

## REVIEW ARTICLE



# Myopia prediction: a systematic review

Xiaotong Han<sup>1</sup>, Chi Liu<sup>1,2</sup>, Yanxian Chen<sup>1,3</sup> and Mingguang He<sup>1</sup>✉

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2021

Myopia is a leading cause of visual impairment and has raised significant international concern in recent decades with rapidly increasing prevalence and incidence worldwide. Accurate prediction of future myopia risk could help identify high-risk children for early targeted intervention to delay myopia onset or slow myopia progression. Researchers have built and assessed various myopia prediction models based on different datasets, including baseline refraction or biometric data, lifestyle data, genetic data, and data integration. Here, we summarize all related work published in the past 30 years and provide a comprehensive review of myopia prediction methods, datasets, and performance, which could serve as a useful reference and valuable guideline for future research.

Eye (2022) 36:921–929; <https://doi.org/10.1038/s41433-021-01805-6>

## INTRODUCTION

Myopia is the most common refractive error and a leading cause of reversible visual impairment and blindness worldwide [1, 2]. The “myopia boom” has raised significant international concern in the twenty-first century. The prevalence of myopia keeps increasing in recent years, especially in East Asia where 80–90% of 18-year olds are myopic, and 10–20% are highly myopic [3, 4]. Based on current trends, it is estimated that by 2050, there will be 4758 million people with myopia and 938 million people with high myopia globally [4]. Prevention of myopia onset and progression is critical, since myopia poses an ever-present threat to the quality of life, and high myopia can be further complicated by a number of vision-compromising diseases, including myopic maculopathy and glaucoma [5, 6]. The good news is that various interventions have been proposed to control myopia progression effectively [7–9]. Spectacles are the most widely used method for the correction of myopia [10]. Many newly introduced contact lenses [11] and spectacles [10, 12, 13] also appear to be very effective in slowing progression. Other commonly used effective myopia control methods include low-dose atropine eye drops and orthokeratology [14, 15]. Unlike increased outdoor time which should be encouraged for everyone, the above-mentioned medical interventions differ in expenses, effects and adverse effects, and clinical decision of treatment options should be based on an individual basis [13, 16]. Accurate myopia prediction could help identification of high-risk children for more timely and effective intervention, slowing the pace of myopia onset or progression and leading to improved visual outcome and quality of life [17].

Our understanding of myopia aetiology has been advanced with growing evidence reporting risk factors for myopia onset or progression, including age, gender, parental myopia, susceptible genes, and outdoor activities [7, 18–20]. On this basis, researchers had developed various types of prediction models to foreshadow the risk of myopia or high myopia in different populations [21–24]. Despite the growing interests in this area, no model can be

considered for wide application in clinical practice at present. Existing models differ in definitions of myopia, subsets of predictors and statistical prediction methods, and with varying prediction performance [21, 25–27]. This study aims to provide an evidence-based update on existing myopia prediction models and discuss future challenges.

## LITERATURE SEARCH

We conducted a systematic search of all published articles related to myopia prediction model published between January 1, 1990, and February 1, 2021, by searching the online databases, including PubMed, Embase, and Google Scholar. The search terms contained (“myopia” or “high myopia” or “spherical equivalent” or “spherical equivalence” or “refractive error” or “refraction”) and (“predict” or “prediction” or “predictor” or “predictive”). Detailed search terms could be found in the supplementary file. Published studies were included if they were prospective observational studies conducted on humans and reported the use of a certain method to predict the future myopia risk, including myopia onset, myopia progression, and specific spherical equivalence (SE). Only full-text studies published in English were included. Unpublished studies and meeting abstracts were not included due to uncertainty of methodological quality. Studies evaluating refraction prediction after treatments, including orthokeratology, atropine eyedrop and cataract surgery, were further excluded. A total of 3581 articles were identified in the initial search, and titles of the articles were screened by XH and YC, independently. After excluding duplicates papers and those that did not meet the inclusion criteria, a total of 25 full-text articles were subsequently screened by XH and YC. After full-text review, an additional six articles were excluded: one that constructed a risk score system for myopia symptom based on cross-sectional data [28], six estimated the current myopia status or SE instead of future myopia risk [24, 29–33], and one that used a myopia growth chart to estimate myopia progression but no statistical prediction model

<sup>1</sup>State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. <sup>2</sup>School of Computer Science, University of Technology Sydney, Sydney, NSW, Australia. <sup>3</sup>Department of Ophthalmology, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China. ✉email: [mingguang\\_he@yahoo.com](mailto:mingguang_he@yahoo.com)

Received: 22 October 2020 Revised: 21 August 2021 Accepted: 1 October 2021

Published online: 13 October 2021

was established [23]. In a nutshell, this review was based on 17 core papers that utilized different data types for predicting future myopia risk (Table 1).

### MYOPIA PREDICTION WORKFLOW

To develop a myopia prediction model, there are five things to consider: (1) outcome definition; (2) data acquisition; (3) predictor selection, (4) model development, and (5) model evaluation (Fig. 1).

The first step in myopia prediction is to properly define the outcome, which could be the incidence of myopia/high myopia onset or progression over a certain period of time or the specific SE value. In previous studies, the most commonly used definition was  $SE \leq -0.75D$  in the right eye for myopia, and  $\leq -6D$  in the right eye for high myopia [22, 34, 35]. Future studies should follow the definition of myopia ( $SE \leq -0.50D$  when ocular accommodation is relaxed) and high myopia ( $SE \leq -6D$  when ocular accommodation is relaxed) in the recent international Myopia Institute report [36].

Data acquisition is the most crucial step as myopia prediction is essentially a data-driven problem. It is also a hypothesis-driven problem and researchers need to have a comprehensive understanding of existing literature on myopia prediction models to decide which data to include. Proper selection and definition of potential predictive factors before study initiation is the cornerstone for successful and effective data acquisition. Multiple data types could be used for myopia prediction, including socioeconomic factors, ocular biometry, lifestyle factors, genetic and imaging data. The vast majority of previous studies collected data for myopia prediction from large cohort studies [20, 22, 37], due to the accessibility of diverse structured data from long-term follow-up. Other available data sources include clinical data (e.g., electronic medical records) and publicly available datasets.

Many statistical methods could be used to select candidate predictors from the specific dataset. Most previous studies leveraged linear regression or logistic regression model (depending on whether the outcome is numeric type or categorical type) for correlation analysis, and chose factors with statistically significant associations with outcome ( $P < 0.05$ ) as predictors for further analysis [27, 38]. Other studies also employed tree-based machine learning (ML) models to cope with more complex datasets with numerous factors, such as random forest and gradient boosting regression tree, which can provide the importance of each input factor [21, 26].

