



# Exploring Healthy Retinal Aging with Deep Learning

Martin J. Menten, PhD,<sup>1,2</sup> Robbie Holland, MSc,<sup>1</sup> Oliver Leingang, PhD,<sup>3</sup> Hrvoje Bogunović, PhD,<sup>3,4</sup> Ahmed M. Hagag, MD,<sup>5,6</sup> Rebecca Kaye, MD,<sup>7</sup> Sophie Riedl, MD,<sup>3</sup> Ghislaine L. Traber, MD,<sup>8,9</sup> Osama N. Hassan, MSc,<sup>1</sup> Nick Pawlowski, PhD,<sup>1,10</sup> Ben Glocker, PhD,<sup>1</sup> Lars G. Fritsche, PhD,<sup>11</sup> Hendrik P.N. Scholl, MD,<sup>8,9</sup> Sobha Sivaprasad, MD,<sup>5,6</sup> Ursula Schmidt-Erfurth, MD,<sup>3</sup> Daniel Rueckert, PhD,<sup>1,2</sup> Andrew J. Lotery, MD,<sup>7</sup> for the PINNACLE Consortium

**Purpose:** To study the individual course of retinal changes caused by healthy aging using deep learning.

**Design:** Retrospective analysis of a large data set of retinal OCT images.

**Participants:** A total of 85 709 adults between the age of 40 and 75 years of whom OCT images were acquired in the scope of the UK Biobank population study.

**Methods:** We created a counterfactual generative adversarial network (GAN), a type of neural network that learns from cross-sectional, retrospective data. It then synthesizes high-resolution counterfactual OCT images and longitudinal time series. These counterfactuals allow visualization and analysis of hypothetical scenarios in which certain characteristics of the imaged subject, such as age or sex, are altered, whereas other attributes, crucially the subject's identity and image acquisition settings, remain fixed.

**Main Outcome Measures:** Using our counterfactual GAN, we investigated subject-specific changes in the retinal layer structure as a function of age and sex. In particular, we measured changes in the retinal nerve fiber layer (RNFL), combined ganglion cell layer plus inner plexiform layer (GCIPL), inner nuclear layer to the inner boundary of the retinal pigment epithelium (INL-RPE), and retinal pigment epithelium (RPE).

**Results:** Our counterfactual GAN is able to smoothly visualize the individual course of retinal aging. Across all counterfactual images, the RNFL, GCIPL, INL-RPE, and RPE changed by  $-0.1 \mu\text{m} \pm 0.1 \mu\text{m}$ ,  $-0.5 \mu\text{m} \pm 0.2 \mu\text{m}$ ,  $-0.2 \mu\text{m} \pm 0.1 \mu\text{m}$ , and  $0.1 \mu\text{m} \pm 0.1 \mu\text{m}$ , respectively, per decade of age. These results agree well with previous studies based on the same cohort from the UK Biobank population study. Beyond population-wide average measures, our counterfactual GAN allows us to explore whether the retinal layers of a given eye will increase in thickness, decrease in thickness, or stagnate as a subject ages.

**Conclusion:** This study demonstrates how counterfactual GANs can aid research into retinal aging by generating high-resolution, high-fidelity OCT images, and longitudinal time series. Ultimately, we envision that they will enable clinical experts to derive and explore hypotheses for potential imaging biomarkers for healthy and pathologic aging that can be refined and tested in prospective clinical trials.

**Financial Disclosure(s):** Proprietary or commercial disclosure may be found after the references. *Ophthalmology Science* 2023;3:100294 © 2023 by the American Academy of Ophthalmology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



Supplemental material available at [www.opthalmologyscience.org](http://www.opthalmologyscience.org).

Many retinal diseases, such as age-related macular degeneration and diabetic retinopathy, develop gradually over time.<sup>1,2</sup> Clinicians are able to track their progression using OCT imaging, which provides high-resolution images of the retina.<sup>3</sup> However, the retina also undergoes age-related physiologic changes.<sup>4</sup> A good understanding of how healthy aging manifests itself in the retina is a crucial prerequisite to distinguish between normal and pathologic changes and effectively diagnose, prognose, and treat ocular diseases.

The retina has been extensively studied by retrospectively or prospectively collecting large amounts of OCT images from representative populations.<sup>5–14</sup> The pooled images are analyzed by measuring the shape and thickness of individual

retinal layers. By identifying population-wide correlations between the eyes' structure and demographic, lifestyle, and medical information, researchers are able to find and validate imaging biomarkers. Supported by the emergence of large population studies and automated tools for processing of medical images,<sup>15</sup> these approaches have successfully found links between age and changes in the nerve fiber layer,<sup>5,6,8,14</sup> ganglion cell complex,<sup>7–9,14</sup> photoreceptor layers,<sup>8,13</sup> and retinal pigment epithelium (RPE).<sup>12</sup>

However, these population-based studies have several shortcomings. Usually, pooled data sets only include a single scan of each eye. Even if time series data are available, it is rare that a subject is monitored for longer than a

couple of years. Furthermore, the imaging conditions change between subsequent visits. The retina may appear differently because of varying levels of pupil dilation, changes in OCT scanner hardware and software, and different orientations of the eye. Consequently, population-based studies are limited in their ability to evaluate the development of the eye at a subject-specific level and resolve subtle retinal changes that occur over the course of decades.

In this study, we used deep learning to study the individual course of retinal changes caused by healthy aging. Our counterfactual generative adversarial network (GAN), a type of neural network, learns from cross-sectional retrospective data. It then synthesizes high-resolution counterfactual OCT images and longitudinal time series. These counterfactuals reflect hypothetical scenarios in which certain characteristics of the imaged subject, such as age or sex, are altered, whereas other attributes, crucially the subject's identity and image acquisition settings, remain fixed. Such counterfactual images allow the investigation of what-if questions that are impossible to answer in population-based studies. Examples of such counterfactual queries are "how will this person's eye look in 20 years?" or "how would this eye look if the subject was born as the opposite sex?" We extensively benchmark the visual fidelity and realism of the generated counterfactual images before ultimately demonstrating the utility of our proposed method by quantifying the subject-specific retinal layer structure as a function of age and sex.

## Methods

An overview of our method and study workflow is presented in [Figure 1](#). After introducing the used data set of OCT images, we describe the counterfactual GAN. Next, we present our experiments to measure the visual fidelity and realism of the artificial OCT images, respectively. Finally, we describe how to extract and analyze the retinal layer structure from the counterfactual images.

### Participants and OCT Image Data Set

We used the OCT image data set that has been acquired as part of the UK Biobank population study. The UK Biobank has collected extensive demographic, lifestyle, health, and medical imaging information from > 500 000 members of the United Kingdom's general public.<sup>16</sup> In its scope, 175 844 retinal OCT scans of 85 709 participants were acquired using a Topcon 3D OCT-1000 Mark II device (Topcon Corporation).<sup>17,18</sup> The UK Biobank population study has been reviewed and approved by the North West Multicentre Research Ethics Committee in accordance with the tenets of the Declaration of Helsinki so that additional ethical approval was not required for our study.

