

**The efficacy of different biomarkers and endpoints to refine referrals for suspected prostate cancer.
The TARGET study (Tiered integrAted tests for eaRly diaGnosis of clinically significant ProstatE
Tumours)**

Artitaya Lophatananon¹, Kenneth R Muir¹, and Vincent J Gnanapragasam^{2,3#}

¹Division of Population Health, Health Services Research & Primary Care Centre, University of Manchester

²Division of Urology, Department of Surgery, University of Cambridge, Cambridge, UK

³Cambridge Urology Translational Research and Clinical Trials Office, Cambridge Biomedical Campus, Addenbrooke's Hospital, Cambridge, UK

#Corresponding author email: vjg29@cam.ac.uk.

Correspondence to Professor Vincent J Gnanapragasam, Cambridge Prostate Cancer Research and Clinical Trials Group, University of Cambridge Box 193, Cambridge Biomedical Campus Cambridge CB20QQ

Corresponding author ORCID: 0000-0003-4722-4207

Abstract

Background

The majority of men referred with a raised PSA for suspected prostate cancer will receive unnecessary tertiary investigations including MRI and biopsy. Here we compared different types of biomarkers to refine tertiary referrals and when different definitions of clinically significant cancer were used.

Methods

Data and samples from 798 men referred for a raised PSA (≥ 3 ng/ml) and investigated through an MRI-guided biopsy pathway were accessed for this study. Bloods were acquired pre-biopsy for liquid biomarkers and germline DNA. Variables explored included PSA+Age (base model), Free/Total PSA (FTPSA), Prostate Health Index (*phi*), PSA density (PSAd), Polygenic Risk Score (PRS) and MRI ($\geq$ LIKERT 3). Different diagnostic end-points for significant cancer ($\geq$ Grade Group 2 [GG2], $\geq$ GG3, $\geq$ Cambridge Prognostic Group 2 [CPG2], $\geq$ CPG3) were tested. The added value of each biomarker to the base model was evaluated using regression models, AUC and Decision Curve Analysis (DCA) plots.

Results

The median age and PSA was 65y and 7.13 ng/ml respectively. Depending on definition of clinical significance $\geq$ Grade Group 2 (GG2) was detected in 57.0% (455/798), $\geq$ GG3 in 27.5% (220/798), $\geq$ CPG2 in 61.6% (492/798) and $\geq$ CPG3 in 42.6% (340/798). In the pre-MRI context, the PSA+Age (base model) AUC for prediction of $\geq$ GG2, $\geq$ GG3, $\geq$ CPG2 and $\geq$ CPG3, was 0.66, 0.68, 0.70 and 0.75 respectively. Adding *phi* and PSAd to base model improved performance across all diagnostic end-points but was notably better when the composite CPG prognostic score was used: AUC 0.82, 0.82, 0.83, 0.82 and AUC 0.74, 0.73, 0.79, 0.79 respectively. In contrast neither FTPSA or PRS scores improved performance especially in detection of $\geq$ GG3 and $\geq$ CPG3 disease. Combining biomarkers did not alter results. Models using *phi* and PSAd post-MRI also improved performances but again benefit varied with diagnostic endpoint. In DCA analysis models which incorporated PSAd and *phi* in particular were effective at reducing use of MRI and/or biopsies especially for $\geq$ CPG3 disease.

Conclusion

Incorporating *phi* or PSAd can refine and tier who is referred for tertiary imaging and/or biopsy after a PSA test. Incremental value however varied depending on the definition of clinical significance and was particularly useful when composite prognostic endpoints are used.

Keywords

Prostate cancer

Early detection

PSA

PSA density

Free Total PSA

Polygenic risk scores

Prostate Health Index (*phi*)

MRI

Biopsy

Cambridge Prognostic Groups

Background

Prostate Cancer has a high incidence but a relatively low mortality rate (1). To date, screening studies have failed to show a clear benefit of survival against the counter-balances of over-diagnosis and over-treatment (2). While this paradigm has been improved with the introduction of MRI (magnetic resonance imaging) pre-biopsy, it remains the case that many men need to be tested and referred to a tertiary centre before scans and biopsies are done to confirm or refute if a significant cancer is present (3). Indeed, it is estimated that for every man diagnosed with a significant cancer, up to 2 others will have unnecessary referral and investigations (3-4). Natural history studies have shown that almost all men will develop cancer foci in the prostate with age but relatively few go on to cause morbidity (5). Our understanding of what is a prognostically important cancer has also evolved greatly and the concept of not detecting cancers too early is now widely accepted (6).

Current approaches to early detection have tended to view men as a group from which the one at risk needs to be picked out, and referred. The key conundrum is the balance between detecting something too early for which treatment is not (or will ever be) needed and hence confer no survival gain; but on the other hand, not detecting cancers too late when curative treatment is not possible. All within the context of a naturally growing prostate gland and aging male. Whatever the approach, cost and cost-benefit and linking with survival outcomes is going to be key for any prostate cancer early detection programme in an ever more expensive health care environment (7). These aspects argue strongly for a more rational and tiered strategy to the use of investigations and referrals.

