OPEN ACCESS

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

Tyler M Seibert,[1,2] Chun Chieh Fan,[1,3] Yunpeng Wang,[4] Verena Zuber,[4,5] Roshan Karunamuni,[1,2] J Kellogg Parsons,[6] Rosalind A Eeles,[7,8] Douglas F Easton,[9] ZSofia Kote-Jarai,[7] Ali Amin Al Olama,[9,10] Sara Benlloch Garcia,[9] Kenneth Muir,[11,12] Henrik Grönberg,[13] Fredrik Wiklund,[13] Markus Aly,[13,14,15] Johanna Schleutker,[16,17,18] Csilla Sipeky,[16,17] Teuvo LJ Tammela,[19] Børge G Nordestgaard,[20,21] Sune F Nielsen,[20,21] Maren Weischer,[21] Rasmus Bisbjerg,[22] M Andreas Røder,[23] Peter Iversen,[20,23] Tim J Key,[24] Ruth C Travis,[24] David E Neal,[25,26] Jenny L Donovan,[27] Freddie C Hamdy,[25] Paul Pharoah,[28] Nora Pashayan,[28,29] Kay-Tee Khaw,[30] Christiane Maier,[31] Walther Vogel,[31] Manuel Luedeke,[31] Kathleen Herkommer,[32] Adam S Kibel,[33] Cezary Cybulski,[34] Dominika Wokolorczyk,[34] Wojciech Kluzniak,[34] Lisa Cannon-Albright,[35,36] Hermann Brenner,[37,38,39] Katarina Cuk,[37] Kai-Uwe Saum,[37] Jong Y Park,[40] Thomas A Sellers,[41] Chavdar Slavov,[42] Radka Kaneva,[43] Vanio Mitev,[43] Jyotsna Batra,[44] Judith A Clements,[44] Amanda Spurdle,[44,45,46] Manuel R Teixeira,[47,48] Paula Paulo,[47] Sofia Maia,[47] Hardev Pandha,[49] Agnieszka Michael,[49] Andrzej Kierzek,[49] David S Karow,[1,50] Ian G Mills,[4,25,51] Ole A Andreassen,[4] Anders M Dale[1,50,52]; The PRACTICAL Consortium*

For numbered affiliations see end of article.

Correspondence to: T M Seibert
tseibert@ucsd.edu
A M Dale
amdale@ucsd.edu

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2018;360:j5757
http://dx.doi.org/10.1136/bmj.j5757

Accepted: 4 December 2017

ABSTRACT

OBJECTIVES
To develop and validate a genetic tool to predict age of onset of aggressive prostate cancer (PCa) and to guide decisions of who to screen and at what age.

DESIGN
Analysis of genotype, PCa status, and age to select single nucleotide polymorphisms (SNPs) associated with diagnosis. These polymorphisms were incorporated into a survival analysis to estimate their effects on age at diagnosis of aggressive PCa (that is, not eligible for surveillance according to National Comprehensive Cancer Network guidelines; any of Gleason score ≥7, stage T3-T4, PSA (prostate specific antigen) concentration ≥10 ng/L, nodal metastasis, distant metastasis). The resulting polygenic hazard score is an assessment of individual genetic risk. The final model was applied to an independent dataset containing genotype and PSA screening data. The hazard score was calculated for these men to test prediction of survival free from PCa.

SETTING
Multiple institutions that were members of international PRACTICAL consortium.

PARTICIPANTS
All consortium participants of European ancestry with known age, PCa status, and quality assured custom (iCOGS) array genotype data. The development dataset comprised 31 747 men; the validation dataset comprised 6411 men.

MAIN OUTCOME MEASURES
Prediction with hazard score of age of onset of aggressive cancer in validation set.

RESULTS
In the independent validation set, the hazard score calculated from 54 single nucleotide polymorphisms was a highly significant predictor of age at diagnosis of aggressive cancer ($z=11.2$, $P<10^{-16}$). When men in the validation set with high scores (›98th centile) were compared with those with average scores (30th-70th centile), the hazard ratio for aggressive cancer was 2.9 (95% confidence interval 2.4 to 3.4). Inclusion of family history in a combined model did not improve prediction of onset of aggressive PCa (P=0.59), and polygenic hazard score performance remained high when family history was accounted for. Additionally, the positive predictive value of PSA screening for aggressive PCa was increased with increasing polygenic hazard score.

