Merriel, S. W. D., Turner, E., Walsh, E., Young, G., Metcalfe, C., Hounsome, L., ... Martin, R. (2016). Validation of the National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry with data from the CAP study. In *Public Health Science: A National Conference Dedicated to New Research in UK Public Health* [S77] (The Lancet; Vol. 388, No. Suppl 2). Lancet Publishing Group. https://doi.org/10.1016/S0140-6736(16)32313-3

Elsevier Editorial System(tm) for The Lancet
Manuscript Draft

Manuscript Number: THELANCET-D-16-04732R1

Title: National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry validation utilising data from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP) study

Article Type: Conference Abstracts

Corresponding Author: Dr. Sam Merriel,

Corresponding Author's Institution: University of Bristol

First Author: Sam Merriel

Order of Authors: Sam Merriel; Emma Turner; Eleanor Walsh; Grace Young; Chris Metcalfe; Luke Hounsome; Isobel Tudge; Jenny Donovan; Freddie Hamdy; David Neal; Richard M Martin

Manuscript Region of Origin: UNITED KINGDOM

Abstract: Background
Prostate cancer is the second most common cancer in the UK, with 39,741 cases diagnosed in 2014. The NCRAS collects cancer data from across England, including prostate cancer cases. The CAP study is a cluster randomised controlled trial investigating the effectiveness and cost-effectiveness of prostate-specific antigen (PSA) testing. The value of cancer registries is maximised by receiving and processing quality data, and they can be evaluated in a number of domains, including completeness and validity. This studied aimed to compare the completeness and accuracy of the NCRAS registry and CAP study diagnosis and staging/grade data.

Methods
We compared the percentage completeness and agreement of prostate cancer diagnosis Union Internationale Contre le Cancer (UICC) Tumour-Node-Metastasis (TNM) stage and Gleason grade data in the NCRAS cancer registry with information collected via independent medical record review on 1,356 participants in CAP, a large prostate cancer screening trial. CAP study participants were matched to NCRAS registry entries using their NHS number. Agreement was assessed using Cohen's Kappa.

Findings
The mean age of the 1,356 men included in this study was 75.15 years (SD 5.09). Both the NCRAS (97.86%) registry and CAP study (98.53%) had high levels of completeness for diagnosis date. Kappa agreement was 0.90 (95% CI 0.89, 0.92) for combined Gleason score, and 0.48 (0.43, 0.53) for American Joint Committee on Cancer (AJCC) group. TNM staging agreement was 0.35 (0.31, 0.37) for T, 0.51 (0.45, 0.57) for N, and 0.58 (0.51, 0.66) for M stage overall. Agreement was moderate when considering local (T1-3N0M0) vs metastatic disease (T4NxMx / TxN1Mx / TxNxM1) (k=0.54 95% CI 0.44, 0.64).

Interpretation
The NCRAS prostate cancer registry appears to have a high level of completeness for case registration, and strong accuracy for Gleason

grade. Agreement of exact TNM staging and AJCC group appears to be low, which could be explained if staging data was collected from different sources (i.e. pathological vs. imaging staging methods) and needs to be explored further. Prostate cancer stage and grade data accuracy for the NCRAS registry need repeated evaluation to drive improvements in data quality.

National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry validation utilising data from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP) study

Dr Sam Merriel[1] MBBS, Dr Emma L Turner[1] PhD, Ms Eleanor Walsh[1] MSc, Ms Grace Young[1] MSc, Dr Chris Metcalfe[1] PhD, Dr Luke Hounsome[2]PhD, Ms Isobel Tudge[2], Prof Jenny Donovan[1] PhD, Prof Freddie Hamdy[3] MD, Prof David Neal[3] FRCS, Prof Richard M Martin[1] PhD
[1]School of Social & Community Medicine, University of Bristol, Bristol, UK
[2]National Cancer Registration and Analysis Service, Public Health England, UK
[3]Nuffield Department of Surgical Sciences, Oxford University, Oxford, UK

Corresponding author – Dr Samuel W D Merriel, School of Social & Community Medicine, University of Bristol, 39 Whatley Road, Clifton BS8 2PS

Background
Prostate cancer is the second most common cancer in the UK, with 39,741 cases diagnosed in 2014. The NCRAS collects cancer data from across England, including prostate cancer cases. The CAP study is a cluster randomised controlled trial investigating the effectiveness and cost-effectiveness of prostate-specific antigen (PSA) testing. The value of cancer registries is maximised by receiving and processing quality data, and they can be evaluated in a number of domains, including completeness and validity. This studied aimed to compare the completeness and accuracy of the NCRAS registry and CAP study diagnosis and staging/grade data.

Methods
We compared the percentage completeness and agreement of prostate cancer diagnosis Union Internationale Contre le Cancer (UICC) Tumour-Node-Metastasis (TNM) stage and Gleason grade data in the NCRAS cancer registry with information collected via independent medical record review on 1,356 participants in CAP, a large prostate cancer screening trial. CAP study participants were matched to NCRAS registry entries using their NHS number. Agreement was assessed using Cohen's Kappa.

