A genetic risk score and number of myopic parents independently predict myopia

Neema Ghorbani Mojarrad¹, Cathy Williams², Jeremy A. Guggenheim¹

Affiliations:
¹ School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK.
² Population Health Sciences, Bristol Medical School, University of Bristol, UK.

Keywords:
Refractive error, myopia, genetic prediction, ALSPAC

Corresponding author:
Professor Jeremy A. Guggenheim,
School of Optometry & Vision Sciences,
Cardiff University,
Maindy Road, Cardiff, CF24 4HQ, UK.
Tel +44 (0) 29 2087 4904
Fax +44 (0) 29 2087 4859
Email. GuggenheimJ1@cardiff.ac.uk
Financial disclosures:
No Financial Disclosures. The authors indicate no competing or commercial interests.

Acknowledgements:
We are extremely grateful to all the families who took part in the ALSPAC study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and JG and CW will serve as guarantors for the contents of this paper. A comprehensive list of grants funding (PDF, 459KB) is available on the ALSPAC website. This research was specifically funded by Cardiff University, a PhD Studentship grant from the College of Optometrists (to JG and CW; supporting NGM) entitled “Genetic prediction of individuals at-risk for myopia development”). CW is funded by an NIHR Senior Research Fellowship award SRF-2015-08-005. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.
ABSTRACT

PURPOSE. To investigate whether a genetic risk score (GRS) improved performance of predicting refractive error compared to knowing a child's number of myopic parents (NMP) alone.

METHODS. This was a retrospective analysis of data from the Avon Longitudinal Study of Parents and Children (ALSPAC) birth cohort study. Refractive error was assessed longitudinally between age 7-15 using noncycloplegic autorefraction. Genetic variants (n=149) associated with refractive error from a CREAM consortium genome-wide association study were used to calculate a GRS for each child. Using refractive error at ages 7 and 15 years as the outcome variable, coefficient of determination ($R^2$) values were calculated via linear regression models for the predictors: NMP, GRS and a combined model.

RESULTS. NMP was weakly predictive of refractive error in children aged 7 years, $R^2 = 3.0\%$ (95% CI 1.8-4.1%, $P<0.0001$) and aged 15 years, $R^2 = 4.8\%$ (3.1-6.5%, $P<0.0001$). The GRS was also weakly predictive; age 7 years, $R^2 = 1.1\%$ (0.0-1.9%, $P<0.0001$) and 15 years $R^2 = 2.6\%$ (1.3-3.9%, $P<0.0001$). Combining the 2 variables gave larger $R^2$ values at ages 7, $R^2 = 3.7\%$ (2.5-5.0%, $P<0.0001$) and 15, $R^2 = 7.0\%$ (5.0-9.0%, $P<0.0001$). The combined model improved performance at both ages (both $P<0.0001$).

CONCLUSION. A GRS improved the ability to detect children at risk of myopia independently of knowing the NMP. We speculate this may be because NMP captures information concerning environmental risk factors for myopia. Nevertheless, further gains are required to make such predictive tests worthwhile in the clinical environment.
Introduction

The prevalence of refractive error, particularly myopia (short-sightedness) has increased in recent years\(^1, 2\), and is the leading cause of preventable blindness\(^3\) with some countries in East and Southeast Asia reporting prevalence rates of over 90% in the younger population\(^4\). Current estimates suggest that this increase in prevalence will continue to rise, with a possible 2.5 billion people becoming myopic within the next decade\(^5\).

Myopia has been associated with ocular disease and irreversible blindness, even at low levels of severity\(^6\), and hence there is concern that as myopia is rising in the population so will the levels of sight loss. Refractive error and myopia also increase the financial burdens on healthcare providers and individuals, which have been growing for several years\(^7, 8\). This may lead to individuals avoiding an update to their ocular prescription, which risks accelerating myopia progression\(^9\).

The aetiology of myopia can be described as complex. It has been shown myopia can be produced by a solely environmental cause (the deprivation of form vision during infancy)\(^10-12\), whilst other studies have shown that myopia can develop from a genetic cause alone\(^13, 14\). In the case of purely genetic causes, i.e. syndromic myopia, the effects are not typical for the majority of the myopic population. As such, much attention in the research literature has been to the relative importance of genes versus environment, with a growing consensus that myopia is predominantly caused by a combination of both genetic and environmental factors\(^15-18\).

