university of groningen

University Medical Center Groningen

University of Groningen

Development and validation of a questionnaire-based myopia proxy in adults

Asefa, Nigus G.; Neustaeter, Anna; Vehof, Jelle; Nolte, Ilja M.; Snieder, Harold; Jansonius, Nomdo M.

Published in:
BRITISH JOURNAL OF OPHTHALMOLOGY

DOI:
10.1136/bjophthalmol-2021-319166

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2023

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Asefa, N. G., Neustaeter, A., Vehof, J., Nolte, I. M., Snieder, H., & Jansonius, N. M. (2023). Development and validation of a questionnaire-based myopia proxy in adults: the LifeLines Cohort Study. *BRITISH JOURNAL OF OPHTHALMOLOGY*, *107*, 1035-1042. https://doi.org/10.1136/bjophthalmol-2021-319166

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 01-02-2024

Development and validation of a questionnaire-based myopia proxy in adults: the LifeLines Cohort Study

Nigus G Asefa ,[1] Anna Neustaeter,[2] Jelle Vehof,[2] Ilja M Nolte,[1] Harold Snieder,[1] Nomdo M Jansonius[2]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bjophthalmol-2021-319166).

[1]Department of Epidemiology, University Medical Centre Groningen, Groningen, The Netherlands
[2]Department of Ophthalmology, University Medical Centre Groningen, Groningen, The Netherlands

Correspondence to
Dr Nigus G Asefa, Department of Epidemiology, University Medical Centre Groningen, Groningen 9713 GZ, The Netherlands; n.g.asefa@umcg.nl

Received 1 March 2021
Accepted 13 February 2022

Check for updates

© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Asefa NG, Neustaeter A, Vehof J, *et al*. *Br J Ophthalmol* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/bjophthalmol-2021-319166

ABSTRACT

Aims To build a questionnaire-based myopia proxy and to validate the proxy by confirming its association with educational attainment and a Polygenic Risk Score (PRS) for myopia.

Methods Data were collected between 2014 and 2017 from 88 646 Dutch adults from the LifeLines Cohort. First, we performed principal component analysis (PCA) to responses of five refraction-status questions. Second, we measured the refractive state in a subset of LifeLines participants (n=326) and performed logistic regression using myopia (mean spherical equivalent <−0.5 D) as a dependent variable and the principal components (PCs) as independent variables. We identified specificity, sensitivity and the classification threshold. Third, the classification equation was applied to the remaining LifeLines participants. The value of the proxy was then explored by calculating its association with educational attainment and a PRS of myopia.

Results A total of 77 096 participants (58.1% women) were eligible for the PCA. The first two PCs had a specificity of 91.9% (95% CI 87.8% to 95.4%) and a sensitivity of 90.4% (95% CI 84.3% to 96.4%) for myopia. The area under the receiver operating characteristic curve was 95.0% (95% CI 92.2% to 97.8%). The age-standardised prevalence of proxy-inferred myopia was 33.8% (95% CI 33.4% to 34.3%). Compared with low education level, the ORs of proxy-inferred myopia were 1.66 (95% CI 1.58 to 1.74, $p=5.94\times10^{-90}$) and 2.54 (95% CI 2.41 to 2.68, $p=4.04\times10^{-271}$) for medium and high education levels, respectively. Similarly, individuals at the top 10% of PRS (vs lower 90%) had an OR of 2.18 (95% CI 1.98 to 2.41, $p=6.57\times10^{-56}$) for proxy-inferred myopia, whereas those at the highest decile had an OR of 4.51 (95% CI 3.9 to 5.21, $p=1.74\times10^{-89}$) when compared with the lowest decile.

Conclusion Self-administered refractive error-related questions could be used as an effective tool to capture proxy-inferred myopic cases in a population-based setting.

INTRODUCTION

Myopia (short-sightedness) is a refractive error of the eye that causes the image to focus in front of the retina. A refractive error can be corrected with glasses, contact lenses or refractive surgery. More importantly, myopia is associated with an increased risk of other ocular complications including myopic macular degeneration, glaucoma and retinal detachment.[1 2] The global prevalence of myopia differs significantly based on ethnic background and geographical region. For example, the prevalence of myopia in Europe ranges from 30% to 36%,[3 4] whereas in high-income countries in East and South-east Asia, it has reached epidemic proportions, affecting 80%–90% of young adults.[5] Regardless of these differences, myopia is dramatically increasing worldwide; 20 years ago, myopia affected approximately a quarter of the world's population, and it is projected to be doubled by 2050.[3] The combination of visual impairment and economic loss from myopia-related complications poses a major public health challenge.

Large biobank studies are becoming increasingly important to understand how the biological, psychosocial and behavioural processes that operate across an individual's life course determine health outcomes.[6–8] Such studies collect clinical, genetic, sociodemographic and exposure information in mega-cohorts and provide a unique opportunity to further our understanding of the complex interaction between genetic and environmental factors in myopia. However, ophthalmic examinations, including refractive error assessment, in such large numbers of individuals are challenging and costly.

An alternative approach to collect information in large populations is the use of questionnaires; questions related to age at and/or reasons for first glasses use, impairment of near or distance sight, and assessment of prescribed lenses (minifying, magnifying or no difference to a standard viewed image) have been used to determine the refractive status.[9–13] For example, taking autorefraction as a gold standard, self-reported responses to questions about the reason for first glasses use in the UK Biobank had a specificity of 83.7% (95% CI 83.4% to 84.0%) and a sensitivity of 89.1% (95% CI 88.7% to 89.4%) for myopia.[13] However, different studies used different questions for their myopia definition, and a systematic comparison/synthesis of questions seems lacking thus far.

The objective of this study was to develop and validate a questionnaire-based myopia proxy that can be used in large-scale population-based epidemiology. For this purpose, we (1) performed a principal component analysis (PCA) of the answers to refraction-related questions posed to LifeLines participants, a population-based study conducted in the Northern Netherlands with ~167 000 participants; (2) measured the refractive status in a subset of LifeLines participants, performed a logistic regression with myopia as a dependent variable and the principal components (PCs) as independent variables to construct the proxy; and (3) confirmed

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

Table 1 Near and distance vision-related questions used for the myopia proxy creation

Questions	Possible responses	(Dummy) variables included in the PCA
Q1. Do you wear glasses, reading glasses or contact lenses?	Yes/no	–
Q2. Can you see in the distance without glasses or contact lenses (eg, watching television)?	Yes/no	Yes or NA=1, no=0
Q3. Can you see close by without glasses or contact lenses (eg, read a book)?	Yes/no	Yes or NA=1, no=0
Q4. How old were you when you started wearing glasses or contact lenses?	Numerical response	Q4.1. ≥30 years old=1, <30 years old or NA=0.
		Q4.2. <30 years old=1, ≥30 years old or NA=0.
Q5. What was the main reason for buying these glasses or contact lenses?	1. Difficulty in seeing at a distance.	Q5.1. Difficulty in seeing at a distance=1, difficulty in seeing close by or NA=0.
	2. Difficulty in seeing close by.	Q5.2. Difficulty in seeing close by=1, difficulty in seeing at a distance or NA=0.
	3. Strabismus.	–
	4. I do not know.	–
Q11. Age at data collection.	Numerical response	Numerical data

NA, not applicable; PCA, principal component analysis.

the value of the proxy for epidemiological research by determining the prevalence of proxy-inferred myopia and using the proxy as an outcome measure for analyses with two well-known myopia risk factors, that is, education and a Polygenic Risk Score (PRS) of myopia.