Hitherto, groups working in early detection have generally fallen into separate camps e.g. using SNP baseline population testing, bio-marker panels or imaging tests to visualise lesions (8-10). None have gained traction to replace the simple PSA test. Thus, PSA will likely remain the first and most important baseline test and this is the case in many current renewed efforts to re-evaluate screening. Here we tested the incremental value of different types of biomarkers in addition to PSA to detect significant cancer and how performance may vary depending on the definition of “clinical significance”. Our goal was to derive the optimal sequence or testing tiers to (i) Reduce the overall investigated population in men with a raised PSA (ii) hence refine or reduce use of imaging and/or biopsies and (iii) explore the impact of using different definitions of clinically significant disease.

Methods

Patients

Data and samples from men referred for a raised PSA ($\geq 3\text{ng/ml}$), investigated through an MRI-guided biopsy pathway and who consented to participating in the DIAMOND biomarker study (research ethics 03/18.) were accessed for this study (period from 2013-2019). All men have had pre-biopsy MRI, Likert scores assigned and image-guided fusion biopsies (or systematic biopsies if no target seen) with resultant accurate disease characterisation. All men were biopsy naïve. Diagnostic outcomes of any cancer and significant cancers were recorded as well as a benign outcome. Different diagnostic end-points for significant cancer ($\geq$ Grade Group 2 [GG2], $\geq$ GG3, and composite prognosis group $\geq$ Cambridge Prognostic Group 2 [CPG2] and $\geq$ CPG3) based on the UK NICE guidelines were tested (11). The Cambridge Prognostic Groups are a composite prognostic stratification system that has been shown to outperform older definitions of low, intermediate and high-risk and more accurately predict prognosis from a diagnosis of prostate cancer (6,11).

Biomarkers studied

Research bloods were acquired pre-biopsy for liquid biomarkers and germline DNA. Variables explored included Polygenic Risk Score (PRS), PSA (ng/mL) + Age (base model), Free/Total PSA (FTPSA), Prostate Health Index (ϕ), PSA density (PSAd) and MRI ($\geq$ LIKERT 3). The added value of each biomarker to the base model was evaluated using regression models and AUC.

Polygenic Risk scores: Genotyping was carried out using Illumina Onco-array of $\sim 500\text{k}$ SNPs (single nucleotide polymorphism). Data quality control was carried out by filtering out any SNPs with Mean Allele Frequencies (>0.001) and Hardy-Weinberg equilibrium ($p>1\text{E-}10$). Samples were also checked for missingness ($\leq 1\%$), sex (male sex with no chromosome aneuploidy and no discrepancy between self-reported and genetic sex). PRS were derived from a panel of 269 SNPs previously published (12). For missing SNPs, we carried out imputation from Chr1 to Chr X using the Michigan Imputation platform. Total SNPs included in our analyses were 245 SNPs. The genetic data file (PED and MAP files) were converted into STATA to generate a single weighted polygenic risk score ($\text{PRS}_j = \sum N\beta_i \times \text{dosage}_{ij}$) where N was the number of SNPs in the score, β_i is the effect size (or beta) of variant i and dosage_{ij} is the number of copies of SNP i in the genotype of individual j . For each outcome, a standardised score was generated by dividing the raw score by the SD in control group (no cancer or benign group).

PSA, FTPSA and PSAd: DiaSorin Liaison assays were used for total and free PSA assays and performed by the NIHR Cambridge Biomedical Research Centre, Core Biochemistry Assay Laboratory. PSA density (PSAd) was calculated using MRI-defined prostate volumes using the ellipsoid formula.

phi: *phi* assays were handled according to the manufacturer's recommendations (Beckman Coulter). Samples were centrifuged and frozen at – 80 °C within 3 h before dispatch to a central laboratory and performed on a Beckman Coulter Access Autoanalyzer. Quality Assurance samples were analysed before and after each batch to ensure the validity of the results. All quality control results were within Beckman Coulter's target ranges.

MRI: MRI on 1.5 T or 3 T systems with multi-channel surface phased array coils were performed including standard anatomical and functional imaging (diffusion weighted and contrast enhanced). Image acquisition and processing was performed in accordance with local standard clinical protocols. Sequences were evaluated and scoring was performed according to Prostate Imaging Reporting and Data ver. 2 (PI-RADS version 2) and using a Likert scale (as this is the standard practice in our department). Likert 1–2 lesions were considered MRI negative for this study and positive lesions denoted as 3–5, the latter denoted as MRI Likert ≥ 3 . The quality of our MRI programme in our unit has been previously documented (3).

