

# Mining Electronic Health Records to Extract Patient-Centered Outcomes Following Prostate Cancer Treatment

**Tina Hernandez-Boussard, PhD, Panagiotis D. Kourdis, PhD, Tina Seto, MS, Michelle Ferrari, RN, Douglas W. Blayney, MD, Daniel Rubin, MD, James D. Brooks, MD**  
Stanford University, School of Medicine, Stanford, CA

## Abstract

*The clinical, granular data in electronic health record (EHR) systems provide opportunities to improve patient care using informatics retrieval methods. However, it is well known that many methodological obstacles exist in accessing data within EHRs. In particular, clinical notes routinely stored in EHR are composed from narrative, highly unstructured and heterogeneous biomedical text. This inherent complexity hinders the ability to perform automated large-scale medical knowledge extraction tasks without the use of computational linguistics methods. The aim of this work was to develop and validate a Natural Language Processing (NLP) pipeline to detect important patient-centered outcomes (PCOs) as interpreted and documented by clinicians in their dictated notes for male patients receiving treatment for localized prostate cancer at an academic medical center.*

## Introduction

Prostate cancer is the most common malignancy in men with an estimated 21% prevalence among new cancer cases in males for 2016.(1) Given the excellent survival rates, patients undergoing clinical treatment for prostate cancer often focus on patient-centered outcomes of care to guide treatment choices, such as rates of urinary incontinence (UI) or irritative voiding symptoms and erectile/sexual dysfunction (ED).(2, 3) However, the measurement of these outcomes at a population level has been hindered due to the lack of availability of these outcomes in generalizable, large-scale study cohorts.

Under the US healthcare reform, there is increased focus on quality of health care delivery.(4) Increasingly, standardized quality metrics have been developed and proposed to measure key components of care across the full continuum of care delivery, including patient-centered outcomes.(5) Benchmarking and reporting such quality metrics from multiple health care providers offers a unique opportunity to improve clinical practice and ensure that patients are receiving high-quality care and treatment options that correspond to their personal values.

Granular clinical information important for quality assessment is routinely collected within electronic health records (EHRs).(6) Due to their significant recent adoption,(7) healthcare workers have identified remarkable benefits and significant challenges in using these data to approach a learning health-care system.(8) Particularly challenging is the fact that most information in EHRs is stored as unstructured free text.(9) However, many quality measures, including patient-centered outcomes, are captured in EHRs *only* as free text.(10)

We have designed and developed the infrastructure that can leverage routinely collected information from EHRs to efficiently and accurately assess clinicians' documentation of important patient-centered outcomes following treatment for prostate cancer. Using urinary incontinence and erectile dysfunction as the example, we developed a Natural Language Processing (NLP) pipeline using the Java-based open source software GATE (General Architecture for Text Engineering) to parse strings containing clinical notes stored in the EHRs. This work incorporated electronic *phenotypes* for each PCOs that can improve its generalizability across systems.

## Methods

### Data Source

Our pipeline identified patients within a large academic EHR-system using ICD-9/10 and CPT codes. The healthcare system provides inpatient, outpatient and primary care services and has a fully functional Epic system installed since 2008. (Epic Systems Corporation, Verona Wisconsin) To improve cohort identification and validation, the EHR records were linked to the state California Cancer Registry, which includes detailed information on patients' histology, pathology, disease progression and survival, as well as treatments received outside of our academic health care system. An Oracle relational database was internally deployed to organize the internal and external structured and unstructured data elements. The patient cohort includes demographics, healthcare encounters, diagnoses/problem lists, clinical reports (narratives & impressions), encounter notes/documents, lab &

diagnostics results, medications (down to the ingredient level), treatment plans, procedures, billing summaries, patient history, patient surveys, and follow-up/ survival information. To develop vocabularies related to urinary incontinence, irritative voiding symptoms and erectile dysfunction, a minimum of 100 charts selected at random were reviewed and terminology extracted. We performed manual chart review to estimate the positive and negative predicted values of the workflow. We will randomly select 200 records. A urology research nurse manually validated these reports to create a gold standard.