CONCLUSIONS
Polygenic hazard scores can be used for personalised genetic risk estimates that can predict for age at onset of aggressive PCa.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Screening for prostate cancer (PCa) with tests for prostate specific antigen (PSA) can lead to early detection and allow for curative treatment, but universal screening also has considerable disadvantages for the men who might 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 onset is not yet available

WHAT THIS STUDY ADDS

This study presents and validates a novel polygenic hazard score that is an indicator of age at onset of aggressive PCa

The score is a relatively inexpensive assessment of an individual man's age specific 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 a million new cases and over 300 000 associated deaths estimated worldwide in 2012.[1] An international randomised controlled trial showed that screening for prostate specific antigen (PSA) resulted in a 27%

reduction in PCa mortality.[2] Because of concerns over a high rate of false positive results, in addition to aggressive treatment of apparently indolent disease, however, 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 men at high risk for whom screening and early detection can reduce morbidity and mortality.

A patient's genetic predisposition could be critical to the decision of whether and when to offer screening. Genome-wide association studies (GWAS) have shown genetic variants associated with increased risk of PCa.[6 7] These developments, combined with the recent accessibility of genotyping, provide an opportunity for cancer screening informed by genetic risk.[8] With a combination of 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 who to screen for PCa and at what age.[10 11]

We used 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 (for example, with PSA) by providing a risk stratification strategy to maximise screening efficiency. The hazard score was tested in data from an independent screening study (UK ProtecT[12]), with the hypothesis that it 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 cancer warrants treatment.[13 14] When possible, we therefore focus validation in this study on prediction of aggressive disease, defined as any tumour that would require radical treatment for a typical healthy man according to guidelines from the National Comprehensive Cancer Network (NCCN)—that is, not eligible for active surveillance.[14] This includes cancers with any of Gleason score ≥7, stage T3-T4, PSA concentration ≥10 ng/mL, nodal metastasis, or distant metastasis. Stage T2 tumours were classified without a subcategory in our database, so a patient with low Gleason score and low PSA concentration 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 tumours were included as cases of aggressive cancer.

Some additional analyses used age of diagnosis of any PCa (rather than only aggressive PCa) as complementary information. 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

To develop the polygenic hazard score model, we obtained genotype and data on age from 21 studies from the PRACTICAL consortium (table A in appendix 1), representing 31 747 men (18 868 with any PCa, 10 635 with aggressive PCa, 5406 with very aggressive PCa, 12 879 controls) of genotypic European ancestry. Age was either at diagnosis or last follow-up (for controls). Genotyping, performed with a custom Illumina array (iCOGS), and quality control steps have been described previously.[6] A total of 201 043 single nucleotide polymorphisms were available for analysis. We could not categorise cancer as aggressive or not in 4803 of the men with cancer because of incomplete data on staging; these were excluded from analyses of aggressive cancer.

Validation set

An independent study examined performance of the model. The validation set came from the ProtecT study, which screened 82 429 men with PSA testing and found 8891 men with PSA concentration over than the specified threshold of 3.0 ng/L or higher, 2896 of whom received a diagnosis of PCa.[12] Among those individuals, we obtained data on genotype and age for 6411 men (1583 with any PCa, 632 with aggressive PCa, with 220 very aggressive PCa, 4828 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 appendix 1.

Missing data

During model development, we excluded single nucleotide polymorphisms with call rates less than 95%. We imputed missing calls for the remaining polymorphisms with the mean genotype count for that allele across all participants.

Polygenic hazard score

The polygenic hazard score was developed previously as a parsimonious survival analysis model to predict the time to event outcome (in this case, age of onset of PCa). It has been published elsewhere,[15] and further details of application here are described in appendix 1.