Statistical analysis

Data analysis was carried out using STATA 15 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC.). For biomarkers except *Phi*, we imputed missing values and arbitrarily created 10 imputation datasets. We used *mi* command to generate imputed datasets and ran logistic regression and obtain *MI* estimates AUC. For *phi* (n=284 men), logistic regression was performed using complete case analysis and subsequently derived AUC for each fitted model. To ensure no selection bias, comparison of the *phi* sub-cohort versus the rest of the cohort was done and did not show any differences in baseline attributes. Decision curve analysis to assess net reduction of imaging/biopsy was carried out using *dca* command.

Results

Cohort demographics

Data were available for analysis in 798 men. The median age was 65 years and median PSA and prostate volume were 7.13 ng/ml and 43.9 mls respectively (Table 1). Overall any cancer was detected in 76.5% (611/798) of men. Depending on definition of clinical significance $\geq$ Grade Group 2 (GG2) cancer was detected in 57.0% (455/798), $\geq$ GG3 in 27.5% (220/798), $\geq$ Cambridge Prognostic Group 2 (CPG2) in 61.6% (492/798) and $\geq$ CPG3 in 42.6% (340/798) of men. 2.5% (20 men) were already metastatic at diagnosis. The distributions of Grade Groups, clinical stage and CPG are shown in Table 1. The biomarkers tested in this study were evaluable for the vast majority of men except for *phi* which was only available in a subset of 284 men and FTPSA available in 678 (Table 2). Median Polygenic Risk Score (PRS) was 20.6, PSA density (PSAd) 0.16 ng/mL², Free/Total ratio 0.11 and *phi* score 41.9. MRI Likert score was considered an independent variable using a Likert score of ≥ 3 as a positive scan (Table 2). Only 2 men were not able to have MRI in this study. Notably over a quarter of men had no visible lesion on MRI (28.7%) affirming that this cohort was not particularly selected or biased towards men with a positive MRI.

Performance in prediction of disease detection

Polygenic risk scores (PRS)

PRS in this study did add predictive value to detection of any cancers when compared to the base model of PSA+Age in both the pre-MRI setting (AUC 0.73 vs 0.65, $p < 0.00001$) (Additional File Table S1A) and post MRI setting (AUC 0.79 vs 0.74, $p = 0.0002$) (Additional File Table S1B). However, PRS scores did not add incremental value in predicting $\geq$ GG2 or $\geq$ GG3 disease in the pre-MRI setting (AUC 0.68 vs 0.66 and 0.69 vs 0.68, $p = 0.08$ and $p = 0.59$ respectively) (Table 3). Using detection of $\geq$ CPG2 disease as the endpoint in the pre-MRI setting showed some modest benefit (AUC 0.73 vs. 0.70) but not for detection of $\geq$ CPG3 disease (0.76 vs. 0.75, $p = 0.16$). In the post-MRI setting (MRI Likert ≥ 3) this lack of benefit for significant cancer detection was also evident with only marginal and mostly non-significant gains in AUC (Table 4). These data do not appear to support use of PRS scores in reliably predicting significant cancer presence in a PSA elevated population. The PRS panel used in particular did not add value in detection of $\geq$ GG3 or $\geq$ CPG3 i.e. higher grade and poorer prognosis disease

Free to Total PSA ratio (FTPSA)

FTPSA, in common with all the biomarkers tested in this study, added some value in predicting detection of any cancer (Additional File Table S1A&B) though incremental benefit did not match those of PSAd or *phi*. For detection of $\geq$ GG2, FTPSA only added a modest incremental benefit in the pre-MRI

context (AUC 0.71 vs. 0.66) but no significant enhanced value for $\geq$ GG3 detection (AUC 0.69 vs. 0.68, $p=0.59$) (Table 3). In detection of $\geq$ CPG2, adding FTPSA provided an increase in AUC (0.75 vs. 0.70) but again only a small incremental value in detection of $\geq$ CPG3 disease (AUC 0.76 vs. 0.75). In the post-MRI setting (MRI Likert ≥ 3) FTPSA did show some incremental value in detection of $\geq$ GG2 (AUC 0.75 vs. 0.72) and $\geq$ CPG2 (0.79 vs. 0.75) disease but not for $\geq$ GG3 or $\geq$ CPG3 disease (Table 4). Thus, overall, like for PRS, the addition of FTPSA failed to show much benefit over PSA+Age or PSA+Age+MRI Likert ≥ 3 base especially in detection of higher-grade/poorer prognosis disease.

PSA density (PSAd)

PSAd demonstrated consistent improvement in detection of any cancer and significant cancers regardless of definition used in the pre-MRI context (all comparisons $p<0.001$) (Additional File Table S1A and Table 3). Notably the incremental gain was highest when the end point was detection of composite prognostic detection endpoints i.e. $\geq$ CPG2 (AUC 0.77 vs. 0.70) and $\geq$ CPG3 disease (0.79 vs. 0.75) (Table 3). Of note PSAd added to the base model of PSA+Age performed similarly to a base model of PSA+Age+MRI Likert ≥ 3 in detection of $\geq$ GG2, $\geq$ GG3, $\geq$ CPG2 and $\geq$ CPG3 disease (Table 3 and 4). The added value of PSAd was also evident in the post-MRI setting regardless of endpoint definition (Table 4). Here again the incremental value was highest when the end point was a composite prognostic endpoint; $\geq$ CPG2 (AUC 0.80 vs. 0.75) and $\geq$ CPG3 disease (0.81 vs. 0.78) (Table 4). Further adding FTPSA to PSAd did not add incremental value in either the pre or post MRI models (Table 3 and 4).

