Title: The long and short of it: a comprehensive assessment of axial length estimation in myopic eyes from ocular and demographic variables.

Running title: Estimating axial length for myopia management

Authors: Gareth Lingham[1,2,3], James Loughman[1,2], Davoud Shariat Panah[2,4], Siofra Harrington[1,5], Kathryn J. Saunders[6], Gui-Shuang Ying[7], Hongguang Cui[8], Emmanuel Kobia-Acquah[1], Daniel Ian Flitcroft[1,2,9]

1. Centre for Eye Research Ireland, Environmental Sustainability & Health Institute, Technological University Dublin, Dublin, Ireland

2. Ocumetra Ltd., Dublin, Ireland

3. Centre for Ophthalmology and Visual Science (incorporating the Lions Eye Institute), the University of Western Australia, Perth, Australia

4. School of Computer Science, Technological University Dublin, Dublin, Ireland

5. School of Physics and Clinical and Optometric Sciences, Technological University Dublin, Dublin, Ireland

6. Centre for Optometry and Vision Science, School of Biomedical Sciences, Ulster University, Coleraine, Northern Ireland

7. Center for Preventive Ophthalmology and Biostatistics, Department of Ophthalmology, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

8. The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, People's Republic of China

9. Department of Ophthalmology, Children's Health Ireland at Temple Street Hospital, Dublin, Ireland

Corresponding author:

Daniel Ian Flitcroft

ian.flitcroft@tudublin.ie

ABSTRACT

Background/Objectives

Axial length, a key measurement in myopia management, is not accessible in many settings. We aimed to develop and assess machine learning models to estimate the axial length of young myopic eyes.

Subjects/Methods

Linear regression, symbolic regression, gradient boosting and multilayer perceptron models were developed using age, sex, cycloplegic spherical equivalent refraction (SER) and corneal curvature. Training data were from 8135 (28% myopic) children and adolescents from Ireland, Northern Ireland and China. Model performance was tested on an additional 300 myopic individuals using traditional metrics alongside the estimated axial length vs age relationship. Linear regression and receiver operator characteristics (ROC) curves were used for statistical analysis. The contribution of the effective crystalline lens power to error in axial length estimation was calculated to define the latter's physiological limits.

Results

Axial length estimation models were applicable across all testing regions ($p \geq 0.96$ for training by testing region interaction). The linear regression model performed best based on agreement metrics (mean absolute error [MAE]=0.31mm, coefficient of repeatability=0.79mm) and a smooth, monotonic estimated axial length vs age relationship. This model was better at identifying high-risk eyes (axial length >98th centile) than SER alone (area under the curve 0.89 vs 0.79, respectively). Without knowing lens power, the calculated limits of axial length estimation were 0.30mm for MAE and 0.75mm for coefficient of repeatability.

Conclusions

In myopic eyes, we demonstrated superior axial length estimation with a linear regression model utilising age, sex and refractive metrics and showed its clinical utility as a risk stratification tool.

INTRODUCTION:

Myopia is typically caused by abnormal axial eye growth between childhood and young adulthood[1] and this excessive ocular elongation is strongly linked to irreversible vision loss among myopic adults.[2] This is one of the reasons why axial length has been proposed as the key metric for assessing myopia control treatment efficacy.[3]

To maximise accessibility for patients and efficiently utilise expertise and workforce, the management and treatment of myopia progression will likely take place in primary and community eye care settings.[4] However, unlike specialised and secondary eye care settings, access to biometric devices to measure axial length in primary and community care settings is limited. Recent guidelines and research publications have emphasised the value of axial length in managing myopia progression.[5–7] Thus, the inability to measure axial length is a potential barrier to the uptake of myopia control treatments and limits a clinician's ability to identify patients most in need of treatment.

Spherical equivalent refractive error (SER) and axial length are intrinsically linked, with SER related to the mismatch between the axial length and focal power of the eye, the latter being predominantly derived from the cornea and crystalline lens. Previous research has shown that axial length can be estimated from SER alone or a combination of SER and corneal curvature.[8–12] However, challenges in accurately estimating axial length from SER remain,[3,13,14] with published models[8–11] generally performing below a clinically desirable level and failing to account for the impact of age and sex on axial length,[11] potentially non-linear associations[13,15] and alterations in the axial length to SER relationship that occur with age.[11,16]

Machine learning aims to develop predictive models from data and spans a range of techniques from linear regression to more complex deep learning models.[17] The full potential for machine learning techniques to improve axial length estimation models has not been explored. We aimed to: 1) develop and compare new machine learning models using age, sex, SER and corneal power to estimate axial length, 2) compare these models to published algorithms and evaluate their application in the clinical management of myopia and 3) explore the physiological limits of axial length estimation, using the known biological sources of error in the estimation of axial length from refractive variables.

