



Accuracy of a new intraocular lens power calculation method based on artificial intelligence

David Carmona González ^{1,2} · Carlos Palomino Bautista^{1,2}

Received: 28 September 2019 / Accepted: 8 April 2020 / Published online: 28 April 2020
© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2020

Abstract

Purpose The purpose of this study is to develop and assess the accuracy of a new intraocular lens (IOL) power calculation method based on machine learning techniques.

Methods The following data were retrieved for 260 eyes of 260 patients undergoing cataract surgery: preoperative simulated keratometry, mean keratometry of posterior surface, axial length, anterior chamber depth, lens thickness, and white-to-white diameter; model and power of implanted IOL; and subjective refraction at 3 months post surgery. These data were used to train different machine learning models (*k*-Nearest Neighbor, Artificial Neural Networks, Support Vector Machine, Random Forest, etc). Implanted lens characteristics and biometric data were used as input to predict IOL power and refractive outcomes. For external validation, a dataset of 52 eyes was used. The accuracy of the trained models was compared with that of the power formulas Holladay 2, Haigis, Barrett Universal II, and Hill-RBF v2.0.

Results The SD of the prediction error in order of lowest to highest was the new method (designated Karmona) (0.30), Haigis (0.36), Holladay 2 (0.38), Barrett Universal II (0.38), and Hill-RBF v2.0 (0.40). Using the Karmona method, 90.38% and 100% of eyes were within ± 0.50 and ± 1.00 D respectively.

Conclusions The method proposed emerged as the most accurate to predict IOL power.

Introduction

When planning cataract surgery, there are many formulas and methods available to calculate the refractive power of an intraocular lens (IOL). These calculation methods can be divided into three broad groups based on Gaussian-beam optics, real ray tracing, or artificial intelligence algorithms. However, there is no consensus regarding the best formula to use, as no single formula has proved to be highly accurate across a range of eye characteristics. According to most authors, the formula selected needs to be based on the anatomical and optical parameters of the patient.

In a study by Melles et al. [1] conducted in 18,501 eyes, SDs of the prediction error for the different formulas were

from lowest to highest: Barrett Universal II (0.404), Olsen (0.424), Haigis (0.437), Holladay 2 (0.450), Holladay 1 (0.453), SRK/T (0.463), and Hoffer Q (0.473). Values reported by Cooke et al. [2] for 9 IOL power formulas in 1079 eyes were as follows: OlsenStandalone (0.361), Barrett Universal II (0.365), Haigis (0.393), Super Formula (0.403), Holladay 2 (0.404), Holladay 1 (0.408), Hoffer Q (0.428), and SRK/T (0.433).

According to the recent literature, the formulas and methods providing the most predictable outcomes are Barrett Universal II, Haigis, Holladay 2, Olsen, and Hill-RBF [3–6]. Fourth- and fifth-generation formulas can use as many as seven variables to predict effective lens position and IOL power. For example, Holladay 2 uses keratometry, axial length (AXL), anterior chamber depth (ACD), lens thickness (LT), white-to-white (WTW), age, and preoperative refraction.

Hill-RBF [7] v2.0 is a data-driven IOL calculation approach and is therefore free of the limitation of lens-position estimation. Its most recent update from 3445 to 12,419 eyes has meant a significant improvement [8]. The novelty of this method is that it is based on machine learning techniques and, therefore, the model learns and improves as more cases are introduced. The Hill-RBF

✉ David Carmona González
davcargon@hotmail.com

¹ Hospital Universitario QuironSalud, Madrid, Spain

² Department of Medicine, School of Biomedical Sciences and Health, Universidad Europea de Madrid, Madrid, Spain

method has an outlier detection feature that excludes certain cases in which the calculator is likely to be out of bounds.

The new IOL power calculation method proposed here is also data driven. It is an ensembled model that combines predictions from separate models. Unlike Hill's model, WTW distance and central LT are taken into account, and it also includes a variable defined as the ratio between the curvature of the anterior and posterior surface of the cornea.

