Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large-scale cohorts

Tyler M. Seibert, resident physician[1,2]
Chun Chieh Fan, PhD candidate[1,3]
Yunpeng Wang, postdoctoral fellow[4]
Verena Zuber, postdoctoral fellow[4,5]
Roshan Karunamuni, postdoctoral fellow[1,2]
J. Kellogg Parsons, professor[6]
Rosalind A. Eeles, professor[7,8]
Douglas F. Easton, professor[9]
ZSofia Kote-Jarai, researcher[7]
Ali Amin Al Olama, research associate[9,10]
Sara Benlloch Garcia, researcher[9]
Kenneth Muir, professor[11,12]
Henrik Grönberg, professor[13]
Fredrik Wiklund, associate professor[13]
Markus Aly, researcher[13,14]
Johanna Schleutker, professor[15,16]
Csilla Sipeky, adjunct professor[17]
Teuvo LJ Tammela, professor[18]
Børge G. Nordestgaard, professor[19,20]
Sune F. Nielsen, researcher[19,20]
Maren Weischer, resident physician[20]
Rasmus Bisbjerg, consultant[21]
M. Andreas Røder, researcher[22]
Peter Iversen, professor[19,22]
Tim J. Key, professor[23]
Ruth C. Travis, associate professor[24]
David E. Neal, professor[25,26]
Jenny L. Donovan, professor[27]
Freddie C. Hamdy, professor[28]
Paul Pharoah, professor[29]
Nora Pashayan, clinical reader in applied health research[30,29]
Kay-Tee Khaw, professor[31]
Christiane Maier, researcher[32]
Walther Vogel, professor[32]
Manuel Luedeke, researcher[32]
Kathleen Herkommer, researcher[33]
Adam S. Kibel, professor[34]
Cezary Cybulski, professor[35]
Dominika Wokolorczyk, assistant professor[35]
Wojciech Kluzniak, assistant professor[35]
Lisa Cannon-Albright, professor[36,37]
Hermann Brenner, professor[38,39,40]
Katarina Cuk, postdoctoral fellow[38]

Kai-Uwe Saum, researcher[38]
Jong Y. Park, associate professor[41]
Thomas A. Sellers, cancer center director[42]
Chavdar Slavov, researcher[43]
Radka Kaneva, researcher[44]
Vanio Mitev, researcher[44]
Jyotsna Batra, researcher[45]
Judith A. Clements, professor[45]
Amanda Spurdle, professor[46]
Australian Prostate Cancer BioResource , researcher[45,47]
Manuel R. Teixeira, professor[48,49]
Paula Paulo, researcher[48]
Sofia Maia, researcher[48]
Hardev Pandha, professor[50]
Agnieszka Michael, researcher[50]
Andrzej Kierzek, professor[50]
David S. Karow, associate professor[1,51]
Ian G. Mills, associate professor[4,52]
Ole A. Andreassen, professor[4]
Anders M. Dale, professor[1,51,53]
The PRACTICAL consortium*

Affiliations

[1]Center for Multimodal Imaging & Genetics, University of California, San Diego, La Jolla, CA, USA
[2]Department of Radiation Medicine & Applied Sciences, University of California, San Diego, La Jolla, CA, USA
[3]Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA
[4]University of Oslo, Oslo, Norway
[5]MRC Biostatistics Unit, Cambridge Biomedical Campus, Cambridge CB2 0SR, UK
[6]Department of Surgery, University of California, San Diego, La Jolla, CA, USA
[7]The Institute of Cancer Research, London, SM2 5NG, UK
[8]Royal Marsden NHS Foundation Trust, London, SW3 6JJ, UK
[9]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK
[10]Department of Clinical Neurosciences, Stroke Research Group, University of Cambridge, R3, Box 83, Cambridge Biomedical Campus, Cambridge CB2 0QQ, UK
[11]Institute of Population Health, University of Manchester, Manchester, UK
[12]Warwick Medical School, University of Warwick, Coventry, UK
[13]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[14]Department of Molecular Medicine and Surgery, Karolinska Institutet, and Department of Urology, Karolinska University Hospital, Solna, 171 76 Stockholm
[15]Department of Medical Biochemistry and Genetics, Institute of Biomedicine, Kiinamyllynkatu 10, FI-20014 University of Turku; and Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital
[16]BioMediTech, 30014 University of Tampere, Tampere, Finland
[17]Department of Medical Biochemistry and Genetics, Institute of Biomedicine, Kiinamyllynkatu 10, FI-20014 University of Turku, Finland
[18]Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Finland
[19]Faculty of Health and Medical Sciences, University of Copenhagen, Denmark
[20]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[21]Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark
[22]Copenhagen Prostate Cancer Center, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[23]Cancer Epidemiology Unit, Nuffield Department of Population Health University of Oxford, Oxford, OX3 7LF, UK
[24]Cancer Epidemiology, Nuffield Department of Population Health University of Oxford, Oxford, OX3 7LF, UK
[25]Nuffield Department of Surgery, University of Oxford, Room 6603, Level 6, John Radcliffe Hospital, Headley Way, Headington, Oxford, OX3 9DU, UK
[26]University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Hills Road Cambridge CB2 0QQ, UK
[27]School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK
[28]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK
[29]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, UK
[30]University College London, Department of Applied Health Research, 1-19 Torrington Place, London, WC1E 7HB
[31]Clinical Gerontology Unit, University of Cambridge, Cambridge UK
[32]Institute of Human Genetics, University Hospital of Ulm, Ulm, Germany
[33]Department of Urology, Klinikum rechts der Isar der Technischen Universitaet Muenchen, Munich, Germany
[34]Division of Urologic Surgery, Brigham and Womens Hospital, Dana-Farber Cancer Institute, 75 Francis Street, Boston, MA 02115, USA
[35]International Hereditary Cancer Center, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland
[36]Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA
[37]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah, USA