The score is defined as the vector product of a patient's genotype (X_i) for n selected single nucleotide polymorphisms and the corresponding parameter estimates ($_i$) from a Cox proportional hazards regression (see equation).

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

Genetic prediction specifically for aggressive PCa has proved elusive, with most single nucleotide polymorphisms associated with aggressive disease also

doi: 10.1136/bmj.j5757 | *BMJ* 2018;360:j5757 | the bmj

showing association with any PCa.[16] Therefore, in the interest of maximising power to select polymorphisms associated with age of onset, we decided to initially include all cases from the development set (that is, any PCa) for generation of the model. We then tested an alternate strategy that limited generation to cases of aggressive cancer for comparison. The primary metric for validation in both instances remained prediction for aggressive cancer in the independent validation set.

To verify whether the polygenic hazard score accurately predicts age at onset of aggressive PCa, we calculated the score for all patients in the validation set and tested it as the sole predictive variable in a Cox proportional hazards regression model for age of diagnosis. Patients in the validation set with a diagnosis of low risk disease (Gleason score ≤6, PSA concentration <10 ng/L, and stage T2N0M0 or lower) were censored at time of diagnosis, reflecting the fact that it is unknown if they would later receive a diagnosis of aggressive disease or at what age that might have occurred. Significance was set at α 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 scores (>98th centile) with those with average risk (30th-70th centile). All hazard ratios presented here refer to the same pattern: high versus average risk.

Because of evidence that initially low risk disease often progresses to require treatment,[17-19] and because this might be particularly important for men with a diagnosis at a young age, we performed a secondary analogous analysis to test for prediction of age of diagnosis of any PCa. We did a secondary analysis for prediction of very aggressive disease.

To further assess the clinical significance of the polygenic hazard score, we looked at the positive predictive value (PPV) of PSA testing within the validation set, with clinical diagnosis (including biopsy result) as the ideal. We posited that risk stratification by centiles of the score would reflect the underlying incidence of PCa and therefore also affect the positive predictive value of PSA testing. Details on the calculation of positive predictive value are in appendix 1. Categories of the score were designated by centile compared with the young healthy population within the development set—that is, those controls aged <70. All centiles reported in this manuscript refer to this population.

To visualize distribution of the polygenic hazard score among cases of aggressive PCa in the validation set, we generated a Lorenz curve.[20-22]

Comparison with family history

One of the most important risk factors used currently for screening decisions is family history.[3] We compared family history and polygenic hazard score for prediction of onset of aggressive PCa using the same Cox model approach as before, with the 5703 men (1405 with any PCa, 554 with aggressive PCa, 4298 controls) from men in the validation set with known family history status (none or one or more affected first degree relatives). Models were constructed with family history alone, hazard score alone, or with both. These were compared via log likelihood tests.

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 PRACTICAL consortium website, and the respective principal investigators of each contributing study will be provided the results to disseminate to individual participants when possible.

Results

Model development

Of the 201 043 single nucleotide polymorphisms included in the dataset, 2415 were associated with increased risk of PCa in the trend test, with $P<10^{-6}$. The stepwise regression framework then identified 54 of these polymorphisms that were incorporated into the Cox proportional hazards model (table B in appendix 1). The 54 parameter estimates (for the hazard of developing PCa) were combined with individual genotype to generate the polygenic hazard score. Figure 1 shows Kaplan-Meier and Cox regression estimates for the final model. The final model performed well for prediction of age at onset of aggressive PCa in the development set ($z=37.5$, $P<10^{-16}$, hazard ratio 2.3, 95% confidence interval 2.2 to 2.4).

We excluded only 43 polymorphisms (0.02%) for low call rate during model development and used imputation for missing calls for 0.4% of calls in the final model. Of the 6411 participants in the validation set, the median individual polymorphism call rate was 100%, with a minimum of 98%.

