 2020P002969 and Dana-Farber Harvard Cancer Center IRB protocols 13-055 and 17-669). The IRBs granted a waiver of written informed consent because the study involved minimal risk, relied on existing clinical data, did not influence patient management, and a substantial fraction of the cohort consisted of deceased patients.

Data sources

FAHR-FaceAge was fine-tuned in a two-step training process, first by use of eight oversampled public datasets

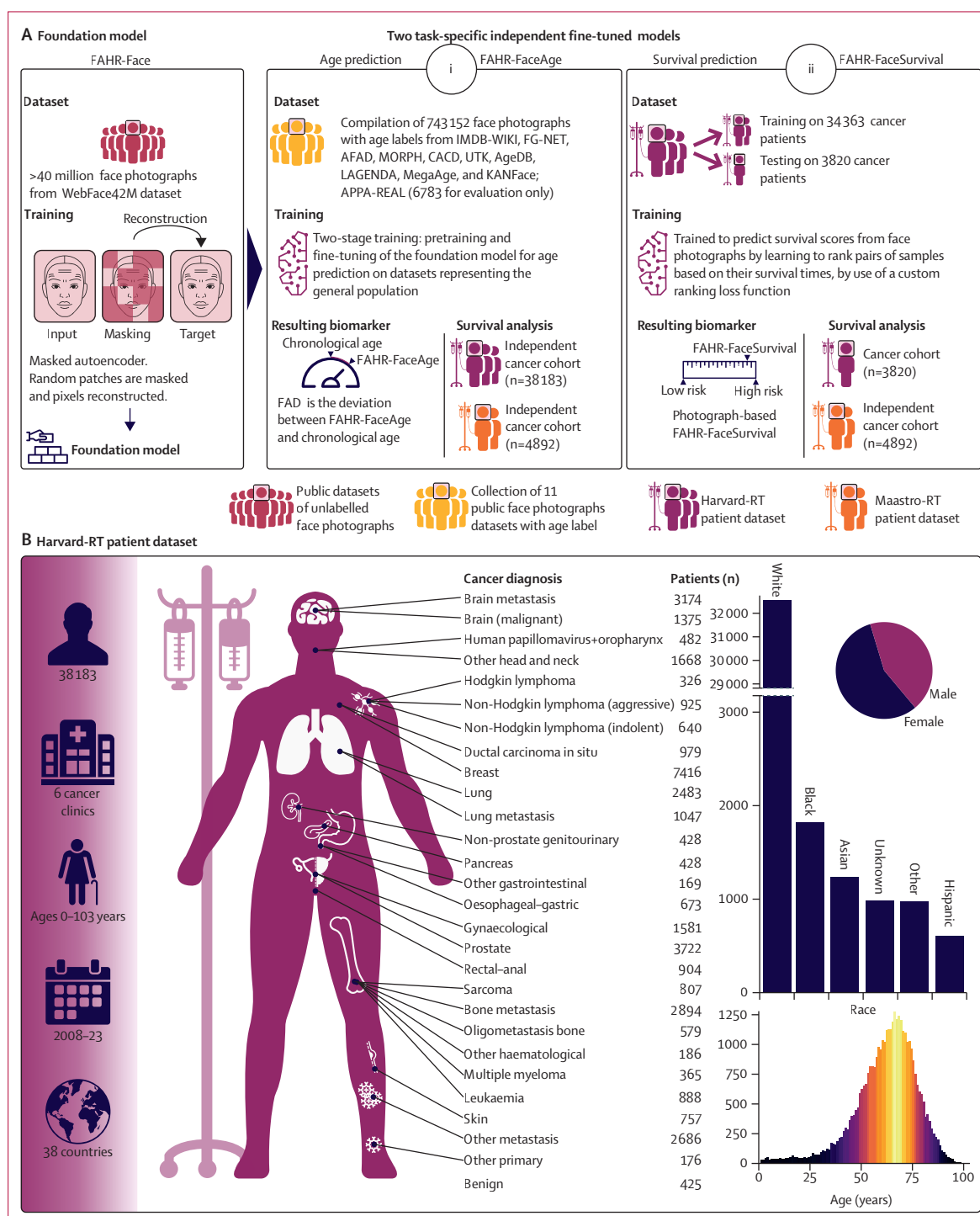
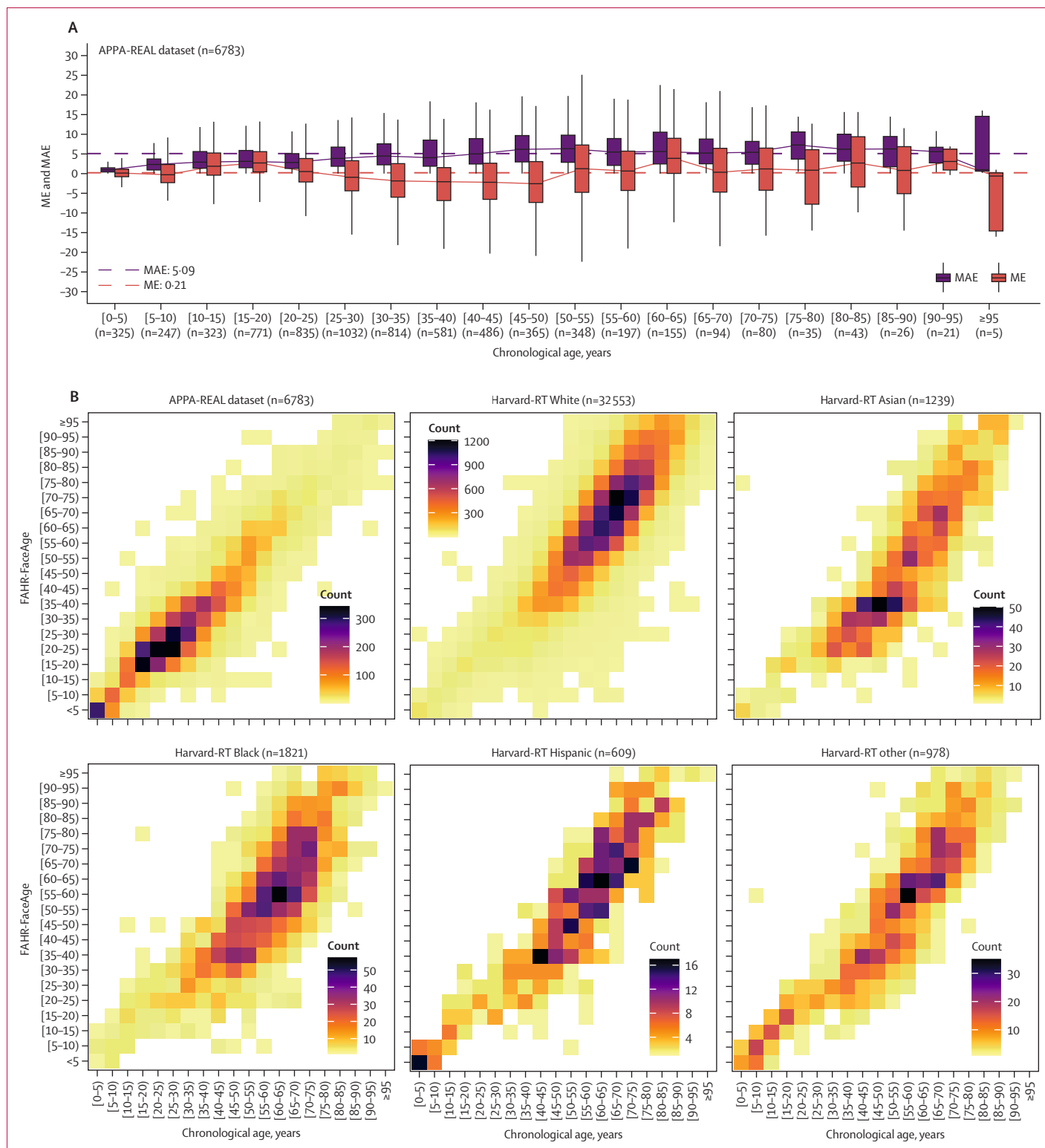
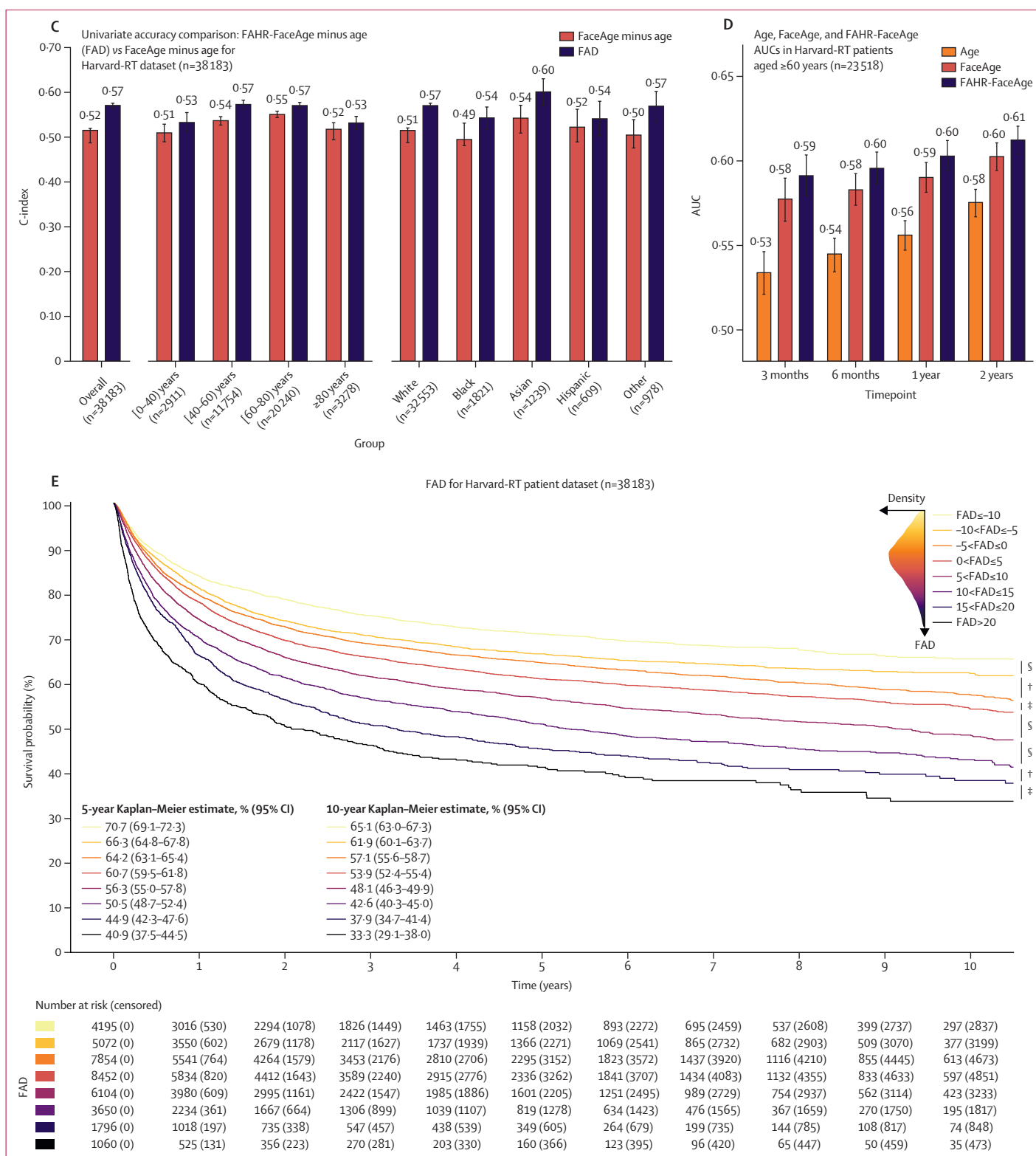