Subjects and Methods

Model development

Four machine learning approaches were used to develop new models for estimating axial length: linear regression, symbolic regression, gradient boosting and multilayer perceptron. Multilayer perceptron is a form of artificial neural network, while gradient boosting is an ensemble method that combines a series of weak machine learning models. Neither model produces a defined equation. Symbolic regression iteratively searches mathematical relationships to determine the simplest equation for accurately describing relationships within a dataset. For each model, input variables were cycloplegic SER (dioptres), average corneal radius of curvature (mm), biological sex, age (years) and, due to the known logarithmic relationship between axial length and age,[18] the base 10 logarithm of age. Corneal curvature was included as widely available autorefractors measure both SER and corneal curvature.

The standard multiple linear regression model was built using R (R Foundation for Statistical Computing, Austria) and interaction terms were tested and retained if they improved prediction performance on validation data (see below). Symbolic regression, gradient boosting and multilayer perceptron models were implemented using the Scikit-learn Python library.[19] For the gradient boosting model, the number of estimator parameters was fixed at 600 and learning rate, maximum depth and sub-sampling fraction parameters were fixed at 0.02, 2, and 0.2, respectively. For the multilayer perceptron model, the maximum number of iterations was fixed at 1000 and the logistic function used as the activation function for hidden layers. Symbolic regression was implemented using the GP-GOMEA library[20] with the number of generations fixed at 10.

The four newly developed machine learning models were compared to published algorithms, which used either the Gullstrand schematic eye[10] or least-squared techniques such as simple[21] or multiple linear regression. [8,9,11,12] No previous algorithms used sex, while only Queirós et al. included age.[12]

Training data

Training data were derived from three epidemiological studies, the Ireland Eye Study,[22] the Northern Ireland Childhood Errors of Refraction (NICER) study[23,24] and a refractive error study from the Zhejiang province of China.[25] The Ireland Eye Study (IES) included a population-based sample of 1626 children aged 6-7 or 12-13 years. The NICER study is a population-based cohort study of children in Northern Ireland with data from 2 424 examinations of 1 068 children aged 6-20 years. The Zhejiang refractive error survey was a school-based cross-sectional study of 3 429 participants aged 5-18 years. The training data additionally included data from myopic children aged 6-19 years enrolled in studies, including myopia control clinical trials, at the Centre for Eye Research Ireland (CERI; 806 visits from 240 participants).[26] Participant region was defined as either the island of Ireland (referred to as Ireland) or China. Across all studies, informed consent was obtained from legal guardians, SER was measured using cycloplegic autorefraction and axial length, corneal curvature and anterior chamber depth (ACD) measured using an ocular biometer (details published elsewhere[22–27]). For model training, the training data were randomly divided, with 95% used for model training and the remaining 5% for validation, where validation data were used for iterative assessment of model performance during training to avoid overfitting.

Testing data

Final models were tested on myopic participants (SER ≤-0.50 dioptres) only, using a separate, randomly selected subset of 150 participants of the cross-sectional Zhejiang study and 150 participants of the CERI clinical trial, the Myopia Outcome Study of Atropine in Children (MOSAIC; ISRCTN: 36732601). The myopic testing population reflects the proposed clinical application of an axial length estimation tool, which is to assess and monitor the axial length of myopic children and adolescents. MOSAIC participants were enrolled in a randomised placebo-controlled clinical trial of nightly atropine 0.01% eye drops and data were available at baseline and the 12-month follow-up. At the 12-month follow-up, 2/3 of MOSAIC participants were using atropine 0.01% eye drops, which could reflect a clinical setting.

Applicability of models

To investigate whether models trained on European data could be applied to Asian populations and vice versa, models were trained using either all data combined, or data from Ireland or China alone. The performance of these models was then assessed, separately, using testing data from either Ireland, China or both.

Model performance

We used a mixture of traditional and non-traditional statistics for assessing model performance:

1. Agreement between estimated and actual axial length, including 12-month change.
2. Ability to identify high-risk eyes with long axial length or fast axial length growth
3. Monotonicity and smoothness of axial length estimations as a function of age.

Agreement between estimated and actual axial length

Agreement between estimated and actual axial length was assessed by analysing the mean bias and standard deviation of the prediction errors, the mean absolute error (MAE), the concordance correlation coefficient,[28] which measures the correlation along a line of perfect agreement, and the coefficient of repeatability. The coefficient of repeatability indicates the interval around the estimated axial length in which the true measurement would lie with 95% confidence and was calculated as 2.77 × within-subject standard deviation.[29] Change in axial length was estimated by subtracting the estimated axial length at baseline from that at the 12-month follow-up.

