UC San Diego
UC San Diego Previously Published Works

Title

Genetic Stratification of Age-Dependent Parkinson's Disease Risk by Polygenic Hazard Score

Permalink

https://escholarship.org/uc/item/6j76s343

Journal

Movement Disorders, 37(1)

ISSN

0885-3185

Authors

Pihlstrøm, Lasse
Fan, Chun C
Frei, Oleksandr
et al.

Publication Date

2022

DOI

10.1002/mds.28808

Peer reviewed

RESEARCH ARTICLE

Genetic Stratification of Age-Dependent Parkinson's Disease Risk by Polygenic Hazard Score

Lasse Pihlstrøm, MD, PhD,[1]* Chun C. Fan, PhD,[2,3] Oleksandr Frei, PhD,[4,5] Manuela Tan, PhD,[1]
Roshan A. Karunamuni, PhD,[6] Cornelis Blauwendraat, PhD,[7] Sara Bandres-Ciga, PhD,[7] Ziv Gan-Or, MD, PhD,[8,9,10]
Donald G. Grosset, BSc, MBChB, MD,[11] International Parkinson's Disease Genomics Consortium (IPDGC),
Anders M. Dale, PhD,[2,3,4,12,13] Tyler M. Seibert, MD, PhD,[3,4,6,12,14] and Ole A. Andreassen, MD, PhD[4]

[1]*Department of Neurology, Oslo University Hospital, Oslo, Norway*
[2]*Department of Cognitive Science, University of California San Diego, La Jolla, California, USA*
[3]*Center for Multimodal Imaging and Genetics, School of Medicine, University of California San Diego, La Jolla, California, USA*
[4]*NORMENT Centre, Division of Mental Health and Addiction, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway*
[5]*Center for Bioinformatics, Department of Informatics, University of Oslo, Oslo, Norway*
[6]*Department of Radiation Medicine and Applied Sciences, University of California San Diego, La Jolla, California, USA*
[7]*Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA*
[8]*Montreal Neurological Institute, McGill University, Montreal, Québec, Canada*
[9]*Department of Human Genetics, McGill University, Montreal, Québec, Canada*
[10]*Department of Neurology and Neurosurgery, McGill University, Montreal, Québec, Canada*
[11]*Department of Neurology, Queen Elizabeth University Hospital, Glasgow, United Kingdom*
[12]*Department of Radiology, University of California San Diego, La Jolla, California, USA*
[13]*Department of Neurosciences, University of California San Diego, La Jolla, California, USA*
[14]*Department of Bioengineering, University of California San Diego, La Jolla, California, USA*

ABSTRACT: Background: Parkinson's disease (PD) is a highly age-related disorder, where common genetic risk variants affect both disease risk and age at onset. A statistical approach that integrates these effects across all common variants may be clinically useful for individual risk stratification. A polygenic hazard score methodology, leveraging a time-to-event framework, has recently been successfully applied in other age-related disorders.

Objectives: We aimed to develop and validate a polygenic hazard score model in sporadic PD.

Methods: Using a Cox regression framework, we modeled the polygenic hazard score in a training data set of 11,693 PD patients and 9841 controls. The score was then validated in an independent test data set of 5112 PD patients and 5372 controls and a small single-study sample of 360 patients and 160 controls.

Results: A polygenic hazard score predicts the onset of PD with a hazard ratio of 3.78 (95% confidence interval 3.49–4.10) when comparing the highest to the lowest risk decile. Combined with epidemiological data on incidence rate, we apply the score to estimate genetically stratified instantaneous PD risk across age groups.

Conclusions: We demonstrate the feasibility of a polygenic hazard approach in PD, integrating the genetic effects on disease risk and age at onset in a single model. In combination with other predictive biomarkers, the approach may hold promise for risk stratification in future clinical trials of disease-modifying therapies, which aim at postponing the onset of PD. © 2021 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society

Key Words: Parkinson's disease; age at onset; genetics; polygenic score; prediction

---

*Correspondence to: Dr. Lasse Pihlstrøm, Department of Neurology, Oslo University Hospital, PO Box 4950 Nydalen, 0424 Oslo, Norway; E-mail: lasse.pihlstrom@medisin.uio.no

A full list of IPDGC members is available as an online Supplementary file.

Relevant conflicts of interest/financial disclosures: None.

Age is one of the strongest-known risk factors for Parkinson's disease (PD). Consequently, a dramatic increase in PD prevalence is expected over the next decades, as a larger proportion of the population survives into old age.[1] A deeper understanding of the mechanisms relating aging to neurodegeneration and improved tools for individual risk stratification are immediately needed to meet this major public health challenge. Early detection and prediction through precision medicine[2] will be increasingly important for the development of novel disease-modifying therapies in PD, as neurodegeneration is shown to start long before the onset of clinical symptoms.

Common genetic variation also accounts for a substantial proportion of variability in sporadic PD risk, with an estimated heritability of about 16%–36%, based on meta-analyses of genome-wide association studies (GWAS).[3] A recent twin study reported 27% heritability when all age groups were included.[4] Variants detected in PD GWAS are also associated with age at onset of PD in linear regression models.[5-7] Only a few common variants have reached genome-wide significance as individual age-at-onset association signals, but models based on polygenic risk scores (PRSs) have consistently found that a higher cumulative burden of genetic PD risk correlates with earlier onset.[8-10] Although the findings clearly overlap, the association of genetic variants with disease risk and age at onset in PD has thus far been studied only as independent questions, using separate statistical frameworks.