Figure 1: Study design and Harvard-RT patient dataset overview

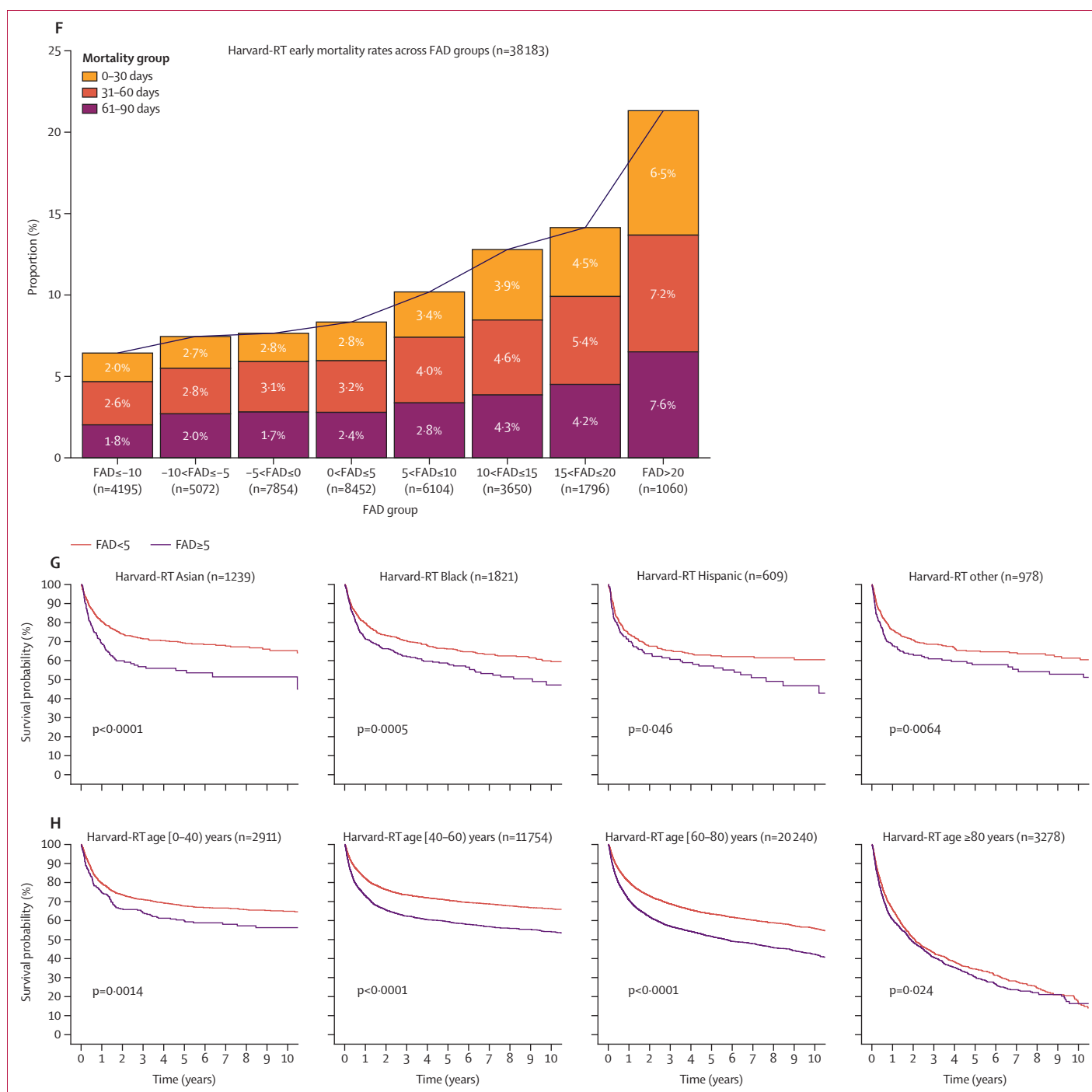
(A) Study design flowchart. The FAHR-Face foundation model is trained on more than 40 million face photographs from the WebFace42M dataset. Two applications are derived: FAHR-FaceAge for age prediction and FAHR-FaceSurvival for survival prediction. FAHR-FaceAge uses 743 152 face photographs with age labels from ten public datasets and is evaluated on the public dataset APPA-REAL (n=6783) and the independent Harvard-RT patient dataset (n=38 183). FAHR-FaceSurvival is trained on 34 363 patients from the Harvard-RT patient dataset and tested on 3820 patients (90/10 split). Both models are evaluated on the independent Maastro-RT patient dataset (n=4892). (B) Overview of the Harvard-RT patient dataset. 38 183 patients were treated with radiotherapy from 2008 to 2023. The dataset includes cancer diagnosis and patient demographics. The pie chart shows sex distribution. The bar chart displays race distribution. The histogram illustrates the age distribution. FAD=FAHR-FaceAge deviation (FAHR-FaceAge minus chronological age). AFAD=Asian Face Age Dataset. APPA-REAL=Apparent and Real Age. CACD=Cross-Age Celebrity Dataset. Harvard-RT=Harvard radiotherapy. KANFace=Kinship, Age, geNder. LAGENDA=Layer Age and Gender Dataset. Maastro-RT=Maastro radiotherapy.



(Figure 2 continues on next page)



(Figure 2 continues on next page)



(IMDB-WIKI, KANFace, FG-NET, CACD, AFAD, Mega-Age, MORPH, and LAGENDA), followed by an age-balanced fine-tuning stage on two public datasets (UTK and AgeDB) to ensure unbiased age estimation across the human lifespan (0–116 years; see appendix pp 2–3). The model was benchmarked against state-of-the-art

architectures¹⁹ on the APPA-REAL dataset and assessed for robustness to image variations. It was evaluated on the entire Harvard radiotherapy (Harvard-RT) patient dataset, a multicentre cohort of 38 183 patients treated with radiotherapy for cancer or benign conditions at six Harvard-affiliated sites, and externally validated on the independent

Figure 2: FAHR-FaceAge and FAD analysis in the clinical dataset (Harvard-RT) and publicly available dataset (APPA-REAL)

(A) MAE and ME per 5-year age bin for age estimation on the public APPA-REAL dataset. This panel shows consistent performance across all age bins, indicating that the model's accuracy in predicting age is stable throughout different age groups. (B) FAHR-FaceAge versus chronological age heatmaps for different race groups in the Harvard-RT patient dataset and APPA-REAL datasets (Pearson correlation 0–92). The heatmaps illustrate that although different race groups exhibit different age distributions, the model shows similar diagonal patterns, with high correlations across all groups (White 0–86, Black 0–81, Asian 0–86, other 0–90, and Hispanic 0–91). (C) Concordance index (C-index) for FAD compared with FaceAge¹⁰ minus age in the overall Harvard-RT patient dataset and in age and race subgroups. Error bars indicate 95% CIs obtained from non-parametric bootstrap with 1000 resamples at the patient level. (D) AUC comparison of Age, FaceAge¹⁰, and FAHR-FaceAge for survival prediction in those in the Harvard-RT patient dataset aged 60 years and older (n=23 518). AUC values were calculated at four timepoints: 3 months, 6 months, 1 year, and 2 years. The analysis was restricted to patients aged 60 years and older because the FaceAge¹⁰ model was optimised for individuals aged 60 years and older. (E) Kaplan–Meier survival curves stratified by FAD groups for the Harvard-RT patient dataset. The FAD groups range from –10 to 20+ years. Numbers at risk, cumulative censored counts in parentheses, and 5-year and 10-year Kaplan–Meier estimates with 95% CIs are shown below the plot. The sample density plot shows the relative distribution of patients per FAD group. (F) Percentage of patients with early mortality (within 30, 31–60, and 61–90 days) across different FAD groups. The stacked plot represents a continuous increase in the proportion of patients who die within 90 days as the FAD value increases. (G) Kaplan–Meier survival curves for the Harvard-RT patient dataset, split by FAD $\geq$ 5 and FAD<5, across different race groups (Asian, Black, Hispanic, and other), with two-sided log-rank p values. (H) Kaplan–Meier survival curves for the Harvard-RT patient dataset, split by FAD $\geq$ 5 and FAD<5, across different age groups (0–40, 40–60, 60–80, and $\geq$ 80 years), with two-sided log-rank p values. AUC=area under the receiver operating characteristic curve. FAD=FAHR-FaceAge deviation (FAHR-FaceAge minus chronological age). Harvard-RT=Harvard radiotherapy. MAE=mean absolute error. ME=mean error. $\dagger p<0.01$. $\ddagger p<0.001$. $\S p<0.0001$.

Maastro radiotherapy (Maastro-RT) patient dataset, comprising 4892 patients with cancer treated with radiotherapy in Maastricht, Netherlands (appendix p 6).

FAHR-FaceSurvival was developed by fine-tuning FAHR-Face on a subset of the Harvard-RT patient dataset. The Harvard-RT images were randomly partitioned (90:10 split) to maximise training data while retaining sufficient data for testing, resulting in a training set of 34 363 images and a held-out test set of 3820 images. Training used a ranking-based loss approach to predict survival outcomes. The final FAHR-FaceSurvival model performance was evaluated on the held-out test set and independently validated by use of the same Maastro-RT patient dataset (see appendix pp 5–6).

Outcomes

Chronological-age accuracy was summarised with mean absolute error (overall and averaged across 5-year bins). Biological age deviation was expressed as FAHR-FaceAge deviation (FAD; defined as FAHR-FaceAge minus chronological age).²⁰ Survival discrimination was assessed with concordance index (C-index), multivariate Cox models, Kaplan–Meier curves and time-dependent area under the receiver operating characteristic curve (AUC; 3, 6, 12, and 24 months).