Findings
The mean age of the 1,356 men included in this study was 75.15 years (SD 5.09). Both the NCRAS (97.86%) registry and CAP study (98.53%) had high levels of completeness for diagnosis date. Kappa agreement was 0.90 (95% CI 0.89, 0.92) for combined Gleason score, and 0.48 (0.43, 0.53) for American Joint Committee on Cancer (AJCC) group. TNM staging agreement was 0.35 (0.31, 0.37) for T, 0.51 (0.45, 0.57) for N, and 0.58 (0.51, 0.66) for M stage overall. Agreement was moderate when considering local (T1-3N0M0) vs metastatic disease (T4NxMx / TxN1Mx / TxNxM1) (k=0.54 95% CI 0.44, 0.64).

Interpretation
The NCRAS prostate cancer registry appears to have a high level of completeness for case registration, and strong accuracy for Gleason grade. Agreement of exact TNM staging and

AJCC group appears to be low, which could be explained if staging data was collected from different sources (i.e. pathological vs. imaging staging methods) and needs to be explored further. Prostate cancer stage and grade data accuracy for the NCRAS registry need repeated evaluation to drive improvements in data quality.

Word count – 343 (excluding authors)

Funding
This research project did not receive any funding. The CAP trial is funded by Cancer Research UK and the UK Department of Health (C11043/A4286, C18281/A8145, C18281/A11326 and C18281/A15064).

Contributions
ET co-ordinated CAP trial data collection, overseen by RM, FH, DN and JD. EW extracted CAP trial data for this study. LH and IT extracted NCRAS data. SM performed analysis with assistance from GY and CM. SM drafted abstract, with input from all authors.

Conflicts of interest
We declare we have no conflicts of interest

Acknowledgements
The authors would like to acknowledge Cancer Research UK, the Public Health England National Cancer Registration and Analysis Service South-West, and the CAP trial group.

National Cancer Registration and Analysis Service (NCRAS) prostate cancer registry validation utilising data from the Cluster randomised triAl of PSA testing for Prostate cancer (CAP) study

Dr Sam Merriel[1] MBBS, Dr Emma L Turner[1] PhD, Ms Eleanor Walsh[1] MSc, Ms Grace Young[1] MSc, Dr Chris Metcalfe[1] PhD, Dr Luke Hounsome[2]PhD, Ms Isobel Tudge[2], Prof Jenny Donovan[1] PhD, Prof Freddie Hamdy[3] MD, Prof David Neal[3] FRCS, Prof Richard M Martin[1] PhD
[1]School of Social & Community Medicine, University of Bristol, Bristol, UK
[2]National Cancer Registration and Analysis Service, Public Health England, UK
[3]Nuffield Department of Surgical Sciences, Oxford University, Oxford, UK

Corresponding author – Dr Samuel W D Merriel, School of Social & Community Medicine, University of Bristol, 39 Whatley Road, Clifton BS8 2PS

Background
Prostate cancer is the second most common cancer in the UK, with 39,741 cases diagnosed in 2014. The ~~National Cancer Registry Service (NCRS)~~NCRAS collects cancer data from across England, including prostate cancer cases. The ~~C~~Cluster randomised Tri~~A~~l of PSA testing for ~~P~~Prostate cancer (CAP~~)~~ study is a cluster randomised controlled trial investigating the effectiveness and cost-effectiveness of prostate-specific antigen (PSA) testing. The value of cancer registries is maximised by receiving and processing quality data, and they can be evaluated in a number of domains, including completeness and validity. This studied aimed to compare the completeness and accuracy of the NCRAS registry and CAP study diagnosis and staging/grade data.

Methods
We compared the percentage completeness and agreement of prostate cancer diagnosis Union Internationale Contre le Cancer (UICC) Tumour-Node-Metastasis (TNM) stage and Gleason grade data in the NCRAS cancer registry with information collected via independent medical record review on 1,356 participants in CAP, a large prostate cancer screening trial. CAP study participants were matched to NCRAS registry entries using their NHS number. Agreement was assessed using Cohen's Kappa.
~~Patients' data in NCRS prostate cancer registry and/or the CAP trial between Feb 2002 and Dec 2015 were merged using a unique identifier. Chi square tests and Kappa statistic calculations were performed to confirm the accuracy and completeness of NCRS prostate cancer registry data, relative to the CAP trial data.~~

Findings
~~Data were obtained from NCRS or CAP (or both) for~~The mean age of the 1,356 men included in this study was 75.15 years (SD 5.09). Both t~~T~~he NCRAS (97.86%) registry and CAP study (98.53%) had high levels of completeness for diagnosis date.~~, with 26 (1.92%) of men from NCRS missing a date of diagnosis.~~ Kappa agreement ~~on staging~~ was 0.90 (95% CI 0.89, 0.92) for combined Gleason score, and 0.48 (0.43, 0.53) for American Joint Committee on Cancer (AJCC) group. TNM staging agreement was 0.35 (0.31, 0.37) for T, 0.51 (0.45, 0.57) for N, and

0.58 (0.51, 0.66) for M stage overall. Agreement was moderate when considering local (T1-3N0M0) vs metastatic disease (T4NxMx / TxN1Mx / TxNxM1) (k=0.54 95% CI 0.44, 0.64).