Polygenic hazard scores (PHSs) leveraging a time-to-event, or survival, framework have been successful in Alzheimer's disease (AD) and prostate cancer, both highly age-related complex disorders.[11-13] This approach integrates the association with disease risk and age at onset into a common concept, under the hypothesis that genetic variation acts as a modulator of age-dependent risk. In the present article, we apply this PHS methodology in PD for the first time.

Although PD incidence is strongly dependent on age, some caution is warranted from the outset. A minor fraction of PD patients have a monogenic cause, recently estimated at approximately 1% in the United Kingdom,[14] yet higher in specific populations.[15] Autosomal recessive disease is strongly overrepresented among early-onset patients (eg, <40),[16] and even patients negative for mutations in known Mendelian genes may have a different genetic architecture from the common late-onset form.[9] In the oldest age groups, epidemiological studies have reported mixed results with respect to the trends in sporadic PD incidence.[17]

Using individual-level case–control genotype data from the International Parkinson's Disease Genomics Consortium (IPDGC) with information about age of onset, we model a PHS on reference data from sporadic PD patients and healthy controls in the age group from 40 to 75 years and validate it in two independent data sets. We further demonstrate how hazard ratios obtained using this method are directly interpretable in terms of stratified annualized incidence rates, with potential implications for clinical trial design. At present, a standard PRS approach based on summary statistics from case–control GWAS has the advantage of a far larger sample size to model from. The current results highlight the importance of collecting age-at-onset data for future improved PHS modeling in large-scale collaborative efforts.

Patients and Methods

All statistical analyses were performed using Matlab R2019a or R v3.6.1. Plink v1.9 was used to prepare the genetic data. An outline of the study workflow is provided in Figure 1. A set of recommended reporting standards for polygenic studies were published while this article was under review, and we have aimed to adhere to these guidelines.[18]

Genetic Data Sets

We used individual-level genotype data generated as part of previous genetic association studies by the IPDGC.[5] Standard quality checks have been performed on site-level data before initiation of the current study, including filtering for individual and variant missingness, excess heterozygosity, relatedness, Hardy–Weinberg equilibrium, and sex-check failures, as previously reported.[3] All included samples were of European ancestry. Genotypes have been imputed using reference data from the Haplotype Reference Consortium before data from individual sites were combined and duplicated/related samples from the level of first cousins were excluded from the common data set.

For the present study we included samples from sites with only available information on age at onset for patients or age at recruitment for controls. For some of the site-level data sets, age at subjective symptom onset was not available, and for these we used age at diagnosis as a surrogate for onset, similar to a recent age at onset GWAS based on the same raw data.[5] Notably, a high correlation between age at reported symptom onset and age at diagnosis (Pearson's $r^2 = 0.965$) was found in the Oslo data set where both variables where available. Patients with onset age and controls with recruitment age below 40 years were excluded. We selected the largest single data set available, including 5112 PD patients and 5372 controls, as a test data set for validation of the PHS model. Previous analyses have shown this sample size to afford high statistical power for testing of PHSs in a polygenic disorder.[19] The same IPDGC data set was used for replication in a 2014 PD GWAS and PRS validation in the 2019 PD meta-GWAS and is available in the database of Genotypes and Phenotypes (dbGaP phs000918.v1.p1).[20] Similarly,
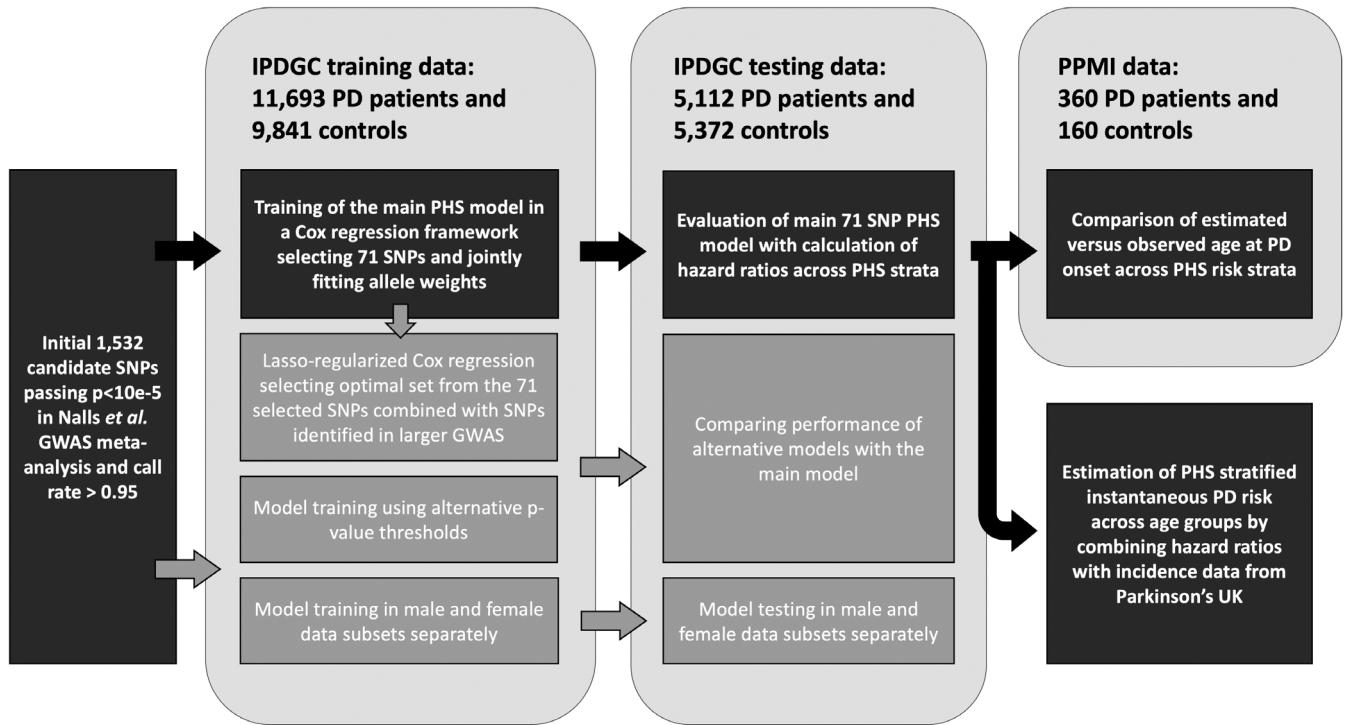
FIG. 1. Overview of the study workflow. The figure shows the different analysis steps of the study and what data sets were used. The main study workflow is shown at the top in bold font and darker boxes. Supplementary analyses are shown below in lighter boxes.