Statistical analysis

Baseline characteristics were summarised descriptively. Survival was modelled with Cox proportional hazards; multivariate models included age (per decade), sex, race or ethnicity, cancer site, treatment intent, calendar year, and radiation technique, retaining covariates with $p<0.05$ in univariate screening. Discrimination was quantified with C-index and time-dependent AUCs at 3, 6, 12, and 24 months, comparing chronological age, FaceAge, and FAHR-FaceAge. Risk stratification used Kaplan–Meier curves and log-rank tests; subgroup hazard ratios (HRs) with 95% CIs were reported. Complementarity and independence of the two models were assessed in both feature and geometric space. Analyses used R 4.3.1; two-sided $p<0.05$ was considered significant. Reporting followed TRIPOD+AI (appendix pp 32–33).

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The FAHR-FaceAge model, which was fine-tuned from the FAHR-Face foundation model, achieved the lowest mean absolute error of 5.1 years on the public APPA-REAL dataset of presumed healthy individuals (figure 2A; appendix p 4). In comparison as a benchmark, a baseline Vision Transformer (ViT) architecture trained from random initialisation reached 6.92 years mean absolute error (see appendix p 5 for additional benchmark details and results). Fine-tuned with age-balanced data to reduce bias, FAHR-FaceAge showed generalisability across diverse age ranges and racial groups in publicly available, real-world photographs and photographs obtained in clinical settings (figure 2B). The heatmaps reveal consistent age estimation accuracy across racial subgroups with high correlations (White 0–86, Black 0–81, Asian 0–86, Hispanic 0–91, and other 0–90).

Next, we tested FAHR-FaceAge as a biological age biomarker in clinical cohorts and showed its superiority to our previously published age estimation algorithm (FaceAge¹⁰). The clinical application of the FAHR-FaceAge model was first evaluated by analysing the FAD. FAD was tested as a clinical biomarker on a large independent cohort of 38 183 patients with a diverse range of cancer diagnoses (Harvard-RT patient dataset). The median FAD was 1.1 years, with 12 610 patients (33.0%) having an FAD of at least 5 years and 6506 patients (17.0%) having an FAD of 10 years or more—ie, patients appearing much older than their chronological age. Conversely, 9267 patients (24.3%) had an FAD of –5 years or less—ie, patients appearing much younger than their chronological age. FAD was only weakly correlated with chronological age ($R=0.16$). The median follow-up time in this cohort was 4.3 years.²¹ The patient characteristics are summarised in the appendix (pp 20–21).

Patients with higher FADs exhibited lower survival probabilities across all chronological age groups, with significant differences observed between patients with FAD differences exceeding 5 years or more; patients appearing younger than their chronological age showed progressively higher survival rates, whereas those appearing older had increasingly poorer health outcomes (figure 2E). This association between FAD and survival is evident across all age and race subgroups (figure 2G, H). Additionally, FAD is a robust early mortality biomarker (figure 2F); patients in the highest FAD group (>20 years) had a 90-day mortality of

226 (21.3%) of 1060 patients, compared with 270 (6.4%) of 4195 patients for those with an FAD of ≤ 10 years or less.

Benchmarking against the previously published biological face age estimation model (FaceAge¹⁰), in single-predictor analyses, our new FAD biomarker achieved a C-index of 0.57, whereas the previous model's age difference biomarker (FaceAge¹⁰ minus chronological age) achieved 0.52, with FAD maintaining higher C-indices across age and racial subgroups (figure 2C). Examining different survival endpoints in patients aged 60 years and older, FAHR-FaceAge had higher performance for 3 months, 6 months, 1 year, and 2 years survival prediction than FaceAge¹⁰ and chronological age (figure 2D).

In univariate Cox regression analyses (table 1), FAD (per decade) emerged as a significant predictor of survival (HR 1.30, 95% CI 1.27–1.32, $p < 0.0001$). Multivariate analysis adjusting for age, sex, race or ethnicity, cancer site, treatment intent, treatment year, and radiation technique confirmed FAD's significance (adjusted HR 1.14, 1.12–1.16, $p < 0.0001$). Notably, a FAD of 5 years or more (≥ 5 years) was linked to a 21% higher hazard (adjusted HR 1.21, 1.17–1.26, $p < 0.0001$). Conversely, FAD ≤ 5 years or less was associated with a 17% lower hazard compared with an FAD greater than ≤ 5 years (adjusted HR 0.83, 0.79–0.86, $p < 0.0001$).

Subgroup analyses (appendix p 22) were generally consistent. Although both males and females exhibited significant FAD-survival associations, females showed higher HRs. White and Asian patients had stronger FAD-based risk stratification than Black and Hispanic groups. Across cancer diagnoses, patients with FAD of 5 years or more (≥ 5 years) consistently had poorer survival (appendix p 12), and this difference was significant in 18 of 28 cancer diagnoses, with the remaining subgroups showing the same directional separation but not reaching statistical significance.

Lifestyle factors were significantly associated with FAD: heavy alcohol consumption and drug use corresponded with increased FAD. The strongest association was observed with smoking, which showed consistent dose-dependent relationships across both smoking-related and non-smoking-related cancers: both in smoking status (current $>$ former $>$ never smokers) and lifetime exposure (moderate $>$ heavy $>$ light $>$ never smokers; appendix p 14).

Next, we independently trained the FAHR-FaceSurvival model from patient photographs by fine-tuning the FAHR-Face foundation model for survival prediction. The model was internally validated on a subcohort of 3820 patients from the Harvard-RT patient test dataset. In the test cohort, the median age was 64.1 years (range 0.8–98.5). The test cohort was 2178 female patients (57.0%), with a racial distribution of 3255 White (85.2%), 182 Black (4.8%), 116 Asian (3.0%), 59 Hispanic (1.5%), and 106 other-race patients (2.8%). Most patients had non-metastatic cancer (2630 patients, 68.8%), followed by metastatic cancer (1,045 patients, 27.4%), and benign or

ductal carcinoma in situ (DCIS) conditions (145 patients, 3.8%). The median follow-up time was 4.2 years. The training and test patient characteristics are summarised in the appendix (pp 20–21).

Higher FAHR-FaceSurvival was associated with more aggressive cancer risk groups, with a median of 0.13 (IQR 0.02–0.32) for benign or DCIS conditions, 0.25 (0.07–0.46) for non-metastatic cancer, and 0.45 (0.25–0.68) for metastatic cancer ($p < 0.0001$; figure 3A). Among age groups, FAHR-FaceSurvival predicts the highest mortality risk in the age 60–80 years and age 80 years and older groups (figure 3A). FAHR-FaceSurvival was also significantly higher in males (0.36, IQR 0.19–0.56 for males ν 0.23, 0.05–0.48 for females; $p < 0.0001$), with no significant difference between most races ($p > 0.05$ for all comparisons except Asian ν White, $p = 0.036$; figure 3A). Higher FAHR-FaceSurvival quartiles were associated with significantly lower survival probabilities (figure 3B).

On univariate Cox regression, the continuous FAHR-FaceSurvival was associated with survival (HR for every 0.1 increase = 1.32, 95% CI 1.30–1.35, $p < 0.0001$). Multivariate analysis, adjusting for age, sex, race or ethnicity, cancer site, treatment course intent, year treated, and radiation technique, confirmed that FAHR-FaceSurvival remained significantly associated with survival (adjusted HR 1.17, 1.14–1.20, $p < 0.0001$; table 2).

Subgroup analysis showed FAHR-FaceSurvival association with survival across various demographic and clinical subgroups. As shown in the appendix (p 24), both continuous (per 0.1 increase) and dichotomised (≥ 0.5 ν < 0.5) FAHR-FaceSurvival was significantly associated with survival in univariate and multivariate analyses for most subgroups. Kaplan–Meier plots of dichotomised FAHR-FaceSurvival for different racial and age subgroups are shown in figure 3E and figure 3F, respectively.

FAHR-FaceSurvival was also significantly associated with early mortality risk. In the highest FAHR-FaceSurvival group (0.9–1.0), 64 (52.5%) of 122 patients died within 90 days, compared with 20 of 975 patients (2.1%) in the lowest FAHR-FaceSurvival group (0.0–0.1; figure 3D). The C-index for race, age, and diagnosis subgroups showed varied performance with an overall C-index of 0.72 and ranging from 0.53 to 0.88 across different subgroups (figure 3G). Notably, a benchmark model trained on photographs of patients with cancer alone without the FAHR-Face foundation model, which used a ViT with random initialisation, had a C-index of 0.66 (appendix p 7).