#### Patient-Centered Outcome Phenotypes

Phenotypic algorithms were developed for identifying and extracting both cases and controls of UI and ED assessment from EHRs. Input categories include ICD-9/ICD-10 codes, billing codes, medications and vocabularies matched with existing ontologies from the National Center for Biomedical Ontology (NCBO) and Unified Medical Language System (UMLS) concepts.(11) To improve accuracy, vocabularies were also manually curated with EHR terms found during manual chart review. The phenotypes of each of the three PCOs are available through the publicly accessible repository PheKB (<https://phekb.org>) a knowledgebase for EHR-based phenotypes.(12)

#### NLP Extraction of PCOs

Our NLP pipeline analyzes the clinical narrative text of EHRs using GATE software.(13) GATE provides several customizable processing resources that perform specific NLP processing tasks, i.e. tokenizers, sentence splitters, gazetteers which annotate documents based on look-up lists of keywords, parsers etc. Below, we describe the building components of our GATE-based NLP application for extracting PCOs from clinical notes, which comprises (1) an ANNIE module to detect PCO mentions in narrative texts, (2) a ConText module to determine the semantic context of the PCO mentions, and (3) A JAPE module to annotate the PCO mentions in the text.

We first used the English tokeniser from GATE's ANNIE plugin component that splits the text into simple constituent tokens such as numbers, punctuation and words of different types. Next, sentence splitting was performed using the RegEx sentence splitter, which is suitable when faced with irregular inputs, such as those regularly encountered in the contents of EHR clinical notes.

Following tokenization and sentence splitting, a Hash Gazetteer (part of the ANNIE) was run on the clinical note to find occurrences of PCO mentions. We developed three types of Gazetteer list files containing the various PCO-related keywords in text that indicate PCO mentions. In addition to detecting the PCO mentions, our pipeline detects the context of the PCO mention using the ConText algorithm.(14) The ConText algorithm determined whether the PCO mentioned in the clinical report was negated, hypothetical, historical, or experienced by someone other than the patient. PCO mentions were Type I if the ConText algorithm did not negate the term but the text indicated a negated PCO. PCO mentions were Type II if the ConText algorithm did not negate the term, but the text indicated an “affirmed” PCO. Finally, PCO mentions were Type III if the ConText algorithm negated them and the text indicated a negated PCO. (Table 1)

PCOs were classified as follows: “Affirmed”, meaning the patient had the symptom; “Negated”, the patient did not have the symptom; or “Discussed Risk”, the clinician documented the discussion regarding risks of PCOs by treatment with the patient.

**Table 1.** Example of sentences in the clinical note assigned to different types of Gazetteer List files.

Type	Description	Example sentences	Annotation Value	ConText Negated
I	No urinary incontinence	His urinary control is good.	Negative	No
II	Urinary incontinence	He reports urinary leakage.	Affirmed	No
III	No urinary incontinence	No urinary complaints.	Negative	Yes

A gazetteer list file was also used to include keywords related to discussion of post-operative risk. When PCO terms appear with these terms in a sentence, the corresponding sentence is considered as a discussion of post-operative risk with the patient, rather than a PCO per se. Additionally, a Gazetteer list is used to exclude terms that are used in an alternative context (e.g. the term leakage related to urinary incontinence vs. leakage around the foley catheter). Finally, the regular expression pattern-matching engine called Java Annotation Patterns Engine (JAPE) was used to

create annotations in the text of the clinical notes based the patterns in the text detected by the Gazetteer files and ConText algorithm.

### Evaluation

The gold standard was used to evaluate our pipeline. The 200 records in the gold standard were processed and recorded all mentions of PCOs and classified each according to the annotation value. These data were compared with the annotations in the gold standard and precision and recall metrics were calculated. The university's Internal Review Board approved the study

### **Results**

We identified 7,109 male individuals who received treatment for prostate cancer from 2008 to 2016 from a single, large, academic medical center. Patients' demographics are displayed in Table 2. Patients had a mean age at time of diagnosis of 65.1 (SD 8.9). The majority of patients had localized disease, with 65% Stage II and 83% with a Gleason  $\leq 7$ .

**Table 2.** Patient Demographics for Prostate Cancer Patients in the EHR Database who received treatment between 2008-2016

<b>Variable</b>	<b>Value</b>
N	7,109
Age, Mean (SD)	65.19 (0.11)
Ethnicity, %	
White	69.68
Hispanic	2.15
Black	4.03
Asian	9.95
Other	14.18
Year of Diagnosis, %	
2005	9.20
2006	11.11
2007	11.45
2008	10.92
2009	10.23
2010	9.44
2011	9.36
2012	7.98
2013	6.96
2014	7.05
2015	6.30
BMI, mean (SD)	18.87 (0.16)
Stage, %	
0	0.44
1	10.82
2	64.65
3	11.76
4	8.55
Unknown	3.77
Gleason, %	
5	0.01
6	6.99
7	6.89
8	1.70
9	1.11
10	1.83

Unknown	83.15
Charlson Score, mean (SD)	3.31 (0.03)

Table 3 demonstrates the number of PCOs extracted from the different types of clinical notes, either history & physical or progress. The majority of PCO information extracted from the NLP pipeline came from the progress notes for both urinary incontinence and erectile dysfunction as compared to the information found in the history & physical note set.