In the present study, the new model is developed and its predictive accuracy is compared to that offered by the popular methods Holladay 2, Barret Universal II, Haigis, and Hill-RBF.

Materials and methods

This retrospective study was performed at the Hospital Universitario QuironSalud, Madrid, Spain. The study protocol adhered to ethics codes based on the tenets of the Declaration of Helsinki and received institutional review board approval (Ref. EO153-18_HUQM).

Data were retrieved for 260 eyes of 260 patients undergoing uneventful cataract surgery. The preoperative biometric data collected were simulated keratometry, mean keratometry of posterior corneal surface, AXL, ACD, LT, and WTW distance. Further data recorded for each eye were the model and power of the IOL implanted, and subjective refraction 3 months after cataract surgery.

Patient exclusion criteria were amblyopia, corneal astigmatism > 1 D measured with simulated keratometry (simk), and a history of ocular disease or surgery and intraoperative or postoperative complications.

Ten different models of non-toric IOLs were implanted in the capsular bag. The lens constants used were those optimized by each surgeon for each IOL during their routine clinical practice. All patients were operated on with phacoemulsification (Centurion Alcon, Inc., Fort Worth, USA). Surgery was performed through a clear 2.2 mm temporal corneal incision.

Preoperative biometric measurements were made by two optometrists using the IOL Master 700 (Carl Zeiss Meditec AG, Jena, Germany) and corneal topography with the Pentacam v1.20r87 (Oculus Optikgerate, Wetzlar, Germany).

Follow-up visits were scheduled for 1 day, 1 month, and 3 months after surgery. In the 3-month visit, subjective refraction and best-corrected distant visual acuity were measured by an optometrist. The target of IOL power calculation was always emmetropia.

Model variables

Biometric variables for the whole study sample are provided Table 1. Anterior segment depth (ASD) was calculated as

Table 1 Descriptive statistics for the biometric variables examined ($n = 260$).

Biometric Variables	Mean \pm SD	Range
Ratio B/Fx100	82.80 \pm 1.73	[77.88, 88.04]
AXL (mm)	23.41 \pm 1.51	[19.33, 28.67]
ACD (mm)	3.10 \pm 0.41	[1.96, 4.19]
LT (mm)	4.57 \pm 0.38	[3.18, 5.52]
WTW (mm)	11.94 \pm 0.41	[10.90, 13.60]
K-mean anterior (mm)	7.73 \pm 0.28	[6.68, 8.76]
K-mean posterior (mm)	6.40 \pm 0.28	[5.54, 7.41]
ASD (mm)	7.66 \pm 0.38	[6.96, 8.85]

the sum of ACD and LT. The variable R_B/F defined as the ratio of the back to front corneal surface central radius was also introduced in the model. Therefore, the average radius in the 3 mm zone on the anterior and posterior corneal surfaces were measured with Pentacam. For example: Radius back = 6.88 mm and Radius front = 7.80 mm, R_B/F = 6.88/8.30 = 0.83

Refractive outcomes

The refraction prediction error (RPE) was calculated as the difference between the postoperative spherical equivalent and that predicted by each formula. In addition, the mean absolute error (MAE) or absolute value of the predictive error and the median of the absolute error (MedAE) were determined for the new model and the rest of the formulas. The percentage of eyes within ± 0.50 D, ± 0.75 D, and ± 1.00 D was also calculated for each formula.

The mean values of RPE for each formula were zeroed out [9, 10] to eliminate the systematic error derived from the different constants used in the calculation and their possible non-optimization. After adjustment, MAE and MedAE were recalculated for each formula.

Modeling

We worked with RStudio version 1.1.423 (R Foundation, Boston, USA) and the caret library (Classification And REgression Training) of Max Kuhn [11] for development, adjustment, validation, and comparison of each model. The RStudio software was also used for comparative statistical analysis of the new model with the rest of the formulas.