Prostate Health Index (phi)

Of all the biomarkers test, *phi* added the greatest incremental value for any cancer detection scenario with AUC consistently high for any cancer or significant cancer endpoints in both the pre-MRI and post MRI setting (Additional File Tables S1 A&B, 3 and 4). In the pre-MRI context performance was equally strong for detection of $\geq$ GG2 and $\geq$ GG3 (AUC 0.82 and 0.82 vs. 0.66) disease as it was for detection of $\geq$ CPG2 and $\geq$ CPG3 disease (AUC 0.83 and 0.82 vs. 0.66) ($p<0.0001$ for all comparisons). Importantly, including *phi* in the pre-MRI base model outperformed the predictive accuracy afforded by using MRI Likert ≥ 3 for detection of $\geq$ GG2, $\geq$ GG3, $\geq$ CPG2 and $\geq$ CPG3 disease (Table 3 and 4). Adding *phi* to the post-MRI model (PSA+Age+MRI Likert ≥ 3) further improved overall performance achieving AUC of 0.84-0.86.

Decision curve analysis (DCA) in reducing over-investigation

To understand the value in reducing over-investigation, we generated DCA for PSAd and *phi*, the 2 most promising biomarkers in our study. Using PSAd and with $\geq$ CPG2 as an endpoint; above a threshold of a 30% risk of missing cancer using PSA+Age+PSAd dominated with net reductions in MRI

and the need for subsequent biopsies which is probably too high a risk in routine clinical practice (Figure 1A). With $\geq$ CPG3 as an endpoint the net benefit however was gained at much lower threshold probabilities ($\geq 10\%$ risk of missing cancer) (Figure 1B). In models to only reduce biopsies (post-MRI), the dominant model was a combination of PSA+Age+PSAd+MRI Likert ≥ 3 , with net reductions in biopsies above threshold probabilities of 20% (Figure 1C). With $\geq$ CPG3 as an endpoint the net benefit of PSA+Age+PSAd+MRI ≥ 3 was evident regardless of the threshold probability and dominated all other models (Figure 1D). With the *phi* assay, it was clear that all models which incorporated *phi* were effective at reducing use of both MRI and biopsies. In the pre-MRI setting whether the endpoint was detection of $\geq$ CPG2 or $\geq$ CPG3, use of *phi* resulted in a net reduction in number of investigations across all risk thresholds (Figure 2A and B). Similarly, in models to only reduce biopsies, the dominant model was a combination of PSA+Age+*phi*+ MRI Likert ≥ 3 , with net reductions in biopsies evident across all threshold probabilities and in detection of both $\geq$ CPG2 or $\geq$ CPG3 disease (Figure 2C and D). Taken together these results suggest that PSAd and *phi* may be useful to refine and reduce use of MRI and biopsies as part of a stepwise tiered diagnostic pathway but this is dependent on the definition of significance used.

Discussion

In this study we compared and modelled different biomarkers to refine the predictive value of the ubiquitous PSA test. We find that the performance of biomarkers added to PSA differs based on the definition of clinical significance used and also whether in the pre or post MRI context. We further identified that using a composite definition of significant cancer based on prognosis resulted in better biomarker performance and a more contextualised value for decisions to investigate. Consistently, all biomarkers were able to better detect composite CPG2 and CPG3 disease compared to using a histological Grade Group endpoint alone. MRI has been a game change in modern prostate diagnostics but remains a resource expensive and interpreter dependent tool. Maintaining high quality services require regular quality assessment and standardisation (13). In contrast, liquid biomarker tests are less subjective and simpler to adopt and standardise though measurements are known to vary between providers using different biochemical assays (14). Considering the health resource needs of the future as well as environmental imperative to reduce energy consumption, it is important to identify tests that can refine and reduce the use of imaging as well as biopsy (15). In our opinion the MRI and biopsy should therefore be considered a single integrated tertiary referral endpoint of the diagnostic pathway.

