

# Validating the doubly weighted genetic risk score for the prediction of type 2 diabetes in the Lifelines and Estonian Biobank cohorts

Katri Pärna<sup>1,2</sup>  | Harold Snieder<sup>1</sup> | Kristi Läll<sup>2</sup> | Krista Fischer<sup>2,3</sup> | Ilja Nolte<sup>1</sup>

<sup>1</sup>Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

<sup>2</sup>Institute of Genomics, University of Tartu, Tartu, Estonia

<sup>3</sup>Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia

## Correspondence

Katri Pärna, Department of Epidemiology, University Medical Center Groningen, Groningen, the Netherlands and Institute of Genomics, University of Tartu, 9700 RB Groningen, Tartu, Estonia.  
Email: k.parna@umcg.nl

## Funding information

Netherlands Organization of Scientific Research NWO, Grant/Award Number: 175.010.2007.006; Estonian Research Council, Grant/Award Numbers: GP1GV9353, IUT20-60; Centre of Excellence for Genomics and Translational Medicine (GENTRANSMED); the University of Tartu (SP1GVARENG); Archimedes Foundation, Grant/Award Number: 3.2.1001.11-0033; EU 2020, Grant/Award Number: 692145; European Regional Development Fund, Grant/Award Number: 2014-2020.4.01.16-0024; Horizon 2020, Grant/Award Number: 777107-PRESICE4Q; Estonian Research Council, Grant/Award Number: PUT PRG687

## Abstract

As many cases of type 2 diabetes (T2D) are likely to remain undiagnosed, better tools for early detection of high-risk individuals are needed to prevent or postpone the disease. We investigated the value of the doubly weighted genetic risk score (dwGRS) for the prediction of incident T2D in the Lifelines and Estonian Biobank (EstBB) cohorts. The dwGRS uses an additional weight for each single nucleotide polymorphism in the risk score, to correct for “Winner’s curse” bias in the effect size estimates. The traditional (single-weighted genetic risk score; swGRS) and dwGRS were calculated for participants in Lifelines ( $n = 12,018$ ) and EstBB ( $n = 34,129$ ). The dwGRS was found to have stronger association with incident T2D (hazard ratio [HR] = 1.26 [95% confidence interval: 1.10–1.43] and HR = 1.35 [1.28–1.42]) compared to the swGRS (HR = 1.21 [1.07–1.38] and HR = 1.25 [1.19–1.32]) in Lifelines and EstBB, respectively. Comparing the 5-year predicted risks from the models with and without the dwGRS, the continuous net reclassification index was 0.140 (0.034–0.243;  $p = .009$  Lifelines), and 0.257 (0.194–0.319;  $p < 2 \times 10^{-16}$  EstBB). The dwGRS provided incremental value to the T2D prediction model with established phenotypic predictors. It clearly distinguished the risk groups for incident T2D in both biobanks thereby showing its clinical relevance.

## KEYWORDS

genetic risk score, incidence, personalized prediction, type 2 diabetes

## 1 | INTRODUCTION

If no immediate action is taken, it is estimated that the current number of 463 million diabetes cases will increase to 700 million by the year 2045, with type 2

diabetes (T2D) as the most common type, accounting for 90% of the cases (International Diabetes Federation, 2019). T2D is often accompanied by severe comorbidities and possible complications resulting in premature mortality and a major burden on healthcare systems

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Genetic Epidemiology* Published by Wiley Periodicals LLC

(Kahn, Cooper, & Del Prato, 2014; World Health Organization, 2014). Furthermore, it is speculated that one-third or even half of T2D cases are undiagnosed, since it usually starts without acute symptoms and is difficult to detect at an early stage (International Diabetes Federation, 2019; Langenberg et al., 2014). Therefore, developing a powerful tool for early detection of high-risk individuals for T2D would allow postponing or even preventing T2D. Genetic markers have high potential for early detection of high-risk individuals since they are fixed for life (Lyssenko et al., 2008). Heritabilities of T2D vary widely with estimates between 20% and 80% (Meigs, Cupples, & Wilson, 2000; Poulsen, Ohm Kyvik, Vaag, & Beck-Nielsen, 1999). Only since the advent of genome-wide association studies (GWASs) progress has been made in finding the genetic risk factors for T2D (Burton et al., 2007; Palmer et al., 2012; Scott et al., 2007; Visscher et al., 2017). Nonetheless, identified single nucleotide polymorphisms (SNPs) have only been able to account for a small proportion (approximately 10–15%) of the total heritability of T2D (McCarthy, 2010; Prasad & Groop, 2015). Therefore, it is crucial to further improve the genetic prediction and risk assessment of T2D.

Individually each SNP has only a small effect on disease risk (Reisberg, Iljasenko, Läll, Fischer, & Vilo, 2017). Several methods have been developed and tested to combine the effects of multiple disease-associated SNPs into one genetic risk score (GRS) representing part of each individual's genetic susceptibility for a complex disease (Morris et al., 2012; Reisberg et al., 2017). The most commonly used GRS includes only the genome-wide significant SNPs, and each included SNP is typically weighted by the effect size from the genome-wide meta-analysis in which it was discovered (Läll, Mägi, Morris, Metspalu, & Fischer, 2017). A limitation of this GRS is the exclusion of genetic variants truly associated with the disease (Morris et al., 2012; Reisberg et al., 2017) for which GWASs have too low power to detect due to the stringent significance threshold (usually  $p < 5 \times 10^{-8}$ ; Wray et al., 2013). Hence, another common method is to combine thousands of SNPs in a polygenic risk score by applying a more lenient significance threshold ranging from  $p < 5 \times 10^{-8}$  up to even  $p = 1$  depending on the polygenic nature of the disease hypothesizing that many of those SNPs were false negative in the discovery GWAS (Wray et al., 2013). Some recent publications have shown that such polygenic risk scores have considerably better predictive power (Läll et al., 2017; Reisberg et al., 2017).

Recently, a novel polygenic method for risk prediction called the doubly weighted GRS (dwGRS) was developed (Läll et al., 2017). Compared to the usual polygenic GRS, a second weight for the SNPs included in the dwGRS is added to adjust for the Winner's curse bias, a phenomenon stating that the effect of those SNPs reaching genome-wide

significance in GWAS may be overestimated by chance (Läll et al., 2017). It has been shown that the dwGRS has better predictive power than the traditional single-weighted GRS (swGRS) by not only increasing the number of SNPs beyond the genome-wide significance threshold as in a polygenic risk score, but also by applying more accurate weighting (Läll et al., 2017). However, the superior performance of this novel dwGRS has not yet been validated in independent cohorts.

In this study, we aim to investigate the added value of the novel dwGRS to the prediction model of incident T2D with main established risk factors in the Lifelines and Estonian Biobank (EstBB) cohorts.

## 2 | METHODS

### 2.1 | Study population

Data for the current study were derived from the Lifelines Cohort Study (Lifelines) and the EstBB. Both cohorts are large prospective cohort studies in Europe with similar aim to improve the current understanding of genetic, environmental, and phenotypic factors involved in the development of common complex diseases (Leitsalu, Haller et al., 2015; Scholtens et al., 2015). Lifelines is a multidisciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviors of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, sociodemographic, behavioral, physical, and psychosocial factors which contribute to the health and disease of the general population, with a special focus on a multimorbidity and complex genetics. The Lifelines participants were recruited between 2006 and 2013 and the follow-up is ongoing (Klijs et al., 2015; Scholtens et al., 2015). Every 5 years, biomaterials are collected, a physical examination is done, and extensive questionnaires are completed. In between, participants fill in questionnaires approximately every 1.5 years.

