



# The utility of a type 2 diabetes polygenic score in addition to clinical variables for prediction of type 2 diabetes incidence in birth, youth and adult cohorts in an Indigenous study population

Lauren E. Wedekind<sup>1,2</sup> · Anubha Mahajan<sup>3,4</sup> · Wen-Chi Hsueh<sup>1</sup> · Peng Chen<sup>1,5</sup> · Muideen T. Olaiya<sup>1,6</sup> · Sayuko Kobes<sup>1</sup> · Madhumita Sinha<sup>1</sup> · Leslie J. Baier<sup>1</sup> · William C. Knowler<sup>1</sup> · Mark I. McCarthy<sup>3,4,7</sup> · Robert L. Hanson<sup>1</sup>

Received: 8 August 2022 / Accepted: 29 November 2022 / Published online: 2 March 2023

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023

## Abstract

**Aims/hypothesis** There is limited information on how polygenic scores (PSs), based on variants from genome-wide association studies (GWASs) of type 2 diabetes, add to clinical variables in predicting type 2 diabetes incidence, particularly in non-European-ancestry populations.

**Methods** For participants in a longitudinal study in an Indigenous population from the Southwestern USA with high type 2 diabetes prevalence, we analysed ten constructions of PS using publicly available GWAS summary statistics. Type 2 diabetes incidence was examined in three cohorts of individuals without diabetes at baseline. The adult cohort, 2333 participants followed from age  $\geq 20$  years, had 640 type 2 diabetes cases. The youth cohort included 2229 participants followed from age 5–19 years (228 cases). The birth cohort included 2894 participants followed from birth (438 cases). We assessed contributions of PSs and clinical variables in predicting type 2 diabetes incidence.

**Results** Of the ten PS constructions, a PS using 293 genome-wide significant variants from a large type 2 diabetes GWAS meta-analysis in European-ancestry populations performed best. In the adult cohort, the AUC of the receiver operating characteristic curve for clinical variables for prediction of incident type 2 diabetes was 0.728; with the PS, 0.735. The PS's HR was 1.27 per SD ( $p=1.6 \times 10^{-8}$ ; 95% CI 1.17, 1.38). In youth, corresponding AUCs were 0.805 and 0.812, with HR 1.49 ( $p=4.3 \times 10^{-8}$ ; 95% CI 1.29, 1.72). In the birth cohort, AUCs were 0.614 and 0.685, with HR 1.48 ( $p=2.8 \times 10^{-16}$ ; 95% CI 1.35, 1.63). To further assess the potential impact of including PS for assessing individual risk, net reclassification improvement (NRI) was calculated: NRI for the PS was 0.270, 0.268 and 0.362 for adult, youth and birth cohorts, respectively. For comparison, NRI for HbA<sub>1c</sub> was 0.267 and 0.173 for adult and youth cohorts, respectively. In decision curve analyses across all cohorts, the net benefit of including the PS in addition to clinical variables was most pronounced at moderately stringent threshold probability values for instituting a preventive intervention.

**Conclusions/interpretation** This study demonstrates that a European-derived PS contributes significantly to prediction of type 2 diabetes incidence in addition to information provided by clinical variables in this Indigenous study population. Discriminatory power of the PS was similar to that of other commonly measured clinical variables (e.g. HbA<sub>1c</sub>). Including type 2 diabetes PS in addition to clinical variables may be clinically beneficial for identifying individuals at higher risk for the disease, especially at younger ages.

✉ Lauren E. Wedekind  
Lauren.Wedekind@linacre.ox.ac.uk; wedekindle@nih.gov

<sup>1</sup> Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ, USA

<sup>2</sup> Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>3</sup> Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

<sup>4</sup> Present address: Genentech, San Francisco, CA, USA

<sup>5</sup> College of Basic Medical Sciences, Jilin University, Changchun, China

<sup>6</sup> School of Clinical Sciences, Monash University, Clayton, VIC, Australia

<sup>7</sup> Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Headington, UK

## Research in context

### What is already known about this subject?

- Previous studies, using polygenic scores (PSs) constructed from 15 variants to over six million common variants, have generally found that PSs were significantly associated with type 2 diabetes incidence but contributed little to overall prediction of type 2 diabetes beyond that provided by clinical variables
- Most previous studies were performed in adults (age  $\geq 18$  years) and in European-ancestry populations

### What is the key question?

- What is the utility of type 2 diabetes PSs in identifying individuals at risk for type 2 diabetes, and how do they perform when compared with commonly measured clinical variables?

### What are the new findings?

- Analyses were conducted in an Indigenous population from the Southwestern USA in three age cohorts: adults (age at baseline  $\geq 20$  years), youth (5–19 years) and birth (baseline examination at birth)
- Type 2 diabetes PSs were significantly associated with incident type 2 diabetes in all cohorts; the additional contribution of the PS to the overall prediction was modest
- The contribution of the PS to classification of type 2 diabetes risk—as measured by net reclassification improvement—was similar to that provided by HbA<sub>1c</sub> and BMI

### How might this impact on clinical practice in the foreseeable future?

- Including a type 2 diabetes PS in addition to clinical variables may be beneficial for selecting higher-risk patients for preventive interventions, particularly at younger ages

**Keywords** Clinical prediction · Decision curve analysis · Incidence analysis · Polygenic score · Type 2 diabetes

### Abbreviations

DIAGRAM	Diabetes Genetics Replication And Meta-analysis consortium
DIAMANTE	Diabetes Meta-Analysis of Trans-Ethnic association studies consortium
FPG	Fasting plasma glucose
GWAS	Genome-wide association study
2hPG	2 h plasma glucose
MAF	Minor allele frequency
NRI	Net reclassification improvement
PS	Polygenic score
$p_t$	Threshold probability

## Introduction

Type 2 diabetes-associated genetic variants, derived from genome-wide association studies (GWASs), have largely been reproducible across populations. There is limited information on how polygenic scores (PSs) based on these variants add to clinical variables for predicting type 2 diabetes incidence. Such prediction could help identify individuals at

increased risk of type 2 diabetes for targeted prevention efforts.

Previous studies assessing contributions of a type 2 diabetes PS for prediction of type 2 diabetes incidence have mostly been conducted in European-ancestry populations [1, 2, 3, 4, 5, 6]. These studies, using PSs constructed from 15 variants to over six million common variants, have generally found that PSs were significantly associated with type 2 diabetes incidence but contributed little beyond clinical variables to overall prediction of type 2 diabetes [1, 2, 3, 4, 5, 6].

Previous studies were largely conducted in adults, but the utility of PSs for prediction of subsequent type 2 diabetes may be greater earlier in life (in youth or even at birth). The present study employed a PS for prediction of type 2 diabetes incidence in an Indigenous population from the Southwestern USA with a high prevalence of type 2 diabetes and obesity, and in which long-term follow-up data are available. In this population, the age-adjusted prevalence of diabetes is approximately six times higher than in non-Hispanic white people in the USA [7]. We aimed to analyse how genetic and clinical variables could inform strategies for screening and prevention in three cohorts of individuals in different age groups (birth, youth and adulthood) at baseline.

## Methods

**Study design and participants** A longitudinal study of diabetes (1965–2007) was conducted in an Indigenous study population from the Southwestern USA; methods for this study have been described previously [8]. Before participation, volunteers were fully informed of the nature and purpose of the study and adult participants provided written informed consent, including consent for genetic studies; minor participants provided written assent. Protocols were approved by the institutional review board of the National Institute of Diabetes and Digestive and Kidney Diseases, and research was conducted in accordance with the principles of the Declaration of Helsinki.

Briefly, individuals at least 5 years old were invited for health examinations every 2 years. At each exam, a 75 g oral glucose tolerance test was administered with measurement of HbA<sub>1c</sub> and fasting and 2 h plasma glucose (FPG, 2hPG). Diabetes was diagnosed using 1997 American Diabetes Association criteria (FPG  $\geq 7.0$  mmol/l, 2hPG  $\geq 11.1$  mmol/l or clinical diagnosis) [9]. Height and weight were measured to calculate BMI and birthweight was collected from clinical information and Arizona state birth certificates. Participants had not been directly asked to report parental diabetes; however, since many participants' parents had also participated in the study, we were able to approximate the information that would be available in clinical encounters by using information from direct examination in the parents. We defined parental diabetes using three categories (yes, no or unknown) per parent. Characteristics of participants are summarised in electronic supplementary material (ESM) Tables 1 and 2.