During image preprocessing, we filtered out scans of poor image quality using an intensity-histogram-based score ([Fig 2](#)).<sup>19</sup> We also excluded any subjects that reported being affected by age-related macular degeneration, diabetic retinopathy, glaucoma, cataracts, previous eye trauma, or other serious eye diseases. Next, 11 retinal layer surfaces of the 3-dimensional OCT scans were segmented using the Iowa Reference Algorithms (Retinal Image Analysis Laboratory, Iowa Institute for Biomedical Imaging).<sup>20–22</sup> The obtained layer segmentations were used to flatten and register

all images. During flattening, the images were sheared so that the outer boundary of the RPE is orientated horizontally. The center of the fovea was defined as the position with the minimal distance between the inner limiting membrane and the outer plexiform layer. We extracted the transverse 2-dimensional slice that passed through this position. Finally, all images were resampled to  $224 \times 224$  pixels with a pixel size of  $23.4 \times 7.0 \mu\text{m}^2$ , half the median resolution.

Preprocessing yielded 117 246 highly standardized images of 65 831 subjects. The data set was then split into 3 subdata sets ([Fig 2](#)). Overall, 46 444 scans of the right eye of 46 444 subjects were used for training of the counterfactual GAN. A total of 20 000 images of 10 000 eye pairs were used to train a set of refereed neural networks that evaluated the generated images. Overall, 2112 images of 528 subjects, for which initial and follow-up scans of both eyes were available, were used for final testing. Because images of left eyes were only used for training of the referee networks and final testing but not for GAN training, we ended up not using 48 688 images from our data set.

### GAN to Synthesize Counterfactual OCT Images

The task of generating counterfactual OCT images was formulated as image translation using a GAN.<sup>23</sup> Our counterfactual GAN consists of 2 neural networks, a generator, and a discriminator ([Fig 1](#)). The generator is provided with a real OCT image and a counterfactual query, which consists of the target age and sex encoded as a vector. Given these inputs, the generator is tasked with creating a counterfactual image. These images are provided to the discriminator together with a set of real images from the training data set. For each image, the discriminator has to establish whether it is real or artificially generated. The discriminator also has to estimate the age and sex of the subject in each case. Finally, the counterfactual images are passed through the generator once again with the goal of changing their appearance back to their original state.

Based on these training objectives, the generator and discriminator are trained simultaneously in a zero-sum game.

During inference, the trained generator receives an existing real OCT image and a counterfactual query and creates a corresponding counterfactual image ([Fig 1](#)). The neural network framework was adapted from work by Choi et al.<sup>24</sup> We describe our modifications to it and the full network architecture, training procedure, and used hyperparameters in the [supplemental material](#) (available at [www.ophtalmologyscience.org](http://www.ophtalmologyscience.org)).

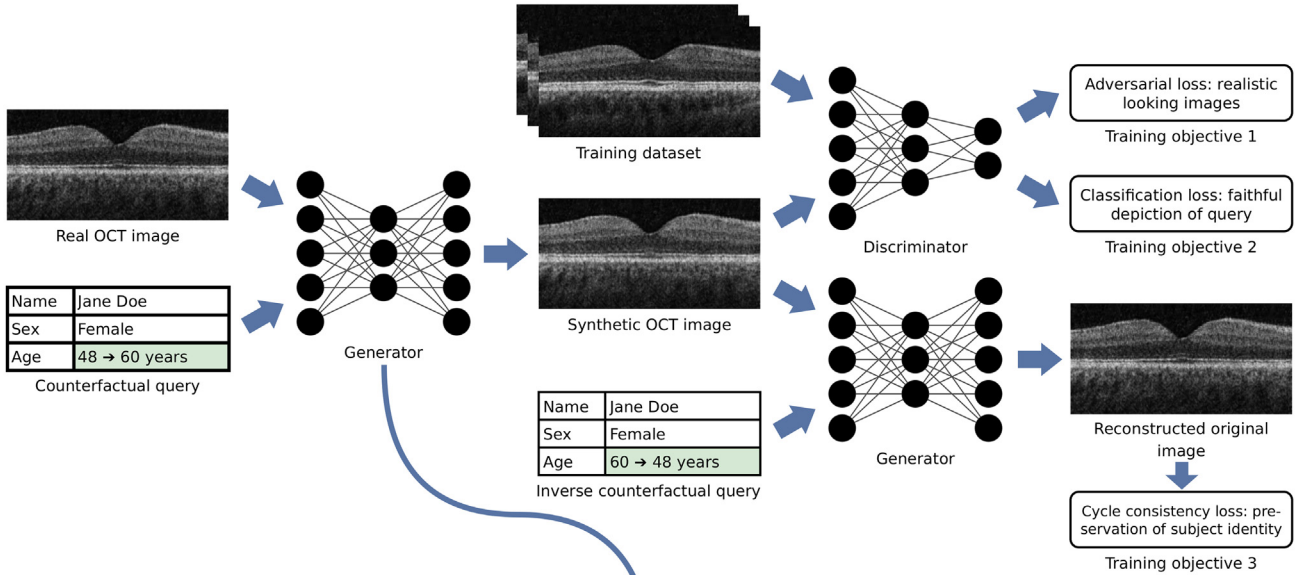
### Visual Turing Test to Assess Image Realism

To quantitatively assess the realism of the counterfactual images, we conducted a visual Turing test. The test measured the ability of 5 expert ophthalmologists (A.M.H., R.K., S.R., G.L.T., A.J.L.) to distinguish between real and artificially generated OCT images. To ensure a fair comparison, all real images were downsampled and flattened according to the previously reported image preprocessing steps. Initially, the participants were given the option to review up to 100 real images. Afterward, they were shown 50 real OCT images and 50 artificially generated images in random order and had to determine which ones were real and which ones were fake. We report the average accuracy of all ophthalmologists.

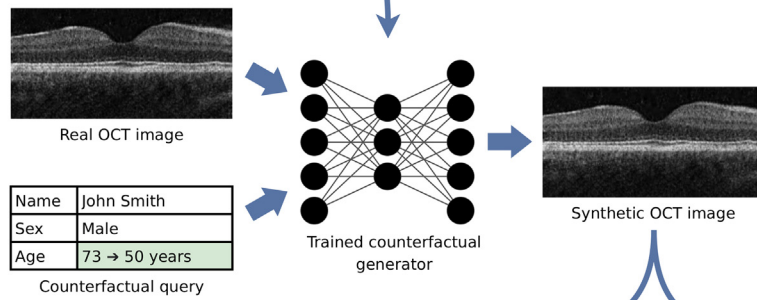
### Neural-Network–Based Quantification of Counterfactual Age, Sex, and Identity

It is crucial that the counterfactual GAN faithfully models the effect of subject age and sex, whereas simultaneously preserving the subject identity. To measure this capability, we trained 3 referee