The EstBB cohort represents the Estonian adult population with individuals ( $n = 51,515$ ) recruited between 2002 and 2011 (Leitsalu, Haller et al., 2015). Biomarkers were collected, extensive phenotypic questionnaires completed, and physical measurements taken at baseline. Follow-up data are available via linkage with national health-related registries and via re-examination of participants. Furthermore, electronic health records are updated for phenotypic outcome information every half year (Leitsalu, Alavere, Tammesoo, Leego, & Metspalu, 2015; Leitsalu, Haller et al., 2015).

In the current study, subsets of individuals from Lifelines and from EstBB, respectively, with genetic data

available were analyzed (Figures S1 and S2). In Lifelines, individuals were excluded if they had been diagnosed at baseline with T1D, T2D, or another type of diabetes, in case of pregnancy (without gestational diabetes), there were missing values for body mass index (BMI), diabetes type was unspecified, or no follow-up data were available. In EstBB, individuals who were included in the original dwGRS paper (Läll et al., 2017) or their first-degree relatives ( $\pi\text{-hat} \geq 0.35$ ), and other first-degree relatives from the remaining sample were excluded ( $n = 3,121$ ) to get an independent validation dataset. In addition, individuals were excluded if T1D or T2D had been diagnosed at baseline, or in case of pregnancy, extreme age values (age  $< 18$  years or  $> 90$  years), missing phenotype information, and no follow-up information available. In case of duplicate samples in the remaining sample, one of them was removed. In Lifelines, duplicate samples and first-degree relatives were excluded as part of the quality control of the genotyping. This yielded 12,018 and 34,129 individuals from Lifelines and from EstBB, respectively, for analysis. Additional sensitivity analyzes including only largely unrelated individuals (i.e.,  $\pi\text{-hat} < 0.125$ ) were conducted in EstBB ( $n = 26,669$ ).

Lifelines and EstBB have been approved by the Medical Ethics Committee of the University Medical Center Groningen and by the Ethics Committee of Human Studies, University of Tartu, Estonia, respectively, and all participants signed an informed consent.

## 2.2 | Type 2 diabetes

In the Lifelines cohort, the information about T2D was collected at four time points: at baseline, and at 1.5, 3, and 5 years after recruitment. At baseline, the participant was diagnosed with T2D if at least one of the four following T2D diagnosis criteria was met (American Diabetes Association, 2017): (a) fasting plasma glucose (FPG)  $\geq 7.0$  mmol/L; (b) glycated hemoglobin (HbA1c)  $\geq 6.5\%$ ; (c) use of T2D medication (anatomical therapeutic chemical [ATC] codes A10A or A10B); or (d) self-reported T2D in combination with self-reported T2D medication use. At 1.5 and 3 years, the T2D diagnosis was based only on self-reported T2D as only questionnaire data were available. At 5 years, diagnosis was based on FPG and/or HbA1c measures and/or on self-reported T2D, since medication use was not recorded (Van Zon et al., 2018). FPG was determined in fasting state and measured using a hexokinase method (Integra analyzer; Roche). HbA1c level was measured with a turbidimetric inhibition immunoassay (Modular Roche; Jansen et al., 2013).

In EstBB, the data on T2D diagnoses were obtained via linkage with the Electronic Medical Records database

of the Estonian Health Insurance Fund. T2D was in the first place diagnosed by the general practitioner according to World Health Organization guidelines (WHO; World Health Organization, 2011) and in the EstBB database, it was recorded as International Classification of Diseases (ICD-10) code E11.

## 2.3 | Phenotypic prediction model

One of the most well-known noninvasive prediction models is the Finnish Diabetes Risk Score (FINDRISC; Lindstrom & Tuomilehto, 2003). It has been tested and validated worldwide (Meijnikman et al., 2016; Witte, Shipley, Marmot, & Brunner, 2010; Zhang, Zhang, Zhang, Hu & Chen, 2014) and it contains age, BMI, waist circumference, physical activity, fruit and vegetable consumption, use of antihypertensive medication or blood pressure, history of high blood glucose, and family history of diabetes. Therefore, our selection of phenotypic predictors was based on the FINDRISC model with some modifications imposed by the current data sets. To construct the prediction model in Lifelines, the following variables were available: (a) age; (b) BMI (Janssen, Katzmarzyk, & Ross, 2002); (c) waist circumference (Jansen et al., 2013); (d) physical activity (measured with the Short Questionnaire to Assess Health; Wendel-Vos, Schuit, Saris, & Kromhout, 2003); (e) fruit and vegetable consumption (measured with the extensive semi-quantitative baseline food frequency questionnaire; Siebelink, Geelen, & de Vries, 2011); (f) antihypertensive medication usage (ATC code C02, C03, C07, C08, C09; WHO, 2017); (g) blood pressure (Messerli, Williams, & Ritz, 2007), whereas participants were categorized as hypertensive with blood pressure higher than 140/90 mmHg, and/or antihypertensive medication usage was recorded (Amini et al., 2016). With respect to physical activity, for better comparison of two biobanks we categorized participants into “low activity” (0–3 days a week active for  $\geq 30$  min) or “high activity” ( $\geq 4$  days a week active for  $\geq 30$ ) in Lifelines. Two variables for fruit and vegetable intake (in grams per 1,000 kcal) were divided into quintiles and awarded with scores ranging from 0 to 4 (subscale score of the original Lifelines Diet Score; Vinke et al., 2018). The original FINDRISC predictors “history of high glucose levels” and “family history of diabetes” were dropped from the model as they were not available in Lifelines.

In EstBB, data on all the phenotypic FINDRISC variables were obtained from the 16-module questionnaire about lifestyle, diet, and clinical diagnosis based on the WHO International Classification of Diseases (WHO, 2016), which was filled in at baseline during a

computer-assisted personal interview (Leitsalu, Haller et al., 2015). Most of the FINDRISC variables were measured and categorized in the same way as in Lifelines with the exceptions for the following: (a) physical activity was categorized as “active” and “non-active” based on the question of “Have you done or are you doing physical exercises (Sunday sports)?”; (b) fruit and vegetable consumption were considered as separate variables and categorized based on consumption frequency (days per week): not once, 1–2, 3–5, 6–7.

For some of the FINDRISC variables (waist circumference, physical activity, and fruit–vegetable consumption), the number of missing values in at least one of the cohorts was considered to be too high (>5%). Because of their minor importance as predictors of incident T2D (Lindstrom & Tuomilehto, 2003), these variables were excluded from the analyzes to provide comparable results between the two cohorts.

## 2.4 | Genetic risk score

For both biobanks, each participant’s genetic predisposition for T2D was expressed by a GRS constructed according to the traditional swGRS and novel dwGRS method (Läll et al., 2017). The swGRS included 1,000 independent SNPs (Läll et al., 2017) with their risk alleles weighted by the effect size from the genome-wide meta-analysis on T2D by the DIAGRAM consortium (Morris et al., 2012). The number of 1,000 SNPs was chosen because in the original GRS paper (Läll et al., 2017), this number of SNPs provided a fit for the dwGRS (see next paragraph) that was not significantly different from the best-fitting swGRS in either the BMI-unadjusted or BMI-adjusted analysis and we wanted to have the same number of SNPs in the swGRS.