**Genotypic data** Of the study participants, 7701 had genotypes available from previous GWASs, generated using a custom Axiom array designed to capture common variation in members of this community (minor allele frequency (MAF)  $\geq 0.05$ , or  $\geq 0.01$  for coding variants), using methods described previously (Affymetrix, Santa Clara, CA, USA) [10]. Missing and ungenotyped variants were imputed with whole genome sequence data for 266 community members as a reference panel using Impute 2, resulting in 6.6 million variants with MAF  $> 0.01$  and imputation quality score  $> 0.5$  (median 0.95) [11]. Previous work in this population suggests that a population-specific reference panel is optimal for imputing common variants, with little value from including samples from outside populations [12]. Variants were excluded from analyses if they had an imputation quality score  $< 0.5$  or MAF  $< 0.01$  (ESM Method 1).

**Study cohorts** Of the 7701 individuals with genotypes available, we constructed three cohorts based on age at baseline

examination for those who had data for at least two exams with availability of clinical variables. There were 2333 participants followed from first examination in adulthood (age  $\geq 20$  years); 640 cases of type 2 diabetes occurred over 16,686 person-years of follow-up. There were 2229 participants followed from first examination in youth (age 5–19 years); 228 cases of type 2 diabetes occurred over 17,803 person-years of follow-up. There were 2894 participants with birthweight data available who were considered to be followed from birth; 438 cases of type 2 diabetes occurred over 61,591 person-years of follow-up. Individuals were included in multiple cohorts if suitable data were available.

**Construction of type 2 diabetes PSs** We compared associations of ten different constructions of type 2 diabetes PS, derived from GWASs conducted for populations from various world regions. We used 'pruning and thresholding' methods to select variants for the PS, selecting independent genome-wide significant variants from the GWASs for other populations (ESM Tables 3–6). These PSs included the following, each named for the meta-analysis from which it was derived: Diabetes Genetics Replication And Meta-analysis consortium (DIAGRAM) 2018 (constituting 293 variants derived from European populations) [13], Asian Genetic Epidemiology Network consortium (AGEN) 2020 (125 variants derived from East Asian populations) [14] and Diabetes Meta-Analysis of Trans-Ethnic association studies consortium (DIAMANTE) 2022. The seven DIAMANTE PSs are constructions of multi-ancestry PSs (287 variants) with weights taken from meta-analyses of populations representing the following ancestry groups: multi-ancestry, African, East Asian, European, Hispanic/Latino and South Asian [15], in addition to a 'population-specific weight' PS from these same DIAMANTE 2022 variants with weights derived from the present population, using tenfold cross-validation to address overfitting. Finally, we also derived a 'population-specific variant' PS by selecting 287 type 2 diabetes-associated variants from the 515,692 variants typed in the type 2 diabetes GWAS in the study population, using twofold cross-validation (ESM Method 2). While PSs can be constructed using a larger number of variants, by using less stringent significance thresholds or accounting for linkage disequilibrium, applicability across populations with different linkage disequilibrium patterns is uncertain. We, thus, employed the widely used method of selecting significant variants.

We constructed each PS using imputed genotypes available in the present study population. The products of the number of risk alleles for each individual with the effect size (logarithm of the OR) from the corresponding GWAS were summed across variants. PSs were standardised across the entire study population to have mean of 0 and SD of 1: HRs for PSs were expressed in terms of SD of that PS.

**Statistical analyses** Analyses were completed in SAS 9.4 (SAS Institute, Cary, NC, USA). For each cohort, individuals were followed from inception (first examination with clinical data available for youth and adult cohorts; birth for birth cohort) until they developed type 2 diabetes or until their last examination, whichever came first. We evaluated the relative contributions of various combinations of clinical variables and/or the PS in the following analyses: cumulative incidence, survival, AUC of the receiver operating characteristic curve, net reclassification improvement (NRI) and decision curve. Cumulative incidence, survival, decision curve and NRI analyses required calculation of the predicted occurrence of type 2 diabetes at a specified follow-up time for all individuals to ensure comparability: a follow-up of 10 years was used for the adult and youth cohorts, and 30 years for the birth cohort.

The variables available for the adult cohort included: age, sex, parental diabetes, BMI, HbA<sub>1c</sub>, FPG, 2hPG and the type 2 diabetes PS [16]. Those for the youth cohort included: age, sex, parental diabetes, modified BMI *z* score [17], HbA<sub>1c</sub>, FPG, 2hPG and PS. Those for the birth cohort included: sex, parental diabetes, birthweight and PS. This specific set of clinical variables was chosen because the US Preventive Services Task Force focuses on measures of obesity, family history and hyperglycaemia in recommendations for screening and prevention of type 2 diabetes [18]. Most previous studies have assessed prediction with control for a similar set of clinical predictors as used here, but many studies have also included measurement of lipids [1, 2, 3, 4, 5, 6]. Measurements of serum HDL and triglycerides/triacylglycerols were available in a subset of the present cohort (measured since 1993) and adjustment for these variables yielded similar results to those observed in primary analyses (ESM Table 7).

Since HbA<sub>1c</sub> was only measured for examinations after 1989, we conducted additional analyses that did not require HbA<sub>1c</sub> to allow for longer follow-up and greater sample size: these analyses returned similar findings to analyses that included HbA<sub>1c</sub> (ESM Table 8). Given the U-shaped relationship between birthweight and type 2 diabetes in this population [19], we analysed birthweight using two binary variables, one denoting birthweight <3000 g and another denoting birthweight >4000 g. We also conducted analyses that included a continuous birthweight variable and its squared term to capture the quadratic relationship between birthweight and type 2 diabetes. While these analyses gave similar results, the dichotomised birthweight variables yielded a better fit according to Akaike's information criterion.

We also conducted analyses including stated admixture as a covariate; its inclusion returned virtually the same results as without. To further control for population stratification, we conducted additional analyses after adjustment of the PS for the first ten genetic principal components derived from the

GWAS, with separate estimation of principal components in each of the three target cohorts; results were similar to those of the primary analyses (ESM Table 9). Previous studies in this population demonstrated that genetic variants at *KCNQ1* rs2237895 (risk allele frequency=0.49, OR 1.31; exhibits parent-of-origin effects) [20] and *ABCC8* rs1272388614 (risk allele frequency=0.017, OR 2.02) [21] are significantly and strongly associated with type 2 diabetes. We conducted further analyses to assess the contributions of these genotypes in addition to the type 2 diabetes PS for prediction of type 2 diabetes incidence.

**Cumulative incidence and survival analyses** We used Cox proportional hazards regression to evaluate associations of clinical variables and PS with type 2 diabetes incidence. Cumulative incidence of type 2 diabetes was calculated as the proportion of individuals that developed type 2 diabetes over the specified follow-up time, using Breslow's method (PROC PHREG in SAS). To assess separate contributions of PS and clinical risk, we calculated predicted cumulative incidence according to different levels of PS and of clinical risk, as determined by linear predictors from the clinical variables in the proportional hazards model.

**AUC analyses** We compared the Harrell's C statistic [22] of models that included clinical variables alone with the C statistic of those that included clinical variables and the PS. The C statistic expresses the probability within a pair of individuals, one who developed type 2 diabetes and one who did not, that the individual who developed type 2 diabetes had a higher predicted probability of doing so [23]. In the context of survival analysis (e.g. in the proportional hazards models used here), the C statistic is equivalent to the AUC of the receiver operating characteristic curve [23], and we refer to it as 'AUC' throughout the manuscript.

**NRI analyses** Continuous-variable NRI quantifies the amount of correct reclassification introduced by using a model with an additional variable [24]. We analysed NRI by calculating the net proportion of events reclassified correctly (assigned a higher probability value) plus the net proportion of nonevents reclassified correctly (assigned a lower probability value) [25]. Confidence intervals for the NRI were calculated by a bootstrap method.

