Analyzing and Predicting Visual Acuity Outcomes of Anti-VEGF Therapy by a Longitudinal Mixed Effects Model of Imaging and Clinical Data

Wolf-Dieter Vogl,[1,2] Sebastian M. Waldstein,[2] Bianca S. Gerendas,[2] Thomas Schlegl,[1,2] Georg Langs,[1] and Ursula Schmidt-Erfurth[2]

[1]Computational Imaging Research Lab, Department of Biomedical Imaging and Image-guided Therapy, Medical University Vienna, Austria
[2]Christian Doppler Laboratory for Ophthalmic Image Analysis, Department of Ophthalmology and Optometry, Medical University Vienna, Austria

Correspondence: Ursula Schmidt-Erfurth, Department of Ophthalmology, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria; ursula.schmidt-erfurth@meduniwien.ac.at.

GL and US-E contributed equally to the work presented here and should therefore be regarded as equivalent authors.

Submitted: March 16, 2017
Accepted: June 12, 2017

Citation: Vogl W-D, Waldstein SM, Gerendas BS, Schlegl T, Langs G, Schmidt-Erfurth U. Analyzing and predicting visual acuity outcomes of anti-VEGF therapy by a longitudinal mixed effects model of imaging and clinical data. *Invest Ophthalmol Vis Sci*. 2017;58:4173–4181. DOI:10.1167/iovs.17-21878

Purpose. We develop a longitudinal statistical model describing best-corrected visual acuity (BCVA) changes in anti-VEGF therapy in relation to imaging data, and predict the future BCVA outcome for individual patients by combining population-wide trends and initial subject-specific time points.

Methods. Automatic segmentation algorithms were used to measure intraretinal (IRF) and subretinal (SRF) fluid volume on monthly spectral-domain optical coherence tomography scans of eyes with central retinal vein occlusion (CRVO) receiving standardized anti-VEGF treatment. The trajectory of BCVA over time was modeled as a multivariable repeated-measure mixed-effects regression model including fluid volumes as covariates. Subject-specific BCVA trajectories and final treatment outcomes were predicted using a population-wide model and individual observations from early follow-up.

Results. A total of 193 eyes (one per patient, 12-month follow-up, 2420 visits) were analyzed. The population-wide mixed model revealed that the impact of fluid on BCVA is highest for IRF in the central millimeter around the fovea, with −31.17 letters/mm^3 (95% confidence interval [CI], −39.70 to −23.32), followed by SRF in the central millimeter, with −17.50 letters/mm^3 (−31.17 to −4.60) and by IRF in the parafovea, with −2.87 letters/mm^3 (−4.71 to −0.44). The influence of SRF in the parafoveal area was −1.24 letters/mm^3 (−3.37–1.05). The conditional R^2 of the model, including subject-specific deviations, was 0.887. The marginal R^2 considering the population-wide trend and fluid changes was 0.109. BCVA at 1 year could be predicted for an individual patient after three visits with a mean absolute error of six letters and a predicted R^2 of 0.658 using imaging information.

Conclusions. The mixed-effects model revealed that retinal fluid volumes and population-wide trend only explains a small proportion of the variation in BCVA. Individual BCVA outcomes after 1 year could be predicted from initial BCVA and fluid measurements combined with the population-wide model. Accounting for fluid in the predictive model increased prediction accuracy.

Keywords: statistical machine learning, anti-VEGF therapy, image analysis, optical coherence tomography, intraretinal fluid, subretinal fluid

Intravitreal anti-VEGF therapy is the standard of care for a range of exudative macular diseases, including neovascular age-related macular degeneration and macular edema secondary to diabetic retinopathy and retinal vein occlusions.[1] In clinical practice, the functional response to anti-VEGF treatment is markedly heterogeneous and ranges from substantial vision gains with subsequent stability, to loss of vision despite aggressive therapy.[2] The individual functional potential to respond to VEGF inhibition is determined by several factors, including baseline visual acuity (VA) and variables obtained by high-resolution imaging such as optical coherence tomography (OCT).[3] While some changes observed within the retina are amenable to treatment (such as exudative fluid), others may become permanent despite therapeutic intervention (e.g., photoreceptor damage).[3] However, the extent of each factor's contribution is unclear and the interplay between imaging biomarkers, visual function, and treatment response remains poorly understood. Conventional statistical methods reporting outcomes (e.g., mean change in best-corrected visual acuity [BCVA]) cannot describe the longitudinal trajectories of visual function or the relation between vision and retinal morphology.

Longitudinal analysis enables individual development trajectories as well as the population-wide normative development to be captured and modeled.[4] In this study, we integrated longitudinal clinical and imaging data to model the influence of pathologic changes in the retina on the clinical outcome variable and to predict individual future therapeutic outcomes based on the current state. We demonstrated our method on

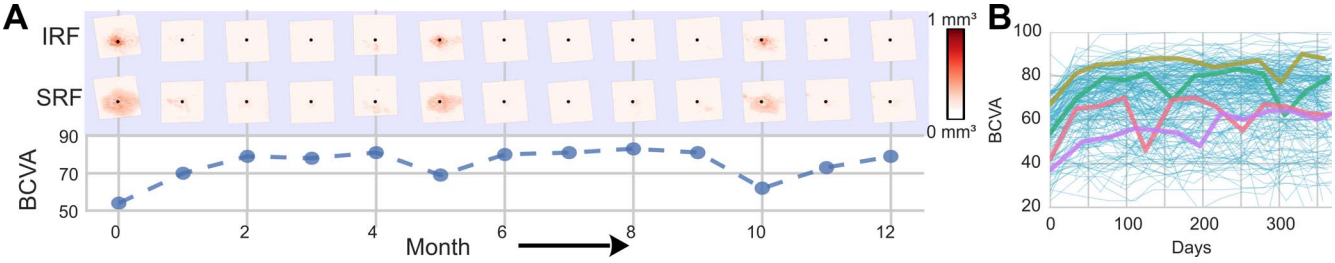

Figure 1. (**A**) Visits of an example study patient over 1 year. The *top rows* contain en face maps of automatically segmented intraretinal and subretinal fluid volumes. The *bottom row* shows the corresponding BCVA score measured in letters. Note the drop in BCVA score when fluid recurs. (**B**) BCVA trajectories of the whole study population, where each *blue line* represents one patient. Trajectories for four exemplary patients are highlighted to illustrate the variation in the data caused by different disease states at first visit, varying speed of BCVA recovery and temporary drops in BCVA caused by recurring fluid.

automatically segmented OCT characteristics and BCVA letter score as a clinical variable in a large population of patients with macular edema secondary to central retinal vein occlusion (CRVO) receiving anti-VEGF therapy (Fig. 1). We assumed that the individual VA and its development during therapy are affected by two major factors: reversible damage from which the retina recovers over time (where fluid in the retina is the major contributor) and irreversible damage. Hence, we modeled the population-wide temporal BCVA trajectory as a function of time and retinal fluid. By introducing random effects into the model, we accounted for the subject-specific deviation from the population-wide BCVA trajectory due to variation in reversible and irreversible damage as well as speed of recovery. We demonstrated that the combination of imaging and clinical data allows accurate modeling as well as prediction of treatment outcomes.