The selection of SNPs for the dwGRS in the Lifelines sample was based on the list of 7,502 independent SNPs that were selected in the original study (Läll et al., 2017), where also the more detailed SNP selection process was described. Briefly, summary statistics from the large scale meta-analysis on T2D unadjusted for BMI performed by the DIAGRAM consortium (Morris et al., 2012) were used for GRS construction. As the Estonian MetaboChip (includes approximately 200,000 SNPs in genes associated with cardiac and metabolic diseases) was part of the DIAGRAM consortium, the meta-analysis was reran without the EstBB data to obtain independent effect sizes. These new results were used to clump the EstBB MetaboChip genotyping data (based on a  $p$ -value for association with T2D  $p < .5$ ;  $r^2 \leq 0.05$  and a minimal distance of 2 Mb) to get a set of independent SNPs (Purcell et al., 2007). This yielded a final set of 7,502 SNPs to

construct the GRSs. Six of these SNPs were not available in Lifelines leaving 7,496 SNPs for further analysis. These SNPs were extracted from the genotyped and imputed GWAS data in Lifelines. Genotyping was done with the Illumina CytoSNP 12 v2 chip and imputation was performed with Minimac (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) using the 1000 Genomes Phase 1 global reference panel (The 1000 Genomes Project Consortium, 2015). The median imputation quality of the 7,496 SNPs was 0.94 and for 77.8% of them the imputation quality was >0.8. However, as low-quality SNPs still contribute to the GRS (Nolte et al., 2017), we included all SNPs in the analysis. Standard quality control was done for genetic data in Lifelines (Nolte et al., 2017) and in EstBB (Läll et al., 2019).

In the dwGRS, additional to the traditional weight (i.e., the regression coefficient from the meta-GWAS), a second weight is introduced  $\hat{\pi}_i$ , which is the estimated probability that the specific marker belongs to the set of 1,000 top SNPs with the strongest association with a disease (Equation 1).

$$\text{dwGRS} = \sum_{i=1}^N \hat{\pi}_i (1000) \hat{\beta}_i X_i. \quad (1)$$

The motivation for this is that the traditional GRS systematically includes more SNPs whose effects are overestimated by chance and excludes more of those with an underestimated effect, which may therefore unjustly be excluded from the risk score calculation. In this study we used the same estimated probabilities as earlier (Läll et al., 2017), since these do not depend on the validation cohort. To better interpret the effect sizes of the GRSs, they were standardized within both cohorts.

## 2.5 | Statistical analyzes

Normally distributed continuous variables were described with mean and standard deviation ( $SD$ ) and nonnormally distributed with median and interquartile range ( $IQR$ ). For categorical variables counts and percentages were presented.

Throughout the study, data from the two biobanks—Lifelines and EstBB—have been analyzed separately. In EstBB, the Cox proportional hazard model was used for survival analysis since the exact time of diagnosis was available. Proportional hazards assumptions were confirmed by Schoenfeld’s test. In Lifelines, data of T2D diagnosis were interval-censored, as it was only known whether the event occurred within a certain time interval. Time intervals (in months) between baseline, 1.5, 3, and 5-year follow-up were used. We used accelerated

failure time (AFT) analysis to test association of the FINDRISC variables and the GRSs with incident T2D. For better interpretation of the results of the AFT, the AFT coefficient of each predictor was converted to a hazard ratio (HR) using the formula  $HR_i = \exp(-\alpha_i/\sigma)$ , where  $\alpha_i$  is the coefficient from the AFT model and  $\sigma$  is the scale factor of the residuals. The corresponding 95% confidence intervals (CIs) for HR were calculated similarly using  $\alpha_i \pm 1.96 \times SD(\alpha_i)$ .

To assess the added value of the dwGRS in risk prediction of incident T2D, three models were fitted in both cohorts: (a) model adjusted for the baseline phenotypic risk factors available in both cohorts (i.e., baseline model/Model 1); (b) baseline model plus the swGRS (Model 2); (c) baseline model plus the dwGRS (Model 3).

To test the prediction improvement by using dwGRS after 5 years of follow-up, the following analyzes were applied on smaller subsets of cases that were diagnosed with T2D at most 5 years after baseline plus controls with a follow-up time of at least 5 years (Lifelines:  $n = 5,243$ ; EstBB:  $n = 33,057$ ). To assess the prediction improvement, the Harrell's C-statistic was determined for all the models with an increase in the statistics value reflecting the incremental predictive value of the genetic profile. Additionally, in EstBB, the predictive value of the dwGRS was tested among a high-risk group (age: 35–79 and BMI: 25–30) for incident T2D to check whether the performance of the dwGRS improved in case of selecting a high-risk sample. In Lifelines, the high-risk group analysis was not possible due the smaller sample size. To obtain Harrell's C-statistics and to present cumulative incidence in Lifelines, a Cox proportional hazard model was applied using the middle of the time interval as time of diagnosis. Confidence intervals for a change in C-statistics were estimated with a bootstrapping procedure. Additionally, continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI) indices were calculated to assess the prediction improvement (Kundu, Aulchenko, Van Duijn, & Janssens, 2011; Pencina, D'Agostino, D'Agostino, & Vasan, 2008). Continuous NRI denotes the proportion of cases correctly assigned a higher probability by the improved model compared to the old model without GRS minus the corresponding proportion of controls, multiplied by 2 (Pencina, Steyerberg, & D'Agostino, 2011). IDI denotes the average increase in risk estimates for participants who were diagnosed with T2D during follow-up plus the average decrease in risk estimates for participants without diagnosis according to improved model (Fischer et al., 2014; Pencina et al., 2011).

Finally, a Cox proportional hazard model was applied to assess the risk of incident T2D by dwGRS quintiles and Kaplan–Meier graphs of cumulative incidence of T2D on all individuals in Lifelines and EstBB, respectively.

The statistical package IBM SPSS for Windows (version 22.0; IBM Corp., Armonk, NY) and R version 3.5.2 (in Lifelines) and version 3.6.0 (in EstBB) for Windows (R Development Core Team, 2008) were used for statistical analyzes.  $p < .05$  was considered significant.

### 3 | RESULTS

Baseline characteristics of participants in both studies are presented in Table 1. In total, 12,018 individuals from Lifelines and 34,129 from EstBB were studied. The age range was similar in both cohorts and the proportion of males was lower than the proportion of females. However, there were more females in EstBB than in Lifelines (58.1% compared to 48.6%, respectively). In Lifelines, a

**TABLE 1** Descriptives for Lifelines and Estonian Biobank cohorts

Characteristics	Lifelines cohort $n = 12,018$	Estonian Biobank $n = 34,129$
Incident cases, $n$ (%)	255 (2.1)	1,565 (4.6)
Follow-up time (y)	4.7 (3.8–5.5)	7.0 (5.7–7.9)
Age range (y)	18–89	18–90
Sex, $n$ (male %)	4,971 (41.4)	10,904 (31.9)
BMI ( $\text{kg}/\text{m}^2$ )	$26.2 \pm 4.1$	$25.9 \pm 5.0$
Hypertension, $n$ (%)	3,479 (28.9)	9,530 (27.9)
Waist circumference (cm)	$91.3 \pm 11.8$	$85.4 \pm 14.0$
Physical activity <sup>a</sup>		
Low/inactive	3,828 (34.6)	11,803 (43.3)
High/active	7,224 (65.4)	15,448 (56.7)
Fruit consumption <sup>b</sup>		
Low	2,201 (35.6)	10,595 (31.1)
Medium	2,496 (40.3)	12,426 (36.4)
High	1,489 (24.1)	11,101 (32.5)
Vegetable consumption <sup>b</sup>		
Low	2,222 (35.9)	23,241 (68.1)
Medium	2,712 (43.9)	8,587 (25.2)
High	1,252 (20.2)	2,296 (6.7)

Note: Mean  $\pm$  SD; median (interquartile range);  $n$  (%). Low consumption = not eating at all or 1–2 days per week, medium consumption = eating 3–5 days per week, and high consumption = eating 6–7 days per week in Estonian Biobank.