The fourth step is to develop a myopia prediction model based on the optimal subset of factors. In some cases, the regression model used for predictor selection can be directly applied as the prediction model or with mild adjustments. Other statistical models such as the discrete-time survival model and generalized estimating equations were also employed per the type of the expected outcome [39, 40]. Recently, with the increasing availability of medical data and computational power, growing attentions have been paid to ML and deep learning algorithms when facing high-dimensional and large-scale datasets [26, 41]. The model development is generally conducted in a 'learning' fashion, i.e., the model is driven by existing data to best fit the ground truth of the given outcome.

The final step is to evaluate the best-fit model. Receiver-operating characteristic (ROC) and area under the ROC curve (AUC) are the most commonly used metrics for assessing the discriminative ability of models in predicting the future presence of myopia [42]. Other metrics include sensitivity, specificity, accuracy, mean square/absolute error, R square and so on. Some studies also imposed resampling processes (such as bootstrapping) or cross-validation methods to gain higher statistical significance [43]. An efficient prediction model is expected to be able to accurately estimate myopia risk for not only the existing data but also unseen data, thus additional evaluations in external

independent datasets are generally preferred to demonstrate the generalization performance.

### DATA-DRIVEN MYOPIA PREDICTION

When predicting myopia, it should be understood that data plays a crucial role in analyzing risk factors, identifying myopia incidents and modelling prediction problems. Driven by the increasingly available data sources, predicting myopia seems to be much more feasible than before. In the literature, myopia prediction can be generally classified into four categories from a data-driven perspective, including prediction based on baseline refraction or biometric data, prediction based on lifestyle data, prediction based on genetic data and prediction based on data integration, as shown in Fig. 2.

#### Prediction based on baseline refraction or biometric data

Baseline refraction and ocular biometry have been consistently reported as risk factors for myopia onset and progression [44]. These factors are relatively easy to collect and have been widely used in previous studies for myopia prediction.

The prediction of myopia based on baseline refraction or biometric data can be traced back to 1999, when Zadnik et al. tried to predict the onset of juvenile myopia based on 554 children who were not myopic at baseline from the Orinda Longitudinal Study of Myopia (OLSM) [34]. Myopia was defined as at least  $-0.75D$  in the right eye measured by cycloplegic autorefractometry. They found that the best single predictor of future myopia onset was the spherical refractive error at baseline. The AUC for the mean sphere was 0.880, with a sensitivity of 86.7% and a specificity of 73.3%. Further combining of corneal power, Gullstrand lens power, and AL in the logistic model resulted in minor improvements (AUC = 0.893). Later in 2015, Zadnik performed a more comprehensive prediction model for myopia onset under the same definition based on the Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error (CLEERE) Study [37]. Thirteen risk factors from 4512 ethnically diverse, nonmyopic school-aged children were assessed. Eight were associated with the onset of myopia, including SE at baseline, parental myopia, AL, corneal power, crystalline lens power, and the ratio of accommodative convergence to accommodation (AC/A ratio), horizontal/vertical astigmatism magnitude, and visual acuity (VA). Multiple prediction models were constructed based on different combinations of these eight factors, and backward stepwise selection and tenfold cross-validation were used for model comparison. These models achieved an AUC of 0.87–0.93 in the prediction of myopia onset. SE was also found to be the single best predictive factor, given that the AUC only decreased by 0.01 or 0.02 when the number of predictors reduced from all eight to SE only. Findings by Ma et al. [45] also supported that single baseline SE could provide effective prediction for future myopia risk. In this study of 1856 students from Shanghai, China, the AUC of baseline AL, AL/CR and SE to predict 4-year incident myopia (cycloplegic  $SE \leq -0.5D$ ) was 0.585, 0.740, and 0.839, respectively. Combining baseline SE, AL/CR, age, gender and parental myopia resulted in an AUC increment of only 0.022 compared with using baseline SE alone.

In addition, Jones et al. [46] used SE of first-grade students and parental myopia to predict myopia onset (cycloplegic  $SE < -0.75D$ ) between the second and eighth grades based on the CLEERE Study. A total of 1854 nonmyopic first graders were included, and the sensitivity and specificity were 62.5% and 81.9%, respectively. They did not perform the ROC curve to evaluate the model performance. Zhang et al. [47] built a 3-year myopia onset prediction model using baseline data from 236 children in Xiamen, China, and further validated its performance in 1979 ethnically Chinese children from Singapore where the myopia prevalence was significantly higher. The model including gender, height, VA,

**Table 1.** A summary of prediction models for myopia.

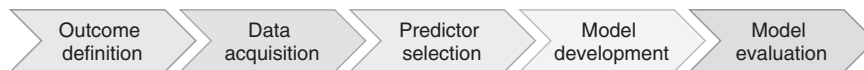
Article	Myopia definition	Outcome definition	Data source	Participants (age, gender)	Predictors	Prediction/analysis method	Evaluation
Zadnik et al. [34]	≤ -0.75D in the right eye (cycloplegic)	Later onset of juvenile myopia	The Orinda Longitudinal Study of Myopia (OLSM)	554 children in grade 3	Refractive error, CP, LP, and AL	Logistic regression model	AUC = 0.893
Jones et al. [35]	≤ -0.75D in the right eye (cycloplegic)	Later onset of myopia	The OLSM	514 children in grade 3	SE, AL, CP, sports/outdoor activity hours, number of myopic parents	Logistic regression model	AUC = 0.90
Jones et al. [46]	≤ -0.75D in the right eye (cycloplegic)	Onset of myopia from grade 2–8	The Collaborative Longitudinal Evaluation of Ethnicity and Refractive Error (CLEERE) Study	1854 children in grade 1	SE at grade 1, number of myopic parents	Discrete-time survival model	Sensitivity of 62.5%; specificity of 81.9%
Zhang et al. [47]	< -0.75D in either eye (cycloplegic)	3-year incident myopia	Study cohort from Xiamen and Singapore	236 Xiamen children (mean 7.82 years); 1979 Singapore children (mean 7.83 years)	Gender, height, presenting VA, AL, ACD, LT, VCD, CC	Generalized estimating equations	AUC = 0.974 in Xiamen; 0.815 in Singapore <sup>a</sup>
French et al. [20]	< -0.50D in either eye (cycloplegic)	Later onset of myopia	The Sydney adolescent Vascular and Eye Study (SAVES)	2059 children	SE, time outdoors, parental myopia, ethnicity, near work	Logistic regression model	AUC = 0.89
Zadnik et al. [37]	≤ -0.75D in the right eye (cycloplegic)	Onset of myopia from grade 2–8	The CLEERE Study	4512 children in grade 1–6 (6–11 years, 49.6% female)	Age, gender, ethnicity, SE, parental myopia, AL, CP, LP, AC/A ratio, horizontal/vertical astigmatism magnitude, and VA.	Discrete-time survival analysis	AUC = 0.84–0.89 for younger cohort; = 0.88 for older cohort
Chua et al. [38]	High myopia: ≤ -5D in the right eye (cycloplegic)	Onset of high myopia	The Singapore Cohort Of the Risk factors for Myopia (SCORM)	928 children (7–9 years)	age at myopia onset, gender, race, school, parental myopia	Linear regression model	AUC = 0.87
Chen et al. [49]	High myopia: ≤ -6D (cycloplegic)	Onset of high myopia before the age of 15 years in GTES	The Guangzhou Refractive Error Study in Children (RESC) and the Guangzhou Twin Eye Study (GTES)	4218 children (5–15 years) and 354 first-born twins (7–15 years)	Reference centile curves for refraction based on the RESC data	Quantile regression model	The 5th centile was most effectively (sensitivity 92.9%, specificity 97.9%, positive predictive value 65.0%)
Ma et al. [45]	≤ -0.50D in the right eye (cycloplegic)	2-year and 4-year incident myopia	A 4-year school-based cohort study in Baoshan, Shanghai, China	1856 children in grade 1–3 (mean 8.1 years, 47.6% female)	Age, gender, baseline SE, parental myopia, AL/CR ratio	Multiple linear regression model	AUC = 0.880 (2 year) and 0.861 (4 year)
Ghorbani Mojarrad et al. [27]	≤ -1.00D (non-cycloplegic)	Incident myopia	The UK Biobank and the Avon Longitudinal Study of Parents and Children (ALSPAC)	3320 children aged 7 years; 2273 children aged 15 years	Age, gender, number of myopic parents, PRS from CREAM and 23andMe Inc, and the UK Biobank	Linear mixed models, cox proportional hazards model	Inclusion of genetic risk score in the survival model improved the model fit compared to using number of myopic parents alone ( $p < 0.0001$ ) <sup>a</sup>
Lin et al. [26]	High myopia: ≤ -6D in the right eye (cycloplegic)	Presence of high myopia at 18 years; onset of high myopia over 10 years	Data from 8 ophthalmic Centres in China, and 2 datasets from population-based cohort studies	Electronic medical records of 129242 individuals (6-20 years); 17113 follow-up records of 3215 participants	Age, SE, annual progression rate	Random forest model	Onset of high myopia for 8 years: AUC = 0.852–0.888 in internal validation; 0.802–0.886 in external validation, 0.752–0.869 in multi-resource testing (4 years); Onset of high myopia by the age of 18: AUC = 0.801 to 0.837 (over 8 years) <sup>a</sup>