To our knowledge this is the first study to explore the relative value of different tests including genomics, serum-based markers and imaging within the same cohort. Our work suggests that a primary additional value is gained by serum biomarkers based on PSA isoforms and prostate volume. The *phi* and PSA density tests consistently improved the predictive ability to detect significant cancer regardless of the endpoint definition. Indeed, use of these biomarkers added to base models of PSA+Age out-performed MRI. However, it is worth noting that the DCA showed that the benefit gain for PSAd depends very much on the definition of clinical significance used and context (i.e. whether trying to reduce MRI and biopsies or biopsies alone). As an example, a biopsy threshold of $\geq 30\%$ is too high for normal standard practice for missing CPG2 cancers (if this was the definition of clinical significance used). However, with using CPG3 as a definition of clinical significance then the threshold is 10% which is more acceptable for clinical practice. For *phi* the thresholds were all lower (10% or less) across different scenarios (reducing MRI and/or biopsies and whether pre or post MRI) and for different definitions of clinical significance. In contrast, the PRS panel used did not seem to add much value. Free to total PSA performance was variable and added only modest incremental value especially in detection more aggressive cancers. Our data also suggest that PRS likely do not add much value in improving the detection of significant cancers, at least, in men already referred with a raised PSA. Most previous studies have either looked at individual variables or a fixed panel. Perhaps the most well-known combination of the latter is the SThM3 test which includes PSA, different kallikreins,

genomic SNPs, molecular markers as well as clinical factors (16). These are combined into an algorithmic score to predict the likelihood of cancer (defined as histological grade $\geq$ GG2). In a non-MRI population, the AUC achieved was 0.74 versus 0.5 with PSA alone, not dissimilar to some of the individual biomarkers in this present study (16). Subsequent work in MRI imaged populations have shown similar incremental value gains in AUC (17). Strom *et al* 2018 published the individual contributions of each component to the overall performance of the SThM3 test (18). Clinical variables (i.e. prostate volume, previous biopsies, and prostate volume etc.) added the most to the cumulative AUC with prostate volume producing the greatest increase. Notably individual molecular markers and genomic scores added very little to the test's final predictive ability and this is consistent with our own findings here (18).

The definition of significant cancer and its impact on biomarker performance has not been previously well explored in early detection research. The standard endpoint is usually based on histological grade ($\geq$ GG2). Yet the clinical validity of this endpoint is debatable especially as there is good evidence that many GG2 cancers (especially if detected by MRI targeting) also can have a very slow natural history (19-20). Moreover prognosis (the risk of dying of cancer) is long established to be based on not just histopathology alone but also on stage and PSA level, and is what is used in clinical decision making after a diagnosis (6,21). Indeed, in nearly all prostate cancer guidelines, management is recommended based on risk/prognostic group rather than just grade group (11,22). Hence, it is logical to consider a composite endpoint based on prognosis as the definition of significance as this directly informs clinical decision making and improved survival outcomes (i.e. the *raison d'être* for any early detection or screening programmes). In this study when CPG2 and CPG3 was used as an endpoint it was notable that the performance of liquid biomarkers as well as imaging all improved. Previous work from our unit has demonstrated that $\geq$ CPG3 disease really represents when prostate cancer treatment has a significant impact in reducing mortality (23). We therefore advocate that a detection endpoint of $\geq$ CPG3 disease is the most useful for prostate cancer detection strategies. Our modelling would further suggest that using this endpoint was particularly useful in differentiating who needs to proceed to an MRI and if needed, to subsequent biopsies.

Our study does have limitations being a retrospective series based on an already referred population. This was thus a selected population and relatively small cohort with little racial diversity. It may also not be applicable to cohorts with lower cancer prevalence. It does therefore need replication in larger studies and further testing in the context of early detection/screening. In this regard there are currently initiatives looking at new prostate cancer detection methods and we hope our data will be of use in informing their efforts (24). Our cohort of men had biopsies regardless of MRI lesion visibility

so we know the true value of biopsies without excluding men with a negative MRI. Despite this we cannot rule out occult selection bias we were not aware of as men were seen by different clinical team members. We also focused on PRS and not pathogenic genetic variants, which could potentially improve the predictive accuracy of genomics. This work is also based on the current available PRS and there are major ongoing initiatives to find improved PRS. *phi* data was also only available in a sub-cohort of men but our findings do mirror that of many other papers who have reported on its strong predictive performance and ability to refine referrals (25). We further note that other biomarkers have recently shown some value in pre-MRI prostate triage testing in the screening setting e.g. the recently reported Pro-Screen study incorporating the 4K panel (26). Our work however does underpin efforts in the European Union (amongst others) to incorporate PSAd into organised prostate testing programmes (27). The key question is how prostate volume (to derive the PSA density) could be reliably measured in a community clinic setting and before an MRI. Other research from our unit has now also addressed this with a recent study that has shown that prostate volumes (comparable to MRI) can be reliably measured using standard non-invasive surface measurements with transabdominal/transperineal ultrasound (US) (28). Sensitivity of 100% and specificity 85% were obtained in our study and US could even be done using very low-cost hand-held devices now available.

Conclusions

In summary we present data comparing different biomarkers strategies to increase discrimination of cancer detection in a PSA referred population. We find no place for PRS but identify that PSAd and *phi* especially can reduce and refined the use of MRI and biopsies. PSAd offers a low or no cost option for a reflex test to refine onward referral for imaging/biopsy. Importantly, the incremental gain using these markers however is dependent on the definition of significant cancers used as the endpoint for detection. In this regard, we support use of a composite prognostic endpoint classification that mirrors the post-diagnosis classification used for decision making. Taking all this together we advocate that the future of prostate cancer detection should adopt a tiered strategy with each test forming a gateway to the next investigation. In this scenario the burden of over-imaging and over-biopsy can be simultaneously addressed while enriching detection of prognostically important disease.