**Decision curve analyses** We employed decision-analytic methods to assess consequences of clinical decisions and expected outcomes of alternative clinical management (i.e. including various combinations of clinical variables with and without the PS in prediction models). These analyses assume that the threshold probability ( $p_t$ ) of developing type 2



diabetes at which one would opt for an intervention is informative of how one weighs the relative benefits and harms of true-positive and false-positive predictions, and the net benefit of using a predictive model to select individuals above a given  $p_t$  is calculated accordingly [26]. We used extensions to decision curve methods for survival analysis to plot net benefit across a range of  $p_t$  values to evaluate for which  $p_t$  ranges and what corresponding proportion of the population the PS had marginal net benefit [27].

**Comparisons of associations of PSs among cohorts** To compare the effects of the PSs (e.g. HRs) among the different age cohorts, a bootstrap analysis was conducted as previously described [28]. In brief, the 4770 individuals who were included in at least one cohort were resampled 2000 times, and the analyses were repeated for each iteration. The resulting differences in the logarithm of the HR between each pair of cohorts and their standard errors were calculated and used to test statistical significance of the differences. Since the availability and predictive power of different clinical covariates may affect the HR estimates, these analyses were conducted without any covariates.

## Results

**Ten constructions of type 2 diabetes PSs** All ten PSs for type 2 diabetes, constructed using the overlap of published type 2 diabetes GWAS summary statistics and genotypes available in this study population, had significant associations with type 2 diabetes incidence in the study population. HRs for the PSs in models adjusted for clinical variables (age, sex, BMI, FPG, HbA<sub>1c</sub> and parental diabetes for the adult cohort; age, sex, modified BMI  $z$  score, FPG, HbA<sub>1c</sub> and parental diabetes for the youth cohort; and sex, birthweight and parental diabetes for the birth cohort) ranged from 1.13 to 1.27 per SD for the adult cohort, from 1.19 to 1.49 for the youth cohort and from 1.27 to 1.48 for the birth cohort (ESM Table 10). The PS that consistently had the strongest associations with type 2 diabetes incidence (largest HRs) was constructed using the DIAGRAM 2018 GWAS. Thus, for the rest of this text, we present results for the DIAGRAM 2018 PS. Calibration plots for models for this PS for each of the three cohorts show that these models are well-calibrated (ESM Fig. 1). The DIAMANTE 2022 multi-ancestry PS and the population-specific variant PS also had strong associations with type 2 diabetes incidence, though not as strong (ESM Tables 10–12 and ESM Figs 2–7).

**Association of PS with incidence of type 2 diabetes** The best-performing PS was significantly associated with type 2 diabetes incidence in adult, youth and birth cohorts (Fig. 1). In the

adult cohort, 10 year cumulative incidence of type 2 diabetes in the lowest decile of PS was 20.5%; in the highest, 42.5% (unadjusted HR=1.31 per SD,  $p=6.9 \times 10^{-11}$ ). In the youth cohort, 10 year cumulative incidence of type 2 diabetes in the lowest decile of PS was 2.4%; in the highest, 21.5% (HR=1.59 per SD,  $p=6.8 \times 10^{-12}$ ). In the birth cohort, 30 year cumulative incidence of type 2 diabetes in the lowest decile of PS was 15.1%; in the highest, 37.3% (HR=1.47 per SD,  $p=1.7 \times 10^{-15}$ ). The clinical predictors were also strongly associated with incidence of type 2 diabetes (ESM Fig. 8).

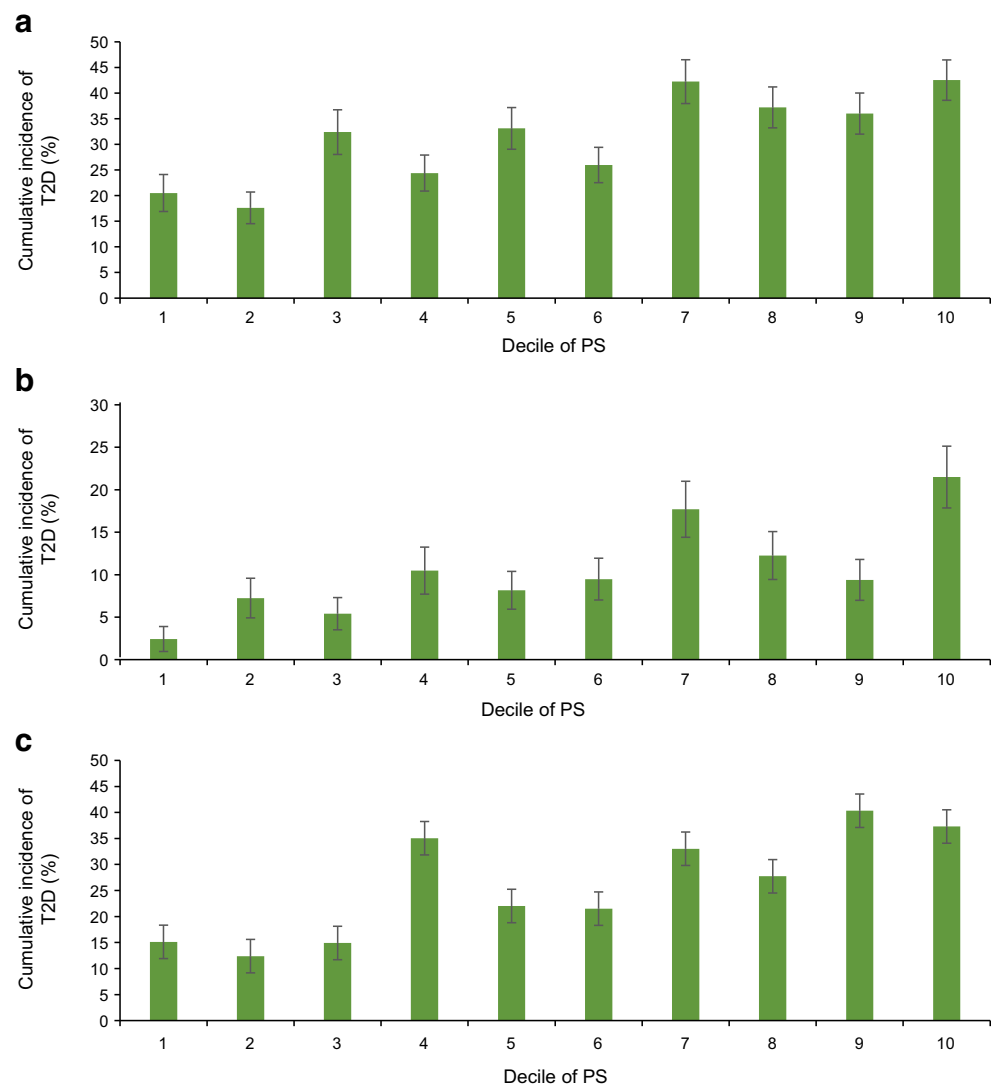
**Survival analyses with adjustment for clinical predictors** We conducted survival analyses to assess associations of individual clinical variables and the PS with type 2 diabetes incidence. In the adult cohort, in the model with clinical variables, the HR of the PS was 1.27 per SD ( $p=1.6 \times 10^{-8}$ ; 95% CI 1.17, 1.38; Table 1). In the youth cohort, in the model with clinical variables, the HR of the PS was 1.49 ( $p=4.3 \times 10^{-8}$ ; 95% CI 1.29, 1.72) (Table 2). In the birth cohort, in the model with clinical variables, the HR of the PS was 1.48 ( $p=2.8 \times 10^{-16}$ ; 95% CI 1.35, 1.63) (Table 3). Adding 2hPG to adult and youth cohorts' models did not substantially alter the HRs of the PS. In general, the HRs associated with clinical variables were only modestly affected with the addition of the PS.

**AUC analyses** We conducted AUC analyses to evaluate the predictive accuracy of models containing combinations of clinical variables and the PS. In the adult cohort, the AUC for the model with age and sex was 0.590 (95% CI 0.566, 0.615); with the PS, 0.619 (95% CI 0.596, 0.643); the difference in AUC (i.e.  $\Delta$ AUC) was 0.029 ( $p=0.003$ ). In the youth cohort, corresponding AUCs were 0.625 (95% CI 0.587, 0.663) and 0.682 (95% CI 0.648, 0.716); the  $\Delta$ AUC was 0.057 ( $p=3.96 \times 10^{-4}$ ). In the birth cohort, AUC for the model with sex was 0.537 (95% CI 0.512, 0.562); with the PS, 0.638 (95% CI 0.610, 0.666); the  $\Delta$ AUC was 0.101 ( $p<10^{-5}$ ).

