

https://dspace.org/

A retrospective analysis of laboratory data to identify clinical practice trends in prostate cancer screening and diagnosis by level of care in the Gauteng Province between 2006 and 2016

Cassim, Naseem

2021
https://hdl.handle.net/10539/33567

Downloaded from a DSpace repository. https://dspace.org/

Abstract

Background:

Prostate cancer (PCa) is the leading male neoplasm in South Africa, with an age-standardised incidence rate (ASIR) of 68.0 per 100,000 population reported in 2018. PCa is also the most diagnosed neoplasm among men globally. Local studies have reported that Black African men present with higher grade and stage disease, higher serum PSA and less often receive potentially curative treatment than men of other race groups. Updated local guidelines recommend informed patient-based screening for males with a life expectancy ≥10 years, commencing at 40 years for Black Africans. PCa risk categories are defined using the total prostate specific antigen (PSA), Gleason score (GS) and clinical stage. The GS is the strongest predictive factor for treatment. The updated grade group (GG) was developed based on the GS as follows: (i) GS ≥6: GG1, (ii) GS 3 + 4 = 7: GG2, (iii) GS 4 + 3 = 7: GG3, (iv) GS = 8: GG4 and (v) GS ≥9: GG5. One of the challenges is the tedious manual extraction of an adenocarcinoma histological finding and the GS as this information is embedded within the semi-structured narrative prostate biopsy report. There is a paucity of local data, with most studies reporting data in a urological setting with small sample sizes, based at one or more academic hospitals. Data mining has the potential to turn narrative reports into information by applying various computational techniques.

Aims:

The aim of this thesis was to describe PSA testing and histological diagnosis of PCa using laboratory data for men attending public-health facilities in the Gauteng Province. A further aim was to use text mining to extract and describe GS, specifically for Black Africans. A further aim was to assess trends in PSA testing for primary health care facilities. Automated methods to extract PCa information to reduce the burden of manual coding were a further aim.

Methods:

A retrospective descriptive study design was used to analyse prostate biopsy and PSA laboratory data between 2006 and 2016. The inclusion criteria were men ≥30 years. The Systematized Nomenclature of Medicine (SNOMED) clinical terms (CT) morphology (M) and topography (T) codes were used to develop lookup tables to assign the following histological findings: (i) diagnosis (benign/malignant) (ii) sub-diagnosis (adenocarcinoma) and (iii) sub-result (inflammation type). Two experts manually coded the diagnosis after reading the narrative prostate biopsy report to assess the positive predictive value (PPV) of the pathologist assigned SNOMED CT codes. For 1000 randomly selected prostate biopsies with PCa, predictive analytics and text mining was used to automate the extraction of the GS, reporting precision, recall and the F score. For the prostate biopsy data, data was reported for only prostatic biopsies. M or T codes were done manually. The GS was manually coded for an adenocarcinoma sub-diagnosis. PCa was defined as an adenocarcinoma histological finding

with the GS reported. We reported associations of PCa with a GG ≥4 for race group. Both the biopsy and PSA were de-duplicated using the CDW unique patient identifier to report: (i) ASIR and (ii) develop a presentation cohort (first-ever PSA). The PSA data was analysed to report patient numbers by calendar year, age category and race group as well as descriptive statistics. We used logistic regression to assess any association for race group and age with a PSA ≥4µg/L.

Results:

SNOMED CT lookup tables were able to report the histological findings for 88% of biopsies. The manual coding revealed a PPV of 0.96 for the pathologist assigned SNOMED CT codes. The predictive analytics and text mining accurately extracted the GS for all 1000 biopsies (F score of 1). There were 22 937 prostatic biopsies referred to NHLS between 2006 and 2016. Of these, a PCa finding was reported for 39% of Black Africans. A high-risk GS was reported for 46% Black Africans. Multiple logistic regression revealed that Black Africans were more likely to have PCa with a GG ≥4 with an odds ratio of 1.45. The ASIR increased from 44.9 in 2006 to 57.3 per 100 000 population by 2016. There were239 506 (86.2%) patients with a first-ever PSA, compared to 277 983 tests. Between 2006 and 2016, numbers of men tested increased from 1 782 to 67 025 respectively, with 186 984/239 506 (78.1%) of tests from clinics. The majority of testing was for men in the 50 - 59 age category and Black Africans. The logistic regression reported that the odds of having a PSA ≥4µg/L was significantly lower for Indian/Asians, Coloureds and Whites than for Black Africans.

Conclusions:

The feasibility of using SNOMED CT codes to automate PCa data was confirmed. In addition, the reliable extraction of the GS to assess late presentation and prognosis was confirmed. These approaches can be applied to national PCa data. Our findings reveal that Black African men are significantly more likely to present with a PCa with a GS ≥8 compared with other racial groups. Our data suggest that predominantly healthy patients were tested which in turn is indicative of some population-based screening. Local public-sector guidelines need to be aligned to the 2017 urological evidence-based recommendations. There is a need for additional research to understand why Black African men present with higher grade disease.