Abbreviation: BMI, body mass index.

<sup>a</sup>Physical activity categorization: low = 0–3 days a week active for  $\geq 30$  min and high =  $\geq 4$  days a week active for  $\geq 30$  min in Lifelines. Active versus inactive in Estonian Biobank.

<sup>b</sup>Fruit and vegetable consumption: Low consumption = 1st and 2nd quintile, medium consumption = 3rd quintile, high consumption = 4th and 5th quintile of Lifelines Diet Score in Lifelines cohort.

total of 255 (2.1%) and in EstBB 1,565 (4.6%) individuals developed T2D during follow-up (with median follow-up of 4.7 years [IQR: 3.8–5.5] and 7.0 years [IQR: 5.7–7.9], respectively). The mean BMI ( $26.2 \pm 4.1$  and  $25.9 \pm 5.0$ , respectively) in both cohorts fell in the overweight category and waist circumference was slightly larger in Lifelines than in EstBB (91.3 vs. 85.4, respectively). Prevalence of hypertension was similar in both cohorts (28.9% and 27.9%, respectively). For the physical activity and fruit and vegetable consumption, a solid comparison cannot be based because the underlying questions were different between the cohorts.

### 3.1 | Association between traditional baseline risk factors and incident T2D

The results of model 1, which includes only the established phenotypic risk factors, are shown in Table 2 for Lifelines and Table 3 for EstBB, respectively. All the models demonstrated significant association between the characteristics measured at baseline and higher incident T2D. Male participants of Lifelines and EstBB had a significantly higher risk for incident T2D during the follow-up time (HR = 1.49, 95% CI: 1.15–1.93 and HR = 1.23, 95% CI: 1.11–1.38, respectively). With every unit increase in BMI, the risk of having T2D increased 1.15 (95% CI: 1.11–1.18) and 1.13 (95% CI: 1.12–1.14) times, and having hypertension increased the risk of T2D 2.63 (95% CI: 1.94–3.56) and 1.91 (1.68–2.16) times in Lifelines and EstBB, respectively.

### 3.2 | Association between the GRS and incident T2D

The dwGRS showed a stronger effect on incident T2D (Model 3) than the swGRS (Model 2) in both cohorts.

Every SD increase in dwGRS increased the risk of incident T2D 1.26 times (95% CI: 1.10–1.43,  $5.25 \times 10^{-04}$ ) and 1.35 times (95% CI: 1.28–1.42,  $2.85 \times 10^{-32}$ ) in Lifelines and EstBB, respectively, compared to the swGRS (HR = 1.21 [95% CI: 1.07–1.38] and HR = 1.25 [95% CI: 1.19–1.32] in Lifelines and EstBB, respectively, while accounting for other traditional risk factors at baseline. When adding the GRS to the model (Models 2 and 3), the effect sizes of age, sex, BMI, and hypertension remained similar to Model 1 and were still associated with significantly higher incident T2D (Tables 1 and 2). Analyses on the subset of largely unrelated individuals in EstBB (only fourth or lower degree relatives) showed similar results (Table S1).

### 3.3 | Assessing the incremental value of the dwGRS for risk prediction of incident T2D

In both cohorts, the Harrell's C-statistic value became larger when a GRS was added to the prediction model with baseline phenotypic risk factors. After adding the dwGRS to the prediction model in Lifelines, the C-statistic increased by 0.003 (95% CI: –0.003 to 0.022) compared to model with baseline risk factors, and in EstBB, it increased by 0.007 (95% CI: 0.004–0.010). When focusing on a high-risk group for incident T2D with age 35–79 and BMI 25–30 in EstBB, the change in C-statistics was much larger: the C-statistic increased by 0.021 (95% CI: 0.009–0.035) when dwGRS was added to the prediction model compared to the model with baseline risk factors. Comparing the 5-year prediction estimates of model with dwGRS and without GRS, the continuous NRI was 0.140 (95% CI: 0.034–0.243;  $p = .009$ ), and 0.257 (95% CI: 0.194–0.319;  $p < 2 \times 10^{-16}$ ) in Lifelines and EstBB, respectively, which indicates more accurate prediction performance of the model with the dwGRS than without a GRS, and a significant incremental

**TABLE 2** Effects of the single and doubly weighted genetic risk scores on incident type 2 diabetes in the Lifelines cohort

Characteristic	Model 1 HR		Model 2 HR		Model 3 HR	
	(95% CI)	p-value	(95% CI)	p-value	(95% CI)	p-value
Age	1.32 (1.18–1.48)	$1.71 \times 10^{-06}$	1.32 (1.18–1.48)	$1.39 \times 10^{-06}$	1.32 (1.18–1.48)	$1.45 \times 10^{-06}$
Sex (male)	1.49 (1.15–1.93)	$2.83 \times 10^{-03}$	1.48 (1.14–1.92)	$3.12 \times 10^{-03}$	1.47 (1.13–1.90)	$3.81 \times 10^{-03}$
BMI	1.15 (1.11–1.18)	$2.94 \times 10^{-21}$	1.15 (1.11–1.18)	$1.94 \times 10^{-21}$	1.15 (1.11–1.18)	$2.36 \times 10^{-21}$
Hypertension	2.63 (1.94–3.56)	$4.40 \times 10^{-10}$	2.62 (1.93–3.54)	$5.29 \times 10^{-10}$	2.62 (1.93–3.55)	$5.22 \times 10^{-10}$
swGRS	–	–	<b>1.21 (1.07–1.38)</b>	<b><math>3.58 \times 10^{-03}</math></b>	–	–
dwGRS	–	–	–	–	<b>1.26 (1.10–1.43)</b>	<b><math>5.25 \times 10^{-04}</math></b>

Note: All models account for age<sup>2</sup>.

Abbreviations: BMI, body mass index; CI, confidence interval; dwGRS, doubly weighted genetic risk score; HR, hazard ratio; swGRS, standardized single-weighted genetic risk score.

**TABLE 3** Effects of swGRS and dwGRS on incident T2D in the EstBB

Characteristic	Model 1 HR (95% CI)	p-value	Model 2 HR (95% CI)	p-value	Model 3 HR (95% CI)	p-value
Age	1.17 (1.14–1.20)	$1.04 \times 10^{-27}$	1.16 (1.13–1.20)	$3.40 \times 10^{-27}$	1.16 (1.13–1.20)	$4.96 \times 10^{-27}$
Sex (male)	1.23 (1.11–1.38)	$2.44 \times 10^{-04}$	1.23 (1.11–1.38)	$2.06 \times 10^{-04}$	1.24 (1.11–1.39)	$1.29 \times 10^{-04}$
BMI	1.13 (1.12–1.14)	$1.06 \times 10^{-153}$	1.13 (1.12–1.14)	$3.93 \times 10^{-152}$	1.14 (1.12–1.15)	$5.36 \times 10^{-155}$
Hypertension	1.91 (1.68–2.16)	$7.86 \times 10^{-25}$	1.89 (1.67–2.14)	$3.13 \times 10^{-24}$	1.89 (1.68–2.14)	$2.14 \times 10^{-24}$
swGRS	–	–	<b>1.25 (1.19–1.32)</b>	<b><math>6.63 \times 10^{-19}</math></b>	–	–
dwGRS	–	–	–	–	<b>1.35 (1.28–1.42)</b>	<b><math>2.85 \times 10^{-32}</math></b>

Note: All models account for age<sup>2</sup> and for genotyping platforms.