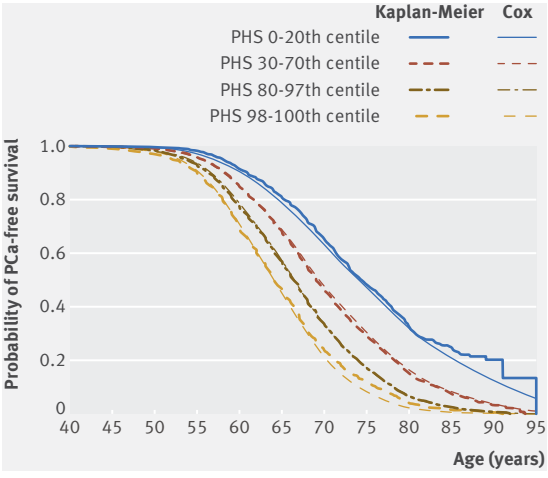

Fig 1 | Kaplan-Meier and Cox estimates of prostate cancer-free survival for patients in development set by centile ranges of polygenic hazard score. Centiles are in reference to distribution of score within 11 190 controls aged under 70 in development set. Time of "failure" is age at any diagnosis of prostate cancer. Controls were censored at age of observation. Formal testing of proportionality is described in appendix 1

Risk prediction with polygenic hazard score

In the independent validation set from the ProtecT study, a Cox proportional hazards model showed that the polygenic hazard score was a significant predictor of age at onset of aggressive PCa (z=11.2, $P<10^{-16}$). Compared with average risk, the hazard ratio for men with a high score (>98th centile) was 2.9 (95% confidence interval 2.4 to 3.4). The score was also predictive of any PCa (z=15.4, $P<10^{-16}$; hazard ratio 2.5, 2.2 to 2.8) and very aggressive PCa (z=6.8, $P<10^{-11}$; 3.0, 2.2 to 4.0).

An alternate model used only cases of aggressive PCa from the development set to select polymorphisms. Prediction for onset of aggressive PCa was still significant (z=9.4, $P<10^{-16}$, 2.6; 2.1 to 3.1) but did not outperform the original model, so we used the original for all subsequent analyses as planned.

As the polygenic hazard score was predictive of risk of PCa, we expected it to modulate the positive predictive value of PSA testing in the validation set. Indeed, the positive predictive value of PSA was lower among patients with a low score and higher among patients with progressively higher scores (fig 2). This pattern held for the positive predictive value for any PCa, as well (fig B in appendix 1).

The Lorenz curve in figure C in appendix 1 shows the distribution of the polygenic hazard score among cases of aggressive PCa in the validation set. Patients with scores above the 50th centile accounted for 76% of cases of aggressive PCa, and the upper fifth accounted for 42%.

Family history

With the subset of the validation set with known family history status (1405 cases, 4298 controls), we repeated the Cox test with adjustment for family history. Family history alone was not predictive of age of onset of aggressive PCa (z=0.9, P=0.37; hazard ratio 1.1, 95% confidence interval 0.9 to 1.4), though there was a trend toward prediction for any PCa (z=2.0, P=0.05; 1.2, 1.0 to 1.3). Inclusion of family history did not improve prediction over the polygenic hazard score alone for aggressive PCa (P=0.59) or any PCa (P=0.14), and the score remained predictive when adjusted for family history.

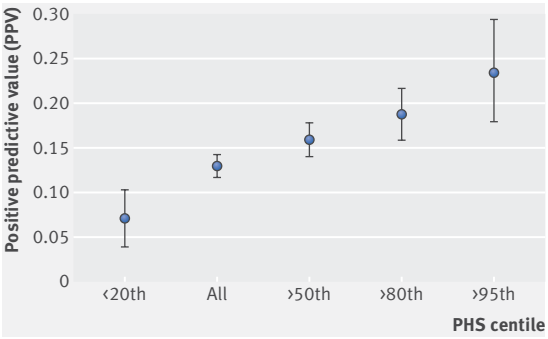

Fig 2 | Positive predictive value of PSA testing for aggressive PCa in validation set. Centiles refer to distribution of polygenic hazard score among young controls in development set. 95% confidence intervals are from random samples of cases in validation set (see methods)

Discussion

PCa risk prediction with polygenic hazard score

Genetic information can guide the decision of whether an individual patient needs PCa screening.[8] The polygenic hazard score described here represents a personalised genetic assessment of a man's age related risk that could inform both whether and when to order screening tests. When applied to data from an independent clinical trial, the score was a highly significant predictor of age at diagnosis of aggressive PCa. Men in the top 2% of the score had a hazard ratio of 2.9 for aggressive PCa compared with men with average risk. As the score 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.

