tvst

Artificial Intelligence

Improved IOL Power Calculation With Femtosecond Laser Enhanced Refractive Outcome Prediction

Jeroen Van Der Donckt[1], Joshua A. Young[2], Michael Rademaker[1], Saurabh Menon[3], Chin-Wen Chang[4], Gilles Vandewiele[1], Benjamin Straker[5], David Dewey[5], George Dai[5], Javier Gonzalez[5], Joseph R. Free[5], Sofie Van Hoecke[1], Wendell Scott[6], Shachar Tauber[6], H. Burkhard Dick[7], and Charles Scales[5]

[1] IDLab-imec, Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium
[2] Department of Ophthalmology, New York University School of Medicine, New York, NY, USA
[3] Mu Sigma Business Solutions LLC, Bengaluru, India
[4] Johnson & Johnson, Global Commercial Data Science, Raritan, NJ, USA
[5] Johnson & Johnson, Research and Development, Milpitas, CA, USA
[6] Department of Ophthalmology, Mercy Clinic, Springfield, MO, USA
[7] Department of Ophthalmology, Ruhr University Eye Hospital, Bochum, Germany

Correspondence: Charles Scales, RWD Operations & Analytics, Johnson & Johnson, 7500 Centurion Parkway, Jacksonville, FL 32256, USA. e-mail: cscales@its.jnj.com

Received: April 22, 2025
Accepted: September 23, 2025
Published: November 14, 2025

Keywords: cataract; surgery; artificial intelligence (AI); intraocular lens (IOL); optical coherence tomography (OCT); femtosecond; laser

Citation: Van Der Donckt J, Young JA, Rademaker M, Menon S, Chang CW, Vandewiele G, Straker B, Dewey D, Dai G, Gonzalez J, Free JR, Van Hoecke S, Scott W, Tauber S, Dick HB, Scales C. Improved IOL power calculation with femtosecond laser enhanced refractive outcome prediction. Transl Vis Sci Technol. 2025;14(11):15, https://doi.org/10.1167/tvst.14.11.15

Purpose: The purpose of this study was to introduce and evaluate the femtosecond laser enhanced refractive outcome (FLERO) prediction method, an intraocular lens (IOL) calculator that augments Barrett Universal II (BUII) by integrating novel anterior segment optical coherence tomography (OCT) biometric predictors obtained during femtosecond laser-assisted cataract surgery (FLACS).

Methods: Two thousand, three hundred sixty-three (2363) eyes of 1720 patients (mean age = 71.33 years, 60.26% women) undergoing FLACS were analyzed. FLERO was developed by selecting the most predictive subset of OCT-derived biometry features using a "genetic algorithm" and combining them with BUII predictions in a linear model. Internal validation was performed through cross-validation, and prediction errors (PEs) were compared with BUII and Kane errors.

Results: Compared to BUII, FLERO increased the proportion of eyes achieving postoperative refraction within ±0.25 diopter (D), ±0.50 D, and ±1.00 D of target from 0.470 to 0.507, 0.781 to 0.824, and 0.962 to 0.970, respectively. Mean absolute error decreased from 0.345 D for BUII and 0.338 D for Kane to 0.315 D for FLERO. FLERO outperformed BUII and Kane across (short, medium, and long) eyes, where proportions of eyes achieving refraction within ±0.50 D were 0.696, 0.831, and 0.782 for FLERO, 0.468, 0.796, and 0.718 for BUII, and 0.595, 0.798, and 0.718 for Kane. Wilcoxon Signed-Rank testing indicated significant reductions in absolute PEs for FLERO versus BUII and Kane ($P < 0.0001$). PE regression revealed FLERO made significantly smaller errors.

Conclusions: FLERO enhances BUII by incorporating novel OCT-derived FLACS biometric parameters across short, medium, and long eyes.

Translational Relevance: FLERO combines advanced FLACS-derived intraoperative biometry with established IOL formulae to refine refractive outcome prediction.

Introduction

Residual refractive error remains a leading cause of dissatisfaction after cataract surgery.[1] The chief determinant in choosing an intraocular lens (IOL) that will produce the desired refractive result is the application of an accurate IOL power calculation formula, itself dependent on biometry. The best IOL calculation formulae demonstrate competitively low mean errors (MEs; zero or very nearly so) and mean absolute errors (MAEs).[2] To this end, machine learning has

Copyright 2025 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Downloaded from tvst.arvojournals.org on 05/04/2026

played a key role in the refinement of IOL calculations, with the introduction of useful formulae, such as the Hill Radial Basis Function,[3,4] Kane,[5,6] and others.[7,8] However, these formulae remain limited by the small set of biometric inputs typically collected in routine clinical practice. Further improvement may be realized by incorporation of more predictive biometric measurements.

Femtosecond laser-assisted cataract surgery (FLACS) not only provides surgical precision but also captures extensive intraoperative biometric data, including high-resolution anterior segment 2D and 3D optical coherence tomography (OCT) images. These datasets contain rich biometric detail that is not typically used in IOL power calculations. This study introduces the femtosecond laser enhanced refractive outcome (FLERO) approach, a new hybrid IOL prediction method that enhances the widely used Barrett Universal II (BUII) formula by integrating its predictions with novel OCT-derived anterior segment parameters in a linear regression model. A key step in FLERO's design is the use of a two-step "genetic algorithm" (GA) process to identify the most informative subset from the vast number of FLACS-derived features. We evaluated FLERO against BUII and Kane, two widely studied fourth-generation IOL formulae,[9] in a retrospective analysis of cataract procedures by Mercy Clinic ophthalmologists using the CATALYS Precision Laser System (Johnson & Johnson, Milpitas, CA).

Methods

Overall Study Design

To build an IOL calculation algorithm incorporating the large OCT data feature set captured by the CATALYS, an artificial intelligence (AI) approach toward feature selection was used prior to cross-validation evaluation of the resulting predictive model. The CATALYS itself provides more than 1800 biometric values for each eye undergoing FLACS. Combining these with conventional biometry captured with Zeiss IOL Master (both 500 and 700), the total number of biometric values equaled 1968. In the language of machine learning, these variables are referred to as "features" and the processes of paring down and thereby transforming these features into a most predictive feature subset is known as feature selection.