**Table 1** continued

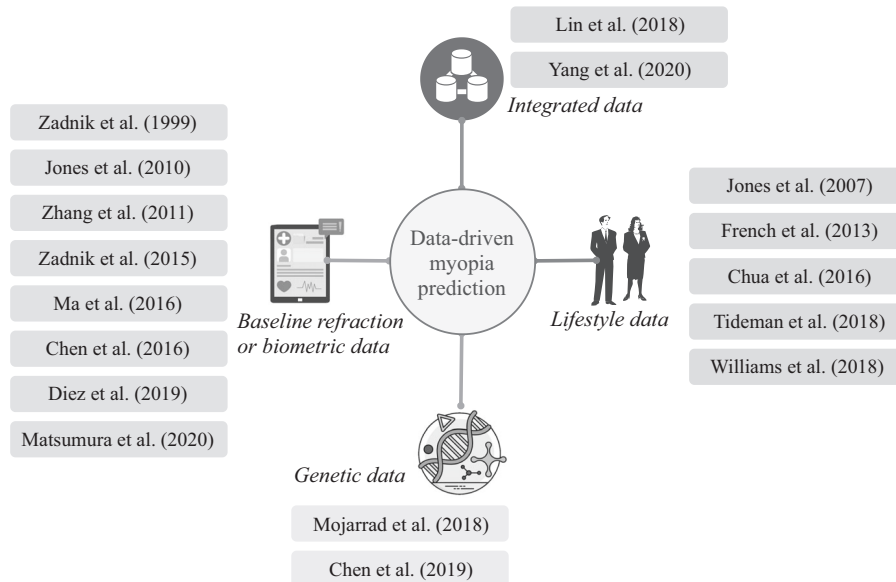
Article	Myopia definition	Outcome definition	Data source	Participants (age, gender)	Predictors	Prediction/analysis method	Evaluation
Williams et al. (2018)	≤ -0.75D, mean of two eye (non-cycloplegic)	Adolescent myopia	The UK-based Twins Early Development Study	1991 twin participants (born between 1991 and 1996)	Age, gender, maternal education, fertility treatment, summer birth, hours spent playing computer games	Logistic regression model	AUC = 0.68, R square = 4.4%
Tideman et al. [54]	≤ -0.50D (cycloplegic)	Incident myopia	The Generation R Study	4734 children (6-9 years, 50.7% female)	Parental myopia, 1 or more books read per week, time spent reading, no participation in sports, non-European ethnicity, time spent outdoors, and baseline AL/CR ratio	Linear regression model	AUC = 0.78
Chen et al. [22]	High myopia: ≤ -6D in the right eye (cycloplegic)	Presence of high myopia at 18 years old	The GTES	1063 children (7-15 years, 51.4% female)	Age, gender, parental SE, outdoor and near work time, PRS from CREAM	Linear mixed effect model	AUC = 0.9845
Diez et al. [25]	High myopia: ≤ -0.50D in the right eye (cycloplegic)	3-year incident high myopia in 226 children	A study cohort from Wuhan, China	12780 children (5-16 years, 48.2% female)	Axial length growth curves based on 12554 children	Analysis of variance (ANOVA); Kruskal-Wallis tests; Least significant difference (LSD) test	9 years or older: 25th percentile had AUC = 0.876 for female and 0.781 for male; 6-9 years: 50th percentile had AUC = 0.875 for female and 0.835 for male <sup>a</sup>
Yang et al. [21]	Not given	Presence of myopia at grade 6	A primary school in Henan, China	3112 children in grade 1-6 (42.2% female)	No. of parents wearing glasses, gender, indoor and outdoor activities, AL, CC, frequency of eating white meat, red meat and eggs, close and remote adjustment ability, pulse, frequency of drinking carbonated drinks, dioptres	Support vector machine	Accuracy = 0.93, precision = 0.95, sensitivity = 0.94, f1-score = 0.94, AUC = 0.98
Matsumura et al. [48]	≤ -0.5D in the worse eye (cycloplegic)	Fast subsequent 2-year myopia progression, defined based on median cut-off	The SCORM study, Singapore	618 myopia children (7-9 years, 46.8% female)	Year 1 myopia progression, baseline SE, age of myopia onset. All models include age, gender, ethnicity, maternal education, parental myopia, near work and outdoor time	Logistic regression model	AUC = 0.77 (year 1 myopia progression), AUC = 0.70 (baseline SE), AUC = 0.65 (age of myopia onset)

D dioptre, SE spherical equivalence, VA visual acuity, AL axial length, ACD anterior chamber depth, LP lens power, CP corneal curvature, VCD vitreous chamber depth, ROC receiver-operating curve, AUC area under the ROC curve, A/C/A ratio of accommodative convergence to accommodation, AL/CR axial length to corneal radius, PRS polygenic risk score, MAE mean absolute error, CNN convolution neuron network.