[38]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany
[39]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany
[40]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany
[41]Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA
[42]Office of the Center Director, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA
[43]Department of Urology and Alexandrovska University Hospital, Medical University, Sofia, Bulgaria
[44]Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University, Sofia, 2 Zdrave Str., 1431 Sofia, Bulgaria
[45]Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia
[46]Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia
[47]Australian Prostate Cancer BioResource, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia
[48]Department of Genetics, Portuguese Oncology Institute, Porto, Portugal
[49]Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal
[50]The University of Surrey, Guildford, Surrey, GU2 7XH
[51]Department of Radiology, University of California, San Diego, La Jolla, CA, USA
[52]Centre for Cancer Research and Cell Biology, Queens University Belfast, Belfast, UK
[53]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

*Additional members from the *Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome consortium* (PRACTICAL, http://practical.ccge.medschl.cam.ac.uk/) provided in the Supplementary Material.

Corresponding authors

Tyler M. Seibert, MD, PhD
Center for Multimodal Imaging & Genetics
Altman CTRI Building 4W 102
9500 Gilman Drive, Mail Code 0841

La Jolla, CA 92093-0841
United States
tseibert@ucsd.edu
(858) 255-1458

Anders M. Dale, PhD
Center for Multimodal Imaging & Genetics
Altman CTRI Building 4W 228
9500 Gilman Drive, Mail Code 0841
La Jolla, CA 92093-0841
United States
amdale@ucsd.edu
(858) 822-6671

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no support from any organisation for the submitted work except as follows:

DSK and AMD report a research grant from the US Department of Defense. OAA reports research grants from KG Jebsen Stiftelsen, Research Council of Norway, and South East Norway Health Authority.

Authors declare no financial relationships with any organisations that might have an interest in the submitted work in the previous three years except as follows, with all of these relationships outside the present study:

TMS reports honoraria from WebMD, Inc. for educational content, as well as a research grant from Varian Medical Systems. ASK reports advisory board memberships for Sanofi-Aventis, Dendreon, and Profound. AK reports paid work for Certara Quantitative Systems Pharmacology. DSK reports paid work for Human Longevity, Inc.

Authors declare no other relationships or activities that could appear to have influenced the submitted work except as follows:

OAA has a patent application # U.S. 20150356243 pending; AMD also applied for this patent application and assigned it to UC San Diego. AMD has additional disclosures outside the present work: founder, equity holder, and advisory board member for CorTechs Labs, Inc.; advisory board member of Human Longevity, Inc.; recipient of non-financial research support from General Electric Healthcare.

Additional acknowledgments for the PRACTICAL consortium and contributing studies are described in the Supplementary Material.

Ethics approval
No specific approval was required for this retrospective study of de-identified data.

Funding for this study
This study was funded in part by a grant from the United States Department of Defense (#W81XWH-13-1-0391), the Research Council of Norway (#223273), KG Jebsen Stiftelsen, and South East Norway Health Authority. Funding for the PRACTICAL consortium member studies is detailed in the Supplementary Material.

Author independence
No funders or outside organisations had any role in the study design; data collection, analysis, or interpretation; writing of the report; or the decision to submit the present article for publication. All researchers have acted independently from funders, and this work is solely that of the authors. All authors had full access to the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Transparency declaration
The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Patient involvement
No patients were directly involved in designing the research question or in conducting the research. A link to the published results will be posted on the publically available PRACTICAL consortium website, and the respective principal investigators of each contributing study will be provided the results to disseminate to individual participants where possible.

Data sharing
No additional data available.

Abstract

Objectives: Prostate-specific-antigen (PSA) screening resulted in reduced prostate cancer (PCa) mortality in a large clinical trial, but due to many false positives and overdiagnosis of indolent disease, many guidelines do not endorse universal screening and instead recommend an individualized decision based on each patient's risk. We sought to develop and validate a genetic tool to predict age of aggressive PCa onset and to guide decisions of whom to screen and at what age.