Given the vast number of data features collected by CATALYS, the study used a GA with backward feature elimination (BFE) to (1) objectively identify informative features and (2) "weed out" or "prune" features that did not contribute significantly to the performance of the IOL calculation model. This form of feature selection removes not only noncontributory and nonpredictive features but also those that are so highly correlated with each other as to be redundant. The GA enabled an efficient reduction of the large number of features, which was further refined by backward feature elimination, ensuring optimal model complexity and generalization. A GA is an optimization procedure that starts with a set of random tentative solutions to a given problem, in this case, feature subsets used for training predictive models. Through iterative rounds, termed "generations", the algorithm selects and adapts the solutions until no more improvements can be made.

Within this study, the GA operated with a population of 400 tentative solutions and ran for 500 generations. Tentative solutions competed via tournament selection with a tournament size of three random solutions. The best performing solutions were chosen for potential modification and inclusion in the next generation's competition. Modification of solutions occurred through stochastic "mutation" and crossover operations with an aim to maintain population diversity. The performance or "fitness" of a candidate solution was evaluated using the area under the diopter (D) error curve: "$|E| < d$" with d in [0, 1] increasing with steps of 0.01.

Following the above feature selection process, internal validation was conducted using seven-fold grouped cross-validation. A linear ensemble model incorporating the selected features, which include the BUII prediction, was then trained and validated. The grouped cross-validation procedure ensured that eyes from the same patient were kept within the same fold, and therein, further confined within either the test or training set for that fold. This grouping strategy eliminated potential data leakage associated with related eyes (from a common patient) being used within both test and training sets during cross-validation.

Incorporation of BUII

In addition to conventional biometry data from the IOL Master and the manifold OCT data from the CATALYS, the predicted post-operative SE for a given implanted IOL power was calculated from the BUII formula and itself was selected (by the GA) and thus incorporated as a feature in the model. This makes FLERO a hybrid IOL calculator: it builds on the theoretical foundation of BUII while simultaneously integrating FLACS-derived features in a data-driven manner. Specifically, this design makes FLERO

Downloaded from tvst.arvojournals.org on 05/04/2026

function as an ensemble (boosting) model, learning from BUII's prediction errors (PEs) to provide more accurate refractive outcome predictions.

Data Source

The study utilized the Mercy dataset, comprising data from patients with cataract admitted to Mercy Clinic, Springfield, Missouri, USA, between 2015 and 2020, and was conducted retrospectively using the available data that met the eligibility criteria outlined below. Ethics approval from an institutional review board and informed consent were not required for this study as it used data from an anonymous, de-identified, electronic health record (EHR) database. The study adhered to the US Department of Health and Human Services guidance regarding "Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act Privacy Rule" and complied with the tenets of the Declaration of Helsinki.

Inclusion Criteria

The study cohort was derived from a database of cataract surgeries performed between 2015 and 2020, with inclusion and exclusion rules conducted at the eye (rather than patient) level. The cohort included eyes that underwent routine phacoemulsification with IOL implantation (CPT 66984) or complex phacoemulsification requiring a specialized approach (CPT 66982) and received a Johnson & Johnson (J&J) IOL. No specific inclusion or exclusion criteria were applied to differentiate between eyes belonging to patients having one or both eyes treated, as the grouped cross-validation procedure ensured that both eyes from the same patient were assigned together to either the training or test set. Eyes were excluded if they had a history of cataract surgery in the same eye or had prior intraocular procedures, including corneal, retinal, or glaucoma surgeries, except for laser trabeculoplasty, laser iridotomy, or IOL exchange. Only eyes with valid laterality information were retained. Cases where the procedure was prematurely terminated, as indicated by billing procedure modifiers 73 and 74, were excluded. Eyes were also removed if postoperative spherical equivalent was not recorded within 15 to 120 days after surgery or if postoperative visual acuity exceeded a LogMAR value of 0.1 (equivalent to Snellen 20/25), as this threshold ensured sufficient visual function to allow accurate manifest refraction and thereby improved the reliability of postoperative spherical equivalent outcomes. Although prior guidelines[10] recommended a more lenient threshold of 20/40 (LogMAR 0.3), subsequent evidence has shown that stricter thresholds of LogMAR 0.12 to 0.20 yield significantly lower mean PEs.[11] For eyes having multiple SE measurements collected within the 15-to-120-day postoperative period, the latest (last date-time) value within that date range was selected. A similar approach was used for eyes having multiple preoperative biometry measurements within 90 days before surgery, that is, the measurement made closest to the surgery date was selected; however, for eyes where multiple biometry measurements were made on the same day closest to surgery, the median values from that day were calculated and used. Eyes with missing or invalid biometric measurements—including anterior chamber depth (ACD), axial length (AL), and steep and flat keratometry within 90 days before surgery—were excluded, as were eyes with a lens thickness below 3 mm. Additional exclusions included eyes classified as pseudophakic or aphakic at baseline, those with afferent pupillary defect, and those with a central corneal thickness below 0.4 mm. Only eyes with confirmed linkage to the CATALYS Laser Precision System records, indicating fully completed surgery, were retained. Deduplication of records was performed. Finally, only eyes with valid BUII predictions were kept, whereas those with extreme input values or implausible predictions were excluded.

This selection process resulted in a well-defined cohort of 2363 eyes from 1720 patients, comprising 1196 right and 1167 left eyes. Approximately 95% of the procedures were Current Procedural Terminology (CPT) 66984, whereas the remaining were CPT 66982. The demographic and clinical characteristics are shown in Table 1.

Table 1. Patient Demographic and Clinical Characteristics

Demographics	Mean (SD)
Female sex	60.26%
Age	71.33 (8.31)
Axial length (AL)	23.92 (1.17)
Average keratometry	43.87 (1.61)
Anterior chamber depth (ACD)	3.16 (0.38)
IOL power	20.70 (3.32)
AL subgroups, *n* (%)	
Short (≤22 mm)	79 (3.3)
Medium (22 mm < AL < 26 mm)	2160 (91.4)
Long (≥26 mm)	124 (5.2)
Observed refractive outcome	−0.351 (0.62)

ACD, anterior chamber depth; AL, axial length; IOL, intraocular lens; SD, standard deviation.