The total sample was randomized and split into two stages: training (on 208 eyes) and external validation (on 52 eyes) following an 80/20 ratio. Thus, 52 eyes were used for comparisons between the new model and the remaining formulas. The validation dataset provides an unbiased assessment of model fit on the training dataset, while tuning

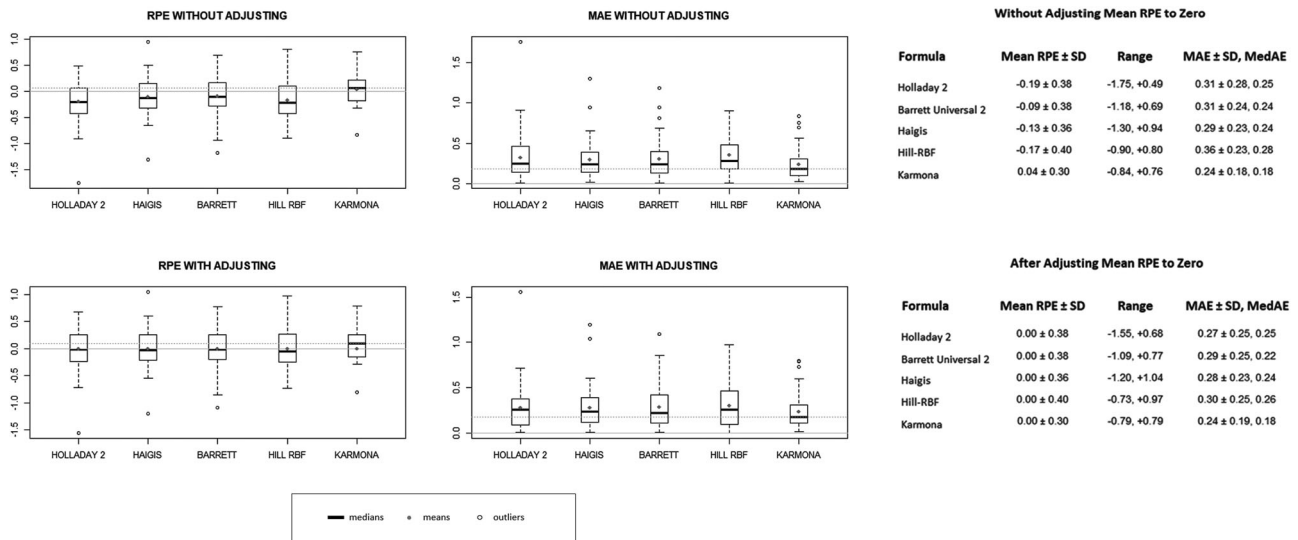


Fig. 1 Refractive predictions and mean absolute errors with and without adjusting the mean numerical refractive prediction to zero.

the model's hyperparameters. The model (e.g., a neural network or k -Nearest Neighbors (KNN)) is trained on the training dataset using a supervised learning method.

Eleven nonlinear regression models based on machine learning techniques were constructed: generalized linear model, KNN, linear discriminant analysis, Support Vector Machines (SVMs) with radial basis function (RBF), SVM with linear kernel, neural networks testing several hidden layers, random forest, decision trees, LASSO regression, multivariate adaptive regression spline (MARS), and stochastic gradient boosting.

Of these 11 models, we selected the two that showed best performance. The first of these models is based on SVM with a Gaussian kernel RBF and the second one is based on MARS with second-order polynomials. These two models were hyperparametrized and tuned to optimize their performance, improving their metrics with respect to raw models. Finally, an ensemble model was generated by combining the two models through a stacking technique [12] and designated Karmona. Ensembles are machine learning methods that combine predictions from separate models.

Although it is known that validating a model by means of cross-validation or bootstrapping techniques yields a good estimate of the model error, a final prediction based on new non-touch observations (test sample) serves to ensure that no overfitting has been generated during the optimization procedure.

As the resampling method, we used LOOCV (leave-one-out cross-validation) also known as "jackknife". LOOCV involves separating the data so that for each iteration we have a single sample for the test data and the rest comprises the training data.