Design: Genotype, PCa status, and age were analyzed to select single-nucleotide polymorphisms (SNPs) associated with PCa diagnosis. These SNPs were incorporated into a survival analysis to estimate their effects on age at diagnosis of aggressive PCa (i.e., not eligible for surveillance per NCCN Guidelines; any of: Gleason score ≥7, stage T3-T4, PSA ≥10, nodal metastasis, distant metastasis). The resulting polygenic hazard score (PHS) is an assessment of individual genetic risk. The final model was applied to an independent dataset containing genotype and screening PSA data. PHS was calculated for these men to test prediction of PCa-free survival.

Setting: Multiple, international PRACTICAL consortium member institutions.

Participants: All PRACTICAL consortium participants of European ancestry with known age, PCa status, and quality-assured iCOGS array genotype data. Development dataset comprised 31,747 men. Validation dataset comprised 6,411 men.

Main outcome measures: PHS prediction of age of onset of aggressive PCa in validation set.

Results: In the independent validation set, PHS calculated from 54 SNPs was a highly significant predictor of age at diagnosis of aggressive PCa (z=11.2, $p<10^{-16}$). When men in the validation set with high PHS (>98th percentile) were compared to those with average PHS (30th-70th percentile), the hazard ratio for aggressive PCa was 2.9.

Conclusions: Polygenic hazard scores give personalized genetic risk estimates that predict for age of onset of aggressive PCa.

Key words
Prostate cancer, genetic, risk, prediction, screening, age, PSA

What this paper adds (box)

What is already known on this subject

- Prostate cancer (PCa) screening with prostate-specific-antigen (PSA) testing can lead to early detection of PCa and allow for curative treatment, but universal screening also has considerable disadvantages for men who may never develop aggressive disease.
- Ideally, physicians would identify and screen patients at high risk of developing aggressive PCa or PCa at a young age.
- A practical, clinically useful tool to predict age of PCa onset is not yet available.

What this study adds

- This study presents and validates a novel polygenic hazard score (PHS) that is an indicator of age of onset of aggressive PCa.
- PHS is a relatively inexpensive assessment of an individual man's *age-specific* PCa risk and provides objective information on whether a given patient might benefit from PSA screening.

Introduction

Prostate cancer (PCa) is a major health problem, with over one million new cases and over 300,000 PCa deaths estimated worldwide in 2012[1]. An international, randomized, controlled trial showed that prostate-specific-antigen (PSA) screening resulted in a 27% reduction in PCa mortality[2]. However, due to concerns over a high rate of false positives, in addition to aggressive treatment of apparently indolent disease, many clinical guidelines do not endorse universal screening and instead stress the importance of taking into account individual patient risk factors to decide whether to screen[3–5]. The goal is to avoid unnecessary screening while still identifying high-risk men for whom screening and early PCa detection can reduce morbidity and mortality.

A patient's genetic predisposition could be critical to the decision of whether and when to offer him PCa screening. Genome-wide association studies (GWAS) have revealed genetic variants associated with increased risk of PCa[6,7]. These developments, combined with the recent accessibility of genotyping, provide an opportunity for genetic risk-informed cancer screening[8]. By combining risk information from an array of single-nucleotide polymorphisms (SNPs), polygenic models can estimate an individual's genetic risk for developing the disease[9]. Predicted polygenic risk could improve clinical decisions such as whom to screen for PCa and at what age[10,11].

Here we use data from 31,747 men of European ancestry from the international PRACTICAL consortium (http://practical.ccge.medschl.cam.ac.uk/) to develop a polygenic hazard score (PHS) for predicting *age-related* risk of developing aggressive PCa. This is designed for use before the decision of whether to screen (e.g., with PSA) by providing a risk stratification strategy to maximize screening efficiency. The PHS was therefore tested in data from an independent, screening study (UK ProtecT[12]), with the hypothesis that PHS would be an indicator of a patient's inherent genetic risk for developing PCa at various ages in his lifetime, and thus could guide PSA screening.

Methods

Definition of aggressive disease

Concerns about overdiagnosis and overtreatment of indolent disease have influenced discussion of PCa screening, whereas there is consensus that aggressive PCa warrants treatment[13,14]. Where possible, we therefore focus validation in this study on prediction of aggressive disease, defined as any tumor that would require radical treatment for a typical, healthy man according to National Comprehensive Cancer Network (NCCN) Guidelines (i.e., not eligible for active surveillance)[14]. This includes cancers with any of: Gleason score ≥7, stage T3-T4, PSA ≥10, nodal metastasis, or distant metastasis. Note that stage T2 tumors were classified without a sub-category in our database, so a patient with low Gleason score and low PSA, but stage T2b or T2c would be considered low risk in this analysis even though NCCN Guidelines would indicate treatment for intermediate risk; this was to ensure that no low-risk tumors were included as cases of aggressive PCa.

Some additional analyses used age of diagnosis of any PCa, either as complementary information or because available data did not permit exclusive focus on aggressive

PCa. This is noted, where applicable. Another secondary analysis tested prediction for 'very aggressive disease,' defined as any of: Gleason score ≥8, stage T3-4, positive nodes, or distant metastases.