<sup>a</sup>Availability of independent external validation.



**Fig. 1 The workflow for myopia prediction.** These five steps indicate the general workflow to develop a myopia prediction model.



**Fig. 2 A taxonomy of myopia prediction methods from a data-driven perspective.** Myopia prediction can be generally classified into four categories from a data-driven perspective, and the contents shown in the colored blocks indicate related publications in the literature.

AL, anterior chamber depth (ACD), lens thickness, vitreous chamber depth, and CC achieved an AUC of 0.974 in Xiamen and 0.815 in Singapore. It is noteworthy that this study conducted external validation for the myopia prediction model and proved that models based on one population could be potentially used for other populations with greater myopia prevalence. Matsumura et al. [48] found that year 1 myopia progression is the best predictor for subsequent 2-year myopia progression based on schoolchildren aged 7–9 years from the Singapore Cohort Of the Risk factors for Myopia (SCORM). Reported AUC for year 1 myopia progression, baseline SE, and age of myopia onset were 0.77, 0.70 and 0.66, respectively.

Centile curve was first used as a prediction tool in ophthalmic research by Chen et al. [49] where the outcome of interest was the onset of high myopia. This research group generated reference centile curves based on cycloplegic refraction data of 4218 children aged 5–15 years from the Guangzhou Refractive Error Study in Children (RESC) study and 354 first-born twins from the Guangzhou Twin Eye Study (GTES) [49]. The curves were used to predict the risk of developing high myopia (cycloplegic  $SE \leq -6D$ ) using the follow-up data in GTES. They found that the 5th centile showed the most effective diagnostic value with a sensitivity of 92.9%, a specificity of 97.9% and a positive predictive value of 65.0%. This method was adopted by another study in 2019. Diez et al. built percentile curves of AL from 12,554 children aged 6–15 years in Wuhan China, and used it to predict the likelihood of suffering high myopia (cycloplegic  $SE \leq -5D$ ) in 3 years in an independent subset of 226 children [25]. Reported AUC in this study ranged from 0.781 to 0.876 for children of different ages and genders.

The studies mentioned above suggested that baseline SE was the single best predictor for myopia onset, and preceding SE change was the best predictor for future myopia progression. In addition to statistical models, the centile curve is another more straightforward method for risk estimation on high myopia development. One might reckon that myopia or high myopia, as disease modalities, are all defined by SE. It would not be too surprising to identify SE collected some years ago as the best

predictor, just as adult height could be accurately predicted based on height during childhood [50]. The challenge is how early and how accurate this prediction can be performed. Nevertheless, most previous studies included school-age children and lacked external validation, whether these findings could be applied to younger and older subjects or populations with different ethnicities needs further investigation.

#### Prediction based on lifestyle data

Lifestyle data, including more time spent on near work, less outdoor activity, and attending cram schools, were found to be risk factors for myopia [51–53]. Effect of adding lifestyle data on myopia prediction ability had also been evaluated in several previous studies. Given that lifestyle factors are modifiable, a better understanding of its role in myopia prediction would be very meaningful.

The benefit of adding lifestyle data into myopia prediction model was suggested in two studies. Jones et al. [35] reported that outdoor activities and parental myopia were important predictors of myopia based on the OLSM study (AUC = 0.75), which significantly improved the performance of traditional prediction model which included sphere, AL and corneal power (to AUC = 0.90) [35]. In Tideman et al. [54] included children aged 6–9 years from the Generation R study in Rotterdam, and environmental factors were found to be associated significantly with both increase in AL and incident myopia. They predicted myopia (cycloplegic  $SE \leq -0.5D$ ) based on seven parameters, including parental myopia, one or more books read per week, time spent reading, no participation in sports, non-European ethnicity, less time spent outdoors, and baseline AL/CR ratio, with an overall prediction accuracy of 0.78 (AUC) [54]. In addition, William et al. [55] analyzed 1991 twin participants from the longitudinal UK-based Twins Early Development Study [55]. They found that age, gender, maternal education, fertility treatment, summer birth and hours spent playing computer games were crucial predictors of myopia which in total could explain 4.4% of the variance in SE, with an AUC of 0.68.



However, several other studies failed to confirm the effect of lifestyle factors in myopia prediction. Studies by French et al. [20] reported that baseline SER (AUC = 0.84), time spent outdoors (AUC = 0.64), near work (AUC = 0.61), parental myopia (AUC = 0.65), and ethnicity (AUC = 0.67) were all significant predictors of incident myopia (cycloplegic SE  $\leq$  -0.50D). Adding lifestyle factors to the basic model consisting of only baseline SE could only slightly improve the prediction power in the younger cohort (aged 6 at baseline), and had little effect in the older cohort (aged 12 at baseline). Similarly, using data from the SCORM study, Chua et al. [38] found that age of myopia onset alone could effectively predict high myopia (AUC = 0.85), while the addition of other factors including school and books per week only marginally improved the prediction power (to AUC = 0.87).

The limited effect of adding lifestyle data on myopia prediction model could be due to that many environmental effects had already been reflected in baseline SE or age at myopia onset. On the other hand, our lack of understanding of how lifestyle factors cause or exacerbate myopia imposes a limit on the performance of prediction models. Hence, the role of lifestyle factors in myopia prediction still warrants further study.

### Prediction based on genetic data

During the last decades, genetic investigation had largely improved our understanding of the molecular mechanisms underlying myopia and impaired vision [19, 56]. The Consortium of Refractive Error and Myopia (CREAM) and 23andMe provide the largest combined genome-wide association studies for refractive error [57, 58]. However, these genes could only explain less than 10% of the inheritance variance of myopia. A recent study by Hysi et al. further combined the data from UK Biobank and the Genetic Epidemiology Research on Adult Health and Ageing, and identified 336 novel genetic loci associated with refractive error, which increased the explanation ability of myopia inheritance to 18.4% [33]. Despite the large quantity of genes identified, the role of genetic data in myopia prediction had been less investigated.

Several studies had used genetic data to estimate the current SE in adults [24, 33]. Ghorbani Mojarrad et al. [24] conducted a meta-analysis of three GWAS including a total of 711,984 individuals. A polygenic risk score (PRS) derived from GWAS data for refractive error was evaluated in 1516 adults aged 24–51 years from the Avon Longitudinal Study of Parents and Children (ALSPAC) mothers cohort. The PRS had an AUC of 0.67 for myopia (SE  $\leq$  -0.75D), 0.75 for moderate myopia (SE  $\leq$  -3D), and 0.73 for high myopia (SE  $\leq$  -5D). In addition, a recently published study by Hysi et al. [33] performed a large meta-analysis of GWAS which involved 542,934 European participants. They reported that a combination of age, gender and 890 significant SNPs achieved an AUC of 0.67, 0.74 and 0.75 for myopia definition of  $\leq$ -0.75D,  $\leq$ -3.00D, and  $\leq$ -5D respectively.

