Self-reported walking pace, polygenic risk scores and risk of coronary artery disease in UK biobank

F. Zaccardi a,b,*, I.R. Timmins c, J. Goldney b, F. Dudbridge c, P.C. Dempsey b,d, M.J. Davies b,d, K. Khunti a,b,e, T. Yates a,b,d

a Leicester Real World Evidence Unit, University of Leicester, Leicester General Hospital, Gwendolen Rd, Leicester, LE5 4PW, UK
b Diabetes Research Centre, University of Leicester, Leicester General Hospital, Gwendolen Rd, Leicester, LE5 4PW, UK
c Department of Health Sciences, University of Leicester, George Davies Centre, University Road, Leicester, LE1 7RH, UK
d NIHR Leicester Biomedical Research Centre, University Hospitals of Leicester NHS Trust and University of Leicester, Leicester, LE5 4PW, UK
e NIHR Collaboration for Leadership in Applied Health Research and Care - East Midlands, University as Leicester, Leicester, LE1 7RH, UK

Received 27 April 2022; received in revised form 15 August 2022; accepted 23 August 2022

Handling Editor: P. Russo

Available online

KEYWORDS
Walking pace; Walking speed; Coronary artery disease; Genetics; Polygenic risk score; UK Biobank

Abstract  Background and aims: Both polygenic risk scores (PGS) and self-reported walking pace have been shown to predict cardiovascular disease; whether combining both factors produces greater risk differentiation is, however, unknown.

Methods and results: We estimated the 10-year absolute risk of coronary artery disease (CAD), adjusted for traditional risk factors, and the C-index across nine PGS and self-reported walking pace in UK Biobank study participants between Mar/2006–Feb/2021. In 380,693 individuals (54.8% women), over a median (5th, 95th percentile) of 11.9 (8.3, 13.4) years, 2,603 (1.2%) CAD events occurred in women and 8,259 (4.8%) in men. Both walking pace and genetic risk were strongly associated with CAD. The absolute 10-year risk of CAD was highest in slow walkers at high genetic risk (top 20% of PGS): 2.72% (95% CI: 2.30–3.13) in women; 9.60% (8.62–10.57) in men. The risk difference between slow and brisk walkers was greater at higher [1.26% (0.81–1.71) in women; 3.63% (2.58–4.67) in men] than lower [0.76% (0.59–0.93) and 2.37% (1.96–2.78), respectively] genetic risk. Brisk walkers at high genetic risk had equivalent (women) or higher (men) risk than slow walkers at moderate-to-low genetic risk (bottom 80% of PGS). When added to a model containing traditional risk factors, both factors separately improved risk discrimination; combining them resulted in the greatest discrimination: C-index of 0.801 (0.793–0.808) in women; 0.732 (0.728–0.737) in men.

Conclusion: Self-reported slow walkers at high genetic risk had the greatest risk of CAD, identifying a potentially important population for intervention. Both PGS and walking pace contributed to risk discrimination.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of The Italian Diabetes Society, the Italian Society for the Study of Atherosclerosis, the Italian Society of Human Nutrition and the Department of Clinical Medicine and Surgery, Federico II University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding author. Leicester Real World Evidence Unit, University of Leicester, Leicester General Hospital, Gwendolen Rd, Leicester, LE5 4PW, UK
E-mail address: frazac@fastwebnet.it (F. Zaccardi).

https://doi.org/10.1016/j.numecd.2022.08.021
0939-4753/© 2022 The Author(s). Published by Elsevier B.V. on behalf of The Italian Diabetes Society, the Italian Society for the Study of Atherosclerosis, the Italian Society of Human Nutrition and the Department of Clinical Medicine and Surgery, Federico II University. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Please cite this article as: Zaccardi F et al., Self-reported walking pace, polygenic risk scores and risk of coronary artery disease in UK biobank, Nutrition, Metabolism & Cardiovascular Diseases, https://doi.org/10.1016/j.numecd.2022.08.021
1. Introduction