Patients and Methods

Patient Population and Study Procedures

This was post hoc analysis of a comprehensive prospective clinical trial database. Anonymized spectral-domain (SD)-OCT images and clinical data of patients enrolled in the CRYSTAL study (clinicaltrials.gov identifier: NCT01535261), available at the Vienna Reading Center (Vienna, Austria), were included. The trial's main outcome measures, inclusion and exclusion criteria, and its imaging and BCVA assessment procedures of the CRYSTAL trial have been reported.[5] Therefore, only procedures immediately relevant to the current analysis are described herein. The study was conducted in compliance with the Declaration of Helsinki and all participants provided written informed consent before inclusion. Approval was obtained from the ethics committee at the Medical University of Vienna and each study site participating in the CRYSTAL trial.

Inclusion/Exclusion Criteria, Treatment and Imaging

Patients, one eye from each, with a complete monthly follow-up, available from baseline to month 12, for whom automatic segmentations of retinal fluid were available ($n = 193$, 13 visits) were included in our retrospective analysis of the prospective study. All patients underwent standardized SD-OCT imaging by Cirrus HD-OCT ($n = 58$; Carl Zeiss Meditec, Dublin, CA, USA) and Spectralis OCT ($n = 135$; Heidelberg Engineering, Dossenheim, Germany) at monthly visits by certified operators. A 6×6 mm macular cube scan pattern with a resolution of 200×200 for Cirrus and 512×49 for Spectralis was used. A total of 89 scans that failed preprocessing steps described below due to poor image quality were removed from the analysis (see Fig. 2 for examples). Overall, 2420 scans were analyzed.

BCVA was measured at each monthly visit based on early treatment diabetic retinopathy study (ETDRS) charts by certified examiners. Patients were treated with 0.5-mg monthly ranibizumab injections (minimum of 3 injections) until stable BCVA was maintained for 3 consecutive months. Thereafter,

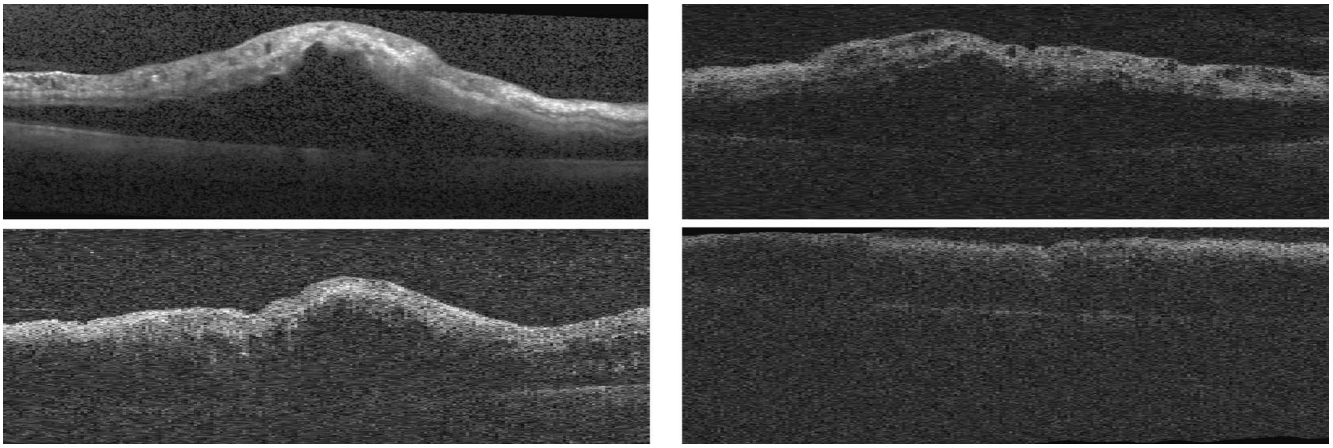

Figure 2. Four examples of SD-OCT scans that were excluded from analysis due to failures in preprocessing caused by poor image quality. Note the severe swelling of the neurosensory retina that precludes visualization of the retinal pigment epithelium layer. *Top left* scan was acquired from Spectralis OCT, and the other scans from Cirrus HD-OCT.

