This is an Accepted Manuscript of an article published by Taylor & Francis in Current Eye Research on 30 Nov 2023, available at https://doi.org/10.1080/02713683.2023.2282938

Optimizing Refractive Outcomes of SMILE: Artificial Intelligence versus Conventional State-of-the-Art Nomograms

Nikolaus Luft, MD[1,2], PhD, Niklas Mohr, MD[1], Dr. Elmar Spiegel[3], Hannah Marchi[3,4], Jakob Siedlecki, MD[1], Lisa Harrant[1], Wolfgang J. Mayer, MD[1], Martin Dirisamer, MD[1,2], Siegfried G. Priglinger, MD[1,2]

[1]Department of Ophthalmology University Hospital, LMU Munich, Mathildenstraße 8, 80336 Munich, Germany
[2]SMILE Eyes Clinic, Weißenwolffstraße 13, 4020 Linz, Austria
[3]Core Facility Statistical Consulting, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Ingolstädter Landstraße 1, 85764 Neuherberg, Germany
[4]Faculty of Business Administration and Economics, Bielefeld University, Germany

3,692 words

No author reports any financial interest relevant to this study.

Corresponding author & reprint requests:
Nikolaus Luft, MD, PhD, FEBO
University Eye Hospital, Ludwig-Maximilians-University
Mathildenstraße 8, 80336 Munich, Germany
Tel.: +49 89 4400 53811; Fax.: +49 89 4400 55160
e-mail: nikolaus.luft@med.uni-muenchen.de

Abstract

Purpose: AI (artificial intelligence)-based methodologies have become established tools for researchers and physicians in the entire field of ophthalmology. However, the potential of AI to optimize the refractive outcome of keratorefractive surgery by means of machine learning (ML)-based nomograms has not been exhausted yet. In this study, we wanted to comprehensively compare state-of-the-art conventional nomograms for Small-Incision-Lenticule-Extraction (SMILE) with a novel ML-based nomogram regarding both their spherical and astigmatic predictability.

Methods: A total of 1,342 eyes were analyzed for creation of three different nomograms based on a linear model (LM), a generalized additive mixed model (GAMM) and an artificial-neuronal-network (ANN), respectively. A total of 16 patient- and treatment-related features were included. Each model was trained by 895 eyes and validated by the remaining 447 eyes. Predictability was assessed by the difference between attempted and achieved change in spherical equivalent (SE) and the difference between target induced astigmatism (TIA) and surgically induced astigmatism (SIA). The root mean squared error (RMSE) of each model was computed as a measure of overall model performance.

Results: The RMSE of LM, GAMM and ANN were 0.355, 0.348 and 0.367 for the prediction of SE and 0.279, 0.278 and 0.290 for the astigmatic correction, respectively. By applying the created models, the theoretical yield of eyes within ±0.50 D of SE from target refraction improved from 82% to 83% (LM), 84% (GAMM) and 83% (ANN), respectively. Astigmatic outcomes showed an improvement of eyes within ±0.50 D from TIA from 90% to 93% (LM), 93% (GAMM) and 92% (ANN), respectively. Subjective manifest refraction was the single most influential covariate in all models.

Conclusion: Machine learning endorsed the validity of state-of-the-art linear and non-linear SMILE nomograms. However, improving the accuracy of subjective manifest refraction seems warranted for optimizing ±0.50 D SE predictability beyond an apparent methodological 90% limit.

Keywords

Small incision lenticule extraction; SMILE; nomogram; artificial intelligence; machine learning

Introduction

In recent years, a surge in applications for artificial intelligence (AI) – particularly the subset of machine learning (ML) and its subset deep learning – has been observable in ophthalmology. AI-based methodologies have become established tools for researchers and physicians alike in the posterior segment, predominantly in diabetic retinopathy, macular and glaucomatous disease.[1] In contrast, its employment in the anterior eye segment seems to bear more potential than currently realized.[2]

In refractive surgery, ML has been successfully utilized to identify eyes with an increased risk for developing progressive post-laser in situ keratomileusis (LASIK) ectasia.[3-5] Furthermore, a multiclass machine learning model has recently been created that can select the optimal laser surgery option between photorefractive keratectomy (PRK), LASIK and small incision lenticule extraction (SMILE) for a particular patient on an expert level.[6] Moreover, ML has proven highly accurate in refining and optimizing intraocular lens power calculations in order to reduce "refractive surprises" after cataract or refractive lens exchange surgery.[7, 8] In the near future, ML may also be beneficial for intraocular lens (IOL) power calculation in post-keratorefractive surgery eyes, which still pose a particular challenge with variable refractive outcomes.[9]

Surprisingly, the potential of AI to optimize the refractive outcome of keratorefractive surgery by means of ML-based nomograms has not been exhausted yet. Nomograms represent mathematical approaches for adjusting the surgical refractive correction in reference to the patient's preoperative subjective refraction, and have been an integral component of refractive surgery since its very beginning.[10] In the first study of its kind, Yang et al.[11] used neural network computing as early as 1998 to determine PRK nomograms but failed in face of their small sample size of only 44 eyes. Taking the ML-based nomogram approach one step further, Cui et al. conducted a larger scaled, prospective study of 510 myopic spherocylindrical SMILE treatments using a ML-based nomogram.[12] Despite overall favorable predictability of the ML-based nomogram, shortcomings of the model became evident for high myopic and astigmatic treatments owing to the limited number of those eyes in the sample that was used for training of the model (training dataset). Moreover, nomograms were analyzed only for their prediction of spherical equivalent (SE) of manifest refraction but not for the astigmatic component of the surgical correction. In addition, no actual comparison with well-established conventional statistical methods for refractive surgery nomograms was conducted. Instead, an undisclosed "surgeon nomogram" was used in the control group based on the surgeons' personal experience.[10, 13]

Hence, the aim of the present study was to comprehensively compare both the spherical and astigmatic predictability of a novel ML-based nomogram with state-of-the-art conventional refractive surgery nomograms in a big data sample of 1,342 SMILE treatments.

Participants and Methods

Patient Selection and Matching

For the purpose of this cross-sectional study, our institution's database was screened for patients who underwent uneventful SMILE for treatment of myopia or myopic astigmatism. A further inclusion criterion was a minimum age of 18 years as well as a minimum of 3 months of post-surgical follow-up with complete records of all investigated parameters. The study protocol was approved by the ethics committee of the Ludwig-Maximilians-University (approval ID: 22-1001). Consent to use their data for analysis and publication was obtained from all subjects and all study-related procedures adhered to the tenets outlined in the Declaration of Helsinki.