Coronary artery disease (CAD), the most common form of cardiovascular disease, is heritable: over 300 independent genetic loci are known to influence CAD risk, with overall heritability estimated between 40% and 60% [1–5]. With the advent of large-scale genetic sequencing, there has been a proliferation of research in screening for alleles which are associated with CAD in order to explain this heritability. Polygenic risk scores (PGS) harness this genetic information to determine an individual's genetic risk with future implications for early intervention and stratified medicine within clinical care [6]. Whilst PGS have been shown to predict the risk of CAD, their clinical utility depends on risk prediction over and above routinely or easily collected traditional risk factors. To date, there have been mixed findings for the value of PGS in predicting CAD events when added to traditional risk scores, ranging from no to modest improvement [6–9]. However, PGS have proven more accurate than other clinical risk factors when assessed individually and recent developments within PGS research have increased their complexity to >1 million single-nucleotide polymorphisms (SNPs), resulting in a larger proportion of the population identified at very high risk than rare monogenic mutations [10,11].

Along with the rapid progress in PGS development and complexity, other research suggests that simple and easy to collect measures of health status or lifestyle behaviors also have utility in identifying high-risk groups for CAD events, which could aid risk prediction and stratified medicine. One of the strongest candidates to date has been self-reported walking pace, which is associated with cardiorespiratory fitness and is a stronger independent predictor of survival and cardiovascular mortality than a wide range of other lifestyle factors, including handgrip strength [12–14]. Indeed, self-reported walking pace has better predictive discrimination for cardiovascular mortality than traditional clinical risk factors, such as serum cholesterol and blood pressure, and within UK Biobank provides better predictive discrimination for all-cause and cardiovascular disease mortality than most other clinical or sociodemographic variables [14,15].

Despite the potential utility of both PGS and walking pace in CAD risk prediction, previous research has not investigated their comparative or combined relevance. In this study, we aimed to investigate the absolute risk of CAD with PGS across different categories of self-reported walking pace within UK Biobank and to compare their prognostic importance. We hypothesized that walking pace would act to differentiate the risk of CAD across the spectrum of genetic risk, with slow walkers at high generic risk having the highest incidence of CAD.

2. Methods

2.1. Cohort definition

We used data from UK Biobank, an ongoing prospective cohort study, collected between March 2006 and July 2010 in women and men aged between 38 and 73 years recruited from 22 centers throughout England, Wales, and Scotland (UK Biobank Application Number 33266). Individuals were recruited from family practices within 25 miles (40 km) of the assessment centers. Written consent was obtained. From the initial sample of 502,599 participants, we excluded participants who withdrew consent during the study and, at baseline: were pregnant; self-reported a doctor diagnosis of cancer or chronic kidney disease; or had prevalent CAD. CAD events before the baseline visit were identified using information on hospital admissions (Hospital Episode Statistics, HES) linked to UK Biobank, based on the International Classification of Diseases (ICD) diagnostic codes (ICD-9: 410–412; ICD-10: I21-I24, I25.2) or coronary artery bypass graft (CABG) and percutaneous transluminal coronary angioplasty (PTCA) procedure codes (OPCS-4: K40 to K46, K49, K50.1, or K75), either in the primary or secondary position in the hospital records. Of the 452,072 remaining participants, a further 67,148 were excluded due to missing covariate data (Supplementary Fig. S1).

2.2. Self-reported walking pace

A touchscreen questionnaire was used to capture usual walking pace at baseline. Participants were asked to answer the following question: “How would you describe your usual walking pace: slow; steady/average; brisk; none of the above; prefer not to answer” Further information was available to participants which clarified a slow pace as <3 miles per hour (mph), a steady/average pace as 3–4 mph, and a brisk pace as >4 mph.

2.3. Genetic data processing

In the UK Biobank, genotyping was performed using the UK BILIVE Axiom Array and the UK Biobank Axiom arrays, with imputation to the Haplotype Reference Consortium panel [16]. We further excluded 4,231 samples following the genotype quality filtering performed centrally by the UK Biobank — the details of which have been described extensively elsewhere [17], leaving 380,693 subjects for the analyses (Supplementary Fig. S1).