MATERIALS AND METHODS

Study population

The LifeLines Cohort Study and Biobank is a multidisciplinary prospective population-based cohort study of the Northern Netherlands.[14 15] The cohort employed a broad range of investigative procedures in assessing the sociodemographic, biomedical, physical, behavioural and psychological factors which contribute to the health and disease of the general population, with a special focus on multimorbidity and complex genetics. Baseline cross-sectional data were collected between 2006 and 2013 (n=167 729). During the first follow-up visit (median of 3.8 years after baseline), 110 759 LifeLines participants aged 18 and older were asked to complete an eye questionnaire. This questionnaire comprised questions on diagnosis and treatment of eye conditions and included the National Eye Institute Visual Function Questionnaire.[16]

Myopia proxy development

Step 1: PCA

We used a subset of the questions from the LifeLines eye questionnaire, all targeting refractive status, as input to build the proxy. This subset is presented in table 1. We excluded (1) participants with conditions that could affect refractive error status, that is, those who underwent cataract surgery or refractive (laser) surgery; and (2) participants who wore glasses/contact lenses for strabismus or who responded 'I do not know' to the question 'Why do you wear glasses/contact lenses?' (question 5 (Q5) in table 1). We also excluded participants with missing or conflicting information (eg, participants who responded 'no' to glasses use (question 1 (Q1)) but responded to 'age at first glasses use' (question 4 (Q4)). The total number of participants invited, the number of individuals excluded at each stage and the total number of participants included in the final analysis are summarised in figure 1.

During data collection, individuals who answered 'no' to Q1 (glasses non-users) were ordered to skip question 2 (Q2)–Q5 (table 1). To be able to analyse those with and without glasses together, glasses non-users were assumed as if they responded 'yes' to Q2 and question 3 (Q3) (table 1), and two dummy variables were created for each of the binary/binarised variables in Q4 and Q5 (table 1). Q4 was binarised because a linear relationship between age and onset of glasses use was a priori not be expected. We selected age 30 as a threshold value (table 1), because the distribution of age at first glasses use was bimodal at this age (this is also in agreement with clinical observations that myopia development occurs mainly <20 (−25) years and the typical subject with hyperopia (of which the vast majority in the general population is limited to one or two diopters) becomes presbyopic >30 years of age). Finally, using the variables of Q2–Q5 together with age at data collection to adjust for cohort effects, we performed a PCA. For this analysis, we excluded the 326 participants with refractive error measurements who were included in Step 2 (see further below). Data were centred and standardised, and we selected PCs with an eigenvalue of ≥1 for further analyses.[17]

Step 2: from PCs to proxy

Refractive error measurements of both eyes were obtained from a subset of selected (≥55 years old) LifeLines participants (n=326). The data of those participants were obtained from the EyeLife project, a randomised controlled screening trial within the LifeLines Cohort, investigating the added value of genetic prescreening in glaucoma detection. Participant selection was not based on refractive error, and the spherical equivalent (SE) distribution (mean of both eyes) of our sample (mean=−0.06 (SD 2.41) D) was representative of the general population.[18] More information is provided elsewhere.[19] Refractive data were an average of three measurements, performed using the Nidek ARK-1S (NIDEK Co., Gamagori, Japan).[20] We first calculated the SE for each eye using the formula SE=spherical power+½ cylindrical power. The right and left SEs were then averaged to yield mean SE (MSE), which was used in the subsequent analyses. In addition to the exclusion criteria implemented in step 1, participants with (1) |right eye SE minus left eye SE|>1 D and (2) |cylinder power|>1.5 D in both eyes were excluded from the analysis (figure 1). Participants with MSE<−0.5 D were defined as myopic.[3] Responses to Q2–Q5 were modified according to the methods mentioned under step 1. We used the loadings of each question from step 1 to calculate the values of the PCs with eigenvalue of ≥1 for each of the participants.

We built a classification equation by performing a logistic regression with myopia as a dependent variable and PCs with an eigenvalue of ≥1 as independent variables. Using the resultant classification equation, we calculated a receiver operating

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

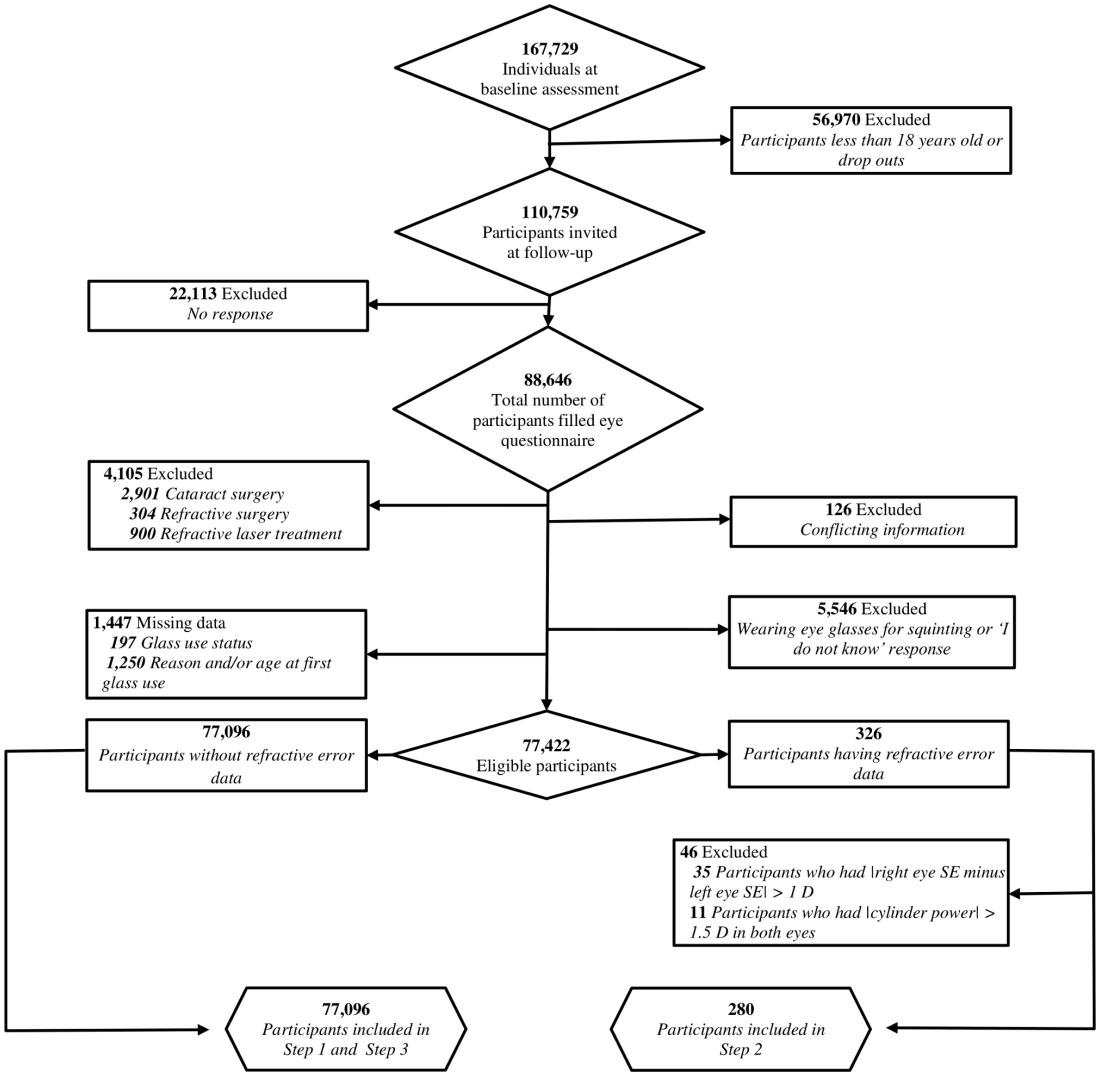

Figure 1 Flow diagram showing exclusion criteria and the final number of individuals included in each step (steps 1–3).