Abbreviations: BMI, body mass index; CI, confidence interval; dwGRS, doubly weighted genetic risk score; EstBB, Estonian Biobank; HR, hazard ratio; swGRS, standardized single-weighted genetic risk score; T2D, type 2 diabetes.

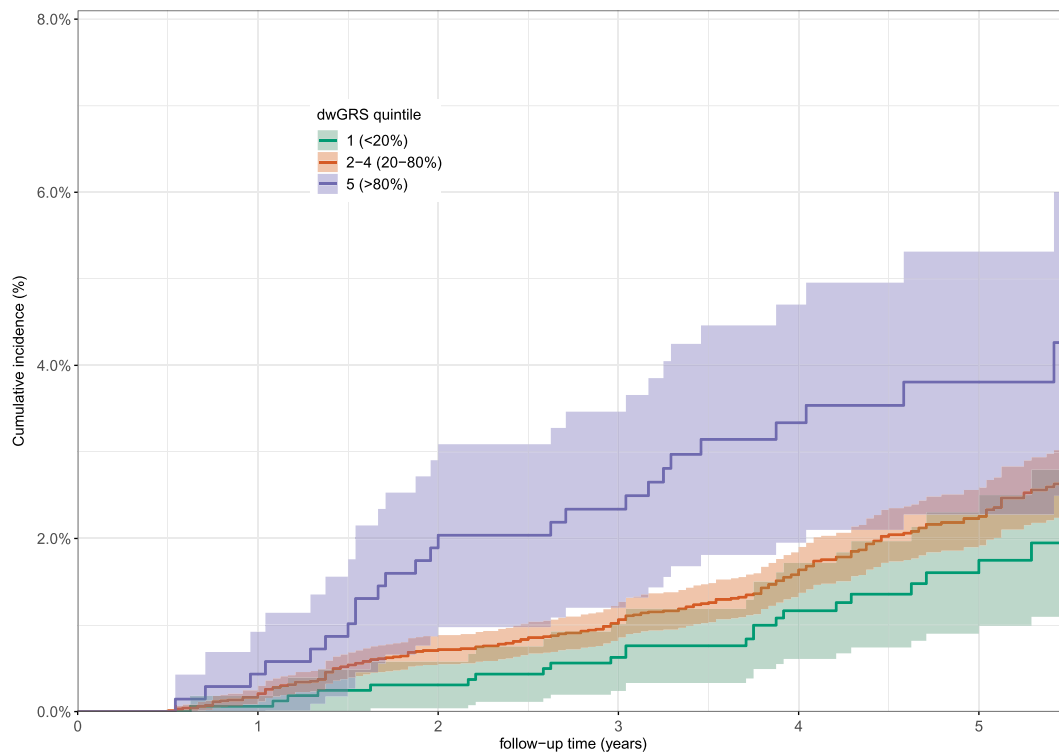
predictive value of dwGRS on incident T2D in both cohorts. The IDI was 0.0028 (−0.0001 to 0.0070;  $p = .138$ ) and 0.003 (0.002–0.004;  $p = 5 \times 10^{-05}$ ) in Lifelines and EstBB accordingly.

### 3.4 | dwGRS risk categories

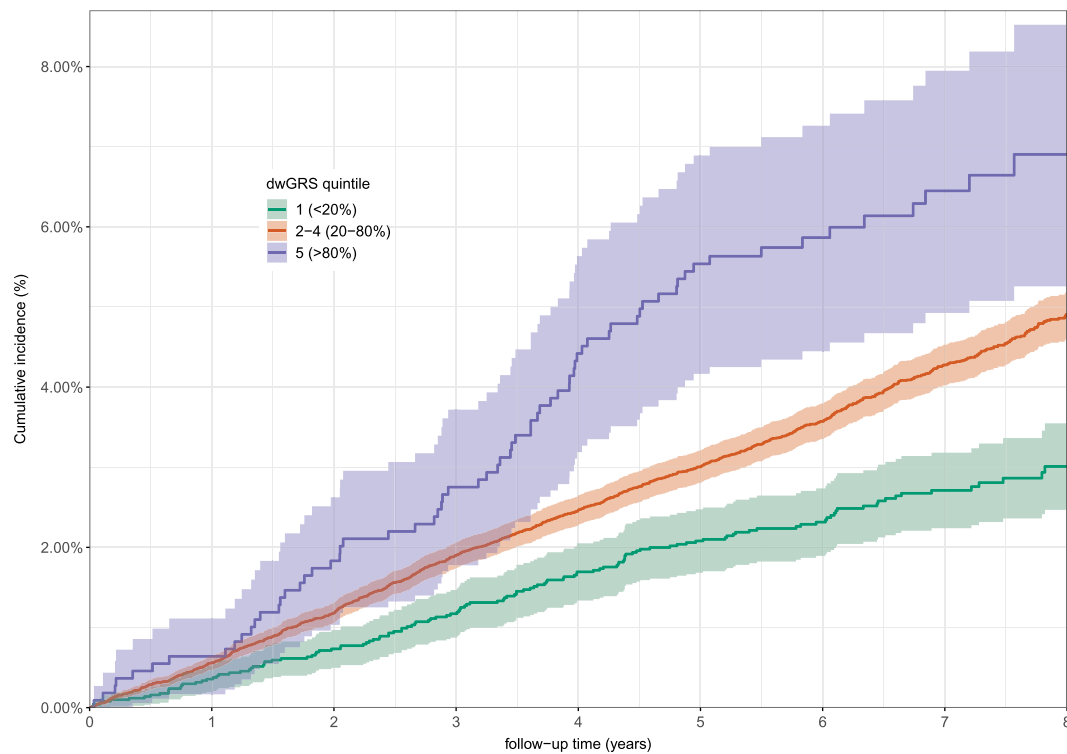
The cumulative incidences of T2D stratified by dwGRS categories are shown in Figure 1 (Lifelines) and 2

(EstBB). Categories were formed based on the dwGRS quintiles: three middle quintiles were combined to fulfill the proportionality assumption. In Lifelines, the logrank test  $p$ -value was .009 implying significant differences in T2D incidences between the dwGRS categories. In EstBB, there is a clear distinction in cumulative incidence of T2D between the dwGRS categories (logrank  $p = 2.95 \times 10^{-12}$ ).

Using the categorical dwGRS in the survival analysis in Lifelines reveals a risk of getting T2D that is 2.31 times



**FIGURE 1** Cumulative incidence of type 2 diabetes (T2D) by three categories of doubly weighted genetic risk score (dwGRS) in the Lifelines cohort. The dwGRS was stratified into quintiles, with quintiles 2–4 combined into one middle category. In the figure, only follow-up of 5.5-years is presented since only 25% of individuals were followed for longer



**FIGURE 2** Cumulative incidence of T2D by three categories of dwGRS in the Estonian Biobank cohort. The dwGRS was stratified into quintiles, with quintiles 2–4 combined into one middle category. In the figure, only follow-up of 8 years is presented since only 25% of individuals were followed for longer. dwGRS, doubly weighted genetic risk score; T2D, type 2 diabetes

(95% CI: 1.33–4.01;  $p = .002$ ) higher in the highest dwGRS quintile than in the lowest quintile of the dwGRS while accounting for the baseline phenotypic risk factors. In the middle category of quintiles 2–4 of the dwGRS, the risk of having incident T2D is 1.34 times (95% CI: 0.89–2.01;  $p = .163$ ) higher than in the lowest quintile. In EstBB, participants who belong to the highest dwGRS quintile, have a 2.81 times (95% CI: 2.15–3.68;  $p = 5.51 \times 10^{-14}$ ) higher risk for incident T2D compared to the lowest quintile of the dwGRS. The category including quintiles 2–4 of the dwGRS is again associated with a 1.66 times (95% CI: 1.40–1.96;  $p = 4.37 \times 10^{-09}$ ) higher risk. These results demonstrate the ability of the dwGRS to identify high-risk individuals for incident T2D in both cohorts.