One widely investigated risk factor is the refractive status of a child’s parents\(^19-21\). Clinicians can offer parents advice on a child’s risk of myopia by questioning them about whether or not they (the parents) have myopia. However this approach has limitations; firstly parents may not be aware of their own refractive status, or they may confuse terms like ‘short-sightedness’ and ‘long-sightedness’, secondly, the approach does not allow an estimation of the level of myopia that may develop, and thirdly, there are conflicting accounts of the relative risk posed by having one or two myopic parents, in different populations\(^21-26\). These conflicting risk estimates make it difficult to give precise, evidence-based advice to parents about their child’s likely refractive status in the future.
Numerous genetic variants associated with refractive error have been discovered\textsuperscript{27-31}. Researchers have used these variants to calculate genetic risk scores (GRS) using beta-coefficients (an effect size calculated for a specific risk allele in a genome-wide association study) and numbers of risk alleles inherited by a particular individual, to predict that individual’s refractive error and their risk of myopia development. For example, Verhoeven et al.\textsuperscript{27} found a tenfold increase in the risk of myopia between individuals in the first versus the tenth decile of a GRS comprised of information on 26 myopia-associated genetic variants.

Here we sought to test two closely-related hypotheses: (1) that a genetic risk score would enhance the prediction of refractive error over and above knowing the number of myopic parents, and (2) that a genetic risk score would enhance the prediction of incident myopia over and above knowing the number of myopic parents.

**Methods**

**Study design and ethical approval**

This work was a retrospective analysis of data from a prospective birth cohort study, the Avon Longitudinal Study of Parents and Children (ALSPAC). The study adhered to the Declaration of Helsinki, and was conducted in compliance with laws in the United Kingdom. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. All participants provided written informed consent. Statistical analyses were performed using R.

**ALSPAC study participants**

ALSPAC recruited pregnant mothers with delivery dates between April 1991 and December 1992 who lived in the area of the previously named county of Avon. 15,247 pregnant women were enrolled into ALSPAC, with 14,701 children still alive at one year of age. Cohort profiles of the ALSPAC study have been published\textsuperscript{32,33}, and the study website contains details of all the data that is available through a fully searchable data dictionary ([www.bris.ac.uk/alspac/researchers/data-access/data-dictionary](http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary)). A flowchart of the participant inclusion and exclusion, and sample subsets can be found in Figure 1.
ALSPAC participants attended research clinics approximately once per year from the age of 7 years. A variety of different measurements were obtained, including assessment of non-cycloplegic auto-refraction (Canon R50 instrument, Canon USA Inc., Lake Success, NY) at the ages of 7, 10, 11, 12 and 15 years old. Replicate auto-refraction readings were averaged, and used to calculate an average mean spherical equivalent (MSE) by adding the sphere and half of the cylinder power together. The MSE averaged between the two eyes was used as the measure for refractive error in the statistical analyses.

ALSPAC researchers collected blood samples from the participants and DNA was extracted from immortalised lymphocytes and genotyped using the Illumina HumanHap550 quad chip genotyping platform for approximately 10,000 children. There were 7,981 children with genome-wide genotype data remaining after excluding those whose data failed quality-control assessments, related individuals, participants with non-European ancestry, and individuals who withdrew their consent.

Information regarding the number of myopic parents was collected via questionnaires completed by each parent; the question asked was: “How would you rate your sight without glasses?” Options for response were: “always very good”, “I can’t see clearly at a distance”, “I can’t see clearly close up”, and “I can’t see much at all.” After evaluating the responses of parents whose refractive error was known, the optimum classification was achieved as follows: parents with responses for both eyes as “I can’t see clearly at a distance” or “I can’t see much at all” or a combination of these two responses were classed as being myopic. Parents with both eyes categorized as “always very good” or “I can’t see clearly close up” or a combination of these two responses were classed as being non-myopic. Any other combination of responses resulted in the classification being set as “missing”.