Participants

Development Set: For PHS model development, genotype and age data were obtained from 21 studies of the PRACTICAL consortium (Supplementary Table S1), representing 31,747 men (18,868 any PCa, 10,635 aggressive PCa, 5,406 very aggressive PCa,12,879 controls) of genotypic European ancestry. Age was either at PCa diagnosis or last follow-up (for controls). Genotyping, performed via a custom Illumina array (iCOGS), and quality control steps have been described previously[6]. 201,043 SNPs were available for analysis. Categorization as aggressive or not was impossible for 4,803 of the PCa cases due to incomplete staging data; these were excluded from aggressive PCa analyses.

Validation Set: Model performance was examined in an independent study. The Validation Set comes from the ProtecT study, which screened 82,429 men with PSA testing and found 8,891 men with PSA greater than the specified threshold of 3.0 μg/L or higher, of whom 2,896 were diagnosed with PCa[12]. Among those individuals, we obtained genotype and age data for 6,411 men (1,583 any PCa, 632 aggressive PCa, 220 very aggressive PCa, 4,828 controls). Staging data were available for all cases. This dataset was selected for validation because PSA results were also available for all participants at time of either diagnosis or interview. Further details in Supplementary Methods.

Missing data:

During model development, SNPs with call rates less than 95% were excluded. Missing calls for the remaining SNPs were imputed with the mean genotype count for that allele across all participants.

Polygenic hazard score (PHS)

The polygenic hazard score (PHS) was developed previously as a parsimonious, survival-analysis model to predict the time to event outcome (in this case, age of PCa onset). It has been published elsewhere[15], and further details of application here are described in Supplementary Material.

The polygenic hazard score (PHS) is defined as the vector product of a patient's genotype (X_i) for n selected SNPs and the corresponding parameter estimates (β_i) from a Cox proportional hazards regression:

$$PHS_X = \sum_i^n X_i \beta_i$$

Genetic prediction of *only* aggressive PCa has proven elusive, with most SNPs associated with aggressive disease also showing association with any PCa[16]. Therefore, in the interest of maximizing power to select SNPs associated with age of onset of PCa, we decided to initially include all cases from the Development Set (i.e., any PCa) for generation of the PHS model. An alternate strategy that limited model generation to aggressive PCa cases was then tested for comparison. The primary metric for validation in both instances remained prediction for aggressive PCa in the independent Validation Set.

To verify whether the PHS accurately predicts age of aggressive PCa onset, the PHS was calculated for all patients in the Validation Set and tested as the sole predictive variable in a Cox proportional hazards regression model for age of diagnosis. Patients in the Validation Set diagnosed with low-risk PCa (Gleason score <=6, PSA<10, and stage T2N0M0 or lower) were censored at time of diagnosis, reflecting the fact that it is unknown if they would later be diagnosed with aggressive disease or at what age that might have occurred. Statistical significance was set at alpha of 0.01 for this and all subsequent Cox models. As an indicator of effect size for the model, we calculated a hazard ratio comparing men with high PHS (>98th percentile) to those with average risk (30th-70th percentile). All hazard ratios in this manuscript refer to the same pattern, comparing high to average risk.

In light of evidence that initially low-risk disease often progresses to require treatment[17–19], and because this may be particularly important for men diagnosed at a young age, we performed a secondary, analogous analysis to test for prediction of age of diagnosis of any PCa. Another secondary analysis was done for prediction of very aggressive disease.

To further assess the clinical significance of PHS, we looked at the positive predictive value (PPV) of PSA testing within the Validation Set, with clinical diagnosis (including biopsy result) as the gold standard. We posited that risk stratification by PHS percentiles would reflect the underlying incidence of PCa and therefore also affect the PPV of PSA testing. Details of PPV calculation are in the Supplementary Material. PHS categories were designated by PHS percentile compared to the young, healthy population within the Development Set: i.e., those controls with age <70 years. All percentiles reported in this manuscript refer to this population.

To visualize PHS distribution among aggressive PCa cases in the Validation Set, we generated a Lorenz curve[20–22].

Comparison to family history

One of the most important risk factors used currently for screening decisions is family history[3]. We compared family history and PHS for prediction of aggressive PCa onset using the same Cox model approach as before, using the 5,703 men (1,405 any PCa, 554 aggressive PCa, 4,298 controls) from the Validation Set with known family history status (0 or ≥1 affected first-degree relatives). Models were constructed with family history alone, PHS alone, or with both. These were compared via log-likelihood tests.

Results

PHS model development

Of the 201,043 SNPs included in the dataset, 2,415 were associated with increased risk of PCa in the trend test, with $p<10^{-6}$. The stepwise regression framework then identified 54 of these SNPs that were incorporated into the Cox proportional hazards model (Supplementary Table S2). The 54 SNP parameter estimates (for the hazard of developing PCa) are combined with individual genotype to generate the polygenic hazard score. Kaplan-Meier and Cox regression estimates for the final model are shown in Figure 1. The final model performed well for prediction of age of aggressive PCa onset in the Development Set (z=37.5, $p<10^{-16}$, HR=2.3 [95% CI: 2.2, 2.4]).