SMILE surgery

All SMILE procedures were performed by one of two highly experienced corneal surgeons (M.D., S.G.P.) using the VisuMax 500-kHz femtosecond laser system (Carl Zeiss Meditec AG, Jena, Germany). The technical principles of the SMILE procedure have been outlined in detail previously.[14] An optical zone between 6.0 mm and 7.0 mm was created. The intended cap diameter was 7.8-7.9 mm and the intended cap thickness ranged between 120 µm and 140 µm. For manual extraction of the refractive lenticule, a 4.00 mm incision was created by the femtosecond laser centered at the 135° position in right eyes and at the 45° position in left eyes.[15]

Postoperatively, patients were prescribed dexamethasone 0.1% and tobramycin 0.3% eyedrops 6 times daily for 1 week. Thereafter, dexamethasone 0.1% eyedrops were tapered over the course of 1 month starting with a 4 times daily (QID) regimen. Additionally, patients were encouraged to use preservative-free lubricating eye drops as often as individually required.

Subjective Refraction & Visual Acuity Readings

Subjective manifest and cycloplegic refraction were measured using the Jackson cross-cylinder method. Monocular and binocular uncorrected (UDVA) and best-corrected distance visual acuity (CDVA) was determined using standard ETDRS charts at 4 meters.

Corneal Tomography

Preoperative and postoperative corneal tomography scans were obtained using a high-resolution rotating Scheimpflug camera system (Pentacam HR; Oculus Optikgeräte GmbH, Wetzlar, Germany). All measurements were obtained under standard scotopic ambience light conditions and subjects had to refrain from using eye drops one hour prior to scanning. Angle kappa was calculated as the offset between the corneal intercept with the Pentacam’s optical axis and the geometric pupil center.[16] Apex-vertex (A-V) distance was calculated as the offset between the corneal intercept with the Pentacam's optical axis and the location of the maximum corneal curvature (Kmax).[16]

Nomogram Generation & Statistical Analysis

For generation and testing of the ML-based nomogram, the clinical data of 1,342 eyes of 686 patients were used. A total of 895 eyes of 457 patients were used for model training (training dataset) and the remaining 447 eyes of 229 patients were held back as test the dataset for model validation (validation dataset). When splitting the dataset, it was ensured that paired eyes were included into the same dataset to avoid biased performance estimates for the different models, since similar responses to the laser treatment can be assumed for paired eyes.

The following patient and SMILE treatment-related parameters were included for nomogram creation and analysis: (1) patient age at the time of treatment (years), (2) sex, (3) laterality, (4) mesopic pupil size (mm), (5) pachymetry (µm), (6) Kmax (D), (7) angle kappa (mm), (8) A-V distance (mm), (9) attempted change in manifest refraction spherical equivalent (SE; Diopters; D), (10) target induced astigmatism (TIA; D), (11) axis of refractive astigmatism (°), (12) optical zone size (mm), (13) cap thickness (µm), (14) minimum lenticule thickness (µm), (15) astigmatism axis deviation (defined as the angle between the refractive and topographic axis of astigmatism) and (16) astigmatism power deviation (defined as the difference between the refractive and topographic power of astigmatism). The two separate, dependent outcome variables were (1) the difference between attempted and achieved change in spherical equivalent (SE) and (2) the difference between target induced astigmatism (TIA) and surgically-induced astigmatism (SIA) according to the Alpins power vector methodology.[13] To investigate the effect of the 16 different covariates on the two dependent outcome variables, we generated three different models, all of which were estimated using R (Version 4.2.1).[17] The significance level for all analyses were set to $p<0.05$.

Firstly, a linear model (LM) was created in which the covariates were selected based on a p-value smaller than 5%, which represents the most commonly and well-established approach for refractive surgery nomograms.[10, 13] All potential 16 covariates were included into the LM and insignificant covariates were removed. The estimation was based on the default “lm” function of R.

Secondly, a generalized additive mixed model (GAMM) was created.[18, 19] GAMM represents a more elaborate linear model, in which the shape of the continuous covariates is data driven (using smooth functions / penalized B-Splines) and in which the presence of multiple measurements per patient (paired-eye data) are appropriately taken into account. We relied on the default estimation of the maximal degrees of freedom and used restricted maximum likelihood (REML) for optimizing the parameters. To account for partially paired-eye data we used random intercepts. The covariates are selected based on 10-fold cross-validation on the training dataset while ascertaining to not split paired-eye data. The estimation was based on the gamm4 package in R.[20]

Thirdly, an artificial neural network (ANN), a deep learning method with three hidden layers was created. The parameters of the ANN (number of knots, activation function, penalty, optimizer and learning rate) were optimized during the training. The resulting model is described in the results section. As loss function we used mean squared error (“MSE”) and applied 25 epochs with a batch size of 32 and a 20% validation split. The estimation is based on the packages tensorflow and keras in R.[21,22] The optimization of the ANN parameters was performed using the tfruns package of R.[23] In the ANN, the continuous covariates are scaled in the interval [0,1] while the categorical variables are coded manually as multiple binary variables. The complete analysis code can be obtained from the following open source repository:

https://ascgitlab.helmholtz-muenchen.de/elmar.spiegel/optimizing_refractive_outcomes_of_smile.

After “training” of the three models based on the training dataset, their performance was evaluated on the validation data set. Using various covariates (see above) the models predict the residual refractive error after SMILE. The predicted refractive error after SMILE was then compared with the actually observed postoperative refractive error in the validation dataset. Based on these prediction errors, root mean squared error (RMSE) of each model were calculated as a measure of model performance. Moreover, kernel density estimator graphics and stacked bar graphs were created to compare the actually observed refractive outcome in the validation dataset with the (hypothetically achievable) outcome after correction by the respective models/nomograms.

Results

Surgical Outcomes

A total of 1,342 eyes of 686 patients [364 (53%) females] with a mean age of 33 ± 8 [95% confidence interval (95% CI) 21-52] years were included into this analysis. Mean preoperative SE was -4.74 ± 1.89 D with a 95% CI of -8.38 to -1.62 D. Mean target induced astigmatism (TIA) was 0.91 ± 0.79 with a 95% CI of 0 to 3.00 D. The standard graphs for reporting keratorefractive surgery outcomes (in the complete dataset) are shown in Figure 1.