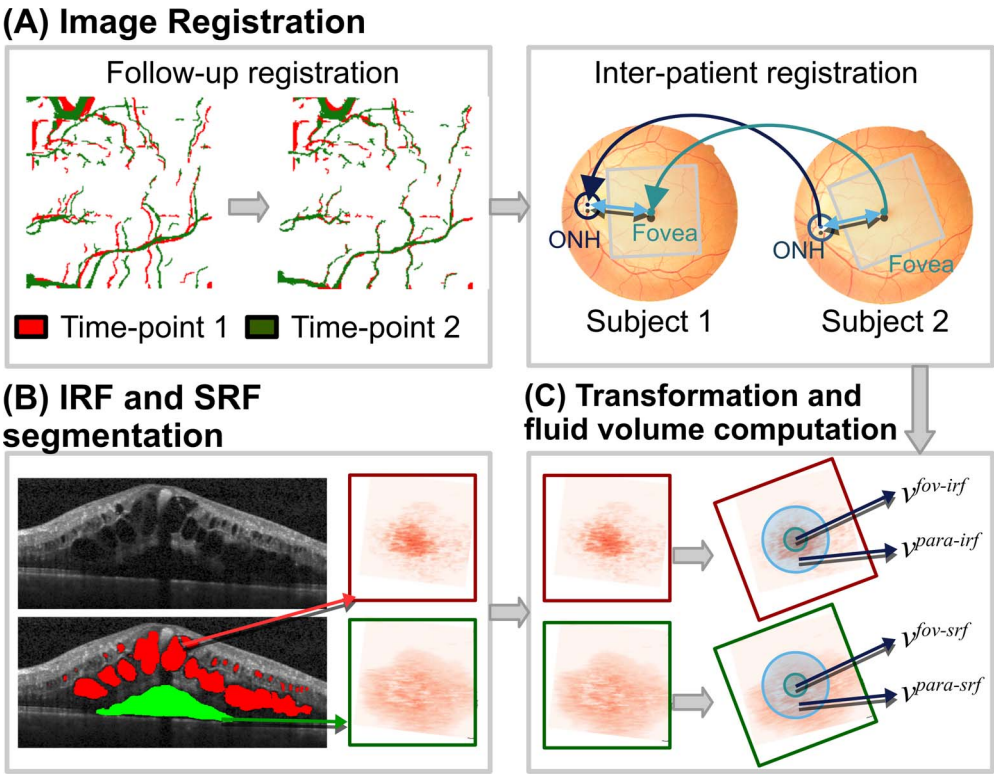

FIGURE 3. Preprocessing steps to obtain fluid volume in foveal and parafoveal regions. (**A**) Follow-up scans from a patient are aligned using the vessel structure. Scans between patients are registered by aligning the foveal center and ONH center. (**B**) Automatic segmentation of IRF and SRF fluid, from which en face projections of the fluid compartments are generated. (**C**) The fluid maps are transformed using the alignment of interpatient registration, followed by computation of the total fluid volume in the central mm ($v^{fov-irf}$, $v^{fov-srf}$) and in the parafoveal region in-between the 1- and 3-mm radius ($v^{para-irf}$, $v^{para-srf}$).

0.5 mg ranibizumab was injected as needed if monthly monitoring indicated a loss of VA resulting from disease activity.

Image Preprocessing

We compensated for variations in anatomy, scanning positions, and scanning resolutions by transforming the SD-OCT scans into a fovea-centered joint coordinate system as described.[6] First, follow-up patient scans were aligned rigidly to match the segmented vessel structures of the two images. Interpatient affine registration was obtained by aligning the center of the retina (fovea) and optic nerve head (ONH) center within patients (Fig. 3A).

Intraretinal fluid (IRF) and subretinal fluid (SRF) were automatically segmented in the SD-OCT images by classifying voxels as being background, IRF, or SRF using a convolutional neural network as reported[7] (Fig. 3B). The segmentations were transformed in the joint coordinate system and the total volume of IRF and SRF was computed within the central 1 mm region around the fovea and parafoveal region in the 1- to 3-mm radius ring (Fig. 3C). In the following, we denote the amount of foveal fluid volume measured as $v^{fov-irf}$ and $v^{fov-srf}$ and the volume in the 1- to 3-mm parafoveal region as $v^{para-irf}$ and $v^{para-srf}$, respectively.

Population-Wide BCVA Development Model

We modeled the development of BCVA over time as a repeated measure mixed-effects regression model (MRM)[8] with a quadratic growth term. BCVA was the outcome variable, time and fluid volume covariates were fixed effects, and individually varying intercept and slope were modeled as random effects. The BCVA value of a single visit j for a patient i at time t then is:

$$bcva_{ij} = \beta_0 + \beta_1 \times t_{ij} + \beta_2 \times t_{ij}^2 + \beta_3 \times v_{ij}^{fov-irf} + \beta_4 \times v_{ij}^{fov-srf} + \beta_5 \times v_{ij}^{para-irf} + \beta_6 \times v_{ij}^{para-srf} + b_{0i} + b_{1i} \times t_{ij} + \varepsilon_{ij} \quad (1)$$

The coefficients β weights the fixed-effects covariates, β_0 to β_2 are the mean intercept and slope coefficients and β_3 to β_6 are the coefficients modeling the influence the fluid in the various sections has on the outcome variable. The random coefficients b_{0i} and b_{1i} are the subject-specific weights modeling the deviation from the general trend in terms of intercept and slope. ε accounts for the model error.

From the general formula we derived four nested models to quantify the benefit of having imaging data in the model and the necessity of the quadratic growth term. The full model, $\mathcal{M}^{quad,img}$, contains the fluid volume and the quadratic growth term as specified in Equation 1. In the nested model $\mathcal{M}^{img}$, we dropped the quadratic term resulting in a linear growth model. By dropping the fluid volume terms, we obtained the models $\mathcal{M}^{quad,noImg}$ with quadratic growth and $\mathcal{M}^{noImg}$ with linear growth.

By using the vector and matrix notation, we can formulate Equation 1 for a series of visits of N patients as:[8]

$$y_i = X_i\beta + Z_i b_i + \varepsilon_i, i = 1, \ldots, N \quad (2)$$

where y_i is a $n_i \times 1$ vector containing the BCVA values observed, and n_i is the number of visits for patient i. X_i is a $n_i \times P$ design matrix containing P fixed effects (intercept,

time, time^2, fluid volumes) as columns, and visits as rows. The coefficients of the fixed effects are subsumized in the $P \times 1$ vector β. The random effects matrix Z_i of size $n_i \times 2$ contains the intercept and time variables as columns, weighted by the two-element vector b_i. Finally, ε_i is the $n_i \times 1$ measurement error.

To do inference on the model, we assumed a Gaussian normal distribution of the model error $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and multivariate Gaussian distribution of the random effects $b_i \sim \mathcal{N}(0, \Sigma)$, with Σ being the 2×2 covariance matrix of the random effects, where b_i and ε_i are independent from each other. The parameter σ can be interpreted as variance in the data within a patient and Σ as covariance between patients. The outcome variable has then a marginal distribution of $y_i \sim (X_i\beta, V_i)$, with the marginal covariance matrix $V_i = \mathrm{cov}(Z_i b_i) + \mathrm{cov}(\varepsilon_i) = Z_i \Sigma Z_i + \sigma^2 I$. The expected value for y_i is the so-called marginal mean $\mathrm{E}(y_i) = X_i\beta$, containing the fixed effects only. Therefore, the coefficients β can be interpreted as the population-wide mean response, similar to coefficient interpretation in linear regression.

Estimates of the fixed-effects, $\hat{\beta}$, and (co-)variance structures, $\hat{\sigma}$ and $\hat{\Sigma}$, are obtained either by maximum likelihood (ML), or restricted maximum likelihood (REML),[9] where ML is known to produce biased estimates. On the other hand, doing likelihood-based model comparisons does not provide meaningful results for REML models with different fixed effects structures.[10] Thus, in our work we used the ML estimates for likelihood-based model selection and comparison only, and the REML estimates otherwise.