Identification of high-risk eyes

High-risk eyes were defined as eyes with axial length in the top 2% for the participant's age and sex. To calculate this threshold, the 98th centiles of axial length were extracted from two reports of European participants[18,30] and a weighted cubic spline fitted to these data to generate age- and sex-specific 98th centiles (Supplementary Figure 1). Fast axial length progression was defined as >0.20mm/year based on the mean axial length growth among myopic eyes in the NICER study.[31]

Monotonicity and smoothness

An estimated axial length that, in the absence of corneal or SER changes, decreases or varies considerably with increasing age is not clinically plausible. We, therefore, examined the monotonicity and smoothness of axial length estimations as a function of age by creating curves in which axial length was estimated for each year of age between 6 and 20 years, while all other variables were held constant. The first derivative (rate of change) of the age curve was calculated, and monotonicity was assessed by testing whether the first derivatives were all ≥0 (a negative value would indicate decreasing axial length with age) and smoothness was evaluated by comparing deviations of the first derivate from a fitted smooth spline curve.

Model extension

Anterior chamber depth (ACD) alters effective lens power and should, therefore, account for some of the variance in axial length not explained by SER or corneal power. We investigated whether adding ACD improved axial length estimation. ACD was either estimated from known relationships with age,[32] to determine whether explicitly defining this relationship in the model could improve its performance, or actual ACD measured from biometry was used.

Impact of sex, age and spherical equivalent on axial length estimation

To illustrate the benefits of including sex in the algorithm, we trained additional models using either female or male training data and tested these models on either males, females or both sexes combined in the myopic testing population. The influence of age and SER on axial length estimation was assessed by quantile regression using axial length estimation error as the outcome.

Limits of axial length estimation

Errors in axial length estimation largely arise from the unmeasured effective crystalline lens power. When the true axial length, corneal power and SER are known, the refractive power of the internal ocular structures (internal refractive power; IRP) can be calculated and, by comparing against age- and sex-matched emmetropic eyes, the contribution of the IRP to SER calculated.[33] Assuming IRP and corneal power are independent, the variance of the IRP contribution to SER can be used to explore the expected variance in axial length estimation error arising from IRP alone.

Spectacle refraction

As cycloplegic autorefraction is not always feasible, we extended the axial length estimating model to work with presenting spectacle refraction by introducing an intermediate step that uses multiple linear regression to estimate cycloplegic autorefraction from spectacle refraction and age. Only data from CERI clinical trials contained both presenting spectacle refraction and cycloplegic autorefraction. Undercorrected participants were excluded by removing examinations with presenting corrected visual acuity >0.1 logMAR.

Statistical analysis

Only right eye data were used, tests are two-sided and significance was set at $p<0.05$. Model performance metrics were compared using linear regression, adjusting for the model type. The effects of training region, testing region and including ACD on model performance were assessed by adding these as covariates. A linear mixed model was used to estimate cycloplegic autorefraction SER from spectacle refraction using random intercept terms to account for any within-subject correlations. Performance metrics were calculated within Python (v3.10.4, Python Software

Foundation, USA), and linear and quantile regression was performed with R (v4.3.0, R Foundation for Statistical Computing, Austria).

RESULTS

Population characteristics

Table 1 provides an overview of participants characteristics in the testing and training data sets. In the training data, participants from China were younger on average, but otherwise had similar characteristics to participants from Ireland. As expected, participants in the testing data were more myopic than the training data.

Applicability of models

Table 2 shows the performance of the four machine learning models when trained and then tested on the combined data or from Ireland or China, separately. While there were significant main effects of model training and testing regions on linear regression, the effect of training region on all model metrics did not vary significantly across testing region and vice versa (interaction for all $p \geq 0.97$).

Model performance

We compared the performance of these newly developed axial length estimation models against previously published algorithms, when applied to the myopic testing data, at either a single visit (cross-sectional, Figure 1a) or when estimating the 12-month change in axial length (longitudinal [Irish data only], Figure 1b). Due to the number of algorithms, data are presented as scatter plots of agreement metrics from Table 2 (all regions), with each plot including a measure of agreement (concordance correlation coefficient, mean bias) and a measure of the spread of estimation errors (coefficient of repeatability, standard deviation of errors), which are often related, on each axis. A good axial length estimator would maximise exact agreement while minimising the spread of errors. Compared to cross-sectional performance, the concordance correlation coefficient for estimation of axial length change was poorer across all models, possibly because rounding of SER to the nearest

0.125 dioptre results in a proportionally larger loss of information. The newly developed models outperformed all previously published models, with the linear regression model performing best or equal best across both cross-sectional and longitudinal axial length estimation.