Additional File Table S1 - Additional value of individual test to improve AUC Base performance for detection of any cancer. A. Base model is pre-MRI standard of care including PSA, age. B. Base model with post-MRI standard of care including PSA, age, MRI and biopsy of LIKERT 3-5. All p value are comparisons against the base models. PRS- Polygenic risk score FTPSA- Free to total PSA PSAd- PSA density, *phi* -Prostate Health Index.

Abbreviations

PSA – prostate specific antigen

FTPSA - Free/Total PSA

phi - Prostate Health Index

PSAd PSA density

PRS- Polygenic Risk Score

GG - Grade Group

CPG - Cambridge Prognostic Group

AUC- Area under the curve

DCA - Decision Curve Analysis

MRI - magnetic resonance imaging

SNP- single nucleotide polymorphism

PI-RADS - Prostate Imaging Reporting and Data

Declarations

Acknowledgements: VJG is grateful for data collation and administration support from Ms Dhakshita Gaweshani and Ms Alexandra Keates.

Consent for publishing: All authors give consent for publishing

Funding: This work was supported by the International Alliance for Cancer Early Detection, an alliance between Cancer Research UK [C1435/A30907], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. Vincent J Gnanapragasam acknowledges infrastructure support from the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Kenneth Muir and Artitaya Lophatananon were in part supported by the NIHR Manchester Biomedical Research Centre.

Availability of Data and Materials: All data is available by application to the study authors

Author contributions: VG, AL and KM conceived the idea and conducted the analysis as well as oversaw the collection and curation of data. All authors read and approved the final manuscript.

Competing Interests: VJG has previously received a speaker honorarium from Beckman Coulter on biomarkers in prostate cancer prior to the conception of this study. The company also provided free reagents for the PHI assay, however the company was not involved in the design, conduct or reporting of this study nor were involved in drafting or review of this paper. VJG, KM and AL conceived and

published the CPG model which is freely available and attracts no fees or royalties of any kind. There are no other competing interests to declare.

Ethical Approval and consent to participate: The study was conducted under UK IRAS ethics REC 03/018. All participants consented to anonymised blood samples and clinical data being used for this study.

Authors' Twitter handle: @VincentGnanapr3 @UniversityCamb

References

1. <https://www.cancerresearchuk.org/about-cancer/prostate-cancer/symptoms> (Accessed March 2024)
2. Ilic D, Djulbegovic M, Jung JH, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ*. 2018;362:k3519.
3. Barrett T, Slough R, Sushentsev N, Shaida N, Koo BC, Caglic I, Kozlov V, Warren AY, Thankappannair V, Pinnock C, Shah N, Saeb-Parsy K, Gnanapragasam VJ, Sala E, Kastner C. Three-year experience of a dedicated prostate mpMRI pre-biopsy programme and effect on timed cancer diagnostic pathways. *Clin Radiol*. 2019;74(11):894.e1-894.e9.
4. Serag H, Banerjee S, Saeb-Parsy K, Irving S, Wright K, Stearn S, Doble A, Gnanapragasam VJ. Risk profiles of prostate cancers identified from UK primary care using national referral guidelines. *Br J Cancer*. 2012;106(3):436-9. doi: 10.1038/bjc.2011.596.
5. Bell KJ, Del Mar C, Wright G, Dickinson J, Glasziou P. Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *Int J Cancer*. 2015 Oct 1;137(7):1749-57. doi: 10.1002/ijc.29538.
6. Zelic R, Garmo H, Zugna D, Stattin P, Richiardi L, Akre O, Pettersson A. Predicting Prostate Cancer Death with Different Pretreatment Risk Stratification Tools: A Head-to-head Comparison in a Nationwide Cohort Study. *Eur Urol*. 2020;77(2):180-188.
7. Sanghera, S., Coast, J., Martin, R.M. et al. Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models. *BMC Cancer* 2018; 84: <https://doi.org/10.1186/s12885-017-3974-1>
8. Huntley C, Torr B, Sud A, Rowlands CF, Way R, Snape K, Hanson H, Swanton C, Broggio J, Lucassen A, McCartney M, Houlston RS, Hingorani AD, Jones ME, Turnbull C. Utility of polygenic risk scores in UK cancer screening: a modelling analysis. *Lancet Oncol*. 2023;24(6):658-668.
9. Farha MW, Salami SS. Biomarkers for prostate cancer detection and risk stratification. *Ther Adv Urol*. 2022;14:17562872221103988. doi: 10.1177/17562872221103988.