Positive predictive value is directly dependent on prevalence, so if the polygenic hazard score predicts age of onset of PCa, the positive predictive value of PSA should vary with the score. Figure 2 shows that this was true in the validation set. Nearly a quarter of the positive PSA test results in men with a high score portended a diagnosis of aggressive PCa. The risk was much lower for men with low scores with a raised PSA concentration. The score is an indicator of the utility of PSA screening and could be influential in the decision whether to order a PSA test for a given patient.

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

The polygenic hazard score is based on hazard ratios and is therefore an estimate of relative 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] Family history, however, was not predictive of age of onset of aggressive PCa in the validation set, and it did not improve prediction over the hazard score alone. This could 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, the polygenic hazard score is not specific for aggressive PCa alone,[16] though the hazard

doi: 10.1136/bmj.j5757 | *BMJ* 2018;360:j5757 | the bmj

ratio was slightly higher for aggressive PCa than for any PCa. The problem of overdiagnosis is compounded by the observation that many men with an initial diagnosis of low risk disease later receive a diagnosis of aggressive disease.[17 19] Active surveillance is one answer to overtreatment that avoids up front treatment but still allows for monitoring for development of indications that treatment is necessary. Indeed, most tumours eventually require treatment,[17 18] and earlier treatment prevents development of metastatic disease.[18] Hence, avoiding screening altogether in patients who might develop PCa at a young age does carry risk of considerable morbidity. The present results show that the polygenic hazard score can help to target screening efforts toward those men at highest risk of early onset PCa or aggressive PCa requiring treatment.

As the score is predictive of aggressive PCa in general, it might also be useful for predicting outcomes of men with a diagnosis of 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

Previous studies have used GWAS-associated polymorphisms to predict risk of PCa with a case-control design.[23-25] Epidemiological data, however, show that risk of PCa is not a simple dichotomy of cases and controls but rather is highly dependent on increasing age. We therefore opted for a survival analysis approach optimised for genetic prediction of age of onset of PCa. The polygenic hazard score can then be used in clinical decisions, when age plays a critical role. If a man has a high risk of developing PCa at age 95, this is a different clinical situation from a man at high risk at age 55. A comparison of the polygenic hazard score with a traditional polygenic risk score is described in appendix 2.

Other PCa risk calculators use clinical variables and are most useful for a man who might already have PCa.[26-28] PSA concentration 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 at which he and his physician have to decide whether he should follow some programme of PCa screening.

The risk stratification metric with best supportive evidence described in the literature is an early midlife PSA concentration measured at a relatively young age (for example, <50). 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 the distribution of concentrations of PSA in tests done in men aged under 50 accounted for 40% of cases of metastatic PCa.[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.[32] A direct comparison of the polygenic hazard score and early midlife PSA for prediction of age at onset of aggressive PCa would be worthwhile. There might also be an advantage to combining the two predictors. Unfortunately, early midlife PSA concentration was not available in the datasets used in the present study so the question is left for future work.

Limitations

The development set was a heterogeneous composite of several studies of varied design (table A in appendix 1), which provides sufficient power to study single nucleotide polymorphisms with relatively small effect sizes but also raises the concern of undetected bias in a retrospective analysis. The validation set, however, came 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.

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

The evidence presented here suggests that the polygenic hazard score can help a physician decide whether to order PSA, based on the pretest probability and positive predictive value of PSA for a given patient. Our study does not, however, deal with an alternate question: how the hazard score might compare to diagnostic tools (including risk calculators) that are part of the clinical investigations after a raised PSA concentration 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 the polygenic hazard score.

The age range of the validation set was limited to 50-70; fortunately, this includes the age at which screening is believed to have the most benefit.[33-36]

Finally, ethnicity in this model was limited to European ancestry. Validation of the score in other ethnic groups—and, if necessary, custom models for each—is needed. We plan to investigate this important question.