Interpretation

The NCRAS prostate cancer registry appears to have a high level of completeness for case registration, and strong accuracy for Gleason grade. Agreement of exact TNM staging and AJCC group appears to be low, which could be explained if staging data was collected from different sources (i.e. pathological vs. imaging staging methods) and needs to be explored further. ~~Accurate cancer registry data is needed to inform allocation of the limited funds available for cancer treatment and prevention.~~Prostate cancer stage and grade data accuracy for the NCRAS registry need repeated evaluation to drive improvements in data quality.

Word count – 343~~297~~ (excluding authors)

Funding

This research project did not receive any funding. The CAP trial is funded by Cancer Research UK and the UK Department of Health (C11043/A4286, C18281/A8145, C18281/A11326 and C18281/A15064).

Contributions

ET co-ordinated CAP trial data collection, overseen by RM, FH, DN and JD. EW extracted CAP trial data for this study. LH and IT extracted NCRAS data. SM performed analysis with assistance from GY and CM. SM drafted abstract, with input from all authors.

Conflicts of interest

We declare we have no conflicts of interest

Acknowledgements

The authors would like to acknowledge Cancer Research UK, ~~and~~ the Public Health England ~~Knowledge Intelligence Team~~National Cancer Registration and Analysis Service South-West, and the CAP trial group.

Responses to reviewer comments

Editorial comments:
Please ensure that your revised abstract is shorter than 350 words (excluding author details and funding, contributors, and competing interest statements).
The abstract word count is 343
Please indicate in your revised Word document whether you are an early career researcher.
The covering letter has been amended to confirm the corresponding author is an early career researcher
Please finalise authorship of your abstract before submitting your revised version.
Authorship has been finalised, and all authors have signed the author statement form
Please state study aim in background
The aim of the study has been added to the background section

Reviewer #1
This abstract is interesting in as far as it goes in that the validity of cancer registries is an important area for public health.
Thank you for your comment
The approach here seem sensible, looking agreement between the bespoke trial data and registry data using kappa test, with appropriate presentation of statistical uncertainty around these estimates.
Thank you for your comment
Reassuring levels of completeness are reported for the registry data. The issue highlighted here is around the relative lack of agreement between the two dated sources around staging. I find it hard to believe that this is anything other than a reflection of the fact that the two data sources are likely to be looking at two different things. I think it highly probable that the two data sources report at different times, using slightly different methods and what this difference is telling us is not so much about the validity of the cancer registry but about the complexity and consequent variation around staging cancers.
Thank you for your comment. Unfortunately this study was not designed to determine why differences between the data sources may exist. This would indeed be an interesting further exploration.
Nevertheless I recommend that the abstract be accepted in its current state as this is highlighting an important potential problem that needs further more detailed investigation.
Thank you for your comment

Reviewer #2
This appears to be a development / proof of principle piece of work validating national cancer registry data against data from a prostate cancer cluster randomised trial. Overall, the abstract is well written and clear, with the interpretation clearly supported by the findings.
Thank you for your comment
What this abstract lacks is a clear statement of objective / hypothesis. It would also benefit the abstract to clearly explain why it is helpful to validate cancer registry data in this way.
The aim of the study has been added to the background section

Reviewer #3:
Prostate cancer is the second most common cancer in England and has the fifth highest burden of all cancers when measured in DALYs. England also has a world leading cancer registry. This study looked a validating registry data against those collected from a

randomized controlled trial. The authors should attempt to address the following:
The aim of this study should be explicitly stated as it's currently unclear.
The aim of the study has been added to the background section
What unique identifier was used to match across the trial and NCRS datasets, and were all records matched?
Participants were matched using their NHS numbers. Not all participants in the CAP study had complete NCRAS registry data.
Were all CAP participants found in the registry as it seems like this was the way the dataset were combined despite wording in abstract (see later point)?
See comment above
Completeness of data in the registry was mentioned, but not the trial dataset.
Completeness for CAP study data has been added to the Findings section
TNM should be defined at first use.
TNM has now been defined in its first in the Methods section
The statement 'or both' in the findings: "Data were obtained from NCRS or CAP (or both)" is confusing. Is it not the case that NCRS data were compared to CAP by only including those individuals in NCRS that also took part in the CAP trial? If not, please clarify how missing data/ participants were dealt with.
Apologies for the confusing wording. Further information about participant selection has been added to the Methods section.
Were there any other missing data in NCRS other than date of diagnosis?
Yes. There was missing data for stage and grade in both the NCRAS registry and CAP study. Unfortunately, we are unable to present all of our findings in the abstract owing to the word limit.
The concluding sentence "Accurate cancer registry data is needed to inform allocation of the limited funds available for cancer treatment and prevention." is rather general. It would be more helpful to propose a way forward for improving NCRS based on the findings of the study.
Thank you. The Interpretation section has been amended to better reflect on the study findings.