10. Padhani AR, Godtman RA, Schoots IG. Key learning on the promise and limitations of MRI in prostate cancer screening. *Eur Radiol.* 2024; doi: 10.1007/s00330-024-10626-6.
11. National Institute for Health and Care Excellence. Prostate cancer: diagnosis and management, NICE guideline [NG131]. 2021 update. Available from: <https://www.nice.org.uk/guidance/ng131> [accessed March 2024].
12. Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, Schumacher FR, Olama AAA, Benlloch S, Dadaev T, Brook MN, Sahimi A, Hoffmann TJ, Takahashi A, Matsuda K, Momozawa Y, Fujita M, Muir K, Lophatananon A, Wan P, Le Marchand L, Wilkens LR, Stevens VL, Gapstur SM, Carter BD, Schleutker J, Tammela TLJ, Sipeky C, Auvinen A, Giles GG, Southey MC, MacInnis RJ, Cybulski C, Wokołorczyk D, Lubiński J, Neal DE, Donovan JL, Hamdy FC, Martin RM, Nordestgaard BG, Nielsen SF, Weischer M, Bojesen SE, Røder MA, Iversen P, Batra J, Chambers S, Moya L, Horvath L, Clements JA, Tilley W, Risbridger GP, Gronberg H, Aly M, Szulkin R, Eklund M, Nordström T, Pashayan N, Dunning AM, Ghoussaini M, Travis RC, Key TJ, Riboli E, Park JY, Sellers TA, Lin HY, Albanes D, Weinstein SJ, Mucci LA, Giovannucci E, Lindstrom S, Kraft P, Hunter DJ, Penney KL, Turman C, Tangen CM, Goodman PJ, Thompson IM Jr, Hamilton RJ, Fleshner NE, Finelli A, Parent MÉ, Stanford JL, Ostrander EA, Geybels MS, Koutros S, Freeman LEB, Stampfer M, Wolk A, Håkansson N, Andriole GL, Hoover RN, Machiela MJ, Sørensen KD, Borre M, Blot WJ, Zheng W, Yeboah ED, Mensah JE, Lu YJ, Zhang HW, Feng N, Mao X, Wu Y, Zhao SC, Sun Z, Thibodeau SN, McDonnell SK, Schaid DJ, West CML, Burnet N, Barnett G, Maier C, Schnoeller T, Luedeke M, Kibel AS, Drake BF, Cussenot O, Cancel-Tassin G, Menegaux F, Truong T, Koudou YA, John EM, Grindedal EM, Maehle L, Khaw KT, Ingles SA, Stern MC, Vega A, Gómez-Caamaño A, Fachal L, Rosenstein BS, Kerns SL, Ostrer H, Teixeira MR, Paulo P, Brandão A, Watya S, Lubwama A, Bensen JT, Fontham ETH, Mohler J, Taylor JA, Kogevinas M, Llorca J, Castaño-Vinyals G, Cannon-Albright L, Teerlink CC, Huff CD, Strom SS, Multigner L, Blanchet P, Brureau L, Kaneva R, Slavov C, Mitev V, Leach RJ, Weaver B, Brenner H, Cuk K, Holleczeck B, Saum KU, Klein EA, Hsing AW, Kittles RA, Murphy AB, Logothetis CJ, Kim J, Neuhausen SL, Steele L, Ding YC, Isaacs WB, Nemesure B, Hennis AJM, Carpten J, Pandha H, Michael A, De Ruyck K, De Meerleer G, Ost P, Xu J, Razack A, Lim J, Teo SH, Newcomb LF, Lin DW, Fowke JH, Neslund-Dudas C, Rybicki BA, Gamulin M, Lessel D, Kulis T, Usmani N, Singhal S, Parliament M, Claessens F, Joniau S, Van den Broeck T, Gago-Dominguez M, Castela J, Martinez ME, Larkin S, Townsend PA, Aukim-Hastie C, Bush WS, Aldrich MC, Crawford DC, Srivastava S, Cullen JC, Petrovics G, Casey G, Roobol MJ, Jenster G, van Schaik RHN, Hu JJ, Sanderson M, Varma R, McKean-Cowdin R, Torres M, Mancuso N, Berndt SI, Van Den Eeden SK, Easton DF, Chanock SJ, Cook MB, Wiklund F, Nakagawa H, Witte JS, Eeles RA, Kote-Jarai Z, Haiman CA. Trans-ancestry genome-wide

association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nat Genet.* 2021;53(1):65-75.

13. Barrett T, Ghafoor S, Gupta RT, Kim CK, Muglia VF, Macura KJ, Purysko AS. Prostate MRI Qualification: AJR Expert Panel Narrative Review. *AJR Am J Roentgenol.* 2022 Nov;219(5):691-702.

14. Falagario UG, Sanguedolce F, Dovey Z, Carbonara U, Crocerossa F, Papastefanou G, Autorino R, Recchia M, Ninivaggi A, Busetto GM, Annese P, Carrieri G, Cormio L. Prostate cancer biomarkers: a practical review based on different clinical scenarios. *Crit Rev Clin Lab Sci.* 2022;59(5):297-308.