Overall Predictive Quality of Models

To compare the overall predictive quality of the three different models (LM, GAMM and ANN), the root mean squared error (RMSE) based on the validation data set was calculated. The results are displayed in Table 1. As compared to models based on a classical statistical approach (LM and GAMM), the ANN showed a propensity towards subpar predictive quality, which can be derived from the slightly higher RMSE values. Only minor differences can be detected when comparing LM and GAMM, which is not surprising since both models contained nearly the same covariates (Table 2).

Covariates and Effects

Table 2 gives an overview of the covariates included into the LM and GAMM models and their respective effects on the model fit for SE outcomes. Including attempted SE in the GAMM improved the predictions the most (RMSE decreases from 0.410 to 0.385). Adding further covariates improved the model fit slightly further, however, the RMSE differences decreased for each additional covariate. The coefficients of the LM can be interpreted as usual. For example, a higher attempted SE by 1 D resulted in a 0.092 D higher predicted residual refractive error. Both classical models have in common that the predicted residual refractive error was higher in left eyes as compared to right eyes. The interpretation of a single covariate effect, however, only describes the direction in terms of higher or lower values for the predicted residual refractive error. To assess, whether the covariates confer a propensity to refractive under- or overcorrection, all relevant covariates as well as their interactions must be considered. In the GAMM, we enabled the algorithm to flexibly decide if the effect of a continuous variable to the response should be linear (as in the LM) or non-linear (any kind of smooth function). In most cases, the GAMM selected a linear trend. However, some effects were estimated as smooth curve (e.g. the effect of TIA). Details on the shape of these curves can be obtained from the supplementary Figures 1 and 2.

Table 3 summarizes the covariates included into the LM and GAMM models and their respective effects on the model fit for astigmatic outcomes, which can be interpreted in the same way as Table 2. Including TIA in the GAMM determined the largest step in the improvement of the model (RMSE decreases from 0.326 to 0.277). Including further variables did not further improve the model fit to this extent. Both classical statistical models included nearly the identical variables in similar order. In the LM, the coefficients of TIA and attempted SE were positive. This indicates that with higher attempted changes the predicted residual refractive error increased. The other variables astigmatism power deviation, astigmatism axis deviation and Kmax had a negative impact with increasing absolute values. In contrast to the models for spherical residual refractive error, the laterality showed no impact on astigmatic outcomes.

From the ANN no classical parameters can be extracted but the complexity of the model can be expressed by the depth of the ANN. During the optimization of the ANN, several hyper-parameters were optimized. In both models (spherical equivalent and astigmatic outcomes) an ANN with 3 hidden layers was selected with Rectified Linear Unit “relu” activation function. To prevent overfitting, the number of units was allowed to become 1 in the final layer and a penalty of 0.001 was added in the MSE loss. The ANN on spherical equivalent resulted in 8 8 8 units for the hidden layers together with “adam” optimizer based on an initial learning rate of 0.01. A similar ANN was selected during the optimization of the hyper-parameters with respect to astigmatism. Here, an ANN with 16 16 16 units was selected together with “relu” activation function and “adam” optimizer based on an initial learning rate of 0.01. These parameters express that high dimensional combinations of the input variables are necessary to fit the data appropriately. However, due to the limited number of input variables and the rather linear relationship of the variables it seems as if the ANN was not able to fully express its benefits. This can be observed from the slightly higher RMSE of the ANN models as compared to LM and GAMM.

Optimizing Clinical Outcomes

Figure 2 shows the actual refractive outcomes as compared with the hypothetically achievable refractive outcomes when optimized by the LM, GAMM and the ANN, respectively. As also observable in Figure 1 (D and E), the original SMILE procedure without application of a nomogram resulted in an undercorrection of approximately 0.20 D of SE (Figure 2A). After correcting with LM or GAMM, the mean diminished towards 0 and also the variance decreased. Adjustment by ANN resulted in a similar variance, but with a small propensity towards overcorrection. Figure 2B shows the astigmatic outcome for the different models. Only subtle differences between models could be detected, as the original SMILE procedure without application of any nomogram already resulted in residual astigmatic errors with a mode of approximately 0.

To further compare the models from a clinical perspective, the absolute values of the differences in SE as well as in astigmatic outcomes were categorized (Figure 3). By applying LM, GAMM and ANN, respectively, the proportion of eyes within ±0.50 D of SE from target refraction could be expanded from 82% to 83%, 84% and 83%, respectively (Figure 3A). In contrast, the yield of eyes within ±1.00 D of SE from target refraction could barely be extended from 98% to 99% with all three models alike. Regarding astigmatic outcomes (Figure 3B), applying LM, GAMM and ANN, respectively, resulted in an increased yield of eyes within ±0.50 D from TIA from 90% to 93%, 93% and 92%, respectively. Only GAMM yielded an improvement of eyes ±1.00 D from 99% (without application of a nomogram) to 100%.

Discussion

Nomogram creation for optimization of functional outcomes is considered a *sine qua non* component of modern keratorefractive surgery. As Mrochen already put forward in 2006 and as later confirmed by Mosquera et al. in 2018, contemporary nomograms are nevertheless limited to a predictability of approximately 90% of eyes within ±0.50 D of SE from the intended surgical refractive change.[10,24] Linear regression and multiple linear regression analyses are considered the statistical gold standard methodologies in contemporary nomogram creation. Thereby, preoperative patient- (e.g. age) or treatment-related parameters (e.g. manifest refraction) are correlated with postoperative refractive outcomes in order to adjust the surgical refractive correction accordingly. This allows for systemic under- and overcorrection to be eliminated.[10]

By applying these conventional nomograms, accurate refractive outcomes for treatment of myopia or myopic astigmatism can readily be achieved with the SMILE procedure. For SMILE, the refractive predictability reported in the peer-reviewed literature commonly varies from 80% to 87% of eyes within ±0.50 D SE of attempted SE.[25,26] Unfortunately, however, in the vast majority of contemporary keratorefractive publications it is not stated whether or not any nomogram correction had been applied at all. As a positive example, using a conventional nomogram, SE predictability of SMILE was remarkably improved in a study by Liang et al.[27] By applying simple linear regression model, this group increased the ±0.50 D predictability from 70% to 86% and the ±1.00 D predictability from 97% to 98%.