### Training of counterfactual GAN



### Generating counterfactual OCT images



### Benchmarking and using counterfactual OCT images

#### Visual Turing test to assess image realism

Turing test

Ophthalmologist

#### Neural-network based quantification of age, sex and identity

Age prediction network: Detected age: 51 years ✓

Sex classification network: Detected sex: male ✓

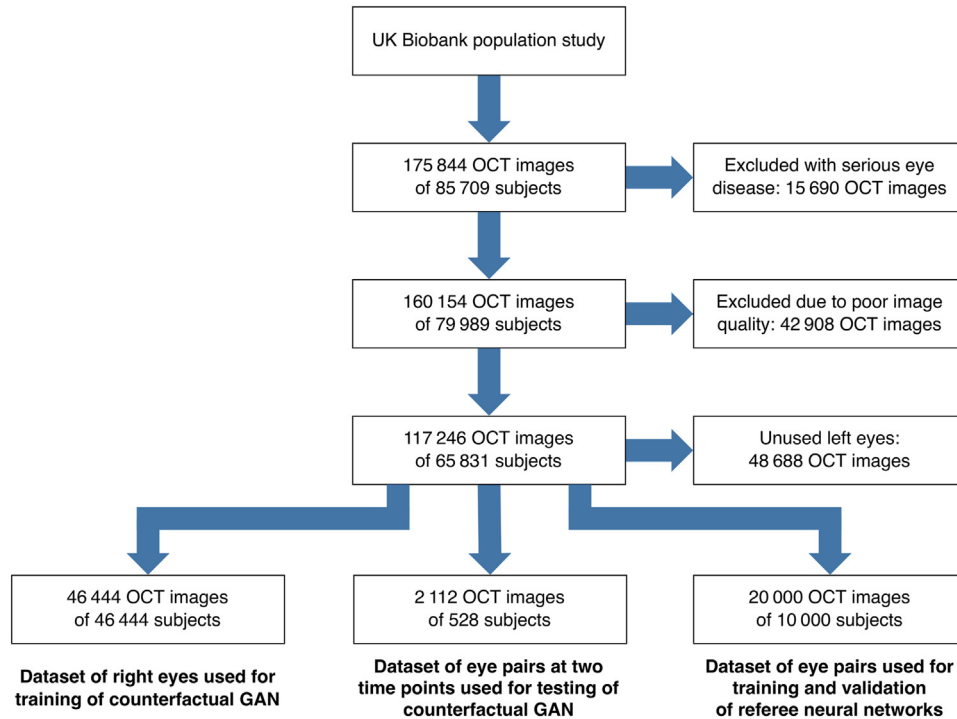
Identity matching network: Detected identity: John Smith ✓

#### Extraction and analysis of the retinal layer structure

Layer	Thickness
RNFL	22.6 μm
GCIPL	67.2 μm
INL-ELM	117.4 μm
ELM-ISOS	11.9 μm
ISOS-RPE	34.6 μm
RPE	22.0 μm

Temporal development of retinal layer structure

**Figure 1.** Workflow diagram explaining how the counterfactual generative adversarial network (GAN) is trained (top part), used to generate counterfactual OCT images (middle part) and benchmarked and utilized (bottom part). ELM = external limiting membrane; GCIPL = ganglion cell layer plus inner plexiform layer; INL = inner nuclear layer; ISOS = inner segment/outer segment junction layer; RNFL = retinal nerve fiber layer; RPE = retinal pigment epithelium.



**Figure 2.** Data flowchart presenting the data inclusion and exclusion criteria as well as the final split into the 3 independent data sets used in the study. GAN = generative adversarial network.

neural networks built according to the well-established Resnet50 architecture.<sup>25</sup> The age prediction network estimates the subject's age from an input OCT image. The sex classification network predicts whether a given OCT image belongs to a male or female subject. The identity-matching network learns to assign a similarity score to an image pair consisting of right and left eyes. A high similarity score indicates that the 2 eyes belong to the same subject. Full network configuration and training procedure are included in the [supplemental material](#).

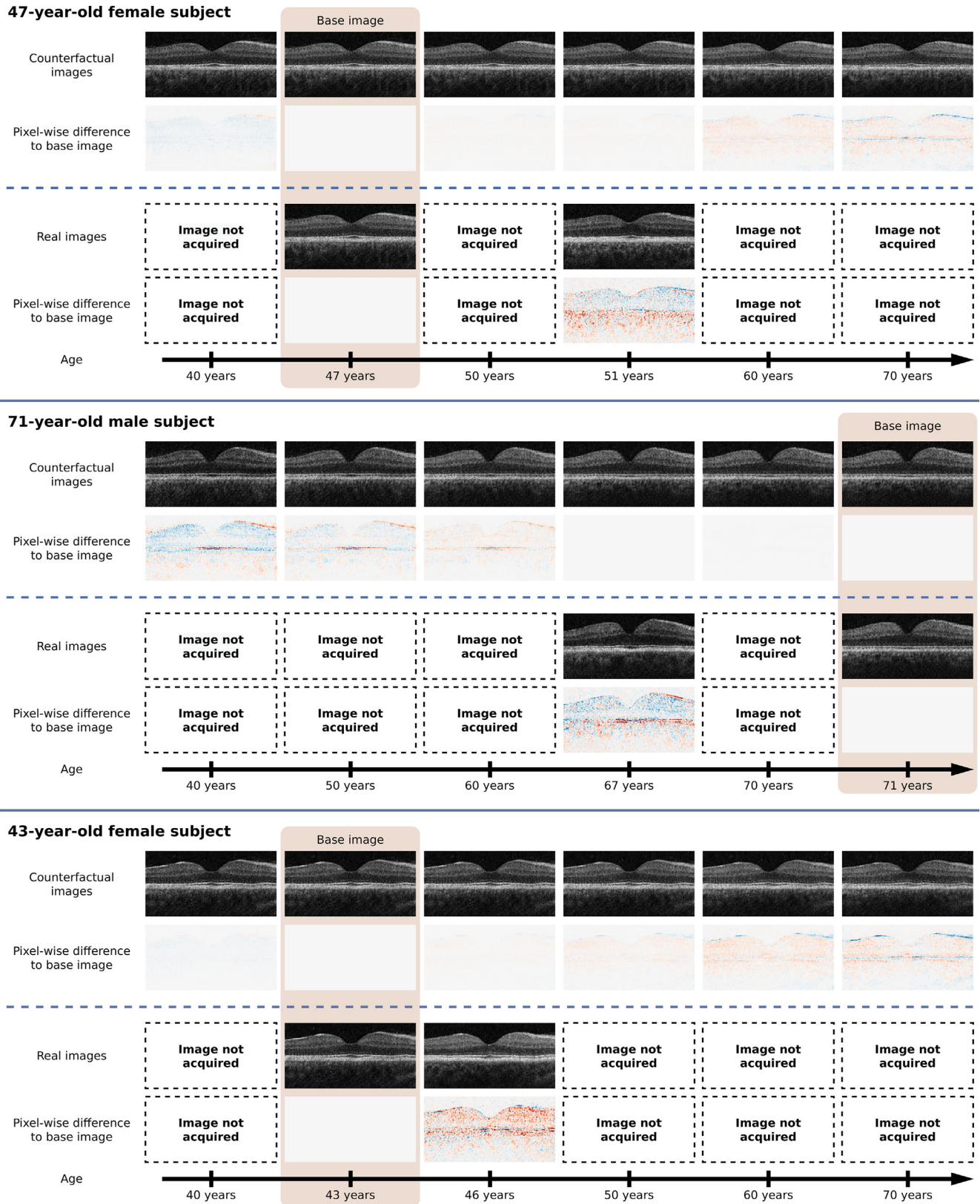
Before ultimately using the 3 referee networks to evaluate the counterfactual images, we benchmarked their performance on an independent subset of real OCT images. The age regression network estimated the subject's age with a mean absolute error of 4.1 years. The sex classification network determined the subject's sex with an accuracy of 79.5% and an area under the receiver operating characteristic curve of 0.90. The identity-matching network was tasked with matching 2000 right and left eyes belonging to 1000 different subjects and achieved a sensitivity of 95.8% and a specificity of 97.7%. During evaluation, the 3 referee networks were shown 10 000 counterfactual images. The counterfactual queries were evenly split between male and female sex and distributed across a uniform age distribution spanning from 40 to 75 years, the age range of the subjects in the training data set. We report whether the determined age, sex, and identity matched the corresponding counterfactual queries.