Though the PS was strongly associated with incident type 2 diabetes, the improvement in AUC compared with clinical variables alone was modest. In the adult cohort, AUC for the full clinical model was 0.728 (95% CI 0.706, 0.750); with the PS, 0.735 (95% CI 0.714, 0.757); and the  $\Delta$ AUC was 0.007 ( $p=0.023$ ) (Table 1). In the youth cohort, AUC for the full clinical model was 0.805 (95% CI 0.778, 0.832); with the PS, 0.812 (95% CI 0.785, 0.839); and the  $\Delta$ AUC was 0.007 ( $p=0.173$ ) (Table 2). For the birth cohort, the increment in AUC with addition of the PS was greater: the AUC for the model including clinical variables was 0.613 (95% CI 0.582, 0.644); with the PS, 0.685 (95% CI 0.657, 0.713); the  $\Delta$ AUC was 0.071 ( $p<10^{-5}$ ) (Table 3).

**NRI analyses** While AUC provides a measure of overall predictive accuracy, it does not fully capture the extent to which addition of a variable can affect individual risk

**Fig. 1** Cumulative incidence of type 2 diabetes by decile of the DIAGRAM 2018 PS. The PS was significantly associated with type 2 diabetes incidence in adult, youth and birth cohorts. **(a)** Cumulative incidence in the adult cohort at 10 year follow-up; at 10 years, 504 individuals had developed type 2 diabetes and 635 remained at risk. **(b)** Cumulative incidence in the youth cohort at 10 year follow-up; at 10 years, 152 individuals had developed type 2 diabetes and 745 remained at risk. **(c)** Cumulative incidence in the birth cohort at 30 year follow-up; at 30 years, 340 individuals had developed type 2 diabetes and 474 remained at risk. T2D, type 2 diabetes



estimates. To examine this, we calculated predicted cumulative incidence of type 2 diabetes according to PS for various levels of clinical risk. Across all cohorts, greater type 2 diabetes PS and greater percentiles of clinical linear predictor were both directly and separately associated with predicted cumulative incidence of type 2 diabetes (Fig. 2).

To further quantify the contribution of each variable to the model's risk classification, we calculated the NRI of each variable. NRI quantifies the extent to which type 2 diabetes cases and non-cases are consequently reclassified upon inclusion of an additional variable. The NRI for adding the PS to clinical variables was 0.270 (95% CI 0.149, 0.392; 0.092 for events, 0.178 for nonevents) in the adult cohort (Table 1); in the youth cohort, 0.268 (95% CI 0.073, 0.464; 0.085 for events, 0.183 for nonevents) (Table 2); in the birth cohort, 0.362 (95% CI 0.222, 0.502; 0.106 for events, 0.256 for nonevents) (Table 3). In comparison, the NRI for HbA<sub>1c</sub> was 0.267 in the adult cohort and 0.173 in the youth cohort.

**Additional genotypic analyses** The effects of some variants strongly associated with type 2 diabetes in this Indigenous study population were not captured in the DIAGRAM 2018 PS. To address this, we assessed the contribution of genotypes for *KCNQ1* rs2237895 (which exhibits parent-of-origin effects) and *ABCC8* rs1272388614 in the adult cohort. For each genotype, associations were significant; however, they contributed modestly to the model of clinical variables and the PS, as assessed by AUC and NRI analyses (ESM Table 13).

**Decision curve analyses** We employed decision curve analyses to estimate the net benefit of including the PS at a range of threshold probabilities (i.e. minimum probabilities of disease that would warrant intervention). When the costs of false-positives are low (i.e. as  $p_t$  approaches 0), population-wide interventions may be favoured; thus, screening by clinical or genetic means would have little net benefit. When false-positive costs are higher (i.e. at higher  $p_t$  values), net clinical

**Table 1** Results of survival, AUC and NRI analyses for the adult cohort

Variable	Clinical variables (AUC=0.728)			Clinical variables + PS (AUC=0.735)			Clinical variables + 2hPG (AUC=0.760)			Clinical variables + 2hPG + PS (AUC=0.765)		
	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI
Age (decades)	1.01 (0.93, 1.10)	0.869	0.011	1.04 (0.95, 1.13)	0.393	0.067	1.00 (0.92, 1.09)	0.992	-0.045	1.03 (0.94, 1.12)	0.540	0.112
Sex (F/M)	1.35 (1.14, 1.59)	$5.48 \times 10^{-4}$	0.115	1.34 (1.13, 1.58)	$7.70 \times 10^{-4}$	0.116	1.13 (0.953, 1.34)	0.157	0.110	1.13 (0.954, 1.35)	0.154	0.106
Mother diabetic/NomDb	1.52 (1.21, 1.90)	$2.75 \times 10^{-4}$	0.161	1.46 (1.17, 1.83)	$2.08 \times 10^{-3}$	0.140	1.43 (1.14, 1.80)	$2.60 \times 10^{-3}$	0.178	1.38 (1.10, 1.73)	0.0131	0.126
Mother unknown/NomDb	1.45 (1.15, 1.82)		1.38 (1.09, 1.74)			1.39 (1.10, 1.74)				1.33 (1.06, 1.67)		
Father diabetic/NomDb	1.45 (1.06, 1.96)		1.38 (1.01, 1.87)			1.38 (1.02, 1.88)				1.33 (0.975, 1.81)		
Father unknown/NomDb	1.20 (0.921, 1.56)		1.17 (0.900, 1.52)			1.20 (0.93, 1.56)				1.17 (0.900, 1.52)		
BMI (kg/m <sup>2</sup> )	1.02 (1.01, 1.03)	$1.22 \times 10^{-4}$	0.158	1.03 (1.01, 1.03)	$1.03 \times 10^{-5}$	0.201	1.02 (1.01, 1.03)	$2.11 \times 10^{-5}$	0.242	1.03 (1.02, 1.04)	$1.88 \times 10^{-6}$	0.269
FPG (mmol/l)	1.99 (1.70, 2.33)	$2.78 \times 10^{-17}$	0.303	1.95 (1.66, 2.29)	$3.13 \times 10^{-16}$	0.295	1.44 (1.22, 1.71)	$2.51 \times 10^{-4}$	0.176	1.44 (1.21, 1.71)	$2.93 \times 10^{-5}$	0.191
HbA <sub>1c</sub> (mmol/mol)	1.08 (1.06, 1.10)	$1.18 \times 10^{-18}$	0.272	1.08 (1.06, 1.10)	$1.14 \times 10^{-17}$	0.267	1.06 (1.05, 1.08)	$9.64 \times 10^{-13}$	0.215	1.06 (1.04, 1.08)	$5.67 \times 10^{-12}$	0.223
2hPG (mmol/l)	-	-	-	-	-	-	1.32 (1.25, 1.40)	$3.74 \times 10^{-24}$	0.420	1.31 (1.24, 1.38)	$6.71 \times 10^{-23}$	0.426
PS (SD)	-	-	-	1.27 (1.17, 1.38)	$1.61 \times 10^{-8}$	0.270	-	-	-	1.24 (1.15, 1.35)	$2.90 \times 10^{-7}$	0.252

‘Clinical variables’ refers to: age, sex, parental diabetes, BMI and FPG for the adult cohort

NRI is calculated for each predictor by comparing the model including the predictor with a model that does not include the predictor