2.4. Polygenic risk scores

We generated PGS for CAD using scores identified and downloaded from nine studies in the publicly available PGS Catalog repository (Supplementary Table S1) [18]. PGS are defined for each individual as the number of risk alleles at each variant in the score, with weights assigned to each variant based on the strength of their association with CAD risk; the greater the score, the higher the genetic risk of CAD.

We used PRSice–2 software to compute PGS for each UK Biobank individual [19]. Ambiguous SNPs with A/T or C/G strands were removed from the scores, as previously recommended [20].
2.5. Confounding variables

Data were also captured for the following putative risk factors: age, sex, social deprivation (Townsend deprivation index, with a higher index indicating a greater degree of deprivation), systolic blood pressure, low-density lipoprotein (LDL) cholesterol, smoking status (current, former, never), history of diabetes mellitus (type 1 or type 2), and family (maternal or paternal) history of myocardial infarction.

2.6. Outcome

We identified incident fatal and non-fatal CAD events using the same ICD and OPCS codes employed to define prevalent CAD at baseline, either in primary or secondary position (see Cohort Definition section). Date and cause of death were obtained with linkage of UK Biobank to NHS Digital in participants from England and Wales and to the NHS Central Register in participants from Scotland. Participants were followed-up between study entry (baseline visit) until the occurrence of the study outcome or censoring (February 28, 2021 for England and Scotland; February 28, 2018 for Wales).

2.7. Statistical analysis

Descriptive values are reported as median and interquartile range for continuous variables and number and percentage for categorical ones. We used the Royston-Parmar-Lambert parametric survival model [21], with study entry to first CAD event as time scale, to investigate the absolute risk of CAD across the continuous PGS, standardized to zero mean and unit variance, and self-reported walking pace. Survival models included an interaction term between self-reported walking pace and PGS; p-values for interactions were estimated with the likelihood ratio test without accounting for multiplicity. Models were adjusted for traditional CAD risk factors: age (continuous), Townsend deprivation index (continuous), systolic blood pressure (continuous), LDL cholesterol (continuous), smoking status (current/former/never), history of diabetes (yes/no), and family history of myocardial infarction (yes/no). We estimated the standardized (adjusted) 10-year risk of CAD quantifying individual risks and averaging them across levels of walking pace [22]. For descriptive purposes and to aid interpretation of results, we defined high genetic risk as the top 20% of PGS distribution and moderate-to-low genetic risk as the bottom 80% of PGS distribution [11]. We compared the risk between the PGS distributions and self-reported walking pace.

We complemented our analysis by investigating the comparative prognostic role of non-genetic and genetic risk factors using Harrell’s C-index [23]. The change in the C-index (ΔC-index) and its confidence interval (CI), obtained from a linear combination of the two estimates [24], were further calculated when walking pace and each PGS were added separately to a base model containing traditional CAD risk factors. The ΔC-index was also estimated from the addition of walking pace to models containing the best performing PGS and traditional CAD risk factors.

All analyses were stratified by sex; there were insufficient ethnic minority participants to allow for further stratification by race or ethnicity. Stata routines, stpm2, and stadsurv commands were used in Stata/BE Version 17.0 (StataCorp, 2021. College Station, TX, USA) and results are reported with 95% CI; graphs were prepared in Stata and Inkscape Version 1.1. Statistical codes are publicly available on GitHub (frazac82) and at UK Biobank, in line with UK Biobank regulations. All aggregate results are reported in the Supplementary Excel file.

2.8. Ethical approval

Ethical approval for the UK Biobank study was obtained from the North West Centre for Research Ethics Committee (MREC, 11/NW/0382). In Scotland, UK Biobank has approval from the Community Health Index Advisory Group (CHIAG). The study complies with the Declaration of Helsinki.

3. Results

3.1. Cohort characteristics

Table 1 shows the cohort characteristics of the 208,627 (54.8%) women and 172,066 (45.2%) men included in the analysis: the large majority (356,299; 93.6%) were white Europeans, followed by mixed ethnicity group (8,672; 2.3%); other ethnicities contributed to the remaining 4.1%. Over a median (5th to 95th percentile) follow-up of 11.9 (8.3–13.4) years, 2,603 (1.2%) CAD events occurred in women and 8,259 (4.8%) in men, equating to 1.1 (95% Confidence Interval [CI]: 1.0 to 1.1) and 4.2 (4.1–4.3) events per 1,000 person-years, respectively. The cohort characteristics of women and men stratified by self-reported walking pace are shown in Table S2.