We divided the participants into five overlapping groups to test our hypotheses, i.e. some participants were assigned to more than one group (see Figure 1). More detail can be found in the relevant sub-sections below.

*Genetic risk modelling*
A genetic risk score was calculated using genetic variants that showed genome-wide significant association with refractive error in a genome-wide association study (GWAS) of 160,420 participants conducted jointly by the CREAM Consortium and 23andMe Inc.\textsuperscript{37}, restricted to those variants whose association with refractive error was replicated UK Biobank participants\textsuperscript{37}. This resulted in 149 variants remaining for calculating the genetic risk scores (Supplementary Table 1), which included both insertion or deletion (INDELs) and single nucleotide polymorphism (SNP) variants.

The ‘--score’ function of Plink 1.9 (version C3)\textsuperscript{38} was used to compute a genetic risk score for each ALSPAC participant. This was done by using the beta-coefficient (effect size) of each of the 149 variants calculated from the UK Biobank replication GWAS\textsuperscript{37} multiplied by the number of risk alleles carried by an individual. The sum of these values was then calculated, as shown in the equation:

\[
P\text{redicted refractive error} = (B_1 \times X_1) + (B_2 \times X_2) + (B_3 \times X_3) + \ldots + (B_n \times X_n)
\]

Where \( B_i \) is the beta-coefficient for variant \( i \) and \( X_i \) is the number of risk alleles of variant \( i \) carried by the participant of interest. For ease of interpretation we standardised the genetic risk score so that it had a mean of zero and a standard deviation of 1 in the sample. We also reversed the sign of the genetic risk score so that a higher score was associated with a higher risk of myopia. An allele score was also calculated using the Plink ‘--score’ function to create a count of the number of risk alleles carried by each participant.

Refractive error linear model predictive analyses
Linear regression models were created with refractive error in Dioptres as the outcome variable and number of myopic parents and/or genetic risk score as the predictor variables. Separate models were created for all children who had refractive error data available at 7 and at 15 years old. The fit of each model was quantified using the coefficient of determination (R\(^2\) value). The number of myopic parents was coded as a categorical variable with three levels (0, 1 or 2), with zero as the reference category. The genetic risk score was coded as a continuous variable. The ‘improvement in model fit’ for the combined model with both predictor variables was tested against a model with only number of myopic parents as predictor, using a likelihood ratio test.
In order to allow a comparison of predictive performance for children of different ages (i.e. age 7 vs. 15 years old) a sample of participants who had refractive data available at both ages was analysed to allow for a direct comparison. As it has been found that participants who consistently attended ALSPAC research clinics throughout the age 7-15 period have a bias towards being more myopic\textsuperscript{33}, we report the results for the samples of participants described above (i.e. all children attending at age 7, and all attending at age 15 years) for all analyses except for the age 7 vs. age 15 comparisons, to minimise the effects of this bias.

**Refractive error trajectory estimation**

Linear Mixed Models (LMM) were used to examine refractive trajectories\textsuperscript{39} to estimate the effects of the genetic risk score and the number of myopic parents, with a combined model also generated for comparison. Each LMM contained terms for the following fixed effects: gender, polynomial terms for age at visit (for the first, second, third, and fourth order), and the predictive variable(s) of interest; namely, the number of myopic parents (coded 0, 1, or 2) and/or the genetic risk score. The refractive error of each child at the baseline visit and the linear trajectory of refractive error with age were modelled as individual-level random effects.

To illustrate the results of the best fitting models, graphs were plotted displaying the refractive trajectories of: (a) children with 0, 1, or 2 myopic parents, (b) children with genetic risk scores, which after being standardised (i.e. converted to Z-scores) were categorised as low, average, or high risk, and (c) children assigned to one of nine groups defined by their number of myopic parents and their standardised genetic risk score category. (In both models containing a term for genetic risk score, “low” and “high” genetic risk were defined as a genetic risk score >1 standard deviation lower or higher than the mean value, respectively, and average genetic risk was defined as having a genetic risk score within ± 1 standard deviation of the mean).

