BMJ Open Retrospective cohort study evaluating clinical, biochemical and pharmacological prognostic factors for prostate cancer progression using primary care data

Samuel William David Merriel,1 Suzanne Marie Ingle,2 Margaret T May,2,3 Richard M Martin2,3

ABSTRACT

Objectives To confirm the association of previously reported prognostic factors with future progression of localised prostate cancer using primary care data and identify new potential prognostic factors for further assessment in prognostic model development and validation.

Design Retrospective cohort study, employing Cox proportional hazards regression controlling for age, prostate specific antigen (PSA), and Gleason score, was stratified by diagnostic stage.

Setting Primary care in England.

Participants Males with localised prostate cancer diagnosed between 01/01/1987 and 31/12/2016 within the Clinical Practice Research Datalink database, with linked data from the National Cancer Registration and Analysis Service and Office for National Statistics.

Primary and secondary outcomes Primary outcome measure was prostate cancer mortality. Secondary outcome measures were all-cause mortality and commencing systemic therapy. Up-staging after diagnosis was not used as a secondary outcome owing to significant missing data.

Results 10,901 men (mean age 74.38±9.03 years) with localised prostate cancer were followed up for a mean of 14.12 (±6.36) years. 2331 (21.38%) men underwent systemic therapy and 3450 (31.65%) died, including 1250 (11.47%) from prostate cancer. Factors associated with an increased risk of prostate cancer mortality included age; high PSA; current or ex-smoker; ischaemic heart disease; high C reactive protein; high ferritin; low haemoglobin; high blood glucose and low albumin.

Conclusions This study identified several new potential prognostic factors for prostate cancer progression, as well as confirming some known prognostic factors, in an independent primary care data set. Further research is needed to develop and validate a prognostic model for prostate cancer progression.

INTRODUCTION

Prostate cancer prognosis and treatment decisions remain a challenging clinical area for clinicians and patients, particularly for men with localised disease at the time of diagnosis. In recent decades, prostate cancer detection rates in many countries have increased markedly, in part, as a result of the rising use of asymptomatic prostate specific antigen (PSA) testing; however, more intensive PSA-based detection of prostate cancer has not been convincingly directly correlated with reductions in prostate cancer mortality for all men, implying increasing overdetection of clinically insignificant tumours. Treatments for prostate cancer carry a significant risk of morbidity for men, underlining the importance of being able to identify which men with tumours confined to the prostate at diagnosis are at higher risk of prostate cancer progression and mortality to inform discussions about management options.

Defining and measuring cancer progression with respect to treatment studies is outlined in the Response Evaluation Criteria in Solid Tumours (RECIST) criteria, which was originally published by the World Health Organisation (WHO) in 2000 and most recently updated in 2009. Evidence of tumour shrinkage on imaging and time to development of disease progression are used to measure treatment response. Definitions of cancer progression that are relevant to...
prognostic studies are less well defined, and numerous clinical, biological and surrogate markers of progression have been proposed in various studies. Prostate cancer mortality appears to be the logical ultimate endpoint of prostate cancer progression, but other measures such as development of metastases, biochemical recurrence, commencing systemic therapy and protein expression have also been reported.

There are a plethora of prognostic factor studies and prediction tools for prostate cancer risk and prognosis in the published literature. The vast majority are not externally calibrated or validated, and very few are established for use in clinical practice. Initiatives such as the MRC PROGnosis RESearch Strategy Partnership (PROGRESS) highlight the importance of high-quality prognostic research to help inform clinical practice and outline methodologically rigorous approaches to achieve this aim. Developing clinically useful risk-prediction rules starts with identifying potentially important prognostic factors, which could be incorporated into a prediction model. The aim of the current study is to confirm the association of previously reported prognostic factors with future progression of localised prostate cancer using primary care data and identify new potential prognostic factors for further assessment in prognostic model development and validation.

MATERIALS AND METHODS

The protocol for this study has been published previously in BMJ Open. In summary, we undertook a retrospective cohort study using a longitudinal data set of prospectively collected electronic primary care medical records from general practices (GPs) in England for the Clinical Practice Research Datalink (CPRD). This data set was linked with cancer registry data from the National Cancer Research and Analysis Service (NCRAS) and mortality data from the Office for National Statistics (ONS). Men were included if they had a diagnosis of prostate cancer entered into their medical record during the 20-year study period (01 January 1987—31 December 2016). Localised prostate cancer was defined as T1-2/N0/M0 based on staging data entered into the NCRAS registry, which is determined from a combination of clinical, pathological and radiological data.