characteristic (ROC) curve, which was used to identify the classification threshold. For the threshold, we employed Youden's Index, that is, the point on the ROC curve with the largest distance between the curve and the chance line, using the pROC package in R V.3.6.1.[21] We subsequently determined the sensitivity and specificity of the myopia proxy using this point. To address potential model overfitting, we performed a *k*-fold (*k*=10) cross-validation analysis[22 23] using the caret package in R[24]; that is, the n=326 dataset was randomly partitioned into *k*=10 mutually exclusive subsets. Of these, *k*-1 subsets were used as the training set, and the remaining subset was used as the testing set. This process was run 10 times; therefore, each subset was used as a testing set once and *k*-1 times as a training set. The cross-validated model was then constructed by averaging the results of the *k*-fold analyses and model performance was estimated by recalculating the area under the ROC curve, sensitivity and specificity. Furthermore, the OR and the explained variances of the models (ie, adjusted R^2 (Nagelkerke's R^2)) of individual questions versus proxy were calculated.

Step 3: application of proxy in confirmation analysis

As a reality check, we determined the prevalence of proxy-inferred myopia by using the proxy derived from step 2 to define cases of myopia in the remaining sample (n=77 096). Subsequently, we used this outcome measure for analyses with two well-known myopia risk factors: educational attainment and a PRS of myopia. Self-reported education was categorised into three categories: low (no education, primary education, lower general secondary education, and lower or preparatory vocational education), medium (higher senior secondary education, intermediate vocational education or apprenticeship, or preuniversity secondary education) and high (higher vocational education or university).

For the PRS and proxy-inferred myopia association analysis, the genetic data of 32 817 participants were obtained from the LifeLines database. For each individual, the number of effect alleles carried at each variant was summed, weighted by its effect size for refractive error. We incorporated only independent autosomal variants (ie, with linkage disequilibrium r^2 <0.1 and a physical distance >1000 kb) with a minor allele frequency of >1%. Because of its complex linkage disequilibrium structure, single-nucleotide polymorphisms (SNPs) from the major histocompatibility complex region were excluded. Variants (n=323 431 SNPs) that passed these criteria were included in the PRS, which was calculated using PLINK.[25] The PRS was constructed based on summary statistics derived from SE Genome-Wide Association Study (GWAS) of 95 619 UK Biobank participants (40–69 years of age, European

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

Table 2 Population characteristics (N=77 096) stratified by glasses, reading glasses or contact lens use

		Q1. Do you wear glasses, reading glasses or contact lenses?*	
		Yes	**No**
n		56 534 (73.3)	20 562 (26.7)
% female		60.3	52.1
Q2. Can you see at a distance without glasses or contact lenses (eg, watching television)?	Yes	19 709 (34.9)	NA
	No	36 825 (65.1)	
Q3. Can you see close by without glasses or contact lenses (eg, read a book)?	Yes	19 761 (35.0)	NA
	No	36 773 (65.0)	
Q4. How old were you when you started wearing glasses or contact lenses?	Age 30 years or more at first glasses use	30 701 (54.3)	NA
	Age <30 years at first glasses use	25 833 (45.7)	NA
Q5. What was the main reason for buying these glasses or contact lenses?	Difficulty seeing at a distance	29 519 (52.2)	NA
	Difficulty seeing close by	27 015 (47.8)	
Q11. Age (years) at data collection, mean (SD).		53.1 (11.4)	40.5 (9.4)

*Unless specified otherwise, data represent the number of participants (%).
NA, not applicable.

ancestry) who underwent non-cycloplegic autorefraction.[26] We generated eight PRSs using increasing liberal GWAS p value cut-offs: PRS 1 ($p<5.0\times10^{-8}$), PRS 2 ($p<5.0\times10^{-7}$), PRS 3 ($p<5.0\times10^{-6}$), PRS 4 ($p<5.0\times10^{-5}$), PRS 5 ($p<5.0\times10^{-4}$), PRS 6 ($p<5.0\times10^{-3}$), PRS 7 ($p<5.0\times10^{-2}$) and PRS 8 ($p<5.0\times10^{-1}$).

We calculated a raw proxy-inferred myopia prevalence in our cohort and a standardised prevalence. For the standardised prevalence, we used data from a large European study on the prevalence of myopia.[27] Following this study, we categorised age (from 20 years to 90 years) into 5-year intervals, and the prevalence of proxy-inferred myopia in each age category was weighted using the proportion of the European population in each category.[28] We performed logistic regression analyses to explore the effects of education and PRS on myopia. The association of educational attainment and PRS with proxy-inferred myopia was adjusted for age and sex. Of the eight PRSs, the PRS with the highest area under the curve (AUC) was selected for the association analysis. The ORs (95% CIs) of proxy-inferred myopia for (1) the top 10% of the PRS and (2) educational attainment were calculated. We also further categorised individuals into 10 groups (deciles) based on their risk score to investigate if there exists a dose–response relation between myopia and the deciles of the PRS. Consequently, the odds of each decile of PRS versus lower decile and the corresponding 95% CIs were also estimated. Statistical significance was set at a p value of <0.05.

RESULTS

Step 1: PCA

Of the 110 759 invited LifeLines participants at follow-up, 88 646 had completed the self-administered eye questionnaire, that is, an 80% response rate. After applying the inclusion and exclusion criteria, 77 096 participants were eligible for the PCA in step 1 (figure 1). Table 2 presents the characteristics of these 77 096 participants. Approximately three-quarters of the participants wore glasses, reading glasses or contact lenses.

Figure 2 presents the results of the PCA. Q5 (reason for first glasses use) had the highest loading onto PC1 (Q5.1= −0.80 and Q5.2=0.86). The first two PCs had an eigenvalue of ≥1 (figure 2). The first PC accounted for 47.3% of the variability in the data. The cumulative variance explained by the first two PCs was 72.5%; all five PCs together explained 97.1% of the variance in the data.