Prediction of Individual BCVA Outcomes

Whereas the model allows to estimate the population-wide mean effects, such as effect of fluid volume on BCVA in terms of fixed effects, β, we used it additionally to predict individual future BCVA development and outcome by estimating the subject-specific random effects, b_i, from initial visits (e.g., the first 3 months) in combination with population estimates $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\Sigma}$. The subject-specific random effects, b_i, were estimated using the empirical best linear unbiased predictor (EBLUP):[11]

$$\hat{b}_i(\hat{\Sigma}, \hat{\sigma}, \hat{\beta}) = \hat{\Sigma} Z_i^T \hat{V}_i^{-1} (y_i - X_i \hat{\beta}) \quad (3)$$

where y_i and X_i contained the individual BCVA and fluid values from the already observed time points, and the fixed-effects and covariance coefficients from the marginal model, $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\Sigma}$, were learned beforehand from a training dataset. By using the estimated $\hat{b}_i$ in Equation 1, we can predict the BCVA for any given time-point t and, thus, construct continuous trajectories of BCVA development under treatment. In particular, it allows us to extrapolate BCVA trajectories based on past observations into predictions of the future. Furthermore, additionally observed time points, that is, from a follow-up visit, can be incorporated by recomputing Equations 3 and 1, resulting in a refined prediction of the random effects and the trajectory.

To predict BCVA for a future time-point, we must estimate the approximate IRF and SRF fluid amount at that time-point for the models $\mathcal{M}^{img}$ and $\mathcal{M}^{quad,img}$, as $\upsilon^{fov-irf}$, $\upsilon^{fov-srf}$, $\upsilon^{para-irf}$ and $\upsilon^{para-srf}$ are coefficients these models. Thus, we used a second growth MRM to predict the future subject-specific fluid development, $\hat{\upsilon}$, analogous to the previously described BCVA prediction. However, the fluid development generally is not linear, as recurrence of an edema causes peaks in the trajectory and the range of fluid volume is large. Therefore, we applied a median filter on the volumes measured for a patient to remove the peaks, where the median volume, $\bar{\upsilon}$, for a patient i and time-point j is: $\bar{\upsilon}_{ij} = \mathrm{median}(\upsilon_{ij-1}, \upsilon_{ij}, \upsilon_{ij+1})$.

The MRM for fluid development is modeled on a logarithmic scale:

$$\log(\bar{\upsilon}_{ij}) = \beta_7 + \beta_8 \times t_{ij} + b_{2i} + b_{3i} \times t_{ij} + \epsilon_{ij} \quad (4)$$

Similar to BCVA, we estimated the fluid trajectories for a patient by estimating the fixed-effects coefficients (intercept and slope) and covariances of Equation 4 from the population first, and then the subject-specific random coefficients from the given initial time points using Equation 2. With the estimated fixed and random-effects coefficients, we predicted fluid volume for any time-point t using Equation 4, and finally the BCVA using the predicted volumes in Equation 1.

Subject-specific prediction intervals, which are the intervals where a future BCVA value will fall with a certain probability, can be approximated using Monte Carlo simulation by sampling from the distributions of the estimated fixed-effects coefficients, random effects and residual error:[12,13]

$$\beta \sim \mathcal{N}\left(\beta, \left[\sum_{i=1}^{N} X_i^T V_i^{-1} X_i\right]^{-1}\right),$$
$$b_i \sim \mathcal{N}\left(b_i, \Sigma - \Sigma Z_i^T V_i^{-1} Z_i \Sigma\right), \varepsilon \sim \mathcal{N}(0, \sigma) \quad (5)$$

New trajectories are created from the coefficients sampled by using them in Equation 1. The α prediction interval for a given time-point t_{ij} is then the interval between the $\alpha/2$ and $1 - \alpha/2$ percentiles of the trajectory values at the time-point. We computed the interval using 1000 samples.

Statistical Evaluation

The evaluation of the proposed models is divided into two parts, with the first part covering the general model fit including the population-wide mean response and the second part covering the predictive power of the model in terms of predicting future BCVA development for new patients.

All four models were fitted on the whole dataset using the package lme4[9] from the statistics software R. We used the EM estimate to compare the models. $\mathcal{M}^{quad,img}$ has been compared to the other nested models $\mathcal{M}^{img}$, $\mathcal{M}^{quad,noImg}$ and $\mathcal{M}^{noImg}$ using a likelihood ratio test on variance components with Satterthwaite's estimation of degrees of freedom[14] and the Akaike information criterion (AIC).[15] As a goodness-of-fit measure, we used the coefficient of determination (R^2) for MRMs,[16,17] which is divided into the marginal R^2 that describes the proportion of variance explained by the fixed factors only and the conditional R^2 that considers fixed and random effects, thus measuring the variation explained by all effects. The coefficients were determined using REML for the final estimates. T-statistics and P values were calculated based on Satterthwaite's approximations and 95% confidence intervals (CIs) were obtained by bootstrap sampling with 500 simulations.

We evaluated the prediction performance of the model in terms of predictability of the VA after 1 year of treatment using a 5-fold cross-validation setup on a patient level. As performance measures, we computed the predicted R^2, which measures how well unseen samples are likely to be predicted by the model, and the mean absolute error (MAE) from the observed and predicted BCVA values. Because patients may suffer from recurring fluid at month 12 causing a drop in the VA, we used the median BCVA from months 10 to 12 as the prediction target to obtain a more reliable treatment outcome variable that is less affected by temporary events. Trajectories for each patient in the test-sets were predicted as described in Section 2.3 by first estimating the model parameters $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\Sigma}$ from the whole time series of training-set patients, followed by the prediction of random intercept and slope coefficients $\hat{b}_i$ from data of the given patient time points using Equation 3.

Table 1. Comparison of Models

Model	AIC	*P* Value Likelihood-Ratio Test	R^2 Marginal/ Conditional
$\mathcal{M}^{quad,img}$	**15,968**		**0.108/0.887**
$\mathcal{M}^{img}$	16,060	<0.0001	0.107/0.881
$\mathcal{M}^{quad,noImg}$	16,815	<0.0001	0.036/0.846
$\mathcal{M}^{noImg}$	17,029	<0.0001	0.022/0.827

In the likelihood-ratio test, all models were compared with the full model $\mathcal{M}^{quad,img}$. Best values are represented by bold numbers.