**Prediction of incident myopia**

Survival analysis was conducted using Cox proportional hazards models to investigate risk factors predictive of the ‘event’ of incident myopia\textsuperscript{34}.
All 3,780 participants with a known number of myopic parents, genotype data available, and who had at least one successful auto-refraction measurement were included in these analyses (Figure 1). Myopia was defined as a non-cycloplegic auto-refraction spherical equivalent $\leq -1.00$D, which corresponds, on average, to a cycloplegic refractive error of approximately -0.75 D in this sample\textsuperscript{40}. The age at myopia onset was defined as the age of the child at the first clinic visit at which they were recorded as being myopic. For children who were classified as non-myopic at all of their clinic visits, the age at their last visit was used as the right-censored time to event.

Three models were created for assessing the risks associated with, (a) number of myopic parents coded as a categorical variable with three levels (0, 1 or 2), (b) genetic risk score categorised as low, average, or high risk, and (c) number of myopic parents and genetic risk score combined. Genetic risk score category was defined as above in the LMM methods. All models included gender as a covariate.

**Results**

*Population demographics & genetic risk score distribution*

We identified a total of 3,320 and 2,273 participants at the ages 7 and 15, respectively, who had a known number of myopic parents, genotype data available, and a valid non-cycloplegic auto-refraction measurement. For analyses comparing effects at the 2 ages, a total of 2,048 participants had refractive data at both the ages of 7 and 15, after meeting the above inclusion criteria. The demographic characteristics of these participants are presented in Table 1. Approximately 40% of participants had no myopic parents, while approximately 45% and 15% had 1 and 2 myopic parents, respectively.

The distribution of the number of genetic risk alleles carried by participants and the distribution of genetic risk scores are shown in Figure 2, in the left and right panels, respectively. The mean number of risk alleles was 121 (range 95 to 151); the theoretical maximum would be $2 \times 149 = 298$. The genetic risk score differed subtly between participants with 0, 1, or 2 myopic parents: the GRS was 0.15 SD units higher in participants with 1 vs. 0 myopic parents (P<0.0001) and 0.36 SD units higher in those with 2 vs. 0 myopic parents (P<0.0001). The distribution (density function) of genetic risk scores in participants with different numbers of myopic parents can be seen in Figure 3.
**Linear model analysis**

*Number of myopic parents as a predictor*

The number of myopic parents was weakly predictive of refractive error in children both at age 7 and at age 15 years. At age 7 years, the variation in refractive error explained (R²) by number of myopic parents was 3.0% (95% CI 1.8%-4.1%, P<0.0001). At the age of 15, the R² value increased to 4.8% (95% CI 3.1%-6.5%, P<0.0001).

*Genetic risk score as a predictor*

The genetic risk score of individuals was also weakly predictive of refractive error at both ages. At age 7 years the variation in refractive error explained by the genetic risk score was 1.1% (95% CI 0.04%-1.9%, P<0.0001). This increased to 2.6% (95% CI 1.3%-3.9%, P<0.0001) at the age of 15 years.

*Combined genetic risk score and number of myopic parents*

Combining the 2 predictor variables improved predictive performance at both ages. At 7 years, the R² was 3.7% (95% CI 2.5%-5.0%, P<0.0001) and at 15 years, the R² was 7.0% (95% CI 5.0%-9.0%, P<0.0001). There was minimal evidence for an interaction between genetic risk score and number of myopic parents in the combined model at age 7 years (P=0.63); however there was support for an interaction between having 2 myopic parents and genetic risk score at age 15 years (interaction effect=-0.17 D per 1 SD increase in genetic risk score; P=0.027). An improvement was found between the combined models versus the prediction using number of myopic parents alone using a likelihood ratio test: P<0.0001, at both 7 and 15 years of age. A summary of these results can be seen in Table 2 and Figure 4.