Mojarrad et al. [27] did a retrospective analysis of non-cycloplegic autorefraction data of children aged 7 and 15 years from the ALSPAC birth cohort study. Genetic variants associated with refractive error from CREAM and 23andMe were used to calculate a genetic risk score (GRS). Three prediction models were built, with model A including age, gender, number of myopia parents (NMP), and model B, including age, gender, GRS, and model C including age, gender, NMP and GRS. Compared to model A and B, the combined model C could better estimate the current refractive error at both aged 7 years ( $R^2 = 3.0\%$  and  $1.1\%$  vs.  $3.7\%$ ) and 15 years ( $R^2 = 4.8\%$  and  $2.6\%$  vs.  $7.0\%$ ). In predicting incident myopia, inclusion of GRS in the Cox proportional hazard model also improved the model fit compared to using NMP alone ( $p < 0.0001$ ).

Chen et al. [22] assessed the effect of adding genetic information to predict future myopia risk, based on cycloplegia data from 1063 first-born twins from the GTES. GRS was calculated based on the CREAM study, and five models were constructed and

compared. Model 1 included age, age square, and gender; model 2 included age, age square, gender, parental SE, and outdoor and near work time; model 3 included age, age square, gender, GRS, and outdoor and near work time; model 4 included age, age square, gender, and parental SE; and model 5 included age, age square, gender, and GRS. Results showed that model 1 had achieved a good performance in predicting high myopia (SE  $\leq$  -6D) at the age of 18 (AUC = 0.95), whereas adding GRS data did not significantly improve the model performance. This has been attributed to the low penetrance and small effect size of included SNPs in this study, and that the SNPs were taken [59] from GWAS studies based on Caucasian instead of Asian populations. This study further suggested that adding more follow-up visits data into the model could enhance the prediction performance, and for participants with baseline data only, the age of 13 appeared to be the earliest age threshold for high myopia prediction.

Though the rapidly increasing myopia prevalence in recent years has mainly been attributed to environmental factors, both nature (genetics and heredity) and nurture (environment and lifestyle) play a part in the myopia aetiology [10]. The critical advantage of genetic data over refraction in myopia prediction is that the genotype is fixed at conception, and genetic data could be combined with age, gender and other risk factors (e.g., parental SE) for myopia risk prediction at very young ages. Existing studies suggested a very limited added value of genetic data in myopia prediction, this could be due to the current limited understanding of myopia-related SNPs and gene–environment interaction, or environment solely plays a dominant role. The genetic background for high myopia and extremely high myopia is stronger, and theoretically might benefit more from genetic risk prediction. Guggenheim et al. has provided a detailed review on the genetic prediction of myopia, they suggested that genetic prediction had the potential to outperform existing prediction models based on baseline SE or biometric data [18, 59, 60]. With the identification of more susceptible genes with larger effects from large-scale genetic studies in the future, more studies are necessary to enhance our understanding and ability of myopia prediction based on genetic data.

### Prediction based on integrated data

With the wide application of electronic medical records and increasing affordability and convenience of clinical examinations, large-scale real-world medical dataset are accumulating in most hospitals. These datasets provide an extraordinary source for comprehensive clinical analysis and validation. Still, the challenge remains in how to effectively remove the “noises” and extract the “signals” from the integrated data.

Lin et al. [26] predicted myopia among Chinese school-aged children using refraction data from electronic medical records of 129,242 individuals from eight ophthalmic centres. Age, SE and annual progression rate were used to develop an ML algorithm to predict SE and high myopia onset (SE  $\leq$  -6D) in the future 10 years. Their algorithm accurately predicted the presence of high myopia in internal validation of data from the Zhongshan Ophthalmic Center (AUC: 0.903–0.986 for 3 years, 0.875–0.901 for 5 years, and 0.852–0.888 for 8 years), external validation of data from the remaining seven centres (AUC: 0.874–0.976 for 3 years, 0.847–0.921 for 5 years, and 0.802–0.886 for 8 years), and multi-resource testing of data from two population-based cohorts (AUC: 0.752–0.869 for 4 years). The algorithm also achieved clinically acceptable prediction of the actual refraction values at future time points (MAE: 0.253–0.395 for 3 years, 0.394–0.496 for 5 years, and 0.503–0.799 for 8 years). Yang et al. [21] built a myopia prediction model based on 3112 primary school students from Henan, China through ML methods. A feature selection method was first used to construct a predictor subset from the 200 factors in the dataset for model training. The prediction model was built based on the SVM model, and achieved an AUC of 0.87 and an accuracy of 0.79 in

predicting whether a sixth-grader would become myopic based on data from grades 1–5. They have also adopted a data transformation technique, which improved the accuracy to 0.93 and AUC to 0.98 in the final model.

These two studies suggested the potential of using a big integrated dataset in myopia prediction. With the emerging new dataset and advancing technique, integrated real-world big datasets are tended to be a rising topic in future medical research.

## CHALLENGES AND FUTURE DIRECTIONS

Successful myopia prediction has an important reference value for effective myopia control in real-world clinical practice. While outdoor activity is encouraged for all children regardless of potential myopia risk, more frequent follow-ups and invasive interventions (e.g., Orthokeratology and atropine) should be reserved for those with increased myopia risk [7, 61]. Data regarding the effect of these inventions on children with different levels of future myopia risk is needed to guide better clinical management. It should also be noted that there is considerable heterogeneity in myopia pathogenesis, and currently there is no reported prediction model specifically for more rarer forms of myopia (e.g., associated with keratoconus or extreme lenticular myopia) which may progress very differently. Recognizing them is also one of the first steps towards better prediction and personalized medicine in relation to myopia. Looking upon existing myopia prediction methods, several challenges remain.

Predicting myopia incident and progression is ultimately a data-driven problem, and accurately forecasting future incidents relies on the availability of a representative, authentic and comprehensive dataset. Hence, how to establish large datasets with abundant and reliable predictors is challenging. The quality of ground truth determines the performance of prediction. Therefore, standardized methods should be used for data collection and myopia definition. Also, the representativeness of the population should be considered, and multi-ethnic population-based studies were preferable for model training and validation. On the other hand, the availability of multiple data sources is important for model construction and comparison. In addition to the above-mentioned data types, imaging data have been increasingly acknowledged as a valuable data source for medical analysis, especially with the rapid development of artificial intelligence [62]. In the last 4 years, multiple studies have proved the ability to successfully estimate current SE based on various sources of images, including fundus images, ocular appearance images, and photorefractive images [29, 30, 32]. Their findings demonstrate the potential of imaging data for myopia risk prediction which could be considered as a future direction. Moreover, most existing studies can only predict risk in 2–4 years, longer-term follow-up datasets are highly desirable for improving the ability of myopia prediction in the more distant future.