genotype data from the Parkinson's Progression Marker Initiative (PPMI) were reserved for independent validation.[21] The remaining data sets were included in the training data set used to generate the PHS model. Previous studies have shown that genetic factors associated with longevity can bias allele frequencies in data sets that include participants from the oldest age groups.[5] We therefore excluded patients and controls aged above 75 years from the training data set, which after exclusion comprised 11,693 PD patients and 9841 controls. Demographics are provided in Table S1, with site-specific details of the reference data in Table S2.

Training and Testing the Polygenic Hazard Model

Standard PRSs use a logistic regression framework to estimate weights for individual single-nucleotide polymorphisms (SNPs) and treat patients and controls as permanent designations. The PHS approach uses Cox proportional hazard models to directly estimate associations with age of onset of the disease, which may be particularly important for conditions, like PD, where incidence is strongly dependent on age. The time to event for patients is the age at diagnosis, and controls are censored at age at last follow-up, allowing for the possibility that they may develop PD at an older age.

Mathematically, the PHS is the vector product of the individual's genotype (X_i) for n SNPs and the corresponding parameter estimates (β_i) from the Cox proportional hazard regression:

$$PHS_x = \sum_{i}^{n} X_i \beta_i$$

The hazard rate at time t for a given subject X_i, $\lambda(X_i)$, is given by $\lambda(X_i) = \lambda_0(t)\exp(\beta_1 X_{i1} + \cdots + \beta_p X_{ip})$, where $\lambda_0(t)$ is the baseline hazard rate function, $\beta_1, \ldots, \beta_p$ are weights optimized from the training data sets, and $X_{i1}, \ldots, X_{ip}$ are covariates for the ith subject. In our case, the covariates include the genotype vectors of all SNPs included in the PHS model, as well as sex and top five principal components of the genotype matrix.

We first used full summary statistics (including data from 23andMe, Inc., Sunnyvale, CA, USA) from the largest meta-analysis of PD GWAS to date[3] to identify a list of SNPs associated with PD risk at significance threshold of $P < 1e-5$. From this list, we extracted only variants with call rate >0.95 across the individual genotype data sets, which included 1532 SNPs. Genotypes were coded 0 for reference allele homozygotes, 1 for heterozygotes, and 2 for alternate allele homozygotes, in line with a standard log-additive model. These SNPs were evaluated in a stepwise forward, greedy procedure, using a PHS to predict time to PD onset in the training data set, applying a Cox proportional hazard model, while controlling for sex and the top five genetic principal components. In each step, the algorithm

selected the SNP that best improved model prediction by minimizing the martingale residuals from Matlab's "coxphfit" function, until no SNP could be included in the PHS that would further significantly improve the model. The P-value threshold interpreted as a significant improvement of the model was arbitrarily set to $P < 10e\text{-}3$ based on our previous experience with the method. We also repeated the workflow testing both stricter ($P < 5e\text{-}4$) and more liberal ($P < 0.05$) thresholds and compared model performance based on z scores. The proportional hazard assumption was assessed using graphical comparisons.

Having defined the model SNPs and allele weights based on the training data, we then used the same model to calculate individual PHS in the independent test data set. We evaluated the performance of the PHS as a predictor of age-dependent PD incidence by model z score and plotted the Kaplan–Meier estimates for PD-free survival stratified by PHS percentile ranges.