Downloaded from tvst.arvojournals.org on 05/04/2026

Data Features From FLACS Workflow

Data features from across the FLACS workflow were leveraged within this study, including preoperative ZEISS IOLMaster data, intraoperative CATALYS OCT data, as well as post- and perioperative data from Mercy's Epic EHR system. EHR data provided demographic details, medical history including comorbidities, and information gathered during ophthalmic examinations. Preoperative biometric measures, including parameters such as AL and keratometry measures, were obtained from ZEISS IOLMaster 500 ($N = 155$, 6.56%) or 700 ($N = 2208$, 93.44%) devices. The CATALYS data features comprised structured OCT data from axial and sagittal OCT scans as well as unstructured 3D OCT data. Derived features from unstructured OCT CATALYS scans included fitted cornea and lens surfaces as well as crystalline lens power (see Supplementary Table S1).

Main Outcome Measure

The postoperative refractive state of the patient was obtained from the Mercy Clinic EHR. Of interest, the distribution of postoperative refractive spherical equivalent in the Mercy data was skewed toward a mild myopic outcome with a mean of −0.351 D (see Table 1; Supplementary Fig. S1) which may suggest surgeons' preferences to target a slightly myopic result. Note that the surgeon's target is not explicitly available in the dataset, nor is it required to evaluate a predictive model.[10] It suffices to have knowledge of the implanted IOL power and the actual postoperative refractive spherical equivalent outcome to calculate model accuracies.

Statistical Analysis

BUII, Kane, and FLERO were compared head-to-head in their ability to predict the actual refractive spherical equivalent outcome of each eye following FLACS. BUII was selected because it is arguably the most widely used IOL calculation formula in clinical practice[12–15] and because FLERO builds upon its predictions, making it a key comparator. Kane was chosen for its consistently high predictive accuracy.[15–19] Comparisons with other AI-based formulas, such as PEARL-DGS and Hoffer QST, are provided in the Supplementary materials.

Besides conventional metrics (e.g., SD and median absolute error [MedAE]) and the proportion of eyes within key dioptric thresholds, a rigorous statistical evaluation of absolute prediction errors was performed to ensure meaningful comparisons between the IOL calculators. Because absolute errors do not follow a normal distribution, nonparametric statistical methods were used. Specifically, the Wilcoxon Signed-Rank test was utilized to evaluate the null hypothesis that the mean absolute PE of the FLERO model is greater than or equal to that of the comparison formulae (BUII and Kane). Furthermore, to validate this analysis with best practice according to literature, the method proposed by Holladay et al.[20] for evaluating non-Gaussian PEs of SE associated with IOL calculators was also applied. Finally, to investigate the clinical significance of FLERO compared with BUII, the same analyses were repeated for the subset of cases in which prediction differences would have led to a different IOL selection.

Surgical Context

All patients with cataract received femtosecond laser cataract surgery (CATALYS, Johnson & Johnson) at the Mercy National Eye Surgery Center in Springfield, Missouri. As all patients undergoing cataract surgery at this site receive CATALYS, the femtosecond procedure is performed in the operating suite adjacent to the operating microscope, allowing both procedures to be performed in the same room and minimizing the time between the laser treatment and the phacoemulsification procedure. All patients received an acrylic intraocular lens with an A-constant of 119.3 (TECNIS Series, Johnson & Johnson).

Results

The feature selection process reduced the feature set to 18 features and included the BUII prediction. Internal validation of these features was conducted using seven-fold grouped cross-validation. The subsequent section presents the results of this internal validation.

FLERO Performance Over BUII and Kane

The final selected model used linear regression with L2 (Ridge) penalty, and a regularization strength set to 1. This model proved to have the dual advantages of excellent performance and a high degree of explainability comprising 19 parameters: one weight for each of the 18 input features and an additional weight for the bias. Using requisite input features from each eye within the Mercy dataset, head-to-head model performance was assessed between FLERO, and adjusted (ME-subtracted) outputs of BUII and Kane (referred

Downloaded from tvst.arvojournals.org on 05/04/2026

Table 2. Metric-Based Comparison of BUII, Kane, and the Proposed FLERO Approaches on the Mercy Dataset

						Absolute Error Threshold, |E|				
Formula	ME	SD	MedAE	MAE	MSE	<0.25	<0.50	<0.75	<1.0	<2.0
BUII (unadjusted)	−0.200	0.470	0.310	0.382	0.261	0.412	0.720	0.892	0.952	0.994
BUII (adjusted)	**0.000**	0.470	0.268	0.345	0.221	0.470	0.781	0.916	0.962	0.995
Kane (unadjusted)	−0.107	0.463	0.270	0.350	0.226	0.460	0.758	0.911	0.963	0.995
Kane (adjusted)	**0.000**	0.463	0.258	0.338	0.215	0.488	0.787	0.916	0.964	0.995
FLERO	0.001	**0.430**	**0.245**	**0.315**	**0.185**	**0.507**	**0.824**	**0.933**	**0.970**	**0.997**

BUII, Barrett Universal II; FLERO, femtosecond laser enhanced refractive outcome; MAE, mean absolute error; ME, mean error; MedAE, median absolute error; MSE, mean squared error.

For BUII and Kane, both the unadjusted and the adjusted (i.e. ME-subtracted) versions of the formulae are analyzed.

The scores in bold represent the best score for each metric.

Table 3. Metric-Based Comparison of BUII, Kane, and the Proposed FLERO Approaches on Subsets of the Mercy Dataset Differentiated by Eye Size

						Absolute Error Threshold, |E|				
Formula	ME	SD	MedAE	MAE	MSE	<0.25	<0.50	<0.75	<1.0	<2.0
Short eyes (AL ≤= 22 mm) [$N = 79$]										
BUII (adjusted)	0.421	0.657	0.514	0.599	0.609	0.253	0.468	0.747	0.835	0.975
Kane (adjusted)	0.406	**0.608**	0.461	0.558	0.535	0.329	0.595	0.722	0.873	0.975
FLERO	**0.105**	0.614	**0.347**	**0.459**	**0.388**	**0.380**	**0.696**	**0.823**	**0.886**	**0.987**
Medium eyes (22 mm ≤ AL ≤ 26 mm) [$N = 2160$]										
BUII (adjusted)	−0.003	0.449	0.263	0.332	0.201	0.479	0.796	0.923	0.968	0.997
Kane (adjusted)	**−0.001**	0.444	0.252	0.327	0.197	0.494	0.798	0.925	0.969	0.997
FLERO	**−0.001**	**0.418**	**0.240**	**0.307**	**0.175**	**0.517**	**0.831**	**0.939**	**0.975**	**0.998**
Long eyes (AL ≥= 26 mm) [$N = 124$]										
BUII (adjusted)	−0.212	0.525	0.303	0.396	0.321	0.452	0.718	**0.895**	0.935	0.984
Kane (adjusted)	−0.233	0.517	0.312	0.400	0.321	**0.468**	0.718	0.887	0.935	0.984
FLERO	**−0.038**	**0.478**	**0.295**	**0.362**	**0.230**	0.427	**0.782**	**0.895**	**0.962**	**1.000**

For BUII and Kane, only the adjusted (i.e. ME-subtracted) versions of the formulae are analyzed.