Smoothness and monotonicity

We next assessed the modelled relationship between estimated axial length and age (Figure 2). The gradient boosting model was monotonic, but not smooth, explaining the poor longitudinal performance (Figure 1b). The multilayer perceptron model was smooth, but not monotonic, predicting that highly myopic eyes of young adult females shorten over time. The linear and symbolic regression models fulfilled the pre-specified criteria of monotonicity and smoothness, indicating the modelled axial length vs age relationship was clinically plausible.

Figure 3 shows a clinically relevant assessment of the top performing linear regression model (Equation 1). Prediction errors were approximately normally distributed around a small mean bias. Despite a reasonable level of uncertainty surrounding estimated axial length, 86% of estimations were within ±0.50mm of the actual axial length for cross-sectional data and 82% were within ±0.10mm on longitudinal testing. The linear regression model was better than SER alone at identifying high-risk eyes with axial length in the top 2% for age and sex (area under the curve [AUC] 0.89 vs 0.79, respectively), correctly identifying 99/121 (82%) high-risk eyes compared to only 88/121 (73%) using the optimum SER threshold of -3.37 D. However, the linear regression model was not better than SER at identifying eyes with fast axial length progression >0.2mm/year (Figure 3).

$$AL = 5.472 - 0.069 \times age + 3.060 \times \log_{10}(age) - 0.265 \times sex -$$

$$0.135 \times SER + 2.019 \times CR - 0.072 \times SER \times CR$$

Equation 1 Top-performing linear regression equation where AL is the axial length in mm, age is in years, sex=0 for males and 1 for females, SER is spherical equivalent refractive error in dioptres, and CR is the corneal radius of curvature in mm.

Model extension with anterior chamber depth

We investigated whether model performance could be improved by including either an estimated ACD, derived from age, or actual ACD. The performance metrics of the original and ACD-extended models are shown in Supplementary Table 1 and, for the linear regression model, in Figure 3. Compared to models without ACD, adding actual ACD significantly improved axial length estimation (MAE difference=-0.03mm, p=0.001; coefficient of repeatability difference=-0.08mm, p<0.001), whereas adding estimated ACD did not (MAE difference=-0.001, p=0.99; coefficient of repeatability difference=0.003mm, p=0.96).

Model performance by sex, age and spherical equivalent

Compared to the model in Equation 1, the performance of linear regression models were worse when trained using either only males or females and then tested on participants of the opposite sex (Supplementary Table 2). Models trained on male data overestimated female axial length (mean bias=-0.19mm) and models trained on female data underestimated male axial length (mean bias=0.33mm).

We used quantile regression to assess whether the linear regression model performance varied with age or SER (Supplementary Table 3, Supplementary Figure 2). Quantile slopes were positive at the 2.5% and 50% centile for SER and the 2.5% centile for age, indicating median bias tended to become more positive with decreasing myopia and that axial length estimation was less precise for highly myopic (<-8 D) and older participants.

Limits to axial length estimation

In the myopic testing population, the distribution of the IRP contribution to SER was Gaussian with a standard deviation of 0.91 dioptres. From this distribution (mean shifted to 0), we randomly sampled 10,000 observations (referred to as IRP error) and calculated the estimated axial length with and without IRP error added to the SER parameter, for ages 6-20 years, males and females, separately, and corneal curvature held constant at 7.76mm. The difference between the two estimations represents the axial length estimation error attributable to the variance in IRP around the population mean. From this, the minimum obtainable coefficient of repeatability was ±0.75mm, and MAE was 0.30mm, comparable to the linear regression model's ±0.80mm and 0.31mm, respectively.

Spectacle refraction

Of the 806 CERI clinical trial training data, 731 (90.7%) examinations had corrected visual acuity of 0.10 logMAR or better and were used to model cycloplegic autorefraction from spectacle refraction with a linear mixed model (Equation 2).

$$SER_{cyclo\ autoref} = -0.153 + 0.955 \times SER_{spectacle} - 0.028 \times age - 0.013 \times SER_{spectacle} \times age$$

Equation 2 Formula for estimating cycloplegic autorefraction (cyclo autoref) spherical equivalent from spectacle spherical equivalent (D) and age (years).

Compared to spectacle refraction alone, the formula in Equation 2 improved the estimation of cycloplegic autorefraction SER in the training dataset (MAE improved from 0.61 D to 0.45 D; concordance correlation coefficient improved from 0.931 to 0.945). This allows an estimate of axial length in the absence of cycloplegic autorefraction, where the measured cycloplegic autorefraction SER in Equation 1 is replaced by the estimated cycloplegic autorefraction SER output by Equation 2. Tested on Irish data alone, axial length estimated from spectacle refraction performed similarly to that estimated using actual cycloplegic autorefraction SER (the latter shown in Table 2; MAE: 0.34 vs 0.34mm, respectively, concordance correlation coefficient: 0.91 vs 0.91, respectively, and coefficient of repeatability: 0.83 vs 0.84mmm respectively) and was better at identifying long eyes compared to using spectacle SER alone (AUC 0.92 vs 0.81, respectively; Supplementary Figure 3).