### Extraction and Analysis of the Retinal Layer Structure

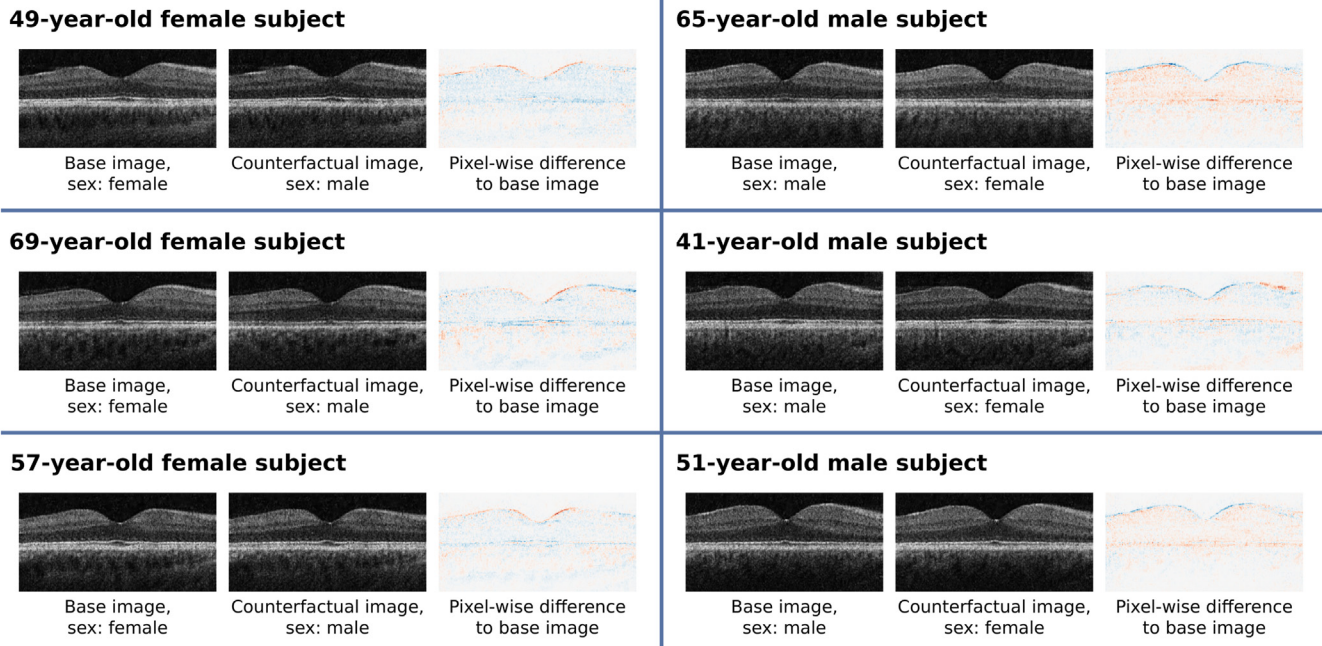
Finally, we analyzed the retinal structure in the OCT images. To this end, we trained a Resnet50 neural network to segment 11 retinal surfaces following the approach by Shah et al.<sup>26</sup> We used

real OCT images and the previously obtained layer segmentations as training data. The network was able to accurately localize the retinal layers in an independent subset of 1000 real OCT images. Across all 11 layers, the mean absolute difference between the predicted and ground truth layer segmentations was  $4.2 \mu\text{m} \pm 7.5 \mu\text{m}$ . Because we did not have ground truth annotations for the counterfactual OCT images, we could not quantitatively assess the segmentation network's performance on these images. However, we visually confirmed that the network is able to delineate the layers in artificially generated images before proceeding to process all counterfactual images. Additionally, we segmented and analyzed the real OCT images that were used to train the counterfactual GAN, resembling conventional population-based studies.<sup>12–14</sup> More details can be found in the [supplementary material](#).

In this study, we focused on the retinal nerve fiber layer (RNFL), combined ganglion cell layer plus inner plexiform layer (GCIPL), inner nuclear layer to the inner boundary of the retinal pigment epithelium (INL-RPE), which contains the photoreceptor layers, and retinal pigment epithelium (RPE). We chose these retinal layers because their age-related changes have previously been researched using the UK Biobank database.<sup>12–14</sup> We report the average thickness as well as the effect of age and sex for each of the 4 layers. We further calculated these measures in each of the following 5 subfields of the retina, the outer temporal subfield, inner temporal subfield, central subfield, inner nasal subfield, and outer nasal subfield. Because the analysis was conducted using 2-dimensional images, we cannot report results for the superior and inferior subfields. To better compare our findings with those of other studies, we corrected our 1-dimensional measurements by assuming the same thickness profile in the entire 2-dimensional subfield.



**Figure 3.** The counterfactual generative adversarial network smoothly visualizes the process of healthy retinal aging at a subject-specific level. In each of the 3 representative examples, the first row presents the counterfactual time series as a function of age. The third row shows the 2 available real images from the UK Biobank data set. The second and fourth rows depict the pixel-wise difference between the time series image and the base image. Red and blue color denote image regions in which the counterfactual is brighter and darker, respectively.



**Figure 4.** Six representative examples in which the counterfactual generative adversarial network alters a real retinal OCT scan (left image) to appear as if the subject was born as the opposite sex (middle image). The pixel-wise difference between the 2 images is also shown (right image). In red areas the counterfactual is brighter than the base image and in blue regions the counterfactual is darker.