As the adoption of AI in IOL power calculation has allowed for enhanced refractive outcomes of intraocular lens surgery, it stands to reason as to whether similar advances could be achieved in keratorefractive surgery nomograms with the help of AI.[7,8] In the first work of its kind, Cui et al. conducted a larger scaled, prospective study of 510 myopic SMILE treatments using a ML-based nomogram.[12] However, methodological shortcomings of the study limited its power. The ANN nomogram, which was mainly affected by the preoperative manifest refraction (and the co-correlated amount of corneal stromal ablation), outperformed the comparator nomogram with regards to its ±0.50 D predictability (93% versus 83%). Unfortunately, however, it remains unclear how the conventional nomogram was created in detail. In addition, the study investigated purely spherical predictability and neglected the astigmatic component of the keratorefractive correction. This resulted in higher dispersion of SE in their ML group. The authors concluded that the limited eye samples with high myopia and astigmatism, the maximum of the latter comprising -3.5 D, were the reason for this limitation.[12]

Hence, the present study set out to create a multi-layer ANN to predict not only spherical but also cylindrical refractive outcomes in a substantial sample of 1,342 SMILE treatments. In a second step, the predictive power of the model was compared with state-of-the art conventional nomograms based on LM and GAMM. With all three models, an improvement in refractive predictability could be achieved with regards to both spherical and astigmatic predictability. Using the different models, the proportion of eyes within ±0.50 D SE from target refraction could be expanded from 82% to 83-84%. In respect to residual refractive cylinder, the yield of eyes within ±0.50 D from TIA could be increased from 90% to 92-93%. Potential reasons as to why ANN was unable to outperform the conventional statistical models may be found in the nature of the input variables. For a ML-algorithm as complex as ANN, a total of 16 input variables is still regarded as quite limited and the data inherited rather linear relationship. Hence, the ANN model may not have been able to fully utilize its benefits. Especially for astigmatic outcomes, only minimal differences between models could be detected. This was probably due to the “zero-inflated” nature of our data (residual astigmatic errors after SMILE showed a mode of approximately 0).

The most influential covariates affecting the nomograms were the attempted SE regarding spherical predictability and TIA regarding cylindrical predictability. By adding further covariates to the models, the improvements in predictability were clinically negligible. Our findings are in good agreement with a recent theoretical study by Park et al. that compared a plethora of ML algorithms (e.g. decision tree, AdaBoost and ANN) for SMILE nomogram development.[28] Based on a dataset of 3,034 eyes, the authors identified preoperative manifest refraction as the single most influential parameter affecting the nomograms.

Our data indicates that - even when incorporating leading-edge AI-based methodology - nomograms still fail to outperform the relatively simpler conventional regression-based models, neither are they able to approximate “perfect” predictability (i.e. 100% of eyes within ±0.50 D from target refraction). Since the preoperative manifest refraction seems to represent the most influential covariate in all aforementioned studies, it might be considered as potential weak spot. Manifest refraction is a highly subjective measurement depending on several incalculable factors, e.g. idiosyncrasies of the refracting examiner, fluctuations in the patients' accommodation, in working distance, trial lens vertex distance or power of the trial lenses.[10] In their study of 150 eyes that underwent keratorefractive surgery, Mosquera et al. computed an uncertainty in subjective refraction of approximately 0.6 D for a measurement which is commonly performed in 0.25 D steps. Hence, subjective refraction may be regarded as the major limitation to improving the accuracy of refractive surgery nomograms.[10] Hence, future research effort should be dedicated towards precise and more “objective” determination of subjective manifest refraction. A promising approach was recently provided by the working group of Damien Gatinel.[29] Fittingly, the authors showed that subjective spherocylindrical refraction could be accurately and precisely predicted by machine learning from polynomially decomposed ocular wavefront aberrometry data, which represents an objective measurement modality.

Limitation to the present study may be found. First and foremost, the nomogram models developed in this study were not tested prospectively in a clinical setting. Instead, the models were virtually validated by assuming that the adjustment in the surgical treatment plan transfers linearly to the achieved surgically induced refractive change - an approach that is commonly used when developing refractive surgery nomograms.[10] Moreover, including further potentially relevant variables into the models (e.g. biomechanical corneal data) might have enabled more precise predictions. A strength of the study is the large sample size as well as the inclusion of combination-parameters (e.g. astigmatism axis deviation) as well as the elaborated statistical analysis including non-linear models and adjustment for paired-eye data.

In conclusion, the present study of 1,342 SMILE treatments showed that machine learning endorses the validity of state-of-the-art conventional nomograms for adjustment of spherocylindrical refractive outcomes. However, it was unable to outperform their predictability. Improving the precision of the single most influential covariate in refractive surgery nomograms - subjective manifest refraction - seems warranted for optimizing refractive outcomes beyond the apparent methodological ±0.50 D accuracy limit of 90%.

Acknowledgements:

We thank Dr. Hannah Spitzer (Institute of Computational Biology, Helmholtz Zentrum München) for her statistical expertise and support.

Competing interests:

No author reports any financial interest relevant to this study.

Jakob Siedlecki received previous speaker fees and travel expenses from Novartis Pharma GmbH, Carl Zeiss Meditec AG, Oculentis OSD Medical GmbH and Pharm-Allergan GmbH. Jakob Siedlecki received personal consultation fees from Bayer AG.

Benedikt Schworm received previous speaker fees and travel expenses from Novartis Pharma GmbH and Topcon Corporation.

Wolfgang Mayer received previous financial, consultation and speaker fee support from Carl Zeiss Meditec AG, Alcon Pharma GmbH, and Ziemer Ophthalmology GmbH. Wolfgang Mayer received previous financial and speaker fee support from Novartis Pharma GmbH, Oculentis OSD Medical GmbH, Pharm-Allergan GmbH and

Bausch & Lomb GmbH. Wolfgang Mayer received previous speaker fees from Heidelberg Engineering GmbH, Oertli AG, and STAAR Surgical AG.

Siegfried G. Priglinger received previous speaker fees and/or travel expenses from Novartis Pharma GmbH, Oertli AG, Bayer AG, Alcon Pharma GmbH and Pharm-Allergan GmbH.

Nikolaus Luft received income from honoraria as a lecturer from Carl Zeiss Meditec AG, Alcon Laboratories Inc., NIDEK Co. Ltd and CenterVue SpA.