With available datasets, the next challenge is how to construct the prediction models. Previous studies mostly applied traditional regression or survival analysis, future research should explore new statistical methods and models for making better use of new data sources and complex datasets to improve the prediction performance. Considering the increasing availability and accessibility of medical data in the big data era, integrating multi-modality data across multiple domains for myopia prediction appeals to the practitioner in this field. By leveraging appropriate ML methods based on an adequate dataset for myopia prediction, identification of a better prediction model is possible.

How to validate or evaluate the model performance is another challenge to consider. Successful validation in independent datasets with stable performance, preferably in multi-ethnic populations, is warranted before the real-world application of the prediction model in clinical practice. In order to directly compare the performance of myopia prediction models from

different studies, a large publicly available dataset should be established for external validation in the future. Furthermore, even after achieving an accurate myopia prediction model, how to properly deploy this tool in the clinical practice and how to protect patient information privacy present future challenges.

## CONCLUSION

In summary, we provided a comprehensive review and research outlook of myopia prediction methods, datasets and performance, which could serve as a useful reference and valuable guideline for future research. We summarized the research methodology as an integrated process of outcome definition, data acquisition, predictor selection, model development and model evaluation. A thorough literature review was conducted which collectively suggested that age-specific SE was the currently known strongest predictor for myopia prediction, while the additive effect of data including lifestyle, genetic and imaging data was inconclusive. Many challenges existed in this emerging field of myopia prediction. With the development of new analytic methods and accumulating real-world medical datasets, future studies hold promise in better prediction ability of myopia onset and progression, leading us one step closer to the ultimate goal of identifying high-risk populations for timely targeted intervention in clinical practice. We want to emphasize that while more research and better prediction models are always helpful and needed, the major task and challenge to fight the current myopia epidemic are successful implementation of currently available effective myopia prevention strategies (e.g., increased outdoor time) and timely diagnosis and treatment for myopia individuals to minimize the risk of progression.

## REFERENCES

- Fricke TR, Jong M, Naidoo KS, Sankaridurg P, Naduvilath TJ, Ho SM, et al. Global prevalence of visual impairment associated with myopic macular degeneration and temporal trends from 2000 through 2050: systematic review, meta-analysis and modelling. *Br J Ophthalmol*. 2018;102:855–62. <https://doi.org/10.1136/bjophthalmol-2017-311266>.
- Morgan IG, French AN, Ashby RS, Guo X, Ding X, He M, et al. The epidemics of myopia: aetiology and prevention. *Prog Retin Eye Res*. 2018;62:134–49. <https://doi.org/10.1016/j.preteyeres.2017.09.004>.
- Wolffsohn JS, Calossi A, Cho P, Gifford K, Jones L, Jones D, et al. Global trends in myopia management attitudes and strategies in clinical practice – 2019 Update. *Cont Lens Anterior Eye*. 2020;43:9–17. <https://doi.org/10.1016/j.clae.2019.11.002>.
- Holden BA, Fricke TR, Wilson DA, Jong M, Naidoo KS, Sankaridurg P, et al. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology*. 2016;123:1036–42. <https://doi.org/10.1016/j.ophtha.2016.01.006>.
- Ha JJ, He M. Preventing myopia in East Asia. *Community Eye Health*. 2019;32:13–14.
- Kandel H, Khadka J, Goggin M, Pesudovs K. Impact of refractive error on quality of life: a qualitative study. *Clin Exp Ophthalmol*. 2017;45:677–88. <https://doi.org/10.1111/ceo.12954>.
- He M, Xiang F, Zeng Y, Mai J, Chen Q, Zhang J, et al. Effect of Time Spent Outdoors at School on the Development of Myopia among Children in China: a randomized clinical trial. *JAMA*. 2015;314:1142–8. <https://doi.org/10.1001/jama.2015.10803>.
- Yam JC, Jiang Y, Tang SM, Law AK, Chan JJ, Wong E, et al. Low-Concentration Atropine for Myopia Progression (LAMP) Study: a randomized, double-blinded, placebo-controlled trial of 0.05%, 0.025%, and 0.01% atropine eye drops in myopia control. *Ophthalmology*. 2019;126:113–24. <https://doi.org/10.1016/j.ophtha.2018.05.029>.
- Ang M, Flanagan JL, Wong CW, Müller A, Davis A, Keys D, et al. Review: myopia control strategies recommendations from the 2018 WHO/IAPB/BHVI Meeting on Myopia. *Br J Ophthalmol*. 2020. <https://doi.org/10.1136/bjophthalmol-2019-315575>.
- Baird PN, Saw SM, Lanca C, Guggenheim JA, Smith EL III, Zhou X, et al. Myopia. *Nat Rev Dis Prim*. 2020;6:99. <https://doi.org/10.1038/s41572-020-00231-4>.
- Prieto-Garrido FL, Hernandez Verdejo JL, Villa-Collar C, Ruiz-Pomeda A. Predicting factors for progression of the myopia in the MiSight assessment study Spain (MASS). *J Optom*. 2021. <https://doi.org/10.1016/j.optom.2020.11.003>.