## Results

### Counterfactual OCT Images to Visualize the Impact of Healthy Retinal Aging and Subject Sex

Our counterfactual GAN can smoothly visualize the individual course of retinal aging. Based on a single input image, it provides a plausible hypothesis of how a specific eye will look several decades into the future or how it appeared in the past (Fig 3). By comparison, population-based approaches are limited with regard to the availability, frequency, and range of time series data. In the case of the UK Biobank data set, follow-up OCT scans were acquired from < 5% of the subjects and were dated only 2 to 4 years after the initial scan. Retinal layer orientation, image brightness, and contrast are preserved in the counterfactual images, whereas it fluctuates in the follow-up scans of the UK Biobank data set. This allows focusing on subtle retinal changes, which are difficult to appreciate in conventionally acquired time series. When visually inspecting the generated counterfactual time series, we found that increased age was associated with changes in several retinal layers, including the RNFL, photoreceptor layers, and RPE. This agrees with the previously reported findings.<sup>5,6,8,10–14</sup>

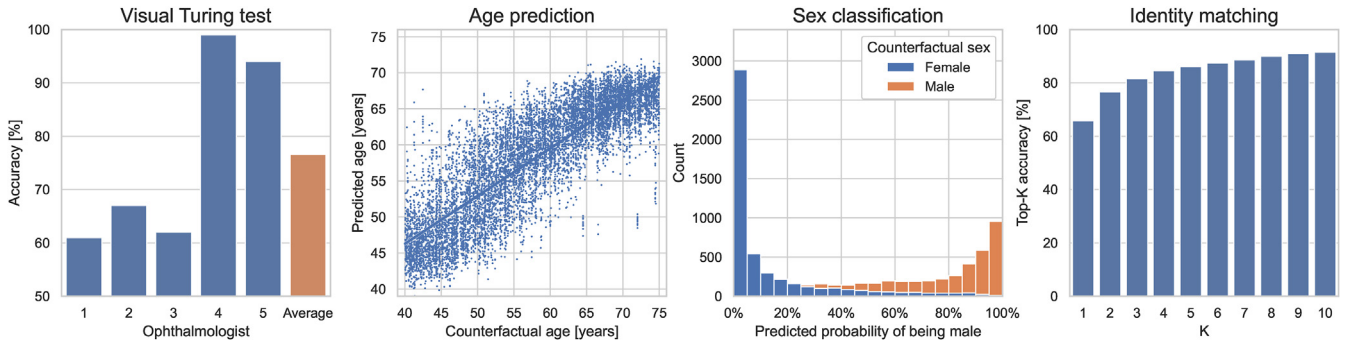
The counterfactual GAN can also simulate how an individual's eye would appear if the subject had been born as the opposite sex (Fig 4). Naturally, this counterfactual scenario is not possible to research in population-based studies. When changing female eyes to a male appearance, we consistently observe that the foveal pit becomes slightly deeper and steeper. In many cases, the overall macular

thickness increases. Conversely, when counterfactually converting male eyes to female, the retina becomes shallower and thinner.

### Benchmarking of Counterfactual OCT Images

The realism of the counterfactual OCT images was quantified in a visual Turing test. The mean accuracy across the 5 ophthalmologists was  $76.6\% \pm 18.4\%$  (Fig 5). Although this is substantially better than random choice (i.e., 50% accuracy), the ophthalmologists were not able to correctly distinguish whether an image was real or artificially generated in many cases. Two ophthalmologists achieved a substantially higher accuracy of 98% and 94%, respectively, by looking at the choroid and vitreous in the background of the images. They also relied on spotting pathologic features as well as shadowing artifacts caused by blood vessels because these would mostly occur in real OCT images. All ophthalmologists agreed that the counterfactual GAN produces samples with a highly realistic-looking retinal layer structure, which we focus on in the remaining study.

Next, we used the referee networks to predict age, sex, and identity from counterfactual images. We measured whether these attributes matched the corresponding counterfactual queries. The age estimated by the age prediction network agreed with the counterfactual query with a mean absolute error of  $4.2 \text{ years} \pm 0.4 \text{ years}$ , while being strongly correlated (Pearson's  $R$  of  $0.86 \pm 0.02$ ; Fig 5). The sex classification network correctly predicted the counterfactual sex with an accuracy of  $79.7\% \pm 5.8\%$  and



**Figure 5.** Benchmarking of the counterfactual OCT images. The visual Turing test assessed the images’ realism (left-most graph). Referee neural networks determined the subject’s age and sex from the counterfactual images. We measured whether their prediction agreed with the counterfactual query (middle 2 graphs). A third referee neural network matched counterfactual right eyes with real left eyes. We assessed whether the correct pairing is among the top K guesses of the network as this indicated that the subject identity was preserved in the counterfactual images (right-most graph).

area under the receiver operating characteristic curve of  $0.92 \pm 0.03$  (Fig 5). Finally, we tested whether the identity was preserved in the counterfactual OCT images. We counterfactually increased the age in 528 OCT scans to match the subject’s age at the time of a follow-up scan. The identity-matching network compared the images of the artificially aged right eyes with images of the real left eyes. In  $65.8\% \pm 9.1\%$  of the cases, the referee network correctly matched the right eye to its corresponding left eye, while being given 1000 candidate eyes (top-1 accuracy; Fig 5). In  $91.5\% \pm 3.3\%$  of cases, the correct eye is among the top 10 guesses (top-10 accuracy). Even considering the residual error of all referee networks, these results quantitatively confirm that the counterfactual GAN is able to faithfully simulate the effect of age and sex on the retina, while preserving the identity of the subject.

### Retinal Layer Structure in Counterfactual Images

We segmented and analyzed the RNFL, GCIPL, INL-RPE, and RPE in the counterfactual images. Figure 6 presents their mean thickness as well as the change per decade aging and effect of subject sex. The thickness of the RNFL, GCIPL, and INL-RPE decreases as we increase the counterfactual age, whereas the RPE grows slightly with age. The RNFL and RPE are thicker in male subjects than in female subjects. We also obtained the same set of measurements directly from the real OCT images that were used to train the counterfactual GAN. This approach is similar to a conventional population-based study. The absolute thickness of the 4 retinal structures is very similar in the 2 different approaches. The counterfactual GAN accurately learned to model the impact of age and sex in the RNFL and RPE, while slightly underestimating the effect in the large GCIPL and INL-RPE structures.




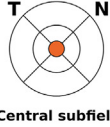


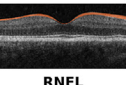
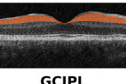
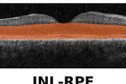
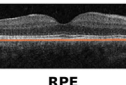
### Discussion

In this study, we have created a counterfactual GAN to investigate the individual course of retinal changes caused

by healthy aging. Learning from a data set with only 1 OCT image per subject, our machine learning algorithm is able to generate synthetic longitudinal time series. It can visualize how the retina may develop with age at a subject-specific level. This allows studying of subtle structural changes that occur over the course of decades and cannot be resolved in conventional population-based studies. The counterfactual GAN can also simulate how a given eye would look if a person was born as the opposite sex, a scenario that cannot be researched naturally. In extensive benchmarking experiments, we confirmed that our tool creates realistic OCT images and faithfully models the influence of subject age, sex, and identity on the retina.

Our results agree well with previous studies based on the same cohort from the UK Biobank population study.<sup>12–14</sup> Khawaja et al<sup>14</sup> reported very similar absolute thicknesses and impact of aging for the RNFL and GCIPL. They found that both structures are thinner in male than in female subjects, whereas we found that sex only has a small impact on the GCIPL and the opposite effect on the RNFL. The association between age and photoreceptor thickness was researched by Chua et al.<sup>13</sup> Our measured INL-RPE thickness was slightly larger than the reported value. The observed relationship between thickness and age was the same in their study and ours. Ko et al<sup>12</sup> measured the RPE-Bruch membrane complex. In both their study and ours, the RPE thickness was barely affected by subject sex. They found that the complex thins with increasing age in subjects that are  $\geq 45$  years old, which we did not observe. These differences are potentially caused by our layer segmentation algorithm not fully outlining the fine Bruch membrane in the lower-resolution images. This hypothesis is supported by the fact that we detect a slightly thinner RPE structure compared with their reported thicknesses. In all 3 comparisons, minor discrepancies may also result from the use of 2-dimensional instead of 3-dimensional images in our study and different algorithms for retinal layer segmentation.