Potentially relevant clinical, biochemical and pharmacological factors measured in CPRD were identified from a review of the existing published literature (See BMJ Open protocol paper for more information about the prognostic factors assessed). The primary outcome of the study was prostate cancer mortality. Secondary outcomes were all-cause mortality and commencing systemic prostate cancer therapy (a measurable proxy for progression and metastatic spread of prostate cancer). Surgery, radiotherapy and brachytherapy were classified as localised therapy, with chemotherapy, hormone treatments (primary or neoadjuvant) and immunotherapy considered systemic therapy. Mortality outcomes were based on primary/immediate cause of death reported in death certification information from the ONS and therapy outcomes from NCRAS data. In our published protocol, up-staging after diagnosis was proposed as a secondary outcome indicating spread of disease; however, this was not used in the final analysis as repeat staging was rarely recorded in the cancer registry.

Descriptive statistics were used to summarise the basic demographic details of the men and the prevalence of the preselected putative prognostic factors. Cox proportional hazards regression was used to estimate crude and mutually adjusted hazard ratios (with 95% CIs) for prostate cancer-specific and all-cause mortality according to the prognostic factors, controlling for variables currently used in clinical practice (age, PSA level, Gleason score). Regression analyses of continuous prognostic factors were standardised using hazard ratios per change in one SD. A proportional hazards test was performed to confirm modelling met regression assumptions. The analysis was also stratified by stage at diagnosis (T1/2N0M0 vs T3+and/or N1 and/or M1). Sensitivity analysis was performed, assuming all men in the overall sample with unknown tumour location had localised disease. In order to achieve 95% power and detect a difference of 0.1 in prostate cancer mortality for a binary risk factor using an alpha of 0.05, a sample of at least 6046 men with prostate cancer would be required, assuming that 10% die over a median 10-year follow-up.

RESULTS

A total of 54500 men within CPRD had a diagnosis of prostate cancer entered into their primary care medical record during the study period. Baseline participant data are shown in Table 1. Tumour–node–metastases (TNM) staging data from the linked cancer registry were available for 7646 (14.03%) of the sample population and treatment data were available for 22766 (41.77%) men. Missing TNM staging data from the cancer registry were lower for men diagnosed in more recent years: there were no TNM stage data for men diagnosed before 1993, rising to 37.2% with TNM stage data (1064/2836) in 2015. This is consistent with a recent validation study of the NCRAS prostate cancer registry that showed low levels of completeness of TNM stage and Gleason score data prior to 2010. Using the available staging and treatment data, 10901 (20%) men were identified as having localised prostate cancer at the time of diagnosis and were included in the final cohort for analysis, with a mean follow-up of 14.12 (±6.36) years. Levels of missing data for selected prognostic factors within CPRD varied.

1250 men with localised disease died of prostate cancer over the course of follow-up, giving a prostate cancer mortality rate of 8.1 per 1000 person-years. The total number of deaths for included men was 3450 (21.11 deaths per 1000 person-years). A total of 2331 (21.38%) men with localised disease received systemic therapy in the follow-up period after diagnosis. For over 90% of the
men, it was unknown whether they were reinvestigated for cancer staging after diagnosis or not (see table 2).

Raised acute phase reactants (C reactive protein (CRP) (adjusted HR per SD 1.35 95% CI 1.02 to 1.77)), ferritin (adjusted HR per SD 2.03; 95% CI 1.21 to 3.39) and random glucose (adjusted HR per SD 1.27; 95% CI 1.06 to 1.54) were associated with prostate cancer mortality. Anaemia (adjusted HR per SD 0.72; 95% CI 0.59 to 0.88) and low albumin (adjusted HR per SD 0.81; 95% CI 0.67 to 0.97) were also associated with this outcome. No medications assessed were associated with prostate cancer mortality. Current and ex-smokers (adjusted HR 1.47; 95% CI 1.05 to 2.05) and patients with a history of ischaemic heart disease (adjusted HR 1.79; 95% CI 1.20 to 2.66) had a higher risk of prostate cancer mortality over the study period.