DISCUSSION

We investigated and compared four machine learning methods for estimating axial length from ocular and demographic variables. When tested on myopic eyes, the newly developed models all performed better than previously published algorithms for axial length estimation. Interestingly, standard multiple linear regression was the superior model, performing best on cross-sectional and longitudinal data and meeting pre-specified criteria relevant to the clinical monitoring of axial length change.

Machine learning model performance typically decreases when applied to new populations, known as data or dataset shift[34,35]; however, in this study we found that training region effects on performance did not vary with testing region and that models were therefore applicable across populations. This is, perhaps, unsurprising as axial length estimators are modelling optical relationships, which are independent of populations. The reason why models tested on Chinese data performed better is unclear, but could be related to differences in the age, SER or crystalline lens power distributions.[36] Models trained using participants of only one sex (e.g. females) had a high mean bias and a higher MAE when applied to the opposite sex (e.g. males), illustrating the importance of incorporating sex into axial length estimation models. The linear regression algorithm became slightly less precise with increasing age and high myopia.

Compared to previously published algorithms, the enhanced performance of the machine learning models in this study is likely the result of the additional inclusion of age and sex as input variables and the large, diverse training data set. Interestingly, the multiple linear regression model outperformed the more complex deep learning models, indicating this is an effective method for developing axial length estimation models. Adding ACD to models improved the estimation of axial length, reducing the coefficient of repeatability by 0.08mm. While not commonly measured in primary and community eye care settings, methods to non-biometrically measure ACD exist[37] and may become more accessible in the future. Adding lens thickness will further improve axial length estimation; however, clinicians are unlikely to have access to measures of lens thickness, but not axial length. Other input variables could potentially be added to improve axial length estimation, for example, axial length can be estimated from fundus photographs,[38,39] but the performance of these convolutional neural networks (MAE=0.74-0.90mm) was much poorer than models in the present study (MAE=0.32 to 0.41mm).

Interestingly, adding ACD did not improve the estimation of axial length change (Figure 3). This is likely because the axial length estimation errors, which largely arise from the ACD and crystalline lens, are highly correlated at baseline and 12 months (Pearson r=0.90 in testing data). Thus, subtracting the estimated axial length at baseline from 12 months cancels out the ACD contribution to estimation error at these two visits.

The optimal linear regression model had a mean bias close to zero and a coefficient of repeatability of ±0.80mm. While this coefficient of repeatability interval may seem large, it is approaching the physiological limits of axial length prediction (±0.75mm), without incorporating effective lens power. The ability of the linear regression model to identify a long eye (axial length in the top 2% for age and sex) was much better than SER alone, representing a clinically useful application for identifying eyes at higher risk of future vision loss. Likewise, the linear regression model could estimate the axial length change over 12-months; however, it was not better than SER at identifying fast axial eye growth. Previous authors have noted the difficulty in using axial length ‘calculators’ to obtain appropriately sensitive measures of axial length change.[13]

Strengths of this study include the large sample sizes, comprehensive assessment of multiple machine learning models and input variables, a focus on the clinical application of the tool and the in-depth exploration of the expected limits of axial length estimation. Limitations of this study generally relate to the application of its findings. Although models were trained using all refractive errors, axial length estimation was only validated for myopic eyes, and our results may be less applicable to non-myopic eyes. This could also be considered a strength given the likely clinical application of an axial length estimator. Our results are similarly limited to ages 6-20 years and model performance was less precise among highly myopic, older eyes. Extending the top-performing linear regression to older ages generated implausible values (Supplementary Figure 4). Thus, further work is needed to extend axial length estimation to older ages, but we have laid groundwork for such an extension. To maximise data availability, we used keratometry measured by optical biometers, rather than from autorefractors; however, strong agreement between these devices means our results should be valid across both.[40,41] The algorithm also requires cycloplegic SER, which can be time-consuming and invasive to measure; however, we showed that extending the model to use spectacle refraction, by adding an intermediate conversion step, produced good results for axial length estimation among the Irish testing data and was better than spectacle SER alone for identifying eyes with a long axial length. Such an intermediate step might also work for estimating axial length from non-cycloplegic autorefraction[42]; although, the increased variance of the internal refractive power contribution to SER means axial length estimation would theoretically be less precise than with cycloplegic autorefraction.