*Comparison of prediction at ages 7 vs. 15*

In the sample of 2,048 individuals with refractive error information at both age 7 and 15 years (Figure 1), the R² for prediction of refractive error using solely the number of myopic parents increased from 2.8% (95% CI 1.4%-4.2%, P<0.0001) at age 7 years to 4.6% (95% CI 2.8%-6.4%, P<0.0001) at age 15 years. For prediction of refractive error using solely the genetic risk score, the R² increased from 0.7% (95% CI 0.0%-1.4%, P<0.0001) at age 7 years to 2.0% (95% CI 0.8%-3.2%, P<0.0001) at age 15 years. For prediction using both
number of myopic parents and genetic risk score, the $R^2$ increased from 3.3% (95% CI 1.8%-4.8%, $P<0.0001$) at age 7 years to 6.1% (95% CI 4.1%-8.0%, $P<0.0001$) at age 15 years. Although all of these results were suggestive of improved prediction at the older age point, the overlapping 95% confidence intervals indicated no statistical support for any of the differences compared to what would be expected by chance.

**Linear mixed model analyses**

After excluding participants lacking genotype data or information about their number of myopic parents, and those who had attended fewer than 3 visits at different ages, there were 3,047 participants available for inclusion in the LMM analyses (Figure 1). Figure 5 shows the refractive error trajectories predicted using the best-fit models for genetic risk score, number of myopic parents, and a combined model, respectively. All models indicated a progression towards a more myopic refractive error with age across all risk categories, although this tendency was less evident in the lower risk categories for both genetic risk score and number of myopic parents. Results from all three models can be found in Table 3.

A broadening in refractive error distribution with age was evident in all three analyses, with number of myopic parents showing a larger distribution between individuals with a different number of myopic parents at the age of 15 years old compared to the high vs. low genetic risk score. Children with 2 myopic parents appeared to show a higher degree of myopia at the age of 15 than those with a high genetic risk score.

The right panel of Figure 5 illustrates the trajectory of participants’ refractive error using the combined model. Consistent with the linear model analyses (Figure 4), the number of myopic parents and the genetic risk score appeared to provide additional information compared to each risk factor in isolation. Thus, the lowest risk was for children with zero myopic parents and a low genetic risk score, whilst the highest risk was for participants with two myopic parents and a high genetic risk. For the combined model, there was evidence for a three way interaction between genetic risk score, number of myopic parents, and age of visit ($P=0.004$).

**Prediction of incident myopia**
Incident myopia was associated with both a higher number of myopic parents and a higher genetic risk score. The survival curves from the Cox proportional hazard models (Figure 6) suggested that the increased risk of myopia conferred by having two vs. zero myopic parents was larger compared to having a genetic risk score in the high vs. low categories. The results from the combined model for incident myopia suggested that participants with two myopic parents and a high genetic risk score had the highest myopia incidence rate, while children with zero myopic parents and a low genetic risk had the lowest incidence rate. Inclusion of genetic risk score in the survival model (Figure 6 right panel) improved the model fit compared to using number of myopic parents alone (likelihood ratio test, P=5.1x10^{-10}). There was minimal evidence for an interaction between the number of myopic parents and the genetic risk score in the combined model (P=0.87). Results from the three survival models can be found in Table 4.

**Discussion**

We studied a large cohort of children whose refractive error was assessed longitudinally between the ages of 7 and 15 years-old to test two closely-related hypotheses: that the accuracy of predicting refractive error (hypothesis #1) and incident myopia (hypothesis #2) based on knowing only the number of myopic parents would be enhanced by also considering a genetic risk score. In support of the hypotheses, predictive performance improved when both predictors were combined, compared to prediction based on the number of myopic parents alone (P<0.0001 for hypothesis #1 and hypothesis #2 at both ages 7 and 15), suggesting that these risk factors are at least partially independent of one another.

Despite the improved predictive performance of linear models incorporating a genetic risk score, the highest R² value achieved using any of our models was less than ~ 7%, which is too low to be clinically useful. This low R² value vividly illustrates the difficulty in predicting the refractive error trajectory of an individual as compared to explaining the refractive trajectories of a group of individuals (i.e. the same phenomenon that causes statistical 95% prediction intervals to be much wider than 95% confidence intervals).