3.2. 10-year risk of CAD

The PGS distributions by incident coronary artery disease event and by walking pace are shown in Fig. S2 and Fig. S3, respectively. Evidence of significant interactions between walking pace and PGS was observed for some, but not all, PGS, with a progressively lower relative risk reduction comparing average or brisk vs slow pace for a higher genetic risk (Table S3).

Both walking pace and genetic risk were strongly associated with the 10-year risk of CAD (Fig. 1; Fig. S4). Among the investigated PGS, the two largest (GPS_CAD [PGS-13] and metaGRS_CAD [PGS-18]) were most strongly associated with the 10-year risk of CAD across categories of walking pace. We used the largest PGS, GPS_CAD [PGS-13], to support the main interpretations from the analyses (Fig. 1); however, all other PGS, particularly metaGRS_CAD [PGS-18], were associated in a similar direction and magnitude (Fig. S4). The risk of CAD was lowest in brisk walkers with low genetic risk and highest in slow walkers at high genetic risk. For GPS_CAD [PGS-13], the 10-
year risk of CAD for slow walkers at high genetic risk (top 20% of PGS) was 2.72% (95% CI: 2.30 to 3.13) in women and 9.60% (8.62–10.57) in men (Fig. 2). In women, the risk of CAD in brisk walkers at high genetic risk was similar to the risk in slow walkers at moderate-to-low genetic risk (bottom 80% of PGS) while in men brisk walkers at high genetic risk had a higher risk of CAD than slow walkers at moderate-to-low genetic risk. The predicted 10-year risk of CAD for brisk walker women at high genetic risk was 1.31% (0.40–2.23), while in slow walkers with moderate-to-low genetic risk it was 0.94% (0.40–1.49). The difference between brisk and slow walkers was greater at higher genetic risk in the best performing PGS (Fig. 1) while it was more variable for the other PGS (Fig. S5).

For GPS_CAD [PGS-13], in women the difference in 10-year risk of CAD between slow and brisk walker was 1.26% (0.81–1.71) at high genetic risk and 0.76% (0.59–0.93) for moderate-to-low genetic risk (Fig. S6), equating to a between-genetic risk difference comparing slow vs brisk of 0.50% (0.02–0.97). For men, the corresponding differences in the absolute risk between slow and brisk walkers were 1.50% (0.85–2.15) at high genetic risk and 1.02% (0.63–1.42) at moderate-to-low genetic risk (Fig. S6), resulting in a between-genetic risk difference comparing slow vs brisk of 0.48% (0.01–0.97).

### 3.3. Risk discrimination

The findings for associations translated into risk discrimination. Walking pace provided a similar risk discrimination for CAD to all but the best performing PGS (GPS_CAD [PGS-13], metaGRS_CAD [PGS-18], and CAD_GRS_204 [PGS-58]) which provided greater risk discrimination in both men and women (Fig. S7). Walking pace also modestly improved risk discrimination when added to a base model including traditional CAD risk factors: in women, the C-index increased from 0.763 (95% CI: 0.755 to 0.771) in the base model to 0.769 (0.761–0.777) upon the inclusion of walking pace, resulting in a ΔC-index of 0.006 (0.003–0.008); corresponding C-index values in men were: 0.688 (0.682–0.693), 0.695 (0.690–0.700) with a ΔC-index of 0.007 (0.006–0.009; Fig. 3). In both women and men, the additional risk discrimination from adding walking pace to a base model of traditional risk factors was similar to the additional discrimination from adding PGS, with the exception of the best performing PGS, i.e. GPS_CAD [PGS-13], metaGRS_CAD [PGS-18], and CAD_GRS_204 [PGS-58], which resulted in the greatest risk discrimination (Fig. 3).