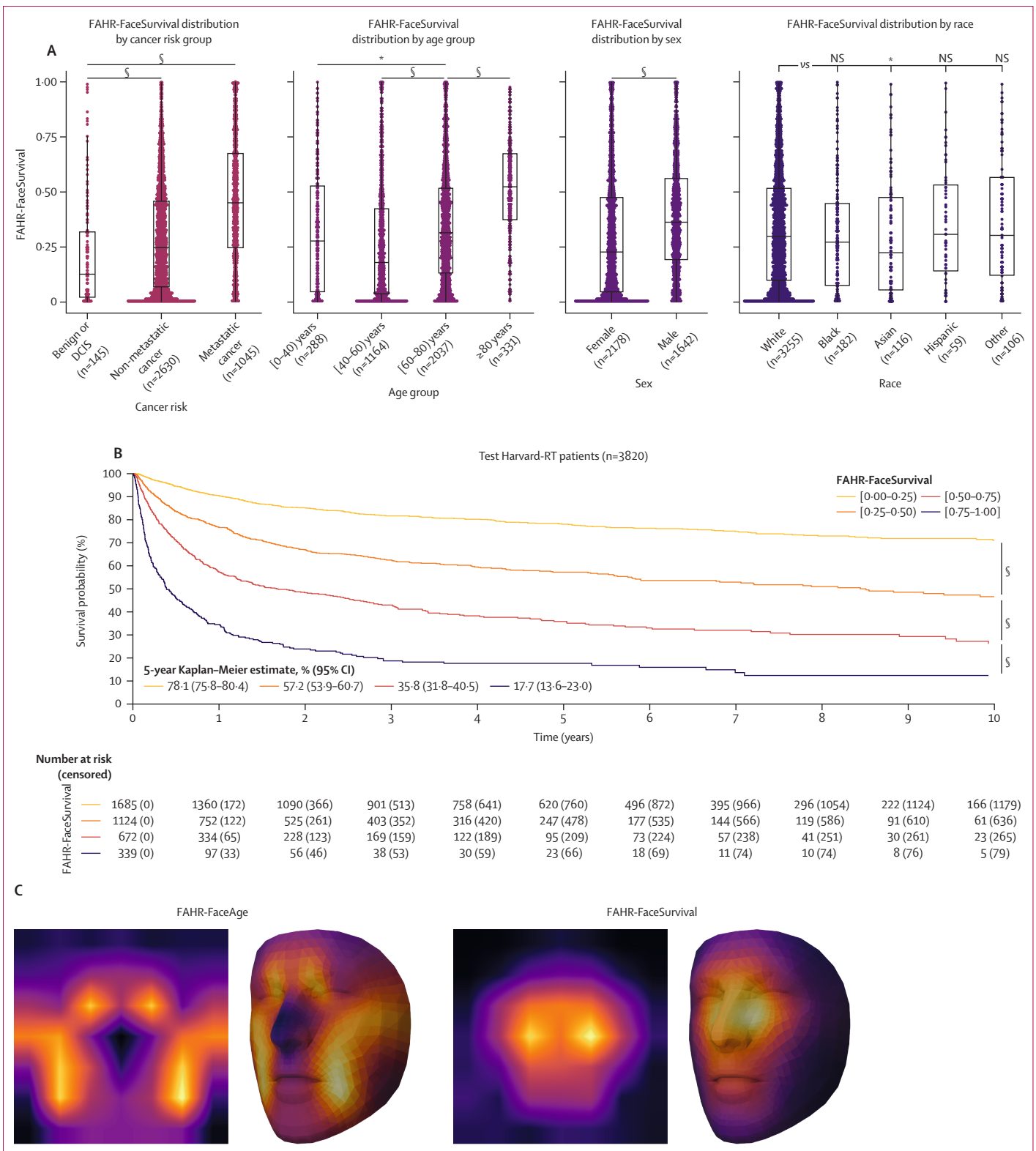
The generalisability of both FAHR-Face models was validated by use of 4892 patients from Maastricht-RT. This cohort had a median age of 66.0 years, with 2309 female patients (47.2%). Most common cancer sites included breast, gastrointestinal, genitourinary, and lung (appendix pp 20–21). The median follow-up time was 4.9 years.

The median FAHR-FaceAge was 70.9 years (range 12.0–97.0), with a median FAD of 4.4 years (range ≤ 31.0 to 39.6). 2288 patients (46.8%) had a FAD ≥ 5 years, and 1103 patients (22.5%) had a FAD ≥ 10 years. The mean

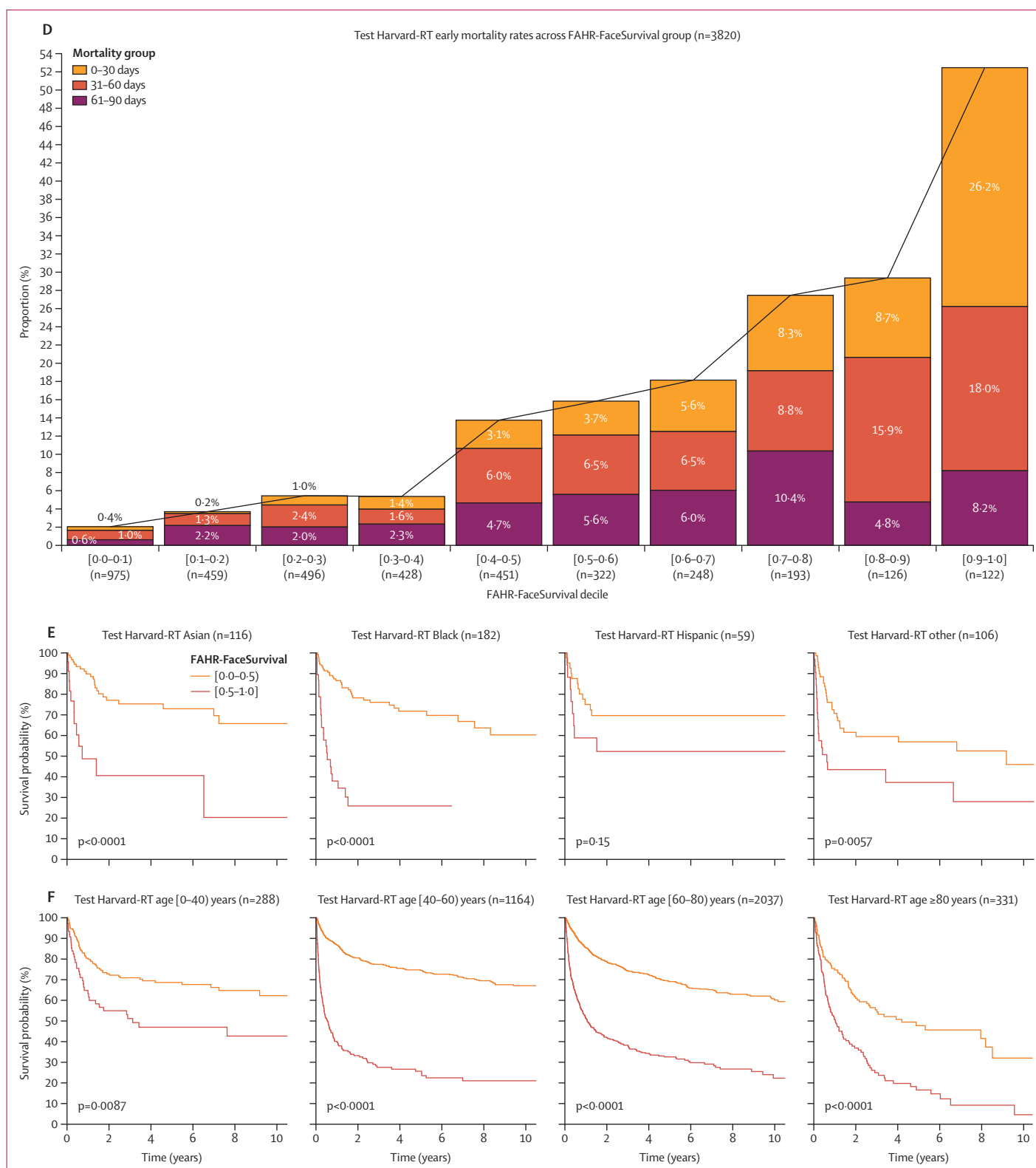
	Univariate*		Multivariate model 1† (FAD per decade)		Multivariate model 2‡ (FAD ≥5 years vs <5)		Multivariate model 3‡ (FAD ≤-5 years vs >-5 years)	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Age, decade	1.19 (1.18–1.21)	<0.0001	1.14 (1.13–1.16)	<0.0001	1.14 (1.13–1.16)	<0.0001	1.15 (1.14–1.17)	<0.0001
FAHR-FaceAge, decade								
FAD, decade	1.30 (1.27–1.32)	<0.0001	1.14 (1.12–1.16)	<0.0001	..	..	..	..
FAD ≥5 vs <5	1.53 (1.48–1.58)	<0.0001	..	..	1.21 (1.17–1.26)	<0.0001	..	..
FAD ≤-5 vs >-5	0.68 (0.65–0.71)	<0.0001	..	..	..	..	0.83 (0.79–0.86)	<0.0001
Sex, male vs female	1.38 (1.34–1.43)	<0.0001	1.12 (1.08–1.16)	<0.0001	1.13 (1.09–1.17)	<0.0001	1.10 (1.06–1.14)	<0.0001
Race or ethnicity‡								
White	1 (ref)	..	1 (ref)	..	1 (ref)	..	1 (ref)	..
Black	0.90 (0.83–0.98)	0.014	1.11 (1.02–1.20)	0.021	1.08 (0.99–1.18)	0.073	1.11 (1.02–1.21)	0.019
Asian	0.83 (0.74–0.92)	0.00040	1.04 (0.93–1.16)	0.48	1.00 (0.89–1.11)	0.93	1.02 (0.92–1.14)	0.72
Hispanic	1.01 (0.88–1.15)	0.94	1.22 (1.06–1.40)	0.0041	1.20 (1.05–1.37)	0.0082	1.21 (1.05–1.38)	0.0068
Other	0.96 (0.86–1.07)	0.47	1.14 (1.02–1.28)	0.022	1.14 (1.02–1.28)	0.024	1.15 (1.03–1.29)	0.013
Cancer risk group§								
Non-metastatic cancer	1 (ref)	..	..	..	..	..	..	..
Benign conditions or ductal carcinoma in situ	0.18 (0.15–0.23)	<0.0001	..	..	..	..	..	..
Metastatic cancer	5.76 (5.57–5.97)	<0.0001	..	..	..	..	..	..
Cancer site								
Breast	1 (ref)	..	1 (ref)	..	1 (ref)	..	1 (ref)	..
Benign	2.42 (1.82–3.21)	<0.0001	2.71 (2.02–3.63)	<0.0001	2.74 (2.04–3.67)	<0.0001	2.72 (2.03–3.64)	<0.0001
Prostate	1.67 (1.44–1.93)	<0.0001	1.73 (1.46–2.05)	<0.0001	1.75 (1.47–2.07)	<0.0001	1.75 (1.47–2.07)	<0.0001
Lung	15.40 (13.83–17.16)	<0.0001	13.72 (12.18–15.45)	<0.0001	13.95 (12.39–15.72)	<0.0001	14.08 (12.50–15.86)	<0.0001
Gastrointestinal	10.71 (9.56–12.01)	<0.0001	9.86 (8.71–11.15)	<0.0001	9.91 (8.76–11.22)	<0.0001	10.00 (8.84–11.31)	<0.0001
Gynaecological	5.47 (4.75–6.29)	<0.0001	6.69 (5.74–7.81)	<0.0001	6.74 (5.78–7.85)	<0.0001	6.74 (5.78–7.86)	<0.0001
Head and neck	5.54 (4.88–6.28)	<0.0001	5.87 (5.09–6.77)	<0.0001	5.92 (5.13–6.83)	<0.0001	5.98 (5.18–6.90)	<0.0001
Sarcoma	6.26 (5.33–7.35)	<0.0001	6.90 (5.81–8.18)	<0.0001	6.93 (5.84–8.22)	<0.0001	6.91 (5.82–8.20)	<0.0001
Haematological	7.53 (6.75–8.41)	<0.0001	4.48 (3.96–5.06)	<0.0001	4.52 (4.00–5.11)	<0.0001	4.52 (4.00–5.11)	<0.0001
Skin	6.54 (5.54–7.72)	<0.0001	6.62 (5.45–8.04)	<0.0001	6.68 (5.50–8.11)	<0.0001	6.69 (5.51–8.12)	<0.0001
Brain (malignant)	24.57 (21.93–27.52)	<0.0001	24.99 (21.99–28.40)	<0.0001	25.06 (22.05–28.48)	<0.0001	25.05 (22.04–28.47)	<0.0001
Other primary	11.71 (10.05–13.64)	<0.0001	10.22 (8.66–12.07)	<0.0001	10.34 (8.75–12.21)	<0.0001	10.29 (8.71–12.15)	<0.0001
Bone metastasis	36.78 (33.21–40.73)	<0.0001	12.93 (11.45–14.59)	<0.0001	12.98 (11.50–14.65)	<0.0001	13.05 (11.56–14.74)	<0.0001
Brain metastasis	44.94 (40.58–49.77)	<0.0001	21.45 (18.91–24.32)	<0.0001	21.73 (19.16–24.64)	<0.0001	21.81 (19.23–24.74)	<0.0001
Other metastasis	24.59 (22.21–27.24)	<0.0001	13.07 (11.61–14.72)	<0.0001	13.18 (11.71–14.84)	<0.0001	13.29 (11.80–14.97)	<0.0001
Treatment course intent								
Curative	1 (ref)	..	1 (ref)	..	1 (ref)	..	1 (ref)	..
Oligometastasis ablation	4.18 (3.96–4.42)	<0.0001	1.83 (1.69–1.98)	<0.0001	1.83 (1.69–1.98)	<0.0001	1.83 (1.69–1.98)	<0.0001
Palliative	8.77 (8.45–9.10)	<0.0001	3.42 (3.23–3.62)	<0.0001	3.42 (3.23–3.63)	<0.0001	3.44 (3.25–3.64)	<0.0001
Year treated								
2008–15	1 (ref)	..	1 (ref)	..	1 (ref)	..	1 (ref)	..
2016–23	0.74 (0.71–0.76)	<0.0001	0.82 (0.79–0.85)	<0.0001	0.82 (0.79–0.85)	<0.0001	0.82 (0.79–0.85)	<0.0001
Radiation technique¶								
Three-dimensional conformal	1 (ref)	..	1 (ref)	..	1 (ref)	..	1 (ref)	..
Intensity modulated radiotherapy	0.69 (0.66–0.72)	<0.0001	0.75 (0.71–0.80)	<0.0001	0.75 (0.71–0.80)	<0.0001	0.75 (0.71–0.80)	<0.0001
Stereotactic body radiotherapy	1.08 (1.00–1.16)	0.061	0.55 (0.50–0.60)	<0.0001	0.55 (0.50–0.60)	<0.0001	0.55 (0.50–0.60)	<0.0001
Stereotactic radiosurgery	2.58 (2.43–2.73)	<0.0001	0.71 (0.65–0.77)	<0.0001	0.71 (0.65–0.77)	<0.0001	0.71 (0.65–0.78)	<0.0001
Brachytherapy	0.20 (0.17–0.24)	<0.0001	0.31 (0.26–0.39)	<0.0001	0.32 (0.26–0.39)	<0.0001	0.32 (0.26–0.39)	<0.0001
Electrons	0.56 (0.49–0.63)	<0.0001	0.61 (0.53–0.71)	<0.0001	0.61 (0.52–0.71)	<0.0001	0.61 (0.52–0.71)	<0.0001
Adaptive	0.72 (0.60–0.86)	0.00029	0.55 (0.45–0.66)	<0.0001	0.54 (0.45–0.66)	<0.0001	0.54 (0.45–0.66)	<0.0001