Statistical analysis

The normality of RPE, MAE, and medAE between formulas and methods was assessed using the Kolmogorov–Smirnov test with Lilliefors correction followed by the paired t - and the Wilcoxon signed-rank tests. Finally, a non-parametric Friedman test was used as suggested by Aristomeu et al. [13] and Benavoli et al. [14]. For post-hoc analysis, we used a Nemenyi test for multiple comparisons. All statistical tests were performed using the package RStudio version 1.1.423 (R Foundation, Boston, USA). Significance was set at $p < 0.05$.

Results

The proposed Karmona stacked regression model showed an R^2 of 0.9955 and a root mean square error of 0.3808.

For the external validation of Karmona and comparisons with each formula, we used a test dataset (52 eyes). Refractive prediction errors for each formula with and without adjusting mean RPE to zero are provided in Fig. 1.

The formula Holladay 2 resulted in the highest mean myopic predictive error (-0.19 ± 0.38 D), followed by Hill-RBF, which gave rise to a mean hyperopic predictive error of $+0.17 \pm 0.40$ D ($p = 0.7147$). The Karmona method achieved the lowest RPE, MAE, and MedAE (0.07 ± 0.30 D, 0.24 ± 0.19 D, and 0.18 D, respectively). Our new model also returned the lowest SD and lowest maximum error, regardless of whether or not we adjusted for residual predictive errors. In terms of RPE without adjustment, differences were detected for Karmona ($p < 0.05$) vs. the other formulas. Pairwise comparisons using the Nemenyi