The addition of both walking pace and PGS to a base model containing traditional risk factors resulted in the model with the best risk discrimination. A model containing GPS_CAD [PGS-13], walking pace and traditional risk factors resulted in a C-index of 0.801 (95% CI: 0.793 to 0.808) in women and 0.732 (0.728–0.737) in men, corresponding to a ΔC-index from the model containing just traditional risk factors of 0.038 (0.032–0.043) and 0.045 (0.041–0.049) in women and men, respectively (Fig. 3).

The ΔC-index for walking pace and all PGS was similar between men and women, with the exception of the addition of PGS-13 (estimated difference between ΔC-index in men vs ΔC-index in women: 0.007 [0.0005 to 0.013]) and of the combination of WP and PGS13 (0.007 [0.001 to 0.014]).

### 4. Discussion

A growing body of knowledge has shown the importance of PCS and walking pace as individual risk factors for CAD. We show that when these risk factors are considered...
together, walking pace acts to differentiate risk across the continuum of PGS, particularly in those at high genetic risk where the difference in risk between slow and brisk walkers was greatest. Nevertheless, brisk walkers at high genetic risk had equivalent (women) or higher (men) risk of CAD than slow walkers at moderate-to-low genetic risk. Walking pace and PGS modestly or moderately improved risk discrimination over traditional risk factors, with the greatest improvement occurring when both factors were added to the same model.

As PGS continue to develop in complexity, their utility as predictors of risk has increased. This was evident in the current study where the largest PGS demonstrated the strongest association with the risk of CAD. Although walking pace has been shown to be one of the best lifestyle predictors of all cause-mortality and cardiovascular

---

**Figure 1** 10-year risk of coronary artery disease across walking pace and GPS_CAD genetic risk score. Top panels: 10-year risk of coronary artery disease for slow (blue), average (orange) and brisk (green) walking pace across standardised (x-axis) polygenic risk score. Bottom panels: 10-year risk difference of coronary artery disease for slow vs average (orange) and slow vs brisk (green) walking pace. Estimates, adjusted for age, Townsend score, systolic blood pressure, LDL cholesterol, smoking status, history of diabetes, and family history of myocardial infarction, are shown for ±3 standard deviations (99.7% of distribution) of the polygenic risk score. Areas indicate 95% confidence interval. (For interpretation of the references to color/colour in this figure legend, the reader is referred to the Web version of this article.)

Please cite this article as: Zaccardi F et al., Self-reported walking pace, polygenic risk scores and risk of coronary artery disease in UK biobank, Nutrition, Metabolism & Cardiovascular Diseases, https://doi.org/10.1016/j.numecd.2022.08.021
disease [12–15], men who were brisk walkers at high genetic risk had a higher risk of CAD than slow walkers at moderate-to-low genetic risk. However, walking pace remained an important discriminator of CAD risk across the continuum of PGS. Indeed, the difference in the risk of CAD between slow and brisk walkers was greatest at high genetic risk. This finding builds upon a previous UK Biobank study, which suggested that the association between cardiorespiratory fitness and risk of coronary heart disease was greatest in those at high genetic risk [25]. Therefore, those with a slow walking pace or low fitness and high genetic risk may carry a particularly elevated risk of CAD. This has important implications. Although genetic risk remains fixed, walking pace and cardiorespiratory fitness are modifiable. Indeed, it has previously been shown that self-reported walking pace is causally associated with CAD and other cardiometabolic risk traits [26], including biological age [27]. In addition, exercise-based cardiac or pulmonary rehabilitation programs, which are commonly based on walking activities and evaluated by walking pace tests [28], have been shown to reduce hospital admissions and mortality [29,30]. Therefore, it is plausible that exercise-based rehabilitation style interventions for slow walkers at high genetic risk could play an important role for the primary prevention of CAD, a possibility that warrants further investigation.