p values were derived from a two-sided Wald test. FAD=FAHR-FaceAge deviation (FAHR-FaceAge minus chronological age). HR=hazard ratio. *n=38 183 with 13 769 events (63.9% censored), unless otherwise specified. †n=34 775 with 12 801 events (63.2% censored). ‡n=37 200. §Excluded from multivariate analysis given significant confounding with cancer site variable. ¶n=35 706.

Table 1: Univariate and multivariate Cox regression for overall survival—FAHR-FaceAge



(Figure 3 continues on next page)



(Figure 3 continues on next page)

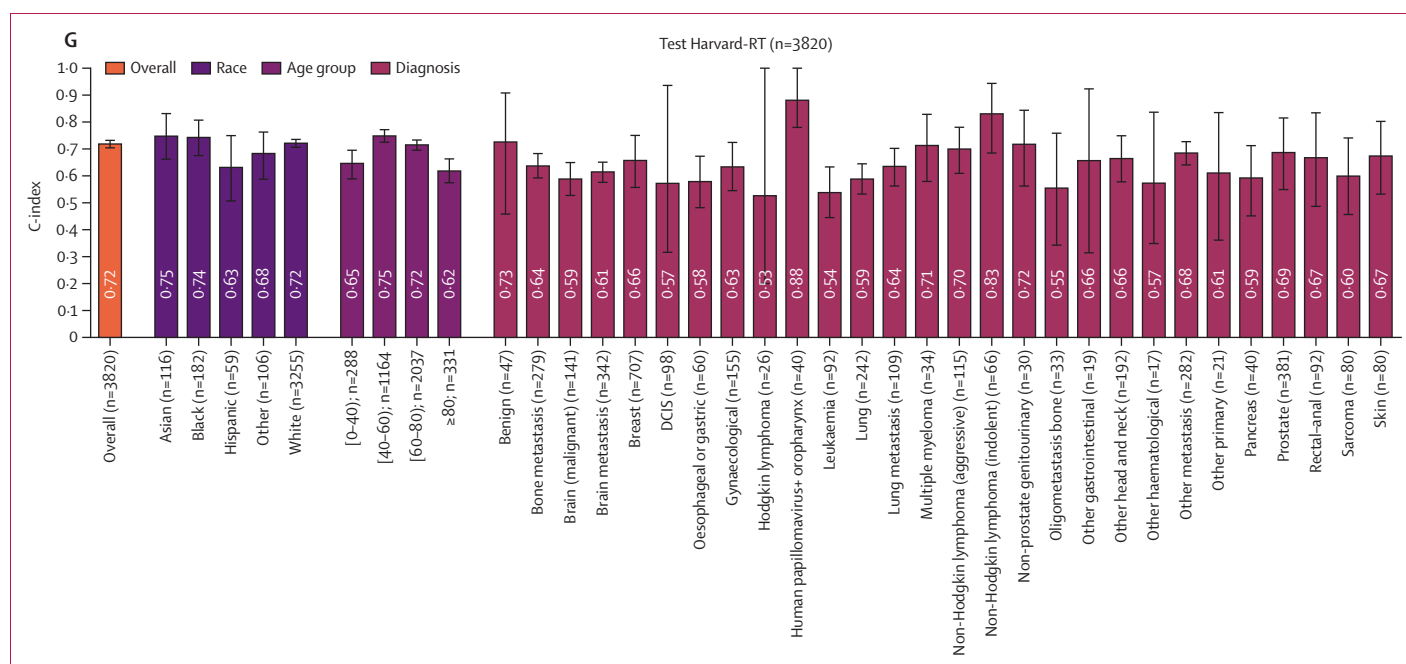


Figure 3: FAHR-FaceSurvival analysis on the Harvard-RT test dataset

(A) Box plots show the variability of FAHR-FaceSurvival across different demographic groups, indicating that the model captures meaningful variations in risk on the basis of these factors (two-sided Wilcoxon rank-sum test). Boxplots show median and IQR; whiskers indicate $1.5 \times$ IQR. (B) Risk stratification by grouping patients into quartiles on the basis of their FAHR-FaceSurvival. The Kaplan-Meier survival curves show significant differences in survival probabilities between the quartiles, indicating the model's ability to effectively stratify patients by risk (two-sided log-rank *p* values). Numbers at risk, cumulative censored counts in parentheses, and 5-year Kaplan-Meier estimates with 95% CIs are shown below the plot. (C) Mean attention maps for FAHR-FaceAge and FAHR-FaceSurvival, showing the facial regions most relevant for each prediction in yellow. Results are displayed on both two-dimensional projections ($n=3820$) and three-dimensional canonical face models ($n=3815$). (D) Early mortality percentages within 30, 31–60, and 61–90 days across different FAHR-FaceSurvival groups, divided into deciles (1/10 steps). The stacked plot shows a continuous increase in early mortality rates with higher FAHR-FaceSurvival. (E) Kaplan-Meier survival curves for the Harvard-RT patient test dataset, split by FAHR-FaceSurvival higher than and lower than 0.5, across different race groups (Black, Asian, Hispanic, and other), with two-sided log-rank *p* values. (F) Kaplan-Meier survival curves for the Harvard-RT patient test dataset, split by FAHR-FaceSurvival higher than and lower than 0.5, across different age groups (0–40, 40–60, 60–80, and ≥ 80 years), with two-sided log-rank *p* values. (G) Concordance index (C-index) for overall age group, sex, race, and cancer diagnosis. This panel shows the performance of the FAHR-FaceSurvival model as a single predictor in estimating overall survival across various subgroups, with higher C-index values indicating better predictive accuracy. Error bars indicate 95% CIs obtained from non-parametric bootstrap with 1000 resamples at the patient level. DCIS=ductal carcinoma in situ. Harvard-RT=Harvard radiotherapy. NS=not significant ($p>0.05$). * $p<0.05$. $\S p<0.0001$.

absolute error for FAHR-FaceAge was 6.99 years, and the mean error was 4.70 years.

Kaplan-Meier survival analysis revealed significant stratification on the basis of FAD (figure 4A). Multivariate analysis, adjusted for age, sex, and cancer site, confirmed FAD as a significant predictor of survival (adjusted HR 1.13, 95% CI 1.06–1.20, $p=0.0003$; figure 4C). The HR for FAD per decade increased after adjustment for age, showing that FAD adds prognostic information beyond chronological age.

For FAHR-FaceSurvival, the median was 0.4 in this cohort. Kaplan-Meier analysis showed significant survival differences among FAHR-FaceSurvival groups (figure 4B). Multivariate analysis, adjusted for age, sex, and cancer site, confirmed FAHR-FaceSurvival (per 0.1 unit) as a significant predictor of survival (adjusted HR 1.09, 95% CI 1.06–1.11, $p<0.0001$; figure 4D).