Evaluating the Potential for Model Improvement by Adding SNPs Identified in Larger GWAS

An important disadvantage of the PHS approach is that the training step requires age data and individual genotypes, typically not available for data sets on the same scale as the largest GWAS meta-analyses. In a recent paper, we have shown that a PHS generated by our Cox regression method can be improved by the incorporation of GWAS-nominated SNPs in an additional step where the optimal SNP set is selected using least absolute shrinkage and selection operator (LASSO)-regularized regression.[22] Following this published approach, we took advantage of results from the latest meta-analysis of GWAS in PD, where 83 SNPs were identified as genome-wide significant loci independently of our test data set in a "leave one out" analysis.[3] A list of SNPs combining our PHS model SNPs with these 83 GWAS SNPs was compiled, and the R package "glmnet" was used to estimate a LASSO-regularized Cox proportional hazard model, where the hyperparameter λ was selected using 10-fold cross-validation (see Supplementary Data). A final LASSO model was estimated at the value of λ that minimized the mean cross-validated error,[22] and the performance of the LASSO model was compared to the basic PHS approach described earlier.

Evaluating the Performance of Sex-Specific Models

In previous studies of AD, training and testing the PHS model in sex-matched data sets have been shown to significantly improve performance,[23] indicating that genetic risk interacts with sex, at least for a relevant subset of common susceptibility loci. To evaluate this possibility in PD, we repeated the same workflow for model training and testing in the same data sets using male-only and female-only subsets of the data, respectively.

Predicting Population Risk of PD Onset

Combined with incidence rates from epidemiological studies, PHSs can be used to calculate genetically stratified estimates of absolute disease risk across the age spectrum.[11,24] We defined the age-dependent baseline risk based on epidemiological incidence rates by age group from a comprehensive 2017 report on "The Incidence and Prevalence of Parkinson's in the UK,"[25] representing to our knowledge the largest and most recent data source to provide the figures of interest. As incidence rates were reported for 5-year intervals, we let values represent the midpoint of each interval and used one-dimensional interpolation to estimate annualized incidence rates. Hazard ratios of PHS percentile strata were used to visualize the influence of polygenic risk on incidence curves and recalculate stratified "instantaneous" risk across age groups, applying sample weight correction to account for different case–control proportions in the sample sets as detailed in a previous report.[12]

Validating the Model for Onset Prediction in PPMI Data

We further tested the accuracy of PHS stratification in the relatively small PPMI data set of 360 PD patients and 160 controls, emulating a clinical trial cohort of premanifest individuals with a high risk of developing PD at some unknown age. PPMI participants were stratified into 10 PHS decile strata, and the age-dependent absolute risk for each stratum was calculated across all 1-year intervals from 50 to 75. For each PHS decile we extracted the age at which the proportion of individuals having developed PD reached 25%. We defined an absolute risk reference threshold (0.005) based on the risk corresponding to the 25% prevalence age in the fifth and sixth deciles, and the expected age for all other decile strata was defined as the year when the same threshold was reached. This age was then compared to the actual observed age when prevalence reached 25% in each decile stratum.[11]

Results

The PHSs and Model Performance

The Cox stepwise regression framework identified 71 SNPs that met criteria for inclusion in the final PHS model (Table S3). A graphical comparison between stratified Kaplan–Meier estimations and Cox proportional hazard models indicated that the proportional hazard assumption holds for the final model (Fig. S1).
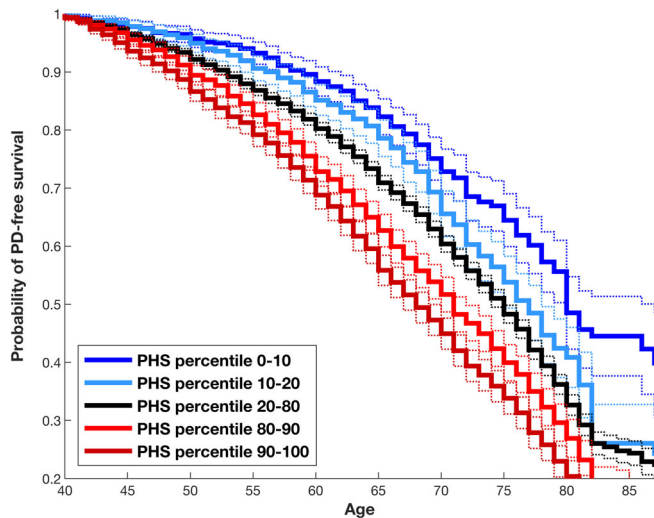
FIG. 2. Kaplan–Meier curves for test data, stratified by PHS (polygenic hazard score) modeled in training data. A polygenic hazard score modeled on the training data was calculated for all individuals in the independent test data set. The figure shows the Kaplan–Meier curves for survival free of Parkinson's disease in the test data set across selected PHS (polygenic hazard score) percentile strata. Thin lines represent 95% confidence intervals calculated by Greenwood's formula.[26] [Color figure can be viewed at wileyonlinelibrary.com]

Notably, several SNPs were included from loci that have also been reported to be individually associated with age at onset, including *SNCA* (three SNPs) and *TMEM175* (four SNPs). PHS was normally distributed (Fig. S2) and showed a strong association with age-related PD incidence in the test data (z score 17.7, $\beta = 0.90$, standard error $= 0.05$, $P < 10e{-}15$). Stratified Kaplan–Meier estimates are shown in Figure 2, demonstrating the effect of PHSs on age at onset in the test data. The hazard ratio comparing the highest to the lowest risk deciles was 3.78 (95% CI [confidence interval] 3.49–4.10) after sample weight correction.