Data availability statement:

The data that support the findings of this study are available from the corresponding author, NL, upon reasonable request.

References

1. Ting DSW, Pasquale LR, Peng L, Campbell JP, Lee AY, Raman R, Tan GSW, Schmetterer L, Keane PA, Wong TY. Artificial intelligence and deep learning in ophthalmology. Br J Ophthalmol 2019;103(2):167-175.
2. Ting DSJ, Foo VH, Yang LWY, Sia JT, Ang M, Lin H, Chodosh J, Mehta JS, Ting DSW. Artificial intelligence for anterior segment diseases: Emerging applications in ophthalmology. Br J Ophthalmol 2020.
3. Lopes BT, Ramos IC, Salomão MQ, Guerra FP, Schallhorn SC, Schallhorn JM, Vinciguerra R, Vinciguerra P, Price FW, Jr., Price MO and others. Enhanced Tomographic Assessment to Detect Corneal Ectasia Based on Artificial Intelligence. Am J Ophthalmol 2018;195:223-232.
4. Xie Y, Zhao L, Yang X, Wu X, Yang Y, Huang X, Liu F, Xu J, Lin L, Lin H and others. Screening Candidates for Refractive Surgery With Corneal Tomographic-Based Deep Learning. JAMA Ophthalmol 2020;138(5):519-526.
5. Ruiz Hidalgo I, Rodriguez P, Rozema JJ, S ND, Zakaria N, Tassignon MJ, Koppen C. Evaluation of a Machine-Learning Classifier for Keratoconus Detection Based on Scheimpflug Tomography. Cornea 2016;35(6):827-32.
6. Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, Lee G, Rim TH. Explainable Machine Learning Approach as a Tool to Understand Factors Used to Select the Refractive Surgery Technique on the Expert Level. Transl Vis Sci Technol 2020;9(2):8.
7. Cheng H, Kane JX, Liu L, Li J, Cheng B, Wu M. Refractive Predictability Using the IOLMaster 700 and Artificial Intelligence-Based IOL Power Formulas Compared to Standard Formulas. J Refract Surg 2020;36(7):466-472.
8. Carmona González D, Palomino Bautista C. Accuracy of a new intraocular lens power calculation method based on artificial intelligence. Eye (Lond) 2020.
9. Koprowski R, Lanza M, Irregolare C. Corneal power evaluation after myopic corneal refractive surgery using artificial neural networks. Biomed Eng Online 2016;15(1):121.
10. Arba Mosquera S, de Ortueta D, Verma S. The art of nomograms. Eye Vis (Lond) 2018;5:2.
11. Yang SH, Van Gelder RN, Pepose JS. Neural network computer program to determine photorefractive keratectomy nomograms. J Cataract Refract Surg 1998;24(7):917-24.
12. Cui T, Wang Y, Ji S, Li Y, Hao W, Zou H, Jhanji V. Applying Machine Learning Techniques in Nomogram Prediction and Analysis for SMILE Treatment. Am J Ophthalmol 2020;210:71-77.
13. Alpins NA, Goggin M. Practical astigmatism analysis for refractive outcomes in cataract and refractive surgery. Surv Ophthalmol 2004;49(1):109-22.
14. Reinstein DZ, Archer TJ, Gobbe M. Small incision lenticule extraction (SMILE) history, fundamentals of a new refractive surgery technique and clinical outcomes. Eye Vis (Lond) 2014;1:3.
15. Luft N, Siedlecki J, Sekundo W, Wertheimer C, Kreutzer TC, Mayer WJ, Priglinger SG, Dirisamer M. Small incision lenticule extraction (SMILE) monovision for presbyopia correction. Eur J Ophthalmol 2018;28(3):287-293.
16. Steinwender G, Shajari M, Mayer WJ, Kook D, Ardjomand N, Vidic B, Kohnen T, Wedrich A. Impact of a Displaced Corneal Apex in Small Incision Lenticule Extraction. J Refract Surg 2018;34(7):460-465.

17. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria; 2020.
18. Lin X, Zhang D. Inference in generalized additive mixed modelsby using smoothing splines. Journal of the royal statistical society: Series b (statistical methodology) 1999;61(2):381-400.
19. Wood SN. Generalized additive models: an introduction with R. CRC press; 2017.
20. Wood S, Scheipl F. Package 'gamm4'. American Statistician 2017;45:339.
21. Allaire J, Tang Y. tensorflow: R Interface to 'TensorFlow'. 2020.
22. Allaire J, Chollet F. keras: R Interface to 'Keras'. 2020.
23. Allaire J. tfruns: Training Run Tools for 'TensorFlow'. 2018.
24. Mrochen M, Hafezi F, Iseli HP, Loffler J, Seiler T. [Nomograms for the improvement of refractive outcomes]. Ophthalmologe 2006;103(4):331-8.
25. Han T, Xu Y, Han X, Zeng L, Shang J, Chen X, Zhou X. Three-year outcomes of small incision lenticule extraction (SMILE) and femtosecond laser-assisted laser in situ keratomileusis (FS-LASIK) for myopia and myopic astigmatism. Br J Ophthalmol 2019;103(4):565-568.
26. Taneri S, Kiessler S, Rost A, Schultz T, Dick HB. Small-incision lenticule extraction for the correction of myopic astigmatism. J Cataract Refract Surg 2019;45(1):62-71.
27. Liang G, Chen X, Zha X, Zhang F. A Nomogram to Improve Predictability of Small-Incision Lenticule Extraction Surgery. Med Sci Monit 2017;23:5168-5175.
28. Park S, Kim H, Kim L, Kim JK, Lee IS, Ryu IH, Kim Y. Artificial intelligence-based nomogram for small-incision lenticule extraction. Biomed Eng Online 2021;20(1):38.
29. Rampat R, Debellemaniere G, Malet J, Gatinel D. Using Artificial Intelligence and Novel Polynomials to Predict Subjective Refraction. Sci Rep 2020;10(1):8565.