The AUC for patients aged 60 years and older (figure 4G) showed that FaceAge¹⁰ and FAHR-FaceAge outperformed chronological age in predicting survival outcomes. At 3 months, FAHR-FaceAge achieved an AUC of 0.66 compared with FaceAge¹⁰ of 0.64 and chronological age of 0.63. This performance advantage persisted at

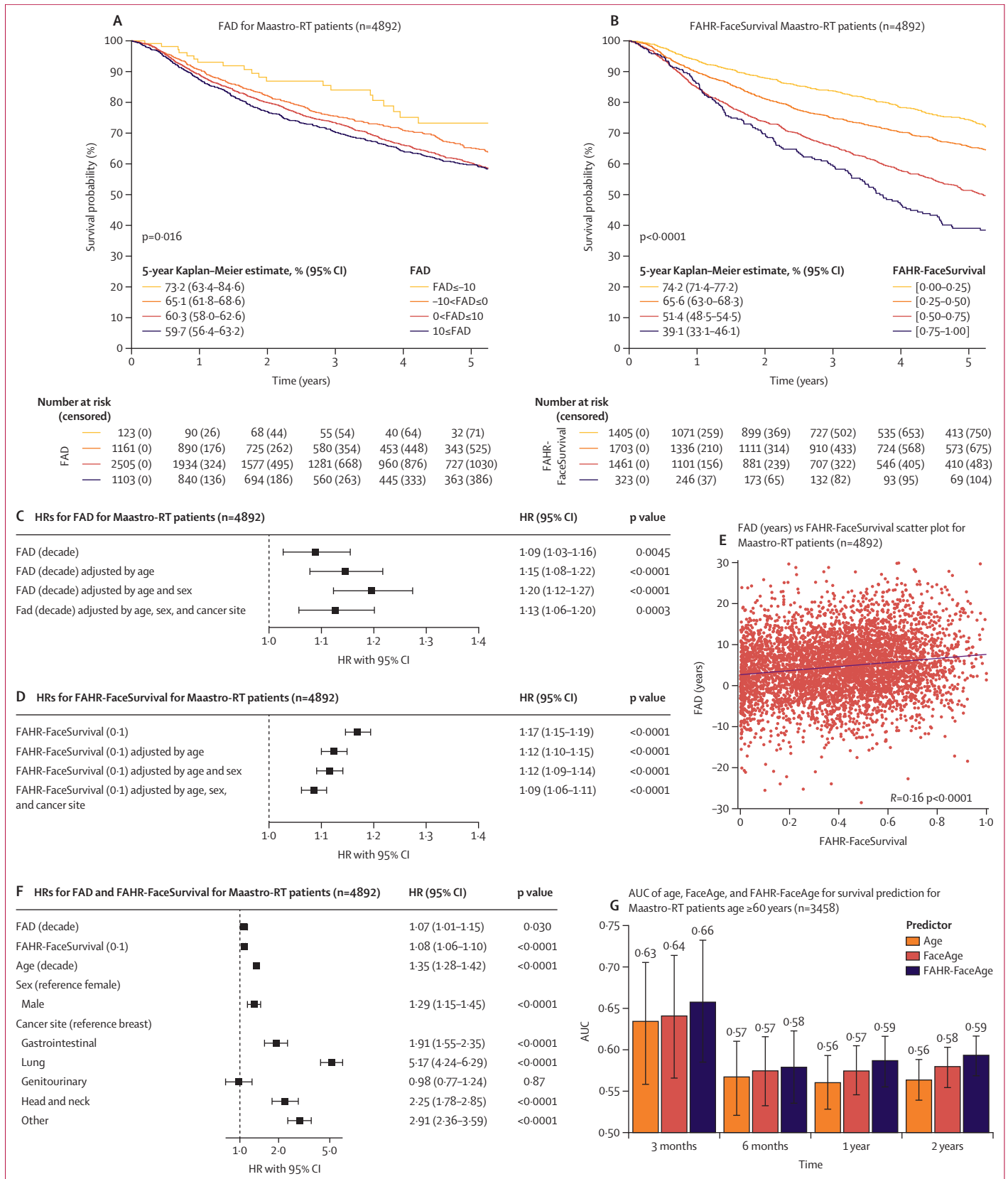
6 months, 1 year, and 2 years. Among patients aged 60 years and older, FaceAge¹⁰ minus age (per decade) showed a weaker association with mortality (adjusted HR 1.11, 95% CI 1.01–1.23, $p=0.026$; figure 4H) than FAD and FAHR-FaceSurvival. In a combined multivariate model including FAD, FaceAge¹⁰ minus age, FAHR-FaceSurvival, age, sex, and cancer site, only FAD (adjusted HR 1.16, 1.05–1.27, $p=0.0026$) and FAHR-FaceSurvival (adjusted HR 1.08 per 0.1 unit, 1.05–1.11, $p<0.0001$) remained associated with survival, whereas FaceAge¹⁰ minus age was not significant (adjusted HR 0.93, 0.82–1.05, $p=0.22$; figure 4I).

FAHR-FaceAge model robustness was evaluated by use of paired images of the same person captured within short timespans from three distinct public datasets, including before and after plastic surgery; with and without make-up; and varying facial expressions and photograph illumination and angles (appendix p 5). FAHR-FaceAge estimates were not significantly different for eyebrow correction, eyelid correction, facelifts, or facial bone correction, but were higher after nose correction (mean difference 1.11 years, $p<0.0001$). No statistical difference was seen with make-up, viewing angle, illumination, or facial expression.

	Univariate*		Multivariate model 1† (FAHR-FaceSurvival continuous)		Multivariate model 2‡ (FAHR-FaceSurvival Quartiles)	
	HR (95% CI)	p value	HR (95% CI)	p value	HR (95% CI)	p value
Age, decade	1.15 (1.11–1.19)	<0.0001	1.08 (1.04–1.12)	<0.0001	1.08 (1.04–1.13)	<0.0001
FAHR-FaceSurvival						
Continuous (0.1)	1.32 (1.30–1.35)	<0.0001	1.17 (1.14–1.20)	<0.0001	..	..
Categorised						
<0.25	1 (ref)	..	..	..	1 (ref)	..
0.25–0.49	2.30 (1.99–2.65)	<0.0001	..	..	1.58 (1.36–1.84)	<0.0001
0.50–0.74	4.29 (3.71–4.97)	<0.0001	..	..	2.24 (1.91–2.63)	<0.0001
≥0.75	8.75 (7.42–10.31)	<0.0001	..	..	3.23 (2.70–3.86)	<0.0001
Sex, male vs female	1.24 (1.11–1.37)	<0.0001	1.03 (0.91–1.15)	0.68	1.03 (0.91–1.16)	0.64
Race or ethnicity‡						
White	1 (ref)	..	1 (ref)	..	1 (ref)	..
Black	0.95 (0.73–1.23)	0.69	1.01 (0.78–1.33)	0.92	0.98 (0.75–1.29)	0.91
Asian	0.79 (0.56–1.12)	0.18	0.72 (0.51–1.02)	0.066	0.73 (0.52–1.04)	0.079
Hispanic	0.88 (0.57–1.37)	0.57	0.67 (0.42–1.07)	0.095	0.65 (0.41–1.04)	0.071
Other	1.42 (1.06–1.90)	0.019	1.38 (1.02–1.85)	0.035	1.39 (1.03–1.87)	0.031
Cancer risk group§						
Non-metastatic cancer	1 (ref)	..	..	..	..	..
Benign conditions or ductal carcinoma in situ	0.23 (0.12–0.43)	<0.0001	..	..	..	..
Metastatic cancer	5.42 (4.87–6.04)	<0.0001	..	..	..	..
Cancer site						
Breast	1 (ref)	..	1 (ref)	..	1 (ref)	..
Benign	1.84 (0.74–4.62)	0.19	1.20 (0.47–3.07)	0.70	1.37 (0.54–3.49)	0.51
Prostate	1.43 (0.91–2.26)	0.12	1.68 (1.00–2.84)	0.052	1.75 (1.04–2.96)	0.036
Lung	12.79 (9.22–17.76)	<0.0001	9.85 (6.78–14.32)	<0.0001	10.26 (7.05–14.93)	<0.0001
Gastrointestinal	8.79 (6.19–12.49)	<0.0001	7.28 (4.96–10.68)	<0.0001	7.66 (5.22–11.24)	<0.0001
Gynaecological	4.26 (2.78–6.51)	<0.0001	3.66 (2.25–5.94)	<0.0001	3.78 (2.33–6.13)	<0.0001
Head and neck	4.54 (3.10–6.65)	<0.0001	4.47 (2.89–6.93)	<0.0001	4.67 (3.01–7.23)	<0.0001
Sarcoma	6.94 (4.34–11.09)	<0.0001	7.18 (4.34–11.87)	<0.0001	7.47 (4.52–12.35)	<0.0001
Haematological	7.31 (5.27–10.14)	<0.0001	3.41 (2.36–4.93)	<0.0001	3.51 (2.43–5.08)	<0.0001
Skin	4.11 (2.37–7.13)	<0.0001	3.27 (1.66–6.43)	0.0006	3.42 (1.74–6.73)	0.00037
Brain (malignant)	21.87 (15.53–30.81)	<0.0001	16.96 (11.42–25.18)	<0.0001	17.37 (11.69–25.80)	<0.0001
Other primary	12.04 (7.54–19.23)	<0.0001	7.46 (4.45–12.53)	<0.0001	8.02 (4.78–13.47)	<0.0001
Bone metastasis	34.27 (25.16–46.67)	<0.0001	8.85 (6.08–12.88)	<0.0001	9.07 (6.23–13.22)	<0.0001
Brain metastasis	35.53 (26.22–48.15)	<0.0001	10.84 (7.33–16.03)	<0.0001	11.28 (7.62–16.70)	<0.0001
Other metastasis	19.07 (14.02–25.93)	<0.0001	7.96 (5.51–11.50)	<0.0001	8.24 (5.70–11.90)	<0.0001
Treatment course intent						
Curative	1 (ref)	..	1 (ref)	..	1 (ref)	..
Oligometastasis ablation	3.99 (3.34–4.77)	<0.0001	1.62 (1.25–2.10)	0.0003	1.60 (1.23–2.07)	0.00045
Palliative	8.63 (7.69–9.68)	<0.0001	3.38 (2.82–4.04)	<0.0001	3.42 (2.86–4.10)	<0.0001
Year treated						
2008–15	1 (ref)	..	1 (ref)	..	1 (ref)	..
2016–23	0.73 (0.66–0.82)	<0.0001	0.79 (0.70–0.90)	0.0002	0.80 (0.71–0.91)	0.00061
Radiation technique¶						
Three-dimensional conformal	1 (ref)	..	1 (ref)	..	1 (ref)	..
Intensity modulated radiotherapy	0.63 (0.55–0.72)	<0.0001	0.67 (0.56–0.80)	<0.0001	0.66 (0.55–0.79)	<0.0001
Stereotactic body radiotherapy	0.99 (0.76–1.29)	0.94	0.58 (0.43–0.77)	0.0002	0.58 (0.43–0.77)	0.00023
Stereotactic radiosurgery	2.48 (2.08–2.96)	<0.0001	0.86 (0.66–1.12)	0.26	0.87 (0.67–1.13)	0.30
Brachytherapy	0.24 (0.14–0.41)	<0.0001	0.49 (0.27–0.88)	0.017	0.48 (0.27–0.87)	0.016
Electrons	0.38 (0.25–0.60)	<0.0001	0.62 (0.37–1.04)	0.069	0.59 (0.35–1.00)	0.049
Adaptive	0.46 (0.24–0.85)	0.014	0.46 (0.24–0.88)	0.018	0.44 (0.23–0.85)	0.014

p values were derived from a two-sided Wald test. HR=hazard ratio. *n=3820 with 1412 events (63.0% censored), unless otherwise specified. †n=3478 with 1318 events (62.1% censored). ‡n=3718. §Excluded from multivariate analysis given significant confounding with cancer site variable. ¶n=3576.