The scores in bold represent the best score for each metric.

to hereafter as BUII and Kane, unless noted otherwise) using metrics ME, SD, MedAE, MAE, and MSE as general model performance indicators, as well as the proportion of eyes within key dioptric thresholds, of 0.25, 0.50, 0.75, 1.00, and 2.00 D (Table 2). Overall, Kane performed at least as well as BUII across all metrics over the dataset. The FLERO approach outperformed both BUII and Kane, with the difference being most pronounced when comparing performance within 0.5 D of error, as demonstrated in Table 2. The MAE and the standard deviation over 7 folds of cross-validation was 0.315 ± 0.0068 D and the fraction of predictions with an absolute error (|E|) of less than 0.5 D was 0.8244 ± 0.0158, demonstrating overall improvement over BUII (MAE = 0.345 ± 0.0092 D, |E| < 0.5 D = 0.781 ± 0.0130) and Kane (MAE = 0.338 ± 0.0084 D, |E| < 0.5 D = 0.787 ± 0.0148).

Table 3 presents the key metrics for FLERO, and the ME-adjusted Kane and BUII formulae on the Mercy dataset for different eye sizes. For short eyes, Kane showed a slightly lower SD (0.006 lower) than FLERO, but FLERO surpasses both BUII and Kane in all other metrics, capturing 10% to 14% more cases within |E| < 0.5 D. For medium eyes, FLERO consistently outperformed the two other formulae. Finally, for long eyes, FLERO again showed the strongest overall performance, achieving the lowest SD; although Kane included 4% more cases within |E| < 0.25 D, FLERO yielded a substantially lower MSE than both alternatives, indicating fewer large mispredictions.

translational vision science & technology

Downloaded from tvst.arvojournals.org on 05/04/2026

Table 4. Parameter Estimates for FLERO Predictors

Model Term	Standardized Coefficient	Standard Error	*P* Value[a]
Intercept[b]	−0.2957	0.0095	<0.0001
Age[b]	−0.0330	0.0117	0.0049
ACG1[b]	−0.1633	0.0252	<0.0001
ACG2[b]	0.4314	0.1838	0.0190
ACG3[b]	−0.4150	0.1788	0.0204
ACD[b]	0.2105	0.0257	<0.0001
AL[b]	−0.5762	0.0983	<0.0001
Flat K[b]	−0.1818	0.0308	<0.0001
Steep K[b]	−0.2391	0.0333	<0.0001
MLP[b]	0.1491	0.0539	0.0057
IOL SE Power[b]	−0.5106	0.0785	<0.0001
PCG1[b]	0.1038	0.0380	0.0063
PCG2[b]	0.1305	0.0185	<0.0001
BUII Prediction[b]	0.2491	0.0076	<0.0001
ALR[b]	0.2648	0.0985	0.0072
ALA[b]	−0.3469	0.1376	0.0117
PLA[b]	−0.1347	0.0357	0.0002
ANP1[b]	0.1642	0.0219	<0.0001
ANP2[b]	−0.1489	0.0368	0.0001

ACD, anterior chamber depth; ACG1, anterior corneal geometry measurement 1; ACG2, anterior corneal geometry measurement 2; ACG3, anterior corneal geometry measurement 3; AL, axial length; ALA, anterior crystalline lens apex position; ALR, anterior crystalline lens radius; ANP1, axial normalization parameter, position 1; ANP2, axial normalization parameter, position 2; IOL, intraocular lens; K, Kane formula; MLP, measured crystalline lens power; PCG1, posterior corneal geometry measurement 1; PCG2, posterior corneal geometry measurement 2; PLA, posterior crystalline lens apex position.

[a]Two-sided *P* value corresponding to a Wald test assuming a null hypothesis of zero effect.

[b]Statistically significant at the usual $\alpha = 0.05$ level.

Overall, these results suggest that FLERO generalizes well across varying ALs, with best performance being observed for medium eyes.

Model Summary and Statistical Analysis

Each of the 18 selected input parameters showed statistical significance, demonstrating that the CATALYS derived features – the distinguishing characteristic of the FLERO model – are significant predictors of refractive outcome, as demonstrated in Table 4.

To determine whether the FLERO model constituted an improvement over BUII and Kane, two additional analyses were used. First, the Akaike Information Criterion (AIC) was used to assess the relative quality of these three models. The AIC considers both predictive performance and model complexity by taking the number of parameters into account, favoring smaller models while penalizing larger models. Models with smaller AIC values are preferred over those with larger AIC values. Given that the FLERO model is a refinement of BUII (unadjusted) and necessarily contains more parameters – 19 for FLERO compared to 8 for BUII (and Kane) – the AIC can thus be conceptualized as a skeptical metric of FLERO's performance. The FLERO predictor demonstrated superior performance with an AIC of 2753, compared with 3156 and 3087 for BUII and Kane, respectively. Second, an analysis of the Extra Sum of Squares (ESS) associated with BUII and Ridley was conducted. This test measures the improvement in model fit when additional parameters are introduced to an existing model and is appropriate for comparing nested models. Because FLERO aims to refine BUII's (unadjusted) prediction by incorporating it as input, ESS is a suitable metric for evaluating this refinement. Table 5 presents the results of ESS analysis in an analysis of variance (ANOVA) table. Again, 19 parameters were used for FLERO, compared with 8 for BUII, which impacted the degrees of freedom. The analysis yielded an ESS value of 86.16, result-

translational vision science & technology

Downloaded from tvst.arvojournals.org on 05/04/2026

Table 5. ANOVA Table for the ESS Analysis

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Extra term	86.16	11	7.833	42.05
FLERO	436.41	2344	0.186	
BUII	522.57	2355		

ANOVA, analysis of variance; ESS, extra sum of squares; FLERO, femtosecond laser enhanced refractive outcome.