15. Moldovan PC, Van den Broeck T, Sylvester R, Marconi L, Bellmunt J, van den Bergh RCN, Bolla M, Briers E, Cumberbatch MG, Fossati N, Gross T, Henry AM, Joniau S, van der Kwast TH, Matveev VB, van der Poel HG, De Santis M, Schoots IG, Wiegel T, Yuan CY, Cornford P, Mottet N, Lam TB, Rouvière O. What is the negative predictive value of multiparametric magnetic resonance imaging in excluding prostate cancer at biopsy? A systematic review and meta-analysis from the European Association of Urology Prostate Cancer Guidelines Panel. *Eur Urol.* 2017;72(2):250–266.

16. Grönberg H, Adolfsson J, Aly M, Nordström T, Wiklund P, Brandberg Y, Thompson J, Wiklund F, Lindberg J, Clements M, Egevad L, Eklund M. Prostate cancer screening in men aged 50-69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol.* 2015;16(16):1667-76.

17. Nordström T, Discacciati A, Bergman M, Clements M, Aly M, Annerstedt M, Glaessgen A, Carlsson S, Jäderling F, Eklund M, Grönberg H; STHLM3 study group. Prostate cancer screening using a combination of risk-prediction, MRI, and targeted prostate biopsies (STHLM3-MRI): a prospective, population-based, randomised, open-label, non-inferiority trial. *Lancet Oncol.* 2021;22(9):1240-1249.

18. Ström P, Nordström T, Aly M, Egevad L, Grönberg H, Eklund M. The Stockholm-3 Model for Prostate Cancer Detection: Algorithm Update, Biomarker Contribution, and Reflex Test Potential. *Eur Urol.* 2018;74(2):204-210.

19. Pekala KR, Bergengren O, Eastham JA, Carlsson SV. Active surveillance should be considered for select men with Grade Group 2 prostate cancer. *BMC Urol.* 2023 ;23(1):152. doi: 10.1186/s12894-023-01314-6.

20. Vickers AJ. Effects of Magnetic Resonance Imaging Targeting on Overdiagnosis and Overtreatment of Prostate Cancer. *Eur Urol*. 2021;80(5):567-572.
21. Schaeffer EM, Srinivas S, Adra N, An Y, Barocas D, Bitting R, Bryce A, Chapin B, Cheng HH, D'Amico AV, Desai N, Dorff T, Eastham JA, Farrington TA, Gao X, Gupta S, Guzzo T, Ippolito JE, Kuettel MR, Lang JM, Lotan T, McKay RR, Morgan T, Netto G, Pow-Sang JM, Reiter R, Roach M, Robin T, Rosenfeld S, Shabsigh A, Spratt D, Teply BA, Tward J, Valicenti R, Wong JK, Shead DA, Snedeker J, Freedman-Cass DA. Prostate Cancer, Version 4.2023, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2023;21(10):1067-1096.
22. Thurtle D, Rossi SH, Berry B, Pharoah P, Gnanapragasam VJ. Models predicting survival to guide treatment decision-making in newly diagnosed primary non-metastatic prostate cancer: a systematic review. *BMJ Open*. 2019;9(6):e029149. doi: 10.1136/bmjopen-2019-029149.
23. Gnanapragasam VJ, Barrett T, Thankapannair V, Thurtle D, Rubio-Briones J, Domínguez-Escrig J, Bratt O, Statin P, Muir K, Lophatananon A. Using prognosis to guide inclusion criteria, define standardised endpoints and stratify follow-up in active surveillance for prostate cancer. *BJU Int*. 2019;124(5):758-767.
24. Bratt O, Auvinen A, Arnsrud Godtman R, et al Screening for prostate cancer: evidence, ongoing trials, policies and knowledge gaps *BMJ Oncology* 2023;2:e000039. doi: 10.1136/bmjonc-2023-00003925.
25. Agnello L, Vidali M, Giglio RV, Gambino CM, Ciaccio AM, Lo Sasso B, Ciaccio M. Prostate health index (PHI) as a reliable biomarker for prostate cancer: a systematic review and meta-analysis. *Clin Chem Lab Med*. 2022;60(8):1261-1277
26. Auvinen A, Tammela TLJ, Mirtti T, et al. Prostate Cancer Screening With PSA, Kallikrein Panel, and MRI: The ProScreen Randomized Trial. *JAMA*. Published online 2024; doi:10.1001/jama.2024.3841
27. Van Poppel H, Albrecht T, Basu P, Hogenhout R, Collen S, Roobol M. Serum PSA-based early detection of prostate cancer in Europe and globally: past, present and future. *Nat Rev Urol*. 2022;19(9):562-572. doi: 10.1038/s41585-022-00638-6.

28. Pantelidou M, Caglic I, George A, Blyuss O, Gnanapragasam VJ, Barrett T. Evaluation of transabdominal and transperineal ultrasound-derived prostate specific antigen (PSA) density and clinical utility compared to MRI prostate volumes: A feasibility study. PLoS One. 2022 ;17(9):e0274014. doi: 10.1371/journal.pone.0274014.