Db, diabetic with regard to parental diabetes; F, female; M, male

**Table 2** Results of survival, AUC and NRI analyses for the youth cohort

Variable	Clinical variables (AUC=0.805)			Clinical variables + PS (AUC=0.812)			Clinical variables + 2hPG (AUC=0.820)			Clinical variables + 2hPG + PS (AUC=0.825)		
	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI
Age (decades)	2.11 (1.46, 3.04)	$6.50 \times 10^{-5}$	0.373	2.21 (1.53, 3.18)	$2.02 \times 10^{-5}$	0.372	2.15 (1.49, 3.11)	$4.96 \times 10^{-5}$	0.367	2.26 (1.57, 3.27)	$1.37 \times 10^{-5}$	0.382
Sex (F/M)	1.40 (1.06, 1.84)	0.0170	0.046	1.32 (1.01, 1.74)	0.0458	0.046	1.09 (0.816, 1.45)	0.562	0.020	1.04 (0.78, 1.39)	0.794	0.032
Mother diabetic/NomDb	2.46 (1.83, 3.29)	$1.21 \times 10^{-11}$	0.524	2.27 (1.70, 3.05)	$1.28 \times 10^{-9}$	0.514	2.26 (1.68, 3.04)	$2.38 \times 10^{-9}$	0.457	2.10 (1.56, 2.83)	$1.29 \times 10^{-7}$	0.472
Mother unknown/NonDb	1.84 (1.27, 2.68)			1.87 (1.29, 2.72)			1.82 (1.26, 2.64)			1.83 (1.26, 2.65)		
Father diabetic/NonDb	1.88 (1.27, 2.78)			1.72 (1.16, 2.55)			1.74 (1.18, 2.57)			1.61 (1.09, 2.39)		
Father unknown/NonDb	1.01 (0.740, 1.37)			0.972 (0.714, 1.32)			1.01 (0.743, 1.37)			0.977 (0.719, 1.33)		
Modified BMI z score	1.50 (1.37, 1.66)	$1.43 \times 10^{-16}$	0.607	1.53 (1.39, 1.69)	$4.48 \times 10^{-17}$	0.587	1.44 (1.30, 1.59)	$6.63 \times 10^{-13}$	0.523	1.47 (1.33, 1.63)	$1.39 \times 10^{-13}$	0.578
FPG (mmol/l)	2.32 (1.66, 3.23)	$7.87 \times 10^{-7}$	0.261	2.06 (1.48, 2.88)	$2.10 \times 10^{-5}$	0.220	1.46 (1.01, 2.10)	0.0435	0.004	1.31 (0.904, 1.88)	0.155	-0.031
HbA <sub>1c</sub> (mmol/mol)	1.06 (1.03, 1.10)	$2.77 \times 10^{-4}$	0.182	1.06 (1.03, 1.09)	$6.29 \times 10^{-4}$	0.173	1.05 (1.01, 1.08)	$6.19 \times 10^{-3}$	0.099	1.04 (1.01, 1.08)	0.00910	0.077
2hPG (mmol/l)	–	–	–	–	–	–	1.35 (1.21, 1.49)	$1.71 \times 10^{-8}$	0.303	1.34 (1.21, 1.48)	$4.20 \times 10^{-8}$	0.310
PS (SD)	–	–	–	1.49 (1.29, 1.72)	$4.31 \times 10^{-8}$	0.268	–	–	–	1.48 (1.28, 1.71)	$1.06 \times 10^{-7}$	0.277

'Clinical variables' refers to: age, sex, parental diabetes, modified BMI z score and FPG for the youth cohort

NRI is calculated for each predictor by comparing the model including the predictor with a model that does not include the predictor

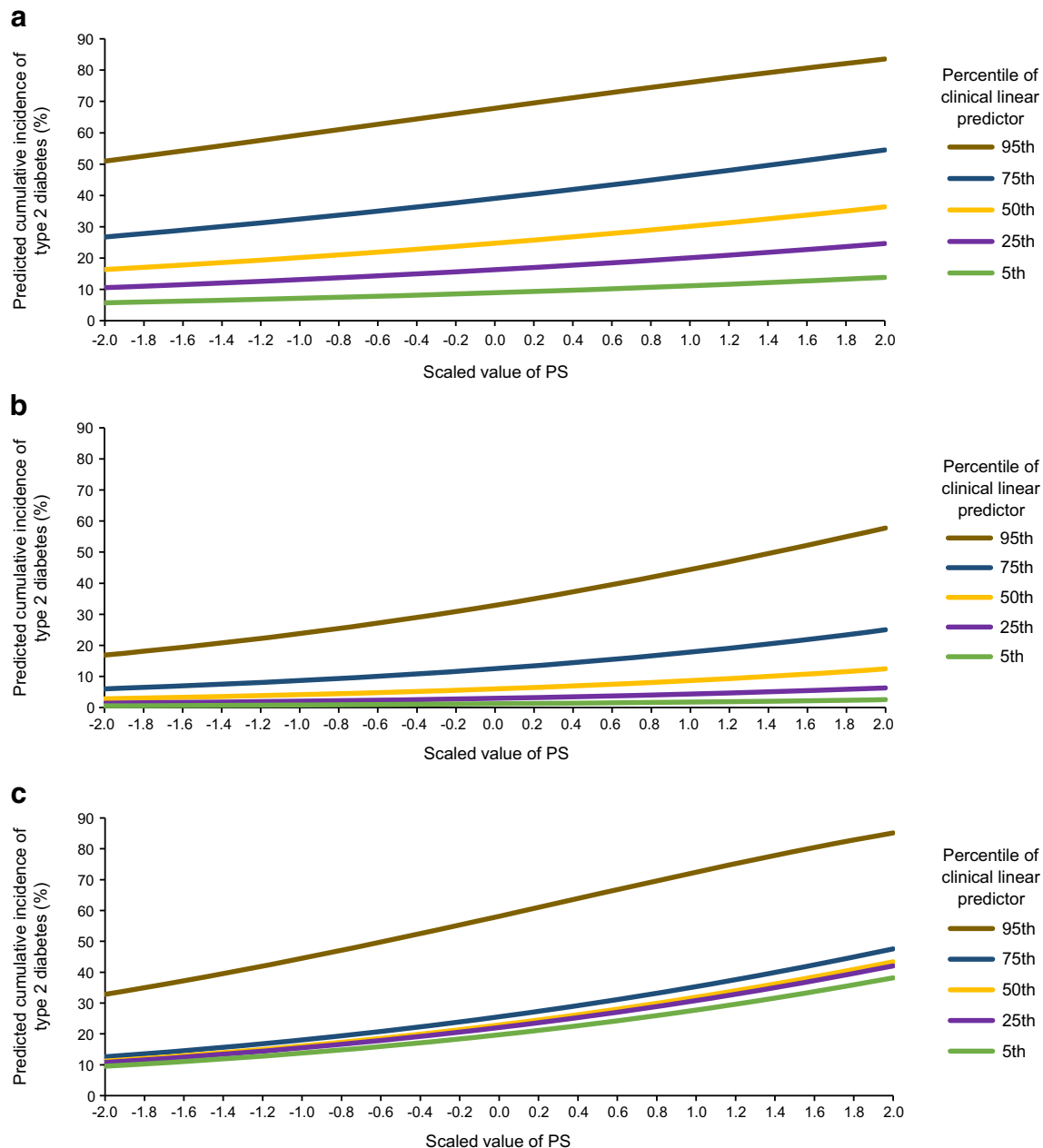
Db, diabetic with regard to parental diabetes; F, female; M, male



**Table 3** Results of survival, AUC and NRI analyses for the birth cohort

Variable	Sex, parental diabetes (AUC=0.597)			Sex, parental diabetes + PS (AUC=0.683)			Sex, parental diabetes, birthweight (AUC=0.613)			Sex, parental diabetes, birthweight + PS (AUC=0.685)		
	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI	HR (95% CI)	p value	NRI
Sex (F/M)	1.18 (0.974, 1.43)	0.0915	0.090	1.18 (0.972, 1.43)	0.0947	0.090	1.15 (0.951, 1.40)	0.146	0.090	1.14 (0.939, 1.38)	0.184	0.090
Mother diabetic/NonDb	7.28 (5.00, 10.6)	$6.79 \times 10^{-30}$	0.212	7.86 (5.39, 11.5)	$1.12 \times 10^{-30}$	0.209	7.40 (5.08, 10.77)	$4.00 \times 10^{-30}$	0.349	7.95 (5.44, 11.6)	$5.20 \times 10^{-31}$	0.317
Mother unknown/NonDb	0.871 (0.711, 1.07)			0.887 (0.724, 1.09)			0.875 (0.715, 1.07)			0.886 (0.723, 1.09)		
Father diabetic/NonDb	3.93 (2.36, 6.54)			3.98 (2.39, 6.64)			3.98 (2.38, 6.65)			4.05 (2.42, 6.78)		
Father unknown/NonDb	1.01 (0.818, 1.26)			1.07 (0.863, 1.33)			0.977 (0.787, 1.21)			1.03 (0.829, 1.28)		
Low birthweight	–	–	–	–	–	–	1.56 (1.24, 1.96)	$6.11 \times 10^{-4}$	0.164	1.16 (0.877, 1.53)	$5.99 \times 10^{-4}$	0.164
High birthweight	–	–	–	–	–	–	1.17 (0.887, 1.55)		–	1.57 (1.25, 1.97)		–
PS (SD)	–	–	–	1.48 (1.35, 1.63)	$2.83 \times 10^{-16}$	0.345	–	–	–	1.48 (1.35, 1.63)	$2.77 \times 10^{-16}$	0.362