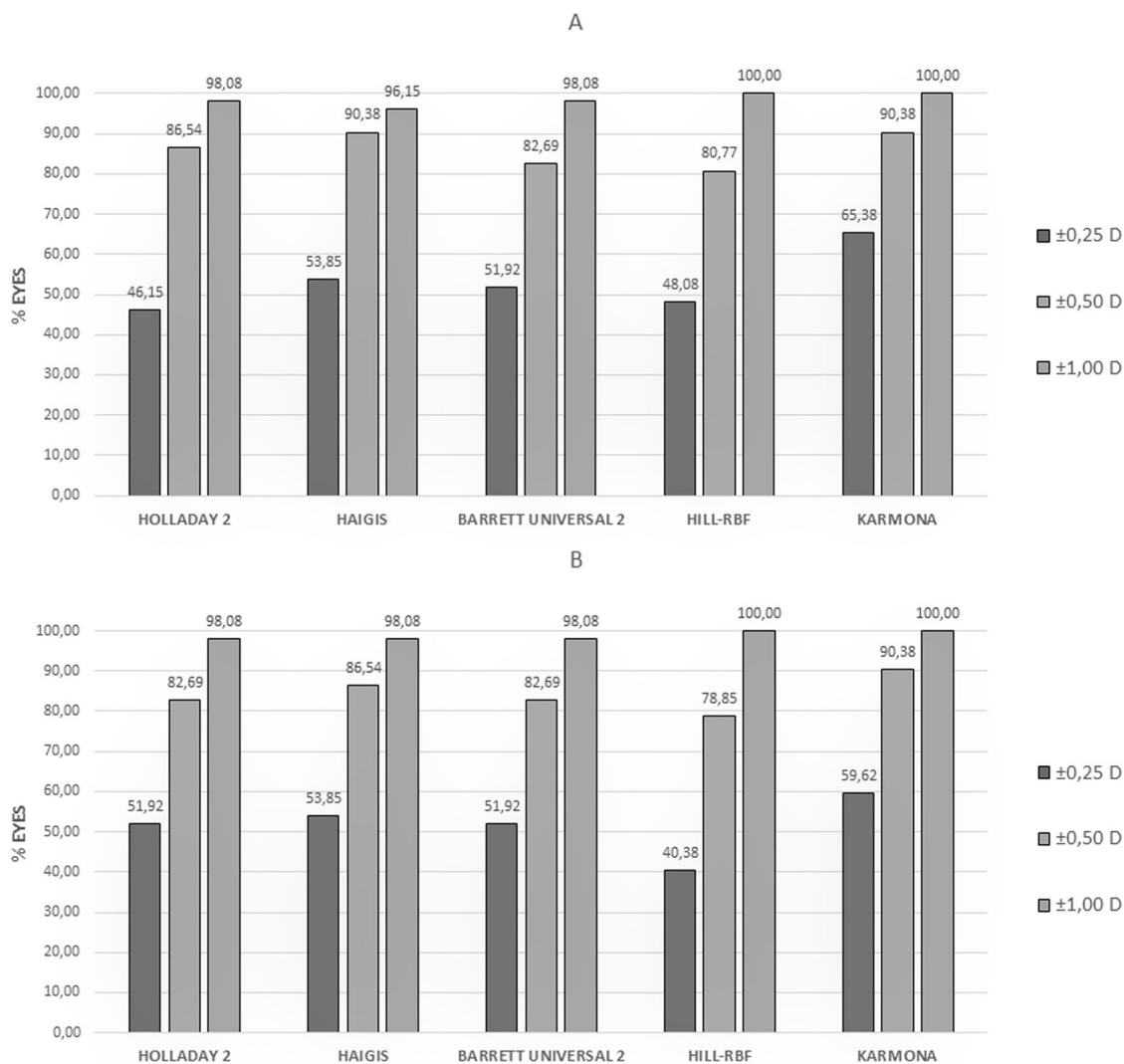


Fig. 2 Percentages of eyes within the ranges. **a** Percentages of eyes with refractive (spherical equivalent) prediction errors within ± 0.25 D, ± 0.50 D, and ± 1.00 D; mean numerical refractive prediction errors

adjusted to zero. **b** Percentages of eyes with refractive (spherical equivalent) prediction errors within ± 0.25 D, ± 0.50 D, and ± 1.00 D; no adjustment.

multiple comparison test for MAE and MedAE revealed significant differences between Karmona and Hill-RBF ($p = 0.012$).

Figure 1 provides a boxplot of RPE and MAE with and without zero adjustment. The solid line shows 0.00 D as the optimal level. The dashed line represents the Karmona median, aiding comparisons with each formula. The dots indicate the means of each category represented. In the case of Karmona, both means and medians were below those provided by the remaining formulas. The interquartile range for Karmona was the smallest in the two categories analyzed (RPE and MAE).

Figure 2a, b shows the percentage of eyes in the ranges ± 0.25 D, ± 0.50 D, and ± 1.00 D, respectively. Highest percentages in the three dioptric ranges analyzed, both with and without zero adjustment, were obtained using Karmona ($p < 0.05$). The second best formula in the ± 0.25 D and

± 0.50 D ranges was Haigis (53.85% and 86.54%, respectively), followed by Barrett Universal II and Holladay 2, which yielded similar percentages. The worst performance was observed for Hill-RBF at 40.38% and 78.85% of eyes in the ranges ± 0.25 D and ± 0.50 D, respectively.

In Table 2, we provide the categories analyzed for all the formulas and their rank order by SD.

DISCUSSION

This study examines the accuracy of a new data-driven IOL power calculation method and compares it with other formulas in terms of predicted refractive outcomes. A major difference between our method, Karmona, and other formulas, is the incorporation of the ratio between the curvatures of the posterior and anterior corneal surfaces. As far as

Table 2 Refractive outcomes sorted by SD after adjusting the mean to zero.