Table 6. Statistical Comparison of Absolute Prediction Errors Between FLERO Versus BUII and Kane

Comparison	Wilcoxon Statistic	Wilcoxon *P* Value	HM Statistic	HM *P* Value
$\overline{PE}_F$ vs. $\overline{PE}_B$	1.11E6	$P < 0.0001$	−0.029	$P < 0.0001$
$\overline{PE}_F$ vs. $\overline{PE}_K$	1.17E6	$P < 0.0001$	−0.023	$P < 0.0001$

HM, Holladay et al. Method; PE, prediction error.
$\overline{PE}_F$, $\overline{PE}_B$, and $\overline{PE}_K$ represent the absolute PEs of FLERO, BUII, and Kane, respectively.
Comparison $\overline{PE}_F$ vs. $\overline{PE}_B$: H0 is $\overline{PE}_F \leq \overline{PE}_B$, H1 is $\overline{PE}_F > \overline{PE}_B$.
Comparison $\overline{PE}_F$ vs. $\overline{PE}_K$: H0 is $\overline{PE}_F \leq \overline{PE}_K$, H1 is $\overline{PE}_F > \overline{PE}_K$.

ing in an F-statistic of 42.05 and a P value < 0.0001. Both analyses collectively affirmed that FLERO represented a significant improvement over BUII (and Kane).

Statistical Comparison With Existing Formulae

Table 6 summarizes the results of the hypothesis tests, demonstrating that both statistical approaches yield the same conclusion: the absolute PEs of FLERO are significantly smaller than those of BUII and Kane. To enable meaningful comparisons between the formulae, the distributions of absolute errors were analyzed. Given that absolute errors do not follow a normal distribution, nonparametric statistical methods were used. Specifically, the Wilcoxon Signed-Rank Test was utilized to evaluate the null hypothesis, that the mean absolute PE of the FLERO model is greater than or equal to those of BUII and Kane in separate pairwise comparisons against the alternative, that the mean absolute PE of FLERO is less than that of BUII and Kane. The results yielded P values < 0.0001 for comparisons between FLERO and BUII, as well as FLERO and Kane, allowing for rejection of the null hypothesis, hence indicating that the absolute PEs of FLERO are significantly smaller for both comparisons. Furthermore, the method proposed by Holladay et al. for evaluating non-Gaussian PEs of SE associated with IOL calculators was also used and produced statistical results consistent with those of the Wilcoxon Signed-Rank test, further supporting the observed differences in PEs.

Clinical Significance

A distinction must be made between statistical and clinical significance. Although FLERO predictions differ significantly from both Kane and BUII, the key question is whether these differences alter clinical decision making. For cataract surgeons, the smallest increment that can meaningfully change IOL selection is the clinical quantum of refractive power, approximately 0.17 D at the spectacle plane. This derives from the fact that IOLs are generally manufactured in 0.50 D steps at the IOL plane, corresponding to 0.335 D at the spectacle plane; hence, the decision to round up or down between available lens powers effectively introduces a half-step quantum of approximately 0.17 D.

When FLERO and BUII predictions diverged by at least this quantum, clinically meaningful differences emerge. Of the 2363 eyes, 764 (32.3%) showed predicted postoperative refractions differing by $\geq$ 0.17 D – enough to potentially change the IOL selected. In this subset, FLERO achieved the smaller predictive error in 453 eyes (59.3%), whereas BUII performed better in 311 eyes (40.7%). Statistical testing with the Wilcoxon Signed-Rank test demonstrated that, in this subset, FLERO's absolute PEs were significantly lower than BUII's ($P < 0.05$). This finding was further confirmed using the method proposed by Holladay et al. for analyzing non-Gaussian PEs.[20]

translational vision science & technology

Downloaded from tvst.arvojournals.org on 05/04/2026

Taken together, these results show that, for most patients, FLERO and BUII converge on the same IOL choice. However, in a substantial minority of cases where predictions differ enough to influence selection, FLERO more often provides the more accurate refractive outcome—underscoring both its statistical and clinical relevance.

Discussion

Positioning FLERO Among AI and Hybrid IOL Power Calculator

Currently, three IOL power calculation formulas based purely on AI (Hill-RBF, Karmona, and Nallasamy) and five hybrid formulas that combine AI with theoretical models (FullMonte Method, Ladas Super Formula AI, PEARL-DGS, Kane, and Hoffer QST) have been published in the peer-reviewed literature.[18] The FLERO calculator positions itself as the sixth hybrid model, as it integrates a theoretical framework—the BUII formula—with a data-driven approach consisting of GA-based feature selection and a linear regression model.

Relative to existing AI-based and hybrid approaches, FLERO is set apart in three aspects: (1) the incorporation of novel FLACS-derived biometric features, (2) a two-step GA-based feature selection pipeline, and (3) the deliberate use of a simple, interpretable linear regression model.

To the best of our knowledge, this is the first study to explore FLACS-derived data for enhancing refractive outcome prediction (point 1). To address the high dimensionality of the FLACS data, we applied a two-step GA + BFE feature selection procedure that reduced more than 1800 candidate variables to 18 (point 2). Whereas GA-based feature selection has also been applied in the development of Hill-RBF 3.0,[3] its scope was limited to a small set of 13 biometric parameters, of which only 3 to 8 were ultimately retained. Finally, the modeling architecture itself is a defining element (point 3). The three AI-based formulae use complex, nonlinear models (e.g., radial basis function neural networks, stacked ensembles, or nonlinear regressors). Similarly, hybrid models, like PEARL-DGS, rely on gradient-boosted trees, support vector regression, or multilayer perceptrons. Although such architectures can capture complex nonlinear interactions, they often lack interpretability and risk overfitting without careful validation. In contrast, FLERO uses a parametric model that is a linear combination (linear regression model) of scaled variables. This structure is much more transparent, with coefficients representing the contribution of each (scaled) input to the predicted IOL power, enabling straightforward analysis and facilitating clinical understanding. As noted by Rudin (2019), in high-stakes domains like healthcare, the use of inherently interpretable models is crucial to ensure trust and accountability in decision making processes.