NRI is calculated for each predictor by comparing the model including the predictor with a model that does not include the predictor  
 Db, diabetic with regard to parental diabetes; F, female; M, male



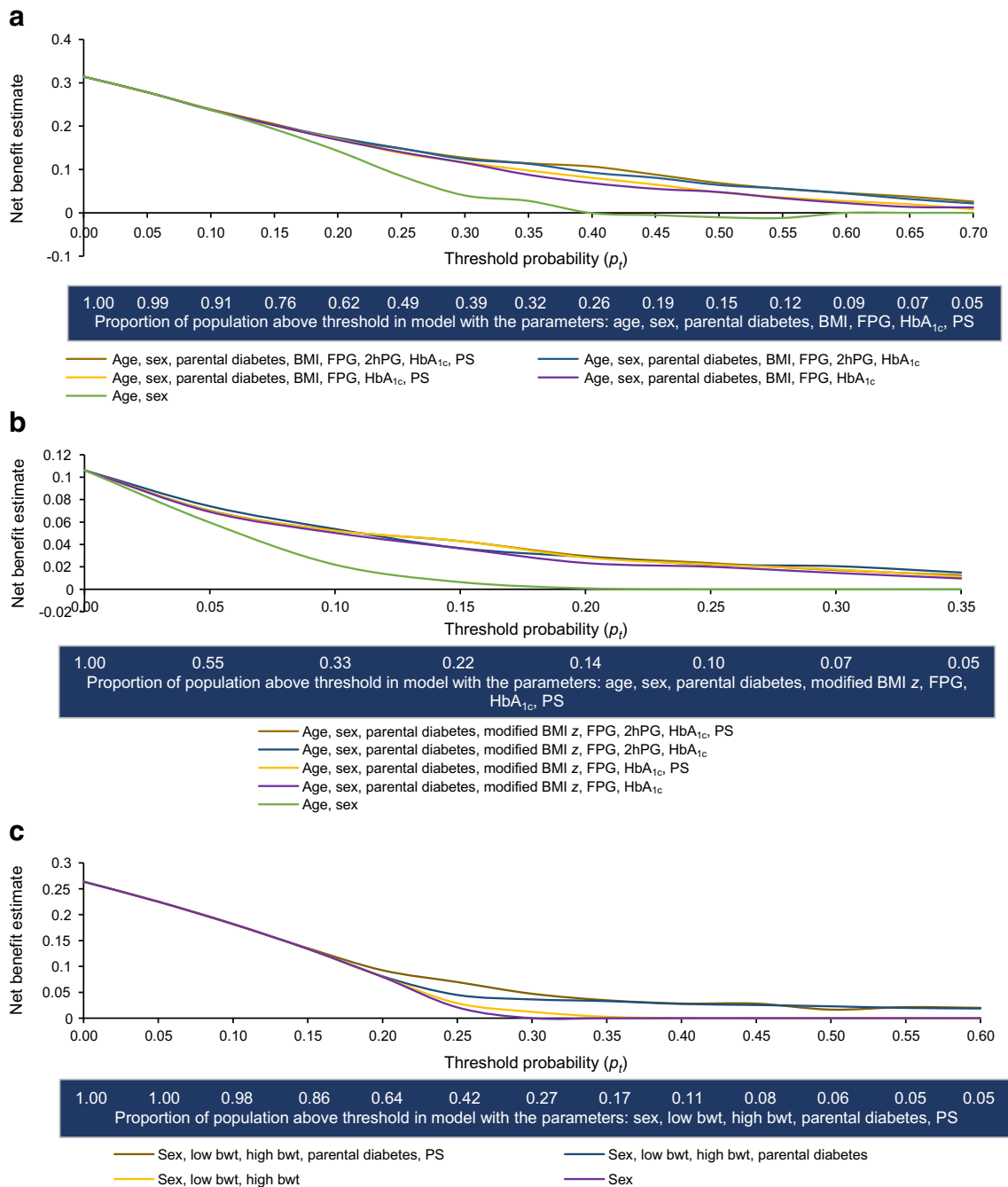
**Fig. 2** Predicted cumulative incidence of type 2 diabetes for the scaled DIAGRAM 2018 PS and specified percentiles of the clinical linear predictor. Cumulative incidence was calculated for various combinations of the PS and clinical risk (based on the linear predictor derived from the clinical variables in the models). Across all cohorts, greater type 2 diabetes PS and greater percentiles of clinical linear predictor were both directly and separately associated with predicted cumulative incidence of type 2 diabetes. **(a)** Predicted cumulative incidence of type 2 diabetes over 10 years of follow-up in the adult cohort: clinical variables include age, sex,

parental diabetes, BMI, FPG, HbA<sub>1c</sub>. At 10 year follow-up, 504 individuals had developed type 2 diabetes and 635 remained at risk. **(b)** Predicted cumulative incidence of type 2 diabetes over 10 years of follow-up in the youth cohort: clinical variables include age, sex, parental diabetes, modified BMI *z* score, FPG, HbA<sub>1c</sub>. At 10 year follow-up, 152 had developed type 2 diabetes and 745 remained at risk. **(c)** Predicted cumulative incidence of type 2 diabetes over 30 years of follow-up in the birth cohort: clinical variables include sex, parental diabetes, birth weight. At 30 year follow-up, 340 had developed type 2 diabetes and 474 remained at risk

benefit can be increased by screening to target the intervention to higher-risk individuals.

In the adult cohort, the net benefit of including the PS in addition to clinical variables was most pronounced at  $p_t$  values 0.3 to 0.5 (up to 18% improvement); this corresponded to 15–40% of the highest-risk individuals

selected for the intervention (Fig. 3). In the youth cohort, the net benefit of including the PS was most pronounced at  $p_t$  values 0.05 to 0.35 (up to 21% improvement) (Fig. 3). In the birth cohort, the net benefit of including the PS was most pronounced at  $p_t$  values 0.15 to 0.35 (up to 56% improvement) (Fig. 3).



**Fig. 3** Net benefit of predictive models with or without the DIAGRAM 2018 PS for type 2 diabetes prediction. For each model in each cohort, the net benefit is plotted on the y-axis against the  $p_t$  for implementing an intervention on the x-axis. The proportion of the population that would be selected at each  $p_t$  is shown below the x-axis. (a) In the adult cohort, the net benefit of including the PS in addition to clinical variables was most

pronounced at  $p_t$  values 0.3–0.5 (up to 18% improvement); this corresponded to 15–40% of the highest-risk individuals selected for the intervention. (b) In the youth cohort, the net benefit of including the PS was most pronounced at  $p_t$  values 0.05–0.35 (up to 21% improvement). (c) In the birth cohort, the net benefit of including the PS was most pronounced at  $p_t$  values 0.15–0.35 (up to 56% improvement). Bwt, birthweight

### Discussion

PSs potentially have utility for identification of individuals with higher risk of type 2 diabetes. Previous studies generally reported significant associations between type 2 diabetes PS and diabetes incidence and modest prediction improvement as

measured by AUC:  $\Delta$ AUC from 0.005 to 0.02 [1, 2, 3, 4, 5, 6]. A limited number of studies include measures of reclassification: continuous NRIs ranged from 0.044 to 0.285 [4, 5, 6]. Most previous studies have been done in European-ancestry populations, but some have been done in non-European populations, including Korean [5], African American [29] and

Iranian [30]. Findings in these populations have generally been similar to those in European-ancestry groups. In the present study, the DIAGRAM 2018 PS was strongly statistically significant in predicting type 2 diabetes incidence in adult, youth and birth cohorts in an Indigenous study population from the Southwestern USA.