DISCUSSION
This retrospective cohort study used primary care medical records data for men with localised prostate cancer from CPRD to confirm prognostic factors associated with prostate cancer progression. Well-known factors already incorporated into clinical guidelines, such as age and PSA, were confirmed as being individual prognostic factors. In addition, further clinical (history of smoking or ischaemic heart disease) and biochemical (anaemia or high ferritin) factors were found to be strongly associated with prostate cancer mortality. Anaemia, low albumin, raised PSA, history of ischaemic heart disease and smoking were also strongly associated with all-cause mortality, as were...
Table 3  Prognostic factors for men with localised disease associated with outcomes

<table>
<thead>
<tr>
<th>Factor</th>
<th>Mean (SD)</th>
<th>Missing (n (%))</th>
<th>Prostate cancer mortality</th>
<th>All-cause mortality</th>
<th>Systemic therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>HR per SD**</td>
<td>95% CI</td>
<td>P</td>
</tr>
<tr>
<td>Age</td>
<td>74.39 (9.03)</td>
<td>0</td>
<td>1.70</td>
<td>1.40 to 2.06</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>BMI</td>
<td>27.43 (4.48)</td>
<td>394 (3.61%)</td>
<td>1.05</td>
<td>0.90 to 1.08</td>
<td>0.52</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>1.45 (0.80)</td>
<td>3856 (35.37%)</td>
<td>0.83</td>
<td>0.64 to 1.08</td>
<td>0.16</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>1.35 (0.43)</td>
<td>3954 (36.27%)</td>
<td>1.05</td>
<td>0.89 to 1.23</td>
<td>0.56</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>2.95 (0.99)</td>
<td>4698 (43.10%)</td>
<td>0.86</td>
<td>0.69 to 1.07</td>
<td>0.18</td>
</tr>
<tr>
<td>Hb</td>
<td>144.28 (14.35)</td>
<td>2696 (24.73%)</td>
<td>0.72</td>
<td>0.59 to 0.88</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Albumin</td>
<td>41.83 (3.94)</td>
<td>2954 (27.10%)</td>
<td>0.81</td>
<td>0.67 to 0.97</td>
<td>0.02</td>
</tr>
<tr>
<td>Random glucose</td>
<td>5.70 (2.11)</td>
<td>4525 (41.51%)</td>
<td>1.27</td>
<td>1.06 to 1.54</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Median (IQR)</th>
<th>Missing (n (%))</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSA</td>
<td>8.4 (5.55, 14.60)</td>
</tr>
<tr>
<td>CRP</td>
<td>3.9 (2, 8)</td>
</tr>
<tr>
<td>Ferritin</td>
<td>108.6 (47, 196)</td>
</tr>
</tbody>
</table>

*Adjusted for age, PSA, Gleason score, TNM stage.
†Proportional Hazards assumption test not met.
BMI, body mass index; CRP, C reactive protein; Hb, haemoglobin; HbA1c, haemoglobin A1c; HDL, high density lipoprotein; LDL, low density lipoprotein; PSA, prostate specific antigen; TNM, tumour–node–metastases.
Table 4 Prognostic factors for men with localised disease associated with outcomes