Assessment of FLERO Generalization

For all folds, nearly no difference between the train and test scores was observed across all metrics. This served as a strong indication that little to no overfitting occurred in the evaluation procedure. Additionally, learning curves for training and cross-validation folds showed significant overlap indicating a low likelihood of model overfitting (data not shown). Furthermore, examination of FLERO's feature weights revealed that the variation of each across the seven folds was minimal. This further supports FLERO's potential for generalizability and low risk of fold-specific overfitting, as overfitting would have resulted in high variability of feature weights across folds.

FLERO Predictive Feature Importance

The GA-BFE-selected feature set consisted of 18 predictors: IOL power; patient age; ZEISS IOLMaster measurements, including AL, steep and flat keratometry, and ACD; and CATALYS-derived OCT values, such as the central apex, depth values, and offset in the depth axis, along with derived values like anterior and posterior lens radii, lens thickness, lens apices, and corneal curvature.

It is notable that AL and IOL power exhibit the most substantial weights, both inversely related to refractive outcome prediction. Significantly, the BUII prediction held a positive weight in the model, indicating that the ridge regression aims to build upon and thus refine the BUII prediction. Indeed, all CATALYS-derived features (9 of the 18 model features) selected by the GA-BFE algorithm were significant at the 95% confidence interval (CI), highlighting their importance in refining the BUII prediction.[21]

Although the inclusion of age may seem out of place compared to anatomic biometry, it is worth noting that age is often used as a numerical predictor in similar formulae, such as ones predicting effective lens position.[22]

Interestingly, the crystalline lens meridian position (LMP) was not retained explicitly by the GA as a predictive feature, implying other combinations of features are at least as informative as this single metric, if not more so; however, the presence of key predic-

Downloaded from tvst.arvojournals.org on 05/04/2026

Table 7. Comparison of (i) BUII and Kane Formulae Performances in the Darcy Study (Upper Part of the Table), and (ii) BUII, Kane, and FLERO Formulae Performances With the Mercy Dataset (Lower Part of the Table) From Table 2

	ME	SD	MedAE	MAE	|E| < 0.25	|E| < 0.5	|E| < 1.0
Darcy study[23]							
BUII	0.000	0.505	0.314	0.390	0.417	0.707	0.947
Kane	0.000	0.490	0.302	0.377	0.426	0.720	0.952
Mercy data							
BUII	**0.000**	0.470	0.268	0.345	0.470	0.781	0.962
Kane	**0.000**	0.463	0.258	0.338	0.488	0.787	0.964
FLERO	0.001	**0.430**	**0.245**	**0.315**	**0.507**	**0.824**	**0.970**

The scores in bold represent the best score for each metric.

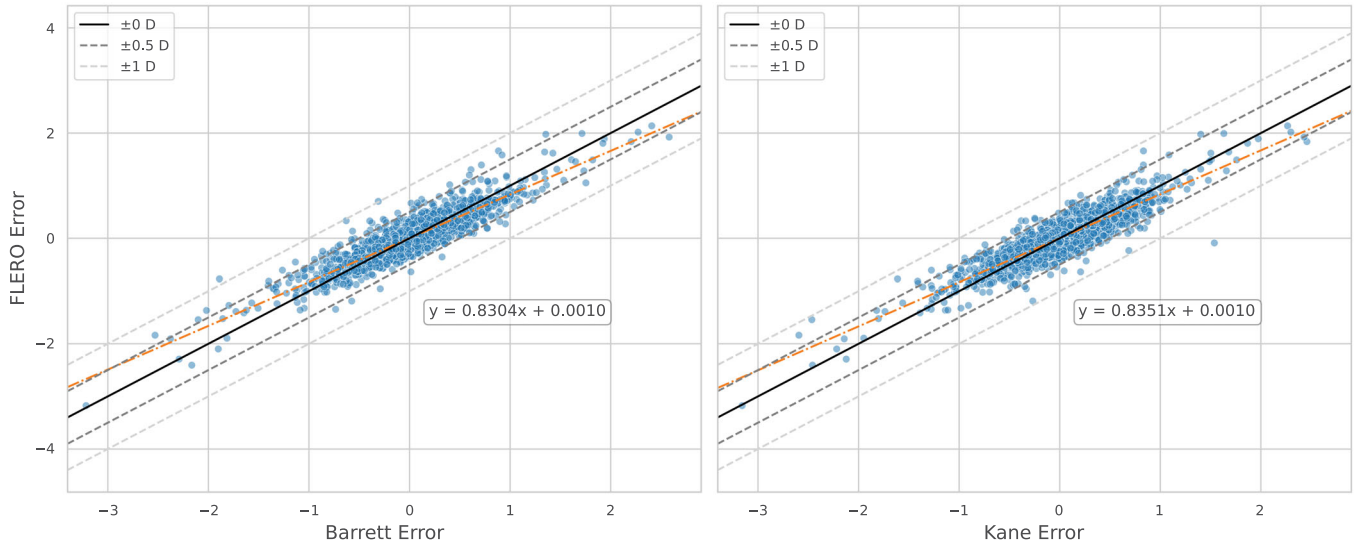

Figure. FLERO versus BUII (*left*) and FLERO versus Kane (*right*) PEs with corresponding linear regression fits and key dioptric error threshold overlays. BUII, Barrett Universal II; FLERO, femtosecond laser enhanced refractive outcome; PE, prediction error.

tors within the predictive features set that can be used to calculate LMP, for example, ALR, PLR, ALA, PLA, etc., suggest its implicit representation. Indeed, further analysis confirmed that LMP can be nearly perfectly reconstructed with an R^2 of 0.965 using a linear combination of the features chosen by the GA-BFE algorithm, which is particularly relevant given that FLERO itself is a linear combination of these input features.

FLERO Performance Versus Standard-of-Care Formulae

To further contextualize FLERO performance with BUII and Kane, the results for BUII and Kane models on the Mercy dataset were compared with one of the largest literature datasets known, the Darcy study (Table 7).[23] BUII and Kane performances were substantially worse in the Darcy study compared to their performance on the Mercy dataset.[23] This is an indication that neither formula performed especially poorly on the dataset on which the FLERO model was built.

