Establishing a Prostate Cancer Data Core in VHA: Early Results

Jeremy B. Shelton M.D., M.S.H.S[1]*, Julie Lynch, PhD, RN, MBA[2] *, Patrick Alba, MS[2], Tori Anglin-Foot MHA[2], Olga Efimova, MD, PhD[2], Olga Patterson, PhD[2], Kung Min Lee, PhD[2], Jeff Ferraro, PhD[2], Brian Robison, MPH[2], Scott DuVall, PhD[2]
[1]GLA VAMC, UCLA , [2]VA Information and Computing Infrastructure (VINCI)
*Co-leads

Background:
Prostate cancer is the most common solid tumor diagnosis among veterans, accounting for 31.8% of all cancer diagnoses at VHA in 2012.[4] 11.8% of men with new prostate cancer diagnoses in VHA have stage III and IV disease. Veterans were nearly twice as likely to be diagnosed with stage IV prostate cancer (7%) than men in SEER (4%). While there has been a sharp increase in the diagnostic and treatment options for men with advanced prostate cancer, enabling the era of precision oncology, the rate of adoption of novel diagnostics and therapeutics, and the requisite changes in the systems of delivery have been slow. The Prostate Cancer Foundation (PCF) and VHA developed a novel partnership to drive rapid implementation of precision prostate cancer care. To support this effort, as well as other VHA operational and research goals, VHA-PCF embarked on an effort to develop a curated prostate cancer database with an initial goal of identifying all veterans in VHA with metastatic disease.

Methods:
A prostate cancer data core was established in Salt Lake City at the VA Informatics and Computing Infrastructure (VINCI) and in partnership with clinicians at GLA, Puget Sound and Durham VAMCs. The VA Corporate Data Warehouse and the VA Central Cancer Registry (VACCR) were queried and an algorithm using available structed data was developed, including coding, laboratory and pharmacy data. Accuracy was determined by iterative chart abstraction. Natural Language Processing (NLP) was also employed and compared to VACCR data. NLP algorithms were validated with iterative and formal chart abstraction.

Results:
The estimated total number of living Veterans in VA with prostate cancer as of 2019 using available structured data vs. NLP was 353,660 vs. 488,984. Estimates of living veterans with metastatic prostate cancer (mPC) as determined by available structured data varied from 417-10,496 and was 16,282 by NLP. 59.9% of men with mPC were White and 27% were African-American, however the proportion of white men with mPC was nearly double that of African-American men (5% vs 3%). The mean age of men with mPC is 66.9, while 62.7% (10,221) live in an urban setting and 28.5% (4643) live in rural settings. The NLP algorithm for identifying men with mPC had a PPV of 0.975 with a sensitivity of 0.828.

Conclusions:
A multispecialty prostate cancer data care was successful established. Some aspects of prostate cancer care, such as metastatic status are not readily discernable with existing structured data, even with VHA's advanced national databases, however focused application of existing technology, such as NLP, appears to have the potential to fill in gaps in structured data to create a near-real time, prostate cancer database.

No conflicts of interest to report.