BCVA at a specific future time-point then was predicted using Equation 1. We evaluated how well the model adapts to new data of increasing number of visits by successively increasing the number of time points available for estimating the random coefficients. The MAE of all four models was compared for their significant difference using a one-sided Diebold-Mariano (DM) test[18,19] with the alternative hypothesis that the prediction results of the nested model is less accurate than the full model in terms of MAE.

Results

Model Comparisons and Coefficient Estimations

Comparison results of $\mathcal{M}^{quad,img}$, $\mathcal{M}^{img}$, $\mathcal{M}^{quad,noImg}$ and $\mathcal{M}^{noImg}$ are provided in Table 1, where best values are represented by bold numbers. Reduction in the AIC values, significance in the likelihood-ratio test, and an increase in the marginal and conditional R^2 indicate that adding fluid volume improves the general model fit. The conditional R^2 of 0.887 indicates a general good model fit. Most of the variance is covered by the subject-specific random intercept and slope, while a marginal R^2 of 0.109 shows that only a rather small proportion of the variance in the data is explained by the fixed effects (intercept, slopes, and fluid volumes).

Table 2. Estimated Coefficients of the Full Model $\mathcal{M}^{quad,img}$

Parameter	Estimate	95% CIs	*P* Value
Intercept	65.81	63.86 to 67.65	<0.0001
Time, per day	0.049	0.041 to 0.058	<0.0001
Time^2, per day	-9.42×10^{-5}	-1.12×10^{-4} to -7.56×10^{-5}	<0.0001
$v^{fov-irf}$, per 1 mm^3	−31.17	−39.70 to −23.32	<0.0001
$v^{fov-srf}$, per 1 mm^3	−17.50	−31.17 to −4.60	0.008
$v^{para-irf}$, per 1 mm^3	−2.87	−4.71 to −0.44	0.006
$v^{para-srf}$, per 1 mm^3	−1.24	−3.37 to 1.05	0.241

As seen in Figure 4, $\mathcal{M}^{noImg}$ was able to model the general trend of BCVA development under treatment but failed to model the temporary deviations in BCVA caused by recurring fluid in the retina. Furthermore, the large amount of fluid present in almost all patients at baseline associated with a lower BCVA at that time-point caused a bias by underestimating BCVA in early months, as observed in the individual estimates (Fig. 4).

As summarized in Table 2, all fixed-effects coefficients were significant in the full model, except $v^{para-srf}$. The mean estimates of fluid coefficients indicate that an increase in IRF causes a higher drop in BCVA than increasing SRF and that the influence of foveal fluid on BCVA is in an order of magnitude higher than the parafoveal fluid influencing BCVA. For instance, the median amount of $v^{fov-irf}$ at baseline is 0.170 mm^3, which causes a drop of 5.30 letters, whereas the median $v^{para-irf}$ of 0.480 mm^3 causes a drop of 1.38 letters. Similarly, the baseline median $v^{fov-srf}$ of 0.056 mm^3 results in a drop of 0.98 letters and $v^{para-srf}$ of 0.135 mm^3 reduces BCVA by 0.17 letters.

Prediction of Individual BCVA Outcomes

We evaluated the prediction performance as described in the Methods section for all four models. We successively increased the number of visits available for prediction for

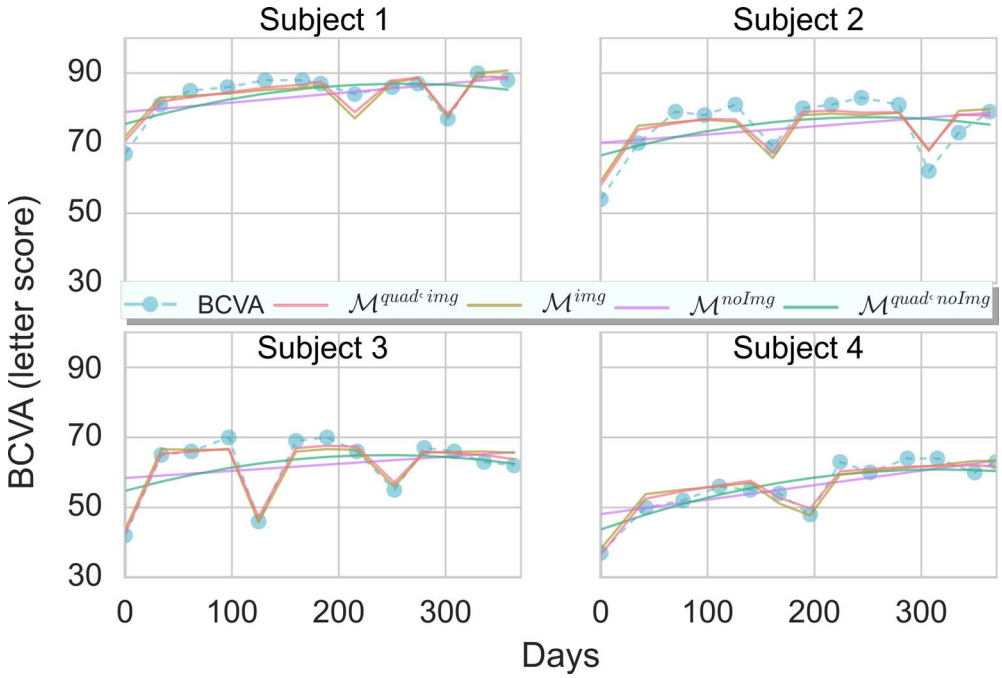

Figure 4. BCVA trajectories (*blue*) measured in four patients and the corresponding subject-specific model fits. *Spikes* in the BCVA trajectory are caused by recurring fluid. Note that the imaging-based models $\mathcal{M}^{quad,img}$ and $\mathcal{M}^{img}$ are able to model these spikes, resulting in a general better model fit.