Previous work has explored the use of GANs for a range of tasks in the field of ophthalmology, such as image denoising, superresolution, and domain transfer.<sup>27–30</sup> In these applications, GANs alter medical images to reflect an

													
		Combined subfields		Temporal outer subfield		Temporal inner subfield		Central subfield		Nasal inner subfield		Nasal outer subfield	
		Counterfactual GAN	Population-based	Counterfactual GAN	Population-based	Counterfactual GAN	Population-based	Counterfactual GAN	Population-based	Counterfactual GAN	Population-based	Counterfactual GAN	Population-based
	Average thickness [ $\mu\text{m}$ ]	23.6 $\pm$ 2.8	23.7 $\pm$ 2.9	34.4 $\pm$ 5.4	35.7 $\pm$ 5.6	25.1 $\pm$ 2.5	25.0 $\pm$ 3.1	10.4 $\pm$ 3.4	9.6 $\pm$ 3.6	19.9 $\pm$ 3.9	19.6 $\pm$ 3.6	15.1 $\pm$ 4.1	14.5 $\pm$ 3.6
	Change per decade aging [ $\mu\text{m}$ ]	-0.1 (-0.1, -0.1)	-0.1 (-0.1, -0.1)	-0.4 (-0.4, -0.3)	-0.5 (-0.5, -0.4)	-0.2 (-0.2, -0.1)	-0.2 (-0.2, -0.1)	+0.2 (+0.2, +0.3)	+0.5 (+0.5, +0.6)	0.0 (-0.1, 0.0)	0.0 (0.0, +0.1)	+0.1 (+0.1, +0.2)	+0.2 (+0.2, +0.3)
	Effect of male sex [ $\mu\text{m}$ ]	+0.5 (+0.2, +0.8)	+0.4 (+0.3, +0.5)	+0.8 (+0.1, +1.4)	+0.7 (+0.5, +0.8)	+0.5 (+0.1, +0.8)	+0.4 (+0.3, +0.5)	+1.0 (+0.6, +1.4)	+1.3 (+1.2, +1.3)	+0.2 (-0.2, +0.7)	+0.2 (+0.1, +0.2)	+0.4 (-0.1, +0.8)	+0.3 (+0.2, +0.3)
	Average thickness [ $\mu\text{m}$ ]	72.8 $\pm$ 8.2	73.9 $\pm$ 8.7	73.9 $\pm$ 8.9	74.5 $\pm$ 9.1	86.5 $\pm$ 9.3	87.7 $\pm$ 9.8	40.6 $\pm$ 7.4	41.2 $\pm$ 8.0	80.8 $\pm$ 10.2	82.4 $\pm$ 10.6	66.1 $\pm$ 9.8	67.2 $\pm$ 10.5
	Change per decade aging [ $\mu\text{m}$ ]	-0.5 (-0.6, -0.4)	-1.6 (-1.5, -1.7)	-0.7 (-0.7, -0.6)	-2.0 (-2.1, -1.9)	-0.8 (-0.9, -0.7)	-1.7 (-1.8, -1.6)	-0.1 (-0.2, 0.0)	-0.3 (-0.4, -0.2)	-0.6 (-0.7, -0.5)	-1.6 (-1.8, -1.5)	-0.3 (-0.4, -0.2)	-1.2 (-1.4, -1.2)
	Effect of male sex [ $\mu\text{m}$ ]	-0.1 (-1.1, +0.9)	+0.5 (+0.3, +0.6)	-2.6 (-3.7, -1.6)	-1.4 (-1.6, -1.2)	+0.6 (-0.6, +1.7)	+1.7 (+1.7, +1.9)	+3.2 (+2.3, +4.1)	+3.9 (+3.6, +4.1)	+1.6 (+0.4, +2.9)	+1.8 (+1.6, +2.0)	+0.9 (-0.3, +2.1)	+0.9 (+0.7, +1.1)
	Average thickness [ $\mu\text{m}$ ]	160.4 $\pm$ 12.2	159.5 $\pm$ 11.8	152.2 $\pm$ 11.8	151.4 $\pm$ 12.4	175.4 $\pm$ 12.4	175.6 $\pm$ 14.1	193.7 $\pm$ 13.0	193.4 $\pm$ 18.2	173.5 $\pm$ 17.1	172.6 $\pm$ 15.7	152.8 $\pm$ 17.3	151.6 $\pm$ 17.1
	Change per decade aging [ $\mu\text{m}$ ]	-0.2 (-0.3, -0.1)	-0.9 (-1.1, -0.8)	-0.1 (-0.2, +0.0)	-1.0 (-1.2, -0.9)	+0.2 (+0.1, +0.3)	-0.7 (-0.9, -0.6)	-0.2 (-0.3, -0.1)	-1.0 (-1.2, -1.8)	-0.2 (-0.4, -0.1)	-0.9 (-1.1, -0.8)	-0.4 (-0.6, -0.2)	-1.0 (-1.2, -0.8)
	Effect of male sex [ $\mu\text{m}$ ]	+0.9 (-0.6, +2.3)	+2.6 (+2.4, +2.8)	+0.5 (-0.9, +1.9)	+1.9 (+1.7, +2.2)	+0.7 (-0.8, +2.2)	+3.1 (+2.8, +3.4)	+1.7 (+0.1, +3.3)	+4.0 (+3.6, +4.3)	+0.7 (-1.3, +2.8)	+2.9 (+2.6, +3.2)	+1.3 (-0.8, +3.4)	+2.9 (+2.6, +3.2)
	Average thickness [ $\mu\text{m}$ ]	22.8 $\pm$ 2.1	22.3 $\pm$ 2.4	23.5 $\pm$ 2.5	23.0 $\pm$ 2.7	22.5 $\pm$ 2.4	21.8 $\pm$ 2.8	23.2 $\pm$ 2.3	22.8 $\pm$ 2.6	21.5 $\pm$ 2.9	20.8 $\pm$ 3.1	22.8 $\pm$ 3.0	22.4 $\pm$ 3.2
	Change per decade aging [ $\mu\text{m}$ ]	+0.1 (+0.1, +0.2)	+0.3 (+0.2, +0.3)	+0.1 (+0.0, +0.1)	+0.2 (+0.1, +0.2)	0.0 (0.0, +0.1)	+0.3 (+0.2, +0.3)	+0.3 (+0.3, +0.3)	+0.6 (+0.6, +0.7)	+0.3 (+0.2, +0.3)	+0.5 (+0.4, +0.5)	+0.2 (+0.1, +0.2)	+0.3 (+0.2, +0.3)
	Effect of male sex [ $\mu\text{m}$ ]	+0.6 (+0.3, +0.8)	+0.2 (+0.1, +0.2)	+0.7 (+0.4, +1.0)	+0.4 (+0.3, +0.4)	+0.3 (0.0, +0.6)	0.0 (-0.1, 0.0)	-0.1 (-0.4, +0.2)	+0.1 (+0.1, +0.2)	+0.3 (0.0, 0.7)	0.0 (0.0, +0.1)	+0.7 (+0.3, +1.0)	+0.2 (+0.0, +0.4)