Figures

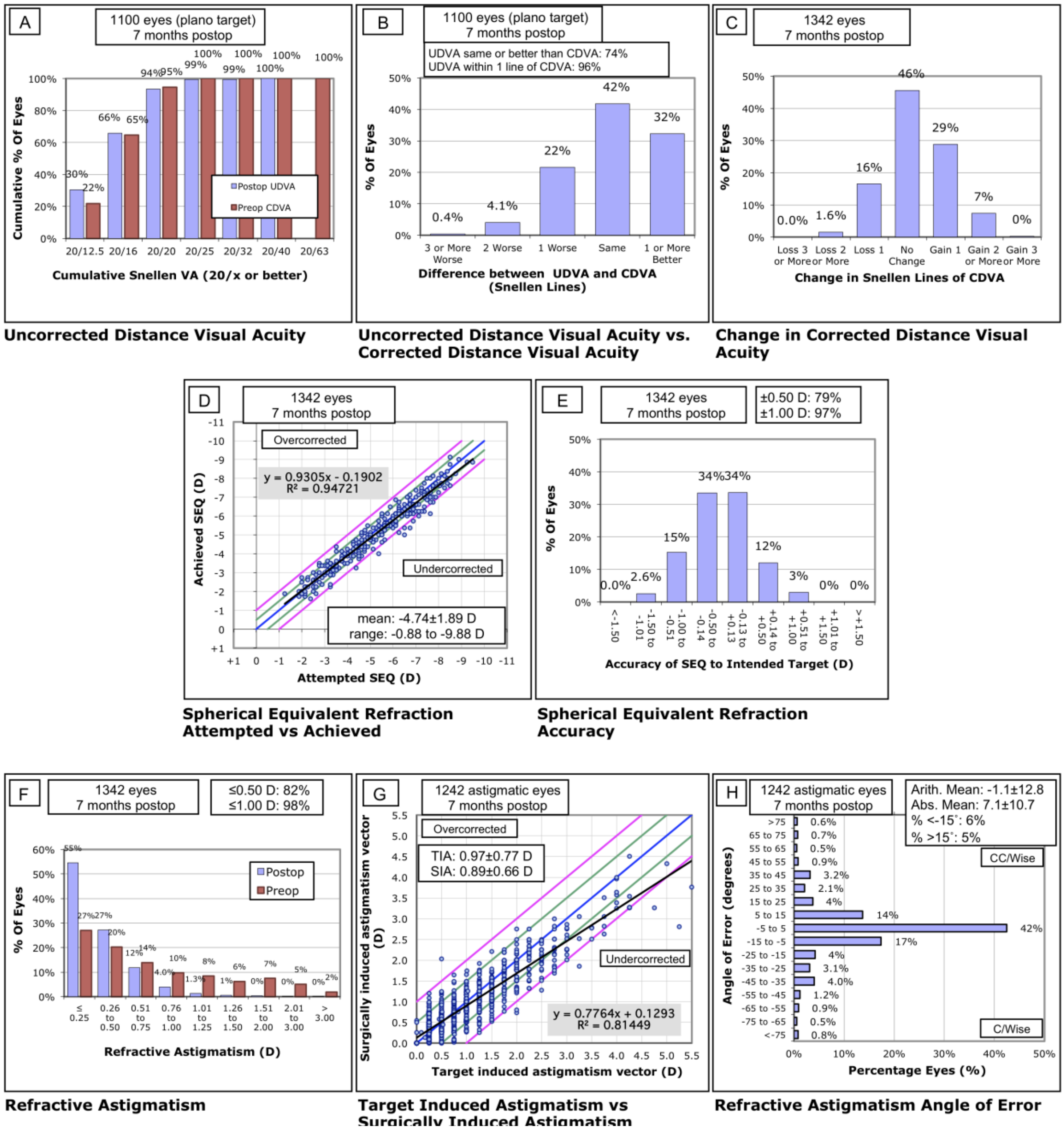


Figure 1: The nine standard graphs for reporting refractive surgery outcomes.

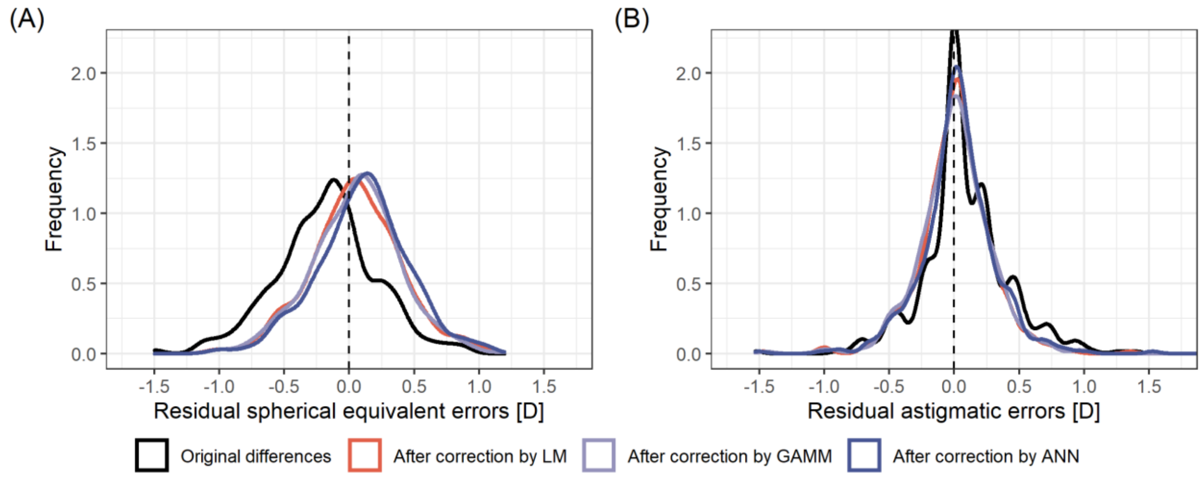


Figure 2: Comparison of kernel density estimates of refractive outcomes without model application and after optimization with the three models: linear model (LM), mixed model (GAMM) and artificial neural network (ANN). Residual spherical equivalent errors are displayed in the left figure (A), residual astigmatic errors are displayed on the right (B).