**Table 3.** Number Of Notes with an Identified PCO, Stratified By Type Of Mention And Note Location

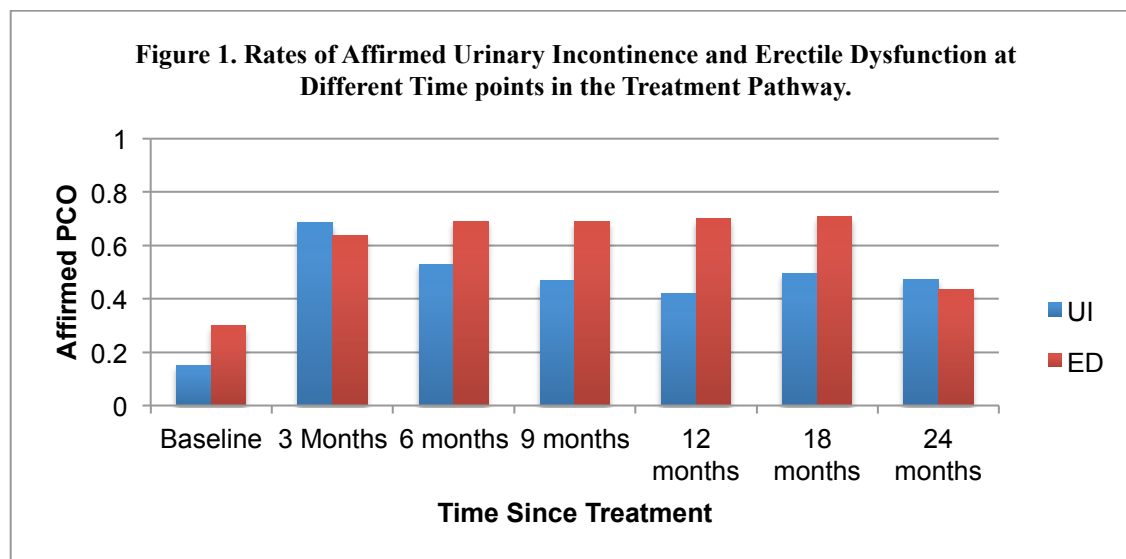
Note type	Urinary Incontinence	Erectile/Sexual Dysfunction
<b>History &amp; Physical (n = 10093)</b>		
Affirmed	362	570
Affirmed-History	75	37
Negated	1019	758
Negated-History	65	8
Discuss Risk	1380	1267
<b>Progress (n =155274)</b>		
Affirmed	7183	6359
Affirmed-History	1156	0
Negated	6734	5770
Negated -History	1122	293
Discuss Risk	3587	3410

The patient-level assessments for the two PCOs in the EHR notes for urinary incontinence and erectile dysfunction within our study cohort are displayed in Table 4. The results indicate that ED was assessed in 77.5% of the patients while UI in 59.4% prior to the primary treatment and 29% and 33% immediately following treatment.

**Table 4.** Rates Of PCO Assessment Pre- And Post-Treatment, Stratified By Affirmed And Negated Mentions

Time Period	PCO	Affirmed	Negated	Assessment Rate (%)
<b>Pre-Treatment</b>	UI	289	1517	59.4
	ED	1409	947	77.5
<b>0-3 Months Post-Treatment</b>	UI	581	450	32.5
	ED	690	218	28.6

We estimated that patients had a pre-treatment rate of UI to be 16%, with varying severity and pre-treatment rate of 30% for ED (Figure 1). The rates for UI remained at approximately 40-50% 1-2 years after surgery and the rates for ED were between 45-70% following treatment.