To further assess FLERO's enhancement over the BUII and Kane formulae, a "prediction error analysis" was conducted. The Figure presents plots of FLERO PEs versus corresponding BUII (left) and Kane (right) PEs. Linear regression of these two plots (orange dashed lines) reveals a slope that is significantly less than 1 for both FLERO versus BUII (slope = 0.830, 95% CI = 0.815 to 0.846) and FLERO versus Kane (slope = 0.835, 95% CI = 0.829 to 0.851), indicating that, on average, FLERO results in more accurate

Downloaded from tvst.arvojournals.org on 05/04/2026

predictions (or a narrower distribution of PEs) of postoperative spherical equivalent compared to BUII and Kane. The implication of these slopes is that a ± 1.00 D error under the BUII or Kane models corresponds to ± 0.830D and ± 0.835D errors, respectively, under the FLERO model. This finding is consistent with the results of the Wilcoxon Signed-Rank test, which demonstrated statistically significant lower absolute PEs of FLERO compared to BUII and Kane. Furthermore, the visualizations in the Figure demonstrate that when predictions of FLERO versus BUII, and FLERO versus Kane are plotted, most predictions fall within ± 1 D (light gray dashed lines) of the observed refractive outcome and the paired models in each respective plot have strong positive correlations with each other.

Importantly, except for a single instance, FLERO never recommended an adjustment to the BUII prediction beyond 1 D. In this exceptional case, FLERO provided a prediction that was considerably closer to the observed outcome compared to the BUII prediction. Comparing FLERO to the Kane formula, only one eye required a significant correction (> 1 D), but again, the FLERO model made a prediction much closer to the observed outcome compared to Kane, reducing the error from 1.5 to almost 0 D. Analysis of the eyes with the largest absolute FLERO PEs showed no anomalous biometric values: the AL was between 21.64 and 25.20 (mean = 23.53), the ACD was between 2.39 and 3.93 (mean = 3.091), and no extremes in their CATALYS-derived biometrics were observed. Both BUII and Kane formulae demonstrated substantial, and in fact larger, PEs for the same eyes as well.

Study Limitations

One of the primary limitations of this study is the lack of external validation. All data came from a single hospital system and involved a limited number of surgeons, which may impact the generalizability of FLERO. Additionally, the large number of features considered in the feature selection step increases the potential risk of overfitting. To address these limitations and maximize the use of available data, this study used a robust feature selection and cross-validation strategy. Whereas all available indicators point to overfitting being absent during cross-validation, validation on independent, multi-center data is an essential next step to confirm the stability and generalizability of the model.

A minor potential limitation arises from the fact that the study does not differentiate between eyes coming from patients who had both eyes treated, and those who had only one eye treated. To prevent information leakage from patients contributing data from both eyes, we used a grouped cross-validation approach, ensuring that all eyes from a single patient were exclusively allocated to either the training or test set. A related concern is the potential correlation of bilateral eyes, because eyes of the same patient are likely not fully independent. To address this, we conducted a bootstrap analysis in which only one eye was randomly selected from each bilateral case. The results showed negligible differences in model performance (mean SD bootstrap = 0.4324 D ± 0.0050 D versus FLERO SD = 0.4297 D; Mann–Whitney U test P value = 0.4428), suggesting that potential correlation between bilateral eyes did not meaningfully influence the study findings. Additionally, there is a potential source of bias: patients who experienced a negative outcome with the first eye might opt not to return for the surgery on the second eye. However, in comparing unilateral to bilateral patients, the former demonstrated no greater errors than the latter as would be expected if patients with poor outcomes did not return for second-eye surgery (analysis not shown).

Another limitation relates to the ability to account for patients with a history of laser vision correction prior to their cataract surgery at Mercy. Whereas this limitation is not expected to advantage FLERO over BUII or Kane (as all PE comparisons were made pairwise between formulae), it is a factor that could be better controlled in future studies assuming availability of this pre-surgical patient context in future real-world datasets.

A minor limitation arises from using both Zeiss IOL Master 500 and 700 devices, as subtle differences in measurement accuracy or device-specific biases could theoretically influence results. In our dataset, only 6.55% of eyes were measured with the IOLMaster 500 (155 eyes vs. 2208 with the IOLMaster 700). A subset analysis comparing outcomes between device groups (data not shown) revealed no meaningful impact on the performance of the BUII, Kane, and FLERO formulas. Furthermore, because the IOL Master 500 collects fewer characteristics than the 700, the feature selection process was restricted to only those features available in both devices, meaning certain potentially predictive features unique to the IOLMaster 700 were not considered.

Yet another limitation of the study relates to the length and variability of the postoperative period prior to measurement of manifest refraction: only measurements of manifest refraction within the window of 15 to 120 days following surgery were used, corresponding to typical real-world postoperative follow-up periods observed in clinical practice. Although this period is considered long enough for a reasonable degree of lens

Downloaded from tvst.arvojournals.org on 05/04/2026

settling to occur,[24] strictly speaking, the model does not capture potential longer-term settling effects, for example, 1 year after surgery. FLERO performance was found not to differ significantly between eyes having a shorter postoperative window prior to measurement of refraction versus those having a longer window (analysis not shown). Correspondingly, this limitation is considered negligible.

A limitation relates directly to FLERO's deliberate dependence on the BUII prediction as an input feature. The rationale for this design is to leverage and refine the complex, domain-specific relationships encoded within BUII, which integrates Gaussian optics and paraxial ray tracing principles. This hybrid approach, however, carries the risk that systematic errors in BUII could propagate into FLERO, particularly in biometric subgroups where BUII is less accurate. This is partially mitigated by FLERO's model design, as it is an outcome-based model that is being trained explicitly for such mispredictions by exploiting novel FLACS features. In this way, errors can become opportunities for refinement rather than simple propagation, similar to boosting methods in machine learning. Nevertheless, external validation across diverse populations is essential to ensure FLERO consistently refines BUII's predictions rather than inheriting its limitations.

Finally, the dataset included a mix of spherical ($N = 2235$, 94.6%) and toric ($N = 128$, 5.4%) IOLs. Theoretically, spherical power IOL calculations should be equally applicable to both types. Indeed, the impact of (minor) toric IOL presence in the training set was assessed and found to be negligible (analysis not shown). Future research could explore whether a model specifically designed for toric IOLs could yield improved performance. Additionally, the study was restricted to J&J IOLs, potentially limiting the generalizability of the results to IOLs from other manufacturers.