Applying either stricter or more lenient *P*-value thresholds for SNP selection changed the number of included SNPs only slightly and did not improve model performance (threshold at $P < 5e{-}4$: 70 model SNPs, z score 17.7; threshold at $P < 0.05$: 82 model SNPs, z score 17.2). Combining the 71 SNPs from our primary PHS model with GWAS hits identified independently of the test data set in the latest meta-analysis of GWAS, we generated 130 SNPs serving as input to our LASSO-regularized Cox regression approach (Table S3). The final LASSO-selected model included 85 of these SNPs and performed slightly better in the test data (z score 19.0). This indicates that a pragmatic incorporation of summary statistics from large case–control GWAS could further improve a PHS model of PD based on Cox regression in individual data. However, as our primary aim in this work was to demonstrate the feasibility of a PHS approach in PD, we went forward with the basic 71 SNP PHS in the rest of our workflow.

Splitting the data into male and female subsets resulted in a smaller sample size for the training (men: 7258 PD patients and 4610 controls; women 4435 PD patients and 5231 controls) and test (men: 3297 PD patients and 3061 controls; women: 1815 patients and 2311 controls) data sets. Using the same threshold ($P < 0.01$) for SNP inclusion in the PHS model, 28 SNPs were selected for the male model and 32 SNPs for the female model. Neither model performed as well as the model combining data from both sexes (male z score $= 11.4$; female z score $= 9.4$). Relaxing the *P*-value threshold to 0.05 and allowing more SNPs to be included did not improve the sex-specific models. Consequently, the potential benefits of a sex-specific PHS model did not outweigh the sacrifice in statistical power in our PD data.

Applying PHSs to Population Risk

Examples of estimated yearly incidence rates for different age groups recalculated for specific PHS strata are provided in Table 1. Incidence rates for selected

TABLE 1 *Baseline incidence rates adjusted for polygenic hazard ratio*

Age	Baseline incidence	PHS 1st decile (95% CI)	PHS 9th decile (95% CI)	PHS 10th decile (95% CI)
45–49	4	1.9 (1.8–2.0)	5.4 (5.3–5.5)	7.3 (7.0–7.5)
50–54	9.1	4.4 (4.2–4.6)	12.3 (12.0–12.5)	16.5 (15.9–17.1)
55–59	18.2	8.7 (8.3–9.1)	24.5 (24.1–25.0)	33.0 (31.8–34.2)
60–64	33.5	16.1 (15.4–16.8)	45.1 (44.3–45.9)	60.7 (58.6–63.0)
65–69	62.5	29.9 (28.5–31.2)	83.9 (82.4–85.4)	112.9 (108.9–117.1)
70–74	113.4	54.3 (51.9–56.8)	152.7 (149.9–155.5)	205.6 (198.2–213.2)
75–79	173.5	83.1 (79.5–867.0)	233.6 (229.4–237.9)	314.5 (303.3–326.1)

The table shows examples of how stratified absolute risk of Parkinson's disease can be recalculated from baseline incidence using hazard ratio. Figures correspond to cases per 100,000 individuals in the population. Baseline incidences are from the Parkinson's UK 2017 report on "The Incidence and Prevalence of Parkinson's in the UK."[25]
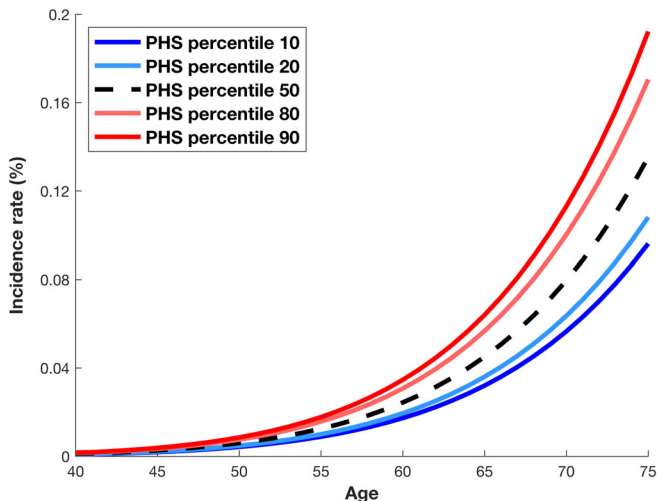Abbreviations: PHS, polygenic hazard score; CI, confidence interval.