The F-measure accuracy scores for the UI annotations against an unseen gold standard manually annotated by our board-certified Urology nurse was 87% affirmed, 96% negated, and 91% discuss risk. For the ED annotations the F-measure was 85%, 92%, and 90%, respectively. (Table 5)

**Table 5.** Performance of NLP extraction application for PCOs using GATE Developer's corpus quality assurance visual resource

Patient-Centered Outcome	Affirmed	Negated	Discuss Risk
<b>Urinary Incontinence</b>			
Precision	0.8667	0.9444	0.9167
Recall	0.8667	0.9714	0.9016
F1-score	0.8667	0.9577	0.9091
<b>Erectile/Sexual Dysfunction</b>			
Precision	0.8567	0.9312	0.9150
Recall	0.8412	0.9011	0.8912
F1-score	0.8489	0.9159	0.9029

## Discussion

Under healthcare reform, stakeholders emphasize the importance of quality and patient-centered healthcare. Accordingly, government officials seek quality metrics that include meaningful outcomes, including those that go beyond mortality and recurrence. However, efforts to establish guidelines for prostate cancer treatment have been difficult to formulate due to insufficient evidence regarding relative benefits and risks of the different treatment options, particularly important patient-centered outcomes such as urinary and erectile dysfunction.(2, 15) Here we found clinicians regularly recorded this information in the unstructured text of EHRs. The current work expands on existing models by focusing on ontology-based dictionaries to annotate free text associated with patient-centered outcomes. We have demonstrated our ability to efficiently and accurately use our methods on EHRs for quality assessment of patient-centered outcomes.

The rates of urinary incontinence and erectile dysfunction identified through our pipeline were highly concordant with rates from surveys that are reported in the literature.(16-18) This highlights the ability of NLP algorithms to leverage routinely collected information from EHRs and efficiently and accurately assess clinicians' documentation

of patient-centered outcomes. Our study suggests that patient-centered outcomes, particularly for prostate cancer patients, are important and monitored not only for the patient but also for the clinician. Given the lack of population-based information on these important outcomes, NLP tools can clearly advance research and evidence in this area.

Our NLP pipeline shows promising results for the task of advancing patient-centered outcomes research (PCOR). To date, most quality metric analyses have used administrative datasets.(16-18) While these datasets are easy to use and readily available, they are generated for billing purposes and lack important clinical details.(19-23) Paper records may have more detail, but cannot be algorithmically assessed.(24-26) Prospective studies are not readily available and often contain ascertainment bias.(27) These issues have limited the efficiency and effectiveness of efforts for quality improvement, organizational learning, and comparative effectiveness research, particularly in the area of PCOR. Using our NLP pipeline, we can obtain population rates on PCOs efficiently and with high certainty. Such work can significantly advance the field and depth of PCOR.

There are limitations to our study. First, our algorithms have been developed and tested in a single academic center. However, the clinical terms used in our algorithms are disseminated with a national repository (pheKB.org) and multiple clinicians across different healthcare settings have vetted the clinical terms. We will be testing our algorithms in another healthcare system to ensure their generalizability. A second limitation is that our system only reports what the clinician documents and does not capture patient-reported outcomes. Many studies have highlighted that clinicians' documentation can be sparse and incomplete. However, our previous work indicated that these patient-centered outcomes are more prevalent in the clinical text than elsewhere in the EHRs. Finally, our algorithms have been developed using American English and vocabularies and language rules would need to be developed for usage where another language is used in the EHR.

A significant benefit of our NLP pipeline is that it leverages multiple sources of data to identify patients and outcomes, including registry data, quality of life surveys, and other information from clinical trials within our institute. Our dataset is updated regularly and it will likely continue to perform well, including the assessment of these outcomes immediately after the introduction of new treatments and technologies.

## Conclusions

We developed an NLP pipeline for detecting clinical mentions of patient-centered outcomes in prostate cancer patients. The current performance of the system appears sufficient to be used for population-based health management and to enhance evidence needed to help patients identify treatment pathways that reflect their healthcare values. Given the importance of these events under the healthcare reform, wide deployment of fully computerized algorithms that can reliably capture PCOs will have numerous applications for the healthcare industry. These approaches are the basis of a learning healthcare system and target in fostering healthcare quality through information technology and data resource utilization.