#### 4 | DISCUSSION

The current study focused on testing the potential of the novel dwGRS when added to the prediction model of incident T2D for adult participants in the Lifelines and EstBB cohorts. Our main findings were that dwGRS had a significant effect on incident T2D risk independently from the traditional phenotypic risk factors and association of dwGRS with incident T2D was stronger than for swGRS in both biobanks.

The stronger association of dwGRS with incident T2D in both biobanks could be explained by two characteristics of the dwGRS that are different from the swGRS. First, the swGRS only incorporates SNPs below a certain significance threshold, but in the dwGRS all SNPs regardless of significance level could be used. Secondly, in the dwGRS an extra weight is added (i.e., the estimated probability of the SNP belonging among the  $k$  strongest associated SNPs) to correct for the Winner's curse bias. This bias arises when the SNPs are selected to be included in the GRS based on their statistical significance, giving stronger weight to the SNPs that have their effect overestimated by chance. If the GRS is constructed in this way, the extra weight shrinks the effect estimates of SNPs that have the strongest association with the phenotype, thus reducing this bias. As a result, a more accurate representation of the genetic risk is achieved, and stronger association between GRS and incident T2D obtained. The swGRS was also significantly associated with incident T2D, but the model fit was slightly lower than for the dwGRS. In other words, the dwGRS appeared to have more clinical value than the traditional swGRS. Therefore, in the current study, only the dwGRS was chosen for the analyzes on prediction improvement.

Our findings on the stronger association of the dwGRS with incident T2D were consistent with the



original study, where the dwGRS methodology was introduced (Läll et al., 2017). That study was also carried out on subjects from EstBB, but those subjects and their relatives were not included in the current study to get independent samples (Läll et al., 2017). The current study was the first one to test and demonstrate the validity of the doubly weighting method on a different part of the same cohort and on another external cohort for the association with incident T2D. The dwGRS improved the fit of the incident T2D prediction model as demonstrated by the increase in Harrell's C-statistics value in both cohorts. Testing the incremental value of dwGRS in a high-risk group (BMI: 25–30 and age: 35–79) in the EstBB resulted in an even larger increase in the value of the C-statistic, which demonstrates the prediction efficiency of the dwGRS, especially in high-risk individuals. We used the continuous NRI to investigate the clinical relevance of the dwGRS and confirmed a significantly better performance of the prediction model with dwGRS than without, in both biobanks. As concluded in previous literature (Läll et al., 2017; Reisberg et al., 2017), the added value of a GRS might seem to be small, but its real clinical value should be the ability to differentiate between the risk categories of incident T2D. For example, our results of having 2.31 and 2.81 times higher risk of incident T2D in the highest quintile of dwGRS compared to the lowest quintile in Lifelines and EstBB, respectively, while adjusting for age, sex, BMI, and hypertension (already established risk factors), clearly differentiates a group of individuals at high risk for T2D. Therefore, they should be under more frequent surveillance to help postpone or avoid the onset of incident T2D.

Strengths of the current study are the availability of two large European cohorts both with a wide range of predictors and follow-up data available. This enabled the current dwGRS validation study in two independent cohorts: (a) external validation in the Lifelines cohort, and (b) internal validation in the EstBB (using a different sample than originally). In addition, the new EstBB dataset was larger than the original one and additionally, due to a longer follow-up time, a larger number of incident T2D cases was available in the current study.

The SNP selection in the original dwGRS study was based on the meta-analysis of the GWASs of samples genotyped with Cardio-MetaboChip and OmniExpress platforms. Even though the Cardio-MetaboChip only contains SNPs from specific genes selected for cardio-metabolic traits and genome-wide data were available in the Lifelines cohort, we used the same selection of SNPs in Lifelines as in the original dwGRS study for comparability purposes (Läll et al., 2017). The similarity of the presented results in both large cohorts provides strong evidence for the additional value of dwGRS on top of

known environmental or phenotypic risk factors. However, future studies should be conducted to construct the optimal dwGRS using the full spectrum of genome-wide SNPs, which might result in an even more accurate prediction. Finally, the current SNP selection was based on the DIAGRAM meta-GWAS results from 2012, but newer meta-GWAS data with approximately eight-fold increase in sample size are already available providing more accurate SNP effects, which will likely result in a substantially improved dwGRS prediction (Mahajan et al., 2018).

Another strength of the current study was the extensive information about T2D diagnosis. In Lifelines, the official T2D diagnosis criteria (WHO, 2011) with supportive T2D medication information and both fasting plasma glucose and HbA1c measures were available at baseline. However, a limitation is that the exact date of T2D diagnosis was not known, that the T2D diagnosis at 1.5 and 3 years was based only on self-report, and that information on T2D medication use was not available at the 5-year follow-up. Nevertheless, it has been shown that the reliability of self-reported T2D is above 90% (Schneider, Pankow, Heiss, & Selvin, 2012), thus we believe that our results are valid. We were able to account for the unknown exact date of diagnosis by applying a more sophisticated survival analyzes method of interval-censored accelerated failure time modeling (Radke, 2003). Nevertheless, these limitations could be the cause of the slightly worse performance of the survival and prediction models in Lifelines. Other putative causes for this worse performance might be the smaller cohort size, the relatively shorter follow-up time available and consequently the lower number of incident T2D cases compared to the EstBB, and a lower imputation quality of the SNPs. The imputation quality for 22.8% of the SNPs included in the Lifelines dwGRS was below 0.8, which is usually regarded as low quality, compared to the imputation quality >0.9 for all the SNPs included in the dwGRS construction in EstBB. Having more SNPs with lower quality is due to the CytoSNP array used for genotyping in Lifelines, which has a low coverage of the whole genome (approximately 250,000 SNPs). Nevertheless, it has been shown in a study previously conducted in Lifelines that even SNPs with poor imputation quality improved the amount of trait variance explained when these were included in the GRS (Nolte et al., 2017). Another limitation of our study is the high number of missing values for some of the predictive environmental variables such as physical activity, fruit and vegetable consumption, and waist circumference. As a consequence, these variables were left out of the prediction models. Nevertheless, even the original FINDRISC study showed that physical activity and fruit and vegetable

consumption did not add much to the predictive power of the model (Lindstrom & Tuomilehto, 2003). Furthermore, waist circumference is highly correlated with BMI (0.82 in both cohorts), so it would not add much to the prediction model. Of note, our aim was not to construct the best prediction model for incident T2D with as many relevant predictive variables included as possible, but to test the additional value of dwGRS in addition to the phenotypic predictors.