<table>
<thead>
<tr>
<th>Factor</th>
<th>n (%)</th>
<th>Missing (n(%)</th>
<th>Prostate cancer mortality</th>
<th>All-cause mortality</th>
<th>Systemic therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>HR* 95% CI</td>
<td>P</td>
<td>HR* 95% CI</td>
<td>HR* 95% CI</td>
</tr>
<tr>
<td>Smoker (current/ ex-)</td>
<td>5112 (46.89%)</td>
<td>1.47 1.05 to 2.05</td>
<td>0.02</td>
<td>1.66 1.39 to 1.98</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Excess alcohol</td>
<td>1829 (16.78%)</td>
<td>0.61 0.36 to 1.04</td>
<td>0.07</td>
<td>0.91† 0.71 to 1.18</td>
<td>0.47</td>
</tr>
<tr>
<td>BPH</td>
<td>1169 (10.72%)</td>
<td>0.64 0.36 to 1.11</td>
<td>0.11</td>
<td>0.81 0.62 to 1.05</td>
<td>0.11</td>
</tr>
<tr>
<td>COPD</td>
<td>862 (7.91%)</td>
<td>0.86 0.47 to 1.57</td>
<td>0.63</td>
<td>1.64 1.29 to 2.09</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>CVA</td>
<td>553 (5.07%)</td>
<td>0.90 0.42 to 1.94</td>
<td>0.79</td>
<td>1.19 0.85 to 1.68</td>
<td>0.30</td>
</tr>
<tr>
<td>IHD</td>
<td>1548 (14.20%)</td>
<td>1.79 1.20 to 2.66</td>
<td>&lt;0.01</td>
<td>1.25 1.02 to 1.55</td>
<td>0.04</td>
</tr>
<tr>
<td>PVD</td>
<td>202 (1.85%)</td>
<td>2.24 0.98 to 5.12</td>
<td>0.06</td>
<td>1.91 1.24 to 2.95</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>T2DM</td>
<td>1508 (13.83%)</td>
<td>0.97 0.62 to 1.51</td>
<td>0.89</td>
<td>0.95 0.76 to 1.19</td>
<td>0.68</td>
</tr>
<tr>
<td>Aspirin</td>
<td>426 (3.91%)</td>
<td>1.88 0.96 to 3.70</td>
<td>0.06</td>
<td>1.58 1.09 to 2.29</td>
<td>0.02</td>
</tr>
<tr>
<td>Metformin</td>
<td>33 (0.30%)</td>
<td>2.74 0.88 to 8.49</td>
<td>0.08</td>
<td>1.73 0.77 to 3.85</td>
<td>0.18</td>
</tr>
<tr>
<td>Alpha-blockers</td>
<td>305 (2.80%)</td>
<td>1.55 0.72 to 3.35</td>
<td>0.26</td>
<td>1.15 0.76 to 1.73</td>
<td>0.52</td>
</tr>
<tr>
<td>Beta-blockers</td>
<td>265 (2.43%)</td>
<td>2.03 0.89 to 4.60</td>
<td>0.09</td>
<td>1.79 1.18 to 2.72</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Statins</td>
<td>339 (3.11%)</td>
<td>1.65 0.87 to 3.15</td>
<td>0.13</td>
<td>1.01 0.66 to 1.53</td>
<td>0.97</td>
</tr>
<tr>
<td>Vitamin D</td>
<td>465 (4.27%)</td>
<td>1.33 0.65 to 2.71</td>
<td>0.44</td>
<td>1.13 0.78 to 1.65</td>
<td>0.51</td>
</tr>
</tbody>
</table>

*Adjusted for age, PSA, Gleason score, TNM stage.
†Proportional Hazards assumption test not met.
BPH, Benign Prostatic Hypertrophy; COPD, chronic obstructive pulmonary disease; CVA, cerebrovascular accident; IHD, ischaemic heart disease; PVD, peripheral vascular disease; T2DM, type 2 diabetes mellitus.
peripheral vascular disease, chronic obstructive pulmonary disease and beta-blocker use. Smoking history was strongly associated with future systemic therapy, as were recent prescriptions of alpha-blockers or vitamin D supplements.

This analysis confirms the prognostic associations of some factors in prostate cancer progression. Smoking has also been found to be a risk factor for prostate cancer progression and mortality in cohort studies and systematic reviews. Low albumin was associated with prostate cancer mortality in the Apolipoprotein-related MOrtality RISK (AMORIS) cohort and, along with anaemia, was a more widely accepted predictor of poor cancer outcomes. The published literature around the prognostic effect of beta-blockers for prostate cancer patients has been more mixed, with this study lending weight to the evidence of increased mortality in patients with cancer. Body mass index (BMI) was not shown to be associated with prostate cancer and overall mortality in this study. While some observational studies of prostate cancer have suggested that an association may exist, reviews of trial data have demonstrated that higher BMI may actually improve the prognosis for men with cancer.

This study attempted to confirm prognostic factors in a primary care data set that could be used in a model to predict prostate cancer progression at the time of diagnosis, prior to any treatment being initiated. This approach could allow the identified prognostic factors to be used to develop a new prognostic tool to inform treatment decisions between a patient and their treating team. There are already examples of similar prognostic tools available for use, including Predict Prostate (https://prostate.predict.nhs.uk/). However, these tools have only been developed using secondary care data, which may not capture all important prognostic factors or have equivalent length of follow-up of patients in their development or calibration cohorts. In the context of on-going challenges with prognostication for men with localised prostate cancer and the increasing numbers of men being diagnosed every year, getting the most accurate information to inform treatment discussions between patients and their treating physicians is vital.