Step 2: from PCs to proxy

Of the step 2 participants who completed questions Q1–Q5 and also had measured MSE data (n=326), 280 were eligible for further analyses (figure 1). Of these 280 participants, 83 were myopic based on the MSE <−0.5 D criterion. The logistic regression analysis with myopia as a dependent variable and PC1 and PC2 as independent variables resulted in the following classification equation:

$$y = 1.19 - 2.36 \times PC1 + 0.23 \times PC2.$$

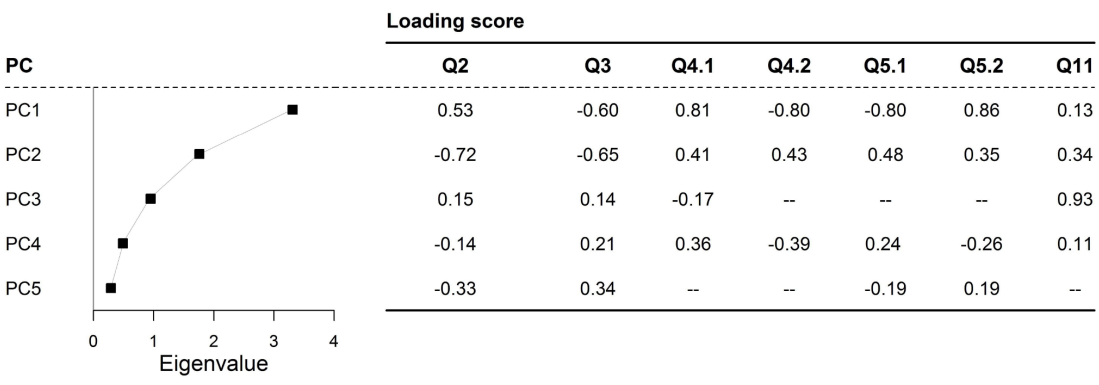

Figure 2 Scree plot and the loading scores of the questions included in the principal component analysis. PC, principal component.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

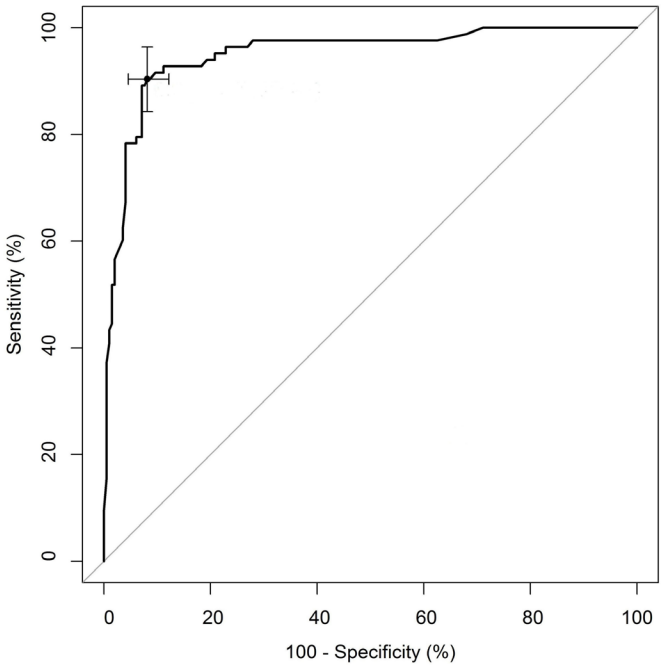

Figure 3 ROC curve for classification equation with PC1+PC2 as independent predictors and myopia as an outcome, depicting the sensitivity and specificity with 95% CI based on Youden's Index. PC, principal component.

where −1.19 is the intercept, and −2.36 and 0.23 are the effect sizes of PC1 and PC2, respectively. Figure 3 shows the resulting ROC curve. The AUC was 95.0% (95% CI 92.2% to 97.8%) for the proxy. Applying Youden's Index as a cut-off value yielded a threshold for y of 0.28. With this threshold, we found a specificity of 91.9% (95% CI 87.8% to 95.4%) and a sensitivity of 90.4% (95% CI 84.3% to 96.4%). In the k-fold cross-validation analysis, these values became 86.5% (95% CI 81.8% to 91.2%) for the AUC, 79.5% (76.8% to 82.3%) for the specificity and 93.4% (91.7% to 95.1%) for the sensitivity. Besides, for Q4+Q5 (the two most contributing questions combined, online supplemental figure S1), the AUC, sensitivity and specificity were 93.2% (95% CI 90.0% to 96.4%), 92.8% (95% CI 86.6% to 97.6%) and 87.3% (95% CI 82.7% to 91.4%), respectively, which changed, after the k-fold cross-validation analysis, to 84.9% (95% CI 80.1% to 89.8%), 90.4% (95% CI 85.3% to 94.1%) and 79.5% (69.2% to 87.6%), respectively. For the 280 participants in step 2, the variance explained by PC1 and PC2 together in the logistic regression model was 70.0% (95% CI 64.2% to 75.8%), whereas for the individual questions, this value ranged from 32.5% (95% CI 23.6% to 41.4%) for Q3 to 66.6% (95% CI 60.3% to 72.9%) for Q5 (online supplemental table S1).

Step 3: application of proxy in confirmation analysis

After excluding step 2 participants (n=326) and applying the classification equation to the remaining LifeLines participants, we identified 26 814 proxy-inferred myopic cases, resulting in an overall crude prevalence of 34.8% (95% CI 34.4% to 35.1%); the standardised prevalence was 33.8% (95% CI 33.4% to 34.3%). Figure 4A shows the ORs of proxy-inferred myopia with educational attainment, after adjusting for age and sex. As can be seen in this figure, we found a strong association of educational attainment with proxy-inferred myopia and also a clear 'dose–response' relationship; those with a medium education level were 1.7 times and those with a high education level 2.5 times more often myopic than those with a low education level.

Of the 77 096 LifeLines participants in step 3, 19 298 had GWAS data. The best predictive accuracy was obtained for PRS 4 ($p<5\times10^{-5}$ threshold; AUC 61.4%, 95% CI 60.6% to 62.3%). Figure 4B shows the ORs of proxy-inferred myopia after categorising PRS 4 into deciles (lowest decile, lowest genetic risk score), after adjusting for age and sex. Those in the top 10%

A (N=77,096)

Risk factors	Odds ratio (95%CI)	P
Sex (female)	1.42 (1.38-1.47)	2.94x10-107
Age per 10 years	0.95 (0.94-0.96)	1.09x10-12
Education		
Low (reference)		
Medium	1.66 (1.58-1.74)	5.94x10-90
High	2.54 (2.41-2.68)	4.04x10-271

B (N=19,298)

PRS		
Top 10% (vs. lower 90%)	2.18 (1.98-2.41)	6.57x10-56
Decile 1 (reference)		
Decile 2	1.42 (1.22-1.66)	1.14x10-2
Decile 3	1.64 (1.41-1.91)	3.02x10-7
Decile 4	1.83 (1.58-2.13)	5.55x10-12
Decile 5	2.12 (1.83-2.46)	8.89x10-20
Decile 6	2.44 (2.11-2.83)	3.34x10-32
Decile 7	2.74 (2.37-3.17)	3.91x10-38
Decile 8	2.71 (2.35-3.14)	2.77x10-37
Decile 9	3.38 (2.92-3.91)	4.18x10-60
Decile 10	4.51 (3.90-5.21)	1.74x10-89