Only 43 SNPs (0.02%) were excluded for low call rate during model development, and imputation for missing calls was used for 0.4% of calls in the final model. Of the 6,411 participants in the Validation Set, the median individual SNP call rate was 100%, with a minimum of 98%.

PCa risk prediction with PHS

In the independent Validation Set from the ProtecT study, a Cox proportional hazards model showed that PHS was a significant predictor of age of onset of aggressive PCa (z=11.2, $p<10^{-16}$). The hazard ratio for high PHS men (>98th percentile) compared to average risk was 2.9 [95% CI: 2.4, 3.4]. PHS was also predictive of any PCa (z=15.4, $p<10^{-16}$, HR=2.5 [2.2, 2.8]) and very aggressive PCa (z=6.8, $p<10^{-11}$, HR=3.0 [2.2, 4.0]).

An alternate model used only aggressive PCa cases from the Development Set to select SNPs. Prediction for aggressive PCa onset was still significant (z=9.4, $p<10^{-16}$, HR=2.6 [2.1, 3.1]) but did not outperform the original model, so the original was used for all subsequent analyses as planned.

As PHS is predictive of PCa risk, we expected PHS to modulate the positive predictive value (PPV) of PSA testing in the Validation Set. Indeed, PPV of PSA was lower among patients with a low PHS, and higher among patients with progressively higher PHS (Figure 2). This pattern held for PPV for any PCa, as well (Supplementary Figure S2).

The distribution of PHS among aggressive PCa cases in the Validation Set is shown as a Lorenz curve in Supplementary Figure S3. Patients with PHS above the 50th percentile accounted for 76% of aggressive PCa, and the upper quintile accounted for 42% of aggressive PCa.

Family History

Using the subset of the Validation Set with known family history status (1,405 cases, 4,298 controls), the Cox test was repeated while accounting for family history. Family history alone was not predictive of age of onset of aggressive PCa (z=0.9, p=0.37, HR=1.1 [0.9, 1.4]), though there was a trend toward prediction for any PCa (z=2.0, p=0.05, HR=1.2 [1.0, 1.3]). Including family history did not improve prediction over PHS

alone for aggressive PCa (p=0.59) or any PCa (p=0.14), and PHS remained predictive when accounting for family history.

Discussion

PCa risk prediction with PHS

Genetic information may guide the decision of whether an individual patient needs PCa screening[8]. The PHS described here represents a personalized genetic assessment of a patient's age-related PCa risk that could inform both whether and when to order screening tests. When applied to data from an independent clinical trial, PHS was a highly significant predictor of age at diagnosis of aggressive PCa. Men in the top 2% of PHS had a hazard ratio of 2.9 for aggressive PCa compared to men with average risk. As PHS is representative of a man's fixed genetic risk, it can be calculated once, long before onset of PCa, and substantially inform the decision of whether he should undergo PCa screening.

PPV is directly dependent on prevalence, so if PHS predicts age of PCa onset, the PPV of PSA should vary with PHS. Figure 2 shows that this was true in the Validation Set. Nearly a quarter of the positive PSA tests in high PHS patients portended a diagnosis of aggressive PCa. The risk was much lower for low-PHS patients with elevated PSA. PHS is an indicator of the utility of PSA screening and could be influential when deciding whether to order a PSA test for a given patient.

These results also add to existing data as further evidence that genetic features are predictive of PCa risk[6–8,11,23–25]. Investigation into the genotypic features described here and elsewhere may give additional insight into biological rationales for the association with PCa.

PHS is based on hazard ratios and is therefore a relative estimate of risk. Absolute risk can be estimated within a given population if the underlying average hazard rate is known. This technique would then allow estimation of an individual PCa-free survival curve for any PHS. An example of these individual curves has been published for Alzheimer's disease[15].

Comparison with family history

Family history of PCa is one of the most commonly used risk factors in clinic to determine screening decisions[3]. However, family history was not predictive of age of onset of aggressive PCa in the Validation Set, and it did not improve prediction over PHS alone. This may reflect a lack of power to detect an association for family history in the relatively small Validation Set.

Concern of overtreatment

A concern with PSA screening is overdiagnosis and overtreatment of indolent disease. As with other genetic prediction tools, PHS is not specific for only aggressive PCa,[16] though the PHS hazard ratio was slightly higher for aggressive PCa than for any PCa.

The problem of overdiagnosis is compounded by the observation that many patients initially diagnosed with low-risk disease are later diagnosed with aggressive disease[19,17]. Active surveillance is one answer to overtreatment that avoids up-front treatment but still allows monitoring for development of indications that treatment is necessary. Indeed, most tumors eventually require treatment[17,18], and earlier treatment prevents development of metastatic disease[18]. Hence, avoiding screening altogether in patients who may develop PCa at a young age does carry risk of considerable morbidity. The present results show that PHS can help target screening efforts toward those men at highest risk of early-onset PCa or aggressive PCa requiring treatment.

Since PHS is predictive of aggressive PCa in general, it might also be useful for predicting outcomes of men diagnosed with low-risk PCa in ProtecT. The clinical data necessary to answer this interesting question have not yet been made available to the PRACTICAL consortium, so it will have to be explored in future analyses.