In line with our research question, we included interaction terms a-priori for all models based on theoretical considerations, rather than a-posteriori on statistical significance [31,32]. Our results indicated, for virtually all PGS, a lower relative risk reduction of CAD associated with average or brisk pace for higher values of the PGS, yet interactions were not "statistical significant" for all PGS. These results, suggesting narrowing differences across walking pace levels in individuals at higher PGS scores, should be interpreted alongside the absolute risk estimates, which are more relevant from a public health perspective [33]. We also stratified the analyses by sex as we expected different absolute risks of CAD in women vs men; while we confirmed a higher risk in men — particularly at high genetic risk, values of C-indices were higher in women in models adding PGS, walking pace, and their combination to the traditional cardiovascular risk factors. This finding, which aligns with previous evidence on sex differences in C-indexes in cardiovascular disease models [9], would suggest that other (causal and non-causal) factors not included in our investigations could improve model discrimination in men. The C-index, however, is only one measure of model performance and other metrics (i.e., calibration) are required for a more comprehensive assessment of a prognostic model [34].

Strengths of this study include the large contemporary population with linkage to genetic data and CAD outcomes. The use of self-reported walking pace is a potential strength as it is a simple to measure risk factor that can easily be incorporated into future research and clinical care pathways at low cost. Furthermore, the self-reported walking pace item used in UK Biobank has been shown to be associated with cardiorespiratory fitness [12], and self-reported walking pace has more generally been observed to be strongly associated with objectively assessed walking pace [35].

The use of self-reported walking pace also has the limitation of being a subjective marker and therefore at risk of reporting bias. This, combined with the observational nature of the study, means that the findings should be interpreted as highlighting a high-risk group (self-reported slow walkers), with risk persisting across genetic susceptibility. Etiological conclusions should be viewed more cautiously and in the context of hypothesis generation. It is also acknowledged that PGS are becoming more complex with improving potential for
greater risk discrimination; therefore, the degree of association with CAD outcomes highlighted here is anticipated to strengthen with time. GPS_CAD [PGS-13] and metaGRS_CAD [PGS-18], two of the PGS showing greatest risk discrimination, were both validated on the UK Biobank population, overlapping the sample used in this analysis. This means that the scores may overfit this population and result in biased estimates compared to their use in the general population. Finally, the UK Biobank cohort is overrepresented by white Europeans and the participants are healthier than the general population [36], which may limit generalizability. Nevertheless, even in this healthier population, men who were slow walkers at high genetic risk had an average absolute 10-year risk of over 9.0%.

In conclusion, this study found that in women and men, brisk walkers at low genetic risk had the lowest risk of CAD, whereas slow walkers at high genetic risk had the greatest risk. The addition of either PGS or self-reported walking pace to a model containing traditional risk factors improved CAD risk discrimination and the greatest discrimination was observed in a model containing both walking pace and PGS. Further research is needed to investigate whether these results can be used to inform and improve established prognostic models. Self-reported walking pace, in particular, has potential utility given that it is easy to collect, having few resource implications. Finally, as self-reported walking pace can also be considered a marker of whole-body physical fitness and function, the efficacy of adapting cardiac rehabilitation type interventions for primary prevention in individuals with poor physical fitness or function and high genetic risk should also be considered.

Funding

National Institute for Health and Care Research (NIHR), Leicester Biomedical Research Centre, Leicester, UK [TY, MJD]. The NIHR Applied Health Research and Care - East Midlands [KK]. TY is supported by a grant from the Medical Research Council (MR/T031816/1).

Declaration of competing interest

The authors declare there are no conflicts of interest and have no financial disclosures to report.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 33266. Statistical codes are publicly available on GitHub (frazac82) and at UK Biobank, in line with UK Biobank regulations. Data from UK Biobank are available to researchers on application https://www.ukbiobank.ac.uk/.

Figure 3  Discrimination of models across non-genetic and genetic risk factors. C-indexes for models including all non-genetic risk factors (NGRF: age, Townsend score, systolic blood pressure, low-density lipoprotein cholesterol, smoking status, history of diabetes, family history of myocardial infarction) are shown as black circles; values upon the inclusion of each polygenic risk scores (PGS) and walking pace (WP) to the NGRF model are shown in blue circles and black squares, respectively. The diamond shows the C-index upon the inclusion of both WP and the best performing PGS (PGS-13) with NGRF. Spikes indicate 95% confidence interval. (For interpretation of the references to color/colour in this figure legend, the reader is referred to the Web version of this article.)
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.numecd.2022.08.021.

References