Simulations have indicated that commonly-occurring genetic variants explain approximately 35% of the variance in refractive error (the so-called “SNP heritability”) and
that, theoretically, genetic prediction accuracy is capable of attaining this level or more\textsuperscript{18, 41}. The reason for the poor performance of the genetic risk score was likely due to (a) the omission of many risk variants with significant effects on susceptibility to refractive error and/or (b) using imprecisely-estimated effect sizes for the 149 variants that were included. Both of these limitations could be improved by conducting a GWAS for refractive error in a larger sample than the 160,420 individuals studied by CREAM and 23andMe. However, there are two additional reasons that may account for the poor performance of the genetic risk scores. Firstly, refractive error in the ALSPAC participants was assessed using non-cycloplegic auto-refraction, which would have introduced measurement error and thereby reduced the prediction accuracy. Secondly, refractive development continues beyond 15 years of age, and therefore SNPs selected predominantly on the basis of effects attained by adulthood are likely to perform better at predicting refractive error in adults than in children. (In support of the latter argument, the CREAM Consortium reported that a genetic risk score derived using a wider set of risk variants explained 7.9\% of the inter-individual variation in refractive error in a group of adults unrelated to those included in the GWAS meta-analysis\textsuperscript{37}.

We speculate that the reason for the improved predictive performance of the combined predictors – number of myopic parents and genetic risk score – is because number of myopic parents not only captures information about genetic risk, but also the risk of myopia from environmental factors. This suggestion that myopic parents not only pass on their myopia-predisposing genes to their children, but also raise their children in a relatively myopia-inducing environment, has also been used to explain why estimates of the heritability of refractive error are higher in sibling-sibling comparisons vs. comparisons of more distant relatives\textsuperscript{42, 43}.

Despite the shortcomings of the genetic risk score used here, a much-improved genetic prediction algorithm would have advantages. For instance, genetic risk scores are entirely objective, and thus do not require information recall or input from the patient. They are also highly personalised, meaning that advice and treatment can be tailored specifically for the individual. Furthermore, genetic profiles are becoming increasingly accessible to clinicians and individual members of the public due to their dramatically reducing cost\textsuperscript{44, 45}.
There are several limitations to this work. The study participants were all born within a few years of each other and recruited from the same small geographic region; results in cohorts with a wider age range or recruited from a larger geographic area may have been different depending on the range and levels of myopia-genic environmental exposures (e.g. the age of starting or leaving schooling). Furthermore, our analysis was restricted to participants or European ancestry; the results would likely have been worse in individuals with a broader range of ethnic backgrounds. ALSPAC participants’ refractive error was measured with non-cycloplegic auto-refraction, which is less accurate than cycloplegic autorefraction and typically leads to an age-dependent over-estimation of the degree of myopia and an under-estimation of the degree of hyperopia. This measurement error would be expected to reduce the accuracy of all of our prediction models, however it would not invalidate the comparison between models using different predictors. Moreover, number of myopic parents was assessed with a simple questionnaire at a single time-point in early childhood. Therefore, parental myopia may have been inferred incorrectly (e.g. some parents may have completed the questionnaire after undergoing refractive surgery or have misunderstood the questions) reducing the reliability of our results.

In conclusion, we found an improvement in the performance of statistical models for predicting refractive error and incident myopia when information about genetic susceptibility and parental myopia was combined. Nevertheless, the predictive performance of even the best model was poor ($R^2 < 8\%$), suggesting that currently such models will be of limited clinical value in detecting children who would benefit most from interventions to slow myopia progression. Further genome-wide association studies with larger sample sizes and the inclusion of greater numbers of genetic risk variants in risk scores will be required to increase performance to a level that would make such predictive tests worthwhile in the clinical environment.
References


17. Rohrer B & Stell WK. Basic fibroblast growth factor (bFGF) and transforming growth factor beta (TGF-β) act as stop and go signals to modulate postnatal ocular growth in the chick. Exp Eye Res 1994; 58: 553-561.