Acknowledgments

Supported by Johnson & Johnson, with locations in Jacksonville, Florida and Milpitas, California, USA. The sponsor participated in the design of the study, conducting the study, data collection, data management, data analysis, interpretation of the data, and preparation, review, and approval of the manuscript.

Disclosure: **J. Van Der Donckt**, None; **J.A. Young**, Johnson and Johnson (C); **M. Rademaker**, None; **S. Menon**, (C); **C.-W. Chang**, Johnson & Johnson (E); **G. Vandewiele**, None; **B. Straker**, Johnson & Johnson (E); **D. Dewey**, Johnson & Johnson (E); **G. Dai**, Johnson & Johnson (E); **J. Gonzalez**, Johnson & Johnson (E); **J.R. Free**, Johnson & Johnson (E); **S. Van Hoecke**, None; **W. Scott**, Johnson and Johnson (C); **S. Tauber**, Johnson and Johnson (C); **H.B. Dick**, Johnson and Johnson (C); **C. Scales**, Johnson & Johnson (E)

References

1. Kirwan C, Nolan JM, Stack J, Moore TC, Beatty S. Determinants of patient satisfaction and function related to vision following cataract surgery in eyes with no visually consequential ocular comorbidity. *Graefes Arch Clin Exp Ophthalmol*. 2015;253:1735–1744.
2. Savini G, Di Maita M, Hoffer KJ, et al. Comparison of 13 formulas for IOL power calculation with measurements from partial coherence interferometry. *Br J Ophthalmol*. 2021;105:484–489.
3. Hill WE, Haehnle J. Hill-RBF: improving IOL power selection by artificial intelligence. In: *Essentials in Ophthalmology*. New York, NY: Springer; 2024:637–648.
4. Hill WE. IOL Power Selection by Pattern Recognition. *ASCRS EyeWorld Corporate Education*. 2016. Available at: https://crstodayeurope.com/articles/new-frontiers-in-iol-prediction-for-improved-refractive-outcomes/iol-power-selection-by-pattern-recognition/.
5. Connell BJ, Kane JX. Comparison of the Kane formula with existing formulas for intraocular lens power selection. *BMJ Open Ophthalmol*. 2019;4:e000251.
6. Liu Y, Wang X, Song L, et al. Advanced artificial-intelligence-based jiang formula for intraocular lens power in congenital ectopia lentis. *Transl Vis Sci Technol*. 2025;14:5.
7. Li T, Stein J, Nallasamy N. Evaluation of the Nallasamy formula: a stacking ensemble machine learning method for refraction prediction in cataract surgery. *Br J Ophthalmol*. 2023;107:1066–1071.
8. Moutari S, Moore JE. An ensemble-based approach for estimating personalized intraocular lens power. *Sci Rep*. 2021;11:22961.
9. Lou W, Zhou W, Wu M, Jin H. A new intraocular lens power formula integrating an artificial intelligence–powered estimation for effective lens position based on Chinese eyes. *Transl Vis Sci Technol*. 2024;13:40.
10. Hoffer KJ, Aramberri J, Haigis W, et al. Protocols for studies of intraocular lens formula

Downloaded from tvst.arvojournals.org on 05/04/2026

accuracy. *Am J Ophthalmol*. 2015;160:403–405. e401.
11. Shammas HJ, Shammas MC, Bahr C, Sahota R, Hall B. Impact of best corrected final visual acuity on the performance of intraocular lens power calculations. *Clin Ophthalmol*. 2025;19:1693–1697.
12. Khatib ZI, Haldipurkar SS, Shetty V, Dahake H, Nagvekar P, Kashelkar P. Comparison of three newer generation freely available intraocular lens power calculation formulae across all axial lengths. *Indian J Ophthalmol*. 2021;69:580–584.
13. Yasar I-I, Yasar S, Al Barri L, Mercea N, Munteanu M, Stanca HT. Comparison of traditional and AI-based methods: Barrett Universal II vs. Ladas Super Formula in IOL power calculation. *J Clin Med*. 2025;14:2023.
14. Stopyra W, Voytsekhivskyy O, Grzybowski A. Accuracy of 20 intraocular lens power calculation formulas in medium-long eyes. *Ophthalmol Ther*. 2024;13:1893–1907.
15. Stopyra W, Langenbucher A, Grzybowski A. Intraocular lens power calculation formulas—a systematic review. *Ophthalmol Ther*. 2023;12:2881–2902.
16. Melles RB, Holladay JT, Chang WJ. Accuracy of intraocular lens calculation formulas. *Ophthalmology*. 2018;125:169–178.
17. Melles RB, Kane JX, Olsen T, Chang WJ. Update on intraocular lens calculation formulas. *Ophthalmology*. 2019;126:1334–1335.
18. Stopyra W, Cooke DL, Grzybowski A. A review of intraocular lens power calculation formulas based on artificial intelligence. *J Clin Med*. 2024;13:498.
19. Hao Y, Fu J, Huang J, Chen D. Comparing the accuracy of intraocular lens power calculation formulas using artificial intelligence and traditional formulas in highly myopic patients: a meta-analysis. *Int Ophthalmol*. 2024;44:242.
20. Holladay JT, Wilcox RR, Koch DD, Wang L. Statistics of prediction error for non-Gaussian dependent and independent datasets. *J Cataract Refract Surg*. 2023;49:440–442.
21. Kim M, Eom Y, Lee H, et al. Use of the posterior/anterior corneal curvature radii ratio to improve the accuracy of intraocular lens power calculation: Eom's adjustment method. *Invest Ophthalmol Vis Sci*. 2018;59:1016–1024.
22. Zou M, Lian Z, Young CA, et al. Improving effective lens position prediction for transscleral fixation of intraocular lens among congenital ectopia lentis patients. *Am J Ophthalmol*. 2023;252:121–129.
23. Darcy K, Gunn D, Tavassoli S, Sparrow J, Kane JX. Assessment of the accuracy of new and updated intraocular lens power calculation formulas in 10 930 eyes from the UK National Health Service. *J Cataract Refract Surg*. 2020;46:2–7.
24. Charlesworth E, Alderson AJ, de JV, Elliott DB. When is refraction stable following routine cataract surgery? A systematic review and meta-analysis. *Ophthalmic Physiol Opt*. 2020;40:531–539.

Downloaded from tvst.arvojournals.org on 05/04/2026