Conclusion

This study demonstrated that the inclusion of age and sex significantly improves axial length estimation, and accuracy is further enhanced by adding anterior chamber depth. Multiple linear regression outperformed more complicated machine learning models, was applicable across both populations, was better than SER at identifying myopic eyes with long, high-risk axial length and approached the physiological limits of axial length estimation without incorporating lens power. In the absence of a biometer, the estimation of axial length from ocular and demographic variables provides a useful tool in enabling the clinical management of myopia.

ACKNOWLEDGEMENTS

COMPETING INTERESTS

GL and DSP are employees of Ocumetra, and JL and DIF are co-founders of Ocumetra, a company providing data analytic tools to assist with the clinical management, including an axial length estimation tool. JL is a consultant/contractor for Dopavision, Topcon, EssilorLuxottica and Ebiga Vision and has received funding from Topcon, Ocumension, Kubota Vision, EssilorLuxottica, Vyluma, Dopavision and Coopervision, all in the area of myopia management. DIF is a consultant/contractor for Vyluma, Coopervision, Essilor, Thea, Ocumension and Johnson & Johnson and has received funding from Topcon, Ocumension and Coopervision in the area of myopia control. KJS is in receipt of research funding from HOYA Vision and Vyluma in the area of myopia management.

FUNDING: None received

DATA AVAILABILITY: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

REFERENCES

1. Flitcroft DI. Emmetropisation and the aetiology of refractive errors. *Eye (Lond)* **28**, 169–179 (2014)

2. Tideman JWL, Snabel MCC, Tedja MS, Van Rijn GA, Wong KT, Kuijpers RAM, et al. Association of axial length with risk of uncorrectable visual impairment for Europeans with myopia. *JAMA Ophthalmol* **134**(12), 1355–1363 (2016).

3. Brennan NA, Toubouti YM, Cheng X, Bullimore MA. Efficacy in myopia control. *Prog Retin Eye Res* **83**, 100923 (2020).

4. Lingham G, Loughman J, Kuzmenko S, Biba M, Flitcroft DI. Will treating progressive myopia overwhelm the eye care workforce? A workforce modelling study. *Ophthalmic Physiol Opt* **42**(5), 1092–1102 (2022).

5. The College of Optometrists.Consensus Statement- Myopia Management.; 2019.

6. Gifford KL, Richdale K, Kang P, Aller TA, Lam CS, Liu YM, et al. IMI – Clinical management guidelines report. *Invest Ophthalmol Vis Sci* **60**(3), M184–M203 (2019).

7. World Council of Optometry. The Standard of Care for Myopia Management by Optometrists.; 2021.

8. Morgan PB, McCullough SJ, Saunders KJ. Estimation of ocular axial length from conventional optometric measures. *Cont Lens Anterior Eye* **43**(1), 18–20 (2020).

9. Dutt DDCS, Yazar S, Charng J, Mackey DA, Chen FK, Sampson DM. Correcting magnification error in foveal avascular zone area measurements of optical coherence tomography angiography images with estimated axial length. *Eye Vis* **9**(1), (2022).

10. Kim HS, Yu DS, Cho HG, Moon BY, Kim SY. Comparison of predicted and measured axial length for ophthalmic lens design. *PLoS One* **14**(1), e0210387 (2019).

11. He X, Zou H, Lu L, Zhao R, Zhao H, Li Q, et al. Axial Length/Corneal Radius Ratio: Association with Refractive State and Role on Myopia Detection Combined with Visual Acuity in Chinese Schoolchildren. *PLoS One* **10**(2), 111766 (2015).

12. Queirós A, Amorim-de-Sousa A, Fernandes P, Ribeiro-Queirós MS, Villa-Collar C, González-Méijome JM. Mathematical Estimation of Axial Length Increment in the Control of Myopia Progression. *J Clin Med* **11**(20), 6200 (2022).

13. Jong M, Sankaridurg P, Naduvilath TJ, Li W, He M. The relationship between progression in axial length/corneal radius of curvature ratio and spherical equivalent refractive error in myopia. *Optom Vis Sci* **95**(10), 921–929 (2018).

14. Galvis V, Tello A, Rey JJ, Serrano Gomez S, Prada AM. Estimation of ocular axial length with optometric parameters is not accurate. Cont Lens Anterior Eye **45**(3), 101448 (2022).

15. Cruickshank FE, Logan NS. Optical 'dampening' of the refractive error to axial length ratio: implications for outcome measures in myopia control studies. *Ophthalmic Physiol Opt* **38**(3), 290–297 (2018).

16. Ip JM, Huynh SC, Kifley A, Rose KA, Morgan IG, Varma R, et al. Variation of the Contribution from Axial Length and Other Oculometric Parameters to Refraction by Age and Ethnicity. *Invest Ophthalmol Vis Sci* **48**(10), 4846–4853 (2007).

17. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* **23**(1), 40–55 (2022).

18. Tideman JWL, Polling JR, Vingerling JR, Jaddoe VWV, Williams C, Guggenheim JA, et al. Axial length growth and the risk of developing myopia in European children. *Acta Ophthalmol* **96**(3), 301–309 (2018).

19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825–2830 (2011).

20. Virgolin M, Alderliesten T, Witteveen C, Bosman PAN. Improving Model-Based Genetic Programming for Symbolic Regression of Small Expressions. *Evol Comput* **29**(2), 211–237 (2021).

21. Grosvenor T. The Inaugural Alan Bott Memorial Lecture: Twenty years of NZOVRF-supported scientific inquiry. *Clin Exp Optom* **88**(6), 415–419 (2005).

22. Harrington SC, Stack J, Saunders K, O'Dwyer V. Refractive error and visual impairment in Ireland schoolchildren. *Br J Ophthalmol* **103**(8), 1112–1118 (2019).

23. McCullough SJ, O'Donoghue L, Saunders KJ. Six Year Refractive Change among White Children and Young Adults: Evidence for Significant Increase in Myopia among White UK Children. *PLoS One* **11**(1), e0146332 (2016).

24. Breslin KMM, O’Donoghue L, Saunders KJ. A Prospective Study of Spherical Refractive Error and Ocular Components Among Northern Irish Schoolchildren (The NICER Study). *Invest Opthalmol Vis Sci* **54**(7): 4843–4850 (2013).

25. Zhao E, Wang X, Zhang H, Zhao E, Wang J, Yang Y, et al. Ocular biometrics and uncorrected visual acuity for detecting myopia in Chinese school students. *Sci Rep* **12**(1), 18644 (2022).

26. McCrann S, Flitcroft I, Strang NC, Saunders KJ, Logan NS, Lee SS, et al. Myopia outcome study of atropine in children (MOSAIC): an investigator-led, double-masked, placebo-controlled, randomised clinical trial protocol. *HRB Open Res* **2**(15), 1–21 (2019).

27. Harrington SC, O’Dwyer V. Ocular biometry, refraction and time spent outdoors during daylight in Irish schoolchildren. *Clin Exp Optom* **103**(2), 167–176 (2020).

28. Lin LI-K. A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* **45**(1), 255–268 (1989).

29. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek ALM. Smallest Real Difference, a Link between Reproducibility and Responsiveness. *Qual Life Res* **10**(7), 571–578 (2001).

30. Truckenbrod C, Meigen C, Brandt M, Vogel M, Sanz Diez P, Wahl S, et al. Longitudinal analysis of axial length growth in a German cohort of healthy children and adolescents. *Ophthalmic Physiol Opt* **41**(3), 532–540 (2021).

31. Saunders KJ, McCullough SJ. Normative data for emmetropic and myopic eye growth in childhood. *Ophthalmic Physiol Opt* **41**(6), 1382–1383 (2021).

32. Jones LA, Mitchell GL, Mutti DO, Hayes JR, Moeschberger ML, Zadnik K. Comparison of Ocular Component Growth Curves among Refractive Error Groups in Children. *Invest Ophthalmol Vis Sci* **46**(7), 2317–2327 (2005).

33. Flitcroft I, Lingham G, Acquah EK, Loughman J. The Refractive Mechanism Map: application of a biometric definition of refractive error in myopia research. In: International Myopia Conference. Rotterdam; 2022. Available at: https://www.internationalmyopiaconference.org/abstract-book/.

34. Sahiner B, Chen W, Samala RK, Petrick N. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol* (2023). https://doi.org/10.1259/bjr.20220878.

35. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The Clinician and Dataset Shift in Artificial Intelligence. *N Engl J Med* **385**(3), 283–286 (2021).

36. Twelker JD, Mitchell GL, Messer DH, Bhakta R, Jones LA, Mutti DO, et al. Children's Ocular Components and Age, Gender, and Ethnicity. *Optom Vis Sci* **86**(8), 918–935 (2009).

37. Zamir E, Kong GYX, Kowalski T, Coote M, Ang GS. A novel method of quantitative anterior chamber depth estimation using temporal perpendicular digital photography. *Transl Vis Sci Technol* **5**(4), 10 (2016).

38. Jeong Y, Lee B, Han J-H, Oh J. Ocular Axial Length Prediction Based on Visual Interpretation of Retinal Fundus Images via Deep Neural Network. *IEEE J Sel Top Quantum Electron* **27**(4), 1–7 (2021).

39. Oh R, Lee EK, Bae K, Park UC, Yu HG, Yoon CK. Deep Learning-based Prediction of Axial Length Using Ultra-widefield Fundus Photography. *Korean J Ophthalmol* **37**(2), 95–104 (2023).