<table>
<thead>
<tr>
<th>Variable</th>
<th>Age 7 years sample (N=3,320)</th>
<th>Age 15 years sample (N=2,273)</th>
<th>Age 7 and 15 sample (N=2,048)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>at age 7 visit</td>
<td>at age 15 visit</td>
<td></td>
</tr>
<tr>
<td>Age (mean ± SD)</td>
<td>7.47 ± 0.17 years</td>
<td>15.41 ± 0.27 years</td>
<td>7.45 ± 0.14 years</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.41 ± 0.26 years</td>
</tr>
<tr>
<td>Refractive error (mean ± SD)</td>
<td>0.17 ± 0.81 D</td>
<td>-0.43 ± 1.24 D</td>
<td>0.16 ± 0.79 D</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-0.43 ± 1.19 D</td>
</tr>
<tr>
<td>Myopic (%)</td>
<td>71 (2.1%)</td>
<td>369 (16.2%)</td>
<td>48 (2.3%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>329 (16.1%)</td>
</tr>
<tr>
<td>Male (%)</td>
<td>1,680 (51%)</td>
<td>1,071 (47%)</td>
<td>978 (48%)</td>
</tr>
<tr>
<td>Number of myopic parents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero (%)</td>
<td>1341 (41%)</td>
<td>886 (39%)</td>
<td>783 (38%)</td>
</tr>
<tr>
<td>One (%)</td>
<td>1535 (46%)</td>
<td>1052 (46%)</td>
<td>963 (47%)</td>
</tr>
<tr>
<td>Two (%)</td>
<td>444 (13%)</td>
<td>335 (15%)</td>
<td>302 (15%)</td>
</tr>
</tbody>
</table>
Table 2. Accuracy of predicting refractive error using a linear regression model with predictor variables (A) number of myopic parents, (B) genetic risk score, or (C) a combined analysis.

<table>
<thead>
<tr>
<th></th>
<th>Model A</th>
<th>Model B</th>
<th>Model C</th>
<th>Model 1 vs. Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of myopic parents</td>
<td>Genetic risk score</td>
<td>Combined analysis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R²</td>
<td>95% CI</td>
<td>P-value</td>
<td>R²</td>
</tr>
<tr>
<td>Children aged 7 years (N=3,320)</td>
<td>0.030</td>
<td>0.018-0.041</td>
<td>&lt;0.0001</td>
<td>0.011</td>
</tr>
<tr>
<td>Children aged 15 years (N=2,273)</td>
<td>0.048</td>
<td>0.031-0.065</td>
<td>&lt;0.0001</td>
<td>0.026</td>
</tr>
</tbody>
</table>
Table 3. Fixed effect parameters for best-fit linear mixed models using predictors (A) number of myopic parents, (B) genetic risk score and (C) a combined analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model A</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>-0.003</td>
<td>-0.063 to 0.058</td>
<td>0.93</td>
</tr>
<tr>
<td>Number of myopic parents</td>
<td>-0.221</td>
<td>-0.265 to -0.177</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 1)</td>
<td>-16.602</td>
<td>-18.522 to -14.682</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 2)</td>
<td>-3.150</td>
<td>-3.951 to 2.349</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 3)</td>
<td>2.647</td>
<td>1.861 to 3.433</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 4)</td>
<td>1.900</td>
<td>0.951 to 2.849</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Number of myopic parents x age</td>
<td>-0.026</td>
<td>-0.032 to -0.019</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Model B</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>0.001</td>
<td>-0.059 to 0.062</td>
<td>0.95</td>
</tr>
<tr>
<td>Genetic risk score (per 1 SD increase)</td>
<td>-0.101</td>
<td>-0.409 to 0.207</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 1)</td>
<td>-22.168</td>
<td>-23.469 to -20.867</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 2)</td>
<td>-3.164</td>
<td>-3.970 to -2.368</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 3)</td>
<td>2.651</td>
<td>1.865 to 3.438</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 4)</td>
<td>1.941</td>
<td>0.984 to 2.882</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Genetic risk score x age</td>
<td>-0.015</td>
<td>-0.019 to -0.011</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Model C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>-0.005</td>
<td>-0.065 to 0.056</td>
<td>0.87</td>
</tr>
<tr>
<td>Number of myopic parents</td>
<td>-0.208</td>
<td>-0.252 to -0.164</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Genetic risk score (per 1 SD increase)</td>
<td>-0.086</td>
<td>-0.131 to -0.042</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 1)</td>
<td>-16.887</td>
<td>-18.798 to -14.987</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 2)</td>
<td>-3.162</td>
<td>-3.963 to -2.361</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 3)</td>
<td>2.645</td>
<td>1.861 to 3.434</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age (polynomial order 4)</td>
<td>1.926</td>
<td>0.971 to 2.868</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Genetic risk score x number of myopic parents</td>
<td>0.002</td>
<td>-0.026 to 0.046</td>
<td>0.94</td>
</tr>
<tr>
<td>Number of myopic parents x age</td>
<td>-0.023</td>
<td>-0.029 to -0.017</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Genetic risk score x age</td>
<td>-0.006</td>
<td>-0.013 to 0.000</td>
<td>0.06</td>
</tr>
<tr>
<td>Genetic risk score x number of myopic parents x age</td>
<td>-0.009</td>
<td>-0.016 to -0.003</td>
<td>0.004</td>
</tr>
</tbody>
</table>
Table 4. Survival analysis for “remaining non-myopic” using predictor variables (A) number of myopic parents, (B) genetic risk score and (C) a combined model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HR</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model A</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>1.15</td>
<td>0.98 to 1.34</td>
<td>0.08</td>
</tr>
<tr>
<td>One myopic parent (vs. zero)</td>
<td>1.57</td>
<td>1.30 to 1.88</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Two myopic parents (vs. zero)</td>
<td>3.10</td>
<td>2.50 to 3.84</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Model B</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>1.13</td>
<td>0.97 to 1.32</td>
<td>0.12</td>
</tr>
<tr>
<td>Genetic risk score (per 1 SD increase)</td>
<td>1.33</td>
<td>1.23 to 1.44</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Model C</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female gender (vs. male)</td>
<td>1.15</td>
<td>0.98 to 1.34</td>
<td>0.08</td>
</tr>
<tr>
<td>One myopic parent (vs. zero)</td>
<td>1.51</td>
<td>1.25 to 1.82</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Two myopic parents (vs. zero)</td>
<td>2.90</td>
<td>2.34 to 3.59</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Genetic risk score (per 1 SD increase)</td>
<td>1.28</td>
<td>1.19 to 1.39</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Figure 1. Flowchart of participant selection.