Acronyms: GAN - generative adversarial network; RNFL - retinal nerve fiber layer; GCPL - combined ganglion cell layer plus inner plexiform layer; INL-RPE - inner nuclear layer to inner boundary of retinal pigment epithelium; RPE - retinal pigment epithelium

**Figure 6.** Average thickness, change per decade aging, and effect of male sex for 4 different retinal layer structures and 5 retinal subfields. In each case, we compare the findings obtained by analyzing the counterfactual images (counterfactual GAN) and real OCT images (population-based). The average thickness is reported as mean with its standard deviation, whereas the changes caused by aging and sex are listed as average difference and their 95% confidence intervals.

improved or functionally different image acquisition process. Ideally, any image transformations would not change information content related to the patient. Conversely, our study researches the setting in which images are altered to reflect changes in the imaged subjects themselves, whereas the acquisition settings are kept fixed. To our knowledge, there is only 1 other study exploring counterfactual image generation for biomarker discovery in ophthalmology. Narayanaswamy et al<sup>31</sup> have previously proposed a counterfactual synthesis of color fundus photographs to discover indicators of diabetic macular edema. They found that the disease state is linked to the presence of exudates, a known biomarker for diabetic macular edema, as well as a darkening of the foveal region, which is currently not being used for clinical predictions. However, they have not quantitatively assessed the quality of the counterfactual images and did not extract imaging biomarkers from the images. Nonetheless, their study showcases an exciting usage for our tool, modeling the effect of ocular disease on the eye.

At the moment, our counterfactual GAN assumes that the eye's appearance in OCT images is governed independently by the subject's age, sex, and identity. Although we aimed to exclude any patients affected by serious eye disease, some eyes with early-stage disease potentially remain in the training data set. The GAN may inadvertently learn to correlate these disease features with age or sex and alter them when generating counterfactual images. To avoid such

artifacts, future work could see the creation of a more sophisticated causal model and its integration with a GAN.<sup>32,33</sup> Such a model could include and explicitly model the relationship between the eye's appearance and subject genotype, lifestyle, or retinal diseases. However, this requires the availability of corresponding labels in the data set that is used to train the GAN. Furthermore, the algorithm cannot learn the relationship for groups of subjects that it has not seen in the data set. For example, our GAN has been trained on subjects between the ages of 40 and 75 years. It is not able to model how the eye develops in children, young adults, or individuals that are > 75 years old. Finally, the counterfactual GAN currently only generates 2-dimensional OCT images. Although generating volumetric images with GANs is challenging,<sup>34</sup> future work should look to increase in the images' dimensionality as well as their field-of-view and resolution.

In conclusion, this study has demonstrated how counterfactual GANs can aid research into retinal aging by synthesizing high-resolution, high-fidelity OCT images, and longitudinal time series. Ultimately, we envision that they will enable clinical experts to derive and explore hypotheses for potential imaging biomarkers for healthy and pathologic aging that can be refined and tested in prospective clinical trials.

### Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 45477.



## Footnotes and Disclosures

Originally received: October 30, 2022.

Final revision: January 24, 2023.

Accepted: February 17, 2023.

Available online: March 1, 2023. Manuscript no. XOPS-D-22-00233R1.

<sup>1</sup> BioMedIA, Imperial College London, London, United Kingdom.

<sup>2</sup> Institute for AI and Informatics in Medicine, Technical University of Munich, Munich, Germany.

<sup>3</sup> Laboratory for Ophthalmic Image Analysis, Medical University of Vienna, Vienna, Austria.

<sup>4</sup> Christian Doppler Laboratory for Artificial Intelligence in Retina, Christian Doppler Forschungsgesellschaft, Vienna, Austria.

<sup>5</sup> Institute of Ophthalmology, University College London, London, United Kingdom.

<sup>6</sup> Moorfields Eye Unit, National Institute for Health Research, London, United Kingdom.

<sup>7</sup> Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom.

<sup>8</sup> Institute of Molecular and Clinical Ophthalmology Basel, Basel, Switzerland.

<sup>9</sup> Department of Ophthalmology, University of Basel, Basel, Switzerland.

<sup>10</sup> Microsoft Research, Cambridge, United Kingdom.

<sup>11</sup> Department of Biostatistics, University of Michigan, Ann Arbor, Michigan.

Presented at the Annual Meeting of the Association for Research in Vision and Ophthalmology, May 1-4, 2022, Denver, Colorado.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The authors made the following disclosures:

A.L.: Support — Roche; Consultant — Roche, Eyebio; Participation — Apellis; Stocks — Gyroscope Therapeutics.

A.M.H.: Employment — Boehringer Ingelheim.

B.G.: Grant — EU Commission, Innovate UK; Participation — Kheiron Medical Technologies; Stock — Kheiron Medical Technologies, HeartFlow.

D.R.: Grant — EU Commission, Innovate UK, German Federal Ministry of Education and Research, Engineering and Physical Sciences Research Council, German Research Foundation; Consultant — HeartFlow, IXICO.

H.B.: Grant — RetInSight, Apellis; Payment — Bayer, Roche, Apellis.

H.P.N.S.: Grant — the Swiss National Science Foundation, National Center of Competence in Research Molecular Systems Engineering, Foundation Fighting Blindness Clinical Research Institute; Consultant — Boehringer Ingelheim, Claris Biotherapeutics Inc, Eluminex Biosciences, Gyroscope Therapeutics Ltd, Janssen Research & Development, LLC (Johnson &

Johnson), Okuvision GmbH, ReVision Therapeutics Inc, Saliogen Therapeutics Inc, Alnylam Pharmaceuticals Inc, Gerson Lehrman Group Inc, Guidepoint Global, LLC, Intergalactic Therapeutics Inc; Patents — Methods to assess cone availability in patients with vision impairments, Human Cone Photoreceptor Optogenetic Constructs, Promoter for specific expression of genes in rod photoreceptors; Participation — Belite Bio, Roche, Novo Nordisk.

N.P.: Employment — Microsoft Research, Google Health.

S.S.: Grants — Boehringer Ingelheim, Bayer, AbbVie, Optos; Consultant — Bayer, Novartis, Biogen, Roche, Apellis, Eyebio, Boehringer Ingelheim; Payment — Bayer, Roche, Boehringer Ingelheim, AbbVie; Support — Bayer, Roche; Participation — Bayer; Leadership — Macular Society Trustee, Trustee of the Royal College of Ophthalmologists.

U.S.E.: Grant — Gentech, Kodiak, Novartis; Royalties — RetInSight; Consultant — Apellis.

Supported by the Wellcome Trust (London, United Kingdom) in the scope of the Wellcome Trust Collaborative Award, “Deciphering AMD by deep phenotyping and machine learning” Ref. 210572/Z/18/Z. The funding organization had no role in the design or conduct of this research.