## References

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA: a cancer journal for clinicians*. 2014;64(1):9-29.
2. Thompson I, Thrasher JB, Aus G, Burnett AL, Canby-Hagino ED, Cookson MS, et al. Guideline for the management of clinically localized prostate cancer: 2007 update. *The Journal of urology*. 2007;177(6):2106-31.
3. Makarov DV, Yu JB, Desai RA, Penson DF, Gross CP. The association between diffusion of the surgical robot and radical prostatectomy rates. *Medical care*. 2011;49(4):333-9.
4. Jha A, Pronovost P. Toward a Safer Health Care System: The Critical Need to Improve Measurement. *JAMA*. 2016;315(17):1831-2.
5. Stelfox HT, Straus SE. Measuring quality of care: considering measurement frameworks and needs assessment to guide quality indicator development. *Journal of clinical epidemiology*. 2013;66(12):1320-7.
6. Tamang S, Hernandez-Boussard T, Ross E, Patel M, Gaskin G, Shah N. Enhanced Quality Measurement Event Detection: an Application to Physician Reporting. *EGEMS*. 2017.
7. Adler-Milstein J, DesRoches CM, Furukawa MF, Worzala C, Charles D, Kralovec P, et al. More than half of US hospitals have at least a basic EHR, but stage 2 criteria remain challenging for most. *Health Aff (Millwood)*. 2014;33(9):1664-71.
8. Coorevits P, Sundgren M, Klein GO, Bahr A, Claerhout B, Daniel C, et al. Electronic health records: new opportunities for clinical research. *J Intern Med*. 2013;274(6):547-60.

9. Grimes S. Unstructured data and the 80 percent rule. Washington: Clarabridge; 2008 [Available from: <http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/>].
10. Hernandez-Boussard T, Tamang S, Blayney D, Brooks J, Shah N. New Paradigms for Patient-Centered Outcomes Research in Electronic Medical Records: An Example of Detecting Urinary Incontinence Following Prostatectomy. *EGEMS (Wash DC)*. 2016;4(3):1231.
11. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, et al. The National Center for Biomedical Ontology. *Journal of the American Medical Informatics Association : JAMIA*. 2012;19(2):190-5.
12. Kirby JC, Speltz P, Rasmussen LV, Basford M, Gottesman O, Peissig PL, et al. PheKB: a catalog and workflow for creating electronic phenotype algorithms for transportability. *Journal of the American Medical Informatics Association : JAMIA*. 2016;23(6):1046-52.
13. Cunningham H, Maynard D, Bontcheva K, Tablan V, editors. A framework and graphical development environment for robust NLP tools and applications. *ACL*; 2002.
14. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of biomedical informatics*. 2009;42(5):839-51.
15. Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. Systematic review: comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Annals of internal medicine*. 2008;148(6):435-48.
16. McDonald KM, Davies SM, Haberland CA, Geppert JJ, Ku A, Romano PS. Preliminary assessment of pediatric health care quality and patient safety in the United States using readily available administrative data. *Pediatrics*. 2008;122(2):e416-25.
17. Agency for Healthcare Research and Quality. Quality Indicators Rockville, MD: U.S. Department of Health & Human Services,; 2012 [Available from: <http://www.qualityindicators.ahrq.gov/>].
18. National Quality Forum. NQF-endorsed standards Washington, DC: NQF,; 2010 [cited 2013 January 14]. Available from: [http://www.qualityforum.org/Measures\\_List.aspx](http://www.qualityforum.org/Measures_List.aspx).
19. Diamond CC, Rask KJ, Kohler SA. Use of paper medical records versus administrative data for measuring and improving health care quality: are we still searching for a gold standard? *Dis Manag*. 2001;4.3:121-30.
20. Naessens JM, Ruud KL, Tulledge-Scheitel SM, Stroebel RJ, Cabanela RL. Comparison of provider claims data versus medical records review for assessing provision of adult preventive services. *The Journal of ambulatory care management*. 2008;31(2):178-86.
21. Persell SD, Kho AN, Thompson JA, Baker DW. Improving hypertension quality measurement using electronic health records. *Medical care*. 2009;47(4):388-94.
22. Baron RJ. Quality improvement with an electronic health record: achievable, but not automatic. *Annals of internal medicine*. 2007;147(8):549-52.
23. Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. *Journal of the American Medical Informatics Association : JAMIA*. 2007;14(1):10-5.
24. Krupski TL, Bergman J, Kwan L, Litwin MS. Quality of prostate carcinoma care in a statewide public assistance program. *Cancer*. 2005;104(5):985-92.
25. Miller DC, Spencer BA, Ritchey J, Stewart AK, Dunn RL, Sandler HM, et al. Treatment choice and quality of care for men with localized prostate cancer. *Med Care*. 2007;45(5):401-9.
26. Miller DC, Litwin MS, Sanda MG, Montie JE, Dunn RL, Resh J, et al. Use of quality indicators to evaluate the care of patients with localized prostate carcinoma. *Cancer*. 2003;97(6):1428-35.
27. Potosky AL, Davis WW, Hoffman RM, Stanford JL, Stephenson RA, Penson DF, et al. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. *Journal of the National Cancer Institute*. 2004;96(18):1358-67.