Conclusions

In conclusion, we describe here the development of a new polygenic hazard score for personalised genetic assessment of individual age associated risk of PCa. This score has been validated in an independent dataset, showing accurate prediction of onset of aggressive PCa. Moreover, the score can predict the utility of PSA testing for an individual man. This genetic risk model might play a role in guiding decisions about whether and when to screen for PCa. Investigation into the relation between the score and early midlife PSA testing is warranted.

AUTHOR AFFILIATIONS

[1]Center for Multimodal Imaging and Genetics, University of California, San Diego, La Jolla, CA, USA

[2]Department of Radiation Medicine and 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]NORMENT, KG Jebsen Centre, Oslo University Hospital and 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]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, 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, Solna, 171 76 Stockholm, Sweden

[15]Department of Urology, Karolinska University Hospital, Solna, 171 76 Stockholm, Sweden

[16]Department of Medical Biochemistry and Genetics, Institute of Biomedicine, Kiinamyllynkatu 10, FI-20014 University of Turku, Finland

[17]Tyks Microbiology and Genetics, Department of Medical Genetics, Turku University Hospital, Turku, Finland

[18]BioMediTech, 30014 University of Tampere, Tampere, Finland

[19]Department of Urology, Tampere University Hospital and Medical School, University of Tampere, Finland

[20]Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[21]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark

[22]Department of Urology, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark

[23]Copenhagen Prostate Cancer Centre, Department of Urology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

[24]Cancer Epidemiology Unit, Nuffield Department of Population Health University of Oxford, Oxford OX3 7LF, UK

[25]Nuffield Department of Surgical Sciences, Faculty of Medical Science, University of Oxford, John Radcliffe Hospital, Oxford, UK

[26]University of Cambridge, Department of Oncology, Box 279, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK

[27]School of Social and Community Medicine, University of Bristol, Bristol BS8 2PS, UK

[28]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK

[29]University College London, Department of Applied Health Research, London WC1E 7HB, UK

[30]Clinical Gerontology Unit, University of Cambridge, Cambridge UK

[31]Institute of Human Genetics, University Hospital of Ulm, Ulm, Germany

[32]Department of Urology, Klinikum rechts der Isar der Technischen Universitaet Muenchen, Munich, Germany

[33]Division of Urologic Surgery, Brigham and Women's Hospital, Dana-Farber Cancer Institute, 75 Francis Street, Boston, MA 02115, USA

[34]International Hereditary Cancer Centre, Department of Genetics and Pathology, Pomeranian Medical University, Szczecin, Poland

[35]Division of Genetic Epidemiology, Department of Medicine, University of Utah School of Medicine, Salt Lake City, Utah, USA

[36]George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, Utah, USA

[37]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany

[38]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

[39]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany

[40]Department of Cancer Epidemiology, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA

[41]Office of the Center Director, Moffitt Cancer Center, 12902 Magnolia Drive, Tampa, FL 33612, USA

[42]Department of Urology and Alexandrovska University Hospital, Medical University, Sofia, Bulgaria

[43]Department of Medical Chemistry and Biochemistry, Molecular Medicine Center, Medical University, Sofia, 2 Zdrave Str, 1431 Sofia, Bulgaria

[44]Australian Prostate Cancer Research Centre-Qld, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia

[45]Molecular Cancer Epidemiology Laboratory, Queensland Institute of Medical Research, Brisbane, Australia

[46]Australian Prostate Cancer BioResource, Institute of Health and Biomedical Innovation and School of Biomedical Science, Queensland University of Technology, Brisbane, Australia

[47]Department of Genetics, Portuguese Oncology Institute, Porto, Portugal

[48]Biomedical Sciences Institute (ICBAS), University of Porto, Porto, Portugal

[49]University of Surrey, Guildford, Surrey, GU2 7XH

[50]Department of Radiology, University of California, San Diego, La Jolla, CA, USA

[51]Centre for Cancer Research and Cell Biology, Queens University Belfast, Belfast, UK

[52]Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

*Details of 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/), acknowledgments, and funding are provided in appendix 2.