Formula	MPE (D)	SD (D)	MAE (D)	MedAE (D)	MaxError (D)	± 0.25 D (%)	± 0.50 D (%)	± 1.00 D (%)	Ranking
Karmona	0.00	0.30	0.24	0.18	0.79	65.38	90.38	100	1
Haigis	0.00	0.36	0.28	0.24	1.20	53.85	90.38	96.15	2
Holladay 2	0.00	0.38	0.27	0.25	1.55	46.15	86.54	98.08	3
Barrett U2	0.00	0.38	0.29	0.22	1.09	51.92	82.69	98.08	4
Hill-RBF	0.00	0.40	0.30	0.26	0.97	48.08	80.77	100	5

we are aware, no other formula takes into account the posterior corneal surface. Another important predictor that was introduced in the model was ASD, which replaces LT and ACD as independent variables, acquiring more specific weight in the radial kernel SVM model than mean keratometry itself. Interestingly, the predictive model of Hill, also based on machine learning, does not take into account WTW distance, central corneal thickness, or central LT. Although these parameters are requested in the calculator data entry form, they have no weight in the calculation.

Our performance indicators contrast with those reported in other recent studies. In a sample of 3122 eyes, Kane et al. [15] recorded MAE values for the formulas Holladay 2, Haigis, Barrett Universal II, and Hill-RBF of 0.410, 0.409, 0.381, and 0.407 D, respectively (SDs were not specified). In our study, MAE values were lower at 0.27 D, 0.28 D, 0.29 D, and 0.30 D for the same formulas. These authors identified Barrett Universal II as the best formula out of 9 examined, 72.80% of eyes being in the range ± 0.50 D using this formula compared with 82.69% found in the present study. As may be observed in Table 2, Barrett Universal II occupied third place below Haigis and Holladay 2, and Karmona obtained higher values in all performance indicators than the values obtained in Kane's work.

In the comparison by Melles et al. [1] of 8 formulas in a sample of 18,501 eyes, MAE \pm SD (MedAE) were 0.350 ± 0.45 (0.29) D, 0.338 ± 0.44 (0.28) D, and 0.311 ± 0.40 (0.25) D for Holladay 2, Haigis, and Barrett Universal II, respectively. Hill-RBF was not included. These values are also higher than those obtained here of 0.27 ± 0.38 (0.25) for Holladay 2, 0.28 ± 0.36 (0.24) for Haigis, and 0.29 ± 0.38 (0.22) for Barrett Universal II. Moreover, our MAE \pm SD (MedAE) of 0.24 ± 0.30 (0.18) for Karmona is better than those reported by Melles' group.

Roberts et al. [5] obtained in 400 eyes, MAE \pm SD (MedAE was not included) values of 0.32 ± 0.27 D for Holladay 2, 0.32 ± 0.24 D for Hill-RBF, and best values of 0.30 ± 0.24 D for Barrett Universal II. Haigis was not included. We obtained similar MAE \pm SD values for Barrett Universal II (0.29 ± 0.38) and also a similar percentage of eyes in the range ± 0.50 D, 82.69% compared with 81% in the study by Roberts et al. [5]. The new Karmona method

offered better MAE \pm SD and higher percentages of eyes in the ranges ± 0.25 D, ± 0.50 D, and ± 1.00 D.

If we compare our data with those reported by Cooke et al. [2], who examined the accuracy of 9 formulas in 1079 eyes and obtained MAE \pm SD (MedAE) of 0.33 ± 0.42 (0.29), 0.32 ± 0.40 (0.27), and 0.31 ± 0.39 (0.26), and percentages of 79.30%, 79.80%, and 80.60% in the range ± 0.50 D for Holladay 2, Haigis, and Barrett Universal II, respectively, again Karmona emerged as the more accurate IOL power calculation method.

Study limitations

Our sample of 260 eyes was split 80 : 20. Eighty percent of eyes were used for model training and 20% for validation or testing. Therefore, the final model obtained was based on data for 208 eyes and the external validation sample used for comparison with the rest of the formulas consisted of a small random dataset derived from 52 eyes. For our whole study sample, AXL was above 25.00 mm (25.00–28.67) in 35 eyes, below 22.00 mm (19.33–22.00) in 31 eyes, and between 22.00 and 25 mm (normal) in the remaining 194 eyes. We are presently expanding the sample size of the model to enhance its prediction power.