Table 2: Univariate and multivariate Cox regression for overall survival—FAHR-FaceSurvival



(Figure 4 continues on next page)

Attention maps were generated from the trained models to understand which regions of the face contribute most to each individual model's predictions in all patients from the Harvard-RT patient test dataset ($n=3820$). The attention maps for FAHR-FaceSurvival were localised to regions below the eyes and including the nasal bridge, as visualised in figure 3C, and three-dimensional projections onto the canonical face model in the appendix (p 16), indicating that these areas are important for predicting patient survival outcomes. In contrast, the FAHR-FaceAge attention maps (appendix p 16) predominantly highlighted the nasolabial folds, forehead, and temporalis area, underlining the importance of age-related changes in these areas. Age-stratified attention maps (appendix p 16) illustrate further the distinct attention for each model: for FAHR-FaceAge, the regions with high amounts of attention shift over time, including a heavier emphasis on the temporalis and forehead areas in older age groups (40 years and onwards), whereas FAHR-FaceSurvival maintains consistent attention patterns across all age groups.

The independence of the two models is further quantitatively supported by a low correlation between FAD and FAHR-FaceSurvival ($R\ 0.22$; appendix p 16) and minimal feature space overlap (median cosine similarity= -0.018 , appendix p 16) in the Harvard-RT patient test dataset. FAD and FAHR-FaceSurvival showed a low correlation ($R\ 0.16$; figure 4E), showing that the models capture different aspects of patient health from face photographs in the independent Maastricht-RT patient dataset, despite a shared foundation model.

In a combined Cox model, both FAD (adjusted HR 1.07, 95% CI 1.01–1.15, $p=0.030$) and FAHR-FaceSurvival (adjusted HR 1.08, 1.06–1.10, $p<0.0001$) remained independent predictors of survival (figure 4F). Separate Cox models were fitted for FAD and FAHR-FaceSurvival individually, as well as for a combined model, each adjusted for age, sex, and cancer site. The combined model exhibited a lower Akaike information criterion (AIC; 26560) relative to the FAD (AIC 26602) and FAHR-FaceSurvival models (AIC 26563), indicating superior model fit.

Discussion

Our study advances non-invasive health assessment by quantitatively analysing facial images by use of a foundation model trained on over 40 million unlabelled facial photographs. Leveraging masked autoencoders and self-supervised learning, our FAHR-FaceAge model showed superior performance and generalisability in biological age estimation compared with previous models, and we showed that the foundation model can enable training of a clinical outcomes-specific facial health algorithm, FAHR-FaceSurvival, overcoming limitations related to clinical dataset size and generalisability. To our knowledge, this is the first report of an AI algorithm that can directly predict cancer patient survival from a simple face photograph.

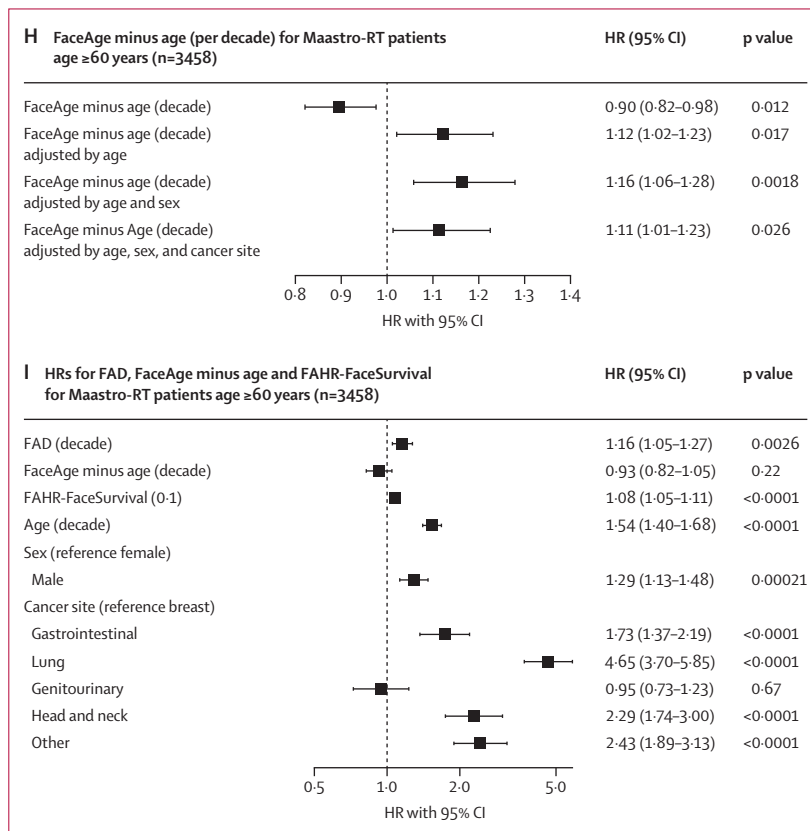


Figure 4: Prognostic performance of FAD and FAHR-FaceSurvival on the external validation dataset

(A) Kaplan–Meier survival curves stratified by FAD groups. Patients with lower FAD values showed higher survival probabilities. The 5-year Kaplan–Meier estimates (95% CI) are provided for each group, with two-sided log-rank p values. (B) Kaplan–Meier survival curves stratified by FAHR-FaceSurvival. Lower FAHR-FaceSurvival was associated with higher survival probabilities. The 5-year Kaplan–Meier estimates (95% CI) are shown for each group, with two-sided log-rank p values. (C) Forest plot showing HRs with 95% CIs for FAD (per decade) in univariate and multivariate analyses, with two-sided Wald test p values. FAD remained a significant predictor of survival after adjusting for age, sex, and cancer site. (D) Forest plot displaying HRs with 95% CIs for FAHR-FaceSurvival (per 0.1 unit) in univariate and multivariate analyses, with two-sided Wald test p values. FAHR-FaceSurvival maintained significant prognostic value after adjusting for age, sex, and cancer site. (E) Scatter plot illustrating the relationship between FAD (years) and FAHR-FaceSurvival. The low but significant correlation ($R=0.16$; $p<0.0001$, two-sided Pearson correlation test) suggests that these measures capture different aspects of patient health. (F) Forest plot comparing HRs with 95% CIs for FAD, FAHR-FaceSurvival, age, sex, and cancer site in a multivariate analysis, with two-sided Wald test p values. Both FAD and FAHR-FaceSurvival remained significant predictors when considered in the same Cox model. (G) AUC of age, FaceAge,¹⁰ and FAHR-FaceAge for survival prediction in those in the Maastricht-RT patient dataset aged 60 years and older ($n=3458$). AUC values were calculated at four timepoints: 3 months, 6 months, 1 year, and 2 years. FAHR-FaceAge consistently showed superior predictive performance across all timepoints, followed by FaceAge,¹⁰ with both outperforming chronological age. The analysis was restricted to patients aged 60 years and older because the FaceAge¹⁰ model was only trained and validated for individuals aged 60 years and older. Error bars indicate 95% CIs obtained from non-parametric bootstrap with 1000 resamples at the patient level. (H) Forest plot showing HRs with 95% CIs for FaceAge¹⁰ minus age (per decade) in univariate and multivariate Cox models among those in the Maastricht-RT patient dataset aged 60 years and older ($n=3458$), with two-sided Wald test p values. FaceAge minus age (per decade) remained a significant predictor of survival after adjustment for age, sex, and cancer site, although the effect size was smaller than for FAD and FAHR-FaceSurvival. (I) Forest plot displaying HRs with 95% CIs from a combined multivariate Cox model including FAD (per decade), FaceAge¹⁰ minus age (per decade), FAHR-FaceSurvival (per 0.1 unit), age, sex, and cancer site for those in the Maastricht-RT patient dataset aged 60 years and older ($n=3458$), with two-sided Wald test p values. In this joint model, FAD and FAHR-FaceSurvival remained independently associated with survival, while FaceAge¹⁰ minus age was no longer significant. AUC=area under the receiver operating characteristic curve. DCIS=ductal carcinoma in situ. FAD=FAHR-FaceAge deviation (FAHR-FaceAge minus chronological age). HR=hazard ratio. Maastricht-RT=Maastricht radiotherapy.

Unlike previous models, which rely on specialised imaging,⁹ our approach effectively uses low-cost, easily acquired consumer-grade facial photographs, which broadens clinical applicability.

Compared with our previously published FaceAge¹⁰ model, which was developed primarily for patients aged 60 years and older in a mostly White cohort, FAHR-FaceAge is built from a foundation model pretrained on more than 40 million faces and age-balanced fine-tuning that spans the full lifespan and all major racial groups. Across every age group and race subgroup we evaluated, the FAD biomarker derived from FAHR-FaceAge showed consistent survival associations for patients who appeared both younger and older than their chronological age, which was not achievable with the previous FaceAge¹⁰ model, particularly in patients younger than 60 years and in Black and other racial subgroups.

Our findings suggest that facial features encode both physiological ageing and imminent mortality risk, showing that consumer-grade photographs carry multiple physiological health signals. Despite their shared backbone, FAHR-FaceAge and FAHR-FaceSurvival offer non-overlapping prognostic information, as supported by attention mapping and multivariate analyses. Attention maps showed their independence; FAHR-FaceSurvival concentrates on the area under the eyes and around the nose, which are regions of the face linked with dying,²² whereas FAHR-FaceAge maps to nasolabial folds and temples, which are regions associated with ageing.^{23,24} Furthermore, the clinical use cases are complementary as well: FAD, given in years, is intuitive and can track ageing or be incorporated into clinical risk stratification tools, whereas FAHR-FaceSurvival might be best suited for near-term mortality prediction.