Contributors: TMS, CCF, VZ, DSK, IGM, OAA, and AMD designed the study. RAE, DFE, ZSKJ, AAAO, SBG, KM, HG, FW, MA, JS, CSi, TLJT, BGN, SFN, MW, RB, MAR, PI, TJK, RCT, DEN, JLD, FCH, PPh, NP, KTK, CM, WV, ML, KH, ASK, CC, DW, WK, LCA, HB, KC, KUS, JYP, TAS, CSI, RKan, VM, B, JAC, AS, APCBR, MRT, PPa, SM, HP, AM, and AK collected the data. TMS, CCF, IGM, OAA, and AMD performed the literature search. TMS, CCF, YW, VZ, RKar, DSK, and AMD performed the data analysis. TMS, CCF, RKar, JKP, DSK, OAA, and AMD interpreted the data. TMS, RKar, JKP, DSK, OAA, and AMD created the figures. TMS, CCF, OAA, and AMD wrote the manuscript. All authors reviewed the manuscript, added appropriate revisions, agreed to submission for publication, and approved the final version. TMS and AMD are guarantors.

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, TMS reports honoraria from WebMD 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, OAA has a patent application (US 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, advisory board member of Human Longevity, recipient of non-financial research support from General Electric Healthcare; no financial relationships with any companies that might have an interest in the submitted work in the

previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

Ethics approval: Not required.

Funding: This study was funded in part by grants from the US Department of Defense (W81XWH-13-1-0391), Prostate Cancer Foundation, 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 appendix 2.

Transparency: 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.

Data sharing: No additional data available.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