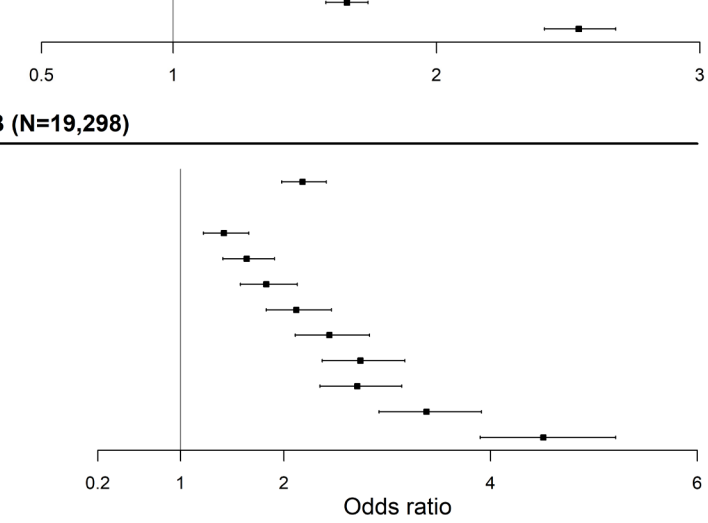

Figure 4 ORs of proxy-inferred myopia with education (A) and myopia PRS (B), both adjusted for age and sex. PRS, Polygenic Risk Score.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

of PRS were approximately two times more often myopic than those in the lower 90%. We also observed a dose–response relationship between myopia and PRS, that is, the ORs of the 9th and 10th deciles were approximately 3.0 and 4.5 when compared to the lowest decile, respectively. The variance explained by the PRS in proxy-inferred myopia was 4.9% (based on 19 298 participants).

DISCUSSION

Using refractive error-related questions from a large, population-based study and refraction data from a subset, we created a questionnaire-based myopia proxy for use in large population-based surveys. Our proxy yielded a prevalence of 35% for proxy-inferred myopia (in a population with a mean age of approximately 50 years and median year of data collection 2015). Our association analyses confirmed the well-known associations of educational attainment and genetic predisposition with myopia.

We standardised our proxy-inferred myopia prevalence to the age distribution of a large, European meta-analysis.[27] This meta-analysis was based on refractive error data collected from 15 European population-based studies. Our standardised prevalence of 33.8% (95% CI 33.4% to 34.3%) agrees reasonably well with their reported prevalence of 30.6% (95% CI 30.4% to 30.9%) with myopia defined as SE ≤−0.75 D (which essentially equals our MSE <−0.5 D). One possible explanation for the difference is that our data collection was more recent, the median year of data collection in LifeLines was 2015, compared with 1990–2013 in the European meta-analysis study—reflecting the increase in myopia prevalence. When compared with individual population-based studies with a matching myopia definition, our crude prevalence of 34.8% (95% CI 34.4% to 35.1%) was similar to the crude prevalence reported in the UK Biobank (33.5%, 95% CI 33.3% to 33.9%), with data collection in 2009,[29] but clearly higher than the prevalence reported in the Blue Mountain Eye Study (15%).[30] The data in the Blue Mountain Eye Study were collected between 1992 and 1994 from older participants (>49 years); this difference in results is expected as the burden of myopia is increasing globally in young people.[3 4]

The results of our association analyses confirm the robustness of the current myopia proxy. Our study is in agreement with previous epidemiological studies reporting educational attainment as one of the strongest risk factors in myopia.[4 31] Historically, strong evidence of a causal role of studying and close work in increasing the prevalence of myopia came from Israeli boys attending Orthodox schools (with a reading load of 16 hours/day) who had a much higher prevalence of 80% compared with the 30% among girls attending the same school. Girls attending Orthodox schools had similar study habits (6 hours' school day) as students attending secular schools, who also showed a myopia prevalence of 30% for both boys and girls.[32–34] Recently, two studies confirmed the causal association between educational attainment and myopia using the Mendelian randomisation approach.[35 36] For example, Cuellar-Partida *et al* reported that an increase of approximately 2 years in education results in a change in SE towards myopia of 0.92±0.29 D.[36]

Myopia is an aetiologically heterogeneous trait resulting from the effects of genetic and environmental risk factors, and the interaction between them.[31 37] Our findings demonstrated that individuals at the highest decile of the PRS were four and a half times more likely to develop myopia. The current findings support those of a previous study reported by Ghorbani Mojarrad *et al*, where individuals in the highest decile had a higher risk (OR 3.47, 95% CI, 2.43 to 4.91) of myopia when compared with the other nine deciles combined.[26] This appears lower than our value, however, we compared the lowest to the highest decile. If we use the same approach, that is, top decile versus nine other deciles combined, then the OR is 2.18 (1.98-2.41, $p=6.57\times10^{-56}$). However, it should be noted that Mojarrad *et al* defined myopia at MSE ≤ −0.75 D, and the analysis was performed on all-female study participants.

Proxy-inferred myopia was more prevalent among female participants in this study. There is no clear evidence on the role of sex on myopia development, and as such, recent studies have hypothesised that the association between sex and myopia is more likely to be a proxy for the differential access, encouragement, or selection to education, near work, or outdoor activities between men and women.[3 5] In our data, sex remained highly significant after adjusting for educational attainment (figure 4A), indicating that education alone is unlikely to fully explain the higher prevalence of proxy-inferred myopia in women. However, no data were available for time spent on near work or outdoor activities, but—with occupation classified in 10 different categories—we noted a significant difference in occupation between men and women (χ^2 test $p<2.0\times10^{-16}$). We added these 10 categories to the analysis presented in figure 4A; a likelihood ratio test showed that the model with occupation performed better than the model without ($p=4.77\times10^{-13}$). Inclusion of occupation in the model resulted in a modest drop of OR (95% CI) of proxy-inferred myopia among women, ie, OR 1.42, 95% CI 1.38 to 1.47 before vs 1.30 95% CI 1.25to 1.35 after adjusting for occupation (online supplemental figure S2). Similarly, the negative association between age and myopia supports previous evidence that reported an increasing burden of myopia among recent birth decades.[4 38] The negative association may be a proxy for the role of lifestyle changes resulting from increased near-work activities and decreased time outdoors in younger adults, similar to the cohort effect observed between cross-sectional studies performed in different time periods.[3]

Our proxy yielded a reasonably high sensitivity and specificity, indicating that it is a useful tool for epidemiological myopia research. Directly comparing the sensitivity and specificity with other questionnaire-based studies is difficult due to different gold standards and different myopia definitions (cut-off values), among other reasons. Our sensitivity value (90%) was similar to that of the UK Biobank (89%)[13] and a study by Ip *et al* (88%)[11] while showing simultaneously a higher specificity of 92% compared with 84%, and 83% in the UK Biobank[13] and Ip *et al*,[11] respectively. A similar specificity (93%) but lower sensitivity (83%) was found in a study by Breslin *et al*[12]; lower values for both sensitivity and specificity were found in a study by Walline *et al*[9] (see online supplemental table S2 for details). As can be seen in this table, the various studies used at least similarly phrased questions for their myopia definition. One possible explanation for our apparently favourable performance is the use of PCA. By using PCA, we maximised the information that could be retrieved from the questions. When comparing individual questions or combinations thereof with PCA in our dataset, the PCA approach yielded the highest AUC point estimate (online supplemental figure S1). Similarly, compared with the best contributing questions (Q5 and Q4), the proxy (PC1 and PC2 together) showed a clear but non-significant improvement in the explained variance (online supplemental table S1). However, as shown in the cross-validation analysis, model overfitting is an issue, and the final AUC, sensitivity and specificity (currently based on n=280) have to be determined in a new, larger sample.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