Previous tools

Prior studies have used GWAS-associated polymorphisms to predict risk of PCa using a case/control design[23–25]. However, epidemiologic data show that PCa risk is not a simple dichotomy of cases and controls, but rather is highly dependent on increasing age. Therefore, we opted for a survival analysis approach optimized for genetic prediction of age of PCa onset. The PHS can then be used in clinical decisions, where age plays a critical role. If a man has a high risk of developing PCa at age 95, this is a very different clinical situation from a man at high risk at age 55. A comparison of PHS with a traditional polygenic risk score (PRS) is described in the Supplementary Material.

Other PCa risk calculators use clinical variables and are most useful for a man who may already have PCa[26–28]. PSA is often included, meaning the decision of whether to screen has necessarily already been made when the tools are to be used. These are less useful for predicting his lifetime risk *before* he reaches an age where he and his physician have to decide whether he should follow some program of PCa screening.

The risk-stratification metric with best supportive evidence described in the literature is an early midlife PSA level measured at a relatively young age (e.g., <50 years). While not currently recommended in many major clinical guidelines[3–5], early midlife PSA has been shown to be predictive of future risk of PCa and lethal PCa[22,29–31]. One nested case-control study showed that just the top 10% of early PSA accounted for 40% of metastatic PCa cases[22]. This has led to a recommendation to consider PSA testing as early as age 45 in men thought by their physician to be at high risk of PCa[32]. A direct comparison of PHS and early midlife PSA for prediction of age of onset of aggressive PCa would be worthwhile. There may also be an advantage to combining the two predictors. Unfortunately, early midlife PSA is not available here, so the question is left for future work.

Limitations

The Development Set is a heterogeneous composite of several studies of varied design (Supplemental Table S1), which provides sufficient power to study SNPs with relatively

small effect sizes but also raises the concern of undetected bias in a retrospective analysis. However, the Validation Set comes from an independent, large, prospective trial, and whatever problems might exist in the Development Set, the most pertinent question is whether the model allows useful predictions.

PHS was applied here to PSA screening alone. PSA is the most prevalent screening test currently for PCa, but PHS could be expected to add value to other screening strategies, too, by predicting underlying risk of PCa for a given age and therefore influencing pre-test probability (and, by extension, PPV). This might include PSA velocity, PSA density, or some screening tool completely independent of PSA.

The evidence presented here suggests PHS can help a physician decide whether to order PSA, based on the pre-test probability and PPV of PSA for a given patient. However, this study does not address an alternate question: how PHS might compare to *diagnostic* tools (including risk calculators) that are part of the clinical work-up *after* an elevated PSA has been found. Adequate data are not available in the present dataset to answer this question, but it could be tried in future work as an additional application of PHS.

The age range of the Validation Set is limited to only 50-70 years; fortunately, this includes the age where screening is believed to have the most benefit[33–36].

Finally, ethnicity in this PHS model is limited to European ancestry. Validation of PHS in other ethnic groups—and, if necessary, custom models for each—is needed. Our group plans to investigate this important question.

Conclusions

In conclusion, we describe here the development of a new polygenic hazard score (PHS) for personalized genetic assessment of individual, age-associated PCa risk. This score has been validated in an independent dataset, demonstrating accurate prediction of aggressive PCa onset. Moreover, PHS is shown to predict the utility of PSA testing for an individual patient. This genetic risk model might play a role in guiding decisions about whether and when to screen for PCa. Investigation into the relationship of PHS and early midlife PSA is warranted.

References

1 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015; **65**: 87–108.

2 Schröder FH, Hugosson J, Roobol MJ, *et al.* Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *The Lancet* 2014; **384**: 2027–35.

3 Wolf AMD, Wender RC, Etzioni RB, *et al.* American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *CA Cancer J Clin* 2010; **60**: 70–98.

4 Horwich A, Hugosson J, de Reijke T, *et al.* Prostate cancer: ESMO Consensus Conference Guidelines 2012. *Ann Oncol Off J Eur Soc Med Oncol ESMO* 2013; **24**: 1141–62.

5 Qaseem A, Barry MJ, Denberg TD, Owens DK, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. *Ann Intern Med* 2013; **158**: 761–9.

6 Eeles RA, Olama AAA, Benlloch S, *et al.* Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 2013; **45**: 385–91.

7 Goh CL, Schumacher FR, Easton D, *et al.* Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *J Intern Med* 2012; **271**: 353–65.

8 Witte JS. Personalized Prostate Cancer Screening: Improving PSA Tests with Genomic Information. *Sci Transl Med* 2010; **2**: 62ps55-62ps55.

9 Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016; **advance online publication**. DOI:10.1038/nrg.2016.27.

10 Pharoah PDP, Antoniou AC, Easton DF, Ponder BAJ. Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer. *N Engl J Med* 2008; **358**: 2796–803.

11 Pashayan N, Duffy SW, Chowdhury S, *et al.* Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer* 2011; **104**: 1656–63.