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. doi:10.3322/caac.21262
2 Schröder FH, Hugosson J, Roobol MJ, ERSPC Investigators. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet* 2014;384:2027-35. doi:10.1016/S0140-6736(14)60525-0
3 Wolf AMD, Wender RC, Etzioni RB, American Cancer Society Prostate Cancer Advisory Committee. American Cancer Society guideline for the early detection of prostate cancer: update 2010. *CA Cancer J Clin* 2010;60:70-98. doi:10.3322/caac.20066
4 Horwich A, Hugosson J, de Reijke T, Wiegel T, Fizazi K, Kataja V, Panel Members, European Society for Medical Oncology. Prostate cancer: ESMO Consensus Conference Guidelines 2012. *Ann Oncol* 2013;24:1141-62. doi:10.1093/annonc/mds624
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. doi:10.7326/0003-4819-158-10-201305210-00633
6 Eeles RA, Olama AAA, Benlloch S, COGS–Cancer Research UK GWAS–ELLIPSE (part of GAME-ON) Initiative, Australian Prostate Cancer Bioresource UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT (Prostate testing for cancer and Treatment) Study Collaborators, PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. *Nat Genet* 2013;45:385-91, e1-2. doi:10.1038/ng.2560
7 Goh CL, Schumacher FR, Easton D. Genetic variants associated with predisposition to prostate cancer and potential clinical implications. *J Intern Med* 2012;271:353-65. doi:10.1111/j.1365-2796.2012.02511.x
8 Witte JS. Personalized prostate cancer screening: improving PSA tests with genomic information. *Sci Transl Med* 2010;2:62ps55. doi:10.1126/scitranslmed.3001861
9 Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016;17:392-406.. 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. doi:10.1056/NEJMsa0708739
11 Pashayan N, Duffy SW, Chowdhury S. Polygenic susceptibility to prostate and breast cancer: implications for personalised screening. *Br J Cancer* 2011;104:1656-63. doi:10.1038/bjc.2011.118
12 Lane JA, Donovan JL, Davis M, ProtecT study group. 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. doi:10.1016/S1470-2045(14)70361-4
13 Loeb S, Bjurlin MA, Nicholson J. Overdiagnosis and overtreatment of prostate cancer. *Eur Urol* 2014;65:1046-55. doi:10.1016/j.eururo.2013.12.062
14 Mohler JL, Armstrong AJ, Bahnson RR. Prostate Cancer, Version 1.2016. *J Natl Compr Canc Netw* 2016;14:19-30. doi:10.6004/jnccn.2016.0004
15 Desikan RS, Fan CC, Wang Y. Personalized genetic assessment of age associated Alzheimers disease risk: Development and validation of a polygenic hazard score. *PLoS Med* 2016;14:e1002289.doi.org/10.1371/journal.pmed.1002258
16 Eeles R, Goh C, Castro E. The genetic epidemiology of prostate cancer and its clinical implications. *Nat Rev Urol* 2014;11:18-31. doi:10.1038/nrurol.2013.266
17 Fleshner NE, Lucia MS, Egerdie B. Dutasteride in localised prostate cancer management: the REDEEM randomised, double-blind, placebo-controlled trial. *Lancet* 2012;379:1103-11. doi:10.1016/S0140-6736(11)61619-X
18 Hamdy FC, Donovan JL, Lane JA, ProtecT Study Group. 10-Year Outcomes after Monitoring, Surgery, or Radiotherapy for Localized Prostate Cancer. *N Engl J Med* 2016;375:1415-24. doi:10.1056/NEJMoa1606220
19 Pashayan N, Pharoah P, Neal DE. PSA-detected prostate cancer and the potential for dedifferentiation--estimating the proportion capable of progression. *Int J Cancer* 2011;128:1462-70. doi:10.1002/ijc.25471
20 Gastwirth JL. A General Definition of the Lorenz Curve. *Econometrica* 1971;39:1037-9doi:10.2307/1909675.
21 Vickers AJ, Cronin AM, Björk T. Prostate specific antigen concentration at age 60 and death or metastasis from prostate cancer: case-control study. *BMJ* 2010;341:c4521. doi:10.1136/bmj.c4521
22 Vickers AJ, Ulmert D, Sjoberg DD. 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. doi:10.1136/bmj.f2023
23 Amin Al Olama A, Benlloch S, Antoniou AC, UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, PRACTICAL Consortium. 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. doi:10.1158/1055-9965.EPI-14-0317
24 Pashayan N, Duffy SW, Neal DE. Implications of polygenic risk-stratified screening for prostate cancer on overdiagnosis. *Genet Med* 2015;17:789-95. doi:10.1038/gim.2014.192
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. doi:10.2217/fon.14.72
26 Thompson IM, Ankerst DP, Chi C. Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial. *J Natl Cancer Inst* 2006;98:529-34. doi:10.1093/jnci/djj131
27 Kranse R, Roobol M, Schröder FH. A graphical device to represent the outcomes of a logistic regression analysis. *Prostate* 2008;68:1674-80. doi:10.1002/pros.20840
28 Kader AK, Sun J, Reck BH. 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. doi:10.1016/j.eururo.2012.05.006
29 Lilja H, Ulmert D, Björk T. 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. doi:10.1200/JCO.2006.06.9351
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. doi:10.1016/j.eururo.2011.07.067
31 Preston MA, Batista JL, Wilson KM. Baseline Prostate-Specific Antigen Levels in Midlife Predict Lethal Prostate Cancer. *J Clin Oncol* 2016;34:2705-11. doi:10.1200/JCO.2016.66.7527
32 Carroll PR, Parsons JK, Andriole G. NCCN Guidelines Insights: Prostate Cancer Early Detection, Version 2.2016. *J Natl Compr Canc Netw* 2016;14:509-19. doi:10.6004/jnccn.2016.0060
33 Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin* 2016;66:7-30. doi:10.3322/caac.21332
34 Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011: the impact of eliminating socioeconomic and racial disparities on premature cancer deaths. *CA Cancer J Clin* 2011;61:212-36. doi:10.3322/caac.20121
35 Jemal A, Siegel R, Ward E. Cancer statistics, 2006. *CA Cancer J Clin* 2006;56:106-30. doi:10.3322/canjclin.56.2.106
36 Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer statistics, 2001. *CA Cancer J Clin* 2001;51:15-36. doi:10.3322/canjclin.51.1.15

Appendix 1: Supplementary methods, tables, and figures

Appendix 2: Details of additional consortium members, acknowledgements, and funding