

EXTRACTING CONCEPTS FROM 1.46 BILLION DOCUMENTS

Dario A. Giuse

Department of Biomedical Informatics, Vanderbilt University Medical Center

WORD CLOUD NLP

- Concept extraction: free text -> UMLS CUIs
- Real-time (650 documents/sec on 500 CPUs)
- 725,000 new documents/day

SCOPE

- All historical and new documents in the HDR (Health Data Repository)
- 4.4M patients, with data back to 1995
- For each patient, create list of all concepts documented, with individual documents and relative timestamps

NLP Features

- Concept recognizer: uses 12,800 UMLS CUIs (good SNOMED coverage)
- Which concepts were documented positively for this patient in this document?
- Fully general (not task-specific NLP)
- 16,500 rules (of which 7,500 are manually created)
- Handles negation, uncertainty, Family History, differential diagnosis, literature/population statements
- Uses validation set of 46,000 documents for regression testing (full regression test takes 1 min 48 sec)

IEC Long Term Clinic Visit
Date of Service: 09/2022
CURRENT PROBLEM: DLBCL at Day # 97 from Yescarta at VUMC.
PREVIOUS TREATMENT:
- R-EPOCH-DA x 4 cycles (discontinued therapy due to poor tolerance even at dose level -1) completed in 02/2020
- R-GEMOX CID1 09/2022 X 3 (last on 07/22).
- Pola R 07/22
PREPARATIVE REGIMEN: FluCy; Day 0: 09/2022
BRIEF CART SUMMARY:
[REDACTED] has successfully completed his CAR T-Cell Therapy with Yescarta for DLBCL. Patient's pre-treatment DSAT was Stable Disease. He underwent lymphodepleting chemotherapy with Flu/Cy. Day zero was on 07/22. Patient's transplant course was complicated by CRS and neurotoxicity. CRS was graded based on the ASTCT Consensus Grading. Maximum CRS grading was 1. Interventions used to manage CRS were tocilizumab and dex. ICANS was graded based on the ASTCT Consensus Grading. Maximum ICANS grading was 1. Interventions used to assess and/or manage ICANS were dexamethasone. Repeat disease assessment with PET/CT is scheduled on 09/23.
INTERVAL HISTORY:
Patient presents with wife for IEC long term visit.
None pain in abdomen.
Notes he had a tick bite last week. Picture of leg in media.
Reports taste is still off but is eating okay. Weight stable.
Denies infectious symptoms.
Answers for HPI/ROS submitted by the patient on 09/2022
Trouble Swallowing: No
Cold intolerance: No
Heat intolerance: No
syncope: No
Difficulty breathing during physical activity: No
Decreased appetite: Yes
Night Sweats: No
Unexpected Weight Change: No
Difficulty Urination: No
Excessive urination: No
Swollen lymph nodes: No
Hair Loss: No
Nail changes: No
Lesions: No
Joint Swelling: No
Stiffness: No
numbness: No
confusion: No
Changes in bowel habits: No
Review of Systems
Constitutional: Positive for malaise/fatigue. Negative for chills and fever.
Eyes: Negative for blurred vision and double vision.
Respiratory: Positive for cough and sputum production. Negative for hemoptysis and wheezing.
Cardiovascular: Negative for chest pain, palpitations and leg swelling.
Gastrointestinal: Positive for abdominal pain. Negative for blood in stool and nausea.
Genitourinary: Negative for flank pain and hematuria.
Musculoskeletal: Positive for joint pain.
Skin: Negative for rash.
IEC Long Term Clinic Visit.
CURRENT PROBLEM: diffuse large b-cell lymphoma at Day # 97 from Yescarta at VUMC.
PREVIOUS TREATMENT:
- R-EPOCH-DA x 4 cycle (discontinued therapy due to poor tolerance even at dose level -1) completed in 02/2020
- R-GEMOX CID1 09/2022 X 3 (last on 07/22).
- Pola R 07/22.
PREPARATIVE REGIMEN: FluCy; Day 0: 09/2022.
BRIEF CART SUMMARY:
[REDACTED] has successfully completed pt_CAR T-Cell Therapy with Yescarta for diffuse large b-cell lymphoma. pt_pretreatment DSAT was Stable Disease. Hs_ underwent lymphodepleting chemotherapy with Flu/Cy. Day zero was on 07/22. pt_transplant course was complicated by CRS and neurotoxicity. CRS was graded based on the ASTCT Consensus Grading. Maximum CRS grading was 1. Interventions used to assess and/or manage ICANS were dexamethasone. Repeat disease assessment with PET/CT is scheduled on 09/23.
INTERVENTION HISTORY:
Intervention used to manage CRS were tocilizumab and dex. ICANS was graded based on the ASTCT Consensus Grading. Maximum ICANS grading was 1. Interventions used to assess and/or manage ICANS were dexamethasone. Repeat disease assessment with PET/CT is scheduled on 09/23.
INTERVAL HISTORY:
Patient presents with rel_m for IEC long term visit.
None abdominal pain.
Notes Hs_ had a tick bite last week. Picture of leg in media.
Reports taste is off; but is eating okay. Weight stable.
Answers: Positive:
anorexia: Yes.
REVIEW OF SYSTEMS:
Constitutional: Positive for malaise, Positive for fatigue.
Respiratory: Positive for cough, Positive for sputum production. C0242104
Gastrointestinal: Positive for abdominal pain.
Musculoskeletal: Positive for [joint pain]=>arthralgia. C0003862
Neurologic: Positive for weakness=>muscle weakness.
Psychiatric Behavioral: The patient.
Facility-Administered Medications
Medication Dose Route Frequency Provider Last Rate Last Admin . F-18 fludeoxyglucose (FDG) injection 10_9 millicurie 10_9 millicurie intravenous Once
[REDACTED] APRN.
The following portion of history were Allergies, Current Medications, Past Family History, Past Medical History, Past Social History, Past Surgical History and Problem List.
PHYSICAL EXAM:
Vitals: BP 108/66 (BP Location: arm, Patient Position: Sitting, Cuff size : 11 Adult)
Pulse 68
Temp 36.8 °C (98.2 °F) (Oral)
Resp 16
Ht 171 cm (67.32")
Wt 67.8 kg (149 lb 6.4 oz)
SpO2 93%
BMI 23.18 kg/m²
Performance status (Karnofsky): 70 Cares for self, activities of daily living impaired or to do active work.
Physical Exam
Constitutional:
General: Hs_ .
Appearance: Normal appearance. Hs_ .
HEENT:
Head: Normocephalic.
Ear: .
Ear: .
Eye:
Conjunctiva/sclera: .
Cardiovascular:

AVAILABLE DATA

- All concepts for all patients, with references to individual documents and timestamps
- Pairwise correlations of any two concepts in the entire population
- All data is sent daily to the Microsoft Azure cloud – can be combined with RD/SD, administrative data, etc.
- Data exported to Epic / eStar for decision support:
 - Implanted Cardiac Devices and positive blood cultures
 - Oncology-related radiology incidental findings
 - Suicide prevention
 - Pediatric Epilepsy

Please contact the author at Dario.Giuse@vumc.org