Table 3. End-Point BCVA Prediction Performance in Terms of MAE of Letters and Predicted R^2, Using an Increasing Number of Time Points (Months) in a 5-Fold Cross-Validation Setup for the Four Models

	$\mathcal{M}^{quad,img}$			$\mathcal{M}^{img}$				$\mathcal{M}^{quad,noImg}$				$\mathcal{M}^{noImg}$			
Months	**MAE**	**STD**	R^2	**MAE**	**STD**	R^2	**DM *P* Value**	**MAE**	**STD**	R^2	**DM *P* Value**	**MAE**	**STD**	R^2	**DM *P* Value**
1	**9.358**	8.362	0.290	9.491	8.479	0.269	**0.020**	9.851	7.682	**0.296**	0.084	10.292	7.797	0.248	0.010
2	**7.218**	7.446	0.515	7.222	7.563	0.507	0.476	7.335	7.310	**0.516**	0.299	7.284	7.417	0.513	0.391
3	**6.008**	6.310	**0.658**	6.124	6.438	0.644	**0.045**	6.223	6.383	0.642	0.107	6.263	6.463	0.635	0.068
4	**5.895**	6.021	**0.680**	6.046	6.174	0.663	0.052	6.128	6.349	0.649	0.072	6.321	6.419	0.634	**0.007**
5	**5.840**	5.452	**0.712**	6.131	5.718	0.683	**0.012**	5.885	5.561	0.704	0.408	6.049	5.791	0.684	0.139
6	5.472	5.046	**0.750**	5.845	5.308	0.719	**0.010**	**5.454**	5.175	0.745	0.537	5.707	5.315	0.726	0.130
7	**4.982**	4.985	**0.776**	5.450	5.151	0.746	**0.004**	5.223	5.027	0.763	0.094	5.492	5.207	0.742	0.017
8	**4.815**	4.653	**0.798**	5.298	5.014	0.760	**0.004**	4.948	4.646	0.792	0.230	5.341	4.820	0.767	0.011
9	**4.145**	4.521	**0.830**	4.601	4.786	0.801	**0.006**	4.597	4.688	0.806	**0.005**	4.826	4.735	0.794	**0.001**
10	**3.735**	4.233	**0.856**	4.141	4.522	0.830	**0.006**	4.222	4.449	0.830	**0.002**	4.248	4.491	0.828	**0.002**
11	**3.094**	3.426	**0.904**	3.472	3.810	0.880	**0.003**	3.533	3.469	0.889	**0.004**	3.508	3.616	0.886	**0.002**
12	**2.580**	2.794	**0.935**	2.933	3.055	0.919	**0.001**	3.039	2.892	0.921	**0.001**	2.911	2.915	0.923	**0.003**
13	**2.197**	2.323	**0.954**	2.480	2.476	0.945	**<0.001**	2.613	2.473	0.942	**0.001**	2.459	2.420	0.946	**0.008**

P values are from a one-sided Diebold-Mariano (DM) test comparing the given model with the full model, $\mathcal{M}^{quad,img}$. STD, standard deviation. Best values at each month are represented by bold numbers.

each patient in the test fold of the cross-validation setup to evaluate how the prediction accuracy adapts to the refined prediction models. All the results are listed in Table 3, and a summary of MAE and predicted R^2 is given in Figure 5. Furthermore, we exemplarily visualize the individual BCVA trajectories and prediction intervals for four patients in Figure 6, where the four models are compared in Figures 6A and 6B. How the predicted trajectory adapts to an increasing number of time points available for prediction is shown in Figures 6C and 6D.

Discussion

In this study, we proposed a framework to model VA outcomes for patients receiving treatment as an MRM incorporating longitudinal clinical data of measured VA and structural SD-OCT data of automatically segmented intraretinal and subretinal fluid. We demonstrated that the marginal part of the MRM is able to model BCVA on a normative population level and the subject-specific conditional part is able to account for variations in retinal anatomy and different individual response patterns. Furthermore, we demonstrate the model can predict the future VA of individual patients receiving treatment by assuming that the individual trajectories can be predicted from early time points and the population-wide model. By accounting for the fluid changes due to treatment and recurrence we can predict the BCVA outcome more precisely. This enables the separation of temporary fluid induced fluctuations in BCVA, and the development of BCVA more relevant for long-time outcome.

Analyzing longitudinal studies is a challenging task. Follow-up observations are correlated and data often are unbalanced due to missing time points and nonuniform visit intervals. Furthermore, the intersubject variability in the disease state at the first visit and diverging responses to treatment must be considered. MRMs on longitudinal data[20] provide a rich statistical framework for approaching these issues. The linear relations between a univariate longitudinal measurement and time points of observations as well as population-wide and individual effects can be modeled by using the particular random slope and intercept model proposed by Laird and Ware.[8] Intersubject variability is accounted for by allowing random deviations from the population-wide mean trajectory on an individual patient level by incorporating random intercepts and slopes. The superior statistical power of MRM compared to alternative methods using longitudinal imaging data of neurodegenerative diseases has been demonstrated.[21] Furthermore, MRM models have been shown to be capable of

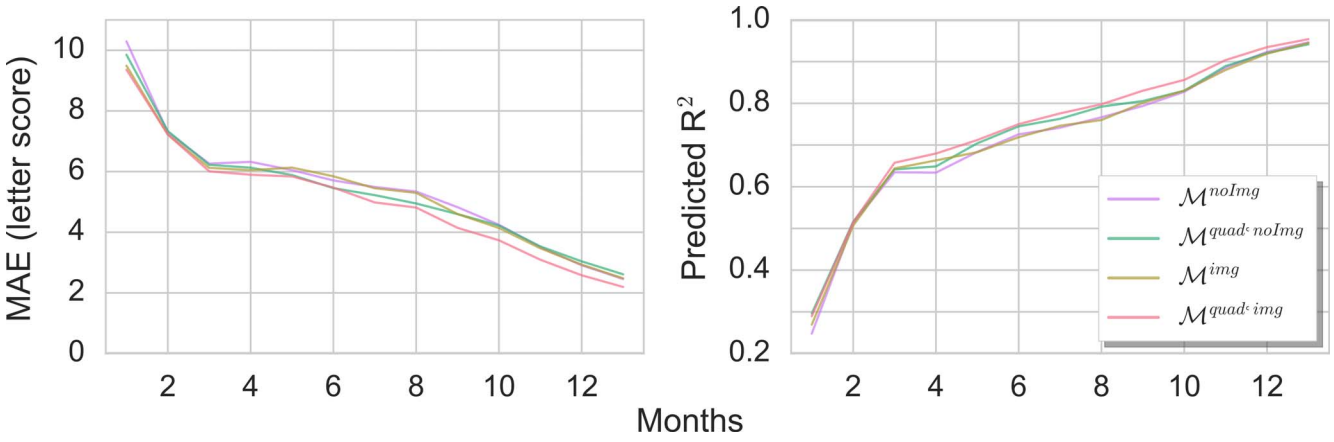

Figure 5. Prediction performance of end-point BCVA for the four models using an increasing number of visits (*Months*). *Left*: MAE in letters between observed and predicted BCVA. *Right*: Predicted R^2.