Results of AUC analyses are consistent with findings of previous studies: improvement in prediction contributed by the type 2 diabetes PS was statistically significant but modest. However,  $\Delta$ AUC does not fully capture the contribution of a single variable to individual risk [31]. We calculated NRIs for individual variables to address this limitation. NRIs for the PS across all cohorts ranged from 0.2 to 0.3, which is considered intermediate power for identifying type 2 diabetes risk [24], and were comparable to those of commonly measured clinical variables (e.g. HbA<sub>1c</sub> and FPG). Our findings are consistent with the evidence that for most chronic diseases PSs generally provide additional predictive information beyond that provided by traditional risk factors [32].

**Implications of decision curve analyses** Ultimately, clinical utility may depend on how the type 2 diabetes PS affects the decision to implement preventive interventions. Across adult, youth and birth cohorts, results of our decision curve analyses suggest modest increases in clinical benefit for using the PS at moderately stringent  $p_t$  values. There are few data on optimal  $p_t$  values for type 2 diabetes prevention: they depend upon preferences of individual patients and clinicians, healthcare system characteristics and the nature of the interventions considered. Many clinicians would recommend lifestyle prevention for individuals with impaired glucose regulation (e.g. FPG  $\geq 5.5$  mmol/l or HbA<sub>1c</sub>  $\geq 39$  mmol/mol (5.7%)); in the adult cohort, the prevalence of impaired glucose regulation at baseline was 35%, and this would correspond to  $p_t=0.32$  (0.21–0.49 based on cumulative incidence). This is in the range in which our analyses suggest meaningful, albeit modest, improvement in clinical benefit from incorporating the type 2 diabetes PS.

Decision curve analysis assumes that the intervention will be equally effective regardless of how risk is determined. There are limited data on how type 2 diabetes PS affects response to preventive interventions. However, a study within the Diabetes Prevention Program Outcomes Study suggested that lifestyle and metformin interventions were both effective, even in those with greater type 2 diabetes PS [33].

**Construction of PS** While all type 2 diabetes PSs we examined were significantly associated with type 2 diabetes incidence across all cohorts, the DIAGRAM 2018 PS, derived from European-ancestry populations, performed slightly better than the others. While we have previously shown modest heterogeneity in effects of established type 2 diabetes variants

between Europeans and this study population [7], the DIAGRAM 2018 PS even out-performed a population-specific variant PS with a comparable number of variants, derived by twofold cross-validation in the present population ( $n \approx 3850$ ). The expectation is that a PS derived from a GWAS in a more closely matched ancestry group would perform better than one from a different ancestry group, if GWAS sample sizes are equal [34], but PSs derived in a large European-ancestry group can outperform ancestry-specific PSs when the sample size available for deriving the ancestry-specific PS is small [15]. In the present study, the DIAGRAM 2018 PS likely performed well due to the large sample size and extensive fine-mapping in the DIAGRAM type 2 diabetes meta-analysis. Achieving adequate sample sizes for GWASs to derive ancestry-specific PSs in Indigenous study populations is challenging, but many Indigenous populations have extensive linkage disequilibrium which may facilitate the ability of PSs to capture causal variants [35]. While further work is needed to optimise type 2 diabetes PSs across diverse populations, the present study suggests that PSs constructed using results of GWASs in larger populations may be suitable for translation across study populations in which well-powered GWASs are not available. Studies in additional populations are needed.

**Optimal age for preventive interventions** Genetic effects of the PSs with respect to type 2 diabetes incidence were greatest in the youth and birth cohorts. This is consistent with the hypothesis that genetic effects for many chronic diseases are strongest earlier in life [36], and consistent with the finding that familial recurrence risk of diabetes in this population is higher when it occurs at younger ages [37]. The present findings could also reflect the limited availability of phenotypic data for study participants at birth or young ages. However, when analysed without any clinical covariates, the HRs associated for the birth cohort (HR=1.47) and the youth cohort (HR=1.59) were significantly higher than that for the adult cohort (HR=1.31); tests for differences in the HRs between the adult and youth cohorts and adult and birth cohorts yielded  $p=0.037$  and  $p=0.006$ , respectively, while differences between birth and youth cohorts were not significant ( $p=0.15$ ). The improvements in AUC and net benefit upon adding the PS to clinical variables were greatest in the birth cohort. The use of type 2 diabetes PS at birth could be particularly beneficial as phenotypic manifestations of risk (e.g. hyperglycaemia and obesity) are less apparent. However, some relevant clinical measures that may be readily obtained at birth (e.g. birth length for calculation of adiposity measures) were not available in the present study. In adults, there is strong evidence that type 2 diabetes can be prevented by lifestyle modification, pharmacologic treatment or bariatric surgery, but there are few data on the efficacy of preventive

efforts initiated in youth or infancy [38]. Thus, while our analyses suggest that the type 2 diabetes PS has the strongest contribution to prediction of type 2 diabetes incidence in the birth cohort, adults may be a more appropriate target population for preventive interventions in the near term.

**Future research** This study shows that type 2 diabetes PSs, as currently constructed, can provide utility for assessing type 2 diabetes risk; as measured by NRI analyses, information from the PS for classifying type 2 diabetes risk is comparable to that from widely used clinical variables (e.g. HbA<sub>1c</sub> and BMI) in this study population. Further optimisation of the PS is expected to provide better prediction in the future [30]. Such investigations could assess whether differences in population genetic characteristics, obesity and incidence of type 2 diabetes are paralleled by differences in performance of PSs. Results from the present study were derived from an Indigenous population from the Southwestern USA with a relatively high prevalence of type 2 diabetes.

Beyond the scientific issues, however, technical, logistical and cultural issues need consideration before PSs can be incorporated into clinical practice. For example, advances in laboratory methods and informatics are required to make PSs and risk algorithms available to clinicians and patients. Health economics studies are needed to investigate which clinical settings and constructions of type 2 diabetes PS would maximise net benefit for prediction of type 2 diabetes incidence. With such knowledge, more informed decisions about the use of genetic information in prevention of type 2 diabetes could be made.

**Supplementary Information** The online version of this article (<https://doi.org/10.1007/s00125-023-05870-2>) contains peer-reviewed but unedited supplementary material.

**Acknowledgements** We thank the study participants and research staff who made the longitudinal and present studies possible. This study used computational resources of the Biowulf system at the National Institutes of Health (NIH), Bethesda, MD, USA.

**Data availability** The data that support the findings of this study are not publicly available due to privacy concerns. Data may be made available upon reasonable request; for more information, refer to dbGAP (<https://www.ncbi.nlm.nih.gov/gap/>) accession number phs002490.v1.p1.

**Funding** The NIH Oxford-Cambridge Scholars Program provided funding support for LEW's doctoral programme. MIM was a Wellcome Senior Investigator and a National Institute for Health Research (NIHR) Senior Investigator. This work was funded in Oxford by the Wellcome Trust (098381, 106130, 203141), the NIH (U01-DK105535; U01-DK085545), the NIHR (NF-SI-0617-10090) and the Oxford Biomedical Research Centre (BRC). The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Intramural Research Program provided support for NIDDK-based co-authors. The views expressed in this article are those of the author(s) and not necessarily those of the NIH, the NHS, the NIHR or the UK Department of Health. The funding sources were not involved in study design; in collection,

analysis and interpretation of data; in writing of the report; or in the decision to submit the paper for publication.

**Authors' relationships and activities** MIM has served on advisory panels for Pfizer, Novo Nordisk and Zoe Global; has received honoraria from Merck, Pfizer, Novo Nordisk and Eli Lilly; and has received research funding from AbbVie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier and Takeda. As of June 2019, MIM and AM are employees of Genentech and holders of Roche stock.