Our study has a number of strengths and some limitations. The novelty of this work is the application of an advanced statistical approach (PCA) to get all available information out of questionnaire data. By using dummy variables, we were able to include both glasses users and non-users in the analyses. Given that the questions used to build the proxy were independent of ethnicity or myopia prevalence, our work should, in principle, be applicable to future studies of any ethnic group or cohort; that is, the proxy can be applied 'as is', as long as studies use the same questions. Otherwise, they may use our methodology to build their own proxy. A caveat is that glasses use will depend on economic factors as well. Another strength is the very large sample size of the population-based LifeLines Cohort. On the other hand, early nuclear cataract may cause a myopic shift, potentially resulting in misclassification. The classification depends on the final PCA score arising from the (now possibly conflicting) answers. Importantly, some of the questions relate to the past, not to the present situation (eg, age at which first glasses were bought (Q4) and reason to buy these glasses (Q5)), and interestingly, these questions were the main contributors to the first PC and this component was—by far—the most important contributor to the classification equation. Also, a significant myopic shift nowadays generally results in an early cataract extraction rather than repetitive adjustments of the glasses. Furthermore, the modest number of participants with refraction data (n=326) may have limited the accuracy of our validation analyses. Finally, in order to optimise the gold standard in terms of having or not having the target condition (myopia), we excluded those with anisometropia or astigmatism exceeding 1.0 and 1.5 D, respectively (see the Materials and methods section). This concerned 46 of the 326 participants with measured refraction data (figure 1). The remaining question is on how far the proxy is able to detect myopia among those with anisometropia or astigmatism. Of the n=46 individuals with anisometropia or astigmatism, 28 had MSE of <-0.5 D and of these, 21 were detected by the proxy as proxy-inferred myopic cases. Of the remaining 18 (46−28) individuals without myopia, only 5 were detected by the proxy as proxy-inferred myopic cases. Hence, the proxy appears to perform reasonably well in subjects with anisometropia or astigmatism.

CONCLUSION

In summary, using self-reported refraction data of those with and without glasses together, we developed a myopia proxy, which may be used effectively in large population-based studies with relatively high specificity and sensitivity. The value of the proxy was confirmed by a realistic prevalence estimate and by using it as an outcome measure for analyses with educational attainment and a PRS of myopia as predictors. Being cost-effective and less time-consuming, the current work could improve myopia research in large population-based samples investigating novel prevention or treatment approaches.

Acknowledgements We thank the University Medical Centre Groningen for their support and for providing access to the high-performance computing cluster (Peregrine and Gearshift).

Contributors NGA, NMJ and HS conceived the idea and designed the study. NGA analysed the data and wrote the manuscript with intellectual inputs from NMJ and HS. IMN performed the Polygenic Risk Score analysis. JV prepared the questionnaire. AN collected refraction data. NMJ, HS, AN, JV and IMN interpreted the data. NGA, HS, and NMJ accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish. All authors approved the final version of the manuscript for submission.

Funding This project has received funding from the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant agreement number 661 883). Additional funding was provided by the Rotterdamse Stichting Belangen (grant number B20150036).

Disclaimer The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The LifeLines data collection was approved by the medical ethics committee of the University Medical Centre Groningen. Written informed consent was obtained from all participants, and the data collection was conducted in accordance with the tenets of the Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Data used in the current study were obtained from the LifeLines Cohort Study. LifeLines data are open for all researchers upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iD

Nigus G Asefa http://orcid.org/0000-0001-5680-5527

REFERENCES

1 Aleman AC, Wang M, Schaeffel F. Reading and myopia: contrast polarity matters. *Sci Rep* 2018;8:10840.
2 Fredrick DR. Myopia. *BMJ* 2002;324:1195–9.
3 Holden BA, Fricke TR, Wilson DA, *et al*. Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050. *Ophthalmology* 2016;123:1036–42.
4 Williams KM, Bertelsen G, Cumberland P, *et al*. Increasing prevalence of myopia in Europe and the impact of education. *Ophthalmology* 2015;122:1489–97.
5 Morgan IG, Rose KA. Myopia: is the nature-nurture debate finally over? *Clin Exp Optom* 2019;102:3–17.
6 Ng JWY, Barrett LM, Wong A, *et al*. The role of longitudinal cohort studies in epigenetic epidemiology: challenges and opportunities. *Genome Biol* 2012;13:246.
7 Kuh D, Ben-Shlomo Y, Lynch J, *et al*. Life course epidemiology. *J Epidemiol Community Health* 2003;57:778–83.
8 Edwards N, Plotnikoff RC. Mobilizing an underused resource: cohort studies for population health intervention research. *Int J Epidemiol* 2018;47:1730–3.
9 Walline JJ, Zadnik K, Mutti DO. Validity of surveys reporting myopia, astigmatism, and presbyopia. *Optom Vis Sci* 1996;73:376–81.
10 Cumberland PM, Peckham CS, Rahi JS. Capturing myopia and hypermetropia 'phenotypes' without formal refraction. *Eye* 2008;22:939–43.
11 Ip J, Robaei D, Rochtchina E, *et al*. Can information on the purpose of spectacle use and age at first use predict refractive error type? *Ophthalmic Epidemiol* 2007;14:88–92.
12 Breslin KMM, O'Donoghue L, Saunders KJ. An investigation into the validity of self-reported classification of refractive error. *Ophthalmic Physiol Opt* 2014;34:346–52.
13 Cumberland PM, Chianca A, Rahi JS, *et al*. Accuracy and utility of self-report of refractive error. *JAMA Ophthalmol* 2016;134:794–801.
14 Scholtens S, Smidt N, Swertz MA, *et al*. Cohort profile: lifelines, a three-generation cohort study and Biobank. *Int J Epidemiol* 2015;44:1172–80.
15 van der Ende MY, Hartman MHT, Hagemeijer Y, *et al*. The lifelines cohort study: prevalence and treatment of cardiovascular disease and risk factors. *Int J Cardiol* 2017;228:495–500.
16 Mangione CM, Lee PP, Gutierrez PR, *et al*. Development of the 25-item National eye Institute visual function questionnaire. *Arch Ophthalmol* 2001;119:1050–8.
17 Cliff N. The eigenvalues-greater-than-one rule and the reliability of components. *Psychol Bull* 1988;103:276–9.
18 Williams KM, Verhoeven VJM, Cumberland P, *et al*. Prevalence of refractive error in Europe: the European eye epidemiology (E3) Consortium. *Eur J Epidemiol* 2015;30:305–15.
19 Neustaeter A, Nolte I, Snieder H, *et al*. Genetic pre-screening for glaucoma in population-based epidemiology: protocol for a double-blind prospective screening study within lifelines (EyeLife). *BMC Ophthalmol* 2021;21:18.
20 Paudel N, Adhikari S, Thakur A, *et al*. Clinical accuracy of the Nidek ARK-1 Autorefractor. *Optom Vis Sci* 2019;96:407–13.
21 Robin X, Turck N, Hainard A, *et al*. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:77.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