#### 4.1 | Clinical implications and future research

Improving the identification of individuals at increased risk for T2D is urgently needed to combat the ongoing T2D epidemic and reduce high healthcare costs. Since in Estonia, there are highly favorable conditions for implementing personalized medicine, several pilot projects have already been initiated, in which EstBB participants are receiving health-related feedback with the dwGRS included in the risk-prediction algorithm for T2D. Furthermore, the dwGRS can be used to classify individuals into risk categories for T2D for application in the clinic. High-risk individuals should be monitored more frequently with the ultimate goal of delaying or even preventing the onset of T2D. In addition, the genetic risk score could be seen as a long-term predictor as opposed to the phenotypic risk factors, which may only become predictive in a relatively short time period before the disease onset. Therefore, we believe that the GRS could be a useful tool to improve public health through postponing or preventing disease onset and it should be added to the disease prediction algorithms if available. Nevertheless, future research focusing on improving the method for constructing the GRS is needed for even more precise T2D prediction. This method should also be applied to other complex diseases and cohorts of different ethnicity to test the generalizability and strengthen the validity of the dwGRS.

In conclusion, the dwGRS was associated with increased risk of incident T2D independently of phenotypic risk factors. This association was stronger than for the swGRS. The dwGRS improved the risk prediction and reclassification while accounting for already established phenotypic risk factors in both biobanks. Categorizing the dwGRS demonstrated the ability of the dwGRS to detect high-risk individuals for incident T2D, thus offering promise for personalized prediction and prevention.

#### ACKNOWLEDGMENTS

The authors wish to acknowledge the services of the Estonian Biobank and the Lifelines Cohort Study, the

contributing research centers delivering data, and all the study participants. The authors would like to thank Petra Vinke for sharing the syntax to calculate the Lifelines Diet Score for our study. The Lifelines Cohort Study, and generation and management of GWAS genotype data for the Lifelines Cohort Study is supported by the Netherlands Organization of Scientific Research NWO (grant 175.010.2007.006), the Economic Structure Enhancing Fund (FES) of the Dutch government, the Ministry of Economic Affairs, the Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports, the Northern Netherlands Collaboration of Provinces (SNN), the Province of Groningen, University Medical Center Groningen, the University of Groningen, Dutch Kidney Foundation and Dutch Diabetes Research Foundation. EGCUT received financing from the Estonian Research Council grants GP1GV9353 and IUT20-60, the Centre of Excellence for Genomics and Translational Medicine (GENTRANSMED), the University of Tartu (SPIGVARENG), the EU structural fund through the Archimedes Foundation, grant 3.2.1001.11-0033, and EU 2020 grant 692145 ePerMed. This work was supported by: European Union through the European Regional Development Fund, project No. 2014-2020.4.01.16-0024 (K. P.), and through the Horizon 2020 grant no. 777107- PRESICE4Q (K. F.). Estonian Research Council grant PUT PRG687 (K. L.).

#### CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

#### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

#### ORCID

Katri Pärna  <http://orcid.org/0000-0002-0013-6077>

#### REFERENCES

- American Diabetes Association (2017). Classification and diagnosis of diabetes. *Diabetes Care, Suppl 1*, S11–S24. <https://doi.org/10.2337/dc17-S005>
- Amini, M., Bashirova, D., Prins, B. P., Corpeleijn, E., Study, L. L. C., Bruinenberg, M., ... Alizadeh, B. Z. (2016). Eosinophil count is a common factor for complex metabolic and pulmonary traits and diseases: The lifelines cohort study. *PLoS One, 11*(12), e0168480. <https://doi.org/10.1371/journal.pone.0168480>
- Burton, P. R., Clayton, D. G., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., ... Compston, A. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature, 447*(7145), 661–678. <https://doi.org/10.1038/nature05911>
- Fischer, K., Kettunen, J., Würtz, P., Haller, T., Havulinna, A. S., Kangas, A. J., ... Metspalu, A. (2014). Biomarker profiling by

- nuclear magnetic resonance spectroscopy for the prediction of all-cause mortality: An observational study of 17,345 persons. *PLoS Medicine*, *11*(2), 1001606. <https://doi.org/10.1371/journal.pmed.1001606>
- Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*, *44*(8), 955–959. <https://doi.org/10.1038/ng.2354>
- International Diabetes Federation. (2019). IDF Diabetes Atlas 2019. In *International Diabetes Federation*. Retrieved from <http://www.idf.org/about-diabetes/facts-figures>
- Jansen, H., Stolk, R. P., Nolte, I. M., Kema, I. P., Wolffenbuttel, B. H. R., & Snieder, H. (2013). Determinants of HbA1c in nondiabetic Dutch adults: Genetic loci and clinical and lifestyle parameters, and their interactions in the lifelines cohort study. *Journal of Internal Medicine*, *273*(3), 283–293. <https://doi.org/10.1111/joim.12010>
- Janssen, I., Katzmarzyk, P. T., & Ross, R. (2002). Body mass index, waist circumference, and health risk. *Archives of Internal Medicine*, *162*(18), 2074–2079. <https://doi.org/10.1001/archinte.162.18.2074>
- Kahn, S. E., Cooper, M. E., & Del Prato, S. (2014). Pathophysiology and treatment of type 2 diabetes: Perspectives on the past, present, and future. *The Lancet*, *383*(9922), 1068–1083. [https://doi.org/10.1016/S0140-6736\(13\)62154-6](https://doi.org/10.1016/S0140-6736(13)62154-6)
- Klijs, B., Scholtens, S., Mandemakers, J. J., Snieder, H., Stolk, R. P., & Smidt, N. (2015). Representativeness of the LifeLines cohort study. *PLoS One*, *10*(9), 1–12. <https://doi.org/10.1371/journal.pone.0137203>
- Kundu, S., Aulchenko, Y. S., Van Duijn, C. M., & Janssens, A. C. J. W. (2011). PredictABEL: An R package for the assessment of risk prediction models. *European Journal of Epidemiology*, *26*, 261–264. <https://doi.org/10.1007/s10654-011-9567-4>
- Langenberg, C., Sharp, S. J., Franks, P. W., Scott, R. A., Deloukas, P., Forouhi, N. G., ... Wareham, N. J. (2014). Gene-lifestyle interaction and type 2 diabetes: The EPIC InterAct case-cohort study. *PLoS Medicine*, *11*(5), e1001647. <https://doi.org/10.1371/journal.pmed.1001647>
- Leitsalu, L., Alavere, H., Tammesoo, M. L., Leego, E., & Metspalu, A. (2015). Linking a population biobank with national health registries—the Estonian experience. *Journal of Personalized Medicine*, *5*(2), 96–106. <https://doi.org/10.3390/jpm5020096>
- Leitsalu, L., Haller, T., Esko, T., Tammesoo, M. L., Alavere, H., Snieder, H., ... Metspalu, A. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *International Journal of Epidemiology*, *44*(4), 1137–1147. <https://doi.org/10.1093/ije/dyt268>
- Lindstrom, J., & Tuomilehto, J. (2003). The diabetes risk score: A practical tool to predict type 2 diabetes risk. *Diabetes Care*, *26*(3), 725–731.
- Lyssenko, V., Jonsson, A., Almgren, P., Pulizzi, N., Isomaa, B., Tuomi, T., ... Groop, L. (2008). Clinical risk factors, DNA variants, and the development of type 2 diabetes. *New England Journal of Medicine*, *359*(21), 2220–2232. <https://doi.org/10.1056/NEJMoa0801869>
- Läll, K., Lepamets, M., Palover, M., Esko, T., Metspalu, A., Tõnisson, N., ... Fischer, K. (2019). Polygenic prediction of breast cancer: Comparison of genetic predictors and implications for risk stratification. *BMC Cancer*, *19*(1), 1–9. <https://doi.org/10.1186/s12885-019-5783-1>
- Läll, K., Mägi, R., Morris, A., Metspalu, A., & Fischer, K. (2017). Personalized risk prediction for type 2 diabetes: The potential of genetic risk scores. *Genetics in Medicine*, *19*(3), 322–329. <https://doi.org/10.1038/gim.2016.103>
- Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., ... McCarthy, M. I. (2018). Fine-mapping of an expanded set of type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *BioRxiv*, 245506. <https://doi.org/10.1101/245506>
- McCarthy, M. I. (2010). Genomics, type 2 diabetes, and obesity. *The New England Journal of Medicine*, *363*(24), 2339–2350.
- Meigs, J. B., Cupples, L. A., & Wilson, P. W. F. (2000). Parental transmission of type 2 diabetes: The Framingham Offspring study. *Diabetes*, *49*(12), 2201–2207.
- Mejnikman, A. S., De Block, C. E. M., Verrijken, A., Mertens, I., Corthouts, B., & Van Gaal, L. F. (2016). Screening for type 2 diabetes mellitus in overweight and obese subjects made easy by the FINDRISC score. <https://doi.org/10.1016/j.jdiacomp.2016.05.004>
- Messerli, F. H., Williams, B., & Ritz, E. (2007). Essential hypertension. *Lancet*, *370*(9587), 591–603. [https://doi.org/10.1016/S0140-6736\(07\)61299-9](https://doi.org/10.1016/S0140-6736(07)61299-9)
- Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segré, A. V., Steinthorsdottir, V., ... McCarthy, M. I. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics*, *44*(9), 981–990. <https://doi.org/10.1038/ng.2383>
- Nolte, I. M., van der Most, P. J., Alizadeh, B. Z., de Bakker, P. I., Boezen, H. M., Bruinenberg, M., ... Snieder, H. (2017). Missing heritability: Is the gap closing? An analysis of 32 complex traits in the Lifelines Cohort study. *European Journal of Human Genetics*, *25*(7), 877–885. <https://doi.org/10.1038/ejhg.2017.50>
- Palmer, N. D., McDonough, C. W., Hicks, P. J., Roh, B. H., Wing, M. R., An, S. S., ... Mooser, V. (2012). A genome-wide association search for type 2 diabetes genes in african americans. *PLoS One*, *7*(1), e29202. <https://doi.org/10.1371/journal.pone.0029202>
- Pencina, M. J., D'Agostino, R. B., D'Agostino, R. B., & Vasan, R. S. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine*, *27*(2), 157–172. <https://doi.org/10.1002/sim.2929>
- Pencina, Steyerberg, W. E., & D'Agostino, B. R., Sr. (2011). Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Statistics in Medicine*, *30*(1), 11–21. <https://doi.org/10.2217/FON.09.6.Dendritic>
- Poulsen, P., Ohm Kyvik, K., Vaag, A., & Beck-Nielsen, H. (1999). Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance: A population-based twin study. *Diabetologia*, *42*(2), 139–145. <https://doi.org/10.1007/s001250051131>
- Prasad, R. B., & Groop, L. (2015). Genetics of type 2 diabetes—pitfalls and possibilities. *Genes*, *6*(1), 87–123. <https://doi.org/10.3390/genes6010087>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based

- linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- R Development Core Team. (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0
- Radke, B. R. (2003). A demonstration of interval-censored survival analysis. *Preventive Veterinary Medicine*, 59(4), 241–256. [https://doi.org/10.1016/S0167-5877\(03\)00103-X](https://doi.org/10.1016/S0167-5877(03)00103-X)
- Reisberg, S., Iljasenko, T., Läll, K., Fischer, K., & Vilo, J. (2017). Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One*, 12(7), 0179238. <https://doi.org/10.1371/journal.pone.0179238>
- Schneider, A. L. C., Pankow, J. S., Heiss, G., & Selvin, E. (2012). Validity and reliability of self-reported diabetes in the atherosclerosis risk in communities study. *American Journal of Epidemiology*, 176(8), 738–743. <https://doi.org/10.1093/aje/kws156>
- Scholtens, S., Smidt, N., Swertz, M. A., Bakker, S. J., Dotinga, A., Vonk, J. M., ... Stolk, R. P. (2015). Cohort profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*, 44(4), 1172–1180. <https://doi.org/10.1093/ije/dyu229>
- Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, L., ... Gonçalo, R. (2007). A genome-wide association study of type 2 diabetes in Finns detect multiple susceptibility variants. *Science*, 316(5829), 1341–1345. <https://doi.org/10.1126/science.1142382.A>
- Siebelink, E., Geelen, A., & de Vries, J. H. M. (2011). Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *British Journal of Nutrition*, 106(2), 274–281. <https://doi.org/10.1017/s0007114511000067>
- The 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68–74. <https://doi.org/10.1038/nature15393>
- Vinke, P. C., Corpeleijn, E., Dekker, L. H., Jacobs, D. R., Navis, G., & Kromhout, D. (2018). Development of the food-based Lifelines Diet Score (LLDS) and its application in 129, 369 Lifelines participants. *European Journal of Clinical Nutrition*, 72(8), 1111–1119. <https://doi.org/10.1038/s41430-018-0205-z>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, 101(1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Wendel-Vos, G. C. W., Schuit, A. J., Saris, W. H. M., & Kromhout, D. (2003). Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *Journal of Clinical Epidemiology*, 56(12), 1163–1169. [https://doi.org/10.1016/S0895-4356\(03\)00220-8](https://doi.org/10.1016/S0895-4356(03)00220-8)
- Witte, D. R., Shipley, M. J., Marmot, M. G., & Brunner, E. J. (2010). Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study. *Diabetic Medicine*, 27(1), 46–53. <https://doi.org/10.1111/j.1464-5491.2009.02891.x>
- World Health Organization. (2011). Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus. 1–25. <https://doi.org/10.1016/j.diabres.2011.03.012>
- World Health Organization. (2014). Global status report on noncommunicable diseases 2014. *World Health Organization*, 176. ISBN 9789241564854
- World Health Organization. (2016). ICD10 – version: 2016. Retrieved September 10, 2017, from World Health Organization website: <http://apps.who.int/classifications/icd10/browse/2016/en#/I25>
- World Health Organization. (2017). WHOC – ATC/DDD Index. <https://doi.org/10.1002/0471684228.egp13486>
- Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14(7), 507–515. <https://doi.org/10.1038/nrg3457>
- Van Zon, S. K. R., Reijneveld, S. A., Van Der Most, P. J., Swertz, M. A., Bültmann, U., & Snieder, H. (2018). The interaction of genetic predisposition and socioeconomic position with type 2 diabetes mellitus: Cross-sectional and longitudinal analyses from the lifelines cohort and biobank study. *Psychosomatic Medicine*, 80(3), 252–262. <https://doi.org/10.1097/PSY.0000000000000562>
- Zhang, L., Zhang, Z., Zhang, Y., Hu, G., & Chen, L. (2014). Evaluation of Finnish diabetes risk score in screening undiagnosed diabetes and prediabetes among U.S. adults by gender and race: NHANES 1999–2010. *PLoS ONE*, 9(5), <https://doi.org/10.1371/journal.pone.0097865>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Pärna K, Snieder H, Läll K, Fischer K, Nolte I. Validating the doubly weighted genetic risk score for the prediction of type 2 diabetes in the Lifelines and Estonian Biobank cohorts. *Genetic Epidemiology*. 2020;44:589–600. <https://doi.org/10.1002/gepi.22327>