Strengths and limitations

This study has a number of unique features. This is the first study that the authors are aware of to use a primary care data set to identify and confirm prognostic factors associated with prostate cancer progression. CPRD contains all data held in the primary care records of millions of UK patients, allowing the inclusion of a range of potentially important prognostic factors. Using a primary care data set from the National Health Service (NHS) also provided long-term data for included patients, with a mean follow-up of over 14 years. Prolonged follow-up for men with prostate cancer is important as many patients can live for years before their cancer progresses. The lack of high-quality prognostic research discussed in the introduction is not limited to prostate cancer, with many other prognostic factor studies being conducted in similarly flawed ways. This study sought to take a confirmatory approach to postulated prognostic factors in prostate cancer in a rigorous manner, following the methodological recommendations of the REMARK guidelines and the PROGRESS partnership.

There are some limitations of this study that need to be considered. Previous research has shown that the prostate cancer registry in England has strong case completeness, but significant missing TNM stage and Gleason score data up until recent years. Data completeness and quality within NCRAS continues to improve, and there is no equivalent UK cancer registry data set with more complete data available at this present time. This level of missing data meant that it was unknown whether the majority of potentially included men had localised disease or not. Even so, the study was still powered to answer the research question, and sensitivity analyses showed minimal changes to almost all relationships between the prognostic factors of interest and the study outcomes. Misattribution of prostate cancer as the primary cause of death may occur in some frail, elderly patients or patients with multimorbidity, affecting the primary outcome of this study. There is evidence of misattribution of prostate cancer as a cause of death in other high-income countries; however, an English study comparing death certification to a blinded; independent panel showed that ONS data on prostate cancer mortality classification are highly accurate. This study uses a retrospective design interrogating electronic primary care records. It relies on accurate coding from GPs, and there was significant missing data for some prognostic factors.

This study took a confirmatory approach to identify which prognostic factors for prostate cancer progression may be relevant, and some new prognostic factors not currently recommended for use in clinical practice were identified. These prognostic factors could be used to generate a more robust clinical risk prediction tool to guide treatment decision-making. Developing an accurate prediction tool for prostate cancer progression, not just mortality, could be more useful for informing management discussions between patients and clinicians.

Twitter Samuel William David Merriel @sammerriel

Contributors SWDM conceived and designed the work that has led to this submission. He acquired the data and performed the analysis. He drafted the manuscript and approves the final version. He agrees to be accountable for all aspects of the work. As corresponding author, he also confirms he has full access to the data in the study and has taken final responsibility for the decision to submit for publication. SMI played an important role in the data analysis and interpretation of the results. She revised the manuscript and approved the final version. She agrees to be accountable for all aspects of the work. MTM helped design the work that has led to this submission, and supported interpretation of the results. She also
provided study supervision to SWDM. She has revised the manuscript and approved the final version. She agrees to be accountable for all aspects of the work. RMM helped to conceive and design the work that has led to this submission. He also provided study supervision to SWDM. He has revised the manuscript and approved the final version. He agrees to be accountable for all aspects of the work.

**Funding** SWDM is supported by the Can Test Collaborative, which is funded by CRUK (C8640/A23385). This work was supported by an Academic Clinical Fellowship in Primary Care for SWDM, funded by the National Institute for Health Research and Health Education England. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, Health Education England or the Department of Health. RMM was supported by a CRUK programme grant, the Integrative Cancer Epidemiology Programme (C18281/A19169). MM was supported by the NIHR Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The funders had no direct role in the planning or undertaking of this study, or the preparation of this manuscript.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** This study received ethical approval from the Independent Scientific Advisory Committee (ISAC) of the Medicines and Healthcare products Regulatory Authority (MHRA)—Protocol reference 17_041. It was conducted in accordance with the Declaration of Helsinki.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data may be obtained from a third party and are not publicly available. This study analysed a CPRD dataset, with linked NCRAS and ONS data. Permission was not sought to share the dataset publicly.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits any non-commercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. See: https://creativecommons.org/licenses/by/4.0/.

**ORCID iD**

Samuel William David Merriel http://orcid.org/0000-0003-2919-9087

**REFERENCES**


Open access