Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200446

Primary Care Prostate Cancer Case Ascertainment

Agnieszka LEMANSKA[a,b,1], Sara FAITHFULL[b], Harshana LIYANAGE[d], Sophie OTTER[c], Marina ROMANCHIKOVA[a], Julian SHERLOCK[d], Nadia A. S. SMITH[a], Spencer A. THOMAS[a] and Simon de LUSIGNAN[d]

[a]*Data Science Department, National Physical Laboratory, Teddington, UK*
[b]*School of Health Sciences, University of Surrey, Guildford, UK*
[c]*Royal Surrey NHS Foundation Trust, Guildford, UK*
[d]*Nuffield Department of Primary Care Health Sciences, University of Oxford, UK*

Abstract: Although routine healthcare data are not collected for research, they are increasingly used in epidemiology and are key real-world evidence for improving healthcare. This study presents a method to identify prostate cancer cases from a large English primary care database. 19,619 (1.3%) men had a code for prostate cancer diagnosis. Codes for medium and high Gleason grading enabled identification of additional 94 (0.5%) cases. Many studies do not report codes used to identify patients, and if published, the lists of codes differ from study to study. This can lead to poor research reproducibility and hinder validation. This work demonstrates that carefully developed comprehensive lists of clinical codes can be used to identify prostate cancer; and that approaches that do not solely rely on clinical codes such as ontologies or data linkage should also be considered.

Keywords. Prostate cancer, primary care, case identification, real-world evidence

1. Introduction

Prostate cancer is a chronic condition, with good prognosis but long-term consequences. Health services worldwide are stretched so it is important to understand the needs of the growing population of people affected by cancer. The increasing use of electronic health records (EHRs) can facilitate epidemiological research. However, the usability of data is limited by the lack of standards and variability in the quality of recording of cancer diagnosis and treatment. In this study, we investigate the usability of clinical codes to identify and describe a cohort of prostate cancer patients from an English database.

2. Methods

A primary care-based, retrospective cohort study using a nationally representative, English database from the Royal College of General Practitioners (RCGP), Research and Surveillance Centre (RSC) was used [1]. A December 2016 data extract was analyzed comprising 164 practices and 1,529,922 EHRs for male patients.

[1]Corresponding Author, E-mail: agnieszka.lemanska@npl.co.uk

Prostate cancer diagnosis was extracted using commonly reported diagnostic codes. This included B46 (*malignant neoplasm of prostate*), B834 (*carcinoma in situ of prostate*) and HNG0200 (*cancer of the prostate*). We also included 4M01 and Xalmc (*Gleason prostate grade 5-7 (medium)*), and 4M02 and Xalmd (*Gleason prostate grade 8-10 (high)*). Date and age at diagnosis was based on the first occurrence of the clinical code. Ethnicity was extracted in five categories using an ontology-based algorithm [2].

3. Results

The average age at diagnosis was 70.8 (SD = 9.5) and the majority of men (11,586, 94.1%) were white (Table 1). Men identified with prostate cancer were spread evenly across the country (Figure 1). The most commonly extracted treatment was hormone therapy, 9324 (47.5%) men received it. The mean age at death was 80.6 (SD = 9.1).

Diagnosis was established for 19,619 (1.3%) men. The three codes: B46, B834 and HNG0200 were used for 18,356 (93.6%), 1,549 (7.9%) and 666 (3.4%) men respectively. Medium Gleason grading was recorded in 1535 EHRs and high in 562 EHRs. This enabled identification of 94 (0.5%) additional cases.

Characteristic	Mean (SD)	*n* (%)
Age at diagnosis (years)	70.8 (9.5)	
Ethnicity		
White		11586 (59.1)
Asian		195 (1.0)
Black		432 (2.2)
Mixed		67 (0.3)
Other		37 (0.2)
Missing		7302 (37.2)
Treatment		
Radiotherapy		634 (3.2)
Hormone therapy		9324 (47.5)
RT and HT		1970 (10.0)
None recorded		7691 (39.2)
Age at death (years)	80.6 (9.1)	
Not recorded		18170 (92.6)

Table 1. Demographic and treatment characteristics

Figure 1. Location of study participants.

4. Discussion and conclusion

Primary care-based epidemiological studies often identify cancer cases using diagnostic codes. In clinical practice, we found variations in the recording of cancer diagnosis. This work contributes to establishing standards for primary care cancer case ascertainment.

References

[1] A. Correa, *et al.*, Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile, *BMJ Open* **6** (2016).

[2] Z. Tippu, *et al.*, Ethnicity Recording in Primary Care Computerised Medical Record Systems: An Ontological Approach, *J Innov Health Inform* **23** (2017), 920.