40. Gill ES, Le C, Joseph J, Munir WM. Comparison of Preoperative Parameter Measurements Using an Optical Biometer, Automated Keratometer, and a Placido-Based Topographer Module. *Eye Cont Lens* **46**(1), 46–51 (2020).

41. Huang J, Savini G, Su B, Zhu R, Feng Y, Lin S, et al. Comparison of keratometry and white-to-white measurements obtained by Lenstar with those obtained by autokeratometry and corneal topography. *Cont Lens Anterior Eye* **38**(5), 363–367 (2015).

42. Du B, Wang Q, Luo Y, Jin N, Rong H, Wang X, et al. Prediction of spherical equivalent difference before and after cycloplegia in school-age children with machine learning algorithms. *Front Public Health* **11** (2023). https://doi.org/10.3389/fpubh.2023.1096330

Figure 1 The performance of the four newly developed machine learning models (linear regression, gradient boosting, multilayer perceptron, symbolic regression; shown in colour), as well as published axial length prediction algorithms (no colour). Absolute mean bias is the absolute of the mean difference between estimated and actual axial length. Better performance is indicated by a higher concordance correlation coefficient and a lower value for all other metrics shown. Due to the high correlation between the mean absolute error and the coefficient of repeatability (Pearson correlation=0.99), only the latter is presented. Models were tested on either cross-sectional axial length data (top; combined Irish and Chinese data) or longitudinal 12-month change in axial length data (bottom; Irish data only).

Figure 2 Relationships between predicted axial length and age in the four machine learning models. The top row shows the predicted axial length at each age for a female with corneal radius of curvature of 7.76mm and with spherical equivalent refraction (SER) of +3 dioptres (D), plano, -3 D and -6 D. The bottom row shows the rate of change (1st derivative) of the plano curve (blue points) compared to a best-fit spline curve (green line). The sum of the distances of all points from the best-fit line are used as a measure of smoothness (uses +3 D, plano, -3 D and -6 D curves; linear regression=0.19, gradient boosting=1.78, multilayer perceptron=0.19, symbolic regression=0.20). An age curve is monotonic when all points for all age curves (+3 D, plano, -3 D and -6 D) lie on or above 0 (red dashed line). All models were monotonic except the multilayer perceptron model, which had 12.5% of first derivative points below 0.

Figure 3 Performance of linear regression models as assessed by (a) agreement between predicted and actual axial length at baseline using a Bland-Altman plot with mean bias and 95% limits of agreement (LOA) shown; (b), ability of models to identify participants with axial length in the top 2% for age and sex using a receiver operating characteristic (ROC) and area under the curve (AUC; AXLest is the AUC for the machine learning model, SER is using spherical equivalent refractive error alone); (c), agreement between predicted and 12-month change in axial length of myopic eyes (Irish data only) using Bland-Altman plot and statistics; (d), ability of models to identify participants (Irish data only) with 12-month axial length change greater than 0.2mm/year using a receiver operating characteristic (ROC) and area under the curve. Linear regression models either (left column) do not use anterior chamber depth (ACD; middle column), ACD as estimated from age and spherical equivalent or (right column) actual ACD measured by biometry.

Supplementary Figure 1 Data on 98th axial length centile thresholds extracted from the LIFE child study and used to define eyes with a high-risk axial length. Points show the actual reported data points and the lines show the smoothing cubic spline fit.

Supplementary Figure 2 Scatter plot of axial length estimation error over age (left) and spherical equivalent (right). Quantile regression slopes are shown for the 2.5% and 97.5% (dashed line), 25% and 75% (dotted line) and 50% quantiles (solid line). Slopes are derived from multivariable quantile regression including age and sex as explanatory variables. Slope positions on the left plot are calculated assuming spherical equivalent of -3.12 dioptres (D) and slopes on the right are plotted assuming an age of 12.5 years (both mean values of the testing data set).

Supplementary Figure 3 Performance of the top-performing linear regression model when applied to spectacle refraction data using Equation 2, using (top) a Bland-Altman plot with mean bias and 95% limits of agreement (LOA) shown, and (bottom), ability of models to identify participants with axial length in the top 2% for age and sex using a receiver operating characteristic (ROC) and area under the curve (AUC; AXLest is the AUC for the machine learning model, SER is using spherical equivalent refractive error alone). The left-hand column is cross-sectional data, while the right-hand column is longitudinal data. Testing data in all plots is from the Irish cohort alone.

Supplementary Figure 4 Examples of the estimated axial length vs age function for females when spherical equivalent and corneal radius are held constant at 0 D and 7.76mm, respectively, within the age range investigated in the present study (left) and for ages up to 80 years (right)