12 Lane JA, Donovan JL, Davis M, *et al.* Active monitoring, radical prostatectomy, or radiotherapy for localised prostate cancer: study design and diagnostic and baseline results of the ProtecT randomised phase 3 trial. *Lancet Oncol* 2014; **15**: 1109–18.

13 Loeb S, Bjurlin MA, Nicholson J, *et al.* Overdiagnosis and Overtreatment of Prostate Cancer. *Eur Urol* 2014; **65**: 1046–55.

14 Mohler JL, Armstrong AJ, Bahnson RR, *et al.* Prostate Cancer, Version 1.2016. *J Natl Compr Cancer Netw JNCCN* 2016; **14**: 19–30.

15 Desikan RS, Fan CC, Wang Y, *et al.* Personalized genetic assessment of age associated Alzheimers disease risk. *BioRxiv - Press PLOS Med* 2016; : 74864.

16 Eeles R, Goh C, Castro E, *et al.* The genetic epidemiology of prostate cancer and its clinical implications. *Nat Rev Urol* 2014; **11**: 18–31.

17 Fleshner NE, Lucia MS, Egerdie B, *et al.* Dutasteride in localised prostate cancer management: the REDEEM randomised, double-blind, placebo-controlled trial. *The Lancet* 2012; **379**: 1103–11.

18 Hamdy FC, Donovan JL, Lane JA, *et al.* 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med* 2016; **0**: null.

19 Pashayan N, Pharoah P, Neal DE, *et al.* PSA-detected prostate cancer and the potential for dedifferentiation—estimating the proportion capable of progression. *Int J Cancer* 2011; **128**: 1462–70.

20 Gastwirth JL. A General Definition of the Lorenz Curve. *Econometrica* 1971; **39**: 1037–9.

21 Vickers AJ, Cronin AM, Björk T, *et al.* Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. *BMJ* 2010; **341**: c4521.

22 Vickers AJ, Ulmert D, Sjoberg DD, *et al.* Strategy for detection of prostate cancer based on relation between prostate specific antigen at age 40-55 and long term risk of metastasis: case-control study. *BMJ* 2013; **346**: f2023.

23 Amin Al Olama A, Benlloch S, Antoniou AC, *et al.* Risk Analysis of Prostate Cancer in PRACTICAL, a Multinational Consortium, Using 25 Known Prostate Cancer Susceptibility Loci. *Cancer Epidemiol Biomarkers Prev* 2015; **24**: 1121–9.

24 Pashayan N, Duffy SW, Neal DE, *et al.* Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* 2015; **17**: 789–95.

25 Mikropoulos C, Goh C, Leongamornlert D, Kote-Jarai Z, Eeles R. Translating genetic risk factors for prostate cancer to the clinic: 2013 and beyond. *Future Oncol* 2014; **10**: 1679–94.

26 Thompson IM, Ankerst DP, Chi C, *et al.* Assessing Prostate Cancer Risk: Results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst* 2006; **98**: 529–34.

27 Kranse R, Roobol M, Schröder FH. A graphical device to represent the outcomes of a logistic regression analysis. *The Prostate* 2008; **68**: 1674–80.

28 Kader AK, Sun J, Reck BH, *et al.* Potential Impact of Adding Genetic Markers to Clinical Parameters in Predicting Prostate Biopsy Outcomes in Men Following an Initial Negative Biopsy: Findings from the REDUCE Trial. *Eur Urol* 2012; **62**: 953–61.

29 Lilja H, Ulmert D, Björk T, *et al.* Long-Term Prediction of Prostate Cancer Up to 25 Years Before Diagnosis of Prostate Cancer Using Prostate Kallikreins Measured at Age 44 to 50 Years. *J Clin Oncol* 2007; **25**: 431–6.

30 Loeb S, Carter HB, Catalona WJ, Moul JW, Schroder FH. Baseline prostate-specific antigen testing at a young age. *Eur Urol* 2012; **61**: 1–7.

31 Preston MA, Batista JL, Wilson KM, *et al.* Baseline Prostate-Specific Antigen Levels in Midlife Predict Lethal Prostate Cancer. *J Clin Oncol Off J Am Soc Clin Oncol* 2016; **34**: 2705–11.

32 Carroll PR, Parsons JK, Andriole G, *et al.* NCCN Guidelines Insights: Prostate Cancer Early Detection, Version 2.2016. *J Natl Compr Cancer Netw JNCCN* 2016; **14**: 509–19.

33 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016; **66**: 7–30.

34 Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011. *CA Cancer J Clin* 2011; **61**: 212–36.

35 Jemal A, Siegel R, Ward E, *et al.* Cancer Statistics, 2006. *CA Cancer J Clin* 2006; **56**: 106–30.

36 Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer Statistics, 2001. *CA Cancer J Clin* 2001; **51**: 15–36.