Another limitation of our work is that we did not stratify the data by eye length. In future work, we will segment the population under study according to axial and keratometric characteristics to analyse the real potential of the proposed model.

In favor of the Karmona model, we could say that its performance was excellent despite the small dataset on which it was based. The addition of further data is likely to improve its accuracy even more. In addition, the high variation of data included in the model such as ten monofocal IOLs used is a strong point.

Conclusions

The Karmona model proposed here emerged as more accurate than the third-generation formula Haigis, the fourth generation Holladay 2 and Barrett Universal II, and the fifth-generation Hill-RBF. We are presently preparing an

open source web site where an updated version of Karmona will be available in Shiny from RStudio [16].

Summary

What was known before

- Intraocular lens power calculation methods and formulas.

What this study adds

- To develop and assess a new intraocular lens power calculation method based on machine learning techniques.

Acknowledgements The contribution of Ana Burton as a reviewer has been supported by Department of Medicine of the School of Biomedical Sciences and Health (Universidad Europea de Madrid).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Melles RB, Holladay JT, Chang WJ. Accuracy of intraocular lens calculation formulas. *Ophthalmology* 2018;125:169–78.
- Cooke DL, Cooke TL. Comparison of 9 intraocular lens power calculation formulas. *J Cataract Refract Surg*. 2016;42:1157–64.
- Wang Q, Jiang W, Lin T, Zhu Y, Chen C, Lin H, et al. Accuracy of intraocular lens power calculation formulas in long eyes: a systematic review and meta-analysis. *Clin Exp Ophthalmol*. 2018;46:738–49.
- Abulafia A, Barrett GD, Rotenberg M, Kleinmann G, Levy A, Reitblat O, et al. Intraocular lens power calculation for eyes with an axial length greater than 26.0 mm: comparison of formulas and methods. *J Cataract Refract Surg*. 2015;41:548–56.
- Roberts TV, Hodge C, Sutton G, Lawless M. Comparison of Hill-radial basis function, Barrett Universal and current third generation formulas for the calculation of intraocular lens power during cataract surgery. *Clin Exp Ophthalmol*. 2018;46:240–6.
- Gökce SE, Zeiter JH, Weikert MP, Koch DD, Hill W, Wang L. Intraocular lens power calculations in short eyes using 7 formulas. *J Cataract Refract Surg*. 2017;43:892–7.
- Hill W. Hill-RBF calculator. Available at: <http://rbfcalculator.com/online/index.html>. Accessed 1 Jan 2018.
- Reitblat O, Levy A, Kleinmann G, Lerman TT, Assia EI. Intraocular lens power calculation for eyes with high and low average keratometry readings: comparison between various formulas. *J Cataract Refract Surg*. 2017;43:1149–56.
- Hoffer KJ, Aramberry J, Haigis W, Olsen T, Savini G, Shammas HJ, et al. Protocols for studies of intraocular lens formula accuracy. *Am J Ophthalmol*. 2015;160:403–5.
- Koch DD, Hill W, Abulafia A, Wang L. Pursuing perfection in intraocular lens calculations: I. Logical approach for classifying IOL calculation formulas. *J Cataract Refract Surg*. 2017;43:717–8.
- Kuhn M. Building predictive models in R using the caret package. *J Stat Software*. 2008;28:1–26.
- Leo B. Stacked regressions. *Mach Learn*. 1996;24:49–64.
- Aristodemou P, Cartwright NEK, Sparrow JM, Johnston RL. Statistical analysis for studies of intraocular lens formula accuracy. *Am J Ophthalmol*. 2015;160:1085–6.
- Benavoli A, Corani G, Mangili F. Should we really use post-hoc tests based on mean-ranks? *J Mach Learn Res*. 2016;17:1–10.
- Kane JX, Van Heerden A, Atik A, Petsoglou C. Accuracy of 3 new methods for intraocular lens power selection. *J Cataract Refract Surg*. 2017;43:333–9.
- Shiny from RStudio. Available at: <https://shiny.rstudio.com/>. Accessed 4 May 2019.