FIG. 3. Incidence rates for Parkinson's disease stratified by polygenic hazard score. The figure shows estimated absolute incidence rates for Parkinson's disease as a function of age and polygenic hazard score percentile. The baseline incidence rates are based on a 2017 report from Parkinson's UK.[25] [Color figure can be viewed at wileyonlinelibrary.com]
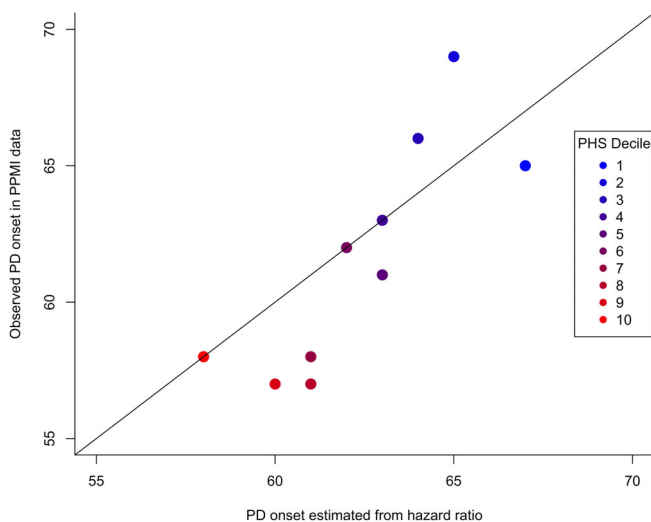

FIG. 4. Observed versus predicted age at onset in PPMI (Parkinson's Progression Marker Initiative) hazard strata. A polygenic hazard score modeled on training data was estimated for both PD patients and controls in the PPMI data set, and participants were stratified into PHS (polygenic hazard score) deciles. The figure shows the observed age at which 25% of participants in each PHS stratum had developed Parkinson's disease plotted against the age estimated from the PHS stratum hazard ratio (see main text). [Color figure can be viewed at wileyonlinelibrary.com]

PHS strata, estimated from hazard ratios and epidemiological data on baseline risk,[25] are shown in Figure 3.

Onset Prediction in PPMI Data

Predicted and observed age at onset stratified for PHS in the PPMI data showed a clear correlation (Pearson's $r = 0.83$, $P = 0.0030$). The results across all PHS decile strata are shown in Figure 4.

Discussion

We show for the first time that the PHS method can be used to estimate PHS-adjusted PD risk. The effect sizes were large enough to achieve significant validation with relatively small sample sizes, such as the PPMI study. Comparing the top and bottom PHS deciles, we observed a hazard ratio of 3.78 (95% CI 3.49–4.10).

We focused our analysis on the age range from 40 to 75, demonstrating that a 71 SNP PHS model trained on the reference data set successfully predicts empirical age at PD onset in the independent test data. The observed hazard ratio for the top and bottom PHS deciles for PD is consistent with previous reports of an association between cumulative burden of genetic disease risk and age at onset in PD.[5,6,8-10] We found no evidence of a sex-specific effect in PD. Smaller sample sizes reduce statistical power, yet the lack of sex-specific effects in our study is also consistent with the mostly negative sex-stratified results reported in a recent GWAS of age at onset in PD.[5]

Methods that summarize the effect of many risk alleles into a polygenic score have gained increasing attention in complex genetics as ever-larger GWAS data sets become available to train more powerful models.[27] The polygenic hazard approach takes advantage of a survival analysis framework to integrate the genetic effects on disease risk and age at disease onset in a single model for age-dependent, complex disorders. A major advantage of the PHS method is that the hazard ratio is directly and intuitively interpretable as a modifier of baseline risk. We have shown how PHS can be combined with epidemiological data on incidence rates per age group to calculate genetically stratified estimates of instantaneous PD risk in a given population.

An important hurdle for the scalability of the PHS approach is that it requires both age data and individual genotype data. In contrast, large-scale meta-analysis of GWAS can take advantage of summary statistics from logistic regression. The standard PRS generated from GWAS results is currently widely used in complex disease research but does not incorporate the age dimension the way our PHS does. We emphasize that we have primarily aimed to demonstrate the feasibility of a PHS approach, not expecting to directly outperform the best current PRSs, where the number of participants contributing to SNP identification and estimation of allele weights is very large. Our LASSO-regularized Cox regression approach indicates that incorporating GWAS results will currently benefit a PHS model. As genetic data sets continue to grow, our results should encourage PD researchers to collect and share participant age data to allow for further improvement in polygenic modeling. We acknowledge that larger training data sets will be required to further improve the accuracy of the PHS model and its potential utility in PD.

It is worth noting that although our study concerns PD onset, the polygenic hazard framework is also relevant for the prediction of progression to specific clinical disease course milestones. While this manuscript was under review, a genome-wide survival study of progression to dementia in PD was published, employing a PHS approach inspired by our previous work in AD.[11,28] Interestingly, a PHS modeled on cognitive progression data was reported to significantly predict PD dementia, whereas a general PRS for PD risk based on GWAS summary statistics did not.