ALSPAC cohort

Removal of participants with consent withdrawn, non-European ancestry, sub-optimal genotype data quality, relatedness or missing genotype data
N = 7,849

Removal of participants with unknown number of myopic parents
N = 4,358

Participants with age 7 refractive data
N = 3,320

Participants with age 15 refractive data
N = 2,273

Participants with both age 7 and 15 refractive data
N = 2,048

Participants with refractive data at 3 different ages
N = 3,047

Participants with refractive data at any age
N = 3,780

Participants used in linear regression models to obtain $R^2$ values for predictive model fit

Participants used linear regression models to identify age-comparative $R^2$ results

Participants used in linear mixed models to estimate refractive error trajectories

Participants used in survival model for incident myopia analysis
Figure 2. Histograms showing the distribution of number of risk alleles carried (Left) and genetic risk Z-scores (Right) for the 4,358 children in the full analysis sample.
Figure 3. Density plot demonstrating the distribution of genetic risk Z-scores for participants with different numbers of myopic parents. It should be noted that the sample sizes were not equal; there were 1,859, 1,946, and 553 participants with zero, one, and two myopic parents, respectively.
Figure 4. Comparison of accuracy of predicting refractive error using number of myopic parents and genetic risk score at the ages of 7 and 15 years old. Error bars show 95% confidence intervals.
Figure 5. Refractive trajectories predicted using (Left) number of myopic parents, (Middle) genetic risk Z-score, and (Right) a combined model with high and low risk genetic risk categories for children with 0, 1, or 2 myopic parents. In the middle and right panels, the high and low genetic risk categories correspond to children with a genetic risk score above or below 1 standard deviation from the mean, respectively.
Figure 6. Survival curves for remaining non-myopic as a function of (Left) number of myopic parents, (Middle) genetic risk score, and (Right) a combined model with genetic risk score and number of myopic parents. In the middle and right panels, the high and low genetic risk categories correspond to children with a genetic risk score above or below 1 standard deviation from the mean, respectively.