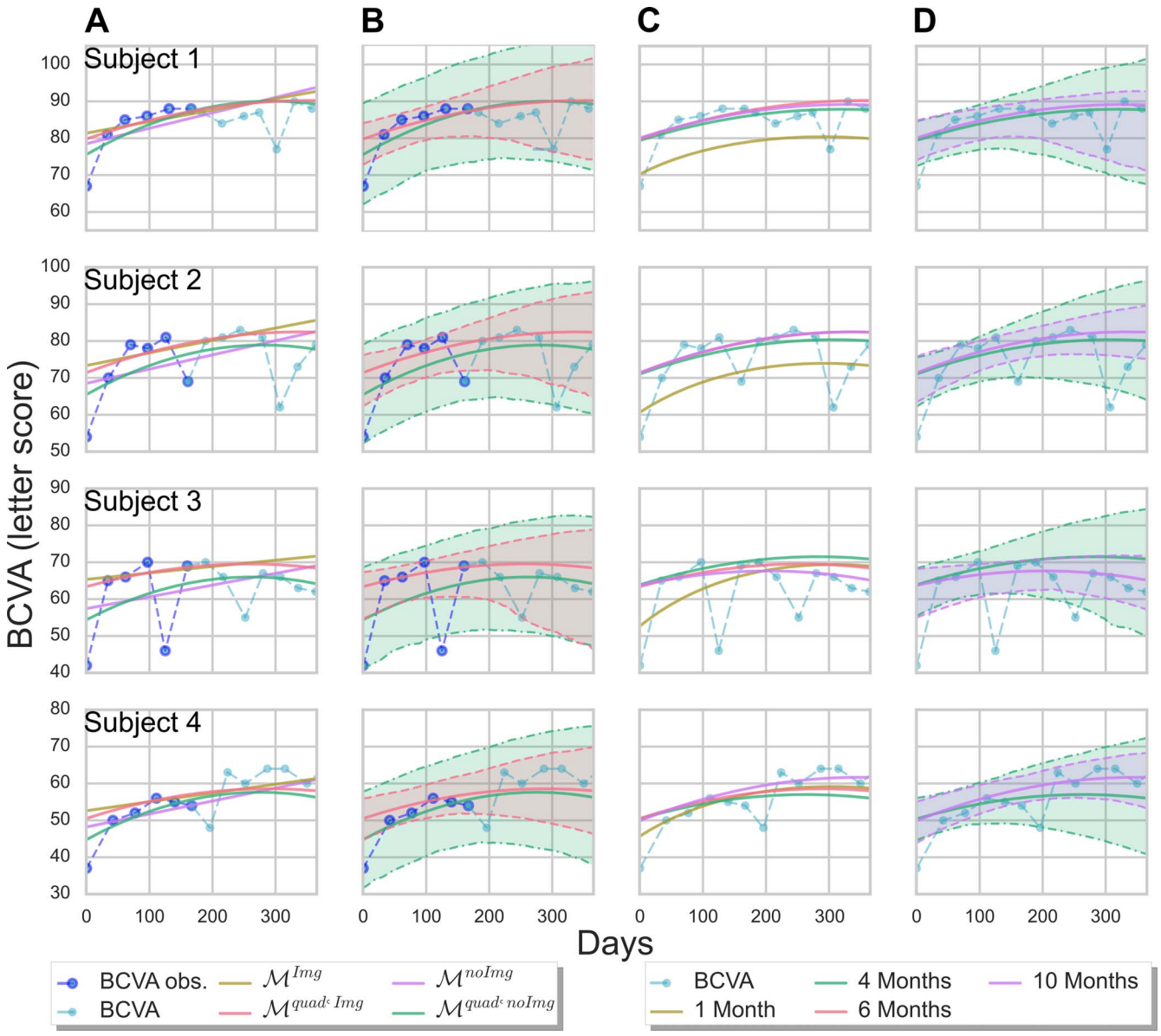

Figure 6. Predicted BCVA trajectories for four patients. **(A)** Comparison of trajectories from the four models with six time points available for prediction (BCVA obs.). **(B)** 95% prediction intervals for $\mathcal{M}^{quad,img}$ and $\mathcal{M}^{quad,noImg}$. **(C)** Illustration of how the BCVA trajectory is refined with an increasing number of time points available. Trajectory has been gained from the full model, $\mathcal{M}^{quad,img}$. **(D)** 95% prediction intervals having four and 10 time points for the full model.

predicting the future development of early infant brain maturation by estimating the growth trajectories from magnetic resonance diffusion tensor images.[13] In previous work,[6] we used SD-OCT data combined with sparse logistic regression to predict disease recurrence from early time-point observations. However, this method is limited to balanced data without missing time points.

In the current study using MRM, we found that intraretinal and subretinal fluid volume explains only a rather small fraction of the BCVA variance in the population (marginal $R^2 = 0.1$), and foveal fluid and IRF have more influence on VA than parafoveal fluid and SRF. Thus, effects other than fluid cause a large fraction of vision. We subsumed these effects in the random effects that models baseline irreversible damage as subject-specific intercept deviation and reversible damage as subject-specific slope deviation. Note that we model only the immediate effects of fluid on BCVA explicitly, and not the secondary effects caused by fluid, such as damage to photoreceptor cells[22,23] or conductive elements of the retina.[24]

Even though the impact of fluid on VA is limited, by incorporating it we are able to model rapid changes in VA caused by fluid resorption after anti-VEGF injections and recurring fluid. Without accounting for fluid, the models tend to underestimate BCVA at early time points and overestimate the slope or treatment response because the low baseline BCVA caused by a large amount of fluid present pulls down the trajectory of the model (Figs. 4, 6A). The prediction interval, not to be confused with confidence interval of the trajectory, is smaller for models accounting for fluid (Fig. 6B). This results from the lower variance of the residuals because the spiky changes in the trajectory are modeled by fluid coefficients. However, the interpretation of prediction intervals differs between nonfluid and fluid models. In the first case, it shows the interval where the future BCVA probably is lying, including

the large deviations caused by recurring fluid, whereas in the second it shows the interval of future BCVA without short-term interruptions.