The growing population of older people makes rapid frailty assessment essential for treatment planning.²⁵ FAHR-Face models could complement or enhance traditional risk calculators, especially where formal assessments (eg, frailty scales, grip-strength, or the Timed Up and Go test)²⁶ are impractical. A single facial photograph analysed by our FAHR-Face models provides a non-invasive, inexpensive alternative that can flag patients unlikely to tolerate aggressive therapy, improving resource allocation and quality of life.^{27,28} As facial features also reflect behaviours such as smoking and alcohol use,^{29,30} the same tool can track lifestyle change or anti-ageing interventions, offering immediate feedback.

Limitations of our study include cancer-only cohorts from mainly centres in high-income countries, which might restrict generalisability to other diseases, regions, and under-represented racial groups. Image quality, lighting, and facial expression can still affect accuracy, underscoring the importance of standardised photograph capture.³¹ Residual demographic bias across racial, ethnic, and age groups also remains a risk, so wider representation, clear regulation and governance, validation in broader cohorts, integration with other biomarkers, and prospective trials that

confirm clinical benefit and inform practical capture standards and clinician training are needed. The clinical role of FAHR-FaceAge and FAHR-FaceSurvival should be refined in further prospective trials that test their use alongside standard prognostic models.

In conclusion, our study shows that quantifying imaging biomarkers from facial photographs through a foundation model approach offers a transformative and accessible method for assessing overall health status and predicting survival outcomes. The foundation model approach provides the ability to train multiple independent models to predict health outcomes, even when relatively limited labelled clinical data are used, which will broaden the potential clinical applications of facial health assessments.

Contributors

FH and GL contributed to conceptualisation, data curation, formal analysis, investigation, methodology, software, validation, visualisation, writing—original draft, and writing—review and editing. JH, LN, DB, DSB, PC, VP, DG, AW, SP, DDR, CG, BHK, and VNG contributed to data curation, formal analysis, investigation, methodology, validation, visualisation, and writing—review & editing. HJWLA and RHM contributed to conceptualisation, funding acquisition, project administration, resources, supervision, writing—review and editing, and methodology. HJWLA and RHM accessed and verified the underlying data. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

RHM reports being on an advisory board for ViewRay and AstraZeneca; consulting for AstraZeneca, Varian Medical Systems, and Sio Capital Management; an honorarium from Novartis and Springer Nature; and research funding from the National Institutes of Health, ViewRay, AstraZeneca, Siemens Medical Solutions USA, and Varian Medical Systems. DSB reports editorial positions, unrelated to this work: Associate Editor of *Radiation Oncology* at HemOnc.org, Associate Editor, *Annals of Oncology*, and Associate Editor, *JCO Clinical Cancer Informatics*; research funding, unrelated to this work: American Cancer Society—American Society for Radiation Oncology (ASTRO-CSDG-24-1244514-01-CTPS) and National Institutes of Health/National Cancer Institute (R01CA294033); and advising and consulting, unrelated to this work: MercuriaAI. HJWLA reports advising and consulting for Onc.AI, Love Health, Sphera, Editas, AstraZeneca, and Bristol Myers Squibb, unrelated to this work. BHK reports research funding from the Botha-Chan Low Grade Glioma Consortium (National Institutes of Health—USA K08DE030216-01). DDR reports institutional financial interests (no personal financial interests) for AstraZeneca (unrestricted research grants and advisory board membership, including consulting or advisory-board fees paid to his institution), Bristol Myers Squibb (unrestricted research grants and advisory board membership, including consulting or advisory-board fees paid to his institution), Siemens (unrestricted research grants and equipment or services), Beigene (unrestricted research grant), Philips (research support and advisory board membership, including consulting or advisory-board fees paid to his institution), Olink (research support and equipment or services), and Eli-Lilly (research support). FH, RHM, and HJWLA report that Mass General Brigham has filed a US provisional patent application on next-generation facial health algorithms: US Provisional Application number 63/769,840. All other authors declare no competing interests.

Data sharing

The WebFace42M dataset used for foundation model training is the cleaned subset of the WebFace260M benchmark and was obtained from the WebFace260M authors under their data-use agreement in December, 2023. Researchers seeking access should contact the WebFace260M dataset authors and follow their access procedures. The datasets used for FAHR-FaceAge model development and evaluation are available through their

respective repositories. The datasets used for robustness evaluation are available online. All URLs listed in this section were last accessed on May 1, 2026. Clinical data from the Harvard-RT patient dataset and Maastricht-RT patient dataset contain protected health information and cannot be made publicly available owing to patient privacy regulations. The code used to train the FAHR-Face, FAHR-FaceAge, and FAHR-FaceSurvival models is available online.

Acknowledgments

This study was funded by the EU European Research Council.

References

- Cox EGM, van Bussel BCT, Campillo Llamazares N, et al, and the SICS Study Group. Facial appearance associates with longitudinal multi-organ failure: an ICU cohort study. *Crit Care* 2024; **28**: 106.
- Gurovich Y, Hanani Y, Bar O, et al. Identifying facial phenotypes of genetic disorders using deep learning. *Nat Med* 2019; **25**: 60–64.
- Rhodes G, Zebrowitz LA, Clark A, Kalick SM, Hightower A, McKay R. Do facial averageness and symmetry signal health? *Evol Hum Behav* 2001; **22**: 31–46.
- Henderson AJ, Holzleitner IJ, Talamas SN, Perrett DI. Perception of health from facial cues. *Philos Trans R Soc Lond B Biol Sci* 2016; **371**: 20150380.
- Gladyshev VN, Kritchevsky SB, Clarke SG, et al. Molecular damage in aging. *Nat Aging* 2021; **1**: 1096–106.
- López-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. Hallmarks of aging: an expanding universe. *Cell* 2023; **186**: 243–78.
- Rothe R, Timofte R, Van Gool L. Deep expectation of real and apparent age from a single image without facial landmarks. *Int J Comput Vis* 2018; **126**: 144–57.
- Bobrov E, Georgievskaya A, Kiselev K, et al. PhotoAgeClock: deep learning algorithms for development of non-invasive visual biomarkers of aging. *Aging (Albany NY)* 2018; **10**: 3249–59.
- Wang J, Gao Y, Wang F, et al. Accurate estimation of biological age and its application in disease prediction using a multimodal image Transformer system. *Proc Natl Acad Sci USA* 2024; **121**: e2308812120.
- Bontempi D, Zalay O, Bitterman DS, et al. FaceAge, a deep learning system to estimate biological age from face photographs to improve prognostication: a model development and validation study. *Lancet Digit Health* 2025; **7**: 100870.
- Haugg F, Lee G, He J, et al. Imaging biomarkers of ageing: a review of artificial intelligence-based approaches for age estimation. *Lancet Healthy Longev* 2025; **6**: 100728.
- Jylhävä J, Pedersen NL, Hägg S. Biological age predictors. *EBioMedicine* 2017; **21**: 29–36.
- Moor M, Banerjee O, Abad ZSH, et al. Foundation models for generalist medical artificial intelligence. *Nature* 2023; **616**: 259–65.
- Pai S, Bontempi D, Hadzic I, et al. Foundation model for cancer imaging biomarkers. *Nat Mach Intell* 2024; **6**: 354–67.
- Lu MY, Chen B, Williamson DFK, et al. A visual-language foundation model for computational pathology. *Nat Med* 2024; **30**: 863–74.
- He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. June 18–24, 2022. <https://doi.org/10.1109/CVPR52688.2022.01553> (accessed Aug 25, 2026).
- Hondru V, Croitoru FA, Minaee S, Ionescu RT, Sebe N. Masked image modeling: a survey. *Int J Comput Vis* 2025; **133**: 7154–200.
- Zhu Z, Huang G, Deng J, et al. WebFace260M: a benchmark unveiling the power of million-scale deep face recognition. Proc IEEE Conf Comp Vis Pattern Recognit; June 20–25, 2021 (abstr 9578817).
- Greco A, Saggese A, Vento M, Vigilante V. Effective training of convolutional neural networks for age estimation based on knowledge distillation. *Neural Comput Appl* 2022; **34**: 21449–64.
- Moqri M, Herzog C, Poganik JR, et al, and the Biomarkers of Aging Consortium. Biomarkers of aging for the identification and evaluation of longevity interventions. *Cell* 2023; **186**: 3758–75.
- Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996; **17**: 343–46.
- Marinella MA. On the hippocentric facies. *J Clin Oncol* 2008; **26**: 3638–40.
- Coleman SR, Grover R. The anatomy of the aging face: volume loss and changes in 3-dimensional topography. *Aesthet Surg J* 2006; **26**: S4–9.
- Coleman SR. Structural fat grafting: more than a permanent filler. *Plast Reconstr Surg* 2006; **118** (suppl): 108S–120S.
- Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet* 2013; **381**: 752–62.
- Chen X, Mao G, Leng SX. Frailty syndrome: an overview. *Clin Interv Aging* 2014; **9**: 433–41.
- Wong YN, Mitra N, Hudes G, et al. Survival associated with treatment vs observation of localized prostate cancer in elderly men. *JAMA* 2006; **296**: 2683–93.
- Hall WH, Jani AB, Ryu JK, Narayan S, Vijayakumar S. The impact of age and comorbidity on survival outcomes and treatment patterns in prostate cancer. *Prostate Cancer Prostatic Dis* 2005; **8**: 22–30.
- Kennedy C, Bastiaens MT, Bajdik CD, Willemze R, Westendorp RG, Bouwes Bavinck JN. the Leiden Skin Cancer Study. Effect of smoking and sun on the aging skin. *J Invest Dermatol* 2003; **120**: 548–54.
- Smith KE, Fenske NA. Cutaneous manifestations of alcohol abuse. *J Am Acad Dermatol* 2000; **43**: 1–16.
- Angulu R, Tapamo JR, Adewumi AO. Age estimation via face images: a survey. *EURASIP J Image Video Process* 2018; **2018**: 42.

For more on **datasets** used see <https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/>, <https://ibug.doc.ic.ac.uk/resources/kanface/>, https://yanweifu.github.io/FG_NET_data/, <https://lbsciurischen.github.io/CARC/>, <https://github.com/John-niu-07/tarball/>, <https://mmlab.ie.cuhk.edu.hk/projects/MegaAge/>, <https://uncw.edu/research/innovation/technology-portfolio/morph/>, <https://wildchlamydia.github.io/lagenda/>, <https://susanqq.github.io/UTKFace/>, <https://ibug.doc.ic.ac.uk/resources/agedb/>, and <https://chalearnlap.cvc.uab.es/dataset/26/description/>

For more on the **robustness evaluation datasets** see <https://dasec.h-da.de/research/biometrics/hda-plastic-surgery-face-database/>, <https://www.kaggle.com/datasets/tapakah68/makeup-detection-dataset>, and <https://fei.edu.br/~cet/facedatabase.html>

For more on the **code used to train the models** see <https://github.com/AIM-Harvard/FAHR-Face>