22 Burman P. A comparative study of ordinary cross-validation, *v* -fold cross-validation and the repeated learning-testing methods. *Biometrika* 1989;76:503–14.
23 Jung Y, Hu J. A *K*-fold Averaging Cross-validation Procedure. *J Nonparametr Stat* 2015;27:167–79.
24 Kuhn M. Caret: classification and regression training. R package version 6.0-85, 2021. Available: https://cran.r-project.org/web/packages/caret/caret.pdf [Accessed 6 Dec 2021].
25 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
26 Ghorbani Mojarrad N, Plotnikov D, Williams C, *et al*. Association between polygenic risk score and risk of myopia. *JAMA Ophthalmol* 2020;138:7–13.
27 Williams KM, Verhoeven VJM, Cumberland P, *et al*. Prevalence of refractive error in Europe: the European Eye Epidemiology (E(3)) Consortium. *Eur J Epidemiol* 2015;30:305–15.
28 Eurostat. Revision of the European Standard Population Report of Eurostat's task force, 2013. Available: https://ec.europa.eu/eurostat/documents/3859598/5926869/KS-RA-13-028-EN.PDF/e713fa79-1add-44e8-b23d-5e8fa09b3f8f [Accessed 5 Nov 2020].
29 Cumberland PM, Bao Y, Hysi PG, *et al*. Frequency and distribution of refractive error in adult life: methodology and findings of the UK Biobank study. *PLoS One* 2015;10:e0139780.
30 Attebo K, Ivers RQ, Mitchell P. Refractive errors in an older population. *Ophthalmology* 1999;106:1066–72.
31 Morgan IG, French AN, Ashby RS, *et al*. The epidemics of myopia: aetiology and prevention. *Prog Retin Eye Res* 2018;62:134–49.
32 Zylbermann R, Landau D, Berson D. The influence of study habits on myopia in Jewish teenagers. *J Pediatr Ophthalmol Strabismus* 1993;30:319–22.
33 Ben-Simon GJ, Peiss M, Anis E, *et al*. Spectacle use and reduced unaided vision in third grade students: a comparative study in different educational settings. *Clin Exp Optom* 2004;87:175–9.
34 Bez D, Megreli J, Bez M, *et al*. Association between type of educational system and prevalence and severity of myopia among male adolescents in Israel. *JAMA Ophthalmol* 2019;137:1.
35 Mountjoy E, Davies NM, Plotnikov D, *et al*. Education and myopia: assessing the direction of causality by Mendelian randomisation. *BMJ* 2018;361:k2022.
36 Cuellar-Partida G, Lu Y, Kho PF, *et al*. Assessing the genetic predisposition of education on myopia: a Mendelian randomization study. *Genet Epidemiol* 2016;40:66–72.
37 Wojciechowski R, Hysi PG. Focusing in on the complex genetics of myopia. *PLoS Genet* 2013;9:e1003442.
38 Williams KM, Hysi PG, Nag A, *et al*. Age of myopia onset in a British population-based twin cohort. *Ophthalmic Physiol Opt* 2013;33:339–45.

Br J Ophthalmol: first published as 10.1136/bjophthalmol-2021-319166 on 10 March 2022. Downloaded from http://bjo.bmj.com/ on May 23, 2022 at University of Groningen. Protected by copyright.

Development and validation of a questionnaire-based myopia proxy in adults: The Lifelines Cohort Study

Nigus G. Asefa (PhD),[1] Anna Neustaeter (PhD),[1,2] Jelle Vehof (MD, PhD),[1,2] Ilja M. Nolte (PhD),[1] Harold Snieder (PhD),[1] Nomdo Jansonius (MD, PhD)[2]

[1]University of Groningen, Department of Epidemiology, University Medical Center Groningen, Hanzeplein 1 (9713 GZ), Groningen, The Netherlands

Email addresses: Nigus G. Asefa: n.g.asefa@umcg.nl, Harold Snieder: h.snieder@umcg.nl, Ilja Nolte: i.m.nolte@umcg.nl

[2]University of Groningen, Department of Ophthalmology, University Medical Center Groningen, Hanzeplein 1 (9713 GZ), Groningen, The Netherlands

Email addresses: Anna Neustaeter: a.neustaeter@umcg.nl, Nomdo Jansonius: n.m.jansonius@umcg.nl, Jelle Vehof: j.vehof@umcg.nl

Supplementary materials
Number of Supplementary Figs: 2
Number of Supplementary Tables: 2

Asefa NG, *et al. Br J Ophthalmol* 2022;0:1–8. doi: 10.1136/bjophthalmol-2021-319166

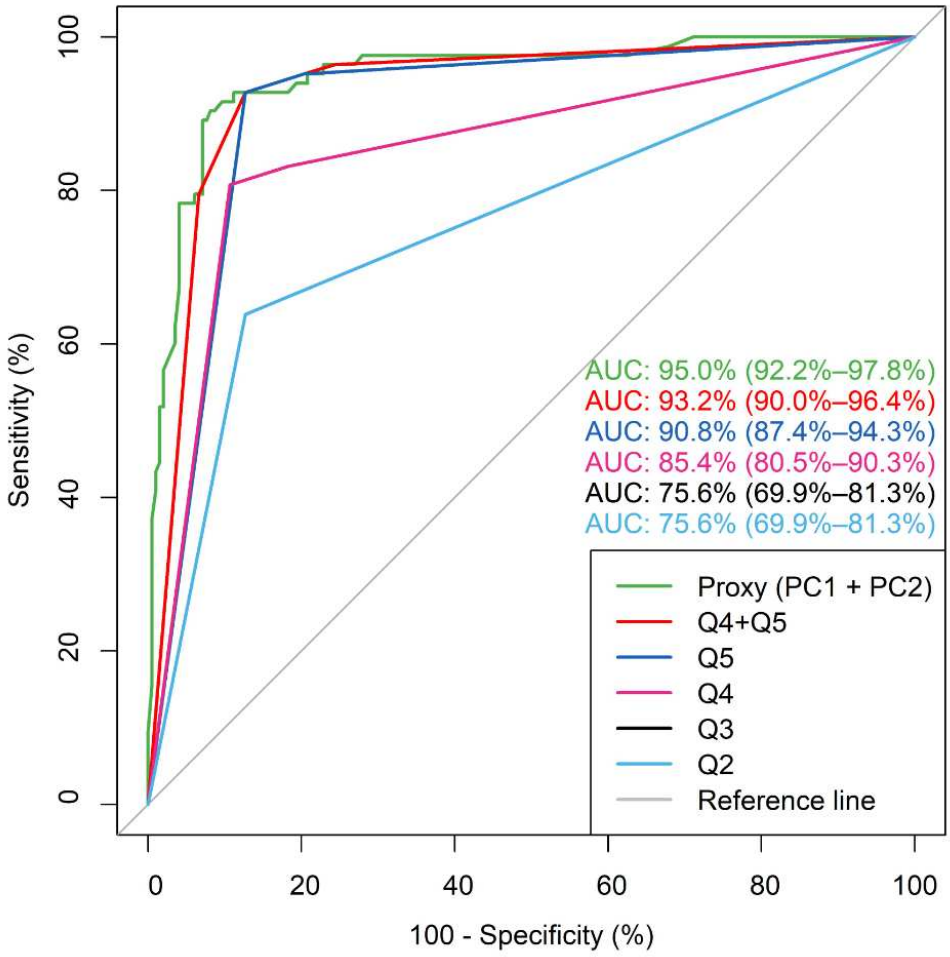

Fig. S1 ROC curve depicting the performance of proxy (PC1 and PC2) and individual or combination of questions in predicting myopia