12. Lam CSY, Tang WC, Tse DY, Lee RP, Chun RK, Hasegawa K, et al. Defocus Incorporated Multiple Segments (DIMS) spectacle lenses slow myopia progression: a 2-year randomised clinical trial. *Br J Ophthalmol*. 2020;104:363–8. <https://doi.org/10.1136/bjophthalmol-2018-313739>.
13. Wildsoet CF, Chia A, Cho P, Guggenheim JA, Polling JR, Read S, et al. IMI - Interventions Myopia Institute: interventions for controlling myopia onset and progression report. *Investig Ophthalmol Vis Sci*. 2019;60:M106–M131. <https://doi.org/10.1167/iov.18-25958>.
14. Chia A, Chua WH, Cheung YB, Wong WL, Lingham A, Fong A, et al. Atropine for the treatment of childhood myopia: safety and efficacy of 0.5%, 0.1%, and 0.01% doses (Atropine for the Treatment of Myopia 2). *Ophthalmology*. 2012;119:347–54. <https://doi.org/10.1016/j.ophtha.2011.07.031>.
15. Huang J, Wen D, Wang Q, McAlinden C, Flitcroft I, Chen H, et al. Efficacy comparison of 16 interventions for myopia control in children: a network meta-analysis. *Ophthalmology*. 2016;123:697–708. <https://doi.org/10.1016/j.ophtha.2015.11.010>.
16. Wolffsohn JS, Flitcroft DI, Gifford KL, Jong M, Jones L, Klaver CC, et al. IMI - myopia control reports overview and introduction. *Investig Ophthalmol Vis Sci*. 2019;60:M1–M19. <https://doi.org/10.1167/iov.18-25980>.
17. Sankaridurg P, Tahhan N, Kandel H, Naduvilath T, Zou H, Frick KD, et al. IMI impact of myopia. *Investig Ophthalmol Vis Sci*. 2021;62:2. <https://doi.org/10.1167/iov.62.5.2>.
18. Fan Q, Guo X, Tideman JW, Williams KM, Yazar S, Hosseini SM, et al. Childhood gene-environment interactions and age-dependent effects of genetic variants associated with refractive error and myopia: The CREAM Consortium. *Sci Rep*. 2016;6:25853. <https://doi.org/10.1038/srep25853>.
19. Tedja MS, Haarman AEG, Meester-Smoor MA, Kaprio J, Mackey DA, Guggenheim JA, et al. IMI - myopia genetics report. *Investig Ophthalmol Vis Sci*. 2019;60:M89–M105. <https://doi.org/10.1167/iov.18-25965>.
20. French AN, Morgan IG, Mitchell P, Rose KA. Risk factors for incident myopia in Australian schoolchildren: the Sydney adolescent vascular and eye study. *Ophthalmology*. 2013;120:2100–8. <https://doi.org/10.1016/j.ophtha.2013.02.035>.
21. Yang X, Chen G, Qian Y, Wang Y, Zhai Y, Fan D, et al. Prediction of myopia in adolescents through machine learning methods. *Int J Environ Res Public Health*. 2020;17. <https://doi.org/10.3390/ijerph17020463>.
22. Chen Y, Han X, Guo X, Li Y, Lee J, He M. Contribution of genome-wide significant single nucleotide polymorphisms in myopia prediction: findings from a 10-year cohort of chinese twin children. *Ophthalmology*. 2019;126:1607–14. <https://doi.org/10.1016/j.ophtha.2019.06.026>.
23. Kim DH, Lim HT. Myopia Growth Chart Based on a Population-Based Survey (KNHANES IV-V): A novel prediction model of myopic progression in childhood. *J Pediatr Ophthalmol Strabismus*. 2019;56:73–7. <https://doi.org/10.3928/01913913-20181003-01>.
24. Ghorbani Mojarrad N, Plotnikov D, Williams C, Guggenheim JA, Eye UKB, Vision C. Association between polygenic risk score and risk of myopia. *JAMA Ophthalmol*. 2019. <https://doi.org/10.1001/jamaophthalmol.2019.4421>.
25. Sanz Diez P, Yang LH, Lu MX, Wahl S, Ohlendorf A. Growth curves of myopia-related parameters to clinically monitor the refractive development in Chinese schoolchildren. *Graefes Arch Clin Exp Ophthalmol*. 2019;257:1045–53. <https://doi.org/10.1007/s00417-019-04290-6>.
26. Lin H, Long E, Ding X, Diao H, Chen Z, Liu R, et al. Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: a retrospective, multicentre machine learning study. *PLoS Med*. 2018;15:e1002674. <https://doi.org/10.1371/journal.pmed.1002674>.
27. Ghorbani Mojarrad N, Williams C, Guggenheim JA. A genetic risk score and number of myopic parents independently predict myopia. *Ophthalmic Physiol Opt*. 2018;38:492–502. <https://doi.org/10.1111/opo.12579>.
28. Du DD, Xu WL, Yang LH, Wang HX, Gu CM, Tang J, et al. A risk score system for myopia symptom warning. *Curr Med Sci*. 2019;39:455–62. <https://doi.org/10.1007/s11596-019-2060-0>.
29. Varadarajan AV, Poplin R, Blumer K, Angermueller C, Ledsam J, Chopra R, et al. Deep learning for predicting refractive error from retinal fundus images. *Investig Ophthalmol Vis Sci*. 2018;59:2861–8. <https://doi.org/10.1167/iov.18-23887>.
30. Chun J, Kim Y, Shin KY, Han SH, Oh SY, Chung TY, et al. Deep learning-based prediction of refractive error using photorefractive images captured by a smartphone: model development and validation study. *JMIR Med Inform*. 2020;8:e16225. <https://doi.org/10.2196/16225>.
31. Rampat R, Debellemanniere G, Malet J, Gatineau D. Using artificial intelligence and novel polynomials to predict subjective refraction. *Sci Rep*. 2020;10:8565. <https://doi.org/10.1038/s41598-020-65417-y>.
32. Yang Y, Li R, Lin D, Zhang X, Li W, Wang J, et al. Automatic identification of myopia based on ocular appearance images using deep learning. *Ann Transl Med*. 2020;8:705. <https://doi.org/10.21037/atm.2019.12.39>.
33. Hysi PG, Choquet H, Khawaja AP, Wojciechowski R, Tedja MS, Yin J, et al. Meta-analysis of 542,934 subjects of European ancestry identifies new genes and mechanisms predisposing to refractive error and myopia. *Nat Genet*. 2020;52:401–7. <https://doi.org/10.1038/s41588-020-0599-0>.
34. Zadnik K, Mutti DO, Friedman NE, Qualley PA, Jones LA, Qiu PH, et al. Ocular predictors of the onset of juvenile myopia. *Invest Ophthalmol Vis Sci*. 1999;40:1936–43.
35. Jones LA, Sinnott LT, Mutti DO, Mitchell GL, Moeschberger ML, Zadnik K. Parental history of myopia, sports and outdoor activities, and future myopia. *Investig Ophthalmol Vis Sci*. 2007;48:3524–32. <https://doi.org/10.1167/iov.06-1118>.
36. Flitcroft DI, He M, Jonas JB, Jong M, Naidoo K, Ohno-Matsui K, et al. IMI - defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. *Investig Ophthalmol Vis Sci*. 2019;60:M20–M30. <https://doi.org/10.1167/iov.18-25957>.
37. Zadnik K, Sinnott LT, Cotter SA, Jones-Jordan LA, Kleinstein RN, Manny RE, et al. Prediction of juvenile-onset myopia. *JAMA Ophthalmol*. 2015;133:683–9. <https://doi.org/10.1001/jamaophthalmol.2015.0471>.
38. Chua SY, Sabanayagam C, Cheung YB, Chia A, Valenzuela RK, Tan D, et al. Age of onset of myopia predicts risk of high myopia in later childhood in myopic Singapore children. *Ophthalmic Physiol Opt*. 2016;36:388–94. <https://doi.org/10.1111/opo.12305>.
39. Ying GS, Maguire MG, Glynn R, Rosner B. Tutorial on biostatistics: linear regression analysis of continuous correlated eye data. *Ophthalmic Epidemiol*. 2017;24:130–40. <https://doi.org/10.1080/09286586.2016.1259636>.
40. Duenas M, Salazar A, Ojeda B, Arana R, Failde I. Generalized Estimating Equations (GEE) to handle missing data and time-dependent variables in longitudinal studies: an application to assess the evolution of Health Related Quality of Life in coronary patients. *Epidemiol Prev*. 2016;40:116–23. <https://doi.org/10.19191/EP16.2.P116.066>.
41. Lee A, Taylor P, Kalpathy-Cramer J, Tufail A. Machine learning has arrived! *Ophthalmology*. 2017;124:1726–8. <https://doi.org/10.1016/j.ophtha.2017.08.046>.
42. Janssens A, Martens FK. Reflection on modern methods: Revisiting the area under the ROC Curve. *Int J Epidemiol*. 2020. <https://doi.org/10.1093/ije/dy2274>.
43. Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, et al. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *Am J Roentgenol*. 2019;212:38–43. <https://doi.org/10.2214/AJR.18.20224>.
44. Morgan IG, Ohno-Matsui K, Saw SM. Myopia. *Lancet*. 2012;379:1739–48. [https://doi.org/10.1016/S0140-6736\(12\)60272-4](https://doi.org/10.1016/S0140-6736(12)60272-4).
45. Ma Y, Zou H, Lin S, Xu X, Zhao R, Lu L, et al. Cohort study with 4-year follow-up of myopia and refractive parameters in primary schoolchildren in Baoshan District, Shanghai. *Clin Exp Ophthalmol*. 2018;46:861–72. <https://doi.org/10.1111/ceo.13195>.
46. Jones-Jordan LA, Sinnott LT, Manny RE, Cotter SA, Kleinstein RN, Mutti DO, et al. Early childhood refractive error and parental history of myopia as predictors of myopia. *Investig Ophthalmol Vis Sci*. 2010;51:115–21. <https://doi.org/10.1167/iov.08-3210>.
47. Zhang M, Gazzard G, Fu Z, Li L, Chen B, Saw SM, et al. Validating the accuracy of a model to predict the onset of myopia in children. *Investig Ophthalmol Vis Sci*. 2011;52:5836–41. <https://doi.org/10.1167/iov.10-5592>.
48. Matsumura S, Lanca C, Htoon HM, Brennan N, Tan CS, Kathrani B, et al. Annual myopia progression and subsequent 2-year myopia progression in Singaporean children. *Transl Vis Sci Technol*. 2020;9:12. <https://doi.org/10.1167/tvst.9.13.12>.
49. Chen Y, Zhang J, Morgan IG, He M. Identifying children at risk of high myopia using population centile curves of refraction. *PLoS ONE*. 2016;11:e0167642. <https://doi.org/10.1371/journal.pone.0167642>.
50. Topor LS, Feldman HA, Bauchner H, Cohen LE. Variation in methods of predicting adult height for children with idiopathic short stature. *Pediatrics*. 2010;126:938–44. <https://doi.org/10.1542/peds.2009-3649>.
51. Pan CW, Qian DJ, Saw SM. Time outdoors, blood vitamin D status and myopia: a review. *Photochem Photobiol Sci*. 2017;16:426–32. <https://doi.org/10.1039/c6pp00292g>.
52. Ku PW, Steptoe A, Lai YJ, Hu HY, Chu D, Yen YF, et al. The associations between near visual activity and incident myopia in children: a nationwide 4-year follow-up study. *Ophthalmology*. 2019;126:214–20. <https://doi.org/10.1016/j.ophtha.2018.05.010>.
53. Liao C, Ding X, Han X, Jiang Y, Zhang J, Scheetz J, et al. Role of parental refractive status in myopia progression: 12-year annual observation from the Guangzhou Twin Eye Study. *Investig Ophthalmol Vis Sci*. 2019;60:3499–506. <https://doi.org/10.1167/iov.19-27164>.
54. Tideman JWL, Polling JR, Jaddoe VVW, Vingerling JR, Klaver CCW. Environmental risk factors can reduce axial length elongation and myopia incidence in 6- to 9-year-old children. *Ophthalmology*. 2019;126:127–36. <https://doi.org/10.1016/j.ophtha.2018.06.029>.
55. Williams KM, Krapohl E, Yonova-Doing E, Hysi PG, Plomin R, Hammond CJ. Early life factors for myopia in the British Twins Early Development Study. *Br J Ophthalmol*. 2019;103:1078–84. <https://doi.org/10.1136/bjophthalmol-2018-312439>.