**HUMAN SUBJECTS:** Human subjects were included in this study. The UK Biobank population study has been reviewed and approved by the North West Multicentre Research Ethics Committee in accordance to the tenets of the Declaration of Helsinki so that additional ethical approval was not required for this study.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Menten, Rueckert, Lotery

Data collection: Menten, Leingang, Hagag, Kaye, Riedl, Traber, Fritsche, Scholl, Sivaprasad, Schmidt-Erfurth, Lotery

Analysis and interpretation: Menten, Holland, Bogunović, Hassan, Pawlowski, Glocker, Lotery

Obtained funding: Fritsche, Scholl, Sivaprasad, Schmidt-Erfurth, Rueckert, Lotery

Overall responsibility: Menten, Rueckert, Lotery

Abbreviations and Acronyms:

**GAN** = generative adversarial network; **GCIPL** = ganglion cell layer plus inner plexiform layer; **INL** = inner nuclear layer; **RNFL** = retinal nerve fiber layer; **RPE** = retinal pigment epithelium.

Keywords:

Aging, Biomarker discovery, Deep learning, Machine learning, Retina.

Correspondence:

Martin J. Menten, Imperial College London, South Kensington Campus, SW7 2AZ, London, United Kingdom. E-mail: [m.menten@imperial.ac.uk](mailto:m.menten@imperial.ac.uk).

## References

- Stitt AW, Curtis TM, Chen M, et al. The progress in understanding and treatment of diabetic retinopathy. *Prog Retin Eye Res.* 2016;51:156–186.
- Mitchell P, Liew G, Gopinath B, Wong TY. Age-related macular degeneration. *Lancet.* 2018;392:1147–1159.
- Adhi M, Duker JS. Optical coherence tomography—current and future applications. *Curr Opin Ophthalmol.* 2013;24:213–221.
- Gao H, Hollyfield JG. Aging of the human retina. Differential loss of neurons and retinal pigment epithelial cells. *Invest Ophthalmol Vis Sci.* 1992;33:1–17.
- Leung CK, Cheung CY, Weinreb RN, et al. Retinal nerve fiber layer imaging with spectral-domain optical coherence tomography: a variability and diagnostic performance study. *Ophthalmology.* 2009;116:1257–1263.
- Sung KR, Wollstein G, Bilonick RA, et al. Effects of age on optical coherence tomography measurements of healthy retinal nerve fiber layer, macula, and optic nerve head. *Ophthalmology.* 2009;116:1119–1124.
- Mwanza JC, Durbin MK, Budenz DL, et al. Profile and predictors of normal ganglion cell—inner plexiform layer thickness measured with frequency-domain optical coherence

- tomography. *Invest Ophthalmol Vis Sci.* 2011;52:7872–7879.
8. Ooto S, Hangai M, Tomidokoro A, et al. Effects of age, sex, and axial length on the three-dimensional profile of normal macular layer structures. *Invest Ophthalmol Vis Sci.* 2011;52:8769–8779.
  9. Koh VT, Tham YC, Cheung CY, et al. Determinants of ganglion cell–inner plexiform layer thickness measured by high-definition optical coherence tomography. *Invest Ophthalmol Vis Sci.* 2012;53:5853–5859.
  10. Gupta P, Sidhartha E, Tham YC, et al. Determinants of macular thickness using spectral domain optical coherence tomography in healthy eyes: the Singapore Chinese Eye study. *Invest Ophthalmol Vis Sci.* 2013;54:7968–7976.
  11. Myers CE, Klein BE, Meuer SM, et al. Retinal thickness measured by spectral-domain optical coherence tomography in eyes without retinal abnormalities: the Beaver Dam Eye Study. *Am J Ophthalmol.* 2015;159:445–456.
  12. Ko F, Foster PJ, Strouthidis NG, et al. Associations with retinal pigment epithelium thickness measures in a large cohort: results from the UK Biobank. *Ophthalmology.* 2017;124:105–117.
  13. Chua SYL, Dhillon B, Aslam T, et al. Associations with photoreceptor thickness measures in the UK Biobank. *Sci Rep.* 2019;9:19440.
  14. Khawaja AP, Chua S, Hysi PG, et al. Comparison of associations with different macular inner retinal thickness parameters in a large cohort: the UK Biobank. *Ophthalmology.* 2020;127:62–71.
  15. Schmidt-Erfurth U, Sadeghipour A, Gerendas BS, et al. Artificial intelligence in retina. *Prog Retin Eye Res.* 2018;67:1–29.
  16. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12:e1001779.
  17. Keane PA, Grossi CM, Foster PJ, et al. Optical coherence tomography in the UK biobank study—rapid automated analysis of retinal thickness for large population-based studies. *PLoS One.* 2016;11:e0164095.
  18. Patel PJ, Foster PJ, Grossi CM, et al. Spectral-domain optical coherence tomography imaging in 67 321 adults: associations with macular thickness in the UK Biobank study. *Ophthalmology.* 2016;123:829–840.
  19. Stein DM, Ishikawa H, Hariprasad R, et al. A new quality assessment parameter for optical coherence tomography. *Br J Ophthalmol.* 2006;90:186–190.
  20. Li K, Wu X, Chen DZ, Sonka M. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Trans Pattern Anal Mach Intell.* 2006;28:119–134.
  21. Garvin MK, Abramoff MD, Wu X, et al. Automated 3-D intraretinal layer segmentation of macular spectral-domain optical coherence tomography images. *IEEE Trans Med Imaging.* 2009;28:1436–1447.
  22. Abramoff MD, Garvin MK, Sonka M. Retinal imaging and image analysis. *IEEE Rev Biomed Eng.* 2010;3:169–208.
  23. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *Adv Neural Inf Process Syst.* 2014;27.
  24. Choi Y, Choi M, Kim M, Ha JW, Kim S, Choo J. Stargan: unified generative adversarial networks for multi-domain image-to-image translation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2018:8789–8797.
  25. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 2016:770–778.
  26. Shah A, Zhou L, Abramoff MD, Wu X. Multiple surface segmentation using convolution neural nets: application to retinal layer segmentation in OCT images. *Biomed Opt Express.* 2018;9:4509–4526.
  27. Costa P, Galdran A, Meyer MI, et al. End-to-end adversarial retinal image synthesis. *IEEE Trans Med Imaging.* 2018;37:781–791.
  28. Halupka KJ, Antony BJ, Lee MH, et al. Retinal optical coherence tomography image enhancement via deep learning. *Biomed Opt Express.* 2018;9:6205–6221.
  29. Zhao H, Li H, Maurer-Stroh S, Cheng L. Synthesizing retinal and neuronal images with generative adversarial nets. *Med Image Anal.* 2018;49:14–26.
  30. Huang Y, Lu Z, Shao Z, et al. Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network. *Opt Express.* 2019;27:12289–12307.
  31. Narayanaswamy A, Venugopalan S, Webster DR, et al. Scientific discovery by generating counterfactuals using image translation. In: *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention.* Springer, Cham; 2020:273–283.
  32. Pearl J. *Causality.* Cambridge Univ Press, Cambridge; 2009.
  33. Pawlowski N, Castro D, Glocker B. Deep structural causal models for tractable counterfactual inference. *Adv Neural Inf Process Syst.* 2020;33.
  34. Hong S, Marinescu R, Dalca AV, et al. 3d-stylegan: a style-based generative adversarial network for generative modeling of three-dimensional medical images. In: *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections.* Springer, Cham; 2021:24–34.