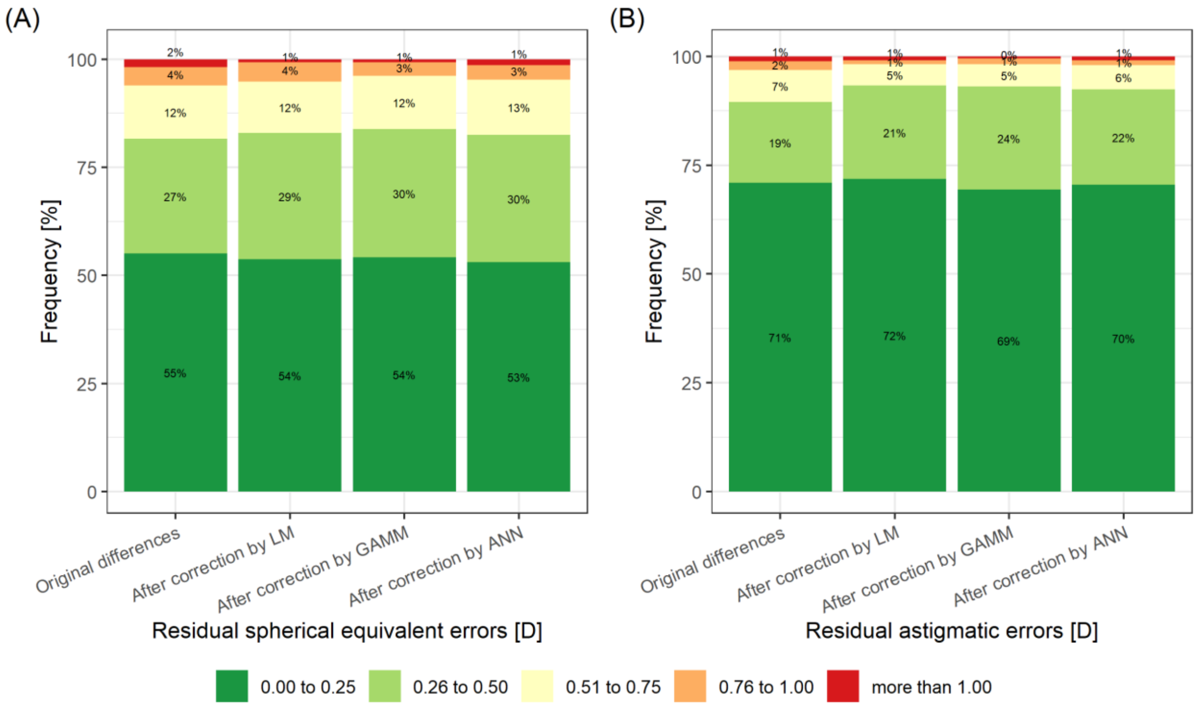
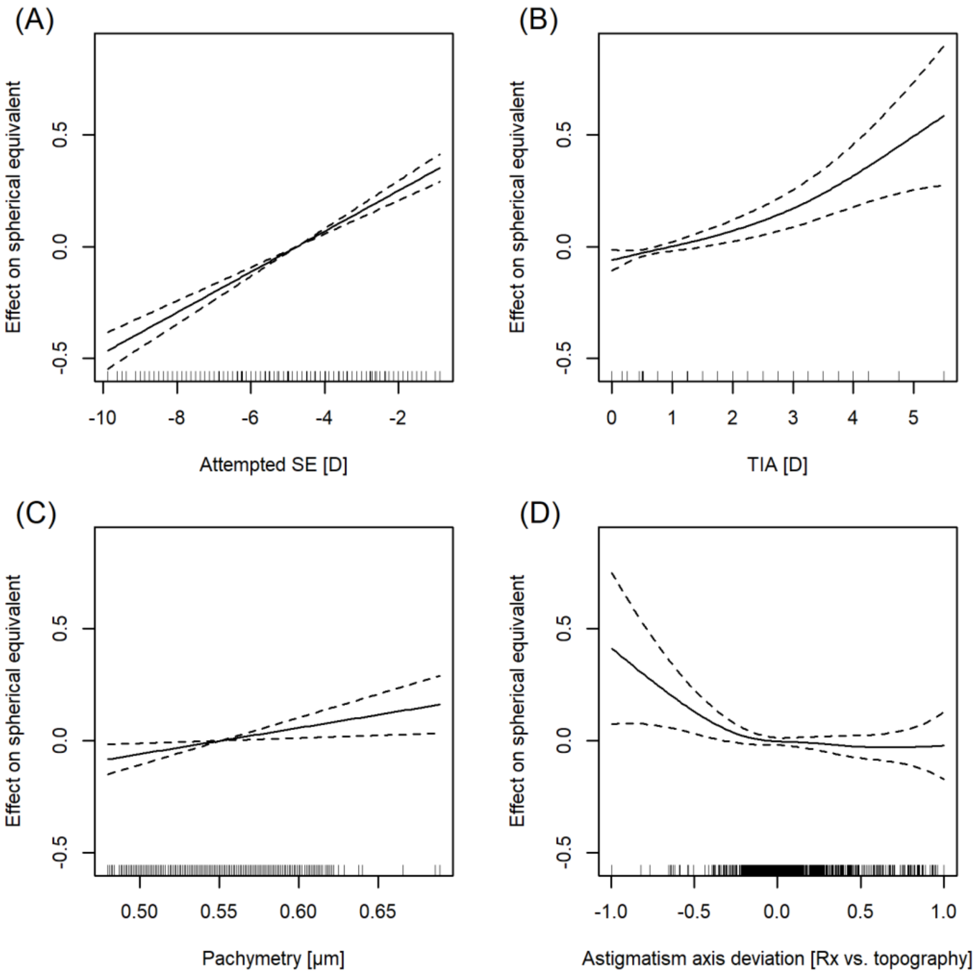
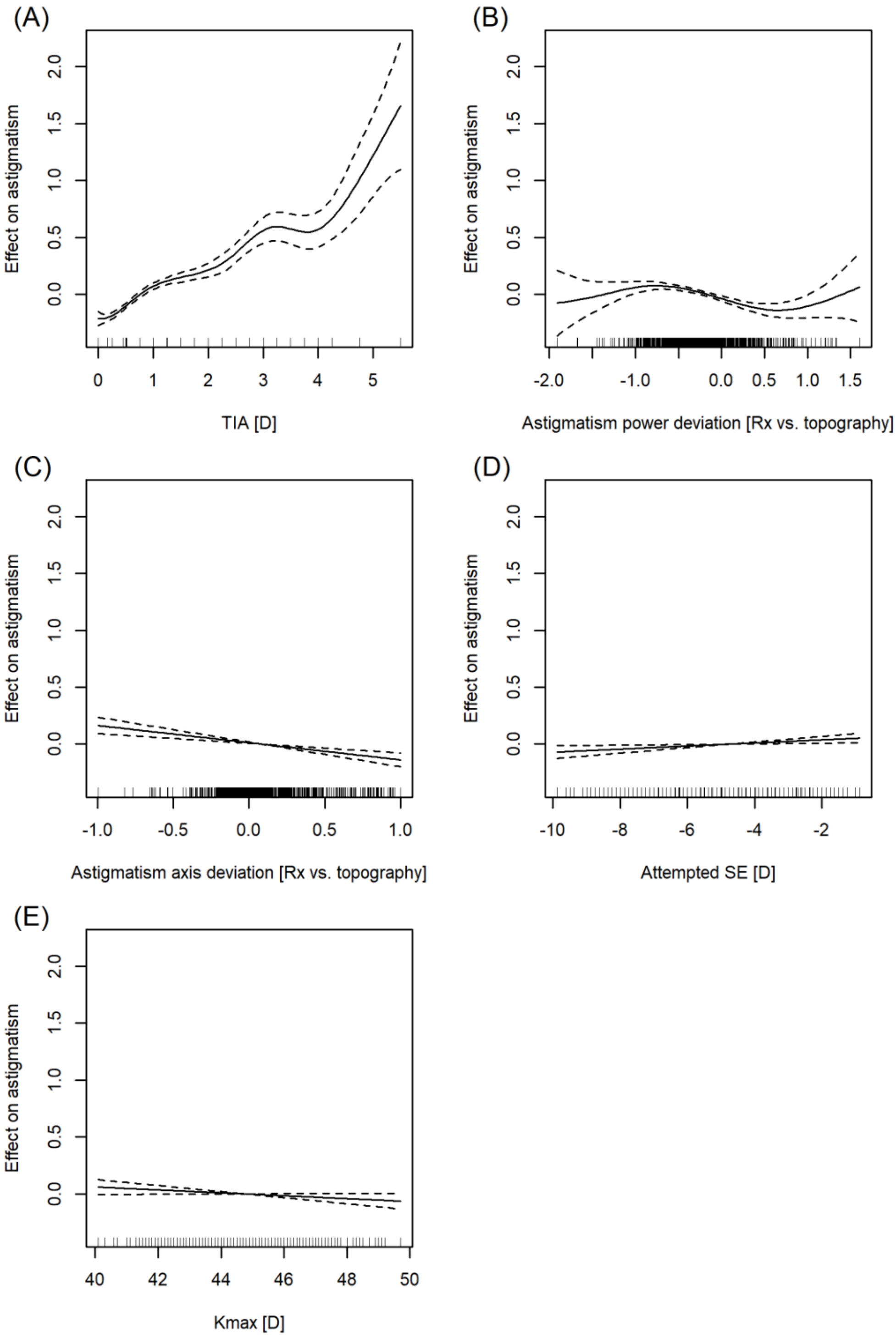