56. Cai XB, Shen SR, Chen DF, Zhang Q, Jin ZB. An overview of myopia genetics. *Exp Eye Res.* 2019;188:107778. <https://doi.org/10.1016/j.exer.2019.107778>.
57. Verhoeven VJ, Hysi PG, Wojciechowski R, Fan Q, Guggenheim JA, Höhn R, et al. Genome-wide meta-analyses of multi-ancestry cohorts identify multiple new susceptibility loci for refractive error and myopia. *Nat Genet.* 2013;45:314–8. <https://doi.org/10.1038/ng.2554>.
58. Tedja MS, Wojciechowski R, Hysi PG, Eriksson N, Furlotte NA, Verhoeven VJ, et al. Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. *Nat Genet.* 2018;50:834–48. <https://doi.org/10.1038/s41588-018-0127-7>.
59. Guggenheim JA, Ghorbani Mojarad N, Williams C, Flitcroft DJ. Genetic prediction of myopia: prospects and challenges. *Ophthalmic Physiol Opt* 2017;37:549–56. <https://doi.org/10.1111/opo.12403>.
60. Chen YP, Hocking PM, Wang L, Považay B, Prashar A, To CH, et al. Selective breeding for susceptibility to myopia reveals a gene-environment interaction. *Investig Ophthalmol Vis Sci.* 2011;52:4003–11. <https://doi.org/10.1167/iovs.10-7044>.
61. Deng L, Pang Y. Effect of outdoor activities in myopia control: meta-analysis of clinical studies. *Optom Vis Sci.* 2019;96:276–82. <https://doi.org/10.1097/OPX.0000000000001357>.
62. Keskinbora K, Guven F. Artificial intelligence and ophthalmology. *Turk J Ophthalmol.* 2020;50:37–43. <https://doi.org/10.4274/tjo.galenos.2020.78989>.

#### AUTHOR CONTRIBUTIONS

XH was responsible for designing the review protocol, conducting the search, extracting and analyzing data, interpreting results, writing the manuscript, updating reference lists and creating tables. C.L. was responsible for designing the review protocol, writing and revising the manuscript, updating reference lists and creating

figures. Y.C. was responsible for conducting the search, extracting and analyzing data and revising the manuscript. MH was responsible for designing the review protocol and providing critical revisions of the manuscript.

#### FUNDING

XH receives support from the China Postdoctoral Science Foundation (2019TQ0365) and the National Natural Science Foundation of China (8210040413). MH receives support from the National Key R&D Program of China (2018YFC0116500), the Fundamental Research Funds of the State Key Laboratory of Ophthalmology, Science and Technology Planning Project of Guangdong Province (2013B20400003). The sponsor or funding organization had no role in the design or conduct of this research.

#### COMPETING INTERESTS

The authors declare no competing interests.

#### ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41433-021-01805-6>.

**Correspondence** and requests for materials should be addressed to Mingguang He.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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