Predicting final VA outcomes from baseline only showed lower accuracy with a mean error of ±9 letters, because the individual slope cannot be estimated from a single time-point and the population-wide trend must be used. Predicting the outcomes with 3 months available showed a mean error of ±6 letters. After the initial phase, the gain in prediction accuracy when incorporating additional visits was not as high as in the first 3 months, which reflects the fact that many patients reached a near-final BCVA response after the loading dose. However, the prediction interval became narrower with additional time points (Fig. 6D) and the model adapted to cases where the patient did not fully recover from a recurrence (e.g., Subject 3, Fig. 6C).

While we only considered fluid segmented in SD-OCT images as an influence on VA and treat other effects as random, this model could be extended easily to measure the effects of any other (pathologic) structure observed in SD-OCT images, such as damage to photoreceptor cells, integrity of the ellipsoid zone and external limiting membrane, disorganization of retinal inner layers (DRIL),[25] and other biomarkers that may be quantified in the future. The limited explanatory value of the fluid parameters on VA emphasizes the necessity of incorporating additional possible biomarkers. The proposed model has the capability to assess and quantify the influence of these biomarkers on VA similarly to the fluid model, by marginal/conditional R^2, and coefficient effect size and significance, and ultimately it may help to obtain additional insights in the disease, its progression, and outcome. Moreover, the specific effects of different pharmacologic agents or other treatments may be compared efficiently. The prediction model also could be applied to other diseases, such as diabetic maculopathy or age-related macular degeneration.

This study has several limitations. The model can predict long-term trends but not short-term events, such as recurring fluid. A data subset from the CRYSTAL study was used with complete follow-ups, which may be a source of bias. Another possible source of bias is that SD-OCT image quality tends to be low for patients with large amounts of fluid and/or poor vision. Segmentation of fluid is difficult or even impossible on such scans and they are more likely to be excluded from the analysis. However, we only had to remove single low-quality scans and not a whole patient time-series, as MRMs are capable to handle unbalanced data.

In conclusion, we demonstrated that MRM can be adapted to create a longitudinal disease model by combining clinical and imaging variables, determine the influence of fluid on visual function from the population-wide model, and determine the treatment outcome after 1 year by predicting individual trajectories from the pooled population-wide model and early time points.

Acknowledgments

Supported by the Austrian Federal Ministry of Economy, Family and Youth, the National Foundation for Research, Technology and Development, and by FWF (I 2714-B31) and OeNB (15356, 15929).

Disclosure: **W.-D. Vogl**, None; **S.M. Waldstein**, None; **B.S. Gerendas**, None; **T. Schlegl**, None; **G. Langs**, None; **U. Schmidt-Erfurth**, None

References

1. Campochiaro PA, Aiello LP, Rosenfeld PJ. Anti-vascular endothelial growth factor agents in the treatment of retinal disease: from bench to bedside. *Ophthalmology*. 2016;123:S78–S88.
2. Tufail A, Xing W, Johnston R, et al. The neovascular age-related macular degeneration database: multicenter study of 92 976 ranibizumab injections: report 1: visual acuity. *Ophthalmology*. 2014;121:1092–101.
3. Schmidt-Erfurth U, Waldstein SM. A paradigm shift in imaging biomarkers in neovascular age-related macular degeneration. *Prog Retin Eye Res*. 2016;50:1–24.
4. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.
5. Larsen M, Waldstein SM, Boscia F, et al. CRYSTAL study group: individualized ranibizumab regimen driven by stabilization criteria for central retinal vein occlusion: twelve-month results of the CRYSTAL study. *Ophthalmology*. 2016;123:1101–1111.
6. Vogl W-D, Waldstein SM, Gerendas BS, et al. Spatio-temporal signatures to predict retinal disease recurrence. *Inf Proc Med Imaging*. 2015;24:152–163.
7. Schlegl T, Waldstein SM, Vogl W-D, Schmidt-Erfurth U, Langs G. Predicting semantic descriptions from medical images with convolutional neural networks. *Inf Proc Med Imaging*. 2015:24:437–448.
8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38:963–974.
9. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Statistical Software*. 2015;67:1–48.
10. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer Science & Business Media; 2009.
11. Robinson GK. That BLUP is a good thing: the estimation of random effects. *Stat Sci*. 1991;6:15–32.
12. Vonesh EF, Chinchilli VM. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker, Inc.; 1997.
13. Sadeghi N, Fletcher PT, Prastawa M, Gilmore JH, Gerig G. Subject-specific predictions using nonlinear population modeling: application to early brain maturation from DTI. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014: 17th International Conference*. Springer International Publishing; 2014:33–40.
14. Satterthwaite FE. An approximate distribution of estimates of variance components. *Biometrics*. 1946;2:110–114.
15. Akaike H. A new look at the statistical model identification. *IEEE Trans Auto Control*. 1974;19:716–723.
16. Nakagawa S, Schielzeth H. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods Ecol Evol*. 2013;4:133–142.
17. Johnson PC. Extension of Nakagawa & Schielzeth's r2glmm to random slopes models. *Methods Ecol Evol*. 2014;5:944–946.
18. Diebold FX, Mariano RS. Comparing predictive accuracy. *J Bus Econ Stat*. 1995;13:253–263.
19. Harvey D, Leybourne S, Newbold P. Testing the equality of prediction mean squared errors. *Int J Forcastt*. 1997;13:281–291.
20. Pinheiro J, Bates D. *Mixed-Effects Models in S and S-PLUS*. New York: Springer Science & Business Media; 2006.
21. Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR, Alzheimer Disease Neuroimaging Initiative. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. *Neuroimage*. 2013;66:249–260.
22. Jaissle GB, Szurman P, Feltgen N, et al. Predictive factors for functional improvement after intravitreal bevacizumab therapy for macular edema due to branch retinal vein occlusion. *Graefe's Arch Clin Exp Ophthalmol*. 2011;249:183–192.

23. Kondo M, Kondo N, Ito Y, et al. Intravitreal injection of bevacizumab for macular edema secondary to branch retinal vein occlusion: results after 12 months and multiple regression analysis. *Retina*. 2009;29:1242–1248.
24. Pelosini L, Hull CC, Boyce JF, McHugh D, Stanford MR, Marshall J. Optical coherence tomography may be used to predict visual acuity in patients with macular edema. *Invest Ophthalmol Vis Sci*. 2011;52:2741–2748.
25. Grewal DS, O'Sullivan ML, Kron M, Jaffe GJ. Association of disorganization of retinal inner layers with visual acuity in eyes with uveitic cystoid macular edema. *Am J Ophthalmol*. 2017;177:116–125.