**Contribution statement** LEW, AM, MIM and RLH contributed to the conceptualisation and design of the manuscript. AM, W-CH, PC, MTO and SK contributed to the acquisition of data. LEW, AM, LJB, MS, WCK, MIM and RLH contributed to the analysis and interpretation of data. All co-authors contributed to the drafting of this article and revised it critically for important intellectual content. All authors gave final approval of the version to be published. RLH is the guarantor of this work.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Meigs JB, Shrader P, Sullivan LM et al (2008) Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med* 359(21):2208–2219. <https://doi.org/10.1056/NEJMoa0804742>
2. Lyssenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNA variants, and the development of type 2 diabetes. *N Engl J Med* 359(21):2220–2232. <https://doi.org/10.1056/NEJMoa0801869>
3. Mars N, Koskela JT, Ripatti P et al (2020) Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 26(4):549–557. <https://doi.org/10.1038/s41591-020-0800-0>
4. Vaxillaire M, Yengo L, Lobbens S et al (2014) Type 2 diabetes-related genetic risk scores associated with variations in fasting plasma glucose and development of impaired glucose homeostasis in the prospective DESIR study. *Diabetologia* 57(8):1601–1610. <https://doi.org/10.1007/s00125-014-3277-x>
5. Park HY, Choi HJ, Hong YC (2015) Utilizing genetic predisposition score in predicting risk of type 2 diabetes mellitus incidence: a community-based cohort study on middle-aged Koreans. *J Korean Med Sci* 30(8):1101–1109. <https://doi.org/10.3346/JKMS.2015.30.8.1101>
6. He Y, Lakhani CM, Rasooly D, Manrai AK, Tzoulaki I, Patel CJ (2021) Comparisons of polyexposure, polygenic, and clinical risk scores in risk prediction of type 2 diabetes. *Diabetes Care* 44(4):935–943. <https://doi.org/10.2337/dc20-2049>
7. Hanson RL, Rong R, Kobes S et al (2015) Role of established type 2 diabetes-susceptibility genetic variants in a high prevalence



- American Indian population. *Diabetes* 64(7):2646–2657. <https://doi.org/10.2337/db14-1715>
8. Knowler WC, Bennett PH, Hamman RF, Miller M (1978) Diabetes incidence and prevalence in Pima Indians: a 19-fold greater incidence than in Rochester, Minnesota. *Am J Epidemiol* 108(6):497–505. <https://doi.org/10.1093/oxfordjournals.aje.a112648>
  9. Kahn R (1997) Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 20(7):1183–1197. <https://doi.org/10.2337/diacare.20.7.1183>
  10. Piaggi P, Masindova I, Muller YL et al (2017) A genome-wide association study using a custom genotyping array identifies variants in GPR158 associated with reduced energy expenditure in American Indians. *Diabetes* 66(8):2284–2295. <https://doi.org/10.2337/db16-1565>
  11. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 5(6):e1000529. <https://doi.org/10.1371/journal.pgen.1000529>
  12. Malhotra A, Kobes S, Bogardus C, Knowler WC, Baier LJ, Hanson RL (2014) Assessing accuracy of genotype imputation in American Indians. *PLoS One* 9(7):e102544. <https://doi.org/10.1371/JOURNAL.PONE.0102544>
  13. Mahajan A, Taliun D, Thumer M et al (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 50(11):1505–1513. <https://doi.org/10.1038/s41588-018-0241-6>
  14. Spracklen CN, Horikoshi M, Kim YJ et al (2020) Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 582(7811):240–245. <https://doi.org/10.1038/s41586-020-2263-3>
  15. Mahajan A, Spracklen CN, Zhang W et al (2022) Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 54(5):560–572. <https://doi.org/10.1038/s41588-022-01058-3>
  16. Vijayakumar P, Nelson RG, Hanson RL, Knowler WC, Sinha M (2017) HbA1c and the prediction of type 2 diabetes in children and adults. *Diabetes Care* 40(1):16–21. <https://doi.org/10.2337/dc16-1358>
  17. Chambers M, Tanamas SK, Clark EJ et al (2017) Growth tracking in severely obese or underweight children. *Pediatrics* 140(6):e20172248. <https://doi.org/10.1542/peds.2017-2248>
  18. Davidson KW, Barry MJ, Mangione CM et al (2021) Screening for prediabetes and type 2 diabetes: US preventive services task force recommendation statement. *JAMA* 326(8):736–743. <https://doi.org/10.1001/JAMA.2021.12531>
  19. Olaiya MT, Wedekind LE, Hanson RL et al (2019) Birthweight and early-onset type 2 diabetes in American Indians: differential effects in adolescents and young adults and additive effects of genotype, BMI and maternal diabetes. *Diabetologia* 62(9):1628–1637. <https://doi.org/10.1007/s00125-019-4899-9>
  20. Hanson RL, Guo T, Muller YL et al (2013) Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in American Indians. *Diabetes* 62(8):2984–2991. <https://doi.org/10.2337/db12-1767>
  21. Baier LJ, Muller YL, Remedi MS et al (2015) ABCC8 R1420H loss-of-function variant in a southwest American Indian community: Association with increased birth weight and doubled risk of type 2 diabetes. *Diabetes* 64(12):4322–4332. <https://doi.org/10.2337/db15-0459>
  22. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA (1982) Evaluating the yield of medical tests. *JAMA* 247(18):2543–2546. <https://doi.org/10.1001/JAMA.1982.03320430047030>
  23. Pencina MJ, D’Agostino RB (2004) Overall C as a measure of discrimination in survival analysis: Model specific population value and confidence interval estimation. *Stat Med* 23(13):2109–2123. <https://doi.org/10.1002/sim.1802>
  24. Pencina MJ, D’Agostino RB, D’Agostino RB, Vasan RS (2008) Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 27(2):157–172. <https://doi.org/10.1002/sim.2929>
  25. Chambless LE, Cummiskey CP, Cui G (2011) Several methods to assess improvement in risk prediction models: extension to survival analysis. *Stat Med* 30(1):22–38. <https://doi.org/10.1002/sim.4026>
  26. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak* 26(6):565–574. <https://doi.org/10.1177/0272989X06295361>
  27. Vickers AJ, Cronin AM, Elkin EB, Gonen M (2008) Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. *BMC Med Inform Decis Mak* 8(1):53. <https://doi.org/10.1186/1472-6947-8-53>
  28. Olaiya MT, Knowler WC, Sinha M et al (2020) Weight tracking in childhood and adolescence and type 2 diabetes risk. *Diabetologia* 63(9):1753–1763. <https://doi.org/10.1007/S00125-020-05165-W>
  29. Vassy JL, Hivert MF, Porneala B et al (2014) Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes* 63(6):2172–2182. <https://doi.org/10.2337/db13-1663>
  30. Moazzam-Jazi M, Najd Hassan Bonab L, Zahedi AS, Daneshpour MS (2020) High genetic burden of type 2 diabetes can promote the high prevalence of disease: a longitudinal cohort study in Iran. *Sci Rep* 10(1):14006. <https://doi.org/10.1038/s41598-020-70725-4>
  31. Pencina MJ, D’Agostino RB, Pencina KM, Janssens ACJW, Greenland P (2012) Interpreting incremental value of markers added to risk prediction models. *Am J Epidemiol* 176(6):473–481. <https://doi.org/10.1093/aje/kws207>
  32. Lambert SA, Abraham G, Inouye M (2019) Towards clinical utility of polygenic risk scores. *Hum Mol Genet* 28(R2):R133–R142. <https://doi.org/10.1093/HMG/DDZ187>
  33. Hivert MF, Jablonski KA, Perreault L et al (2011) Updated genetic score based on 34 confirmed type 2 diabetes loci is associated with diabetes incidence and regression to normoglycemia in the diabetes prevention program. *Diabetes* 60(4):1340–1348. <https://doi.org/10.2337/db10-1119>
  34. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ (2019) Current clinical use of polygenic scores will risk exacerbating health disparities. *Nat Genet* 51(4):584. <https://doi.org/10.1038/S41588-019-0379-X>
  35. Conrad DF, Jakobsson M, Coop G et al (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat Genet* 38(11):1251–1260. <https://doi.org/10.1038/ng1911>
  36. Jiang X, Holmes C, McVean G (2021) The impact of age on genetic risk for common diseases. *PLOS Genet* 17(8):e1009723. <https://doi.org/10.1371/JOURNAL.PGEN.1009723>
  37. Hanson RL, Knowler WC (1998) Analytic strategies to detect linkage to a common disorder with genetically determined age of onset: diabetes mellitus in Pima Indians. *Genet Epidemiol* 15:299–315. [https://doi.org/10.1002/\(SICI\)1098-2272\(1998\)15:3](https://doi.org/10.1002/(SICI)1098-2272(1998)15:3)
  38. Crandall JP, Knowler WC, Kahn SE et al (2008) The prevention of type 2 diabetes. *Nat Clin Pract Endocrinol Metab* 4(7):382–393. <https://doi.org/10.1038/ncpendmet0843>

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