Figure 3: Categorized refractive outcomes without model application and after optimization with the three models: linear model (LM), mixed model (GAMM) and artificial neural network (ANN). Residual spherical equivalent errors are displayed in the left figure (A), Residual astigmatic errors are displayed on the right (B).



Supplementary Figure 1: GAMM-estimated smooth curves of the covariates on the response "spherical equivalent". The curves are per definition centered around 0 and only their shapes are interpretable (e.g. with increasing TIA the predicted difference between attempted and achieved increases).



Supplementary Figure 2: GAMM-estimated smooth curves of the covariates on the response astigmatism. The curves are per definition centered around 0 and only their shapes are interpretable (e.g. with increasing TIA the predicted difference between SIA and TIA increases).

Tables:

Table 1: Root mean square error (RMSE) of the models in the validation data set.

Model	RMSE spherical equivalent	RMSE astigmatism
ANN	0.367	0.290
GAMM	0.348	0.278
LM	0.355	0.279

GAMM= generalized additive mixed model; LM=linear model; RMSE=root mean square error

Table 2: Root mean square error (RMSE) of the selection process of the mixed model (GAMM) and coefficients and p-values of the linear model (LM) for spherical equivalent outcomes

	GAMM	LM		
Variables	**RMSE**	**Coefficient**	**p-value**	**(exact)**
(Intercept)	0.410	-0.563	p=0.0118	(1.18e-02)
Attempted SE [D]	0.385	0.092	p<0.0001	(1.86e-32)
Laterality [OS]	0.380	0.127	p<0.0001	(5.11e-07)
TIA [D]	0.377	n.i.		
Cap thickness [≥ 135 µm]	0.376	0.026	p=0.375	(3.75e-01)
Cap thickness [≤ 125 µm]	0.376	0.139	p=0.0012	(1.22e-03)
Pachymetry [µm]	0.375	1.248	p=0.0024	(2.47e-03)
Astigmatism axis deviation [Rx vs. topography]	0.375	n.i.		
Optical zone [> 6.5 mm]	n.i.	-0.091	p=0.0479	(4.79e-02)
Optical zone [< 6.5 mm]	n.i.	0.104	p=0.328	(3.28e-01)

D=Diopter; GAMM= generalized additive mixed model; LM=linear model; n.i.=not included in the model due to variable selection; RMSE=root mean square error; Rx=Refraction; SE=spherical equivalent; TIA=target induced astigmatism

Table 3: Root mean square error (RMSE) of the selection process of the mixed model (GAMM) and coefficients and p-values of the linear model (LM) for astigmatic outcomes

Variables	GAMM RMSE	Coefficient	p-value	LM (exact)
(Intercept)	0.326	0.556	p=0.0472	(4.72e-02)
TIA [D]	0.277	0.237	p<0.0001	(5.20e-56)
Astigmatism power deviation [Rx vs. topography]	0.274	-0.112	p<0.0001	(2.41e-06)
Astigmatism axis deviation [Rx vs. topography]	0.272	-0.163	p<0.0001	(7.99e-07)
Attempted SE [D]	0.271	0.014	p=0.0057	(5.65e-03)
Kmax [D]	0.270	-0.014	p=0.0255	(2.55e-02)

D=Diopter; GAMM= generalized additive mixed model; Kmax=location of the maximum corneal curvature; LM=linear model; RMSE=root mean square error; Rx=refraction; SE=spherical equivalent; TIA=target induced astigmatism

Supplementary File 1

The complete statistical analysis code and instructional README file can be obtained *via* https://ascgitlab.helmholtz-muenchen.de/elmar.spiegel/optimizing_refractive_outcomes_of_smile