Our study has some limitations. Most important, the model relies on assumptions that may not hold for PD at the extreme ends of the age-at-onset range. Early-onset PD has a higher likelihood of a monogenic cause and may have a different genetic architecture from PD, with later onset also in sporadic cases. The trend for PD incidence late in life remains somewhat controversial, with conflicting results across studies. The 2017 epidemiological report from Parkinson's UK showed a peak in annual incidence rate in the 80 to 84 age group, followed by a decline in the 85 to 89 group and even lower in individuals aged 90 to 94 years.[25] A similar pattern has been observed in other studies.[29,30] In contrast, a study from the Rochester Epidemiology Project, Minnesota, found increasing incidence rates all through the ninth decade of life,[31] in line with a previous meta-analysis.[17] Ascertainment challenges in the oldest age groups could plausibly contribute to these discrepancies, yet the question of a possible PD incidence peak or plateau remains currently unresolved.

We also note that determining the time of PD onset is not trivial. Subjective symptoms present insidiously, often years before the disease is diagnosed by a neurologist, and the accuracy of age at onset in large, heterogeneous data sets is likely to be low. More standardized and homogeneous criteria for determining age at onset would be expected to improve both the estimates of SNP effect sizes and the performance of the PHS. With respect to the genetic data, PHS calculation requires ethnically matched data sets. Consequently, the application of a PHS derived with European ancestry data may have variable performance in individuals of other genetic ancestries, and ancestry-specific SNPs may improve performance.[32,33] We note also that our SNP selection was based on a stepwise forward approach only, and future work should explore stepwise backward selection and other alternative strategies. Furthermore, the approach assumes an additive model and will not capture gene–gene or gene–environment interactions that may contribute to genetic risk.

Our study aims at highlighting the potential future utility of PHSs as biomarkers in PD. That our analyses were based on retrospective data only should be noted as a major limitation, which may not necessarily translate to the prospective context of a clinical trial. We anticipate clinical trials of disease-modifying therapies that aim at postponing the motor onset of PD, where our hope would be that an integrated estimate of disease risk as a function of both age and genetic profile could be a valuable asset for patient selection and stratification. Significant imbalance in genetic risk variants across randomized trial arms has been demonstrated in simulated PD clinical trial cohorts, highlighting the potential risk of genetic heterogeneity confounding true effects.[34] Envisaging pre-motor trials, an estimate of participants' risk of developing PD within the follow-up period is far more relevant than lifetime risk, in line with the framework presented here. We acknowledge, however, that information such as family history and clinical assessments for anosmia or other prodromal symptoms is far more accessible, and how much genetic profiling will independently add to a comprehensive screening battery for pre-motor PD is currently an open question.[35] In our view, our validation in the PPMI data set holds promise that even with relatively small sample sizes, this type of genetic stratification can potentially be clinically meaningful. By ensuring that large ongoing international efforts to generate data for precision medicine[2] also include PD, we expect the polygenic hazard approach will be further improved to obtain clinically relevant performance. ■

investigators have not participated in reviewing the data analysis or content of the manuscript.

Data Availability Statement

IPDGC GWAS summary statistics are available on the IPDGC website (http://pdgenetics.org/resources). Individual genotypes for the IPDGC dataset used as test data in this study are publicly available through dbGAP, accession phs000918.v1.p1.

References

1. G. B. D. Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the global burden of disease study 2016. Lancet Neurol 2018;17:939–953.

2. Denny JC, Collins FS. Precision medicine in 2030-seven ways to transform healthcare. Cell 2021;184:1415–1419.

3. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol 2019;18:1091–1102.

4. Goldman SM, Marek K, Ottman R, et al. Concordance for Parkinson's disease in twins: a 20-year update. Ann Neurol 2019;85:600–605.

5. Blauwendraat C, Heilbron K, Vallagra C, et al. Parkinson's disease age at onset genome-wide association study: defining heritability, genetic loci, and α-synuclein mechanisms. Mov Disord 2019;34(6):866–875.

6. Lill CM, Hansen J, Olsen JH, et al. Impact of Parkinson's disease risk loci on age at onset. Mov Disord 2015;30:847–850.

7. Iwaki H, Blauwendraat C, Leonard HL, et al. Genetic risk of Parkinson disease and progression: an analysis of 13 longitudinal cohorts. Neurol Genet 2019;5:e348.

8. Iwaki H, Blauwendraat C, Leonard HL, et al. Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. Mov Disord 2019;34:1839–1850.

9. Nalls MA, Escott-Price V, Williams NM, et al. Genetic risk and age in Parkinson's disease: continuum not stratum. Mov Disord 2015;30:850–854.

10. Pihlstrom L, Toft M. Cumulative genetic risk and age at onset in Parkinson's disease. Mov Disord 2015;30:1712–1713.

11. Desikan RS, Fan CC, Wang Y, et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. PLoS Med 2017;14:e1002258.

12. Seibert TM, Fan CC, Wang Y, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. BMJ 2018;360:j5757.

13. Huynh-Le MP, Myklebust TA, Feng CH, et al. Age dependence of modern clinical risk groups for localized prostate cancer-a population-based study. Cancer 2020;126:1691–1699.

14. Tan MMX, Malek N, Lawton MA, et al. Genetic analysis of Mendelian mutations in a large UK population-based Parkinson's disease study. Brain 2019;142:2828–2844.

15. Correia Guedes L, Ferreira JJ, Rosa MM, et al. Worldwide frequency of G2019S LRRK2 mutation in Parkinson's disease: a systematic review. Parkinsonism Relat Disord 2010;16:237–242.

