**Extraction of patient-centered outcome for prostate cancer patients – Urinary Incontinence**

Phenotype Algorithm Pseudo Code

**Overview**

Description of a weakly supervised machine learning approach for extracting treatment-related side effects (Urinary Incontinence) following prostate cancer therapy from multiple types of free-text clinical narratives, including progress notes, discharge summaries, history and physical notes. Prostatectomy surgery and radiation therapy are our treatments of interest for prostate cancer.

**Cohort generation**

The cohort is defined with the following criteria:

1. EHR of all male patients of 35 years of age or more, AND
2. For which there is an ICD-9-CM / ICD-10-CM diagnosis of prostate cancer, AND
3. For which there are at least two encounters before first treatment, AND
4. For which there is at least one clinical notes before first treatment, AND
5. For which there is either prostatectomy surgery or radiation procedure performed as identified by CPT codes.

**Common terminology used for UI description**

|  |
| --- |
| incontinence |
| incontinent |
| weak incontinence |
| incontinence of urine  |
| leakage |
| any leaks |
| any leak |
| significant problems with lower urinary tract symptoms |
| significant lower urinary tract symptoms |
| significant problems with urinary tract symptoms |
| lower urinary tract symptom |
| lower urinary tract symptoms |

The whole vocabulary learnt can be found here -

|  |
| --- |
| **Urinary\_Incontinence** |
| incontinence  |
| incontinent  |
| weak incontinence  |
| incontinence of urine  |
| leakage  |
| any leaks  |
| any leak  |
| significant problems with lower urinary tract symptoms  |
| significant lower urinary tract symptoms  |
| significant problems with urinary tract symptoms  |
| lower urinary tract symptom  |
| lower urinary tract symptoms  |
| urinary tract symptoms  |
| any voiding issues  |
| urine leak  |
| urinary leakage  |
| urine leaking  |
| urine leakage  |
| urinary incontinence  |
| urinary and fecal incontinence  |
| urinary or fecal incontinence  |
| urge incontinence  |
| stress incontinence  |
| stress urine incontinence  |
| urinary stress incontinence  |
| stress urinary incontinence  |
| stress or urge incontinence  |
| urge or stress incontinence  |
| involuntary urination  |
| unable to hold urine  |
| unable to control bladder  |
| bladder control  |
| bladder incontinence  |
| unable to prevent bladder emptying  |
| loss of bladder control  |
| significant lower urinary tract symptoms  |
| loss of urinary control  |
| pads  |
| pad  |
| diapers  |
| diaper  |
| wears diaper  |
| wear diapers  |
| wears a diaper  |
| wears a pad  |
| wear a diaper  |
| wears diapers  |
| post micturition dribbling  |
| post-micturition dribbling  |
| post void dribbling  |
| post-void dribbling  |
| post-void leakage  |
| post void leakage  |
| postvoid dribbling  |
| dribbles postvoid  |
| dribbles post-void  |
| post-void dribble  |
| post-void dribbles  |
| postvoid dribble  |
| postvoid dribbles  |
| Incontinence: 0  |

**Characterization**

We classify each note as:

(1) Affirmed: symptom present;

(2) Negated: symptom negated;

(3) Discussed Risk: clinician documented a discussion regarding risk of the symptom.

(4) Not discussed – no information present regarding the symtom

**Algorithm**

We extract urinary incontinence (UI) status from a range of clinical notes without considering manually engineered classification rules or large-scale manual annotations. For learning, the method exploits two sources of pre-existing medical knowledge: (1) [domain-specific dictionaries](https://github.com/imonban/UI_Incontinence/blob/master/dic/terms_dictionary.xls) that have been previously developed for implementing a rule-based information extraction system; and (2) publicly available [CLEVER](https://github.com/stamang/CLEVER/blob/master/res/dicts/base/clever_base_terminology.txt) terminology that represents a vocabulary of terms that often present within clinical narratives. We extracted the UI status before (baseline) and after the treatment (3months, 6months, 12months and 24 months) at different time points based of date of encounters of the clinical notes.

**Flow diagram**



1. A weighted neural word embedding is used to generate sentence-level vectors where term weights are computed using term frequency and inverse document frequency (TF-idf) scoring mechanism, with sentence labels derived from a mapping against domain-specific dictionaries combined with CLEVER (weak supervision).
2. These sentence vectors are used to train a machine learning model to determine whether UI were affirmed or negated, and whether the clinician discussed risk with the patient.
3. Finally, we combine the sentence-level annotations using majority voting to assign a unique liable for the entire clinical note.

***Github code***

Our code base can be clone from here ([git@github.com:imonban/UI\_Incontinence.git](file:///Users/boussard/Library/Containers/com.apple.mail/Data/Library/Mail%20Downloads/A254A1F5-3892-4A95-ADC1-FEACBD0428C9/git%40github.com%3Aimonban/UI_Incontinence.git)) which takes a csv file as input with mandatory fields (case- sensitive): 'PAT\_DEID', 'NOTE\_DEID', 'NOTE', 'NOTE\_DATE', and outputs a csv file with note level annotations.