Figures

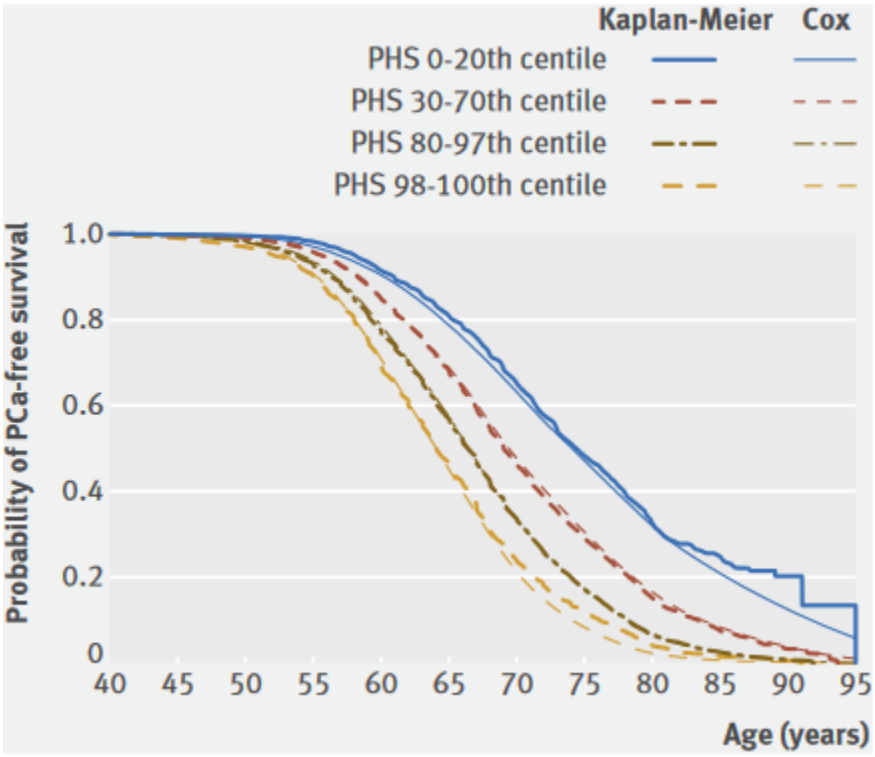

Figure 1: Kaplan-Meier estimates of prostate cancer-free survival for patients in the Development Set, grouped by PHS percentile ranges (as shown in the legend). Cox regression estimates for the same groups are shown as dotted lines of corresponding color. PHS percentiles are in reference to the distribution of PHS within the 11,190 controls in the Development Set who were under 70 years old. Time of "failure" is age at any prostate cancer diagnosis. Controls were censored at age of observation. Formal testing of proportionality is described in the Supplementary Material.

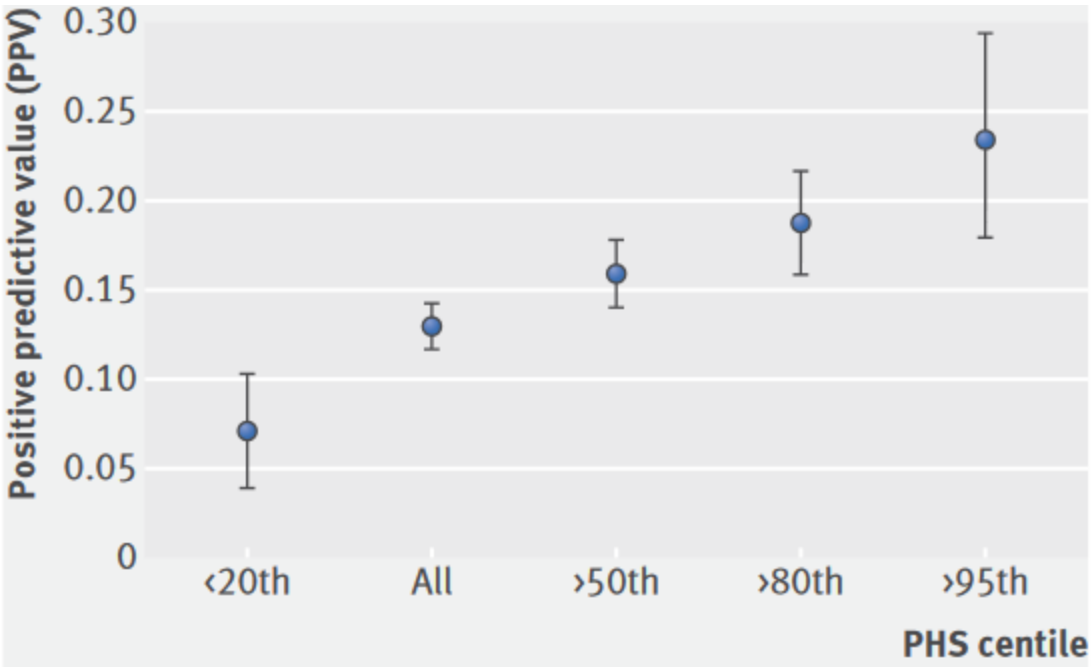

Figure 2: Positive predictive value (PPV) of PSA testing for aggressive PCa in the Validation Set. Percentiles refer to the PHS distribution among young controls in the Development Set. Colored lines are 95% confidence intervals from random samples of cases in the Validation Set (see Methods).