16. Kilarski LL, Pearson JP, Newsway V, et al. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov Disord 2012;27:1522–1529.

17. Hirsch L, Jette N, Frolkis A, et al. The incidence of Parkinson's disease: a systematic review and meta-analysis. Neuroepidemiology 2016;46:292–300.

18. Wand H, Lambert SA, Tamburro C, et al. Improving reporting standards for polygenic scores in risk prediction studies. Nature 2021;591:211–219.

19. Karunamuni RA, Huynh-Le MP, Fan CC, et al. The effect of sample size on polygenic hazard models for prostate cancer. Eur J Hum Genet 2020;28:1467–1475.

20. Nalls MA, Pankratz N, Lill CM, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet 2014;46:989–993.

21. Nalls MA, Keller MF, Hernandez DG, et al. Baseline genetic associations in the Parkinson's progression markers initiative (PPMI). Mov Disord 2016;31:79–85.

22. Karunamuni RA, Huynh-Le MP, Fan CC, et al. Additional SNPs improve risk stratification of a polygenic hazard score for prostate cancer. Prostate Cancer Prostatic Dis 2021;24:532–541.

23. Fan CC, Banks SJ, Thompson WK, et al. Sex-dependent autosomal effects on clinical progression of Alzheimer's disease. Brain 2020;143:2272–2280.

24. Huynh-Le MP, Fan CC, Karunamuni R, et al. A genetic risk score to personalize prostate cancer screening, applied to population data. Cancer Epidemiol Biomarkers Prev 2020;29:1731–1738.

25. Parkinson's UK. The incidence and prevalence of Parkinson's in the UK: Results from the Clinical Practice Research Datalink Reference Report. 2017. https://www.parkinsons.org.uk/professionals/resources/incidence-and-prevalence-parkinsons-uk-report

26. Lawless JF. Statistical Models and Methods for Lifetime Data. 2nd ed. Hoboken NJ: John Wiley & Sons Inc.; 2003.

27. Sugrue LP, Desikan RS. What are polygenic scores and why are they important? JAMA 2019;321:1820–1821.

28. Liu G, Peng J, Liao Z, et al. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. Nat Genet 2021;53:787–793.

29. Duncan GW, Khoo TK, Coleman SY, et al. The incidence of Parkinson's disease in the north-east of England. Age Ageing 2014;43:257–263.

30. Linder J, Stenlund H, Forsgren L. Incidence of Parkinson's disease and parkinsonism in northern Sweden: a population-based study. Mov Disord 2010;25:341–348.

31. Savica R, Grossardt BR, Bower JH, et al. Time trends in the incidence of Parkinson disease. JAMA Neurol 2016;73:981–989.

32. Karunamuni RA, Huynh-Le MP, Fan CC, et al. African-specific improvement of a polygenic hazard score for age at diagnosis of prostate cancer. Int J Cancer 2020;148:99–105.

33. Huynh-Le MP, Fan CC, Karunamuni R, et al. Polygenic hazard score is associated with prostate cancer in multi-ethnic populations. Nat Commun 2021;12:1236.

34. Leonard H, Blauwendraat C, Krohn L, et al. Genetic variability and potential effects on clinical trial outcomes: perspectives in Parkinson's disease. J Med Genet 2020;57:331–338.

35. Nalls MA, McLean CY, Rick J, et al. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. Lancet Neurol 2015;14:1002–1009.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.

Author Roles

O.A.A. conceived the study. L.P., C.B., S.B.-C., Z.G.-O., D.G.G., and the IPDGC provided data. C.B. facilitated data access. L.P. performed statistical analyses with input and assistance from C.C.F., O.F., S.B.-C., R.A.K. and T.M.S. using methodology developed under the leadership of A.M.D. and O.A.A. L.P. drafted the manuscript with input from C.F.F., O.F., T.M.S., and O.A.A. All authors contributed to critical revision of the manuscript.

Full financial disclosures for the previous 12 months

O.A.A. received speaker honorarium from Lundbeck and is consultant for Healthlytix. A.M.D. reports that he was a founder of and holds equity in CorTechs Labs Inc. and serves on its scientific advisory board. He is a member of the scientific advisory board of Human Longevity, Inc., and the Mohn Medical Imaging and Visualization Centre. He received funding through research grants from GE Healthcare to University of California San Diego. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict-of-interest policies. T.M.S. reports honoraria, outside of the present work, from the University of Rochester, Varian Medical Systems, Multimodal Imaging Services Corporation, and WebMD. Z.G.-O. reports personal fees from Idorsia, Neuron23, Handl Therapeutics, Lysosomal Therapeutics Inc., Deerfield, Lighthouse, Prevail Therapeutics, Ono Therapeutics, Denali, and Inception Sciences outside the submitted work. D.G.G. has received honoraria for advisory board meetings from AbbVie and Bial Pharma, speaker fees from Britannia Pharmaceuticals and Bial Pharma, and consultancy fees from the Glasgow Memory Clinic. Other authors have nothing to report. L.P. has served as a consultant for Roche, for which a honorarium was paid to his employer, Oslo University Hospital.