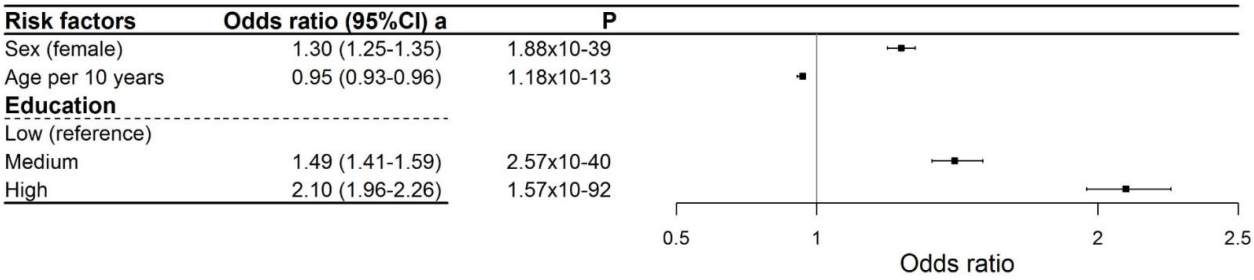

Fig. S2 Association of proxy-inferred myopia with sex and education after adjusting for occupation

[a]*Adjusted for occupation. At the baseline assessment, each Lifelines participant was asked multiple occupation-related questions, including her/his employment status, current (or most recent) occupation, and the number of working hours per week. We coded occupations according to the International Standard Classification of Occupations 2008 system (ISCO-08).*[1] *ISCO-08 classifies occupations into ten major, and 43 sub-major groups, but for this analysis, we used the primary ten major aggregate groups (Managers [1], Professionals[2], Technicians and associate professionals [3], Clerical support workers [4], Service and sales workers [5], Skilled agricultural, forestry and fishery workers [6], Craft and related trades worker [7], Plant and machine operators, and assemblers [8], Elementary occupations [9], Armed forces [10]). The quality of ISCO-08 coding was assessed using the Cascot tool.*[2] *Cascot tool calculates the probability score between zero to 100 (high score high certainty), indicating whether a correct ISCO-08 coding for specific occupation is given. We excluded participants with a probability score < 40.*

Supplementary Table S1 Odds ratios (95% CI) of myopia (< -0.5 D), significance and explained variance (adjusted-R^2[95%CI]) for each individual question and the proxy (*Step 2*, n=280)

Variables	OR (95% CI)*	P	Adjusted-R^2 in percent (95% CI)
Q1. Do you wear glasses, reading glasses or contact lenses?	NA	NA	NA
Q2. Can you see in the distance without glasses or contact lenses (e.g. watching television)?	0.05 (0.01-0.11)	$5.9x10^{-11}$	33.3 (24.5-42.2)
Q3. Can you see close by without glasses or contact lenses (e.g. read a book)?	12.15 (6.67-22.83)	$1.5x10^{-15}$	32.5 (23.6-41.4)
Q4. How old were you when you started wearing glasses or contact lenses?			
Q4.1 ≥ 30 years old vs. 30 years old or No to Q1	0.65 (0.16-4.40)	0.59	54.0 (46.2-61.8)
Q4.2 < 30 years old vs. ≥ 30 years old or 'No' to Q1	23.9 (6.1-159.9)	$6.3x10^{-5}$	
Q5. What was the main reason for buying these glasses or contact lenses?			
Q5.1 Difficulty in seeing at distance vs. difficulty in seeing close by or 'No' to Q1	23.1 (6.0-153.2)	6.6x10-5	66.6 (60.3-72.9)
Q5.2 Difficulty in seeing close by vs. difficulty in seeing at distance or 'No' to Q1	0.19 (0.03-1.46)	0.068	
Proxy			
Being classified as myopic by the proxy	106.1 (43.5-258.3)	$9.9x10^{-25}$	70.0 (64.2-75.8)

******The odds ratios were calculated from a univariate logistic regression of myopia (< -0.5 D) against individual questions and the proxy.*

Supplementary Table S2 Questions used and accuracy of myopia validation analysis at different spherical equivalent thresholds (current versus previous studies)

Variables	Studies				
	Current study (Lifelines)	**Cumberland et al[3] (UK Biobank)**	**Breslin et al[4] (Northern Ireland)**	**Ip et al [5] (Sydney)**	**Walline et al[6] (California)**
Questions used for myopia analysis					
Do you wear glasses or contact lenses?		√		√	√[c]
If yes, are the glasses bifocals?					√[c]
Can you see in the distance without glasses or contact lenses (e.g. watch television)?	√		√[a]		
Can you see close by without glasses or contact lenses (e.g. read a book)?	√		√[a]		
How old were you when you started wearing glasses or contact lenses?	√	√	√[a], √[b]	√	√[c]
What was the main reason for buying those glasses or contact lenses?	√	√		√	√[c]
Are you short-sighted (need spectacles to see far away)			√[b]		√[d]
Are you long-sighted (need spectacles more for close up work e.g. reading, computer)			√[b]		
Do you have myopia?					√[e]
Gold standard, myopia definition, and diagnostic performance					
Measured autorefraction data	√	√	√		√
Prescription paper for glasses				√	
Myopia definition, Dioptre	< −0.5	≤ −1	< 0	≤ −0.5	≤ −0.75
Sensitivity, percent (95% CI)	90.4 (84.3-96.4)	89.1 (88.7-89.4)	83[a] 54[b]	88 (85-92)	54[e] 98[d] 76[c]
Specificity, percent (95% CI)	91.9 (87.8-95.4)	83.7 (83.4-84.0)	93[a] 96[b]	83 (78-87)	83e 48[d] 74[c]

[a]Lay term and [b]Lay term questions combined with questions with more descriptive explanations and the corresponding accuracy estimates (sensitivity or specificity)

[e]Direct, [d]Lay term, and [c]Indirect questions and the corresponding accuracy estimates (sensitivity or specificity) per question

References

1. Choi SB, Yoon J-H, Lee W (2020) The Modified International Standard Classification of Occupations defined by the clustering of occupational characteristics in the Korean Working Conditions Survey. Ind Health 58:132–141
2. Jones R, Elias P (2006) CASCOT: Computer-assisted Structured Coding Tool. Institute for Employment Research.
3. Cumberland PM, Chianca A, Rahi JS, UK Biobank Eye and Vision Consortium. Accuracy and Utility of Self-report of Refractive Error. JAMA Ophthalmol. 2016;134(7):794-801.

4. Breslin KMM, O'Donoghue L, Saunders KJ. An investigation into the validity of self-reported classification of refractive error. Ophthalmic Physiol Opt. 2014;34(3):346-352.
5. Ip J, Robaei D, Rochtchina E, et al. Can information on the purpose of spectacle use and age at first use predict refractive error type? Ophthalmic Epidemiol. 2007;14(2):88-92.
6. Walline JJ, Zadnik K, Mutti DO. Validity of surveys reporting myopia, astigmatism, and presbyopia. Optom Vis Sci. 1996;73(6):376-381.